a variable type (double, int and so forth) and more. The present invention is not limited to such an embodiment and other implementations are possible without deviating from the scope and intent of the present invention. The listing, however detailed, is merely illustrative of the data and functions which are used in the equations and methods described herein.

[0020] Consequently, data and functions from this listing, associated with the various entities, will be used in forthcoming equations, flowcharts and methods. The reader is directed to this listing as reference when reading such equations and examining such drawings.

--- start of entity data model detailed listing ---

Model: Untitled 1 (public)

Contains:

client, session, channel, server, contentSelection.

Component: client (public Class/Interface)

Comment:

A client entity stands for a client presently requesting or receiving service.

Methods:

public static **lookup (id: in int) : client**

public **GetId () : const int&**

public **SetId (val : in int&)**

public **GetCostOfService () : const double&**

public **SetCostOfService (val : in double&)**

Attributes:

private **status: client < int >**

Specifies whether or not a client has been allocated a channel and session.

private **id: int**

Integer-valued identifier that is unique to the client (primary key). Can be obtained from a monotonically increasing counter.

private **costOfService: double**

Dollar charge per Mbyte. This value is the same for all customers under flow optimization. Under cost/charge optimization may be an integer value reflective of the rank; the higher the rank the higher the charge.

Has:

public **selected: contentSelection**

public **assigned a: session**

public **employs: channel**

Component: session (public Class/Interface)

Comment:

A session entity holds various state information about the service being received by an associated customer.

public **GetCurrentPosition () : const double&**

public **SetCurrentPosition (val : in double&)**

public **GetPayloadToGo () : const double&**

public **SetPayloadToGo (val : in double&)**

public **GetStatus () : const int&**

public **SetStatus (val : in int&)**

public **GetMinimumFlowRate () : const double&**

public **SetMinimumFlowRate (val : in double&)**

public **GetFlowRateRange () : const double&**

public **SetFlowRateRange (val : in double&)**

public **GetMaxFlowRate () : const double&**

public **SetMaxFlowRate (val : in double&)**

Attributes:

private **playTimeToGo: double**

Indicates the minutes remaining in the viewing experience. Initialized to contentSelectiont.playTime (see below).

private **currentPosition: double**

Pointer into media content from which content is being delivered.

private **payloadToGo: double**

The amount of media content (in Mbytes) as yet undelivered by the server. Does not include any content presently stored in the client-side buffer.

private **status: int**

Indicates whether sesion is active or paused.

private **minimumFlowRate: double**

This is the minimum flow from server to client required to ensure uninterrupted service over the remaining playTime. Has a value of zero if playloadToGo is zero. Given by (payloadToGo * 8)/(playTimeToGo *60)

private **flowRateRange: double**

Specifies the effective range over which the channel content flow serving a session is constained without consideration for interactions with other flows. Equals maxFlowRate - minimumFlowRate

private **maxFlowRate: double**

Effective maximum bound on flow as expressed in formula (8) which must be re-evaluated periodically.

Has:

public **delivers content through: channel**

Component: channel (public Class/Interface)

Comment:

A channel represents the network resources from server to client associated with an ongoing session, encompassing the client-side buffer if any, and its level.

public **GetBufferLevel () : const double&**

public **SetBufferLevel (val : in double&)**

public **GetFlowRate () : const double&**

public **SetFlowRate (val : in double&)**

public **GetMaxFlowRate () : const double&**

public **SetMaxFlowRate (val : in double&)**

Attributes:

private **bufferSize: double**

Capacity of the client-side buffer (or equivalent)

private **bufferLevel: double**

Current buffer level in MBytes of stored content.

private **flowRate: double**

Flow rate through channel specified by the relevant optimizing flow modulator.

private **maxFlowRate: double**

This value represents the maximum possible flow rate from the server to an individual client over its "channel". This value reflects restrictions on flow that pertain to an individual client. It may be determined by factors such as the bandwidth of client's link to the network, or a limit imposed administratively to ensure balanced network utilization.

Component: server (public Class/Interface)

Comment:

Entity representing the media server and its CAC and flow modulation activities.

public **GetFlowRate () : const double&**

public **SetFlowRate (val : in double&)**

public **GetMaxMinFlowRate[] () : const double&**

public **SetMaxMinFlowRate[] (val : in double&)**

Attributes:

private **maxFlowRate: double**

Maximum possible content flow that is allocated to the server by the network.

private **flowRate: double**

Aggregate content flow rate, summed over all sessions and their associated channels.

private **cac_flowSafetyMargin: double**

Tunable safety margin used by the CAC algorithm to protect sessions-in-progress from being affected by changes in available network bandwidth.

private **maxMinFlowRate[]: double**

Applies when N rate tariffs exist. This array holds the maximum floor level for each category of service. The value for the costliest category N is stored in maxMinFlowRate[N-1], and for the least costliest in maxMinFlowRate[0]. It is the relative magnitude of these ascending values that matters, not their absolute value. Thus the actual maximum floor flow rate for category k is given by server.maxFlowRate * (server.maxMinFlowRate[k-1] /server.maxMinFlowRate[N-1]). Similarly, the maximum floor flow rate for category N is server.maxFlowRate.

Has:

public **contains: contentSelection**

public **modulates: channel**

Component: contentSelection (public Class/Interface)

Comment:

Entity represents a video/sound clip or other bounded unit of content. A continuous data feed does not qualify.

Attributes:

private **averagePlayRate: double**

The average rate at which media content is consumed by the client, as computed by dividing the (payload * 8) by the (playTime * 60)

private **playTime: double**

Duration of play of the media content in minutes.

private **payLoad: double**

total size of the content in Mbytes.

Has:

public **is associated with: session**

--- end of entity data model detailed listing ---

The following table summarizes the highlights of the previous detailed description of each entity in Figure 2.

TABLE 1

Entity	**Description**
client 200	Each client is denoted by an associated unique integer index $_{Id}$. The set of active clients is denoted by $\mathbf{S}_{ActiveClients}$. The set of deferred clients is denoted by $\mathbf{S}_{QdClients}$. Incoming clients are expected to select the content they wish to view prior to being queued for dispatch by the CAC sub-system, which requires knowledge of the client's bandwidth requirements, duration of play, and cost of service, all of which may vary according to the selection.

server 220	Servers sit astride a network and can deliver media content through the network to their clients up to a designated maximum flow rate. The server is responsible for accepting or rejecting clients, launching sessions and associated channels for the former, and modulating content flows over all channels in an optimal manner.
channel 210	A channel represents the data path between the server and the client. The channel buffer is typically located near or within the clients viewing station. The flow of content through the channel is set by the flow modulator sub-system.
contentSelection 230	A server will typically act as a repository for media content, which it can deliver to clients upon demand. For our purposes, media content is characterized by its payload and the play duration, which together imply the ***averagePlayRate*** = (payload*8)/(playTime *60). The averagePlayRate is none other than the streaming rate imposed by real-time just-in-time streaming algorithms.
session 240	Every session represents an instance of media content delivery to an associated client over a designated channel. The ***playTimeToGo*** indicates the time remaining before the content is fully played out to the client. The ***payloadToGo*** is the amount of content data as yet undelivered to the channel. A session terminates when this value reaches zero, at which time ***playTimeToGo*** may still be large, according to the capacity, the level of the channel buffer, and the media play rate.

Constraints On Content Flow

[0021] Before referring to more Figures, it is imperative to establish some formulas and problem statements which are used in the methods which follow.

[0022] The flow of content between entities is subject to the following constraints at all times. Buffer levels are always expressed in Mbytes and data rates in Mbits/sec.

(1)$\Sigma_{\text{i } \epsilon \text{ SactiveClients}}$ (client.lookup(i).channel.flowRate) < = server.maxFlowRate
The sum of all channel flows cannot exceed the imposed maximum throughput capacity of the server.

(2)client.lookup(i).channel.flowRate < = client.lookup(i).channel.maxFlowRate for all $\mathbf{i} \epsilon \Sigma_{\text{activeClients}}$

The data path from server to client is subject to its own constriction.

(3)client.lookup(i).channel.flowRate < =(client.lookup(i).channel.bufferSize - client.lookup(i).channel.bufferLevel)* 8/60 + client.lookup(i).session.mediaContent.averagePlayRate, for all $\mathbf{i} \epsilon \Sigma_{\text{activeClients,}}$

The channel buffer is never allowed to overflow.

(4)client.lookup(i).channel.flowRate < = client.lookup(i).session.payloadToGo * 8/60

for all $\mathbf{i} \in \Sigma_{activeClients,}$

Content that does not exist cannot be delivered. (Constraint 1 will ordinarily prevail except at the very end of a session.)

[0023] The constraints listed above are straightforward applications of common sense in relation to the flow of data through constricted channels, out of finite data sources, and into and out of bounded buffers. By contrast, the following constraint, which imposes a minimum channel flow rate instead of a maximum, is less obvious. The minimum value, termed the ***minFlowRate*** is set to the flow rate which, if sustained over the balance of the play time to go (***playTimeToGo***), ensures that all required content will be available when needed -- and no sooner -- until all content is played out. This floor value can be calculated for $\mathbf{i} \in \Sigma_{activeClients}$ by the formula:

(5)client.lookup(i).session minFlowRate =(client.lookup(i).session.payloadToGo * 8)/ (client.lookup(i).session.playTimeToGo * 60)

Accordingly, the flow rate on the right-hand-side is termed the just-in-time (JIT) flow rate (f^{JIT}).

Thus:

(6)client.lookup(i).channel.flowRate >= client.lookup(i).session.minFlowRate
for all $\mathbf{i} \in \Sigma_{activeClients}$

[0024] The variable constraint bounds (i.e. the values to the right of the inequality symbol) of equations 1 – 4 and 6 are re-evaluated on a periodic basis (e.g. once per second) prior to the execution of the CAC procedure and optimizer. In

particular, the ***minFlowRate*** value starts out at the beginning of a session equal to the streaming rate. By construction the ***minFlowRate*** rate never exceeds this initial value so long as constraint 6 is honored. In fact, constraint 5 implies that the ***minflowRate*** value must be a diminishing function of time that may hold its value for a time but never rises. As seen from equation 6, the actual data rate of the channel, ***flowRate***, is always greater than or equal to the ***minFlowRate***. By design, and virtue of the fact the present invention uses faster-than-realtime transmissions, the system quickly gets ahead of itself and ensures that after initial conditions, the ***minFlowRate*** is always equal to or less than the real-time rate and that it continues to decrease. As we shall see, the CAC procedure exploits this monotonic characteristic of the minimum flow rate over time.

[0025] Constraints 2, 3 and 4 are of like kind, each specifying an upper bound on individual channel flows. Whereas the bound for constraint 2 is typically a constant, the bounds on 3 and 4 will vary over time. Nevertheless, only one of the three bounds is effective at any given time, namely the one with the smallest bound value, given by:

(7) client.lookup(i).session.maxFlowRate = minimum of

1) client.lookup(i).channel.maxFlowRate,

2) (client.lookup(i).channel.bufferSize - client.lookup(i).channel.bufferLevel)*

8/60 + client.lookup(i).session.mediaContent.averagePlayRate,

3) client.lookup(i).session.payloadToGo * 8/60

Consequently, formulas 2, 3, and 4 can be consolidated into a single constraint, the bound for which is computed at every scan to be the smallest bound of associated constraints 2, 3 and 4.

(8) client.lookup(i).channel.flowRate < = client.lookup(i).session.maxFlowRate,

whereby for all $i \in \Sigma_{activeClients}$, maxflowRate is given by equation (7).

At any one time, individual channel flows are constrained over a range, as follows:

(9) client.lookup(i).session.flowRateRange =
client.lookup(i).session.maxFlowRate - client.lookup(i).session.minimumFlowRate

Value Functions

[0026] The value functions introduced previously can be expressed mathematically as linear functions of channel flows, as follows:

Optimizing Throughput (Maximal Flow)

(10) value $= \Sigma_{i \in SactiveClients}$ client.lookup(i).channel.flowRate)

Optimizing Charges (Maximal Charges)

(11) value $= \Sigma_{i \in SactiveClients}$ (client.lookup(i).channel.flowRate * client.lookup(i).costOfService)

Optimization Problem Statement (Figure 3)

[0027] The optimization problem, which in one embodiment is strictly throughput and in another case is charge, can be stated simply as follows:

Find values for

client.lookup(i).channel.flowRate for all $i \in \Sigma_{activeClients}$

constrained by inequalities 1 through 5, such that the value obtained by evaluating expression 10 or 11 assumes a maximum.

[0028] Both of these problem formulations are examples of Linear Programming for which a number of well-known and generally effective computational solutions exist. In linear programming one seeks to optimize a linear cost function of variable **x**

(12) $c*x = c_1*x_1 + \ldots + c_n*x_n$

subject to a set of linear inequality constraints

(13) $A*x <= b$

where $x^T = (x_1,\ldots,x_n)$, $c = (c_1,\ldots,c_n)$ are the state variable & cost vectors, A is an n-by-m matrix, $b^T = (b_i,\ldots b_m)$ is the constraint vector, and the operator '*' stands for matrix or scalar multiplication.

[0029] Figure 3 is introduced as illustrative of the problem statement and the general methods of the prior art, and is not incorporated as an element of the invention.

[0030] The linear programming problem as well as its solution can best be understood with the aid of geometry. Figure 3, depicting a 2-dimensional Cartesian problem space, inequality constraints (13) define a convex hull H 310 over which a search for an optimum value of $\mathbf{x} = (x_1,x_2)$ is permitted to range. The cost vector **c** 350 defines an infinite family of equal cost lines (hyper-planes) which lie orthogonal to **c**. Three examples of such lines are shown in L_1 360, L_2 365, and L_3 370., each of progressively higher value. The supreme value of the cost function is obtained by sliding along **c** 350 till one can go no further, in this instance toward vertex V_4 340 of hull H 310. Many well-known methods (e.g. the Simplex Method) work roughly in this fashion, exploiting the fact that at least one optimum point must be at a vertex. In particular, the Simplex method algorithm begins by finding a vertex (e.g. V_2 320), and then moves along a sequence of vertices (e.g. V_3 330, V_4 340) improving itself each time until no further improvement is possible & the summit is reached.

[0031] Let us suppose instead that V_3 330 were placed along L_3 370 along with V_4 340. According to prior art methods, V_3 330 and V_4 340 are the two possible solutions, but the equally valuable points in between them are not. As we shall soon see, the problem of throughput optimization (6) falls in this category.

[0032] While vertex V_1 300 does not factor into this description, it is depicted in Figure 3 for completeness.

Flow Modulation

Methods for Maximal Flow

The following sections detail two embodiments to optimize total data flow.

Overview (Figure 4-a)

[0033] Figure 4-a depicts a scenario involving two flows. The convex hull is in this instance bounded by line segments L1, L2, L3, L4 and L5. L6 is a boundary used in a different embodiment, however the present embodiment uses L5 and not L6. Flow f_2 can range over the interval separating line segments L1 from L3, namely f_2^{MIN} and f_2^{MAX}; the range is depicted as f_2^{RANGE}. Flow f_1 can range over the interval between lines L2 and L4, namely f_1^{MIN} and f_1^{MAX}, and depicted as f_1^{RANGE}. The sum of flows f_1 and f_2 is constrained to lie inside of line segment L5 which, by construction, is always orthogonal to the cost vector $\mathbf{C_f}$. Cost vector $\mathbf{C_c}$ is also illustrated but used in a distinct embodiment. In the present embodiment only $\mathbf{C_f}$ is used. In the illustrated example of the present embodiment, the constraint on total flow is set to 5, and is therefore low enough to cause L5 to intersect L3 and L4. This would not have been true had the value chosen had been 10 instead of 5. With L5 safely out of contention, the convex hull would instead be a simple rectangle bounded by L1 through L4, thereby permitting both flows to assume their respective maxima without interference. In practice operational constraints exist intrinsically or are imposed from the outside so as to ensure cost effective sharing of potentially costly network resources.

[0034] Supposing Figure 4 to be correct, the well-known methods would select vertex $V_{3\text{-}5}$, which lies at the intersection of L3 and L5, or $V_{4\text{-}5}$, which lies at the intersection of L4 and L5. These solutions, though optimal, are undesirable for the present invention as they fail to spread available bandwidth over all channels as fairly as would a centrally located interior point of L5. For this reason, two alternative optimization methods are taught, which are adapted to the particular needs of this problem and ensure a fairer allocation of constrained bandwidth among all channels.

Iterative Procedure (Figure 5)

[0035] In order to optimize use of all available bandwidth, the following general method is used, with the details illustrated in Figure 5. This method is a solution for the problem illustrated in Figure 4-a, which geometrically illustrates the optimization problem in the limited case of two flows, f1 and f2. The following description expands the problem to an arbitrary number of clients (and therefore flows) and presents a method for solving this optimization problem.

[0036] Referring to Figure 5, in step 500, values are calculated for the ***session maxFlowRate*** and ***session.minFlowRate*** for each client as per the minimum and maximum constraint bound expressions in equations 6 and 8, respectively. This step correlates to the determination of $f_1$min, $f_1$max, $f_2$min and $f_2$max from Figure 4-a.

[0037] The difference between these two yields the ***session.flowRateRange*** of each client. Thus:

session.flowRateRange = session maxFlowRate- session.minimumFlowRate

[0038] In step 505, the active clients are sorted in an ascending fashion based upon their **session.*flowRateRange***. As will be shown this critical step facilitates allocation of the remaining server bandwidth as evenly as possible among all active channels, thus maximizing the number of channels that benefit by use of the total server bandwidth. An arbitrary assignment of remaining bandwidth is likely to saturate the server before all channels have been assigned extra bandwidth, thereby favoring certain channels on an ad-hoc basis. This step correlates to keeping the solution in bounds of the space delinated by line segments L1-L5 of Figure 4-a.

[0039] In step 510, each client's channel flow rate is set to the ***session mimimumFlowRate***.

[0040] By doing so, it is ensured that the minimum flow constraint is met for each session and that the minimum flow rate is a non-increasing function of time, which is critical to the proper functioning of the CAC procedure. This portion of the process also ensures that the solution, starting from the vertex in Figure 4-a as defined by the intersection of L1 and L2, moves generally in the direction of vector c_f . All clients are also marked as unprocessed.

[0041] In the next step, 520, ***server.flowRate*** is set to the sum of each active client's **session.flowRate.**

[0042] Next, the following is iterated over all clients in sorted sequence (during any given iteration the selected client is given by its ***id***) by performing steps 530 through 570. In step 530 evaluating the following expressions test for possible server saturation:

delta = (server.maxFlowRate - server.flowRate) / (qty of un-processed clients)
range = client.lookup(***id***).session.maxFlowRate -
client.lookup(***id***).session.flowRate

[0043] If ***range*** is greater then ***delta***, this implies that the server can be saturated in this iteration by allocating ***delta*** to all unprocessed clients (step 540).

[0044] On the other hand, the 'no' path for step 530 implies that the server is not saturated and that the present client (given by ***id***) will saturate first. Accordingly, in 550 the ***delta*** variable is set as follows:

delta = range

[0045] To again correlate this process back to the geometry of Figure 4-a, server saturation is indicated when the solution which is being sought in the direction of vector c_f goes beyond line segment L5.

[0046] Next, the flow rate is incremented for all unprocessed clients by ***delta***, causing client ***id*** to saturate.

[0047] In step 560 the server flow rate is adjusted accordingly:

server.flowRate = server.flowRate + delta * (qty of unprocessed clients)

In step 570 the client given by ***id***, now saturated, is marked as processed.

Algebraic Procedure (Figures 4-b and 15) By Interpolation

[0048] Referring to Figure 4-b, the iterative method described in the previous section begins its search for a maximum from vertex V_{min} along the direction shown by line C_1. In contrast, the present method follows the diagonal line segment L_{diag} of the rectangle bounded by lines L1 through L4, starting at vertex V_{min} and ending at point $V_{3\text{-}4}$.

[0049] Provided it exists, the intersection between L_{diag} and L5, indicated by V_f, is optimal on the same basis as any other point lying along the intersection of L5 with the rectangle bounded by L1 through L4, as previously discussed.

[0050] Whenever L5 does not intersect this rectangle, the optimal solution is given by vertex V3-4 at which f1 and f2 both assume their respective maxima.

[0051] In the first instance, the coordinates for point V_f can be obtained by elementary vector geometry and algebraic manipulation, as follows:

First, we define a number of abbreviations,

f_k^{min} represents client.lookup(k).session.minimum flow rate

f_k^{range} represents client.lookup(k).session.flow rate range

f_{svf}^{max} represents server.maxflowrate

We seek scale factor x such that the vector sum of $(f_1^{min} \ldots f_n^{min}) + x\ (f_1^{range} \ldots f_n^{range})$ intersects an equi-cost hyperplace for capacity c. The point of intersection is obtained by solving equation:

$$(f_1^{min} + \ldots + f_n^{min}) + x(f_1^{range} + \ldots + f_n^{range}) = c$$

Solving for x, we obtain:

$$x = [f_{svr}^{max} - (f_1^{min} + \ldots + f_n^{min})]/(f_1^{range} + \ldots + f_n^{range})$$

where $f_{svr}^{max} - (f_1^{min} + \ldots + f_n^{min})$ represents the unused bandwidth beyond the minimum allocation for each session. For any given session, the optimal flow rate is given by:

$$f_i = f_i^{min} + x\, f_i^{range}, \text{ if } x < 1,$$

or

$$f_i = f_i^{min} + f_i^{range}, \text{ if } x > 1$$

[0052] This solution by interpolation is interesting by virtue of the great efficiency of the calculation, and the fact that no session saturates ($x >= 1$) unless all sessions saturate. In contrast, the previous method entails considerably more computational effort, and tends to saturate sessions with the lowest reserve capacity.

[0053] A block diagram for this algorithm is depicted in Figure 15. Block 1500 establishes the maxim and minima for each flow. For each flow the minimum is subtracted from the maximum to obtain the range. Available Bandwidth is calculated in block 1520, as the difference between the aggregate flow capacity and the sum of flow minima. Factor x is calculated next in block 1540, and tested against 1 in block 1650. Flow rates for each session are computed in block 1560 if x is less than 1, or 1570 in the contrary case.

A Method for Maximal Charge

[0054] The following sections detail one embodiment to optimize the total monetary charges within the system. The second method is algebraic in nature, and thus very efficient.

Overview (Figure 4-a)

[0055] Referring back to Figure 4-a, cost vector C_c lies orthogonal to line L6, which intersects the convex hull at the vertex formed by the intersection of lines L4 and L5, namely $V_{4\text{-}5}$. This cost vector, and the optimal point that it implies, favors flow f1 over flow f2 . In this example, as the cost of service for f1 equals 2, thus exceeding the cost of service of 1 set for f2 . As the number of flows grow to exceed the number of distinct categories of service (and associated costs of service) the unique optimal solution, depicted in Figure 4 for the case where every flow has a distinct cost of service, no longer applies. Once again a plurality of flows within a service category vie for bandwidth which a method should endeavor to distribute evenly. This method is derived from the previous one, and optimizes one cost category after another, starting with the most costly and ending with the least costly, or when all available capacity is allocated.

An Iterative Search Procedure (Figure 6)

[0056] Let the service categories be denoted by k= 1...N, where k also denotes the cost of service.

[0057] Let $C_1..C_N$ be the partition of $S_{activeClients}$ that places all clients with cost of service k in set C_k. Partition sets C_k can be ordered to form sequence SeqC = $C_N...C_1$.

[0058] Figure 6 depicts the method for implementing the method to maximize the cost of service (service charge).

[0059] This method is nearly identical to the iterative procedure used to maximize flow. The principle difference stems from the partitioning of clients according to their category (cost) of service: clients charged most are allocated bandwidth preferentially. This is accomplished by adding another level of iteration around the method of Figure 5. The inner iteration (steps 650 through 680) functions exactly as before, with the difference that its actions are limited to the clients belonging to the given service category ***k*** (i.e. C_k). This difference also holds true of step 640 which sorts category ***k*** clients according to their flow ranges prior to entry in the bandwidth-allocating inner loop. The outer loop proceeds down a sorted sequence of service categories SeqC (generated in step 630), starting with the category generating the greatest revenue to the service provider. Given a fairly static set of service categories, this sort need be performed only when the categories undergo change. Steps 670, 675 and 680 are identical to their counterparts in the method of Figure 5 (i.e. 570, 575 and 580).

[0060] The net effect of this method is preferential allocation of bandwidth according to category of service, and equitable treatment of clients within the same category of service.

An Iterative Algebraic Procedure (Figure 16)

[0061] The algebraic method used to maximize bandwidth (Figures 4b and 15) can be viewed as a special case of cost optimization in which all costs are equal. By construction, all sessions belonging to a given category, starting with the most expensive category, and ending with the least, and updating available bandwidth after each step in sequence, we obtain an iterative algebraic method that scales much better than the preceding method.

[0062] A block diagram for this algorithm is depicted in Figure 16. Block 1600 establishes the maxima and minima for each flow. For each flow, the minimum is subtracted from the maximum to obtain the range. In block 1610, every session flow is assigned the minimum allowed value, which value will apply by default should the iterative procedure run out of bandwidth before all categories can be considered. Block 1620 computes the initial value for available bandwidth, as the difference between the aggregate flow capacity and the sum of flow minima.

[0063] Flow values are obtained for each cost category in order of decreasing cost within blocks 1640 through 1675. Factor x is calculated next in block 1540, and tested against 1 in block 1650. Flow rates for each sessions of the category under consideration are computed in block 1660 if x is less than 1, or 1670 in the contrary case. In the former case there is no more bandwidth to allocated to the sessions in the category under consideration.

Call Acceptance Control (CAC)

CAC for Maximal Flow Overview (Figures 7-8)

[0064] The CAC procedure applicable to this flow optimization relies on the essential step of accepting a new client if and only if the added load induced thereby does not compromise service to existing clients or the new one. This critical step could not be accomplished without the close integration with previously- described flow-modulation methods of Figures 5, 6, 15 and 16.

[0065] According to the previous discussion, the minimum flow rate is the minimum sustained flow rate that guarantees that the associated viewer will not be subject to interruptions in service due to a shortfall of content from the media server. It follows that whenever data is being delivered at a rate in excess of the minimum flow rate, a downward adjustment toward the minimum level could be accommodated as needed to surrender bandwidth to any newcomer.

[0066] Figure 7 depicts content flow over a channel for the course of a typical session, and also how data is delivered under real-time streaming D. The amount of content delivered is the same in either case, but the manner of delivery differs considerably. A session is launched at time 0 as the network is lightly loaded, and the optimizer sets an accordingly high flow rate. Another client emerges at the end of interval 700, causing a downward adjustment to the flow rate over interval B, as available bandwidth is shared between two sessions. During both of these intervals the minimum flow rate 720 drops quickly, as data accumulates in the client's media buffers. At the end of interval B, a massive influx of clients necessitates that flow be dropped to the minimum flow rate, which now lies substantially below the streaming rate D and is held until all data is delivered at the end of interval C. Note that the minimum flow rate, shown as element 720, diminishes monotonically over time.

[0067] The ***server swing capacity*** is defined as the difference between the maximum capacity of the server and the total minimum flow rates for all active clients. Therefore:

(14) swingCapacity =

server.maxFlowRate - $\Sigma_{\text{i } \epsilon \text{ SactiveClients}}$ (client.lookup(i).session.minFlowRate)

[0068] Given the monotonic decreasing nature of session minimum flow rates, server swing capacity can readily be seen to be a *monotonic increasing* function of time over the intervals separating client admissions, at which point it undergoes a drop as a new load is taken on. This all-important characteristic implies the following:

Any client admitted for service based on the present value of swing capacity is guaranteed to have sufficient bandwidth at its disposal over the entire future course of the session.

[0069] Figure 8 depicts the server swing capacity 800 over the course of the sessions illustrated in Figure 7. Swing capacity rises quickly over intervals A & B as data is delivered at high flow rates over the network. It holds steady over interval C when all channels flow at their minimum rate then jumps at the end of C before resuming its monotonic rise once again.

Procedure (Figure 9)

[0070] In this procedure, which executes on a periodic basis, queued clients awaiting bandwidth are scanned in FIFO order. For each one, the required bandwidth is computed as per the client's prior content selection. If the available swing capacity (reduced by a safety margin) exceeds the amount required, then the client is activated and swing capacity is adjusted accordingly. Otherwise, two possible cases are considered: 1) under the *FirstFit* embodiment, the procedure

continues scanning clients to the end of the queue, activating clients whose requirements can be met; 2) under the *FIFO* embodiment, the procedure ends with the first candidate client whose requirements cannot be met.

[0071] In step 900, available server swing capacity is evaluated according to the formula

swingCapacity =

server.maxFlowRate - $\Sigma_{\text{i} \in \text{SactiveClients}}$ (client.get(i).session.minimumFlowRate)

The bandwidth requirement for client ***id*** in Step 920 is obtained as follows:

required_bandwidth = client.lookup(***id***).contentSelection.averagePlayRate

The predicate evaluated in Step 940 is given by the expression:

(required_bandwidth < = swingCapacity - server.cac_flowSafetyMargin)

[0072] In step 950, client activation entails allocation of a session and a channel, and insertion in the set of active clients eligible for bandwidth allocation by the optimal flow modulator.

[0073] In step 960 the swing capacity is diminished by the amount reserved for the activated client:

swingCapacity = swingCapacity - required_bandwidth;

Responding to Variations in Network Capacity (Maximal Flow)

[0074] In the CAC procedure for maximal flow, a safety margin was introduced, namely *server.cac_flowSafetyMargin*, to provide the means for ensuring that the server's swing capacity will never fall below a minimal threshold value.

[0075] According to this procedure, the following inequality always holds true:
(15) swingCapacity > = server.cac_flowSafetyMargin

[0076] In the previous discussion, a server's swing capacity provided the basis for determining whether or not a prospective client should be allocated bandwidth.

[0077] By holding *server.cac_flowSafetyMargin* in reserve, the CAC algorithm forces delivery of content at faster than real-time rates among accepted clients, even under the heaviest load. The net effect is to apply upward pressure on client side buffer levels, thus promoting continuous accumulation of content and a jitter-free viewing experience once a session is under way.

[0078] In another embodiment, server swing capacity can also be interpreted as specifying the *maximum* amount by which the *server.maxFlowRate* constraint can be dropped without affecting service, should such an adjustment prove necessary due, for instance, to an influx of exogenous network traffic that diminishes the amount available for multi-media services. Parameter *server.cac_flowSafetyMargin* can thus be set so as to *guarantee a minimum capacity to tighten the constraint on maximum server flow* in response to unexpected load changes that affect the server's ability to service its existing clients as well as new ones.

Anticipating Scheduled Variations in Network Capacity (Maximal Flow)

Overview (Figure 10)

[0079] Figure 10 depicts how the constraint on maximum flow might be allowed to vary according to the time of day, day of the week, and so forth, in expectation of time-varying traffic flow levels extrapolated from past experience, traffic flow models, etc. Maximum flow rate 1000 can be seen to vary based upon the time of

day. In practice, defining the right-hand-side of inequality constraint 1 as a time-dependent expression can impose such time-varying capacities. According to the previous description, the optimizer, which executes on a periodic basis, will automatically seek new flow levels for every active session as the constraint varies. There is, however, no guarantee that an acceptable operating point will be found for all sessions (i.e. one that respects the minimal and maximum constraints on session channel flow). One such example is the case where the server is loaded to the limit and total capacity is curtailed in excess of the aforementioned safety margin. Should such a situation arise the only recourse may well be the termination of a number of established sessions (i.e. load shedding).

[0080] The goal is to eliminate service disruptions of this sort by allowing the CAC procedure to look ahead into the future, and accept new clients only if these can be accommodated without any compromise in service in the midst of *previously anticipated changes* in available network capacity. The following CAC procedure generalizes the previous one: before accepting a client, the test on swing capacity is repeated over a sequence of time segments that cover the proposed viewing period.

Definitions -

Let:

(16) t_end(i) = client.lookup(i).session.playTimeToGo + t_now

Let server.maxFlowRate(t) be server flow capacity as a function of time, as exemplified in Figure 10.

[0081] Let $Seq_T(t\_now)$ = advancing sequence of future times, lead by t_now, when server.maxFlowRate(t) undergoes a step change. For instance, at 9:15 in Figure 10 this sequence reads as follows: 9:15, 9:30, 11:30, 13:30, 6:30, 7:30.

[0082] The server swing capacity at a future time t is computed according to the capacity and worst-case client flows at time t.

(17) swingCapacity(t) = server.maxFlowRate(t) -

$$\Sigma_{t\_end(i) > t} \text{(client.lookup(i).session.minFlowRate)}$$

[0083] It is noteworthy that the worst-case client flows at time t are expressed in terms of the present minimum flow rates, which cannot increase over time, but might hold steady. Finally, a predicate is defined that tests whether a prospective customer will cause swing capacity to be exceeded at some time t, as follows:

```
(18) boolean client_fits(i ,t) {

        if(client.lookup(i).contentSelection.averagePlayRate < =

              swingCapacity(t) - server.cac_flowSafetyMargin)

        return true;

    else return false;
      }
```

Procedure (Figure 11)

[0084] This procedure is an adaptation of the first, which has been extended to consider swing capacity at times in the future when capacity undergoes scheduled changes. Before accepting a client, its minimal bandwidth requirement (which by construction of the flow modulator will never increase over time) is checked against *projected swing capacity* at points in time when total available capacity undergoes scheduled step changes, provided these times fall within the proposed content viewing period. A candidate is activated only if all tests succeed.

[0085] Step 1100 builds a sequence of time values (SeqT) at which step capacity changes are scheduled to occur. The first element of this sequence is t_now, representing the present.

[0086] Beyond step 1100 the queue of waiting clients is scanned in FIFO order, yielding a candidate designated by ***id*** at each iteration.

[0087] The bandwidth requirement for client ***id*** in Step 1120 is obtained as follows:

required_bandwidth = client.lookup(***id***).contentSelection.averagePlayRate

[0088] The worst-case end time for content flow to ***id*** is obtained according to the content selected, as follows:

t_end = t_now + client.lookup(***id***).selected.playTime

[0089] Steps 1130 through 1150 are executed within an iteration for each time point t in SeqT falling between t_now and t_end. This iteration is ended in step 1130 if t exceeds the time window of interest, or in step 1150 if the supply of scheduled capacity changes is exhausted.

[0090] For each time value, step 1140 compares required bandwidth to projected swing capacity.

Projected swing capacity at time t is:

$$\text{swingCapacity}(t) = \text{server.maxFlowRate}(t) - \sum_{t\_end(i) > t} (\text{client.lookup}(i).\text{session.minimumFlowRate})$$

[0091] Note that only active clients whose t_end times occur after t are considered in the sum of minimum flow rates.

[0092] The predicate expression used in step 1140 at time t is thus:

(required_bandwidth < = swingCapacity(t) - server.cac_flowSafetyMargin)

Step 1160 performs the same actions as step 660 in the previous CAC flowchart

[0093] The first CAC process is a special case of the present one, in which the set of step change times to server.maxFlowRate is empty (i.e. server.maxFlowRate is constant), and $Seq_T(t\_now) == t\_now$.

[0094] In preparing Seq_T(t_now), one need only consider future times that will pass before the longest possible content is played out if started at t_now. In order to sidestep problems associated with rollover (at midnight, year 2000, etc.), time is best expressed as a monotonically increasing value (e.g. seconds since Jan 1 1990).

CAC for Maximal Charges

Overview (Figure 13)

[0095] Previously, a method for flow modulation was presented that maximizes charges to clients with active sessions. The CAC embodiment presented previously was not sufficient as it does not consider the cost of service as a basis for connection acceptance. As a result, it may turn away higher paying customers while granting service to lower paying ones, thereby defeating the purpose for this particular embodiment. Therefore, another embodiment is defined which offers the following features:

1. Awaiting clients are serviced in order of their respective service categories, higher paying clients first.

2. Once accepted, a client is guaranteed to receive acceptable service irrespective of its service category.

3. Under heavy load conditions higher paying customers are more likely to be accepted than lower paying ones.

4. Lower paying customers will no be starved for service when higher paying ones enjoy a surplus.

5. Available bandwidth is not thrown away needlessly while clients are being denied service.

[0096] The first objective is easily met by dividing the client queue into as many bands as there are service categories, resulting in a banded queue. Bands are ordered within the queue according to their service categories, with the costliest category in front. As prospective clients arrive and make their selection they are placed in their respective queue band according to their service category (which may be set contractually, according to content selection, etc.).

[0097] Our second objective is met by employing a procedure patterned after those presented previously & offering the same guarantee. Toward our third and fourth objectives we propose dividing total available bandwidth in as many strata as there are service categories. As depicted in Figure 12, two service categories are shown, Premium and Basic, each entailing an associated cost of service. A prospective client is accepted only if there is sufficient swing capacity available within its given service category. The swing capacity for a given category is given by the smaller of 1) the difference between its maximum floor flow rate (corresponding to the top of the stratum for the service category) and the sum of the minimum rates of all active sessions in its category or below, and 2) available swing capacity overall. Finally, our fifth objective is met by allowing the flow optimizer to function freely subject to its operational constraints. The imposed ceilings on call acceptance by category relate to *minimum flow rates*, which merely impose a floor on *actual flow rates*. For example, basic clients might well consume all available bandwidth 300 in the absence of any premium customers, yet could be throttled back toward their floor flow rates (which together cannot exceed 200 in this example) at any time should any premium customer suddenly demand service. In contrast, premium customers could consume the entire 300

bandwidth. As lower paying customers appear these would be admitted to the degree that their quota on minimum flow is not exceeded (i.e. 200) and the availability of swing capacity on the system.

Procedure (Figure 13)

[0098] The present procedure requires a number of ancillary definitions, which follow:

Let the service categories be denoted by k= 1...N, where k also denotes the cost of service.

Let server.maxMinFlowRate[k-1] be the top of the stratum for service category k. Note that server.maxMinFlowRate [N-1] = server.maxFlowRate.

Let S_K be the set of active client indices with a service category *equal to or less than* k. Note that S1 is contained in S2, S2 is contained in S3, and so forth, and that $S_N = S_{activeClients}$.

Let swingCapacity(k) denote available swing capacity for service category k. By construction:

(19) swingCapacity(k) = minimum of:
(server. maxMinFlowRate [k-1] - $\Sigma_{i \in Sk}$(client.lookup(i).session.minFlowRate)),
(server.maxFlowRate - $\Sigma_{i \in SactiveClients}$ (client.lookup(i).session.minFlowRate)))

[0099] Referring to Figure 13, this method is used for CAC when multiple rate tariffs are in effect, and there is a desire to maximize economic returns to the service provider while offering acceptable service to all.

[00100] All waiting clients are scanned in FIFO sequence. The actions taken in Steps 1320 and 1360 are identical to those described in connection with earlier CAC flowcharts.

[00101] Step 1340 evaluates a predicate expression that tests whether the required bandwidth can be accommodated without exceeding the lesser of 1) swing capacity available to the client's category of service, and 2) total available swing across all categories of service. The latter factor could be determinative if all available bandwidth were allocated to high paying customers, leaving lower paying ones such as the proposed client unable to draw from their unfilled quota.

Let us suppose that candidate client ***id*** belongs to rate category ***k***.

We define the swing capacity available in rate category ***k*** as:

swingCapacity(***k***) = least of:

(server.maxMinFlowRate[k-1] - $\Sigma_{\text{i in Sk}}$ (client.lookup(i).session.minimumFlowRate))

and

(server.maxFlowRate - $\Sigma_{\text{i} \in \text{SactiveClients}}$ (client.lookup(i).session.minimumFlowRate))

The predicate expression invoked by step 1340 can now be written as follows:

(required_bandwidth < = swingCapacity(***k***) - server.cac_flowSafetyMargin)

[00102] This algorithm processes queued clients in band sequence, and within every band in FIFO order. If the predicate evaluates to true, the client is activated. Otherwise two possible cases are considered: 1) under the *FirstFit* embodiment, the procedure continues scanning clients to the end of the banded queue, activating clients whose requirements can be met; 2) under the *FIFO* embodiment, the procedure ends with the first candidate client whose requirements cannot be met. Many other variations on these two embodiments might also be considered.

Anticipating Scheduled Variations in Network Capacity (Maximal Charge)

Overview

[00103] The procedure applicable to optimization of delivery charges is obtained by blending elements of the CAC method depicted in Figure 13 into the method depicted in Figure 11, which applies without change. To understand how this might work it may be useful to visualize a version of Figure 10 stratified along its length in the manner of Figure 8. As the maximum flow level undergoes a step change, so too do the widths of its constituent strata in equal proportion.

Procedure

[00104] As previously mentioned, the CAC method (Figure 11) applies to this circumstance also, provided we alter the definition of two routines, (17) and (18), upon which that procedure relies, yielding (20) and (21), and adopt the banded queue organization outlined in the previous section.

[00105] The server swing capacity at a future time t is computed according to the capacity and worst-case client flows at time t.

(20) swingCapacity(k, t) = minimum of (

(server.maxFlowRate(t) * (server.maxMinFlowRate[k-1] /server.maxMinFlowRate[N-1])

$\Sigma_{\text{i e Sk \& (t\_end(i) > t)}}$ (client.lookup(i).session.minFlowRate)),

(server.maxFlowRate - $\Sigma_{\text{i e SactiveClients \& (t\_end(i) > t)}}$ (client.lookup(i).session.minFlowRate)))

Finally, we define a predicate that tests whether a prospective customer will cause swing capacity to be exceeded at some time t, as follows:

```
(21) boolean client_fits(i ,t) {
        k = client.lookup(i).costOfService;
        if(client.lookup(i).contentSelection.averagePlayRate <=
                                swingCapacity(k,t) - server.cac_flowSafetyMargin)
                return true;
        else return false;
    }
```

System Description

[00106] Referring to Figure 14, a block diagram illustrating a preferred embodiment for a system implementing the methods presented herein is shown.

[00107] Block 1485 depicts plurality of network-based client computers that receive data from a server over the course of a session. Every client sessions has an associated session block 1435 within the server by means which a flow rate of content from client to server is effected and regulated to a given flow rate set-point. The flow optimizer 1430 manages bandwidth utilization across all sessions that have been admitted by the system call admission control (CAC) and the bandwidth call admission control blocks, labeled 1400 and 1420 respectively. Specifically, the flow optimizer 1430 modulates the flow rate set-point of every active session so as to optimize aggregate flow or, more generally, cost.

Call Admission Control

[00108] Whenever a client contacts a server requesting service, the server must determine whether or not to admit the client. This admission function is performed in two separate phases. The system call admission control block 1400 considers whether or not sufficient internal resources exist to support the request.

Toward this end, the system CAC 1400 estimates the needs of the prospective client with respect to a set of limited system resources such as I/O bandwidth, cache memory, open files, threads, etc. A request is denied if any required resource cannot be obtained. In the contrary case, the request is passed on to the Bandwidth CAC, which tests if the client fits and admits or rejects the client accordingly. The swing capacity is computed by the flow optimizer 1430 on a periodic basis according to equations (14) and (17) as appropriate. The managed bandwidth must reflect the true capacity of the attached network and the adapters leading thereto. This parameter should not be set arbitrarily high to prevent rejection of client requests. Doing so will cause the system to admit calls that cannot be properly handled, thereby diminishing the quality of the viewing experience.

[00109] Having admitted a client, block 1430 creates a client session block 1435, which then begins to converse with its client-resident counterpart, 1480) during their launch phase. Thereafter, the flow optimizer 1430 assumes control, and modulates flow over the session until either all content is delivered, or a failure to deliver is detected, or one of a number of VCR-like stream control events intervene (e.g. pause, rewind, fast-forward).

Server and Client Session Control Blocks

[00110] Data Transport Channel 1450 is responsible for transporting data packets originating from the flow regulator 1440 within the server to the former's peer entity 1460 within the client. The mode of communication between the peer entities may be connection-oriented (e.g. TCP, which delivers a byte stream in order, without duplication or loss) or datagram-oriented (e.g. UDP). Best-efforts datagram service is typically enhanced with acknowledgements to facilitate rapid recovery from occasional packet loss, out-of-sequence delivery, and duplication.

[00111] Peered session control channels 1445 and 1465 permit configuration and status data, together with miscellaneous commands, to be sent from client to server and vice-versa. A reliable connection-oriented transport service is typically employed for this purpose (e.g. TCP). Data and commands alike are multiplexed over this channel in the service of other entities within the clients or server. Thus, a datagram-oriented data transport channel 1460 within a client 1485 might employ the session control channel to ferry selective acknowledgements to its peer in the server. Similarly, the flow meter 1465 within the client 1485 forwards the measured flow rate to the associated flow regulator 1440 within the server.

[00112] The flow regulator 1440 obtains content data from a cache or disk, at the current offset into the content file, which it forwards as packets to the data transfer channel 1450. The size and/or pacing of the packets are determined by the flow regulator 1440 in order to achieve the flow rate set-point imposed by the flow optimizer 1430. The flow regulator 1440 is also responsible for determining the channel flow capacity to the client, which is the least of a configured value obtained from the client and a measured value. Toward this end, during the launch phase, the flow regulator 1440 might send content data packets to the client flow meter 1475 in a series of packet trains at progressively higher peak flow rates ending with the configured maximum flow rate, if any. The highest measured flow rate reported by the client flow meter 1475 is accepted as the flow rate capacity, which is posted for use by the flow optimizer 1430. Subsequently, the flow regulator 1440 will compare the flow rate set-point with the flow rates reported by the flow meter 1465, and down grade capacity as needed should the measured flow rate fall consistently fall below the requested flow rate. Whenever such a downgrading occurs, the flow regulator 1440 will test for subsequent

relaxation of a transient constriction by means of the aforementioned series of packet trains in which the peak packet flow rate exceeds the delivered flow rate.

[00113] The buffer manager 1470 accepts data from the flow meter 1465, which data is retained until it is consumed or passed over the content consumer 1480. The buffer will typically be implemented as a ring buffer in memory or on disk. The write cursor will lag the read cursor by a no less than a specifiable time interval so as to permit rewinds of acceptable scope without requiring selective data re-retrieval from the server.

Flow Optimizer 1430

[00114] The flow optimizer 1430 assumes control over a session once it has been launched, and its channel capacity has been determined. Subsequently, it modulates the flow rate set-point of every session such as to optimize aggregate flow or charges across all active sessions, subject to a variety of min/max constraints on sessions flows, and a maximum constraint on aggregate session flow, as previously discussed.

[00115] The flow optimizer 1430 continues to modulate session flow until all content has been delivered, or one of a number of session anomalies intervene, as follows:

[00116] A downgraded session flow capacity drops below the minimum session flow rate. The session is dropped by the optimizer 1430 as a solution involving this channel does not exist. The user is forced to apply for re-admission, with the option of accepting slower than real-time delivery at the present encoding rate, or accepting a degraded viewing experience at a reduced encoding rate better adapted to the presently available bandwidth.

[00117] A pause VCR command is received from the client. Flow to the client can continue until the client-side buffer is topped off, after which time flow drops to zero. The session is permanently dropped should the pause persists too long. Until that time, the CAC safety margin is temporarily increased by the jit (Just-In-Time) flow rate of the paused session. In this way the session can be resumed without the need for re-admission and the jit bandwidth is made available to currently sessions so long as the pause persists.

[00118] A VCR Jump-Forward command is received from the client. This action is problematic as it causes the just-in-time flow rate (calculated at the new forward offset) to increase, in violation of the assumptions of the CAC algorithm. It must be noted that this violation will occur even if the content that is sought has already been delivered to the client-side buffer. One workable policy is to force the client to apply for re-admission at the forward offset and with a suitably modified buffer level.

[00119] A VCR Jump-Backward (Rewind) command is received. This action may entail the need for re-transmission of content, as previously discussed. The just-in-time flow rate will decrease as per the new position and adjusted buffer contents.

Tunable Minimum Constraint

[00120] A parameterized family of minimum flow constraints can be defined as follows:

(22)

wherein: $f_k(t)$ = Target content flow rate for session k;

$f_k^{JIT}(t)$ = Just-in-time flow rate for session k;

(RHS of Equation (5))

β = Burst tuning factor (positive or zero); and

$f_k^{min}(t)$ = Minimum (reserved) flow rate for session k.

[00121] This family converges on $f_k(t) >= f_k^{JIT}(t)$ when β approaches 0. A β value of .1, for example, ensures a minimum flow rate that exceeds the JIT flow rate by 10%. Thus, non-zero values of β force early delivery of content to a player's media buffer, with beneficial effects on the viewing experience.

Alternative Call Admission Control

[00122] The traditional CAC algorithm compares the flow rate requirement of a prospective client with the unused bandwidth (obtained by subtracting the aggregate flow target from the server flow capacity) and grants service if unused bandwidth is not exceeded. This procedure may break down as the optimizer causes all available bandwidth to be used if it can, thereby reducing unused bandwidth to zero in the best case, even when a new client could be accommodated with a healthy margin to spare.

[00123] Nevertheless, a simple variation on this idea can be used. Available bandwidths can be defined as the difference between server flow capacity and the sum of minimum flow rates for each session computed according to (22) as follows:

(23) $f_{available} = f_{Svr}^{max} - (f_1^{min} + \ldots + f_n^{min})$

wherein: $f_{available}$ = Available bandwidth

f_{Svr}^{max} = Server aggregate flow capacity; and

f_n^{min} = Minimum (reserved) flow rate for session n.

[00124] Service is granted if the computed minimum flow rate of the prospective client is less than available bandwidth even when unused bandwidth is zero.

Toward a Tunable QOS

[00125] In the preceding discussion it is assumed that all clients share the same value of β. One might also consider a scheme whereby a client receives a β value according to the class/cost of service that client has signed up for: the greater the cost (and expected quality of service, or QOS) the greater the value of β.

[00126] Instead of maximizing charges by the methods described above, a variation on the algorithm to maximize flow is considered whereby the reserve flow computation for every session takes into account the individual β value of the associated client. In this fashion a single algorithm can serve either end. The optimizer first ensures that every session receives its required minimum bandwidth (which would now typically depend on β and elapsed session time) before allocating any excess bandwidth more or less evenly to all by interpolation. (According to Figure 15.)

[00127] The scheme is simpler computationally than the iterative interpolation method shown in Figure 16; it is also arguably more fair and stable with respect tot he allocation of available bandwidth.

Deadline-Driven Content Distribution

[00128] In delivering viewable or audible content to a player, the optimization algorithm imposes a minimum flow rate that ensures that all content is delivered by the implied deadline, which is the last moment of a play session. Notwithstanding, the optimizer attempts to deliver content earlier according to bandwidth availability within the server taken as a whole. A similar, though far simpler scenario arises relative to the distribution of content from an optimized server to a target device, whenever the delivery must be accomplished by a designated time, or equivalently, within a given elapsed time, and when overlapped non-blocking consumption of the content at a fixed rate is not required.

[00129] One example concerns the application where a session bundle comprises a session associated with the featured program together with sessions for every advertisement scheduled to interrupt the feature program at a designated time. Each session is endowed with a separate media buffer, which, in the case of ads, would typically be large enough to accommodate the ad in full, thereby permitting a seamless transition to the ad at the appointed time.

[00130] In adapting the optimizer to the needs of content distribution we must revisit the formulation of the flow constraints, as follows:

1. Media buffers are not of a concern, let alone buffer overflow, and it can be assumed that all content received is saved away in a file on disk, or in memory awaiting further use.

2. For these same reasons, the value of β may be zero, at which the minimum flow rate equals the just-in-time rate obtained by dividing the content as yet undelivered by the time till deadline:

(24) $f_k^{JIT}(t_k) = L_k^{TOGO} / (T_k - t_k)$

wherein: L_k^{TOGO} = Content to go for session k, i.e., as yet undelivered by the server at time t, (client.lookup (k).session.payloadToGo).

Given the arbitrary format of the file to be distributed, T no longer signifies the "duration of play" at the average play rate (averagePlayRate); rather, T represents the elapsed time to the deadline measured at the moment the delivery session was launched, and t represents the elapsed time since session launch.

Session Bundles

[00131] A session bundle is a set of sessions that flow along a shared channel, and are thus subject to an aggregate flow constraint equal to the capacity of the shared channel. One important application of session bundles is connected with ad insertion.

[00132] It is not enough to treat each of the constituent sessions as independent with respect to the flow optimizer, for this practice might well result in session flow targets that do not exceed channel capacity when considered singly, yet exceed this same shared capacity when summed together.

[00133] A solution to this problem involves two distinct steps:

1. The sessions bundle is treated as a virtual session by the server optimizer, subject to the minimum flow constraint obtained by summation of the individual session minima, and a maximum constraint that is the least of (a) the sum of the least of constraints (3) and (4) for each session, and (b) the shared

channel capacity client.lookup(k).channel.maxFlowRate (f_k^{cap}). The flow optimizer then generates an aggregate flow target for all channels virtual sessions by interpolation, using:

(25) $Cost(t) = f_1(t)+..+f_n(t)$

[00134] The cost vector in this instance is the n-dimensional unit vector, where n is the number of active clients.

(26) $c=(1, 1,....,1)$

[00135] Now it is possible to have an achievable aggregate flow target for the session bundle k that we must apportion optimally among the constituent sessions. Fortunately for us, our interpolation procedure of Figure 15 can be applied again to the sessions in our bundle, but where the aggregate flow target generated above replaces the aggregate flow constraint f_{svr}^{max}

Specifically:

For every flow f_{ki}, within session bundle k, the optimal flow rate is obtained by interpolation between minimum and maximum session flows:

(27) $f_{ki}^{opt} = f_{ki}^{min} + \alpha * (f_{ki}^{max} - f_{ki}^{min})$

(28) $\alpha = [(f_K^{bundle} - f_{k1}^{min} + ...f_{kn}^{min})]/[(f_{k1}^{max} - f_{k2}^{min}) + + (F_{kn}^{max} - f_{kn}^{min})]$

and the value of f_k^{bundle} is obtained from a higher level optimization in which session bundle k is treated as a virtual session for which:

(29) $f_k^{min} = f_{k1}^{min} + ... + f_{kn}^{min}$

(30) $f_k^{max} = \min(f_{k1}^{max} + .. + f_{kn}^{max}, f_k^{cap})$

It must be noted that session bundles are typically dynamic in nature, outliving many if not all of their constituent sessions. Before a session is added to the bundle two CAC decisions must be reached:

(1) establish bandwidth availability within the bundle; and

(2) bandwidth availability within the server.

Relationship to Adaptive Rate Control

[00136] The flow optimization methods should be viewed as constituting the lowest and perhaps only the first of many supervisory control layers to sit astride all streaming sessions, which comprise their own hierarchical stack of control functions; rate adaptive control sitting above a flow rate regulator which relies on a transport entity.

[00137] Channel capacity is known to vary widely over time. A capacity drop to below the minimum reserve flow rate is problematic as it causes the latter to rise in violation of its monotonicity assumption. In a server with available bandwidth to spare such a rise can be accommodated by the flow optimizer (by bumping the session's minimum flow reservation to the higher value) provided the flow deficit can be made up in the near term by draining pre-delivered content from the media buffer. Should either of these resources approach depletion, stream degradation to a lesser bit-rate or frame rate may be the only viable option short of a pause. Whether up or down, any rate change is a change to *fe*, which must be preceded by change to session's reserve flow rate allocation by the optimizer.

[00138] As can be seen, the optimizer and rate adapter must be integrated with care. One possible approach would be to modify a rate adapter such that it renegotiates its proposed *fe* value with the flow optimizer prior to effecting the rate change it deems appropriate based on information provided by the flow regulator. Rate increases might well be subject to deferral under heavy load conditions, awaiting bandwidth availability. The opposite is true of rate flow decreases: these would be forestalled as long as possible by the optimizer on a system that is not fully loaded, as previously discussed.

Live Events

[00139] Referring to Figure 17, the method of the present application can be extended to the optimization of "live", time-shifted "near-live" events, and cascaded delivery, wherein one server delivers content to another while the latter accepts clients for the same content. The extension assumes the following characteristics of the server and the live performance:

1. The performance is of a fixed duration T known in advance, encoded at an average rate *fe*, with a projected content payload of *fe*T*;

2. The origin server at the point of injection delays the stream by a small dwell time T^{DWELL} on the order of seconds (e.g. 10). Notwithstanding, the delayed content is available for early delivery to mirrored servers and players with the aim of providing these with a modicum of isolation from network jitter affecting the timely delivery of the live video frames.

3. Throughout the live event, captured content is stored in a file, thereby permitting late joiners to view the "live event" starting from any point between the present and the event start time;

4. At the conclusion of the live event, the encoded content, now fully captured to a file, is available for viewing over the balance of the admission window (e.g., 60 minutes in Figure 17).

[00140] The optimisation of live events and video on demand (VOD) differ principally in one respect, namely in the formulation of the server content underflow constraint.

For VOD, this constraint is

(31)

$L_k^{TOGO}(t)$ is initialised to the content size (*fe*T*) at that start of play, and decremented as content is delivered to the client. Δt is the time interval between successive periodic optimizations (e.g. 1 second). In practice this constraint rides

well above other operative maximum flow constraints (e.g. buffer overflow, channel capacity) until the very last calculation of the session flow target by the optimiser.

[00141] For a live stream, this constraint must additionally reflect the fact that content is streaming into the server (where it builds up), even as the server streams it out toward clients, and that the latter activity cannot possibly overtake the former.

[00142] Referring to Figure 17, (which illustrates how optimised live streams behave) line L1 depicts the build-up of freshly captured content to the server. Line L2, which is delayed (time-shifted) relative to L1 by the dwell time T^{DWELL}, represents the live performance as cast and viewable by clients. One such client is admitted at time A and joins the live session at the live moment B. The dashed curved line linking points B and C shows a plausible build-up of content in the client's media buffer in advance of play, as the viewer continues its advance along line L2. Past point C the top of the media buffer tracks the live capture stream and thenceforth advances along line L1. In this instance, the time shift between the live performance as cast (L2) and the live performance as captured (L1) imposes a maximum on the amount of content that could ever accumulate in a client media buffer, irrespective of its actual capacity.

[00143] Another client, admitted at time $t=10$, is interested in viewing the live stream from the beginning without missing anything. As before, the dashed line joining 10 and D depicts a plausible build-up of content in the client's media buffer in advance of play, as the viewer continues its advance along line L3, which is time-shifted by $(10+T^{DWELL})$ minutes relative to L1. The greater time shift (now between L1 and L3) permits more extensive pre-buffering of "near live" content.

[00144] The present constraints do not impose suitable flow limits at and beyond points C and D when the client delivery streams finally links up with the live capture stream, in the live moment. Consequently the optimiser may propose flow targets that cannot be realized.

[00145] Accordingly, the following modification for use in connection with live or cascaded streams is made:

(32) $f_k(t) <= [(L_k^{TOGO}(t) - Live^{TOGO}(t)] / \Delta t$

for a live event delivered at flow rate *fe*, we have

(33) $Live^{TOGO}(t) = \max(fe*(T - T^{DWELL} - t), 0)$

[00146] $Live^{TOGO}(t)$ is the span between projected content size *fe* **T* (level line L0 in Figure 17) and the content capture level line L1, whereas $L_k^{TOGO}(t)$ is the gap between L0 (i.e., *fe* * *T*) and the dashed curved line representing the content level delivered to the media buffer. The latter must always exceed the former. As the two approach one another this constraint takes effect, causing $L_k^{TOGO}(t)$ to always exceed $Live^{TOGO}(t)$ as it decreases at rate *fe*.

[00147] Viewers joining late and willing to see the live performance with some delay relative to the live case enjoy higher potential levels of isolation from network disturbances, and thus QOS, by virtue of their greater effective buffer capacity. By the same token, such viewers afford the optimizer greater latitude to optimize the flow of VOD and "live" content across load fluctuations, thereby enhancing server capacity. Viewers that wish to experience the live event in "real-time" nevertheless benefit by a modestly enhanced QOS conferred by a dwell time's worth of content in their buffers.

Summary

[00148] A method and system for call/connection acceptance and flow modulation for network delivery of video/audio programming is thus provided. Although several embodiments have been illustrated and described, it will be apparent to those skilled in the art that various changes and modifications may be made without departing from the spirit of the invention or the scope of the claims.

CLAIMS:

1. A method of bandwidth allocation for delivery of stored digital content from at least one server device to at least one client device by way of a network, the method comprising the steps of:

a) prescribing a control variable which represents a target flow rate from the server device to each client device;

b) determining time-varying constraints on the flow rate of the content;

c) determining a cost function of the control variables for all clients; and

d) prescribing bandwidth to all clients based upon the value of the control variables that maximize the cost function.

2. The method of Claim 1 wherein step (d) comprises performing periodic computations to update the value of the control variable such that the bandwidth can be continuously allocated to each client.

3. The method of Claim 2 wherein a new client is accepted by:

i) determining an admission capacity of the bandwidth;

ii) admitting a prospective client if the clients minimum allowed value of the control variable is less than the admission capacity; and

iii) wherein a client admitted for service is guaranteed to have sufficient content flow over the entire session.

4. The method of Claim 3 wherein:

The admission capacity equals a server swing capacity reduced by a predetermined safety margin which may be zero, the swing capacity equaling the difference between the server flow capacity and the sum of the minimum allowed flow rate for all clients.

5. The method of Claim 4 wherein the minimum allowed value of the control variable is initialized to the average consumption rate.

6. The method of Claim 4 wherein the minimum allowed value of the control variable is initialized to the average consumption rate multiplied by a factor greater than one.

7. The method of Claim 4 wherein the server flow capacity varies in a step-wise function over time according to a predetermined schedule and the method of accepting a new client further comprises:

i) determining a sequence of future step changes of server flow capacity;

ii) determining at each time in the sequence of future step changes the value for a future worst swing case capacity, the worst case swing capacity being obtained from the server flow capacity at the time by subtracting an extrapolated present minimum allowed flow rate for all active clients that are potentially active in the future;

iii) admitting a prospective client if a minimum allowed value of the control variable is less than the capacity obtained from the sequence and the present capacity;

iv) whereby a new client can be admitted without compromising the service to existing clients in the presence of previously scheduled server flow capacity changes and all clients accumulate content in their buffers to the extent of their respective buffer capacity.

8. The method according to Claim 7 wherein the method of accepting a new client further comprises:

i) partitioning the available bandwidth into partitions representing a prescribed service charge;

ii) determining a maximum-minimum capacity for a service charge to be the maximum allowed sum of all minimum allowed flow rates for all clients incurring an equal or lesser service charge, the maximum-minimum capacity equaling the sum of bandwidth allocations for all service charges less than or equal to the service charge;

iii) determinating a service-charge-specific swing capacity for a given service charge as the difference between the maximum-minimum capacity for the service charge and the sum of all minimum allowed flow rates for all clients incurring the service charge or a lesser service charge;

iv) determining the server swing capacity to be the server flow capacity minus a margin and minus the sum of the minimum allowed flow rate for all clients; and

v) determining the service-charge-specific capacity to be the least of the server swing capacity and the charge specific swing capacity.

9. The method of Claim 8 wherein the method of accepting a new client further comprises:

i) determining the sequence of future times of step changes to server flow capacity over a proposed consumption period;

ii) determining for each time in the sequence of future times, the future worst-case value for the service-charge-specific capacity, the worst case value being obtained by using the extrapolated present minimum allowed flow rate of all presently active clients that are potentially active at the time in the future; and

iii) admitting a client incurring a given service charge if the minimum flow rate of the client is less than the capacity obtained from the sequence and the present service-charge-specific capacity.

10. The method of Claim 3 wherein step (b) comprises determining time-varying constraints wherein:

i) an aggregate flow rate for all clients does not exceed a predetermined server flow capacity;

ii) the flow rate from the server to the client does not exceed a maximum allowed flow rate for the client;

iii) the flow rate to the client will never cause a buffer of the client to overflow;

iv) the flow rate to the client stops when the content is exhausted; and

v) the flow rate from the sever will never be less than the client's minimum allowed flow rate which may vary over time and may be zero.

11. The method of Claim 10 wherein the maximum allowed flow rate to the client is given by the minimum of one selected from the following:

i) a client flow rate ceiling;

ii) the flow rate required to fill the client buffer before the next periodic computation; and

iii) the flow rate required to complete the delivery of the content before the next periodic computation.

12. The method of Claim 11 wherein:

i) the minimum allowed client flow rate cannot increase over time, but may decrease such that the initial value of the control rate is no less than the average consumption rate;

ii) the control variable for the client is always greater than or equal to the current minimum allowed client flow rate; and

iii) the client buffer will never underflow until the content is fully consumed at the average consumption rate.

13. The method of Claim 12 wherein the minimum client flow rate is the product of the constant content flow rate from the client to the server that causes the last piece of content to be delivered at the last possible moment as the client buffer is being drained at the average consumption rate, with a factor greater or equal to one.

14. The method of Claim 13 wherein the minimum client flow rate is initialized to the average consumption rate.

15. The method of Claim 13 wherein the minimum client flow rate is initialized to the average consumption rate multiplied by a factor greater than one.

16. The method of Claim 13 further comprising the step of calculating the minimum client flow rate periodically.

17. The method of Claim 1 wherein the cost function represents an aggregate flow rate and is the sum of the flow rates for all clients.

18. The method of Claim 1 wherein the cost function represents a service charge and is the sum of the client flow rates multiplied by the client's cost of service.

19. The method of Claim 18 wherein:

step (a) comprises:

i) determining a maximum allowed flow rate and a minimum allowed flow rate for each client;

ii) determining a flow rate range for each client as the difference between the maximum allowed flow rate and the minimum allowed flow rate; and

iii) initializing a current flow rate for each client as the minimum allowed flow rate and summing the flow rate into the total server flow rate; and

step (d) comprises:

i) computing remaining server bandwidth as the difference between the maximum server flow capacity and the total server flow rate; and

ii) allocating remaining server bandwidth to remaining clients wherein all clients with a given cost of service receive bandwidth equally to the extent of their respective ranges, and no client with a given cost of service being allocated bandwidth unless all clients with a higher cost of service have received their maximum possible allocation of bandwidth.

20. The method of Claim 17 wherein step (a) prescribing the control variable comprises:

i) forming a hyperplane of control variables that aggregate to the maximum allowed aggregate flow rate;

ii) computing a first vector of minimum allowed client flow rates;

iii) computing a second vector of maximum allowed client flow rates;

iv) determining a multi-dimensional shape of which an interior diagonal spans the first and second vectors; and

v) finding a solution of control variables at the point of intersection between the diagonal and the hyperplane.

21. The method of Claim 20 wherein step (v) comprises finding a solution of control variables at the second vector if there is no point of intersection between the diagonal and the hyperplane.

22. The method of Claim 17 wherein step (a) comprises:

i) computing the sum of the minimum allowed flow rates to all active clients;

ii) computing the sum of the flow rate ranges for all active clients;

iii) computing the difference between the maximum allowed aggregate flow rate and the sum of the minimum;

iv) computing a factor as the ratio of the difference over the sum of the ranges;

v) if the factor exceeds 1, the control variable is set to the maximum allowed client flow rate; and

vi) if the factor is less than 1, computing the control variable by multiplying the client range by the factor and then adding the result to the minimum allowed client flow rate.

23. The method of Claim 17 wherein:

step (a) comprises:

i) determining the maximum allowed flow rate and minimum allowed flow rate for each client;

ii) determining the flow rate range for each client as the difference between the maximum flow rate and the minimum flow rate; and

iii) initializing the control variable for each client as said minimum allowed flow rate and summing the flow rate into a total server flow rate; and

step (d) comprises:

i) computing remaining server bandwidth as the difference between the maximum server flow capacity and the total server flow rate; and

ii) allocating remaining server bandwidth to remaining clients to the extent of their respective ranges.

24. The method of Claim 23 wherein the step of allocating remaining server bandwidth further comprises the clients receiving the bandwidth equally to the extent of their respective ranges.

25. The method of Claim 17 wherein the maximized value of the control variable can be determined by:

i) determining a maximimum and minimum flow rate for each client;

ii) determining a range between the maximimum and mininmum flow rates in order to find a flow rate range;

iii) determining available bandwidth by finding the difference between an aggregate flow capacity and a sum of the minimum flow rate;

iv) determining a flow variable by dividing the unused bandwidth by the flow rate range;

v) prescribing the control variable to be the minimum flow rate added to the flow rate range if the flow varible is less than one or prescribing the control variable to be the minimum flow rate added to the flow rate range corrected by the flow variable if the flow favorable is greater than one such that the control variable is easily calculated for each client.

26. The method of Claim 17 wherein at least one session is a session bundle and:

1) the minimum flow constraint is computed as the sum of the minimum flow rate constraints for all bundled sessions;

2) the maximum flow constraint is the least of 1) the sum of the maximum individual flow rate constraints of each session and 2) the flow capacity of the channel shared among the bundled sessions; and

3) the target flow rate is computed for the session is apportioned among the bundled constituent sessions.

27. The method of Claim 26 wherein the bundled session is an advertisement.

28. A system for allocating bandwidth between a server device and at least one client device, the system comprising:

a call acceptance module operative to receive an incoming request for service;

a flow regulator configured to deliver content at a modulated target flow rate, the content being delivered between the server device and a respective client device when a call is accepted by the call acceptance module; and

a flow optimizer configured to modulate the target flow rate of the flow regulator in order to optimize the aggregate flow of content.

29. The system of Claim 28 wherein the flow optimizer is configured to modulate the flow rate of the flow regulator in order to optimize charges for the content.

30. The system of Claim 29 wherein the flow regulator is configured to deliver the modulated flow rate of content in response to the control variable.

31. The system of Claim 28 wherein the flow optimizer is configured to determine a control variable which corresponds to an optimized flow rate.

32. The system of Claim 31 wherein the flow optimizer is configured to generate a cost function of the control variable that corresponds to a maximized value of the control variable.

33. The system of Claim 28 wherein the flow regulator is configured to deliver the modulated flow rate of content in response to the control variable.

34. The system of Claim 28 wherein the flow optimizer is configured to determine a maximized value of the control variable by:

i) determining a maximum and minimum flow rate for each client;

ii) determining a range between the maximum and minimum flow rates in order to find a flow rate range;

iii) determining unused bandwidth by finding the difference between an aggregate flow capacity and a sum of the minimum flow rate;

iv) determining a flow variable by dividing the unused bandwidth by the flow rate range;

v) prescribing the control variable to be the minimum flow rate added to the flow rate range if the flow variable is less than one or prescribing the control variable to be the minimum flow rate added to the flow rate range corrected by the flow variable if the flow favorable is greater than one such that the control variable is easily calculated for each client.

35. The system of Claim 33 wherein the call acceptance module is configured to accept a call based upon the value of the control variable.

36. A method of bandwidth allocation for delivery of stored digital content from at least one server device to at least one client device by way of a network, the method comprising the steps of:

a) prescribing a control variable which represents a target flow rate from the server to each client device based upon the amount of buffer of the client;

b) determining time-varying constraints on the flow rate of the content;

c) determining a cost function of the control variables for all clients in response to the size of the client's buffer; and

d) prescribing bandwidth to all clients based upon the value of the control variables that maximized the cost function.

37. A method for apportioning a given total flow rate (bandwidth) among a plurality of sessions for clients, the method comprising the steps of:

a) computing a sum of minimum flow rate constraints for each session;

b) computing a sum of flow rate ranges for all active clients;

c) computing a difference between the given total flow rate and a sum of the flow rate minima;

d) computing a factor as the ratio of the difference over the sum of the flow rate ranges;

e) if the factor exceeds 1, setting the flow rate to the maximum allowed client flow rate; and

f) if the factor is less than 1, computing the flow rate by multiplying the client range by the factor and then adding the result to the minimum allowed client flow rate.

38. The method of Claim 37 wherein the clients are distinguished by their cost of service and bandwidth is apportioned to clients iteratively, the amount apportioned to each client in a cost category is the amount remaining from prior allocations to clients in higher cost categories.

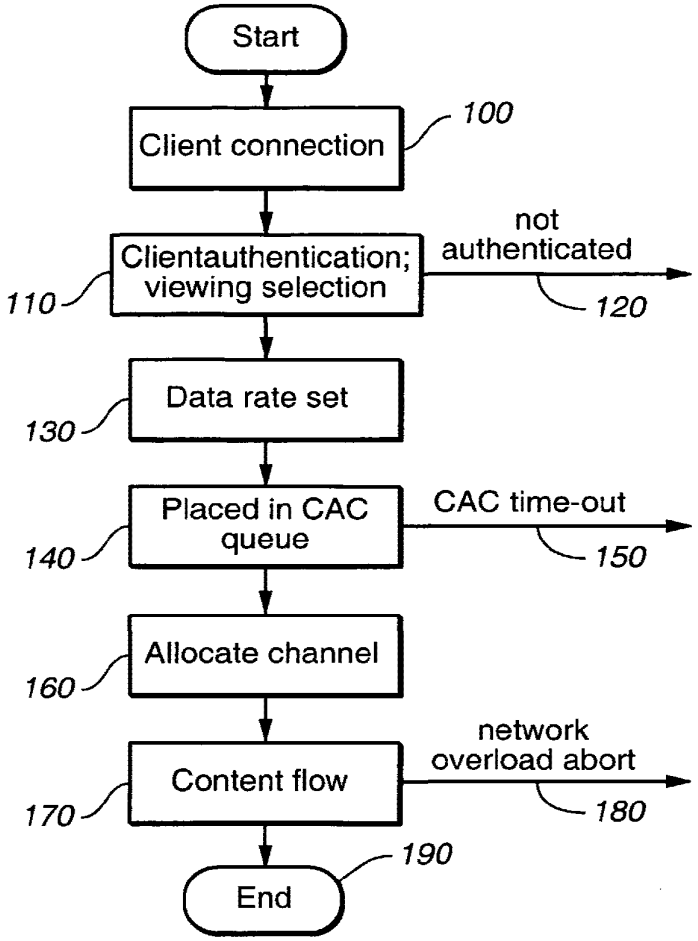

Start
Client connection
100
Clientauthentication; viewing selection
not authenticated
110
120
Data rate set
130
Placed in CAC queue
CAC time-out
140
150
Allocate channel
160
Content flow
network overload abort
170
180
End
190

FIG._1

Aylo Exhibit 1018
Page 1678 of 2308

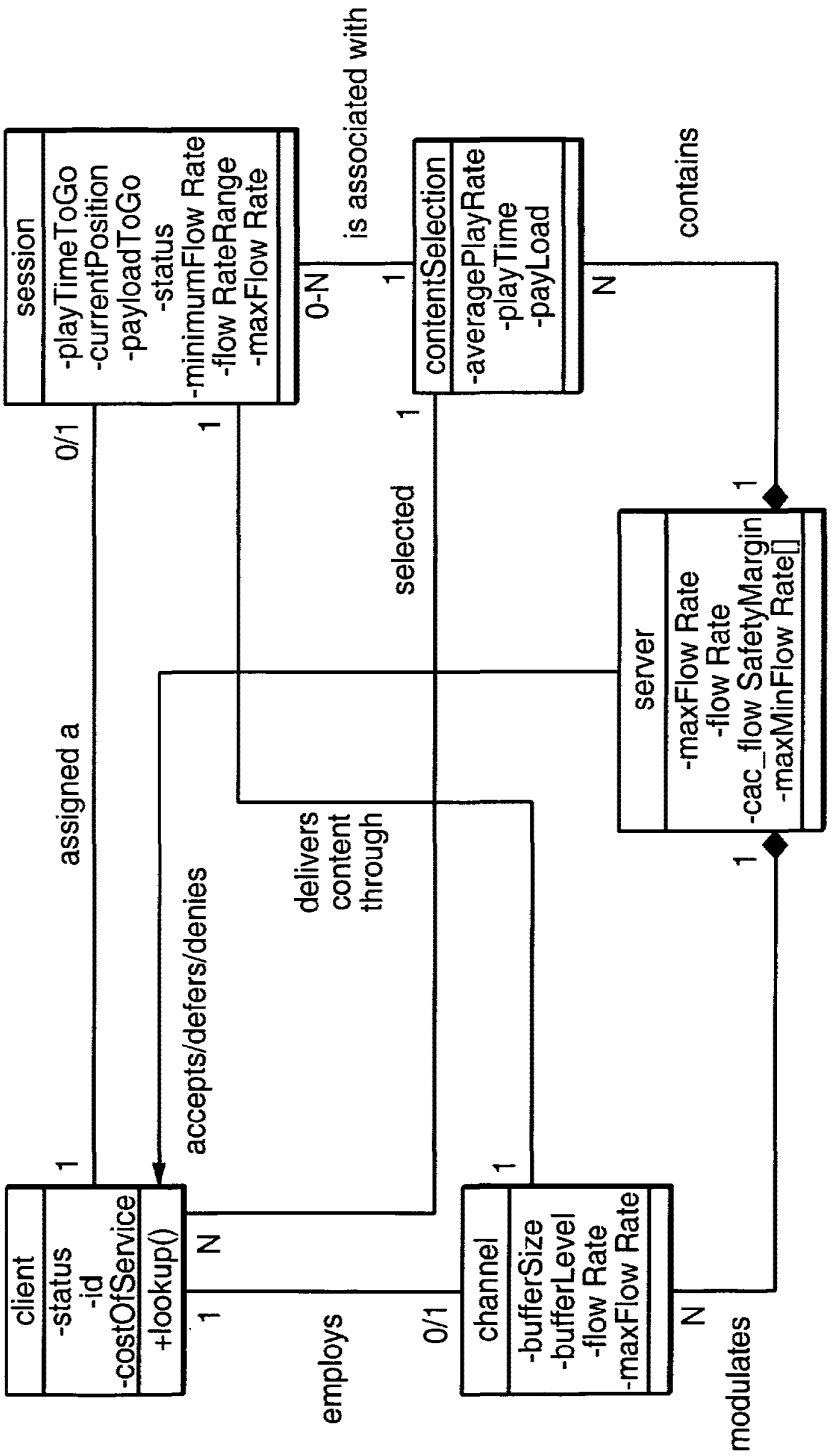


FIG._2

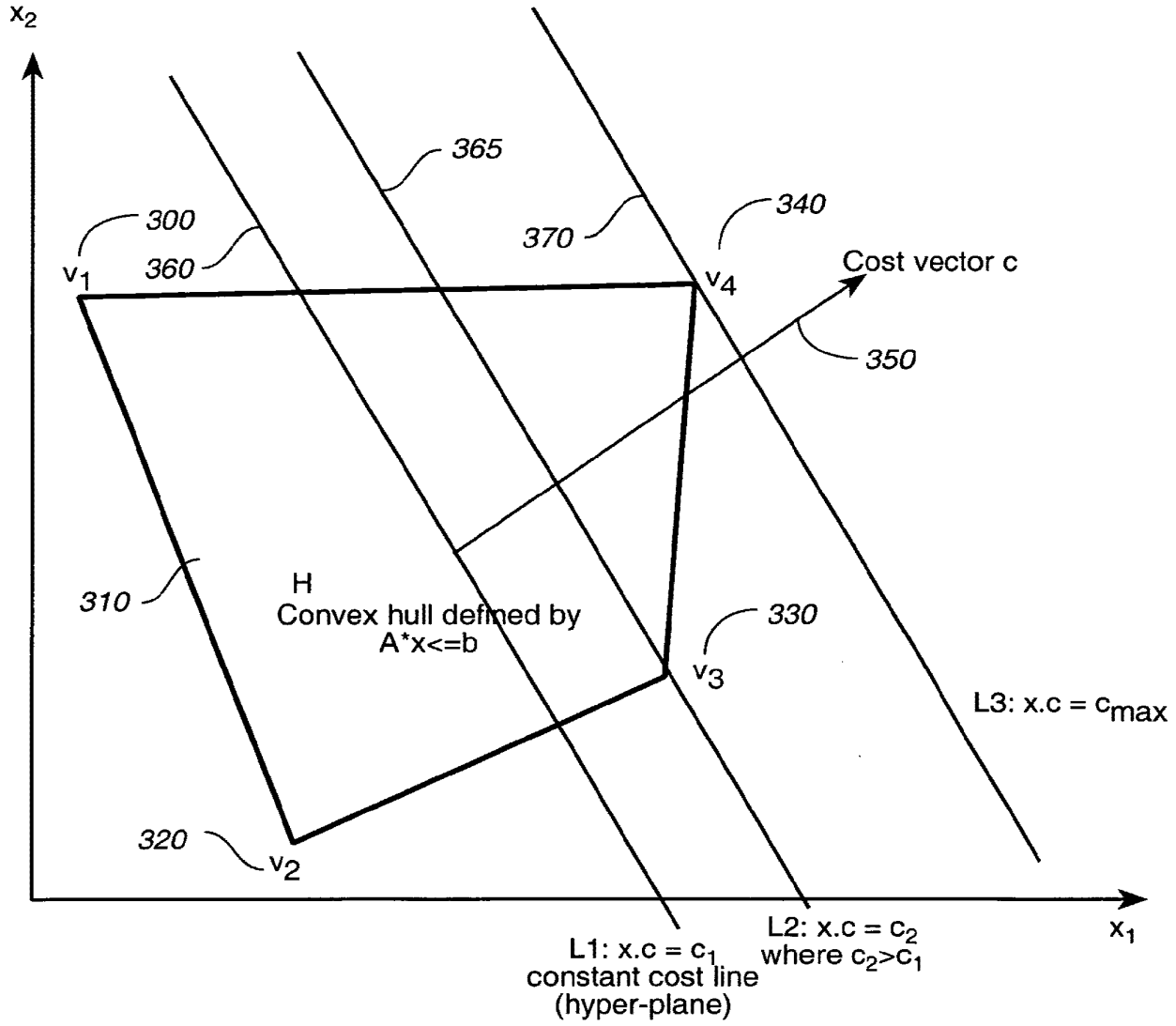

x2
x1
300
v1
360
365
370
340
v4
Cost vector c
350
310
H
Convex hull defined by
A*x<=b
330
v3
L3: x.c = cmax
320
v2
L1: x.c = c1
constant cost line
(hyper-plane)
L2: x.c = c2
where c2>c1

FIG._3

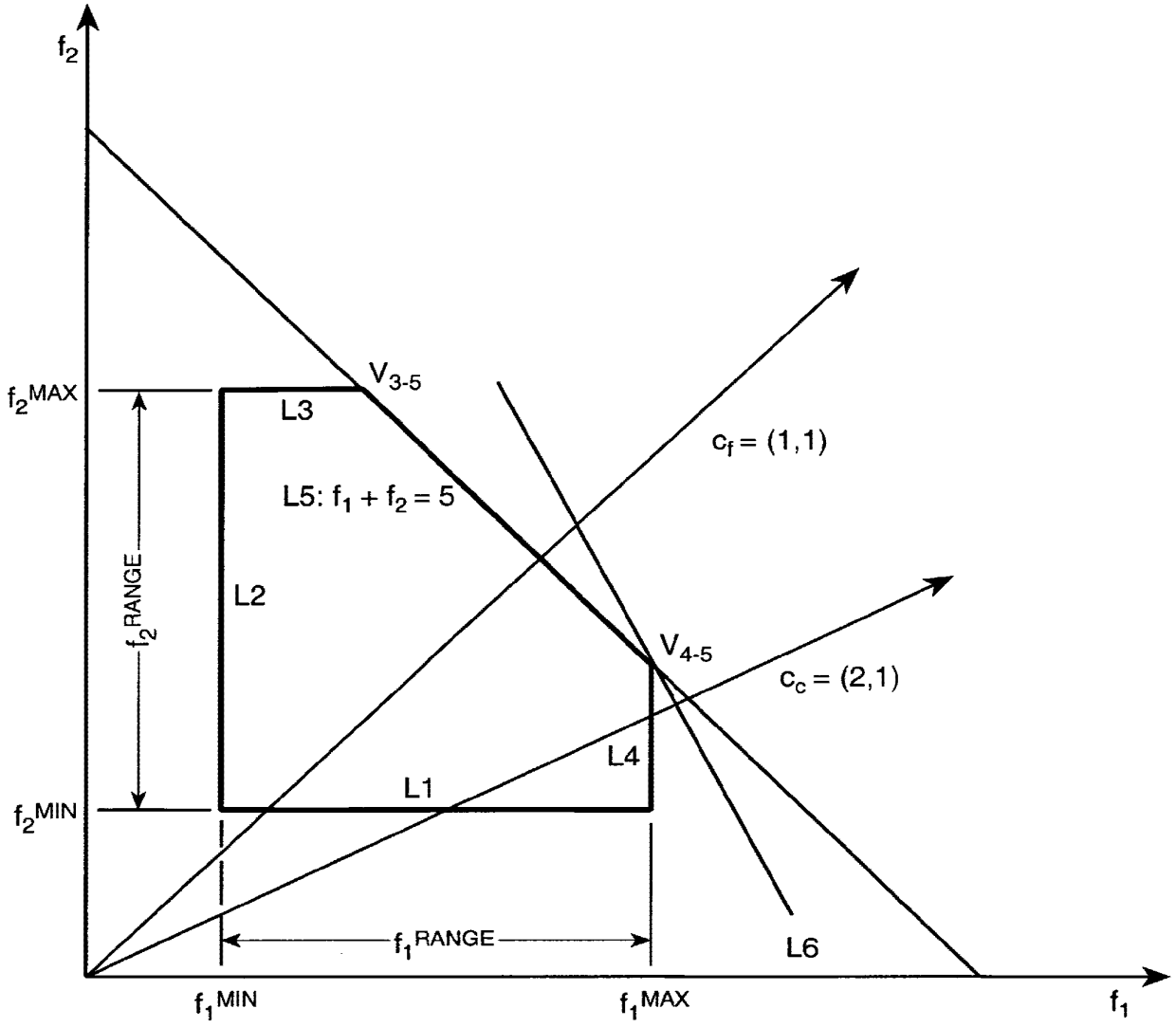

f2
f2MAX
f2MIN
f2RANGE
V3-5
L3
L5: f1 + f2 = 5
L2
V4-5
cf = (1,1)
cc = (2,1)
L4
L1
L6
f1RANGE
f1MIN
f1MAX
f1

FIG._4a

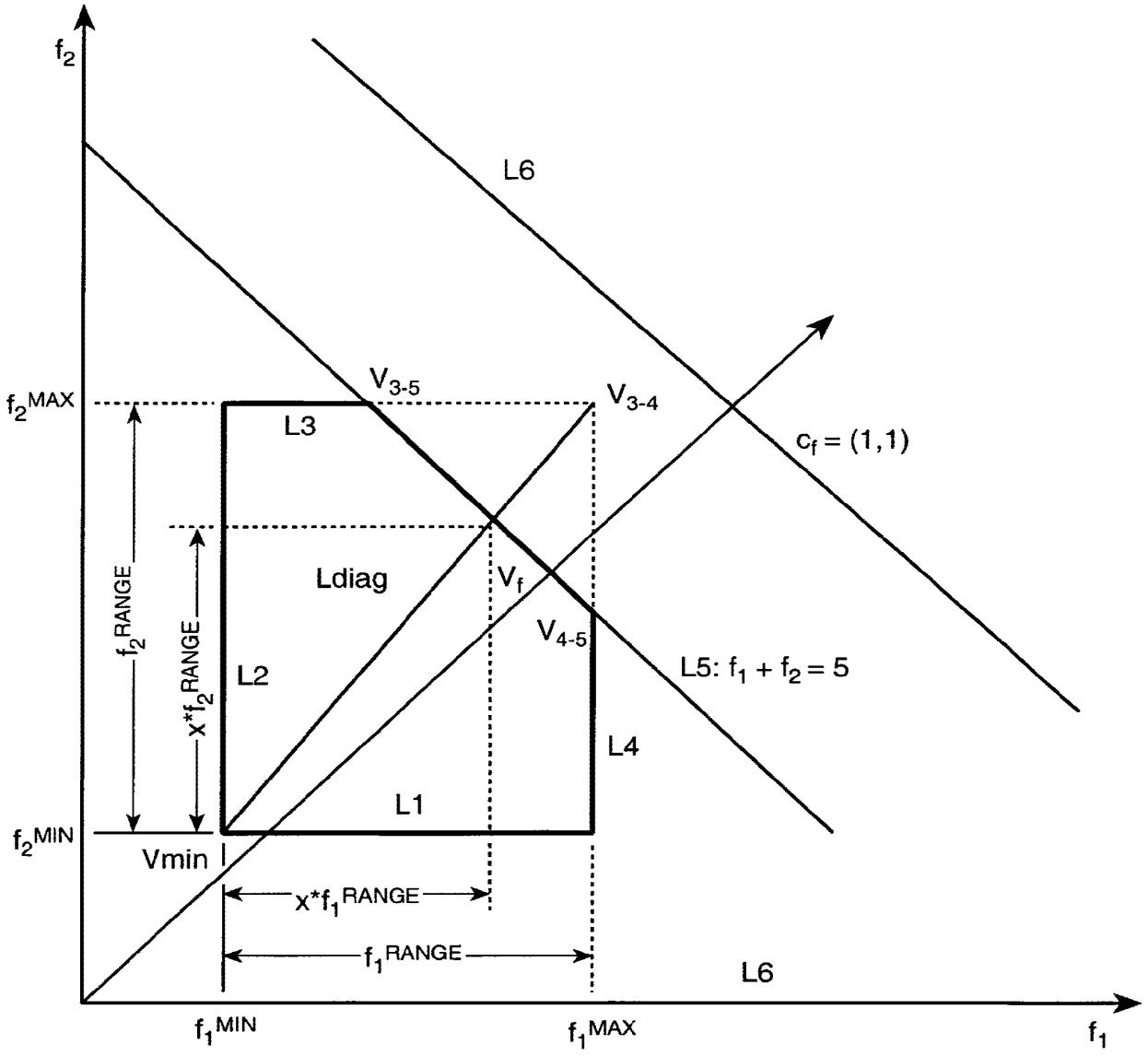

f2
L6
f2MAX
V3-5
V3-4
L3
cf = (1,1)
Ldiag
Vf
f2RANGE
x*f2RANGE
V4-5
L2
L5: f1 + f2 = 5
L4
L1
f2MIN
Vmin
x*f1RANGE
f1RANGE
L6
f1MIN
f1MAX
f1

FIG._4b

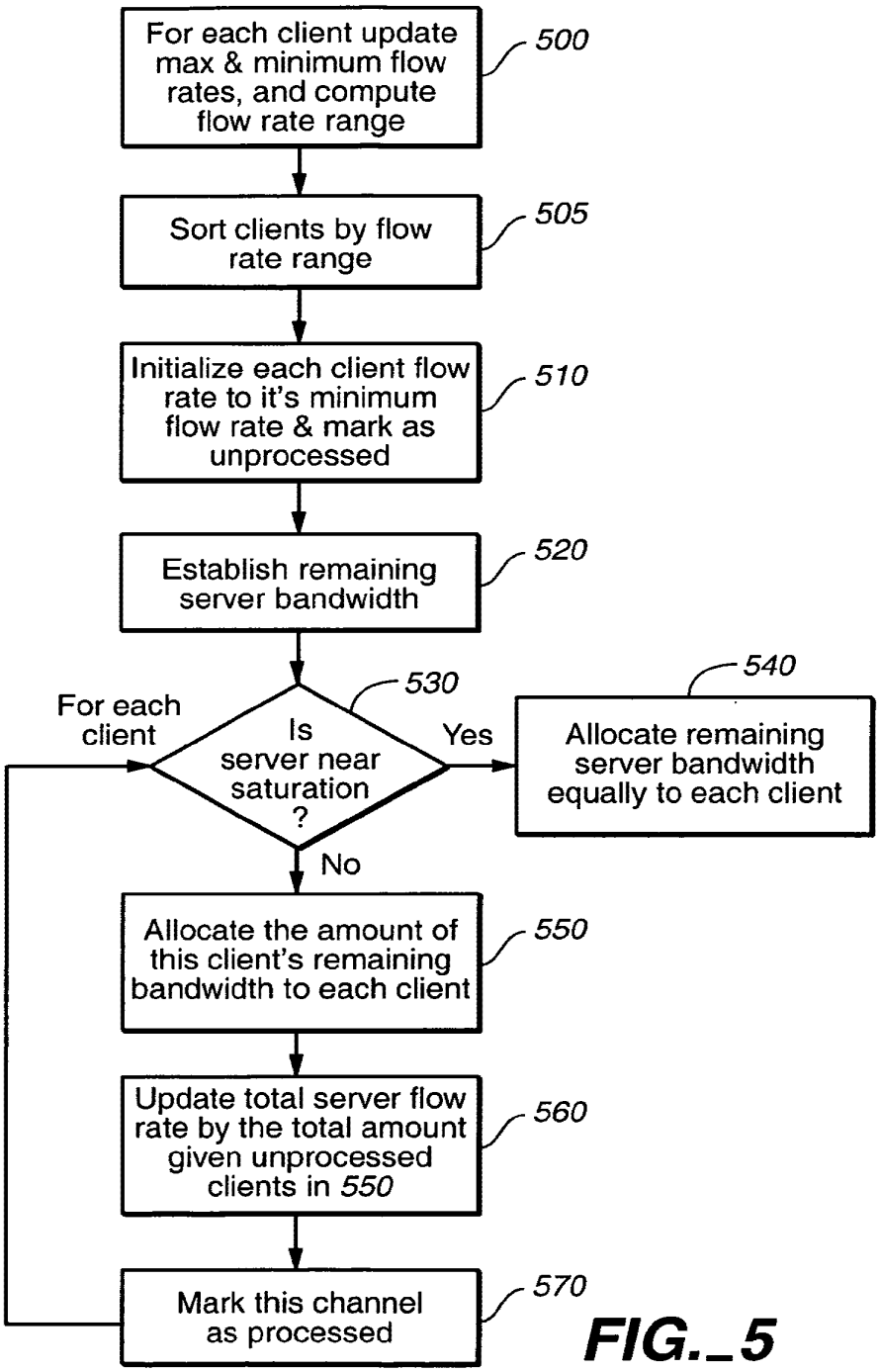

For each client update max & minimum flow rates, and compute flow rate range
500
Sort clients by flow rate range
505
Initialize each client flow rate to it's minimum flow rate & mark as unprocessed
510
Establish remaining server bandwidth
520
For each client
Is server near saturation ?
530
Yes
Allocate remaining server bandwidth equally to each client
540
No
Allocate the amount of this client's remaining bandwidth to each client
550
Update total server flow rate by the total amount given unprocessed clients in 550
560
Mark this channel as processed
570

FIG._5

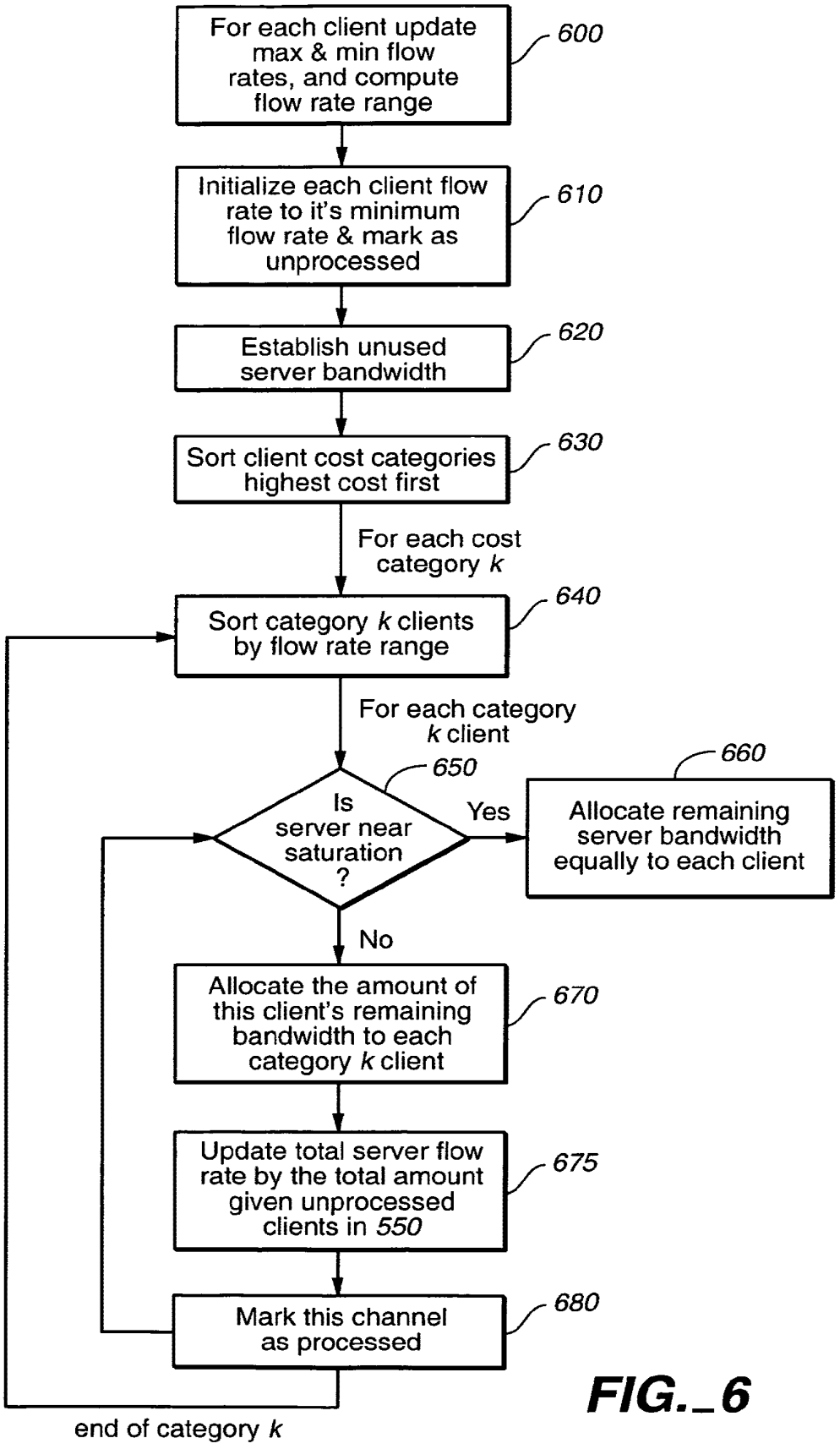

For each client update max & min flow rates, and compute flow rate range
600
Initialize each client flow rate to it's minimum flow rate & mark as unprocessed
610
Establish unused server bandwidth
620
Sort client cost categories highest cost first
630
For each cost category k
Sort category k clients by flow rate range
640
For each category k client
650
Is server near saturation ?
Yes
660
Allocate remaining server bandwidth equally to each client
No
Allocate the amount of this client's remaining bandwidth to each category k client
670
Update total server flow rate by the total amount given unprocessed clients in 550
675
Mark this channel as processed
680
end of category k

FIG._6

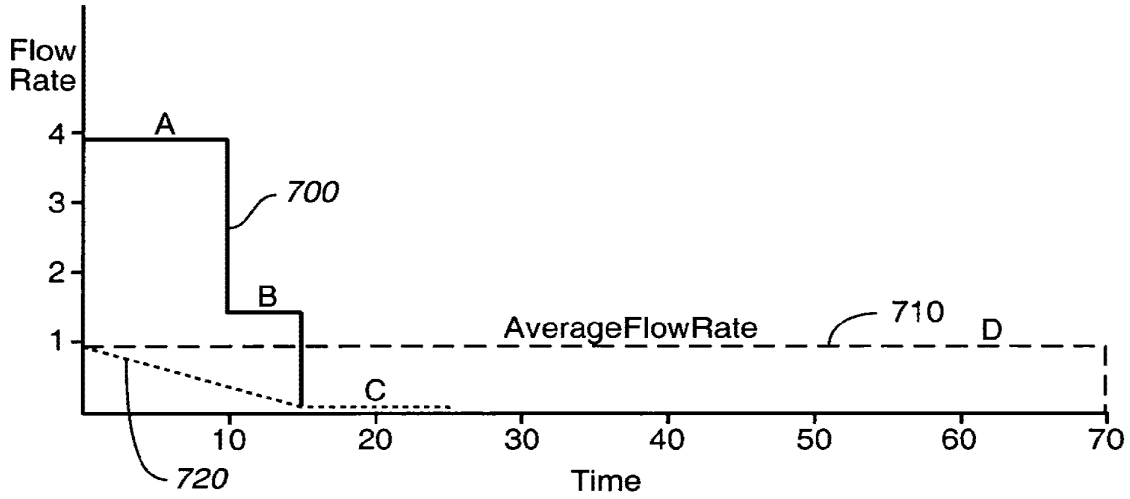

Flow Rate
4
3
2
1
A
B
C
D
700
710
720
AverageFlowRate
10
20
30
40
50
60
70
Time

FIG._7

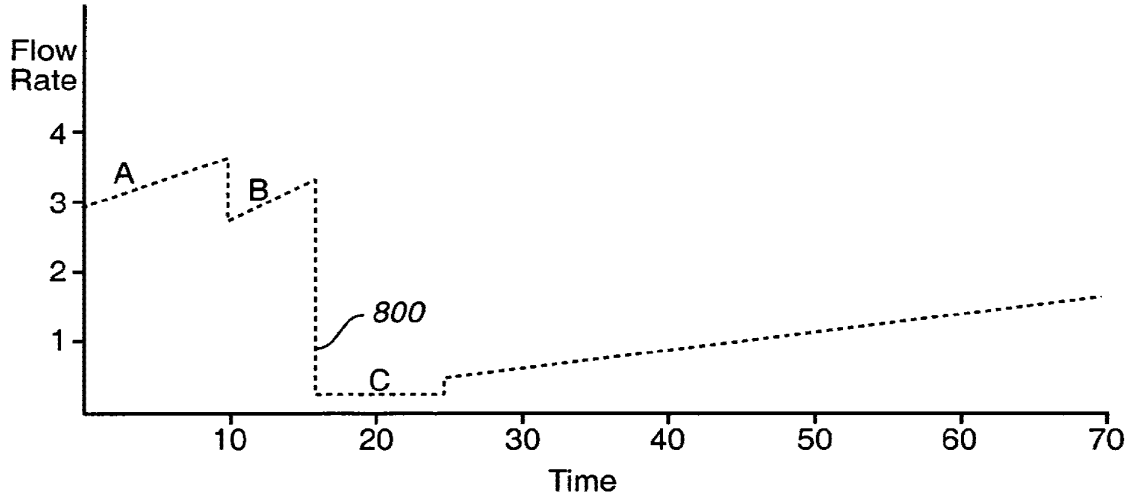

Flow Rate
4
3
2
1
A
B
C
800
10
20
30
40
50
60
70
Time

FIG._8

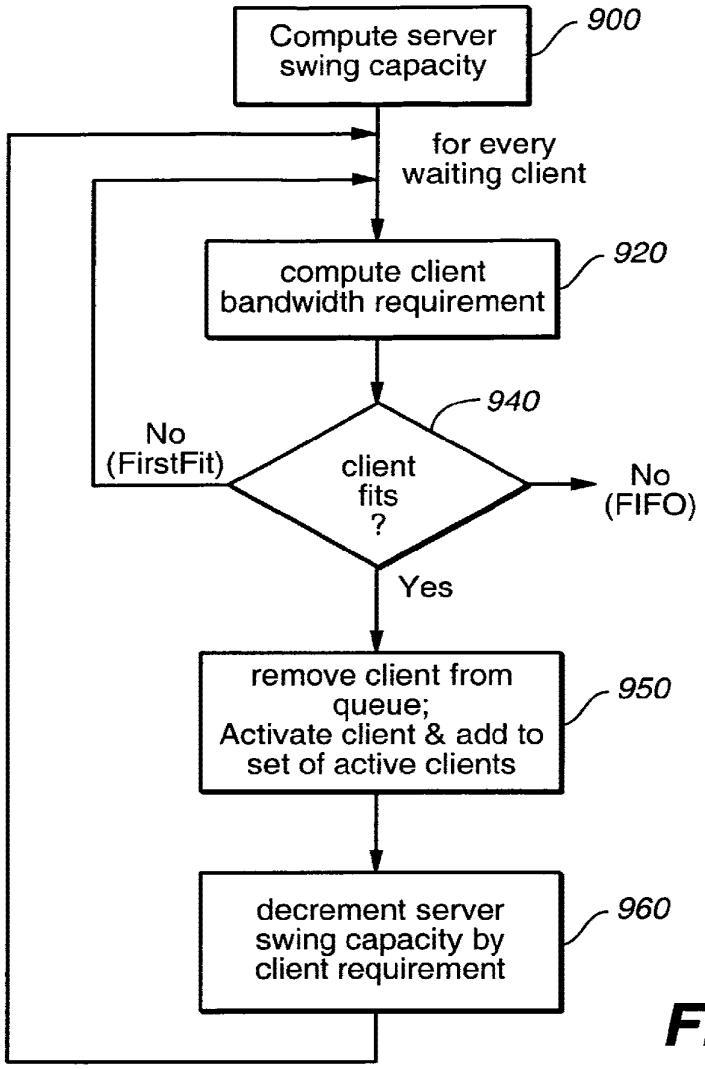

Compute server swing capacity
900
for every waiting client
compute client bandwidth requirement
920
940
client fits ?
No (FirstFit)
No (FIFO)
Yes
remove client from queue; Activate client & add to set of active clients
950
decrement server swing capacity by client requirement
960

FIG._9

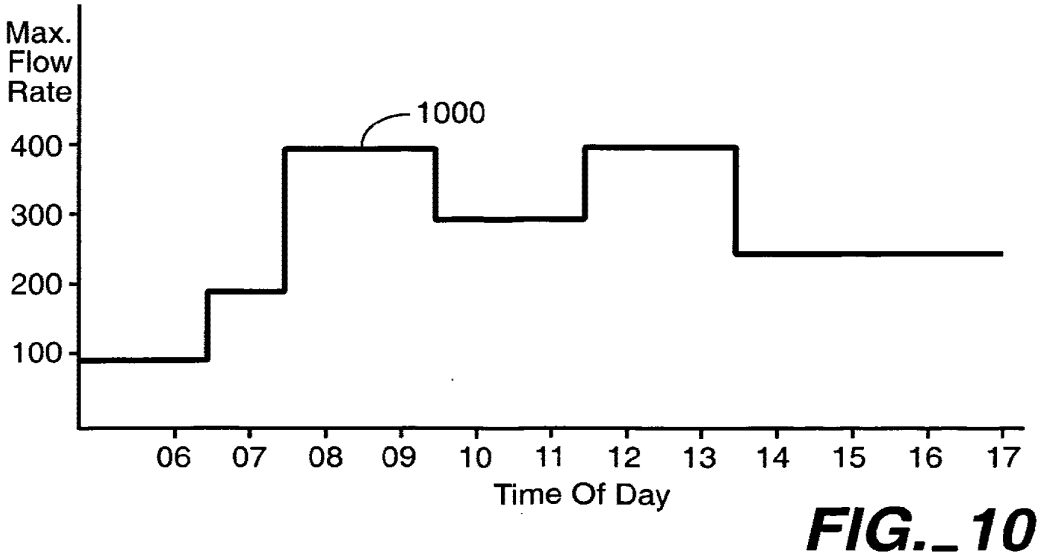

Max. Flow Rate
400
300
200
100
1000
06
07
08
09
10
11
12
13
14
15
16
17
Time Of Day

FIG._10

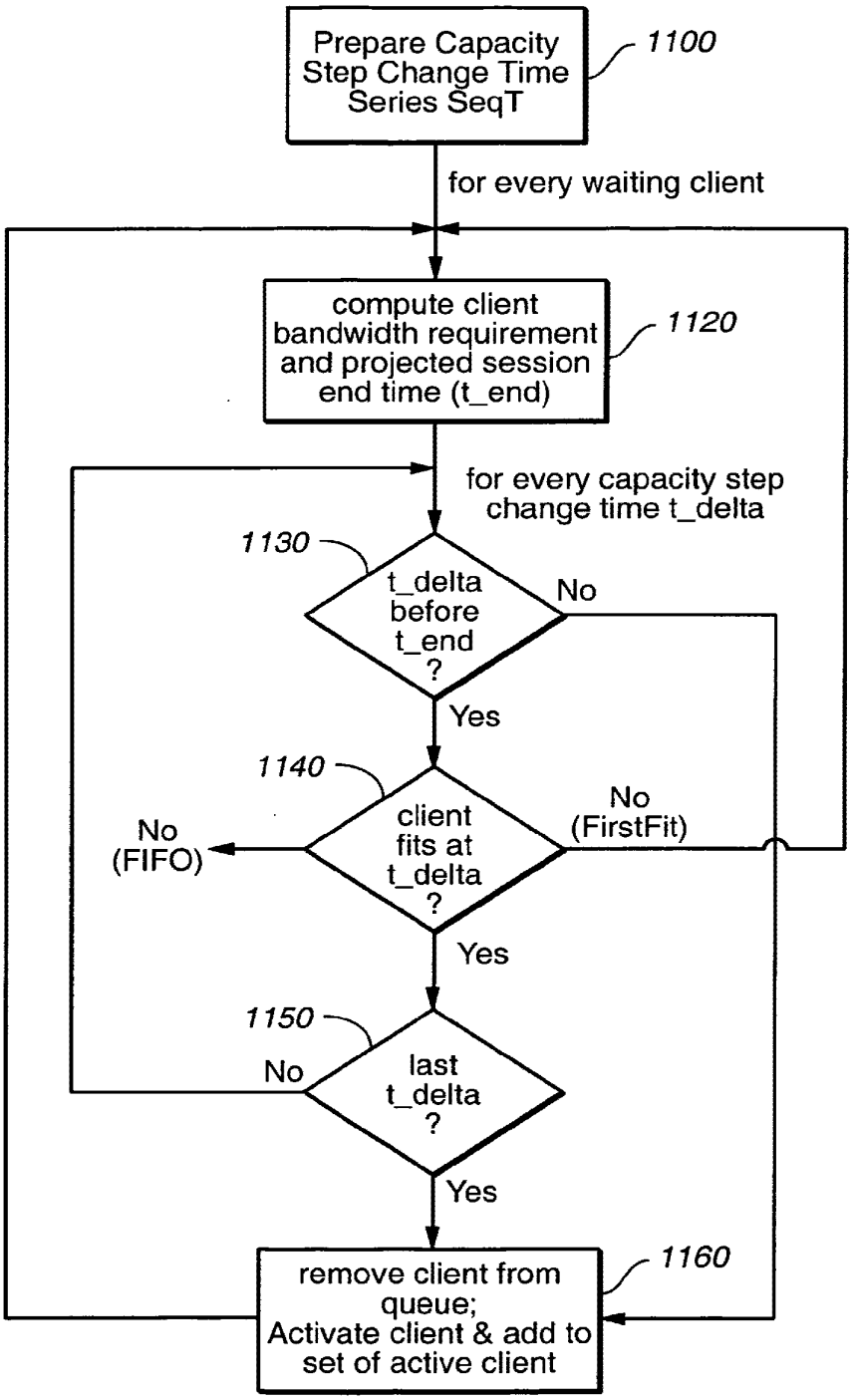

Prepare Capacity Step Change Time Series SeqT
1100
for every waiting client
compute client bandwidth requirement and projected session end time (t_end)
1120
for every capacity step change time t_delta
1130
t_delta before t_end ?
No
Yes
1140
client fits at t_delta ?
No (FIFO)
No (FirstFit)
Yes
1150
last t_delta ?
No
Yes
remove client from queue; Activate client & add to set of active client
1160

FIG._11

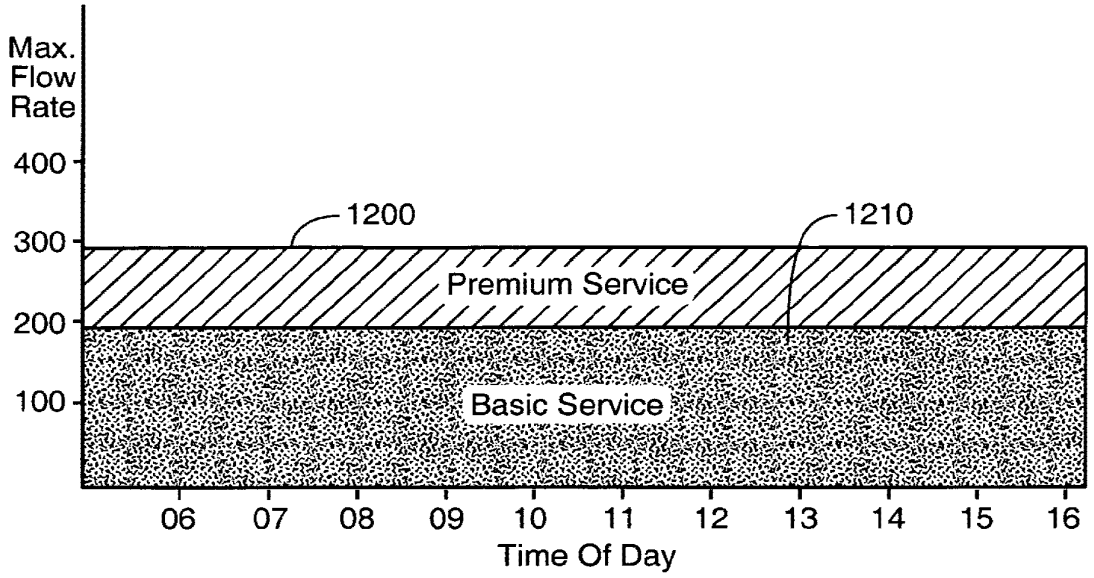

Max.
Flow
Rate
400
300
200
100
1200
1210
Premium Service
Basic Service
06
07
08
09
10
11
12
13
14
15
16
Time Of Day

FIG._12

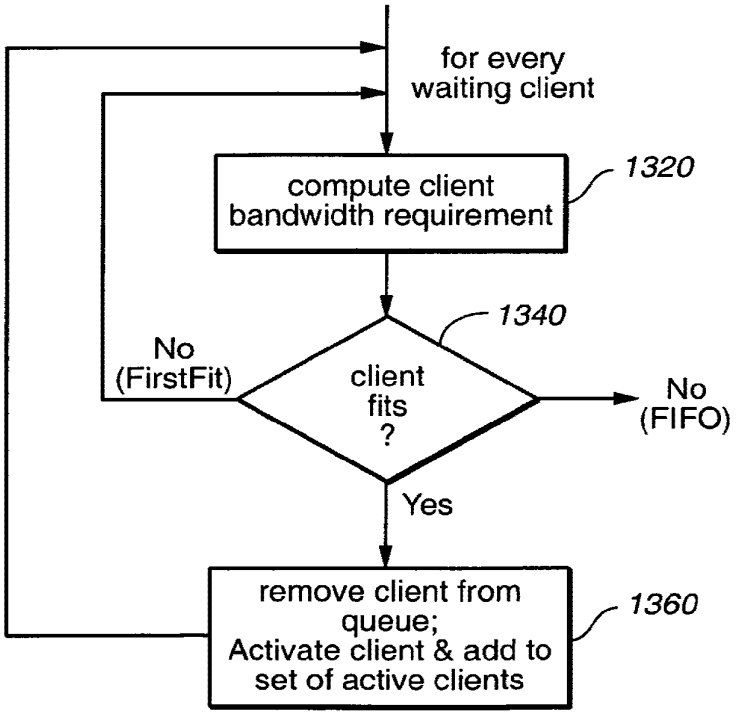

for every
waiting client
compute client
bandwidth requirement
1320
1340
No
(FirstFit)
client
fits
?
No
(FIFO)
Yes
remove client from
queue;
Activate client & add to
set of active clients
1360

FIG._13

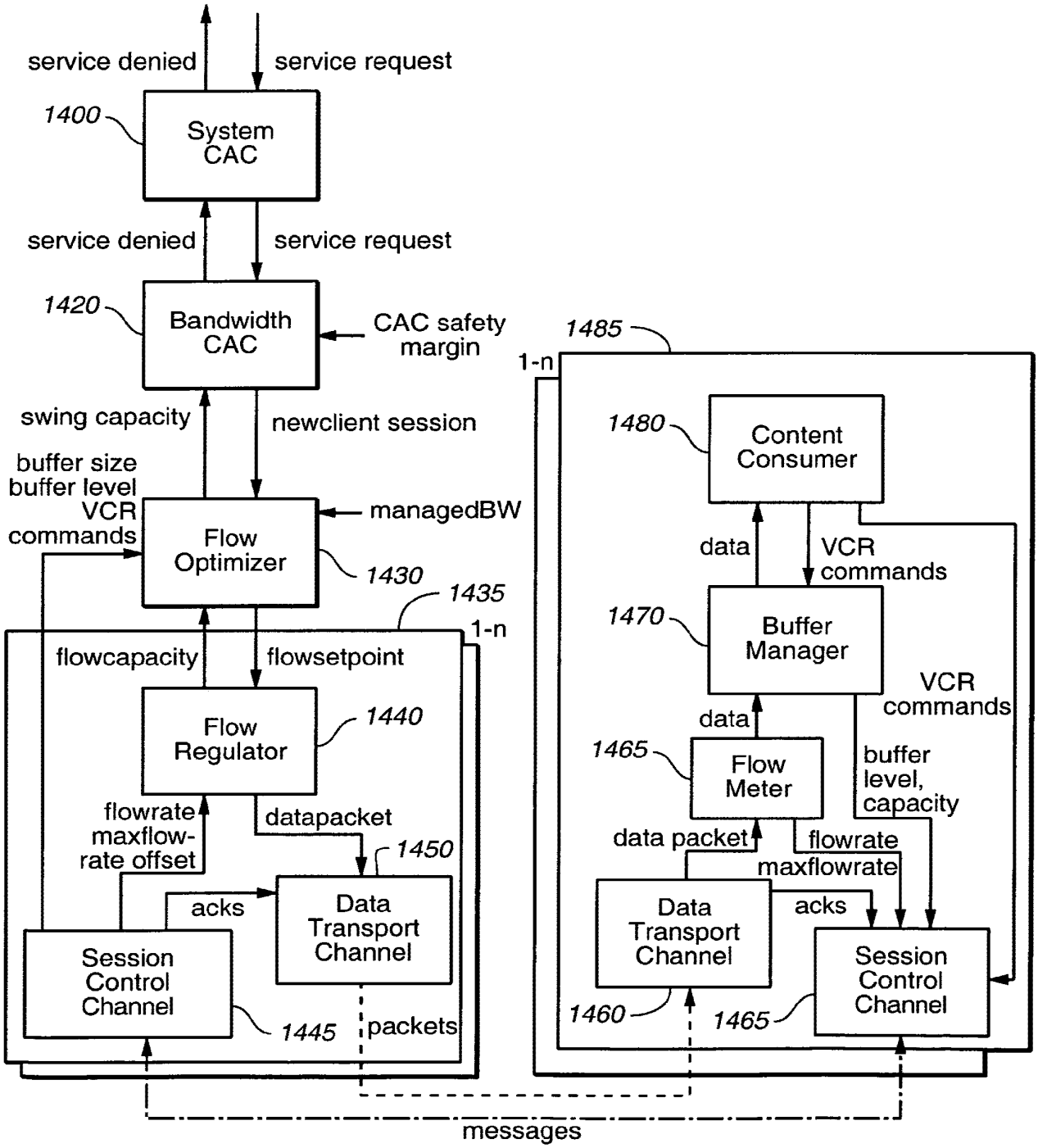

service denied
service request
1400
System CAC
service denied
service request
1420
Bandwidth CAC
CAC safety margin
swing capacity
newclient session
buffer size
buffer level
VCR commands
Flow Optimizer
managedBW
1430
1435
1-n
flowcapacity
flowsetpoint
Flow Regulator
1440
flowrate maxflow-rate offset
datapacket
1450
acks
Data Transport Channel
Session Control Channel
1445
packets
1485
1-n
1480
Content Consumer
data
VCR commands
1470
Buffer Manager
VCR commands
data
1465
Flow Meter
buffer level, capacity
data packet
flowrate maxflowrate
Data Transport Channel
acks
Session Control Channel
1460
1465
messages

FIG._14

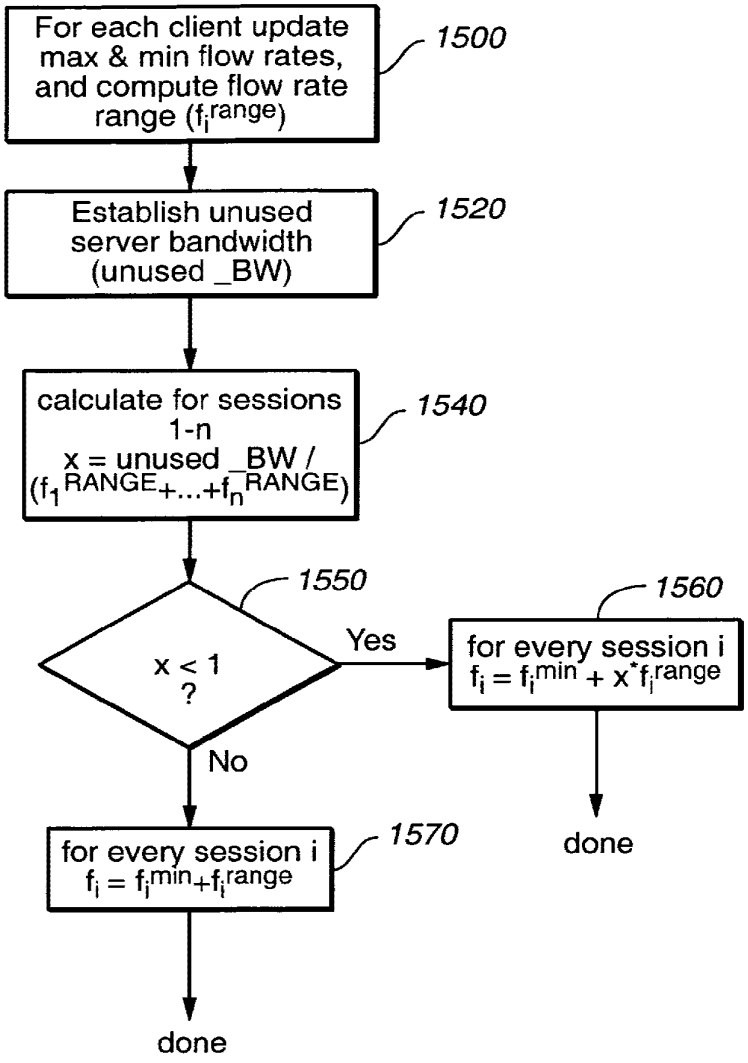

For each client update max & min flow rates, and compute flow rate range (f_i^{range})
1500
Establish unused server bandwidth (unused _BW)
1520
calculate for sessions 1-n
x = unused _BW / (f_1^{RANGE}+...+f_n^{RANGE})
1540
1550
x < 1 ?
Yes
No
1560
for every session i
$f_i = f_i^{min} + x^* f_i^{range}$
done
for every session i
$f_i = f_i^{min} + f_i^{range}$
1570
done

FIG._15

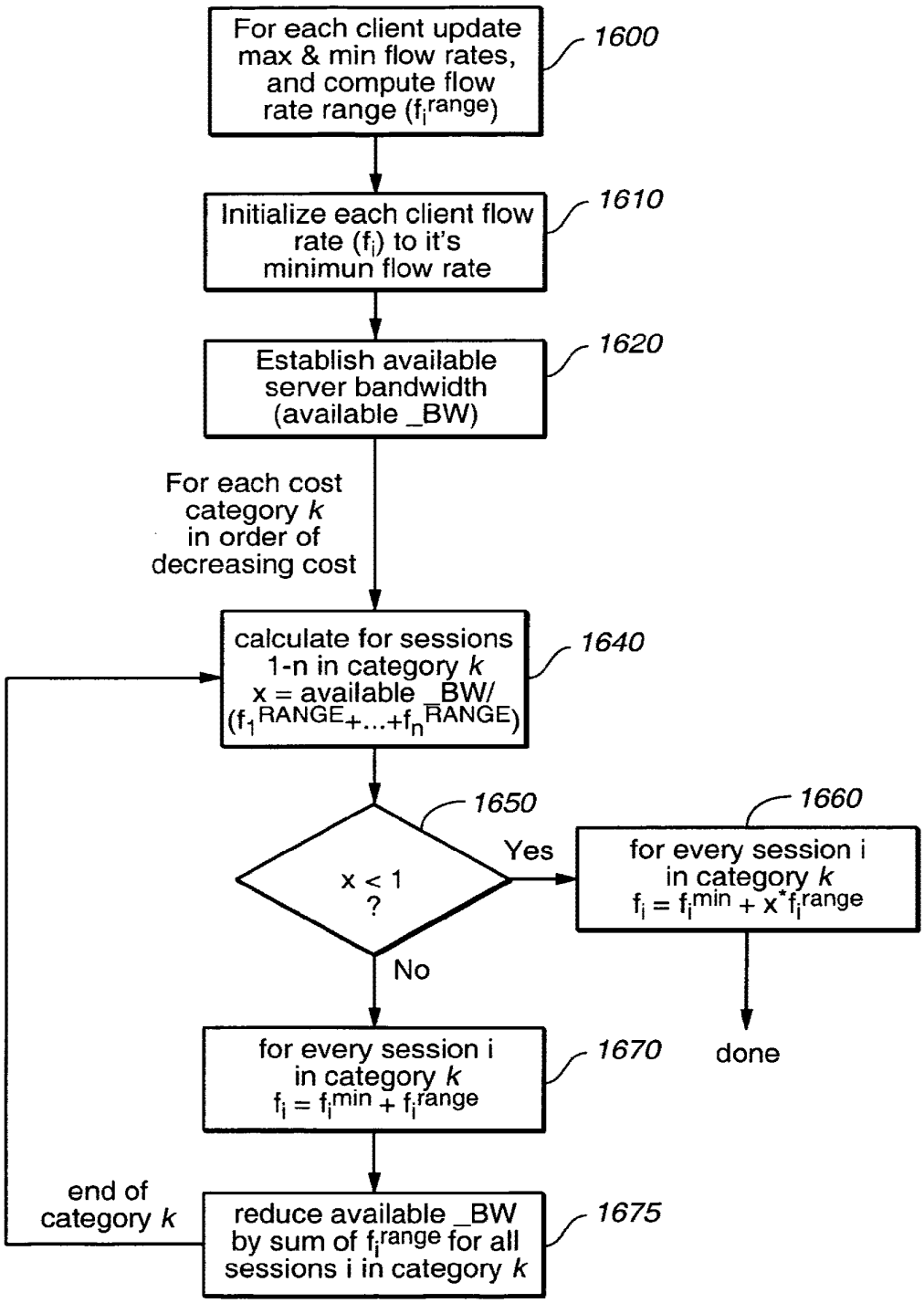

For each client update max & min flow rates, and compute flow rate range (f_i^{range})
1600
Initialize each client flow rate (f_i) to it's minimun flow rate
1610
Establish available server bandwidth (available _BW)
1620
For each cost category k in order of decreasing cost
calculate for sessions 1-n in category k x = available _BW/ (f_1^{RANGE}+...+f_n^{RANGE})
1640
1650
x < 1 ?
Yes
No
1660
for every session i in category k $f_i = f_i^{min} + x^* f_i^{range}$
done
for every session i in category k $f_i = f_i^{min} + f_i^{range}$
1670
end of category k
reduce available _BW by sum of f_i^{range} for all sessions i in category k
1675

FIG._16

Aylo Exhibit 1018
Page 1691 of 2308

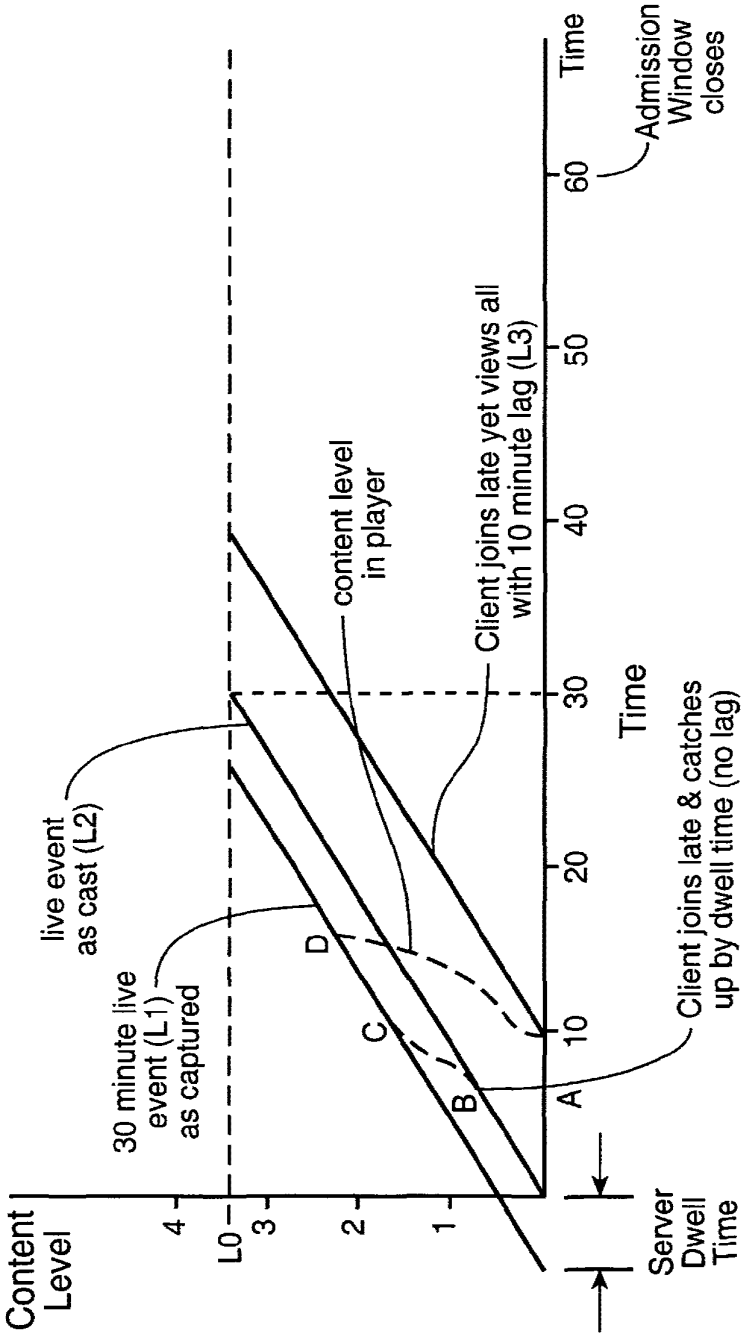


FIG._17

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2006 (26.01.2006)

PCT

(10) International Publication Number
WO 2006/010113 A2

(51) International Patent Classification:
G06F 15/173 (2006.01)

(21) International Application Number:
PCT/US2005/024517

(22) International Filing Date: 11 July 2005 (11.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/586,952 9 July 2004 (09.07.2004) US

(71) Applicant *(for all designated States except US)*: **NETWORK FOUNDATION TECHNOLOGIES, LLC** [US/US]; 2215 Cooktown Road, Ruston, LA 71270 (US).

(72) Inventors: **O'NEAL, Mike**; 2215 Cooktown Road, Ruston, LA 71270 (US). **TALTON, John P.**; 2215 Cooktown Road, Ruston, LA 71270 (US).

(74) Agents: **LANDA, Adam B.** et al.; Greenberg Traurig, MetLife Building, 200 Park Avenue, New York, NY 10166 (US).

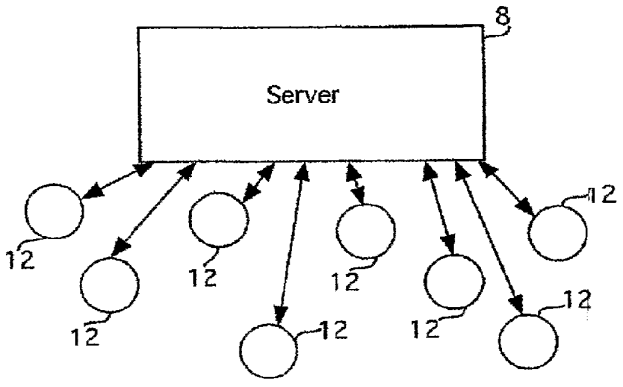
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEMS FOR DISTRIBUTING DATA OVER A COMPUTER NETWORK AND METHODS FOR ARRANGING NODES FOR DISTRIBUTION OF DATA OVER A COMPUTER NETWORK



PRIOR ART

(57) Abstract: Various embodiments of the present invention relate to a system for distributing data (e.g., content data) over a computer network and a method of arranging receiver nodes in a computer network such that the capacity of a server is effectively increased (e.g., the capacity of a server may be effectively multiplied many times over; the capacity of the server may be effectively increased exponentially). In one embodiment the present invention may take advantage of the excess capacity many receiver nodes possess, and may use such receiver nodes as repeaters. The distribution system may include node(s) having database(s) which indicate ancestor(s) and/or descendant(s) of the node so that reconfiguration of the distribution network may be accomplished without burdening the system's primary server. An embodiment of the present invention may include a process for configuring a computer information distribution network having a primary server node and user nodes docked in a cascaded relationship, and reconfiguring the network in the event that a user node departs from the network. In one example (which example is intended to be illustrative and not restrictive), the process may include the steps of providing a new user node (or connection requesting user node) with a connection address list of nodes within the network, having the new user node (or connection requesting user node) go to (or attempt to go to) the node at the top of the connection address list, determine whether that node is still part of the distribution network, and connect thereto if it is, and if it is not, to go to (or attempt to go to) the next node on the connection address list. In another example (which example is intended to be illustrative and not restrictive), when a user node departs from the distribution network, a propagation signal may be transmitted to the nodes below it in the network, causing them to move up in the network in a predetermined order. In another example (which example is intended to be illustrative and not restrictive), the present invention may provide a decentralized approach which provides, to each new user node (or connection requesting user node) a path back to the root server.

WO 2006/010113 A2

SYSTEMS FOR DISTRIBUTING DATA OVER A COMPUTER NETWORK AND METHODS FOR ARRANGING NODES FOR DISTRIBUTION OF DATA OVER A COMPUTER NETWORK

FIELD OF THE INVENTION

This patent application is related to and claims the benefit of the filing date of U.S. Provisional Patent Application Serial No. 60/586,952 for "Systems for Distributing Data Over A Computer Network And Methods For Arranging Nodes For Distribution Of Data Over A Computer Network", filed July 9, 2004.

Various embodiments of the present invention relate to systems for distributing data (e.g., content data) over a computer network and methods of arranging nodes for distribution of data (e.g., content data) over a computer network. In one example (which example is intended to be illustrative and not restrictive), the systems and methods of the present invention may be applied to the distribution of streaming audiovisual data over the Internet.

For the purposes of the present application the term "node" (e.g., as used in the phrase a first node communicates with a second node) is intended to refer to a computer system (or sometimes just "computer") configured to send and/or receive data over a computer network (e.g., a computer network including the Internet, a local-area-network, a wide-area-network, a wireless network). Since a "node" in a computer network would not exist but for a computer system being available on such node, the terms "node" and "computer system" (or sometimes just "computer") may be used interchangeably (i.e., the term "node" should be understood to include an available computer system (or sometime just "computer").

Further, for the purposes of the present application the term "upstream" (e.g., as used in the phrase a first computer system sends data via an upstream connection to a second computer system) is intended to refer to the communication path between a first computer system and the Internet when the first computer system is sending data to a second computer system via the Internet.

Further still, for the purposes of the present application the term "downstream" (e.g., as used in the phrase a first computer system receives data via a downstream connection from a second computer system) is intended to refer to the communication path between a first computer system and the Internet when the first computer system is receiving data from a second computer system via the Internet.

Further still, for the purposes of the present application the term "uptree" (e.g., as

used in the phrase a first node sends data via an uptree connection to a second node) is intended to refer to the network topology communication path between a first node and a second node when a first node is sending data to a second node which is higher up on the network topology tree (that is, closer to the root server).

Further still, for the purposes of the present application the term "downtree" (e.g., as used in the phrase a first node sends data via a downtree connection to a second node) is intended to refer to the network topology communication path between a first node and a second node when a first node is sending data to a second node which is lower down on the network topology tree (that is, closer to the leaves).

Further still, for the purposes of the present application the term "docked" (e.g., as used in the phrase a child node docks with a parent node) is intended to refer to forming a connection between nodes via which data may flow in at least one direction (e.g., at least uptree or downtree).

Further still, for the purposes of the present application the term "apparent available capacity" (e.g., as used in the phrase apparent available capacity to transmit content data) is intended to refer to a nominal capacity (e.g., a node's presumed capacity based upon a design specification but not necessarily actually tested in the network).

BACKGROUND OF THE INVENTION

In a computer network such as the Internet, each node in the network has an address. A computer system resident at a particular address may have sufficient bandwidth or capacity to receive data from, and to transmit data to, many other computer systems at other addresses. An example of such a computer system is a server, many commercial versions of which can simultaneously exchange data with thousands of other computer systems.

A computer system at another location may have only sufficient bandwidth to effectively exchange data with only one other computer system. An example of such a system is an end user's personal computer connected to the Internet by a very low speed dialup modem. However, even typical personal computers connected to the Internet by higher speed dialup modems may have sufficient bandwidth that they may exchange data essentially simultaneously in an effective manner with several other computer systems. Moreover, an end user's personal computer system may have even greater bandwidth when connected to the Internet by ISDN lines, DSL (e.g., ADSL) lines, cable modems, T1 lines or even higher capacity links. As discussed more fully below, various embodiments of the present invention

may take advantage of the availability of such higher capacity end user systems (e.g., those computer systems capable of essentially simultaneously exchanging data with multiple computer systems).

In a typical situation, as shown in FIG. 1, a content provider distributes its data by making the data available on a server node 8 simultaneously to a plurality of users at user nodes 12 (of note, the terms "server", "root server" and "primary server" may be used interchangeably throughout the present application to refer to the same device (i.e., the highest level parent node in a given network). The double-headed arrows show the two-way communication between each end user's computer system and the server. Essentially the content provider's server transmits a separate stream of signals to each receiver node. To accommodate additional users, the content provider would typically either add equipment to increase capacity or it would engage a mirror site to accomplish essentially the same result as adding equipment. The capacities of the end user computers is of virtually no consequence in such a system.

Another system for distributing data is the Napster™ music file exchange system provided by Napster, Inc. of Redwood City, Calif. A schematic of the Napster™ music file exchange system (as its operation is presently understood) is illustrated in FIG. 2.

More particularly, it is believed that in the Napster™ system a copy of the music data is not kept on the server. The server 9 instead maintains a database relating to the various music files on the computers of users who are logged onto the server 9. When a first user 12a sees that a desired music file is available from a second logged on user 12b, the first user causes his computer to query the server 9 for the second user's node address and a connection is made between the first and second user's computers through which the first user's computer notifies the second user's computer of the desired file and the second user's computer responds by transmitting a copy of the desired music file directly to the first user's computer. It is further believed that a first user attempting to download a particular file from a second user must start completely over again if the second user cancels its transmission or goes off line during the data transfer.

In another area, engineers have developed what is known as "streaming media." In summary, streaming media is a series of packets (e.g., of compressed data), each packet representing moving images and/or audio.

To help understand streaming media in more detail, it is helpful to review the traditional Internet distribution method. Each node (whether it is a server node or a user node) in a computer network has a unique identification (sometimes referred to as an "IP" address)

associated with it. On the Internet, the unique address may be referred to as a Uniform Resource Locator ("URL"). A user desiring to obtain data from a particular server enters that server's URL into the user's browser program. The browser program causes a connection request signal to be sent over the Internet to the server. If the server has the capacity to accept the connection, the connection is made between the server and the user node (files requested by the user are typically transmitted by the server in full to the user node and the browser program may store the files in buffer memory and display the content on the user's computer system monitor - some files may be more permanently stored in the computer system's memory for later viewing or playing.) The connection with the server is typically terminated once the files have been received at the user node (or the connection may be terminated a short time thereafter). Either way, the connection is usually of a very short time duration.

With streaming media, the contact between the server and user nodes is essentially continuous. When a connection between a server node and user node is made and streaming media is requested, the server sends streaming media packets of data to the user node. A streaming media player installed on the user's computer system (e.g., software, such as RealMedia™ from RealNetworks, Inc. of Seattle, Wash.,) causes the data to be stored in buffer memory. The player decompresses the data and begins playing the moving images and/or audio represented by the streaming media data on the user's computer system. As the data from a packet is played, the buffer containing that packet is emptied and becomes available to receive a new packet of data. As a result, the memory assets of a user's computer are not overly taxed. Continuous action content, such as, for example, the display of recorded motion picture films, videos or television shows may be distributed and played in essentially "real time," and live events, such as, for example, concerts, football games, court trials, and political debates may be transmitted and viewed essentially "live" (with only the brief delays needed for compression of the data being made available on the server, transmission from the server to the user node, and decompression and play on the user's computer system preventing a user from seeing the event at the exact same moment in time as a person actually at the event). And, when the systems are working as designed, the server node and user node may stay connected to each other until all the packets of data representing the content have been transmitted.

SUMMARY OF THE INVENTION

Various embodiments of the present invention relate to a system for distributing data (e.g., content data) over a computer network and a method of arranging receiver nodes in a computer network such that the capacity of a server is effectively increased (e.g., the capacity of a server may be effectively multiplied many times over; the capacity of the server may be effectively increased exponentially). In one embodiment the present invention may take advantage of the excess capacity many receiver nodes possess, and may use such receiver nodes as repeaters. The distribution system may include node(s) having database(s) which indicate ancestor(s) and/or descendant(s) of the node so that reconfiguration of the distribution network may be accomplished without burdening the system's primary server. An embodiment of the present invention may include a process for configuring a computer information distribution network having a primary server node and user nodes docked (i.e., connected) in a cascaded relationship, and reconfiguring the network in the event that a user node departs from the network. In one example (which example is intended to be illustrative and not restrictive), the process may include the steps of providing a new user node (or connection requesting user node) with a connection address list of nodes within the network, having the new user node (or connection requesting user node) go to (or attempt to go to) the node at the top of the connection address list, determine whether that node is still part of the distribution network, and connect thereto if it is, and if it is not, to go to (or attempt to go to) the next node on the connection address list. In another example (which example is intended to be illustrative and not restrictive), when a user node departs from the distribution network, a propagation signal may be transmitted to the nodes below it in the network, causing them to move up in the network in a predetermined order. In another example (which example is intended to be illustrative and not restrictive), the present invention may provide a decentralized approach which provides, to each new user node (or connection requesting user node) a path back to the root server.

Of note, various embodiments of the present invention may be applied to the transmission (e.g., the "appointment" transmission) of audiovisual content such as, for example (which example is intended to be illustrative and not restrictive), live or pre-recorded concerts, football games, court trials, political debates, motion picture films, videos or television shows. Such transmission of audiovisual content may be made with acceptable levels of quality, wherein "acceptable levels of quality" may vary depending upon the end users and the type of transmission. Further, such transmission of audiovisual content may be

carried out using streaming media transmissions intended to reach large audiences, in much the way that television shows transmitted over television cable and broadcast media reach large audiences (in this regard, the present invention may enable a server to transmit streaming media to the large number of users which would be essentially simultaneously logging on to view a particular audiovisual presentation under this television-type "large audience" transmission).

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic drawing of a prior art computer information distribution network;

FIG. 2 is a schematic drawing of another prior art computer information distribution network;

FIG. 3 is a schematic drawing of an embodiment of a computer information distribution network formed pursuant to the present invention;

FIG. 4 is a schematic drawing of another embodiment of a computer information distribution network formed pursuant to the present invention;

FIG. 5 is a schematic drawing of another embodiment of a computer information distribution network formed pursuant to the present invention;

FIG. 6 is a schematic drawing of a particular topology of a computer information distribution network formed pursuant to an embodiment of the present invention;

FIG. 7 is a schematic drawing of a particular topology of the computer information distribution network formed pursuant to an embodiment of the present invention as shown in FIG. 6 (after the occurrence of an event);

FIG. 8 is a schematic drawing of another topology of a computer information distribution network formed pursuant to an embodiment of the present invention;

FIG. 9 is a schematic drawing of the topology of the computer information distribution network formed pursuant to an embodiment of the present invention as shown in FIG. 8 (after the occurrence of an event);

FIG. 10 is a flow diagram of an embodiment of the present invention showing a Server Connection Routine which may be performed when a prospective child node seeks to join the distribution network;

FIG. 11 is a flow diagram of an embodiment of the present invention showing a Prospective Child Node's Request to Server for a Connection Routine;

FIGS. 12A-12E are schematic drawings showing varying topologies of the computer information distribution network formed pursuant to an embodiment of the present invention under several circumstances;

FIG. 13 is a block diagram of an embodiment of the present invention showing the memory blocks into which the software used may partition a user node's memory;

FIG. 14 is a flow diagram of an embodiment of the present invention showing a Prospective Child Node's Connection Routine - the routine which a new user node (or connection requesting user node) may go through in attempting to connect to a distribution chain (or tree) after receiving a connection address list;

FIG. 15 is a flow diagram of an embodiment of the present invention showing a Prospective Child Node's Connection Routine With Return to Server Subroutine;

FIG. 16 is a flow diagram of an embodiment of the present invention illustrating a Prospective Parent Node's Connection Routine;

FIG. 17 is a flow diagram of an embodiment of the present invention illustrating a Server's Connection Instruction Routine;

FIG. 18 is a flow diagram of an embodiment of the present invention illustrating a Fully Occupied Parent's Connection Instruction Routine;

FIG. 19 is a flow diagram of an embodiment of the present invention illustrating a Multi-Node Selection Subroutine;

FIGS. 20A and 20B are together a flow diagram of an embodiment of the present invention illustrating a Universal Connection Address Routine;

FIG. 21 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention before a new node will be added using the Universal Connection Address Routine;

FIG. 22 is a schematic drawing of a topology of the computer information distribution network of an embodiment of the present invention when a new node is added using the Universal Connection Address Routine;

FIG. 23 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention before a reconfiguration event;

FIG. 24 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention shown in FIG. 23 after a reconfiguration event;

FIG. 25 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention, slightly different from the topology shown in FIG. 23, before a reconfiguration event;

FIG. 26 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention shown in FIG. 25 after a reconfiguration event;

FIG. 27 is a flow diagram of an embodiment of the present invention showing a Child Node's Propagation Routine;

FIG. 28 is a schematic drawing of another topology of a computer information distribution network of an embodiment of the present invention before a reconfiguration event;

FIG. 29 is a schematic drawing of a topology of a computer information distribution network of an embodiment of the present invention shown in FIG. 28 after a reconfiguration event;

FIG. 30 is a schematic drawing of another topology of a computer information distribution network of an embodiment of the present invention before a "complaint" regarding communications;

FIG. 31 is a flow diagram of an embodiment of the present invention showing a Grandparent's Complaint Response Routine.

FIGS. 32, 33A-33C, 34A-34C, 35A, 35B and 36 relate to "Registration Process" examples according to the present invention;

FIGS. 37 and 38 relate to a "Firewall Request / Firewall Status" example according to the present invention;

FIGS. 39A, 39B and 40 relate to a "Propagate Path / Down" example according to the present invention;

FIGS. 41A, 41B and 42 relate to a "Propagate NTM / Up" example according to the present invention;

FIGS. 43 and 44 relate to a "Banned List" example according to the present invention;

FIG. 45 relates to a "Node Interlacing" example according to the present invention; and

FIGS. 46A-46C relate to various distribution network topology examples according to the present invention.

Among those benefits and improvements that have been disclosed, other objects and advantages of this invention will become apparent from the following description taken in conjunction with the accompanying figures. The figures constitute a part of this specification and include illustrative embodiments of the present invention and illustrate various objects and features thereof (of note, the same reference numeral may be used to identify identical elements throughout the figures).

DETAILED DESCRIPTION OF THE INVENTION

Detailed embodiments of the present invention are disclosed herein; however, it is to be understood that the disclosed embodiments are merely illustrative of the invention that may be embodied in various forms. In addition, each of the examples given in connection with the various embodiments of the invention are intended to be illustrative, and not restrictive. Further, the figures are not necessarily to scale, some features may be exaggerated to show details of particular components. Therefore, specific structural and functional details disclosed herein are not to be interpreted as limiting, but merely as a representative basis for teaching one skilled in the art to variously employ the present invention.

Referring now to FIG. 3, an illustration of a linear propagation arrangement 10 of a computer network according to an embodiment of the present invention is shown. In this embodiment, the primary server node (or simply, server) 11 provides content data (e.g., streaming media) to user nodes 12 connected directly to it (sometimes referred to as "first level user nodes"). Each first level user node 12 has a second level user node 13 connected to it and each second level user node 13 has a third level user node 14 connected to it. The computer system at each first level user node 12 passes a copy of the content data received from server node 11 to the computer system at the second level user node 13 attached to such first level user node 12. The computer system at each second level user node 13 in turn passes the content data onto the computer system at the fourth level user node 14 attached to it.

As more fully discussed below, under this embodiment the computer systems at the server and user nodes may have distribution software installed in them which enables the nodes to be arranged as shown and for the computer systems to receive and re-transmit data.

The cascadingly connected arrangement of nodes (i.e., first level nodes are connected to the server, second level nodes are connected to first level nodes, third level nodes are connected to second level nodes and so on) shown in FIG. 3 takes advantage of the bandwidth available in certain nodes to essentially simultaneously receive and transmit data.

In a linear propagation arrangement, the effective distribution capacity of a server is in essence multiplied by the number of levels of nodes linked together. In the example of FIG. 3 (which example is intended to be illustrative and not restrictive), the distribution capacity of the server node is increased from 8 user nodes to 24 in just three levels.

Of note, many user nodes may have at least sufficient bandwidth (e.g., upstream bandwidth as well as downstream bandwidth) to receive data from one node and to re-transmit streams of data essentially simultaneously to two or more other nodes. This capacity could be used in another embodiment to set up a computer network with a cascadingly connected exponential propagation arrangement 16 (as shown in FIG. 4). As the name implies, an exponential propagation arrangement effectively increases the distribution capacity of a server exponentially. For example (which example is intended to be illustrative and not restrictive), with just three levels of user nodes, each having the capacity to retransmit two data streams, the distribution capacity of the server in FIG. 4 is increased from 8 user nodes to 56.

In another embodiment a distribution network may also be set up as a cascadingly connected hybrid linear/exponential arrangement 18, such as shown in FIG. 5.

Of note, the effective distribution capacity grows more quickly in a hybrid linear/exponential arrangement with each new level than does the distribution capacity in a linear propagation arrangement, and less quickly than does the distribution capacity in a pure exponential arrangement. However, any of these arrangements allows a server's distribution capacity to be greatly increased (e.g., with little or no additional investment in equipment for the server). Further, any other desired hybrid arrangement may be used.

Referring again to FIGS. 1-5, all of the connections between nodes are illustrated with arrows on each end. This is intended to signify that data flows in both directions between connected nodes. For example (which example is intended to be illustrative and not restrictive), a user node connected to a server may transmit data to the server indicating the identity of the user node, what the user node wants and/or other data, while the server node may transmits data confirming its identity and containing information and other content to the user node. In the remaining drawings such arrows may not be shown for the sake of simplicity. Further, it is noted that the bulk of data which is transmitted may be content (e.g., web pages, music files, streaming media), and such content will be understood to flow downtree from node to node in the drawings (i.e., in the direction from the root server to a first level node, to a second level node, to a third level node, etc.).

As can be understood from the above, an exponential propagation arrangement may create the most distribution capacity. However, while many (and perhaps even most) user nodes may have sufficient bandwidth to re-transmit acceptable quality multiple copies of the data, a number of user nodes may not have sufficient bandwidth to re-transmit more than one copy of the data with acceptable quality, and some user nodes may not have sufficient bandwidth to re-transmit even a single acceptable copy. Under these conditions, a distribution system employing the present invention may be configured, for example, as a hybrid propagation network. In such a system, a personal computer acting as a server node may reach many (e.g., hundreds, thousands or more) user nodes, even if the server node itself has capacity to transmit content data directly to only one other node.

In the following discussion it will be assumed that the system employing the present invention will take advantage of the capability of many user nodes to simultaneously re-transmit up to two copies of data (e.g., content data). However, it should be understood that the invention could take advantage of user nodes capable of simultaneous re-transmission of even higher numbers of copies.

In one embodiment the length of the distribution chains (i.e., the number of levels of user nodes linked through each other to the server) may be kept as small as possible to reduce the probability that the nodes at the ends of the chains will suffer a discontinuity of service. Thus, under this embodiment, in order to maximize the number of user nodes which may be connected to the server while trying to keep the length of the distribution chains as small as possible, the user nodes having the greatest bandwidth may be placed as far up uptree as possible (i.e., be placed as close as possible to the server, where "close" refers to network node level and not necessarily to geographical proximity).

In this regard, if an unreliable user node is placed near or at the end of a chain, few or no other user nodes would be affected by the unreliable user node's leaving the network. In contrast, if an unreliable user node were placed at or near the beginning of a chain (i.e., be far uptree), many other nodes would be affected by the unreliable node's departure. Because various embodiments of the invention may be used for appointment viewing of streaming media, in which an essentially continuous connection to an audiovisual content source may essentially be a necessity, it may be important (e.g., for this steaming media use) that the most reliable user nodes be placed high up in the chain (i.e., high uptree).

A user node may be deemed unreliable for any of a number of reasons. One example (which example is intended to be illustrative and not restrictive), is that the user node is connected to the Internet by lines having intermittent failures. Another example (which

example is intended to be illustrative and not restrictive), is that the user is merely sampling the content available on the network and intentionally disconnects the user node from the distribution network after discerning that he or she has no interest in the content.

Thus, under an embodiment of the present invention the user nodes may be positioned in the most advantageous positions, taking into account the dynamic nature of the distribution network, in which many user nodes may enter and leave the distribution network throughout the server's transmission of a streaming media show. In addition, the invention may help to preserve the viewing experience of users at user nodes positioned even at the end of a distribution chain (the server and/or user nodes may be enabled to perform the operations required to set up and maintain the distribution network by having data distribution software installed therein).

Referring now to a "User Node Arrival" example (which example is intended to be illustrative and not restrictive), it can be seen from FIGS. 3-5 that the distribution chains can be viewed as a plurality of family trees, each family tree being rooted to the server 11 through one of the first level nodes 12 (each node in the distribution network may have distribution software loaded in it which enables the node to perform the functions described below; before any new node may join the distribution network, such new node may also have such software loaded in it).

Of note, in the bulk of the examples discussed below, the network topologies have a branch factor of two (i.e., no user node is assigned more than two child nodes to be connected directly to it). A network topology with a branch factor of two may be referred to as a binary tree topology. It should be understood, of course, that the teachings set forth herein may be extended to network topologies having other branch factors, for example (which example is intended to be illustrative and not restrictive), branch factors of three, four or more.

In any case, FIG. 6 is a schematic drawing of a distribution network topology useful for describing the current "User Node Arrival" example. This FIG. 6 shows a distribution chain or family tree rooted to the server 11 through user node A, a first level user node 12 (the dashed lines represent connections from the server to other first level nodes or from node A to another user node). User node A could be thought of as a child node of the server and as a parent node for other user nodes connected directly to it. User node B, a second level user node 13, could be thought of as A's child. User nodes C and D, third level user nodes 14, may be thought of as B's children and A's grandchildren (and each other's siblings). User nodes E and F, fourth level user nodes 15, may be thought of as C's children (and each other's

siblings). User node G, also a fourth level user node 15, may be thought of as D's child. And user nodes E, F and G may be thought of as B's grandchildren and A's great grandchildren.

In the present example, whenever a new user node (or connection requesting user node) 19, such as node X in FIG. 6, seeks connection to the distribution network, it will first make a temporary connection to the server node (or a connection server, not shown) in order to begin the process for connecting to the distribution system. The server (or connection server) will discern from the user node a bandwidth rating (discussed below) appropriate to that node and, depending upon the available capacity of the server and any existing distribution chains, the server will either assign the new user node to a spot directly connected to the server or will provide the new user node with a connection path through a tree to the server.

FIG. 10 is a flow diagram associated with this example showing a Server's Connection Routine which is performed when a prospective child node seeks to join the distribution network. In step 101 the server performs the Server's Connection Instruction Routine (discussed below), in which the server determines what connection instructions to give to the new user node (or connection requesting user node). The server then goes to step 102 where it determines whether, as a result of the Server's Connection Instruction Routine, the prospective child node is being instructed to dock with the server. If the prospective child node is being instructed to dock with the server, then the server goes to step 103 in which the server would allow the new user node to dock with it, and the server would begin transmitting data (e.g., streaming media) directly to the new user node.

Of note, as referred to above, two different servers could be used - one to perform the server's connection routine and the other to transmit data (e.g., streaming media) - since both servers would be performing server functions, they will hereinafter sometimes be considered a single server for purposes of the description herein.

Of further note, while the prospective child node may be a new user node or a connection requesting user node, to simplify the following discussion such prospective child node may simply be referred to as a "a new user node".

In any case, if the new user node is not being instructed to dock directly with the server, then the server goes to step 104 in which it provides the new user node with an address connection list and disconnects the new user node from the server.

Essentially contemporaneously with the performance by the server of the Server's Connection Routine, the new user node is performing the Prospective Child Node's Request to Server for a Connection Routine. FIG. 11 is a flow diagram associated with this example

illustrating the Prospective Child Node's Request to Server for a Connection Routine. Upon making the temporary connection to the server, the new user node goes to step 111 in which it provides bandwidth rating and performance rating information to the server. It then proceeds to step 112 in which it receives connection instructions from the server. Then the new user node proceeds to step 113 to determine whether it has been instructed to dock directly with the server. If the answer is "yes," then the new user node proceeds to step 114 in which it erases any information which may have been in its ancestor database and, if the distribution software has a Return to Server Subroutine in it, resets the connection attempt counter to zero. The new user node then proceeds to step 115 in which it docks with the server and begins receiving data (e.g., streaming media). If the new user node is not being instructed to dock directly with the server, then the new user node goes to step 116 in which it receives the new connection address list from the server and loads such list into the user node's ancestor database and begins the Prospective Child Node's Connection Routine (discussed below). If the distribution software has a Return to Server Subroutine in it, the connection attempt counter is reset to one in step 116.

In the examples discussed in connection with FIGS. 6-9 and 12A-12C (which examples are intended to be illustrative and not restrictive), the server has determined that the new user node X will not be allowed to connect directly to the server. Also, for the purposes of these examples, all of the user nodes are presumed to be equally capable of simultaneously re-transmitting two copies of the content data and that the tree rooted through node A is the most appropriate tree through which node X should be connected to the server. In one embodiment the server may rely on chain length in determining to which particular user node, already in the distribution network, that node X should connect.

Referring again to the example of FIG. 6, the topology of the distribution network based on the most recent information available to the server at the moment that node X seeks to join the distribution network is shown. Path D-B-A-S (where "S" is the server) represents the shortest available path from an end of a chain back to the server (e.g., from a node level point of view and not necessarily from a geographical proximity point of view), and the server 11 gives that path information, or connection address list, to node X during step 104 of FIG. 10 (and node X receives such list during step 116 of FIG. 11). That is, node X will be given a connection address list with the URLs and/or IP addresses of each of nodes D, B, A and S. The distribution software in node X causes the path information to be stored in the ancestor portion (or ancestor database) 132 of node X's topology database 131 shown in FIG. 13. The ancestor database may include an ancestor list, a list of node X's ancestors' addresses

from node X's parent back to the server node. (FIG. 13 shows an example block diagram (which example is intended to be illustrative and not restrictive) showing the memory blocks into which the software used in connection with the present invention may partition a user node's memory 130.) Node X then attempts to contact node D first, the user node most distant from server 11 in the path. Note that when the server provides the D-B-A-S connection address to node X, the server is giving what it "believes" to be the complete path information going all the way back to the server. That is, subsequent to the most recent generation of the topological data, node D may have departed from the network, as may have one or more of its ancestors, resulting in a reconfiguration (discussed below) of at least a portion of the tree of which D was a part.

Referring now to FIG. 14 is a flow diagram associated with this example showing the Prospective Child Node's Connection Routine, the routine which a new user node, here node X, will go through in attempting to connect to a distribution chain (or tree) after receiving a connection address list (which node X has stored in the ancestor portion of its topology database) from the server or from a prospective parent node or during a reconfiguration event. In step 141, node X attempts to contact the first node on the connection address list. The first node, and only node, on the connection address list could be the server itself. Here node D is the first node on the list. Node X then proceeds to step 142 and determines whether the first node on the connection address list is still on-line and still part of the distribution network (in one example, if no response is received within a predetermined period of time, from the first node on the connection address list, the answer to the query in step 142 will be deemed to be no). If node D is on-line and still part of the distribution network, node X proceeds to step 143 in which node X inquires whether node D has room for node X. This inquiry may need to be made because the distribution network may have gone through a reconfiguration event resulting in node D's not having sufficient capacity to provide a copy of the content data to node X. If node D has room for node X, then node X proceeds to step 144 in which it connects (or docks) with node D and begins receiving content data from it. This is depicted in FIG. 7. Note that node X is now one of several level four nodes 15.

On the other hand, in step 142, if node D is not on-line (e.g., no response is received from node D within a predetermined period of time) or if node D is on-line but is no longer part of the distribution system (e.g., subsequent to the server's obtaining its topology data the user of the system at node D either caused his or her computer system to go off-line or to leave the distribution system, or there was a line failure causing the computer system to go off line), as depicted in FIG. 8, then node X goes to step 145 in which it deletes the first

address from the connection address list in node X's ancestor database (and, in one example, and for a purpose which will become clear when discussing reconfiguration events below, sets its reconfiguration flag buffer 136 to the lowest propagation rating). At this time node B, in the present example, becomes the first node on the connection address list. Then node X goes back to step 141 and repeats the routine described above. Note that because of node D's leaving the distribution network, a reconfiguration event was triggered which resulted in node G changing from a fourth level node 15 to a third level node 14.

In step 143, if the prospective parent node has no room for the new node (e.g., the capacity of the prospective parent node is fully occupied), the new node goes to step 146, in which it receives a new connection address list from the prospective parent. The prospective parent may then perform a Fully Occupied Parent's Connection Instruction Routine, discussed below in connection with FIG. 18, wherein it creates the new connection address list based on topology data obtained from its progeny. That new list may include the path back to the server through the prospective parent just in case, as discussed above, there are user node departures along the path.

Of note, by having a prospective parent node prepare the new connection address list as described in this example, the burden on the server is reduced and is distributed among the user nodes.

Still referring to the example depicted in FIG. 8, when node X gets to step 143, node B will respond that it has no room and node X will proceed to step 146. When node X performs step 146, the new connection address list it receives from node B will be G-B-A-S. Then node X proceeds to step 141 and repeats the routine from that point on. When the routine is performed on the topology shown in FIG. 8, step 143 will result in node G responding that it has room for node X. Node X will then perform step 144 and be connected to the distribution network through node G as shown in FIG. 9. Here, node X becomes a fourth level node 15.

The distribution software may include a Return to Server Subroutine, comprised of steps 151, 152 and 153 as shown in FIG. 15, as part of the Prospective Child Node's Connection Routine. This subroutine reduces the risk that a prospective child node would enter into an endless cycle of fruitless attempts to dock with nodes in a particular tree. If the answer to the query in step 143 is "no," then node X goes to step 151 in which it increments by one the connection attempt counter 135 in node X's memory. Then node X goes to step 152 in which it determines whether the connection attempt limit has been reached. In one example (which example is intended to be illustrative and not restrictive), the limit may be

preset at any number greater than one and may depend upon what the designer of a particular distribution network determines would be a reasonable amount of time for a node to attempt to make a connection on a particular tree, or a branch of a tree, before that node should be given an opportunity to obtain a completely new connection address list directly from the server. If the connection attempt limit has not been reached, then node X proceeds with step 146 as discussed above. If the connection attempt limit has been reached, then node X goes to step 153, in which it goes back to the server and begins the connection routine again as discussed above in connection with FIG. 6. If docking with a parent node is successful, then after step 144 node X performs step 154 in which the connection attempt counter is set to zero.

Referring now to FIG. 16, a flow diagram associated with this example illustrating the Prospective Parent Node's Connection Routine is shown. Using the example illustrated in FIGS. 6, 8 and 9, when node X queries node B in step 142 during the Prospective Child Node's Connection Routine, node B may begin performing the Prospective Parent Node's Connection Routine. In step 161, in response to node X's query, node B determines whether it is part of the distribution system node X seeks to join. If node B were not part of the distribution network, it would respond in the negative to node X's query and node B would be finished with the Prospective Parent Node's Connection Routine. In the example of FIG. 8, the answer is "yes" and node B proceeds to step 162.

In step 162 node B determines whether it has room for a new node. If the answer were "yes," node B would proceed to step 163 where it would allow node X to dock with it, and node B would begin transmitting to node X data (e.g., streaming media) originating from the server. In the example illustrated in FIG. 8, the answer is "no," and node B, acting as an instructing node, goes to step 164 where it performs the Fully Occupied Parent's Connection Instruction Routine (discussed below) and provides the prospective new child node (here node X) with a new connection address list. As noted above, the new connection address list may include the path back to the server through the node B (the prospective parent in this example) in the event that there are user node departures along the path, which departures may include node B.

After performing step 164 the temporary connection between node B and node X is terminated, and node X is sent on its way. In the example of FIGS. 8 and 9, the new connection address list is G-B-A-S. When node X approaches node G, node G performs the Prospective Parent Node's Connection Routine discussed above. In the example illustrated in FIGS. 8 and 9, node G allows node X to dock with it.

Referring now to a "Distribution Network Construction" example (which example is intended to be illustrative and not restrictive), it is noted that under this example the length of the distribution chains (i.e., the number of levels of user nodes linked through each other to the root server) is aimed to be kept as small as possible.

Assuming that all new user nodes have the same re-transmission capability (or disregarding the different re-transmission capabilities which different new user nodes may have), the user nodes would be distributed in this example through the first level until all the direct connection slots to the server were filled. Then as new user nodes sought connection to the distribution network, they would be assigned to connections with the first level nodes 12 until all the slots available on the first level nodes were filled. This procedure would be repeated with respect to each level as more and more new user nodes attempted to connect to the distribution network. In other words, the server, acting as an instructing node, would perform a Server's Connection Instruction Routine in which one step is determining whether there is room on the server for a new user node and, if so, the server would instruct the new user node to connect directly to the server. If there were no room on the server, then the server would perform the step of consulting its network topology database and devising a connection address list having the fewest number of connection links to a prospective parent user node. After performing the Server's Connection Instruction Routine, the server would either allow the new user node to dock directly with the server or send the new user node on its way to the first prospective parent user node on the connection address list.

In this example (which example is intended to be illustrative and not restrictive), a partially occupied potential parent node in a particular level (i.e., a prospective parent node already having a child but still having an available slot for an additional child node) may be preferred over unoccupied (i.e., childless) potential parent nodes on the same level. This may help to keep the number of interior nodes (i.e., the number of nodes re-transmitting data) to a minimum.

Alternatively, in another example (which example is intended to be illustrative and not restrictive), nodes that have zero children may be preferred over nodes that already have one child. While it is true that this may increase the number of repeat nodes, what the inventors have found is that by filling the frontier in an "interlaced" fashion, connecting nodes build up their buffers more quickly (allowing a reduction in the incidence of stream interruptions).

Referring now to FIGS. 12A-12C, these FIGS. illustrate, in this example, what happens when a partially occupied parent node is preferred over a parent childless node as a

destination address for a new user node (assuming all other factors are equal). FIG. 12A is a schematic diagram associated with this example showing a topology wherein nodes C and D, both third level nodes 14, have remaining capacity for one or more child nodes. Server 11 has the choice of sending new user node X to either node C, as shown in FIG. 12B, or node D, as shown in FIG. 12C, without increasing the length of the longest chain in the distribution network. However, user nodes are free to leave the distribution network at any time. With the topology shown in FIG. 12C, if either of nodes C or D leaves the network, there would be an effect on a child node. With the topology shown in FIG. 12B, there is a significant chance that if one of nodes C and D leaves the network, there would be no impact on any child nodes. That is, if it is node D that leaves the network, there is no child which would be made an orphan.

Of note, in discussing the examples illustrated in FIGS. 6-9 and 12A-12C, it has been presumed that all of the user nodes are equally capable of simultaneously re-transmitting two copies of the content data. However, the user nodes actually joining a distribution network will have various capabilities for re-transmitting data. Some will have no reliable capability because of hardware limitations. Other user nodes will have high nominal capability because of their hardware and type of Internet connection, but may have low reliability because of poor line conditions or the vagaries of the desires of the humans actually using the user nodes. For example, a user node having a T1 or faster connection to the Internet has a significantly large bandwidth, but is not reliable if the user (or viewer) is merely sampling the content available on the distribution network.

As noted above, in one example the user nodes having the greatest reliable capability should be placed as high up in a distribution chain as possible (i.e., as far uptree as possible) because they would have the ability to support the greatest number of child nodes, grandchild nodes and so on.

Thus, to differentiate the reliable capabilities of user nodes, a number of factors may be considered. Examples of such factors (which examples are intended to be illustrative and not restrictive), include following:

- One is time, that is, the number of seconds (or other units of time) since the node made its most recent connection to the network. Everything else being equal, a user node which has been continuously connected to the distribution network for a long period of time may be considered under this example to be likely more reliable

(either because of line conditions or user interest) than a user node which has been continuously connected to the network for a short period of time.

- Another factor is bandwidth rating, which may be determined, for example, by actual testing of the user node when it first attempts to connect to the server or a parent node (e.g., at initial connection time) or determined by the nominal bandwidth as determined by the type of connection made by the user node to the Internet. In one example (which example is intended to be illustrative and not restrictive), we will describe ratings based on nominal bandwidth as follows:

  - A user node with a 56 Kbits/sec dialup modem connection to the Internet is essentially useless for re-transmission of content data because of its small bandwidth. In this example, such a node is assigned a bandwidth rating of zero (0.0).

  - A user node with a cable modem or DSL (e.g., ADSL) connection to the Internet may be given, in this example, a bandwidth rating of one (1.0), because it is a common type of node and has a nominal upstream transmission bandwidth of 128 Kbits/sec, which is large enough to potentially re-transmit two acceptable quality copies of the content data it receives (assuming a 50Kbit/sec content data stream). Such capability fits well into a binary tree topology.

  - A full rate ISDN modem connection nominally has an upstream bandwidth of 56 Kbits/sec and a downstream bandwidth of 56 Kbits/sec, which would potentially support acceptable quality re-transmission of a single copy of the content data stream (assuming a 50Kbit/sec content data stream). For this reason, a user node with a full rate ISDN modem connection to the Internet may be given, in this example, a bandwidth rating of one-half (0.5), or half the rating of a user node connected to the Internet by a DSL (e.g., ADSL) or cable modem connection.

  - User nodes with T1 or greater links to the Internet should be able to support more (e.g. at least twice as many) streams as DSL (e.g., ADSL) or cable modems, and therefore may be given, in this example, a bandwidth rating of two (2.0). In the event that in a distribution network parent nodes

may be assigned more than two child nodes directly connected thereto, bandwidth ratings greater than 2.0 may be assigned under this example to Internet connections having greater bandwidth than T1 connections.

- In another example (which example is intended to be illustrative and not restrictive), other bandwidth networks (e.g., 50 Kbit/sec and 100 Kbit/sec) may be supported.
- In another example, the bandwidth needed for a node to be "repeat capable" may depend on the particular network to which it is attached (upon initial connection the connecting node may first test its upstream bandwidth and then report that bandwidth to the network server; the server may determine whether the node is "repeat capable" or not.
- In another example (which example is intended to be illustrative and not restrictive), a node may be deemed "repeat capable" if the node is not firewalled (e.g., port 1138 is open) and the node has sufficient bandwidth to retransmit two copies of the incoming stream.
- A third factor is performance. A user node's performance rating may, in one example, be zero (0.0) if it is disconnected as a result of a Grandparent's Complaint Response Routine (discussed below in connection with FIG. 31). Otherwise, the user node's performance rating, in this example, may one (1.0).

Further, in this example (which example is intended to be illustrative and not restrictive), a user node's utility rating may determined by multiplying connection time by performance rating by bandwidth rating. That is, in this example:

- Utility Rating = Time x Performance Rating x Bandwidth Rating.

Referring once again to FIG. 13, it is seen that in this example information regarding a user node's time connected to the network, bandwidth rating, performance rating, utility rating and potential re-transmission rating (discussed below) may be stored in the user node's elapsed time buffer 137, bandwidth rating buffer 138, performance rating buffer 139, utility rating buffer 121 and potential re-transmission rating buffer 122, respectively.

In another example (which example is intended to be illustrative and not restrictive), nodes entering the network may be judged to be either "repeat capable" or "non-repeat capable" (wherein non-repeat capable nodes may be called "bumpable" nodes). Repeat

capability may be based on: (1) Upstream bandwidth (e.g., tested at initial connection); and (2) the firewall status, opened or closed, of a node. In one specific example, if a node is either firewalled or has upstream bandwidth less than (approx.) 2.5 times the video streaming rate, that node will be deemed bumpable. All nodes joining the network, whether bumpable or non-bumpable, may be placed as close to the server as possible. However, for the purpose of placement (as described for example in the Universal Connection Address Routine of Figure 20), a repeat capable node can bump a bumpable node when being placed in the network.

Referring now to FIG. 17 a flow diagram associated with this example illustrating the Server's Connection Instruction Routine is shown. This routine need not necessarily rely on the new user node's utility rating. This routine may rely on the potential re-transmission rating of a user node, which is arrived at by multiplying the user node's performance rating by its bandwidth rating. That is, in this example:

- Potential Re-transmission Rating = Performance Rating x Bandwidth Rating.

Thus, in this example, the server would want to put those user nodes with the highest potential re-transmission rating as close to the server in a distribution chain as possible (i.e., as high uptree as possible) because they have the greatest likelihood of being able to re-transmit one or more copies of content data. On the other hand, a potential re-transmission rating of zero in this example indicates that a user node has no ability to (or little expected reliability in) re-transmitting even one copy of content data (the server in this example would want to put a user node with a zero rating as far as reasonably possible from the server in a distribution chain (i.e., as low downtree as possible)). For the purpose of ease of discussion, in the flow diagram example of FIG. 17 illustrating the Server's Connection Instruction Routine, the server is concerned about whether the potential re-transmission rating is zero (i.e., either one or both of the performance rating and bandwidth rating is zero) or greater than zero (i.e., both the performance rating and the bandwidth rating are greater than zero).

In step 171 the server node interrogates the new user node (or connection requesting node) for its bandwidth rating and performance rating. If the new user node is really new to the distribution network, or if it has returned to the server because all of the user node's ancestor nodes have disappeared from the network, the new user node's performance memory will contain, in this example, a performance rating of 1.0 (e.g., the default rating). However, if the new user node has been dispatched to the server for a new connection to the network

because the new user node had failed to provide content data to one of its child nodes, then, in one example, its performance memory will contain a performance rating of zero.

In step 172, the server determines, in this example, whether the potential re-transmission rate of the connection requesting node is greater than zero (i.e., whether both the bandwidth rating and the performance rating are greater than zero, or, if only the bandwidth rating is considered, whether the bandwidth rating is greater than zero (i.e., the connection requesting node is a high-bandwidth node)). If the answer is "yes," then the server goes to step 173 in which the server determines whether it has room for the new user node. If the answer to the query in step 173 is "yes," then the server goes to step 174 in which it instructs the new user node to connect directly to the server. Then the server goes to step 102 in the Server's Connection Routine (see FIG. 10).

If the answer to the query in step 173 is "no" (i.e., the server does not have the capacity to serve the new user node directly), then the server goes to step 175. In step 175, the server, acting as an instructing node, consults its network topology database and devises a connection address list having the fewest number of connection links to a prospective parent node. That is, the server checks its information regarding the nodes in the level closest to the server (i.e., the highest uptree) to determine whether there are any potential parent nodes with space available for the new user node. If there are no potential parent nodes with space available, then the database is checked regarding nodes in the level one link further from the server, and so on until a level is found having at least one potential parent node with space available for the new user node. That is, the server determines which parent node with unused capacity for a child node is closest to the server (i.e., in the highest level, with the first level being the highest), and devises the connection address list from such prospective parent node to the server. The server then goes to step 102 (shown in FIG. 10).

In another embodiment of the invention, the server could skip step 172 and go directly to step 173 (or it could go to step 172 and, if the answer to the query in step 172 is "no" (i.e., either one or both of the bandwidth rating and the performance rating are zero, or, if only the bandwidth rating is considered, whether the bandwidth rating is zero (i.e., the connection requesting node is a low-bandwidth node)), the server could go to step 175). However, this could prevent the server from loading up the highest levels of the distribution chains with nodes capable of re-transmitting at least one acceptable copy of the content data (in contrast to if the server does perform step 172 and does goes to step 176 if the answer to the query in step 172 is "no.").

In step 176 the server consults its network topology database and devises a connection address list having the greatest number of connection links to a prospective parent node. That is, the server checks its information regarding the nodes in the level furthest from the server (i.e., the lowest downtree) to determine whether there are any potential parent nodes with space available for the new user node. If there are no potential parent nodes with space available, then the database is checked regarding nodes in the level one link closer to the server, and so on until a level is found having at least one potential parent node with space available for the new user node. In this manner, user nodes having limited or no reliable re-transmission capability may be started off as far from the server as possible and will have a reduced effect on the overall capacity of the distribution network.

As indicated above, and as discussed more fully below, one or more reconfiguration events may have transpired since the server's topology database was last updated. As a result, the first prospective parent node which is actually present on the distribution network for the new user node to contact may not have room for the new user node. By way of example (which example is intended to be illustrative and not restrictive), when node X tries to join the distribution network having the topology shown in FIG. 12A, the server provides node X with the following connection address list: C-B-A-S. If node C had disappeared from the network between the last update of the server's topology database and node X's attempting to contact node C, then node E, by virtue of a reconfiguration event, would be connected, in this example, to node B as shown in FIG. 12D. Then node X, in performing the Prospective Child Node's Connection Routine discussed in connection with FIGS. 14 and 15, would contact node B. Node B, in the Prospective Parent Node's Connection Routine, discussed in connection with FIG. 16, would have to answer the query of step 162 in the negative and go to step 164, in which it performs the Fully Occupied Parent's Connection Instruction Routine.

Referring now to FIG. 18, a flow diagram associated with this example illustrating that routine is shown. It is similar to the Server's Connection Instruction Routine. Since the fully occupied parent node has already determined that it has no room for the new user node (or connection requesting user node), the Fully Occupied Parent's Connection Instruction Routine does not need to include a step in which a determination is made regarding whether there is room for the new user node (or connection requesting node). In the Fully Occupied Parent's Connection Instruction Routine the fully occupied parent node, acting as an instructing node, first performs step 181 in which it interrogates the new user node for its bandwidth rating and performance rating. In step 182, the fully occupied parent node determines whether the potential re-transmission rate of the new user node is greater than

zero. (If only the bandwidth rating is considered, then it determines whether the bandwidth rating is greater than zero (i.e., the connection requesting node is a high-bandwidth node).) If the answer is "yes," then the fully occupied parent node goes to step 183 in which the fully occupied parent node consults its topology database, which contains the latest information available to that node regarding the links from the fully occupied parent node back to the server and regarding the fully occupied parent node's own progeny (i.e., its children, grandchildren etc.,) and devises a new connection address list having the fewest number of connection links to a new prospective parent node. That is, the fully occupied parent node checks its information regarding the nodes in the level closest to the fully occupied parent node (but not closer to the server) to determine whether there are any potential parent nodes with space available for the new user node. If there are no potential parent nodes with space available, then the database is checked regarding nodes in the level one link further from the fully occupied parent node (and further from the server), and so on until a level is found having at least one potential parent node with space available for the new user node. That is, the fully occupied parent node determines which new prospective parent node with unused capacity for a child node is closest to the fully occupied parent node, and devises the connection address list from such new prospective parent node through the fully occupied parent node and on to the server. The fully occupied parent node then goes to step 184 in which it provides the new user node with the new connection address list and disconnects the new user node from the fully occupied parent.

If the answer to the query in step 182 is no (i.e., either (i) one of the bandwidth rating and performance rating is zero or (ii) if only the bandwidth rating is considered, the bandwidth rating is zero (i.e., the connection requesting node is a low-bandwidth node)), the fully occupied parent node goes to step 185. In step 185 the fully occupied parent node consults its network topology database and devises a connection address list having the greatest number of connection links to a new prospective parent node. That is, the fully occupied parent node checks its information regarding the nodes in the level furthest from it (and farther from the server than it is) to determine whether there are any potential parent nodes with space available for the new user node. If there are no potential parent nodes with space available, then the database is checked regarding nodes in the level one link closer to the fully occupied parent node, and so on until a level is found having at least one potential parent node with space available for the new user node. As discussed above, this helps assure that new user nodes having limited or no reliable re-transmission capability are started off as far from the server as possible. After the fully occupied parent node devises the connection

Aylo Exhibit 1018
Page 1717 of 2308

address list from such new prospective parent node through the fully occupied parent node and on to the server, the fully occupied parent node then goes to step 184 where it performs as discussed above.

In another embodiment of the invention, the distribution software could be designed such that a fully occupied parent performs an abbreviated Fully Occupied Parent's Connection Instruction Routine, in which steps 181, 182 and 185 are not performed. That is, it could be presumed that the server has done the major portion of the work needed to determine where the new user node should be placed and that the fully occupied parent user node need only redirect the new user node to the closest available new prospective parent. In such event, only steps 183 and 184 would be performed.

In the example discussed above in which node C had disappeared from the network when new user node X had been given, by the server, connection address C-B-A-S, and in which node B is a fully occupied parent node as shown in FIG. 12D, node B would appear to have the choice of devising either connection address list D-B-A-S or E-B-A-S regardless of whether the full or abbreviated Fully Occupied Parent's Connection Instruction Routine were performed.

In this regard, to determine which connection address list to choose when there are two or more prospective parent nodes in a particular level having space for a new user node, the distribution software could have an additional subroutine as part of steps 175, 176, 183 and 185. An example of this subroutine, called the Multi-Node Selection Subroutine is illustrated in FIG. 19.

In step 191 the server or fully occupied parent node deciding where to send a new user node determines whether any of the potential new parent nodes is partially occupied. As discussed earlier, in one example a partially occupied potential parent node may be preferred over an unoccupied potential parent node. In this case, if any of the potential parent nodes is partially occupied, then the server or fully occupied parent node goes to step 192. In step 192 the partially occupied prospective parent node with the highest utility rating is selected as the new prospective parent node. If there were only a single partially occupied potential parent node, then that node is selected.

If in step 191 it is determined that there are no partially occupied potential parent nodes, then the server or fully occupied parent node goes to step 193. In step 193 the unoccupied prospective parent node with the highest utility rating is selected as the new prospective parent node.

In another example (which example is intended to be illustrative and not restrictive),the software engineer could have step 193 follow an affirmative response to the query in step 191 and step 192 follow a negative response; in such event, unoccupied prospective parent nodes would be selected ahead of partially occupied prospective parent nodes (all other things being equal, it may be advantageous to select nodes that have zero children over nodes that already have one child; while it is true that this may increase the number of repeat nodes, what the inventors have found is that by filling the frontier in an "interlaced" fashion, connecting nodes build up their buffers more quickly (allowing a reduction in the incidence of stream interruptions).

After whichever of steps 192 and 193 is completed, the server or fully occupied parent node returns to the step from which it entered the Multi-Node Selection Subroutine (i.e., step 175, 176, 183 or 185), and completes that step.

So, in the example shown in FIG. 12D, where node C left the network thereby leaving node B with the chore of devising a new connection address list for node X, node B would perform the Fully Occupied Parent's Connection Instruction Routine. Regardless of the bandwidth and performance ratings of node X, node B would be choosing between nodes D and E in the third level. In step 191 node B would determine that neither D nor E is partially occupied, and therefore node B would go to step 193. Assuming that nodes E and D have equal bandwidth and performance ratings and that node D was connected to the network longer than node E, node D would be selected because it would have the higher utility rating since it was connected to the network longer than node E. Node B would then go to step 194 and then return to the step from which it entered the Multi-Node Selection Subroutine. When node B returns to step 183 or 185, it completes that step and moves on to step 184. In that step, node B provides new user node X with new connection address list D-B-A-S and node X connects to the distribution network as shown in FIG. 12E.

Referring now to an "Alternative Distribution Network Construction" example (which example is intended to be illustrative and not restrictive), it is noted that under this example the length of the distribution chains (i.e., the number of levels of user nodes linked through each other to the server) is aimed to be kept as small as possible.

Also under this example, user nodes having the highest bandwidth capabilities are aimed to be kept closer to the server (e.g., in order to allow the greatest expansion of the distribution system). However, it is possible that zero bandwidth rated nodes may nevertheless appear relatively far uptree (thereby stunting the growth of that chain). The following method may be used in constructing the distribution network both by servers and

by prospective parents which are actually completely occupied, either of which may be thought of as an instructing node (that is, software enabling the routines discussed below could be installed on servers and user nodes alike).

More particularly, in the distribution system of this example, each child node reports directly to (or is tested by) its parent node with respect to information relating to the child node's bandwidth rating, performance rating, potential re-transmission rating and utility rating. In turn, each parent node reports all the information it has obtained regarding its child nodes on to its own parent node (a parent node also reports to each of its child nodes the address list from that parent back to the server, which list forms what may be referred to as the ancestor portion of the topology database – in addition, a parent node reports to each of its child nodes the addresses of their siblings). The reports occur during what may be referred to as a utility rating event. Utility rating events may occur, for example, on a predetermined schedule (e.g., utility rating events may occur as frequently as every few seconds). As a result of all the reporting, each node stores in its topology database the topology of the tree (including all its branches) extending below that node, and the server stores in its topology database the topology of all the trees extending below it. This may be referred to as the descendant portion (or descendant database) 133 of the topology database (see FIG. 13). The descendant database of a particular node may include a descendant list, a list of the addresses of all the nodes cascadingly connected below that particular node. Included in the topology database information may be the utility ratings of the nodes below the node in which that particular topology database resides.

After each reporting event in this example, each parent node (including the server), acting as an instructing node, devises two lists of prospective (or recommended) parent nodes. In this example, the first list, or Primary Recommended Parent List ("PRPL"), stored in the Primary Recommended Parent List buffer 123 (see FIG. 13), lists all the nodes in the descendant portion of that node's topology database which have bandwidth available to support another child node (in one example (which example is intended to be illustrative and not restrictive), in a binary tree system, all nodes in the descendant portion of the topology database having (i) a bandwidth rating of at least one and (ii) less than two child nodes would be listed). They would be listed with those node's which are closest to the node in which that particular topology database resides at the top of the list, and those nodes which are in the same level would be ranked such that the node with the highest utility rating would be listed first, the node with the second highest utility rating would be listed second and so on. By way of example (which example is intended to be illustrative and not restrictive), the PRPL of a

second level node would list a third level node with available bandwidth ahead of a fourth level node with available bandwidth even if the fourth level node's utility rating were higher than that of the third level node.

The second list in this example, or Secondary Recommended Parent List ("SRPL"), stored in the Secondary Recommended Parent List buffer 124 (see FIG. 13), lists all the nodes in the descendant portion of that node's topology database which have the ability to re-transmit content data to child nodes but are fully occupied, and at least one of its child nodes is incapable of re-transmitting content data to another child node (in one example (which example is intended to be illustrative and not restrictive), in a binary tree system, all nodes in the descendant portion of the topology database having (i) a bandwidth rating of at least one and (ii) at least one child node having a bandwidth rating less than one (i.e., being incapable of re-transmitting content data to two child nodes) would be listed). Like the nodes in the PRPL, the nodes in the SRPL would be listed with those node's which are closest to the node in which that particular topology database resides at the top of the list, and those nodes which are in the same level would be ranked such that the node with the highest utility rating would be listed first, the node with the second highest utility rating would be listed second and so on.

Of note, in this example the SRPL lists those parent nodes having the growth of their branches (i.e., their further progeny) blocked or partially blocked by a low-bandwidth child node. This may lead to an unbalanced growth of the distribution system, and a limitation on the total capacity of the system.

Of further note, to the extent that a node (including a server) has room for another child node or is the parent of a low bandwidth node, it may be listed on its own PRPL or SRPL.

Referring now to FIGS. 20A and 20B, an example flow diagram illustrating what may be referred to as the Universal Connection Address Routine is shown. As indicated above, a server or a fully occupied prospective parent node receiving a connection request may be referred to as an "instructing node." When a new node (here node N) approaches an instructing node, the instructing node performs step 201, it receives a connection request. It then goes to step 202 in which it interrogates or tests node N to determine whether it is a low-bandwidth node ("low-bandwidth" may mean, for example, either or both: (1) a node that is incapable of broadcasting two copies of its incoming stream on to its children; and (2) a node that is incapable of rebroadcasting its stream because it is firewalled). In the binary tree system of one example, a low-bandwidth node is a node with a bandwidth rating less than

one. If node N is a low bandwidth node, the instructing node proceeds to step 203 in which the instructing node determines whether there are any available prospective parent nodes which are not fully occupied. Sometimes the distribution network may be fully occupied. If it is, the instructing node's PRPL will be empty. If it is empty, the response to the query in step 203 would be yes. Then the instructing node goes to step 209 in which the new node is sent back to the server to start the connection process from the beginning. If the response to the query in step 203 is no, then the instructing node goes to step 204 in which it selects a prospective (or recommended) parent node for node N. The instructing node then moves on to step 205 in which it consults its topology database and devises a connection address list from the recommended parent node back to the server, and provides such connection address list to node N. (If the instructing node is a user node, then the connection address list leads back to the server through the instructing node.) At this point, node N performs the Prospective Child Node's Connection Routine discussed above in connection with FIG. 14.

As part of step 205, the instructing node inserts node N into its topology database as a child of the recommended parent node. It does this because other new nodes may seek connection instructions prior to the next utility rating event (i.e., before reports providing updated information regarding the topology of the distribution network are received by the instructing node), and such new nodes should not be sent to a prospective parent which the instructing node could know is fully occupied. In this regard, the instructing node then goes to step 206 in which it checks its topology database to determine whether the recommended parent, with node N presumptively docked to it as a child node, is fully occupied (in the example here of a binary tree network, whether it has two child nodes). If the answer is no, then the instructing node has finished this routine. If the answer is yes, then the instructing node goes to step 207 in which it removes the recommended parent from the PRPL and inserts it into the SRPL, and the instructing node has finished the routine.

If the answer to the query in step 202 is no (e.g., node N is not a low-bandwidth node; in the binary tree network example it is a high-bandwidth node capable of re-transmitting content data to two child nodes), the instructing node moves on to step 208. There it determines whether both the PRPL and SRPL are empty (which may occur under certain circumstances, such as, for example, when the number of levels in a distribution system is capped and at least all the nodes on all but the last level are fully occupied with high-bandwidth nodes). If so, the instructing node goes to step 209 in which the new node is sent back to the server to start the connection process from the beginning. If the response to the query in step 208 is no, then the instructing node goes to step 210 in which it selects a

prospective (or recommended) parent node for node N from the PRPL and an alternate recommended parent node from the SRPL (if either the PRPL or SRPL is empty, the instructing node will make a "nil" selection from that list – the instructing node knows from step 208 that at least one of the lists will not be empty). The instructing node then goes to step 211 in which it determines whether the alternate recommended parent node is closer to the server (i.e., higher uptree) than the recommended parent node derived from the PRPL. If the alternate recommended parent node is on the same level as, or on a lower level than the recommended parent node derived from the PRPL (or if the selection from the SRPL is nil), then the answer to the query in step 211 is no.

In such event, the instructing node goes to step 212 in which it consults its topology database and devises a connection address list from the recommended parent node back to the server, and provides such connection address list to node N (if the instructing node is a user node, then the connection address list leads back to the server through the instructing node).

As in step 205 discussed above, at this point, node N performs the Prospective Child Node's Connection Routine discussed above in connection with FIG. 14, and in step 212, the instructing node inserts node N into its topology database as a child of the recommended parent node.

The instructing node moves to step 213 in which it adds node N to an appropriate position on the PRPL. It does this because, as a result of step 202, it knows that the node N is capable of re-transmitting content data to its own child nodes.

The instructing node then goes to step 214 in which it checks its topology database to determine whether the recommended parent, with node N presumptively docked to it as a child node, is fully occupied (in the example here of a binary tree network, whether it has two child nodes). If the answer is no, then the instructing node has finished this routine. If the answer is yes, then the instructing node goes to step 215 in which it removes the recommended parent from the PRPL because it is now deemed to not be an available prospective parent node.

Then the instructing node goes to step 216 in which it consults its topology database to determine whether any of the recommended parent node's children is a low-bandwidth node (in this example, knowing that node N is not a low-bandwidth node and knowing that the recommended parent node has two child nodes, the question is whether the child node other than node N is a low-bandwidth node.) If the answer is no (i.e., all the recommended parent node's children are high-bandwidth nodes), then the instructing node has finished the routine.

If the answer is yes, then the growth of the recommended parent node's line of progeny is partially blocked by a low-bandwidth child node. The instructing node moves on to step 217 in which it adds the recommended parent to the SRPL.

If the answer to the query in step 211 is yes (i.e., the alternate recommended parent (selected from the SRPL) is: (i) on a higher level than the recommended parent (selected from the PRPL) or (ii) the selection from the PRPL is nil), then the instructing node moves to step 218. In that step the instructing node consults its topology database to determine: (i) which of the alternate recommended parent node's child nodes is a low-bandwidth node or (ii) if they both are low-bandwidth nodes, which of the child nodes has been connected to the system a shorter period of time. In one example (which example is intended to be illustrative and not restrictive), the instructing node may send a disconnect signal to that child node with instructions to return to the server to start the connection process from the beginning (as a new user node (or connection requesting user node)). In another example (which example is intended to be illustrative and not restrictive), the bumped node may climb its path to its grandparent which gives it a new connection path – rather than returning to the root server (doing so typically means that the incoming node that bumped the non-repeat-capable node ends up being that node's new parent).

The instructing node moves on to step 219 in which it consults its topology database and devises a connection address list from the alternate recommended parent node back to the server, and provides such connection address list to node N (if the instructing node is a user node, then the connection address list leads back to the server through the instructing node).

As in steps 205 and 212 discussed above, at this point, node N performs the Prospective Child Node's Connection Routine discussed above in connection with FIG. 14, and in step 219, the instructing node inserts node N into its topology database as a child of the alternate recommended parent node.

The instructing node moves to step 220 in which it adds node N to an appropriate position on the PRPL. It does this because, as a result of step 202, it knows that the node N is capable of re-transmitting content data to its own child nodes.

The instructing node then goes to step 221 in which it checks its topology database to determine whether all the child nodes of the alternate recommended parent are high-bandwidth nodes. If the answer is no, then the instructing node has finished this routine. If the answer is yes, then the instructing node goes to step 222 in which it removes the alternate recommended parent from the SRPL because it is now deemed to not be an available

alternative prospective parent node since the growth of its progeny line is not even partially blocked by one of its own children. At this point the instructing node has finished the routine.

With the routines discussed in connection with the above examples, the distribution network will be built with each new node assigned to the shortest tree (or chain), and those with the fewest number of links between it and the server. However, low-bandwidth nodes, which would tend to block the balanced growth of the distribution network, would be displaced by high-bandwidth nodes and moved to the edges of the network where they would have reduced effect on the growth of the network.

Referring now to FIG. 21, this FIG. may be used as an example to illustrate how the Universal Connection Address Routine works. It shows primary server 11 with node A as first level node 12, its child nodes B and C as second level nodes 13, their child nodes D, E, F and G as third level nodes 14, node D and G's child nodes H, I, J and K as fourth level nodes 15, and node I's child nodes L and M as fifth level nodes 17. Assume for this example that all available docking links directly to the server 11 are occupied by high-bandwidth nodes, and that all trees are at full capacity other than that rooted to the server 11 through first level node A. The utility rating for each node is set forth in FIG. 21 under its letter designation. Low-bandwidth nodes (in this example, those nodes having a bandwidth rating less than one) are shown with a bandwidth rating of "Ω," which indicates in this example that such nodes will not have any child nodes assigned to them. Nodes A, B, C, D, E, G, I, J and K are high-bandwidth nodes (or repeater nodes (i.e., they are capable of re-transmitting the content data they receive from their respective parent nodes to their respective child nodes, if any)).

A new user node (or other connection requesting user node), here referred to as "node N", may be directed to node A as a result of various reconfiguration events. When node A receives a connection request from node N, node A will, in this example, either consult both of its PRPL and SRPL buffers, if node N is a high-bandwidth node, or consult just its PRPL buffer (if node N is a low-bandwidth node).

Nodes F, H, L, and M will not appear on any list in this example since they are low-bandwidth nodes. Nodes B, C, D, G, I and A itself will not appear on the PRPL in this example since these nodes are fully occupied. However, nodes C, D, and I will appear on the SRPL in this example because they each have at least one low bandwidth child node.

Using the rules discussed above (i.e., ranking prospective parent nodes by level and within levels by utility rating) the PRPL would be as follows in this example:

- PRPL

- E (a third level node with available capacity)
  - J (the fourth level node with available capacity having the highest utility rating)
- K (a fourth level node with available capacity)

The SRPL in this example would be as follows:

- SRPL
  - C (the fully occupied second level node with a low-bandwidth child node)
  - D (the fully occupied third level node with a low-bandwidth child node)
  - I (the fully occupied fourth level node with a low-bandwidth child node)

If node N is a low-bandwidth node, node A will give, in this example, node N the following as its new connection address list:

- E-B-A-S.

Since node E would not have two child nodes, it would remain on the PRPL.

If node N is a high-bandwidth node, node A will compare (step 211 of FIG. 20B) the first node on the PRPL (the recommended parent node) with the first node on the SRPL (the alternate recommended parent node). Here, node C, the first node on the SRPL is a higher level node than node E, the first node on the PRPL. So, node A will send a disconnect signal (step 218 of FIG. 20B) to node F, node C's low-bandwidth child node. Then it will provide node N with the following new connection address list and add node N to the PRPL (step 219 of FIG. 20B):

- C-A-S.

Since node C would now have two high-bandwidth child nodes (nodes N and G), node C would be removed from the SRPL (step 222 of FIG. 20B).

Further, in this example wherein node N is a high-bandwidth node, when node F returns to the server 11 (or, in another example, goes to Node A) for a new connection, the

server (or Node A) will also use the Universal Connection Address Routine (Node F may go to Node A rather than the server because (in one example) when a node is "kicked" it may climb its internal connection path rather than jumping directly to the root server). Since node F is a low-bandwidth node, the server (or Node A) will give node F the following as its new connection address list:

- E-B-A-S.

FIG. 22 illustrates the new topology. As can be seen, absent intervening events, low-bandwidth node F will end up moving down from the third level to the fourth level and the bandwidth capacity of the third level will increase from six child nodes to eight (its maximum in this binary tree example). The potential bandwidth capacity of the fourth level would also increase, from ten child nodes to twelve.

Referring now to a "Distribution Network Reconfiguration" example (which example is intended to be illustrative and not restrictive), it is noted that under this example the user nodes are free to exit the distribution network at will. And, of course, user nodes may exit the distribution network due to equipment failures between the user node and the communication network (the Internet in the examples discussed herein). This was first discussed in connection with FIGS. 10A-E.

Under this example the present invention may handle the reconfiguration required by a user node departure by having certain user nodes in the same tree as the departing node perform a Propagation Routine. The results of the propagation routine of this example are illustrated in FIGS. 23 and 24. There a tree rooted to the server through first level user node P is illustrated. Node P has two child nodes, second level nodes Q and X. Through node X, P has two grandchild nodes, third level nodes A and B. Based on the relative utility ratings of nodes Q and X, P has sent signals to its children instructing them to set the color of the flag represented by the information in their respective reconfiguration flag buffers to "green" and "red," respectively. The use of colors as designators is merely discretionary. In reality they represent relative ratings of siblings for purposes of determining their roles in a reconfiguration event, and may be referred to as "propagation ratings." In addition, instead of colors, numbers (e.g., 1 and 2) could be used, as well as any of many other nomenclature schemes to describe the grades of propagation ratings. The "red" "green" ratings discussed herein shall be deemed representative of all such schemes. The number of grades of propagation ratings assigned by a parent node may be equal to the number of children each

parent node has. For example, for a branch factor "n" system, the maximum number of grades of propagation ratings may be "n." Since the distribution network in the examples discussed herein is a binary tree distribution network, a parent node will be required to assign at most up to two grades of propagation ratings.

In the example shown in FIG. 23, node P, during the most recent utility rating measurement event, discerned that node Q has a higher utility rating than node X, and therefore P has assigned node Q a green rating, represented by the solid-lined circle surrounding the letter Q in FIG. 23. P has assigned node X a red rating, represented by the dashed-lined circle surrounding the letter X in FIG. 23. In a similar manner, node X has assigned green and red ratings to third level nodes A and B, respectively. In this example (which example is intended to be illustrative and not restrictive), at the same time that a parent node assigns a propagation rating to a child node, it may also provide to such child node the address of the child node's sibling, if there is any. A child node may store information about its sibling (or siblings) in the sibling portion (or sibling database) 134 of its topology database (see FIG. 13). The sibling database includes a sibling list, a list of the addresses of a node's siblings (the data relating to the siblings' addresses may also contain information regarding the propagation ratings of the siblings (e.g., in the event that the distribution system has a branch factor greater than two)). In the example shown in FIG. 23, nodes Q and X know that they are each other's siblings and nodes A and B know that they are each other's siblings.

In another example (which example is intended to be illustrative and not restrictive), nodes do not store information about their siblings. In other words, in this example, a node generally does not know who its sibling is. Instead, cross connections between a red node and its green sibling are implemented via a "priority join" message sent to a red child that includes the IP address of the green sibling as the first entry in the connection path list.

In the event that node X were to leave the distribution network, nodes A and B would of course stop receiving content data from node X. Nodes A and B would consult their topology databases for the address of their grandparent, and each would interrogate node P for instructions. Since node X is out of the distribution network, node P would send out a reconfiguration event signal (sometimes referred to as a "propagation signal") to the nodes which had been node X's children (i.e., the nodes which were node P's grandchildren through node X) and node P would send an upgrade propagation rating signal to its remaining child, here node Q.

In response to the upgrade propagation rating signal, node P's remaining child, node Q, would set its reconfiguration flag buffer 136 (see FIG. 13) to the next highest available propagation rating grade. Since in the example illustrated in FIG. 23, node Q's reconfiguration flag buffer already is set for the highest propagation rating grade (here green), node Q could either do nothing or reset its propagation rating in response to the upgrade propagation rating signal. The result would be the same, node Q's propagation rating would remain green. If node Q's propagation rating were red, then it would set its propagation rating to green in response to the upgrade propagation rating signal. Node Q would do nothing else in response to the upgrade propagation rating signal (note that as a matter of design choice, the software engineer could have the parent of the departed node (here node P is the parent of departed node X) send no upgrade propagation rating signal to its remaining child node (here node Q) if its remaining child node already is at the highest propagation rating).

In this example, the recipients of the propagation signal (i.e., the children of the missing node) would respond thereto as follows. First they would check their own respective propagation ratings. If the propagation signal's recipient has the highest propagation rating grade (here green), it would reset its propagation rating to the lowest rating grade (here red); re-transmit the propagation signal to its own child nodes (if there are any); disconnect the child node which did not have the highest propagation rating prior to its receipt of the propagation signal if the propagation signal's recipient with the green rating has more than one child node; allow its sibling to dock with it as a child node for the purpose of transmitting content data to that child node; and dock (or remained docked), for purposes of receiving content data, with the node sending the propagation signal to it (in systems having a branch factor of more than two, the propagation signal recipient whose propagation rating had been the highest would disconnect its child nodes which did not have the highest propagation rating, prior to the receipt of the propagation signal just sent, and it would allow its (i.e., the formerly green node's) siblings to dock with it as child nodes).

If the propagation signal's recipient has other than the highest propagation rating (i.e., just prior to the receipt of the propagation signal), it would upgrade its propagation rating to the next highest rating grade; dock with its sibling which had the highest rating grade; and begin receiving content data from it. If the propagation signal's recipient has other than the highest propagation rating, it does not re-transmit the propagation signal to its own child nodes.

In the example illustrated in FIG. 23, nodes A and B receive the propagation signal from node P. Since node A has the highest propagation rating grade, here green, it: (i) sets its

reconfiguration flag buffer 136 (see FIG. 13) so that it has the lowest propagation rating grade, here red; and (ii) docks with node P (becoming a second level node 13) to begin receiving content data from node P. Node B changes its propagation rating from red to the next higher propagation rating (and since this example is a binary tree (or branch factor two) system, the next higher rating is the highest, green) and docks with node A to receive content data. The resulting topology is shown in FIG. 24. Note that node B remains a third level node 14.

FIGS. 25 and 26 illustrate what happens in this example when the departing node is green. In FIG. 25, node X is the departing node and it is green. When node P sends the upgrade propagation rating signal to node Q, it changes its propagation rating from red to green. In other respects the reconfiguration event proceeds essentially as described in the paragraph immediately above, and results in the topology shown in FIG. 26 (which is the same as the topology in FIG. 24).

FIG. 27 is a flow diagram associated with an example showing the Child Node's Propagation Routine, the routine which is followed by a child node upon receiving a propagation signal. Of note, this example is provided for illustration only, and is not restrictive, and other examples are of course possible (e.g., the color of a node may assigned by its parent such that nodes assign and know the color of their children but not of themselves).

In any case, still referring to FIG. 27, the child node first performs step 271 wherein it determines whether it has received such a signal. If it has not, then it does nothing. If it has received a propagation signal, it proceeds to step 272 wherein it checks its propagation rating grade in reconfiguration flag buffer 136 (see FIG. 13) . If its propagation rating is at the highest grade (i.e., it is a "green" node in this example), then it proceeds to step 273 where it sets its reconfiguration flag buffer to the lowest propagation rating grade. It then proceeds to step 274 in which it re-transmits the propagation signal to its own child nodes, and which results in the undocking of all its child nodes except for the one with the highest propagation rating. The child node (i.e., the child node referred to in the second sentence of this paragraph) then performs step 275 in which it determines whether it is already docked with the node which sent the propagation signal. If it is not (i.e., it received the propagation signal from its grandparent), then it proceeds to step 276. In that step it: (i) devises a new connection address list, which is the ancestor portion of its topology database with the current nominal parent node removed, resulting in the grandparent node becoming the first node on the connection address list; and (ii) then performs the Prospective Child Node's Connection

Routine (i.e., it goes to step 141 discussed above in connection with FIGS. 14 and 15). The Prospective Child Node's Connection Routine is performed because some likelihood exists that even the grandparent node may have departed from the distribution system between the moment it sent out the propagation signal and the moment that its highest rated grandchild from its missing child attempted to dock with it.

If the answer to the query in step 275 is in the affirmative (i.e., the child node is receiving a propagation signal which has been re-transmitted by its parent), then the child node does nothing more (or as illustrated in FIG. 27, it performs step 277 which is to remain docked to its parent node).

If the answer to the query in step 272 is in the negative (i.e., it does not have the highest propagation rating (it is a "red" node in this example)), then it proceeds to step 278 in which it: (i) sets its reconfiguration flag buffer so that its propagation rating is the next higher grade (in this binary tree system the rating is upgraded from red to green); (ii) undocks from the parent with which it was docked before receiving the propagation signal (if it was docked with a parent before receiving such signal); and (iii) either (a) waits a predetermined period of time during which the node's sibling which was green prior to their receipt of the propagation signal should have re-transmitted the propagation signal to its own child nodes (thereby causing any red child nodes to undock from it) or (b) confirms that the node's sibling which was green prior to their receipt of the propagation signal actual did re-transmit the propagation signal to its child nodes (if it has any). Then the child node performs step 279 in which it: (i) devises a new connection address list, which is the ancestor portion of its topology database with its sibling node having the highest propagation rating placed in front of the previous parent as the first node on the connection address list; and (ii) then performs the Prospective Child Node's Connection Routine (i.e., it goes to step 141 discussed above in connection with FIGS. 14 and 15). The Prospective Child Node's Connection Routine is performed because some likelihood exists that the sibling may have departed from the distribution system between the moment that the propagation signal had been sent to the child node and the moment that the child node attempted to dock with its sibling.

In another example (which example is intended to be illustrative and not restrictive), part (iii) of step 278 may be handled using the "priority join" mechanism. More particularly, in this example, a priority join from a node X cannot be refused by a target node T – it must accept (dock) the incoming node. If the target node T already had two children G & R, the target node T instructs its red child R to connect (via "priority join") to the green node G. The node T then disconnects from its red child R immediately prior to accepting the incoming

node X. The bumped node R then sends a priority join to its own green sibling A as the target. These actions cause the reconfiguration events to cascade until the edge of the tree is reached (of note, the algorithm described in this paragraph may have essentially the same effect as illustrated in Figures 28 and 29 – under certain circumstances the algorithm described in this paragraph may be somewhat more reliable).

FIG. 28 illustrates an example distribution system comprising node P as a first level node 12; nodes Q and X as second level nodes 13; nodes A and B as third level nodes 14; nodes C, D, E an F as fourth level nodes 15; nodes G and H as fifth level nodes 17; nodes I and J as sixth level nodes 20; nodes K and L as seventh level nodes 21 and nodes M and N as eighth level nodes 22. Nodes Q, B, C E, H, I, L and M have red propagation ratings, as symbolized by the dashed circles, and the remaining nodes have green propagation ratings as symbolized by the solid circles. As a result of node X's departing from the system, node P sends out an upgrade propagation rating signal to node Q and a propagation signal to the children of node P's departed child node X (i.e., nodes A and B).

In another example (which example is intended to be illustrative and not restrictive), during a graceful depart of node x, node x may send out messages to its parent and children immediately prior to its departure – rather than the parent of node x initiating these messages as described above.

In any case, continuing the example of FIG. 28, node Q changes its propagation rating to green. Further, in this example nodes A and B begin the Child Node's Propagation Routine. With respect to node A, it answers the query of step 271 in the affirmative. Since its rating is green, it also answers the query of step 272 in the affirmative. In step 273 node A changes the setting of its reconfiguration flag buffer to show a red propagation rating. In step 274 it re-transmits the propagation signal to its child nodes C and D. It answers the query of step 275 in the negative because it is not docked for purposes of receiving content data with node P. Node A then goes to step 276 wherein it consults the ancestor portion of its topology database and creates a new connection address list starting with its grandparent. The new connection address list is P-S. Then node A performs the Prospective Child Node's Connection Routine, starting with step 141. (See FIG. 14.) Assuming that no other intervening events have occurred, node A will successfully dock with node P.

Node B answers the query of step 271 in the affirmative and, because its propagation rating was red when it received the propagation signal, it answers the query of step 272 in the negative. From there it goes to step 278 in which it changes its propagation rating to green. Then in step 279 node B consults the ancestor portion of its topology database and creates a

new connection address list with its sibling, node A, placed in front of node X as the first address on the new connection address list. The new connection address list is A-X-P-S. Then node B performs the Prospective Child Node's Connection Routine, starting with step 141. Assuming that no other intervening events have occurred, node B will successfully dock with node A (if node A has disappeared, node B would attempt to dock with node X, but since it is not there, it would move on to node P.)

Node A, which had the higher propagation rating of the two child nodes of departed node X, moves up one level to become a second level node 13. Since it is a "new" entry to the second level, its initial utility rating and its propagation rating will be lower than that of node Q. As time goes by, at subsequent utility rating measurement events node A's utility rating (and hence its propagation rating) may become higher than that of node Q.

Node B, which had the lower propagation rating of the two child nodes of departed node X, does not re-transmit the propagation signal to its child nodes (nodes E and F), and they will follow B wherever it goes in the distribution system. In the example discussed here, node B becomes the child of node A while remaining third level node 14. At least initially, it will be node A's child with the higher utility rating and propagation rating.

As a result of node A's performing step 274, its child nodes C and D (which were fourth level nodes 15 in FIG. 28 when node X was still in the system) receive a propagation signal. Since node C has a red propagation rating, it, like node B, will remain in its level, change its propagation rating to green and dock with its sibling, with the result being that node D becomes its parent.

Since node D had a green propagation rating when it received the propagation signal, it answers the queries of steps 271 and 272 in the affirmative and changes its propagation rating to red in step 274. It answers the query of step 275 in the affirmative, and remains docked with node A. As a result, node D moves up a level and becomes an third level node 14 (with, at least until the next utility rating event, a lower utility rating and propagation rating than its new sibling, node B).

This process proceeds down the tree, with each child of a node which moves up a level doing one of the following under this example: (i) if it is a node having a green propagation rating, it remains docked with its parent, thereby itself moving up a level, changes its propagation rating to red and re-transmits the propagation signal to its child nodes; or (ii) if it is a node having a red propagation rating, it docks with the node which was its sibling, thereby staying in the same level, and changes its propagation rating to green.

The resulting topology after the reconfiguration event caused by node X's departure is illustrated in FIG. 29.

Of note, it is believed that reconfiguration using the Child Node's Propagation Routine of this example when a node departs the system does not necessarily result, in the long run, in significantly more inter-node reconnection events than any other reconfiguration method (and may result in fewer reconnection events). Moreover, it is believed that reconfiguration using the Child Node's Propagation Routine of this example helps assure that the more reliable nodes are promoted up the distribution network even if many reconfiguration events occur close in time to each other in a particular tree.

Referring now to a specific example of pseudo-code (which example is intended to be illustrative and not restrictive), it is noted that a propagation routine may be set forth as follows (wherein node X is the node departing from the distribution system):

```
Departs(X)
Begin
  If (X.Parent.GreenChild == X)  /* the departing node is green*/
  {
    X.Parent.GreenChild := X.Parent.RedChild
    X.Parent.RedChild := X.GreenChild
  }
  Else /* the departing node is already red */
  {
    X.Parent.RedChild :=X.GreenChild
  }
  Propagate(X.GreenChild, X.RedChild)
End
```

Referring now to another specific example of pseudo-code (which example is intended to be illustrative and not restrictive), it is noted that a propagation routine may be set forth as follows (wherein node promotion during the network reconfiguration event is precipitated by node X's departure):

```
Propagate (X, RedSibling)
Begin
  OriginalGreenChild := X.GreenChild
```

```
OriginalRedChild := X.RedChild
X.RedChild := X.GreenChild
X.GreenChild := RedSibling
If (OriginalGreenChild <> null)
{
  Propagate(OriginalGreenChild, OriginalRedChild)
}
End
```

Referring now to a "Malfunctioning Nodes" example (which example describes a voting system between grandchildren and grandparents about a potentially malfunctioning parent and is intended to be illustrative and not restrictive), it is noted that what happens when a node leaves the distribution network was discussed in the above example (e.g., when a node leaves the network, its parent will, at the next utility rating event, report that it has room for a new node, except when the parent has been contacted about the missing node by its grandchild node or nodes, in which event, a reconfiguration event will proceed as described above). Sometimes, however, a node intends to remain in the network, but there is a failure of communication.

In this regard, each node's topology database may include information regarding that node's ancestors and descendants. In the event that a child node stops receiving signals from its parent node, the child may, in one example, send a "complaint" message to its grandparent. The grandparent may check whether the parent is still there. If it is not, then the grandparent may send a propagation signal to the child nodes of the missing parent.

Further, if the grandparent detects that the parent node is actually still there, then the grandparent node may send a disconnect signal to the parent node (e.g., sending it back to the server to begin the connection process again). This may occur, for example, when one of the two following conditions is exists: (i) the child node is the only child node of the parent, or (ii) the child and its sibling(s) are complaining to the grandparent.

In this example, the grandparent would also send a propagation signal to the child nodes of the disconnected parent, and a reconfiguration event would occur.

However, if the child node has siblings and they are not sending complaint signals to the grandparent, then the grandparent may assume that the problem is with the complaining child node. Thus, grandparent may send a disconnect signal to the complaining child node (e.g., sending it back to the server to begin the connection process again). If the complaining

child node had its own child nodes, they would contact the departed child node's parent to complain, starting a reconfiguration event (in another example (which example is intended to be illustrative and not restrictive), the grandparent would send out signals to the malfunctioning parent (instructing it to depart) and to the children (instructing the green child to climb its path to its grandparent and instructing the red child to cross connect to the green child).

The foregoing example (which example is intended to be illustrative and not restrictive) can be described in connection with FIG. 30 (which depicts a tree in a binary system having a primary server 11, in which node A, a first level node 12, has two child nodes B and B' (which are second level nodes 13); node B has two child nodes C and C' (which are third level nodes 14); node C has two child nodes D and D' (which are fourth level nodes 15); and node D has two child nodes E and E' (which are fifth level nodes 17)) and FIG. 31 (which is a flow diagram showing an example Grandparent's Complaint Response Routine).

Who does node C complain to in this example when node C no longer gets satisfactory service from node B? Node C complains to its grandparent, node A. If node C does not hear back from node A within a predefined amount of time (e.g., if node A has left the network), node C will then exit the network and immediately reconnect as a new node by going to the primary server S for assignment.

What does node A do in response to a complaint from node C in this example? When node A receives a complaint from grandchild node C about node C's parent node B (i.e., node A's child node B), node A will choose to either remove its child node B or its grandchild node C. A will make this determination based on whether node C alone, or both node C and its sibling node C' are experiencing problems with node B, together with the knowledge of whether node A is continuing to receive satisfactory service reports from node B. If node A is continuing to get satisfactory service reports from node B and there is no indication that node C' (the sibling of node C) is experiencing problems with node B, then node A will assign the "fault" to node C and issue a disconnect order for its removal. At this point the green child of node C (i.e., node D or D') will move up a level connecting to parent B, and the red child of node C (the other of node D or D') will connect as a child of its former sibling. The reconfiguration event will then propagate as discussed above.

If, on the other hand, node A is not getting satisfactory service reports in this example from node B and/or a complaint from node C's sibling arrives within a `window` of node C's complaint, then node A will assign the "fault" to node B and issue a disconnect order for its

removal. At this point the green child of node B (i.e., node C or C') will move up a level connecting to grandparent node A, and the red child of node B (the other of node C or C'), will connect as a child of its former sibling. The reconfiguration event will then propagate as discussed above.

An exception to the above is the case where node C is the only child of node B. Under these circumstances node B may be disconnected in this example by node A based solely on the recommendation of node C. In another example, the system will not disconnect a node with one child based solely on the complaint of that child (in this case there may be insufficient data to accurately assign blame - disconnecting the node being complained against and promoting the complainer may have the potential to promote a node that is experiencing problems).

Referring now to FIG. 31 (showing a flowchart associated with this example), node A receives a complaint from node C about node B in step 311. Node A then goes to step 312 in which it checks its topology database to determine whether node C is the only child of node B. If the answer is no (as shown in FIG. 30), then node A goes to step 313 in which node A determines whether there is a communication problem between it and node B. If the answer is no, then node A proceeds to step 314 in which it determines whether it has received a similar complaint about node B from node C' within a predetermined period of time of node A's having received the complaint from node C. If the answer to that query is no, then node A goes to step 315 in which it sends a disconnect signal to node C. At that point node A has completed the Grandparent's Complaint Response Routine (in one example a propagation signal does not necessarily have to sent by node A to node C's child nodes - if any exist, they will complain to node B which, upon discerning that node C is no longer in the distribution network, will send a propagation signal to such complaining nodes).

If the response to any of the queries in steps 312, 313 and 314 is yes in this example, then node A proceeds to step 316 in which it disconnects node B. Node A then proceeds to step 317 in which it sends a propagation signal to its grandchild nodes (here nodes C and C') and a reconfiguration event will occur as described above.

The method of this example described above takes a "2 out of 3" approach on the question of node removal. In this regard, it is believed that many interior nodes in a binary tree distribution network constructed pursuant the present invention will have direct connections to three other nodes: their parent, their green child, and their red child. The connections between a node and its three "neighbor" nodes can be thought of as three separate communications channels. In this example a node is removed from the network

when there are indications of failure or inadequate performance on two of these three channels. When there are indications that two of the channels are working normally, the complaining node is presumed to be "unhealthy" and is removed.

In the case where an interior node has only a single child, the "complaint" of that child to its grandparent is sufficient, in this example, to remove the parent node - even when the node to be removed is communicating perfectly well with it's own parent. In other words, given the communication chain A-B-C where node A is the parent of node B and node B is the parent of node C, and given that node C has no siblings, then a complaint of node C to node A will cause node A to remove node B in this example regardless of the fact that node A and node B are not having communication problems (this course of action may help ensure that a potentially unreliable node does not move up the hierarchy of nodes, even at the cost of occasionally sending a "healthy" node back to the edge of the distribution chain based on uncorroborated complaints of its child). Of course, as mentioned above, there are examples where the system will not disconnect a node with one child based solely on the complaint of that child.

In another example of voting (which example is intended to be illustrative and not restrictive), instead of (or in conjunction with) nodes voting on the behavior of other nodes, nodes may continually grade their own performance and if they detect that there is insufficient bandwidth on two or more of their links (links to their parent and their children) the node will perform a graceful depart from the network (initiating a reconfiguration as described elsewhere) and then the node will either climb its path to an ancestor node (or go directly to the server) to request a connection path to the edge of the tree. This system allows nodes to "step aside" when they detect connection / bandwidth related issues.

Referring now to a "Stepped Delay of Playing Content Data" example (which example is intended to be illustrative and not restrictive) it is noted that in order to have all the user nodes experience the playing of the content data at approximately the same time, regardless of the level in which the node resides, the content data buffer 125 shown in FIG. 13 of a user node may be larger the higher uptree it is in a distribution chain (i.e., the closer it is to the root server). The larger the buffer, the greater the amount of time between a particular event (or from the node's receiving the content data), and the node's actually playing the content data in its content data buffer. To the extent that the content data buffer of a node is sized to vary the time that the playing of the content data is started, the content data buffer is a delay. Alternatively, the delay may be a timer or a delay buffer 126 as shown in

FIG. 13, or a combination of elements. The delay's purpose may be to help assure that all users experience the play of the content data approximately simultaneously.

For a distribution network having n levels of user nodes cascadingly connected to a server node, where n is a number greater than one (and therefore the distribution network comprises the server node and first through nth level user nodes), the period of time (i.e., the delay time) between:

(i) the moment content data is received by a node or, from a predetermined moment such as, for example, a reporting event or utility rating event (the occurrence of which may be based on time periods as measured by a clock outside of the distribution system and would occur approximately simultaneously for all nodes); and

(ii) the playing of the content data by the node;

is greater for a node the higher uptree the node is on the distribution chain. That is, the delays in first level nodes create greater delay times than do the delays in second level nodes, the delays in second level nodes create greater delay times than the delays in third level nodes and so on. Described mathematically, in an n level system, where x is a number from and including 2 to and including n, the delay time created by a delay in an (x-1) level node is greater than the delay time created by a delay in an x level node.

By way of example (which example is intended to be illustrative and not restrictive), first level nodes could have a built in effective delay time in playing content data of one-hundred-twenty seconds, second level nodes could have a built in effective delay time of one-hundred-five seconds, third level nodes could have a built in effective delay time of ninety seconds, fourth level nodes could have a built in effective delay time of seventy-five seconds, fifth level nodes could have a built in effective delay time of sixty seconds, sixth level nodes could have a built in effective delay time of forty-five seconds, seventh level nodes could have a built in effective delay time of thirty seconds, and eighth level nodes could have a built in effective delay time of fifteen seconds.

The delays may take advantage of the fact that transmission of the packets forming the content data from one node to another may take less time than the actual playing of such packets. This may allow for essentially simultaneous playing of the content data throughout all the levels of the distribution network. The delays may also allow nodes to handle reconnection and reconfiguration events without any significant loss of content data.

In another example (which example is intended to be illustrative and not restrictive), the system may not endeavor to have buffers closer to the server contain more data than buffers of nodes that are further away. Instead, every node may have a buffer that can hold,

for example, approximately two minutes worth of A/V content. The system may endeavor to fill every node's buffer as rapidly as possible, regardless of the node's distance from the server, in order to give each node the maximum amount of time possible to deal with network reconfiguration and uptree node failure events.

Of note, nodes may receive a "current play time index" when they connect to the server. This value may be used to keep all nodes at approximately the same point in the stream – to have all nodes playing approximately the same moment of video at the same time – by telling them, essentially, begin playing A/V content at this time index. The play time index may be, for example, approximately two minutes behind "real time" – allowing for nodes to buffer locally up to two minutes of content (the present invention may permit drift between nodes within the buffer (e.g., two minute buffer) or the present invention may reduce or eliminate this drift to keep all nodes playing closer to "lock step").

Turning now to a general description of an example process for Node X to enter a network including server Node S, Node X must contact Node S to initiate the process. For illustrative purposes, in the following example (which example is intended to be illustrative and not restrictive) each node in the network may have no more than two connections, and the root server (i.e., the highest node in the tree) shall be referred to as Node S.

The process of this example begins with Node X making a request of Node S to join the network.

If Node S has an open connection (e.g., including one or more connected bumpable non-repeat-capable nodes which may be replaced due to an incoming connection request from a repeat-capable node), Node S can accept Node X's request to join the network, and Node X becomes a child of Node S.

If, on the other hand, Node S does not have an open connection, Node S may reject Node X's request, and may provide to Node X a node-by-node path from Node S to Node P, which Node S understands to be in the network, and under which, as determined by Node S, Node X may be able to join the network.

The process of this example continues with Node X making a request of Node P to join the network.

If Node P does not respond – regardless of why Node P does not respond – Node X may request connection from the next higher node in the path. For the purpose of this illustration, thereafter the next higher node is now considered Node P. Node X may make such a request of successive Node Ps, until Node X receives a response.

If a particular Node P has an open connection, that Node P can accept Node X's request to join the network, and Node X becomes a child of that particular Node P.

If, on the other hand, a particular Node P does not have an open connection, that particular Node P may reject Node X's request, and may provide to Node X a node-by-node path from that particular Node P to Node P', which the instructing Node P understands to be in the network, and under which, as determined by the instructing Node P, Node X may be able to join the network.

If Node X receives a further rejection, Node X may then treat the Node P' as a Node P and continue the process.

In general, under this example the process is continued until Node X finds a connection.

Referring now more specifically to a "Registration Process" example involving two arbitrary Nodes X and N (neither of which is necessarily a server), FIG. 32 shows Node X sending a request to Node N for acceptance into the distribution network tree (of which Node N is part).

In this example, Node N (which may be thought of as an instructing node) may generally return one of three codes (or messages): Accept, Path, or Reject. FIGS. 33A-33C show example messaging between Node X and Node N under this example. More particularly:

- FIG. 33A shows Node X sending a Register message to Node N and Node N responding to the Register message with an Accept message
- FIG. 33B shows Node X sending a Register message to Node N and Node N responding to the Register message with a Path message (e.g., identifying a path from Node N to a new instructing node)
- FIG. 33C shows Node X sending a Register message to Node N and Node N responding to the Register message with a Reject message (i.e., rejecting Node X's request for a connection to Node N).

In another example, the registration process may be extended to include some system maintenance or configuration activities. Such maintenance or configuration activities may include "flood testing" (discussed in detail below), identification of a bad protocol, and/or Firewall Status (discussed in detail below).

In this regard, FIG. 34A shows the messaging (in connection with registration with a flood test) between Node X, a node requesting to join the network, Node S, a server node

within the network, and Node R (e.g., a random (or predetermined) node that may, but not need be, within the network).

Further, FIG. 34B shows the messaging (in connection with registration with a bad protocol version) between Node X and Node S, a server node within the network. Of note, such a bad protocol message may be sent by the server to a connection requesting node to inform that node that it is running an obsolete version of the client software and must upgrade to the current version before being allowed to join the network.

Further still, FIG. 34C shows the messaging (in connection with registration with standard response) between Node X and Node S, a server node within the network (wherein such standard response includes, for example, Accept / Reject / Path).

Referring now in more detail to the "flood test" example shown in FIG. 34A, this example process begins, as above, with Node X making a request to Node S to join the network. Node S then sends Node X a Meta message (such a Meta message may be sent by the server to a connection requesting node to inform that node of certain characteristics of the network such as, for example: the streaming bandwidth rate used by the network, the server's public key for authenticating that secure packets originated from this server, and any other information the architects of the network might deem necessary or desirable for nodes to know when connecting to the network). Node S may also require that Node X undergo performance or flood testing before being accepted or placed within the network. In such an event, Node S responds to Node X's request to join the network with an instruction that Node X undergo a flood test (e.g., of Node X's upstream channel). Such an instruction may be accompanied by an address of Node R, the illustrative flood test target. Note that any node within or without the network may conduct the flood test. One purpose of the flood test may be to permit Node S to be aware of Node X's upstream capability before determining where (or if) to assign Node X in the network.

Where Node S requires that Node X undergo performance or flood testing before being accepted or placed within the network, Node S provides a message so informing Node X, shown as "DoFlood" in FIG 34A. The DoFlood message may provide to Node X the address of the flood test target, but the flood test target could be Node S itself, or a predetermined flood test target.

In response to the DoFlood message, Node X sends an InitFlood message to the flood test target. In one embodiment, Node X would await an acknowledgement (not shown) of its InitFlood message, but in another embodiment, awaiting an acknowledgement is not

necessary. Node X would then begin transmitting FloodData messages, as shown in FIG 34A.

Since, in this example, FloodData messages are transmitted for a purpose of determining Node X's upstream bandwidth, in one embodiment, Node X would transmit an amount of FloodData to the flood test target as shown in FIG 34A. In various embodiments, the maximum amount of FloodData may be limited to a fixed number of FloodData messages, to a fixed amount of data transmitted within FloodData messages, or to transmission of FloodData or FloodData messages over a fixed period of time.

To protect the flood test target from being inundated with FloodData (e.g., where Node X has a very high upstream bandwidth capability), Node X may cap (e.g., be so directed by Node S) the FloodData transmission to, e.g., 100 Kbits/sec, 200 Kbits/sec, or some other data rate that would be a desired data transmission rate for a node on the particular network of interest.

Thus, by receiving the FloodData messages, the flood test target can determine, for example, an upstream capability rating of Node X (provided that the flood test target has the same or more bandwidth in the downstream channel compared to the upstream capability of Node X.

In another example, both the target and the sender will know the bandwidth of the sender at the conclusion of the flood test. More particularly, the present invention may use TCP/IP (rather than UDP), and all packets sent may receive an acknowledgement. In this example, Node X, the node sending out the flood packets, sends out the next flood packet only after receipt of the first packet has been acknowledged by the target of the flood test. Thus, the number of flood data packets Node X is able to send per unit time (along with the size of those packets) may allow Node X to calculate its upstream bandwidth.

In another example, the target of the "flood test" may be any node including the server but excluding any firewalled or low-bandwidth nodes.

After the "flood test", Node X may then connect with Node S (or be instructed to go to another node if Node S can not/will not accept Node X). Of note, if Node X fails the "flood test", Node X may be sent far downtree or may be refused entry into the network entirely. Further, if node N is accepted under these circumstances it will be deemed a "bumpable" non-repeat-capable node.

Referring now in more detail to the "Firewall Status" example shown in FIG. 34C, this example process begins, as above, with Node X making a request to Node S to join the network. Node S then send Node X a Meta message. A firewall status message is a message

sent over a new connection from (in the case of Figure 34C) Node S to Node X. If Node X receives this message, over a communications channel initiated by Node S, Node X will know that it is not firewalled. Failure to receive a firewall status message when a node is expecting such a message allows that node to conclude that it is in fact firewalled. Thus, if Node X is expecting to receive a firewall status message over new connection initiated by some other node and that message does not arrive, Node X can conclude (with a high degree of likelihood of being correct) that it is firewalled. Also, since (in one example) the present invention uses the TCP/IP messaging protocol the sender, Node S, will also know whether the firewall status message was successfully received by Node X and thus Node S will also know whether Node X is or is not firewalled (one or more other nodes may also receive data regarding the firewall status).

After receiving the firewall status report (not shown) Node S will then respond with a standard response (accept, path, reject) as discussed above.

To now give a specific example of the use of both the "flood test" and the "firewall status" mechanisms during registration of a new node, FIG. 35A shows a state of the network before Node X joins the network and FIG. 35B shows a state of the network after Node X joins the network. Further, FIG. 36 shows the messaging among Node X, Node S, Node B and Node R. More particularly:

- Node X sends a Register message to Node S and Node S responds with a Meta message (providing Node X with basic information about the network) and a DoFlood message (which requests Node X to determine its bandwidth and report back to Node S when Node X has that information). Node S knows to send a "DoFlood" to Node X because the initial Register message sent by Node X to Node S contains an indication that Node X does not yet know its current upstream bandwidth capacity. The DoFlood message will contain the address of a target node (Node R), selected by Node S, that Node X should flood against.
- Node X sends an InitFlood message to Node R and then "floods" Node R with data for some period of time (e.g., 15 seconds). At the conclusion of this test Node X will be able to compute its upstream bandwidth based on the number and size of data packets it was able to successfully transmit to Node R.

- Node X sends another Register message to Node S and Node S responds with a Meta message. Contained within this second Register message is the result of Node X's "flood test" which gives an indication of Node X's upstream bandwidth capacity.
- Node S now tests whether external nodes can initiate communication with Node X by sending Node X a Firewall Status message over a new connection. If this message is successfully received by Node X (and acknowledgement received by Node S), both Nodes X and S will know that Node X is not firewalled (on a network communication port).
- Node S sends a Path message to Node X. This message will include a list of nodes that X should attempt connection with, beginning with the preferred target node (e.g., Node B). As described elsewhere in the application, the choice of target node, and hence the contents of the path, may be influenced by both the current configuration of the network and whether or not Node X is repeat-capable (based on the results of the "flood test" and "firewall test").
- Node X sends a Register message to Node B and Node B responds with an Accept message.

Referring now to a "Firewall Request / Firewall Status" example involving two arbitrary Nodes A and B (neither of which is necessarily a server), FIG. 37 shows a portion of a network including Node A and Node B. In this example (as shown in FIG. 38), a Firewall Request message is sent by Node B to request that the recipient of the message, Node A, test Node B's Firewall Status. The receiving Node A may send (e.g., over a new connection), the Firewall Status message to Node B. If Node B timely receives and interprets the Firewall Status message it can be certain that it is a non-firewalled node. If Node B does not timely receive a Firewall Status message from Node A then Node B may assume that it is firewalled.

In another example (which example is intended to be illustrative and not restrictive), if Node A fails to connect to Node B or to send the Firewall Status message no further action is needed on the part of Node A.

Node B may also augment its assumption about its Firewall Status by using a three strikes (i.e., three attempts) method before considering itself to be firewalled (of course, three strikes is provided as an example only, and any desired number of "strikes" or attempts may be used).

In another example (which example is intended to be illustrative and not restrictive), the multiple attempts may be carried out within a predetermined time "window" (after which receiving no Firewall Status message Node B may consider itself firewalled).

In another example (which example is intended to be illustrative and not restrictive), Firewall Requests may be sent out by all nodes (e.g., at regular time intervals) in order to maintain a current and accurate picture of their Firewall Status.

In another example (which example is intended to be illustrative and not restrictive), Firewall Requests may be sent out by certain node(s) (e.g., at regular time intervals) in order to maintain a current and accurate picture of their Firewall Status.

In another example (which example is intended to be illustrative and not restrictive), the Firewall Status message may be sent by the server on initial registration (e.g., without the need for a Firewall Request message and both the sender (the server) and the receiving node (the node that is registering) may use the successful delivery of the Firewall Status message as an indicator of the connecting node's Firewall Status.

Referring now to a "Propagate Path / Down" example (which example is intended to be illustrative and not restrictive), FIG. 39A shows a portion of a network including Node A and Node B. Further, FIG. 39B shows an example of nodes in a tree sending Propagate Path / Down messages downtree to their children (e.g., at regular time intervals).

In one example (which example is intended to be illustrative and not restrictive), the Propagate Path / Down message may include topology information containing a path back to the root server and the time interval for sending such message may be calculated on a node-by-node and/or level-by-level basis relative to the amount of data to be sent in the Propagate Path / Down message (wherein the amount of data in the Propagate Path / Down message is related to a particular node's (or level's) downtree position relative to the root server). Note, that although this time interval may not be very important for Propagate Path / Down, since the amount of data for each successive downtree node (or level) may increase relatively slowly (e.g., linearly), the time period may be of critical importance for Propagate NTM / Up (which is discussed below), since the amount of data for each successive uptree node (or level) may increase relatively quickly (e.g., exponentially). Of note, the acronym "NTM" refers to the Network Topology Model.

In any case, FIG. 40 shows the messaging between Node A and Node B in connection with a Propagate Path / Down message (i.e., Node A sends a Propagate Path message downtree to Node B).

Referring now to a "Propagate NTM / Up" example (which example is intended to be illustrative and not restrictive), FIG. 41A shows a portion of a network including Node A and Node B. Further, FIG. 41B shows an example of nodes in a tree sending Propagate NTM / Up messages uptree to their parents (e.g., at regular time intervals).

In one example (which example is intended to be illustrative and not restrictive), the Propagate NTM / Up message may include information containing the topology of the subtree(s) below a particular node and the time interval for sending such message may be calculated on a node-by-node and/or level-by-level basis relative to the amount of data being sent in the Propagate NTM / Up message (wherein the amount of data in the Propagate NTM / Up message is related to a particular node's (or level's) uptree position relative to tree height (e.g., the height of a particular node above its lowest leaf level)). Note that in this example (which example is intended to be illustrative and not restrictive), the NTM data that the node sends up is rooted at itself and contains both its children.

In any case, FIG. 42 shows the messaging between Node A and Node B in connection with a Propagate NTM / Up message (i.e., Node B sends a Propagate NTM message uptree to Node A).

In another embodiment, the following process may be employed when a node enters the network. More particularly, when the server adds a node X to the system (gives it a connection path to node Y) the server will add node X to its network topology model as a child of node Y. In other words, the server will assume that the connection will take place normally with no errors. The reason why doing so may be critical to the functioning of the system in this embodiment is that if the server failed to account for the fact that node X is now (probably) a child of node Y, using the "keep the tree balanced" approach embodied in the Universal Connection Address Routine, all new nodes connecting to the system would be assigned as children of node Y up until the point where a propagate up arrives indicating that node X was successfully added to the tree. If the tree is large and node Y is many levels removed from the server it could be some time (e.g., a minute or more, in very large trees several minutes) before confirmation arrives that node X truly connected to node Y as its child.

Now, in addition to simply placing node X into the server's NTM (Network Topology Model) and being done with the matter, in this embodiment the system must account for the fact that it may take a number of propagate NTM/up cycles for the true status of node X to be reported back to the server. Thus, for some number of propagate cycles (or some amount of time) the server may compare the newly propagated (updated) NTM it is receiving from its

children to the list of "recently added nodes". If a recently added node is not in the propagated NTM it will be re-added to the NTM as a child of node Y. At some point node X will no longer be considered "recently added" and the server will cease adding node X back into its NTM. In other words, the server will assume that, for whatever reason, node X never successfully connected to the network.

In another embodiment, the following process may be employed when a node gracefully exits the system. More particularly, when a node (node X) leaves the network by performing a graceful exit (e.g., the user selects "exit" or clicks the "x" or issues some other command to close the application) the node will send signals to its parent and child nodes (if any) immediately prior to actually shutting down. Upon receipt of such a "depart" message, the parent of the departing node (node P) will modify its internal NTM (Network Topology Model) to a configuration that corresponds to the predicted shape of the network after the departing node exits. In this embodiment, node P will assume that its NTM is accurate and that the entire cascade of reconfiguration events initiated by the departure of node X will happen normally. As such, node P will now expect the former green child of node X (node XG -- if such a child exists) to connect – essentially taking node X's place in the network.

At this point, node P's NTM represents a prediction of the (anticipated) state of the network rather than the actual reported configuration. In order to indicate that node P does not (yet) have a physical connection to node XG, node P will mark node XG as "soft" in its NTM – meaning node XG is expected to take up this position in the network but has not yet physically done so. Node P will report this "predicted" NTM (along with node XG's status as a soft node) up during its propagate NTM/Up cycles, until node P physically accepts a child connection in node XG's "reserved" spot and receives an actual report of the shape of the downtree NTM from that child, via propagate NTM/ups.

Of note, the above actions may be taken in order to have the propagated NTM's reflect, as accurately as possible, the topology of the network.

Of further note, because of the distributed nature of a network according to the present invention, a complete and accurate picture of the current shape of the network at any instance in time can never be guaranteed. In fact, though the root server has the "widest" view of the network, without the heuristics described herein, that view would be the "most dated" of any node in the network. Thus, heuristics such as those just described may be vital to producing NTM's that are as close to complete and accurate as possible (e.g., so that the Universal Connection Address Routine can position incoming nodes so as to keep the distribution tree as nearly balanced as possible).

Referring now to a "Banned List" example (which example is intended to be illustrative and not restrictive), it is noted that such a "banned list" (or "banished list") may be used for keeping certain nodes off of the network, removing certain nodes from the network, and/or keeping certain nodes as far downtree as possible (e.g., in the case of a binary tree network keeping certain nodes as "red" nodes).

More particularly, FIG. 43 shows a portion of a network including Node A, Node B, Node C and Node D and FIG. 44 shows the messaging between Node A, Node B, Node C and Node D in connection with Banned List messages (i.e., in this FIG. 44, Node A sends a Banned List message to Node B and Node B sends a Banned List message to Nodes C and D).

In one example (which example is intended to be illustrative and not restrictive), nodes receiving this message should re-transmit the message downtree to all children in an unaltered fashion.

In another example (which example is intended to be illustrative and not restrictive), the Banned List messages may include IP lists and/or URL's.

In another example (which example is intended to be illustrative and not restrictive), the Banned List messages may originate at the server.

In another example (which example is intended to be illustrative and not restrictive), Banned List messages may be sent and/or originated by any desired node(s).

In another example (which example is intended to be illustrative and not restrictive), Banned List messages and may be sent using a Secure Packet format.

In another example (which example is intended to be illustrative and not restrictive), node(s) may store the content of the message (e.g., the banned list itself) for future use.

In another example (which example is intended to be illustrative and not restrictive), the server (or any other desired node(s)) may send out and/or originate the Banned List messages at regular time intervals.

Referring now to a "Bumpable Node" embodiment, it is noted that this "bumpable node" mechanism may be used to "bump" (or send downtree) a particular node (e.g., when a new node is added). More particularly, in one example (which example is intended to be illustrative and not restrictive), a server may instruct Node X to connect with Node P and Node P may know to "bump" its child, Node C, by: (a) instructing Node C to disconnect from Node P; (b) allowing Node X to connect to Node P; and (c) instructing Node C to re-connect as a child of Node X. Of note, such a "bumping" operation may be carried out, for example (which example is intended to be illustrative and not restrictive), because the

bumped node is firewalled and/or because the bumped node has failed an upstream and/or downstream bandwidth test (which test may be carried out by the server or by any other node(s)).

Of note, as used in this application, the terms "bumpable nodes" or "bumpable non-repeat-capable nodes" or "non-repeat-capable nodes" are nodes which either: (a) have insufficient bandwidth to act as repeaters in the network or (b) are firewalled.

Of further note, in one example the more inclusive "is bumpable" test may replace the "is low bandwidth" test in the Universal Connection Address Routine of Figure 20.

Referring now to a "Load Manager" embodiment, it is noted that this "load manager" mechanism may be used to rate limit (i.e., slow or stop) the addition of new nodes to the network. Such rate limiting may be useful because the larger the network grows, the more nodes can join per unit of time.

In one example (which example is intended to be illustrative and not restrictive), the rate limited addition of new nodes may be carried out by a server (or connection server) and may be based upon the length of time that it takes for the buffers associated with one or more nodes (or levels) of the existing network to fill (or reach a certain value). In a more specific example, the rate limited addition of new nodes may be based upon the length of time that it takes for the buffers associated with potential parents of the new nodes to fill (or reach a certain value).

In another example (which example is intended to be illustrative and not restrictive), the rate limited addition of new nodes may be carried out by any desired node(s) (e.g., one or more user nodes and/or a combination of one or more user nodes and a server (or connection server) node).

Referring now to a "Node Interlacing" embodiment, it is noted that this "node interlacing" mechanism may be used to add new nodes to the network by interlacing the new nodes on a level-by-level basis as shown in FIG. 45.

More particularly, as seen in this FIG. 45, new nodes may be added in the order indicated (as can be seen below, this example (which example is intended to be illustrative and not restrictive) relates to a binary tree network). That is, Node 1 may first be added as a child of the server followed by Node 2 as a child of the server – this will fully populate the first level; Node 3 may next be added as a child of Node 1, followed by Node 4 as a child of Node 2, followed by Node 5 as a child of Node 1, followed by Node 6 as a child of Node 2 – this will fully populate the second level; Node 7 may next be added as a child of Node 3, followed by Node 8 as a child of Node 5, followed by Node 9 as a child of Node 4, followed

by Node 10 as a child of Node 6, followed by Node 11 as a child of Node 3, followed by Node 12 as a child of Node 5, followed by Node 13 as a child of Node 4, followed by Node 14 as a child of Node 6 – this will fully populate the third level, etc.

Of note, such node interlacing may be used to efficiently fill (or bring to a certain value) the buffers associated with one or more nodes (or levels) of the existing network while new nodes are added.

Of further note, such node interlacing may help prevent adding a second child to a parent until the first child's buffer is full (or brought to a certain value).

Of still further note, such node interlacing may help to: (a) prevent taxing the parent's downtree channel; and/or (b) prevent two children being left with partially filled buffers in the event that their parent departs.

Of still further note, such node interlacing may be used, for example, in conjunction with or as an alternative to the load balancing mechanism discussed above.

In another embodiment the present invention may be used when a plurality of computer systems are connected (e.g., to the Internet) by one or more of an ISDN line, a DSL (e.g., ADSL) line, a cable modem, a T1 line and/or an even higher capacity link.

In another embodiment a distribution network for the distribution of data from a server node to a plurality of user nodes connected to the server node, wherein each of the user nodes has associated therewith a buffer is provided, which distribution network comprises: a mechanism for rate limiting the addition of new user nodes to the distribution network, wherein the rate limited addition of new user nodes to the distribution network is based at least in part upon the length of time that it takes for at least one buffer associated with at least one user node to reach a predetermined percentage of capacity, wherein the predetermined percentage of capacity is greater than zero.

In another embodiment a distribution network for the distribution of data from a server node to a plurality of user nodes connected to the server node, wherein each of the user nodes is in one of a plurality of levels located downtree from the server node is provided, which distribution network comprises: a mechanism for adding new user nodes to the distribution network in an interlaced manner, wherein the interlacing provides that each sibling user node in a given level of the distribution network is associated with just one child before any siblings in the given level are associated with two children.

The following embodiments will now describe a number of additional example algorithms which may be utilized in whole or in part to implement the present invention:

In one embodiment a process for docking a connection requesting user node with a distribution network is provided, said process including the following steps:

(a) having an instructing node receive a connection request from said connection requesting user node;

(b) forming an instructing node topology database indicating, at a point in time:

(i) which, if any, user nodes are docked with said instructing node as child nodes; which, if any, user nodes are docked with each of said child nodes as grandchild nodes of said instructing node; which, if any, user nodes are docked with each of said grandchild nodes as great-grandchild nodes of said server node; and so on; and

(ii) said docked user nodes respective bandwidth capacities;

(c) selecting from a group of nodes including the instructing node and said docked user nodes a recommended parent node for said connection requesting user node, wherein said recommended parent node has apparent available capacity to transmit content data to said connection requesting user node and is at least as close (i.e., in terms of network topology) to said instructing node as any other docked user node having apparent available capacity to transmit content data to said connection requesting user node (of note, placement may also or alternatively depend, as described elsewhere in the present application, in whole or in part on whether the connection requesting node is repeat-capable or not (i.e., whether it can support children)); and

(d) providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said recommended parent node back to said instructing node.

In one example, the process may further comprise the following steps:

(e) having said connection requesting user node go to (or attempt to go to) the node at the top of said connection address list;

(f) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(g) if the node at top of said connection address list is not part of the distribution network, deleting such node from said connection address list and repeating steps (e) and (f) with respect to the next node at the top of said connection address list;

(h) if the node at the top of said connection address list is part of said distribution network, having said connection requesting user node dock with said node at the top of said connection address list.

In another example, step (d) may further comprise the following step:

(i) if said instructing node is not a server node, including in said connection address list a list of nodes cascadingly connected with each other from the instructing node back to said server node.

In another example, the process may further comprise the following steps:

(e) having said connection requesting user node go to (or attempt to go to) the node at the top of said connection address list;

(f) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(g) if the node at the top of said connection address list is not part of the distribution network, deleting such node from said connection address list and repeating steps (e) and (f) with respect to the next node at the top of said connection address list;

(h) if the node at the top of said connection address list is part of said distribution network, determining whether said node at the top of said connection address list has available capacity to transmit content data to said connection requesting user node; and

(i) if the node at the top of said connection address list is part of said distribution network and has available capacity to transmit content data to said connection requesting user node, having said connection requesting user node dock with the node at the top of said connection address list.

In another example, the process may further comprise the following steps:

(j) if the node at the top of connection address list is part of said distribution network and said node at the top of said connection address list does not have available capacity to transmit content data to said connection requesting user node, repeating steps (a)-(d), with said node at the top of connection address list being said instructing node and a new connection address list being the product of steps (a)-(d).

In another example, the process may further comprise the steps of counting the times that step (j) is performed and, when step (j) is performed a predetermined number of times, having said connection requesting user node repeat steps (a)-(d) with said primary server being the instructing node.

In another embodiment a process for docking a connection requesting user node with a distribution network is provided, said process including the following steps:

(a) having an instructing node receive a connection request from said connection requesting user node;

(b) forming an instructing node topology database indicating, at a point in time:

(i) which, if any, user nodes are docked with said instructing node as child nodes; which, if any, user nodes are docked with each of said child nodes as grandchild nodes of said instructing node; which, if any, user nodes are docked with each of said grandchild nodes as great-grandchild nodes of said server node; and so on; and

(ii) said docked user nodes' respective bandwidth capacities;

(c) assigning utility ratings to each of the docked user nodes, said utility ratings being a function of at least the bandwidth capacity of the docked user node and the elapsed time that the docked user node has been docked with the distribution network;

(d) selecting from a group of nodes including the instructing node and said docked user nodes a recommended parent node for said connection requesting user node, wherein said recommended parent node has apparent available capacity to transmit content data to said connection requesting user node and is at least as close (i.e., in terms of network topology) to said instructing node as any other docked user node having apparent available capacity to transmit content data to said connection requesting user node, and when the closest (i.e., in terms of network topology) docked user node with apparent available capacity to transmit content data to said connection requesting user node is equidistant (i.e., in terms of network topology) to said instructing node with at least one other docked user node with apparent available capacity to transmit content data to said connection requesting user node, said recommended parent node is the docked user node having the highest utility rating among such equidistant (i.e., in terms of network topology) docked user nodes with apparent available capacity; and

(e) providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said recommended parent node back to said instructing node and, if said instructing node is not a primary server node, back to said primary server node.

In another embodiment a process for docking a connection requesting user node with a distribution network is provided, said process including the following steps:

(a) having an instructing node receive a connection request from said connection requesting user node;

(b) forming an instructing node topology database indicating, at a point in time:

(i) which, if any, user nodes are docked with said instructing node as child nodes; which, if any, user nodes are docked with each of said child nodes as grandchild nodes of said instructing node; which, if any, user nodes are docked with

each of said grandchild nodes as great-grandchild nodes of said instructing node; and so on; and

(ii) said docked user nodes' respective bandwidth capacities, including a designation of whether said respective docked user node's bandwidth capacity is below a predetermined threshold (e.g., said respective docked user node is a low-bandwidth node) or at least said predetermined threshold (e.g., said respective docked user node is a high-bandwidth node);

(c) forming a primary recommended parent node list ("PRPL") comprised of the instructing node (if it has available capacity to transmit content data to said connection requesting user node) plus those docked user nodes having apparent available capacity to transmit content data to said connection requesting user node, with said instructing node being placed first on the PRPL (if it is on the PRPL) and said docked user nodes having apparent available capacity to transmit content date to said connection requesting user node being ranked with those docked nodes which are closer (i.e., in terms of network topology) to the instructing node being ranked higher than those docked nodes which are further away (i.e., in terms of network topology) and with equidistant (i.e., in terms of network topology) docked user nodes being ranked such that those docked nodes which have the higher utility ratings being ranked higher than docked nodes having lower utility ratings;

(d) forming a secondary recommended parent node list ("SRPL") comprised of the instructing node (if it has no available capacity to transmit content data to said connection requesting user node but does have at least one low-bandwidth node docked directly to it) plus said docked user nodes which (i) are high-bandwidth nodes with no available capacity to transmit content data to said connection requesting user node and (ii) have at least one low-bandwidth node docked directly to it, with said instructing node being placed first on the SRPL (if it is on the SRPL) and said docked user nodes on the SRPL being ranked with those docked nodes which are closer (i.e., in terms of network topology) to the instructing node being ranked higher than those docked nodes which are further away (i.e., in terms of network topology) and with equidistant (i.e., in terms of network topology) docked user nodes being ranked such that those docked nodes which have the higher utility ratings being ranked higher than docked nodes having lower utility ratings;

(e) determining whether the connection requesting user node is a low-bandwidth node;

(f) if the connection requesting user node is a low-bandwidth node:

(i) selecting the highest ranked node on the PRPL as a recommended parent node; and

(ii) providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said recommended parent node back to said instructing node and, if said instructing node is not a primary server node, back to said primary server node; and

(g) if the connection requesting user node is a high-bandwidth node:

(i) selecting the highest ranked node on the PRPL as a recommended parent node;

(ii) selecting the highest ranked node on the SRPL as an alternate recommended parent node;

(iii) determining whether the alternate recommended parent node is closer (i.e., in terms of network topology) to the server node than the recommended parent node;

(iv) if the alternate recommended parent node is not closer (i.e., in terms of network topology) to the server node than the recommended parent node, providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said recommended parent node back to said instructing node and, if said instructing node is not the primary server node, back to said primary server node;

(v) if the alternate recommended parent node is closer (i.e., in terms of network topology) to the server node than the recommended parent node, (1) forcing the disconnection of a low-bandwidth node from the alternate recommended parent node and (2) providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said alternate recommended parent node back to said instructing node and, if said instructing node is not the primary server node, back to said primary server node.

In one example, step (f)(ii) may include the following steps:

(1) adding said connection requesting user node to the topology database as a child of the recommended parent node; and

(2) if said recommended parent node would have no apparent available capacity to transmit content data to an additional node with said connection requesting user node docked with it, deleting the recommended parent node from the PRPL and adding it to the SRPL.

In another example, step (g)(iv) may include the following steps:

(1) adding said connection requesting user node to the topology database as a child of the recommended parent node;

(2) if said recommended parent node would have no apparent available capacity to transmit content data to an additional node with said connection requesting user node docked with it, deleting the recommended parent node from the PRPL; and

(3) if said recommended parent node is deleted from the PRPL but has at least one low-bandwidth node docked directly to it, adding the recommended parent node to the SRPL.

In another example, step (g)(v) may include the following steps:

(3) adding said connection requesting user node to the topology database as a child of the alternate recommended parent node; and

4) if said alternate recommended parent node would have no apparent available capacity to transmit content data to an additional node with said connection requesting user node docked with it, deleting the alternate recommended parent node from the SRPL.

In another example, the process may further comprise the following steps:

(h) having said connection requesting user node go to (or attempt to go to) the node at the top of said connection address list;

(i) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(j) if the node at top of connection address list is not part of the distribution network, deleting such node from connection address list and repeating steps (h) and (i) with respect to the next node at the top of connection address list; and

(k) if the node at top of connection address list is part of said distribution network, having said connection requesting user node dock with said node at the top of connection address list.

In another example, the process may further comprise the following steps:

(h) having said connection requesting user node go to (or attempt to go to) the node at the top of connection address list;

(i) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(j) if the node at top of connection address list is not part of the distribution network, deleting such node from connection address list and repeating steps (h) and (i) with respect to the next node at the top of connection address list;

(k) if the node at the top of connection address list is part of said distribution network, determining whether said node at the top of connection address list has available capacity to transmit content data to said connection requesting user node; and

(l) if the node at the top of said connection address list is part of said distribution network and has available capacity to transmit content data to said connection requesting user node, having said connection requesting user node dock with the node at the top of said connection address list.

In another example, the process may further comprise the following steps:

(m) if the node at the top of said connection address list is part of said distribution network and the node at the top of said connection address list does not have available capacity to transmit content data to said connection requesting user node, repeating steps (a)-(g), with the node at the top of said connection address list being said instructing node and a new connection address list being the product of steps (a)-(g).

In another example, the process may further comprise the steps of counting the times that step (m) is performed and, when step (m) is performed a predetermined number of times, having said connection requesting user node repeat steps (a)-(g) with said primary server being the instructing node.

In another embodiment a process for docking a connection requesting user node with a distribution network is provided, said process including the following steps:

(a) having an instructing node receive a connection request from said connection requesting user node;

(b) determining whether said connection requesting user node's bandwidth capacity is below a predetermined threshold (e.g., said connection requesting user node is a low-bandwidth node) or at least said predetermined threshold (e.g., said connection requesting user node is a high-bandwidth node);

(c) forming an instructing node topology database indicating, at a point in time:

(i) which, if any, user nodes are docked with said instructing node as child nodes; which, if any, user nodes are docked with each of said child nodes as grandchild nodes of said instructing node; which, if any, user nodes are docked with each of said grandchild nodes as great-grandchild nodes of said server node; and so on; and

(ii) said docked user nodes' respective bandwidth capacities;

(d) if said connection requesting user node is a high-bandwidth user node, selecting from a group of nodes including the instructing node and said docked user nodes a recommended parent node for said connection requesting user node, wherein said recommended parent node has apparent available capacity to transmit content data to said connection requesting user node and is at least as close (i.e., in terms of network topology) to said instructing node as any other docked user node having apparent available capacity to transmit content data to said connection requesting user node;

(e) if said connection requesting user node is a low-bandwidth user node, selecting from a group of nodes including the instructing node and said docked user nodes a recommended parent node for said connection requesting user node, wherein said recommended parent node has apparent available capacity to transmit content data to said connection requesting user node and is at least as far (i.e., in terms of network topology) from said instructing node as any other docked user node having apparent available capacity to transmit content data to said connection requesting user node; and

(f) providing a connection address list to said connection requesting user node, said connection address list listing each node in said instructing node's topology database from said recommended parent node back to said instructing node.

In one example, step (f) may further comprise the following step: (i) if said instructing node is not a server node, including in said connection address list a list of nodes cascadingly connected with each other from the instructing node back to said server node.

In another embodiment a process for connecting a connection requesting user node to a computer information distribution network having a primary server node and user nodes docked therewith in a cascaded relationship is provided, said process further comprising the following steps:

(a) providing said connection requesting user node with a connection address list which sets forth a list of user nodes docked in series with each other back to said primary server;

(b) having said connection requesting user node go to (or attempt to go to) the node at top of said connection address list;

(c) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(d) if the node at the top of said connection address list is not part of the distribution network, deleting such node from said connection address list and repeating steps (b) and (c) with respect to the next node at the top of connection address list; and

(e) if the node at the top of said connection address list is part of said distribution network, having said connection requesting user node dock with the node at the top of said connection address list.

In another embodiment a process for connecting a connection requesting user node to a computer information distribution network having a primary server node and user nodes docked therewith in a cascaded relationship is provided, said process further comprising the following steps:

(a) providing said connection requesting user node with a connection address list which sets forth a list of user nodes docked in series with each other back to said primary server;

(b) having said connection requesting user node go to (or attempt to go to) the node at the top of said connection address list;

(c) having said connection requesting user node determine whether the node at the top of said connection address list is part of the distribution network;

(d) if the node at the top of said connection address list is not part of the distribution network, deleting such node from connection address list and repeating steps (b) and (c) with respect to the next node at the top of said connection address list;

(e) if the node at the top of said connection address list is part of said distribution network, determining whether the node at the top of said connection address list has available capacity to transmit content data to said connection requesting user node; and

(f) if the node at the top of said connection address list is part of said distribution network and has available capacity to transmit content data to said connection requesting user node, having said connection requesting user node dock with the node at the top of said connection address list.

In another embodiment relating to a computer information distribution network comprising a primary server node and user nodes docked therewith in a cascaded relationship, wherein each user node may have no more than a predetermined maximum number of child nodes docked with it, a process for reconfiguring the network in the event of a user node's departure therefrom is provided, said process including the following steps:

(a) providing information at a point in time to a node in the distribution network as follows:

(i) the node's propagation rating if it is a user node having a parent node which is a user node, wherein a propagation rating is one of a predetermined number of grades ranging from highest to second highest to third highest and so on, with the number of grades being equal to said predetermined maximum number;

(ii) an ancestor list setting forth the node's ancestor nodes' addresses (if it has any ancestor nodes) back to the primary server node, with the node's parent node being atop the ancestor list;

(iii) a sibling list of the node's sibling nodes' addresses (if it has any sibling nodes) and their respective propagation ratings (of note, in another example (which example is intended to be illustrative and not restrictive), nodes may not have data about their sibling (i.e., nodes do not know who their sibling is); instead, the mechanism used to get a red node to connect to its green sibling may be for the parent of those two nodes (or a proxy for that parent) to send the red child: (1) a connection path with that red child's green sibling at the top of the path; and (2) a command to "priority join" to that green child - when the green child node accepts the incoming "priority join" connection request from its former red sibling, the green node will kick its own red child (if such a child exists) instructing that red child to join as a child of its own green sibling (these actions will cause a chain reaction of reconfigurations which will cascade until the edge of the tree is reached)); and

(iv) a child list of the node's child nodes' addresses (if it has any child nodes) and their respective propagation ratings;

(b) having a first node send a propagation signal to each former child node of a departed node, wherein said first node is an ancestor of said former child node of said departed node;

(c) with respect to each child node receiving the propagation signal which did not have the highest propagation rating prior to its receiving the propagation signal (hereinafter referred to as a "red node"), upon its receiving the propagation signal:

(i) setting the propagation rating of the red node to the next higher grade above the propagation rating of that red node before it received the propagation signal;

(ii) having the red node undock from its parent node (if it was docked with its parent node before it received the propagation signal); and

(iii) having the red node dock with its sibling node which had the highest propagation rating prior to the red node's receiving the propagation signal;

(d) with respect to each child node receiving the propagation signal which had the highest propagation rating prior to its receiving a propagation signal (hereinafter referred to as a "green node"), upon its receiving the propagation signal:

(i) setting the propagation rating of that green node to the lowest grade;

(ii) having the green node retransmit the propagation signal to its child nodes (if it has any); and

(iii) determining whether the green node is docked with the node from which the green node received the propagation signal and if it is, having the green node remain docked with said node from which the green node received the propagation signal, and if it is not, having the green node dock with said node from which the green node received the propagation signal; and

(e) repeating steps (c) and (d) with respect to each user node which receives a retransmitted propagation signal.

In another embodiment relating to a computer information distribution network comprising a primary server node and user nodes docked therewith in a cascaded relationship, wherein each user node may have no more than a predetermined maximum number of child nodes docked with it, a process for reconfiguring the network in the event of a user node's departure therefrom is provided, said process including the following steps:

(a) providing information at a point in time to a node in the distribution network as follows:

(i) the node's propagation rating if it is a user node having a parent node which is a user node, wherein a propagation rating is one of a predetermined number of grades ranging from highest to second highest to third highest and so on, with the number of grades being equal to said predetermined maximum number;

(ii) an ancestor list setting forth the node's ancestor nodes' addresses (if it has any ancestor nodes) back to the primary server node, with the node's parent node being atop the ancestor list;

(iii) a sibling list of the node's sibling nodes' addresses (if it has any sibling nodes) and their respective propagation ratings; and

(iv) a child list of the node's child nodes' addresses (if it has any child nodes) and their respective propagation ratings;

(b) having a first node send a propagation signal to each former child node of a departed node, wherein said first node is an ancestor of said former child node of said departed node;

(c) with respect to each child node receiving the propagation signal which did not have the highest propagation rating prior to its receiving the propagation signal (hereinafter referred to as a "red node"), upon its receiving the propagation signal:

(i) setting the propagation rating of the red node to the next higher grade above the propagation rating of that red node before it received the propagation signal;

(ii) having the red node undock from its parent node (if it was docked with its parent node before it received the propagation signal);

(iii) devising a first connection address list comprising the address of the red node's sibling node which had the had the highest propagation rating prior to the red node's receiving the propagation signal followed by the red node's ancestor list;

(iv) having said red node go to (or attempt to go to) the node at the top of said first connection address list;

(v) having said red node determine whether the node at the top of said first connection address list is part of the distribution network;

(vi) if the node at the top of said first connection address list is not part of the distribution network, deleting such node from said first connection address list and repeating steps (c)(iv) and (c)(v) with respect to the next node at the top of said first connection address list;

(vii) if the node at the top of said first connection address list is part of said distribution network, determining whether the node at the top of said first connection address list has available capacity to transmit content data to said red node; and

(viii) if the node at the top of said first connection address list is part of said distribution network and has available capacity to transmit content data to said red node, having said red node dock with the node at the top of said first connection address list (with regard to these steps (vii) and (viii), in another example (which example is intended to be illustrative and not restrictive), when a red node connects to a green node as part of a network reconfiguration event, it does not matter whether there is "available capacity" or not for the green node to accept its incoming red sibling – the green node has no choice but to accept the node (these actions are facilitated in this example by the red node issuing a "priority join" to the green node; if the green node does not have "available capacity" it has to kick its own red child to make room for the incoming green node));

(d) with respect to each child node receiving the propagation signal which had the highest propagation rating prior to its receiving a propagation signal (hereinafter referred to as a "green node"), upon its receiving the propagation signal:

(i) setting the propagation rating of that green node to the lowest grade;

(ii) having the green node retransmit the propagation signal to its child nodes (if it has any);

(iii) determining whether the green node is docked with the node from which the green node received the propagation signal and if it is, having the green node remain docked with said node from which the green node received the propagation signal, and if it is not, devising a second connection address list starting with said node from which the green node received the propagation signal and followed by that portion of said green node's ancestor list extending back to the primary server from said node from which the green node received the propagation signal;

(iv) having said green node go to (or attempt to go to) the node at the top of said second connection address list;

(v) having said green node determine whether the node at the top of said second connection address list is part of the distribution network;

(vi) if the node at the top of said second connection address list is not part of the distribution network, deleting such node from said second connection address list and repeating steps (d)(iv) and (d)(v) with respect to the next node at the top of said second connection address list;

(vii) if the node at the top of said second connection address list is part of said distribution network, determining whether said node at the top of said second connection address list has available capacity to transmit content data to said green node; and

(viii) if the node at the top of said second connection address list is part of said distribution network and has available capacity to transmit content data to said green node, having said green node dock with the node at the top of said second connection address list; and

(e) repeating steps (c) and (d) with respect to each user node which receives a retransmitted propagation signal.

In one example, step (c)(iii) may include the following additional step: (1) waiting a predetermined period of time so as to allow the red node's sibling with the highest

propagation rating prior to the red node's receipt of the propagation signal (hereinafter the "green node") to retransmit the propagation signal to the green node's child nodes (if any).

In another embodiment relating to a computer information distribution network comprising an instructing node having a fist user node docked with said instructing node to receive content data therefrom and a second user node docked with said first user node to receive content data therefrom, a process for handling a purported communication interruption between said first user node and said second user node is provided, said process including the following steps:

(a) having said instructing node receive a first complaint about said first user node from said second user node;

(b) if the second user node is the only user node docked with said first user node at a point in time, if said instructing node receives another complaint about the first user node from another user node within a predetermined period of time from when said instructing node received said first complaint or if said instructing node experiences a communication problem between it and said first user node during said predetermined period of time: (i) disconnecting said first user node from said instructing node; and

(c) if said instructing node has not experienced a communication problem between it and said first user node during said predetermined period of time and if a plurality of user nodes are docked with said first user node at said point in time and said instructing node has not received another complaint about said first user node from a user node other than said second user node within said predetermined period of time: (i) disconnecting said second user node from said first user node.

In one example, step (b) may include the following step: (ii) sending a signal to the user nodes docked with said first user node indicating the departure of said first user node from said network.

In another example, each user node may have no more than a predetermined maximum number of child nodes docked with it, and wherein step (b) may include the following steps

(ii) providing information at said point in time to each of said second user node, the other user nodes which had been docked with said first user node (if any) and at least one of any user nodes cascadingly connected in the distribution network under said first user node prior to said first user node's being disconnected from said instructing node (if any) as follows:

(1) the node's propagation rating, wherein a propagation rating is one of a predetermined number of grades ranging from highest to second highest to third highest and so on, with the number of grades being equal to said predetermined maximum number;

(2) an ancestor list setting forth the node's ancestor nodes' addresses (if it has any ancestor nodes) back to the instructing node, with the node's parent node being atop the ancestor list;

(3) a sibling list of the node's sibling nodes' addresses (if it has any sibling nodes) and their respective propagation ratings; and

(4) a child list of the node's child nodes' addresses (if it has any child nodes) and their respective propagation ratings;

(iii) having said instructing node send a propagation signal to each of said second user node and the other user nodes which had been docked with said first user node (if any);

(iv) with respect to each user node receiving the propagation signal which did not have the highest propagation rating prior to its receiving the propagation signal (hereinafter referred to as a "red node"), upon its receiving the propagation signal:

(1) setting the propagation rating of the red node to the next higher grade above the propagation rating of that red node before it received the propagation signal;

(2) having the red node undock from its parent node (if it was docked with its parent node before it received the propagation signal); and

(3) having the red node dock with its sibling node which had the highest propagation rating prior to the red node's receiving the propagation signal;

(v) with respect to each user node receiving the propagation signal which had the highest propagation rating prior to its receiving a propagation signal (hereinafter referred to as a "green node"), upon its receiving the propagation signal:

(1) setting the propagation rating of that green node to the lowest grade;

(2) having the green node retransmit the propagation signal to its child nodes (if it has any); and

(3) determining whether the green node is docked with the node from which the green node received the propagation signal and if it is, having the green node remain docked with said node from which the green node received the propagation signal, and if it is not, having the green node dock with said node from which the green node received the propagation signal; and

(vi) repeating steps (iv) and (v) with respect to each user node which receives a retransmitted propagation signal.

In another embodiment a distribution network for the distribution of content data from a server node to user nodes, wherein said user nodes are connected to said server and each other in cascaded relationship is provided, wherein:

(a) at least one of said user nodes is a repeater node connected directly to said server node, wherein said repeater node retransmits content data received by it to a user node docked to it for purpose of receiving content data from said repeater node (hereinafter referred to as a "child node"); and

(b) wherein each repeater node has the ability to provide to a user node which is attempting to dock with said repeater node connection address instructions.

In one example, each repeater node may include a descendant database indicating:

(i) which child nodes, if any, at a point in time, are docked with it so as to receive content data from said repeater node, and

(ii) which user nodes, if any, at said point in time, are purportedly docked with each of said child nodes.

In another example, the connection address instructions may refer said user node which is attempting to dock with said repeater node to a node in said descendant database.

In another example, each user node may include an ancestor database indicating to which node said user node is docked so that said user node may receive content data therefrom (hereinafter referred to as a "parent node"), and to which node, if any, at said point in time, said parent node is docked so that it may receive content data therefrom.

In another example, if said parent node of said user node departs from said distribution network, said user node may contact another node on its ancestor database.

In another example, each child node of a repeater node may include a sibling database indicating which user nodes, if any, are also child nodes of said repeater node.

In another embodiment a distribution network for the distribution of content data from a server node to user nodes is provided, wherein n levels of user nodes are cascadingly connected to said server node, wherein n is a number greater than one, wherein each user node includes a delay which causes the playing of content data by such node to be delayed a period of time (hereinafter "delay time") from a point in time, and wherein delays in higher level nodes create greater delay times than do delays in lower level nodes.

In another embodiment a distribution network for the distribution of content data from a server node to user nodes is provided, wherein n levels of user nodes are cascadingly

connected to said server node, wherein n is a number greater than one, wherein each user node includes a delay which causes the playing of content data by such node to be delayed a period of time (hereinafter "delay time") from a point in time, wherein x is a number from and including 2 to and including n, and wherein delays in (x-1) level nodes create greater delay times than do delays in x level nodes.

In one example, said point in time is an event experienced approximately simultaneously by substantially all user nodes.

In another example, said point in time is when said node receives content data.

While a number of embodiments and examples of the present invention have been described, it is understood that these embodiments and examples are illustrative only, and not restrictive, and that many modifications may become apparent to those of ordinary skill in the art. For example, any desired branch factor may be used (*see*, e.g., FIG. 46A showing a branch factor of 1 (i.e., each parent node may have only one child node); FIG. 46B showing a branch factor of 2 (i.e., each parent node may have one or two children nodes); and FIG. 46C showing a branch factor of 3 (i.e., each parent node may have one, two or three children nodes)). Of course, other branch factors (e.g., 4, 5, 6, etc) may be used and, as discussed above, the distribution network may be a hybrid network utilizing subtrees having different branch factors). Further still, a modified Utility Rating may be determined as follows: Utility Rating = Time (of connection) x Bandwidth Rating. Further still, a node ID may be based upon: (a) an IP address (external/internal/DHCP); (b) a URL; and/or (c) a Port. Further still, the root server may propagate "its" time upon assignment of a connection and the user node may get a "delta" (i.e., difference) from its clock to use (in the event of a reset the user node may re-acquire the time from the server (or from another time-providing source)). Further still, the various time intervals discussed herein may be predetermined time intervals (e.g., .5 sec., 1 sec, 1 min.) and/or the time intervals may be based upon some varying criteria (e.g., flow of data in the network, number of nodes in the network, number of levels in the network, position of a given node in the network). Further still, partial propagation of data (e.g., ancestor data and/or descendent data) from one or more nodes may be permitted. Further still (as discussed above), one or more of the steps described herein may be omitted (and the various steps which are performed may not necessarily need to be carried out in the order described herein (which description was intended to represent a number of examples, and not be restrictive)). Further still, the root server and each of the user nodes may utilize essentially identical software for carrying out the network configuration, reconfiguration and data transfer operations described herein or the root server may have dedicated server software

while the user nodes have different, dedicated user node software. Further still, the entire distribution network (e.g., the root server and all of the user nodes) may have the same branch factor (e.g., a branch factor of 2) or the user nodes may have a first branch factor (e.g., a branch factor of 2) and the root server may have a second branch factor (e.g., the root server may be capable of acting as a parent to more user nodes than 2). Further still, heuristics may be employed to drive network configuration, reconfiguration and/or data flow (e.g., based upon historic network configuration, reconfiguration and/or data flow).

Aylo Exhibit 1018
Page 1769 of 2308

CLAIMS

1. A method for propagating a network topology model of a binary tree topology network through the network, the network including a plurality of nodes, including a root node and at least one leaf node, operatively networked together to provide a communications path for data, each of the plurality of nodes except the root node being docked downtree of a parent node, the parent node being one of the plurality of nodes, and each of the plurality of nodes adapted to be docked uptree of a first child node and a second child node, each of the at least one leaf nodes having none of the plurality of nodes docked downtree of and therewith, the method comprising the steps of:

each of the at least one leaf nodes providing a propagate network topology model message containing the address of the providing leaf node to the parent node docked uptree of and with the providing leaf node;

for each of the plurality of nodes receiving the propagate network topology model message from each child node docked downtree of and therewith:

forming a network topology model representing one or more paths from the receiving node to each one of the plurality of nodes downtree of the receiving node;

storing in a memory accessible to the receiving node, the network topology model so formed;

providing to the parent node docked uptree of and with the receiving node a propagate network topology model message containing the network topology model so formed,

thereby propagating a network topology model through the network.

2. The method of claim 1, wherein the communications path for data comprises a downtree communications path.

3. The method of claim 1, wherein the communications path for data comprises a downtree communications path for data and an uptree communications path for data.

4. The method of claim 1, wherein the step of each of the at least one leaf nodes providing a propagate network topology model message is carried out a plurality of times.

5. The method of claim 1, wherein the step of each of the at least one leaf nodes providing a propagate network topology model message is carried out periodically.

6. The method of claim 1, wherein a first node being one of the plurality of nodes, but not a leaf node, receives a register message from a second node that is not part of the network, and wherein the first node is docked uptree of a first child node and a second child node, each of the first child node and the second child node being one of the plurality of nodes, the method further comprising the steps of:

searching the network topology model stored in the memory of the first node for a path to an open node, the open node being one of the plurality of nodes that the network topology model shows as not being docked uptree of and with two child nodes;

sending a path message to the second node containing at least an address of the open node.

7. The method of claim 6, wherein the path message comprises a path from the open node to the first node.

8. The method of claim 6, wherein the searching step comprising finding the open node such that a connection to the open node keeps the binary topology network balanced.

9. The method of claim 6, wherein step of searching further comprises the steps of:

searching consecutively further down hierarchical levels of the binary tree represented in the network topology model stored in the memory of the first node:

first to locate a path representing a node not being docked uptree of and with any child nodes, and if no such path is located,

then to locate a path representing a node not being docked uptree of and with two child nodes, and

repeating the searching step until one such path is located; and

designating the located path as a path to the open node.

10. A method for propagating a network topology model of a binary tree topology network through the network, the network including a plurality of nodes, including a root node and at least one leaf node, operatively networked together to provide a communications path for data, each of the plurality of nodes except the root node being docked downtree of a parent node, the parent node being one of the plurality of nodes, and each of the plurality of nodes adapted to be docked uptree of a first child node and a second child node, each of the at least one leaf nodes having none of the plurality of nodes docked downtree of and therewith, the method comprising the steps of:

providing a propagate path message from the root node to each of the plurality of nodes that are docked downtree of and with the root node, the propagate path message indicating an address of the root node;

for each of the plurality of nodes receiving the propagate path message from the root node or from another one of the plurality of nodes:

storing in a memory accessible to the receiving node, the path to the root node stored in the propagate path message;

if the receiving node has at least one of the plurality of nodes docked downtree of and therewith, adding to the propagate path message an address of the receiving node; and

providing the propagate path message to each of the plurality of nodes docked downtree of and with the receiving node, the propagate path message a path to the root node;

thereby propagating a root path of a binary tree topology network through the network.

11. The method of claim 10, wherein the communications path for data comprises a downtree communications path for data.

12. The method of claim 10, wherein the communications path for data comprises a downtree communications path for data and an uptree communications path for data.

13. The method of claim 10, wherein the step of providing a propagate path message is carried out a plurality of times.

14. The method of claim 10, wherein the step of providing a propagate path message is carried out periodically.

15. A method for propagating a network topology model of a binary tree topology network through the network, the network including a plurality of nodes, including a root node and at least one leaf node, operatively networked together to provide a communications path for data, each of the plurality of nodes except the root node being docked downtree of a parent node, the parent node being one of the plurality of nodes, and each of the plurality of nodes adapted to be docked uptree of a first child node and a second child node, each of the at least one leaf nodes having none of the plurality of nodes docked downtree of and therewith, the method comprising the steps of:

each of the at least one leaf nodes, providing a propagate network topology model message containing the address of the providing leaf node to the parent node docked uptree of and with the providing leaf node;

for each of the plurality of nodes receiving the propagate network topology model message from each child node docked downtree of and therewith:

forming a network topology model comprising a path from the receiving node to each one of the plurality of nodes downtree of the receiving node;

storing in a memory accessible to the receiving node, the network topology model so formed;

providing to the parent node docked uptree of and with the receiving node a propagate network topology model message containing the network topology model so formed,

providing a propagate path message from the root node to each of the plurality of nodes that are docked downtree of and with the root node, the propagate path message indicating an address of the root node;

for each of the plurality of nodes receiving propagate path message from the root node or from another one of the plurality of nodes:

storing in a memory accessible to the receiving node, the path to the root node stored in the propagate path message;

if the receiving node has at least one of the plurality of nodes docked downtree of and therewith, adding to the propagate path message an address of the receiving node; and

providing the propagate path message to each of the plurality of nodes docked downtree of and with the root node.

16. The method of claim 15, wherein the communications path for data comprises a downtree communications path for data.

17. The method of claim 15, wherein the communications path for data comprises a downtree communications path for data and an uptree communications path for data.

18. The method of claim 15, wherein the step of providing a propagate network topology model message is carried out a plurality of times.

19. The method of claim 15, wherein the step of providing a propagate network topology model message is carried out periodically.

20. The method of claim 15, wherein the step of providing a propagate path message is carried out a plurality of times.

21. The method of claim 15, wherein the step of providing a propagate path message is carried out periodically.

22. The method of claim 15, wherein a first node being one of the plurality of nodes, but not a leaf node, receives a register message from a second node that is not part of the network, and wherein the first node is docked uptree of a first child node and a second child node, each of the first child node and the second child node being one of the plurality of nodes, the method further comprising the steps of:

searching the network topology model stored in the memory of the first node for a path to an open node, the open node being one of the plurality of nodes that the network topology model shows as not being docked uptree of and with two child nodes;

sending a path message to the second node containing at least an address of the open node.

23. The method of claim 22, wherein the path message comprises a path from the open node to the first node.

24. The method of claim 22, wherein the path message comprises a path from the open node to the root node.

25. The method of claim 22, wherein the searching step comprising finding the open node such that a connection to the open node keeps the binary topology network balanced.

26. The method of claim 22, wherein step of searching further comprises the steps of:

searching consecutively further down hierarchical levels of the binary tree represented in the network topology model stored in the memory of the first node:

first to locate a path representing a node not being docked uptree of and with any child nodes, and if no such path is located,

then to locate a path representing a node not being docked uptree of and with two child nodes, and

repeating the searching step until one such path is located; and

designating the located path as a path to the open node.

27. A method for propagating a network topology model of a binary tree topology network through the network, the network including a plurality of nodes, including a root node and at least one leaf node, operatively networked together to provide a communications path for data, each of the plurality of nodes except the root node being docked downtree of a parent node, the parent node being one of the plurality of nodes, and each of the plurality of nodes adapted to be docked uptree of a first child node and a second child node, each of the at least one leaf nodes having none of the plurality of nodes docked downtree of and therewith, the method comprising the steps of:

providing a banned list message from the root node to each of the plurality of nodes that are docked downtree of and with the root node, the banned list message comprising an identifier for each of a plurality of banned nodes;

for each of the plurality of nodes receiving the banned list message from the root node or from another one of the plurality of nodes:

storing in a memory accessible to the receiving node, the identifiers for each of the plurality of banned nodes; and

providing the banned list message to each of the plurality of nodes docked downtree of and with the receiving node;

thereby propagating the list of banned nodes through the network.

28. The method of claim 27, wherein the communications path for data comprises a downtree communications path for data.

29. The method of claim 27, wherein the communications path for data comprises a downtree communications path for data and an uptree communications path for data.

30. The method of claim27, further comprising the step of rejecting a register message from a new node that is not part of the network if the node is identified on the banned list.

31. A method for propagating a network topology model of a binary tree topology network through the network, the network including a plurality of nodes, including a root node and at least one leaf node, operatively networked together to provide a communications path for data, each of the plurality of nodes except the root node being docked downtree of a parent node, the parent node being one of the plurality of nodes, and each of the plurality of nodes adapted to be docked uptree of a first child node and a second child node, each of the at least one leaf nodes having none of the plurality of nodes docked downtree of and therewith, the method comprising the steps of:

the first node sending a request firewall status message over a first channel to one of the plurality of nodes that is docked with the first node;

the one of the plurality of nodes receiving the request firewall status message sending at least one response message to the first node over a separate channel; and

the first node determining its firewall status based upon whether it receives the response message.

32. The method of claim 31, wherein the communications path for data comprises a downtree communications path for data.

33. The method of claim 31, wherein the communications path for data comprises a downtree communications path for data and an uptree communications path for data.

34. The method of claim 31, wherein the step of sending at least one response message comprises the steps of:

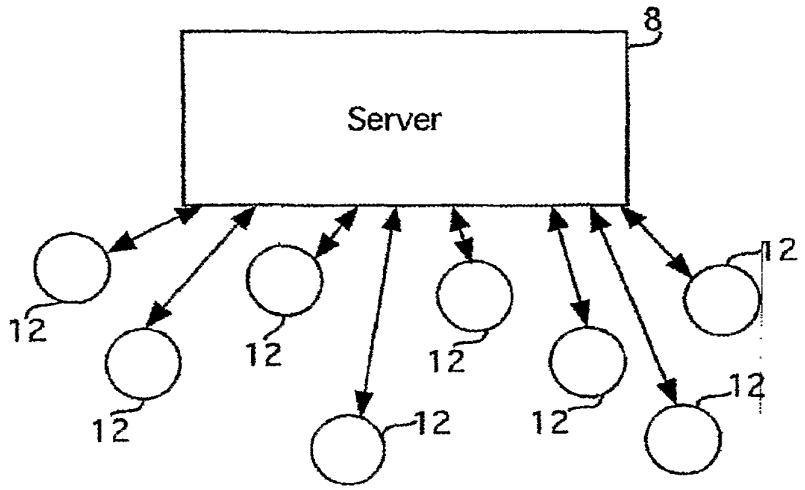
sending a first response message to the first node over a separate channel;

waiting a predefined period for an acknowledgement of the first response message, and if no acknowledgement is received, sending a second response message to the first node over the separate channel;

waiting a predefined period for an acknowledgement of the second response message, and if no acknowledgement is received, sending a third response message to the first node over the separate channel; and

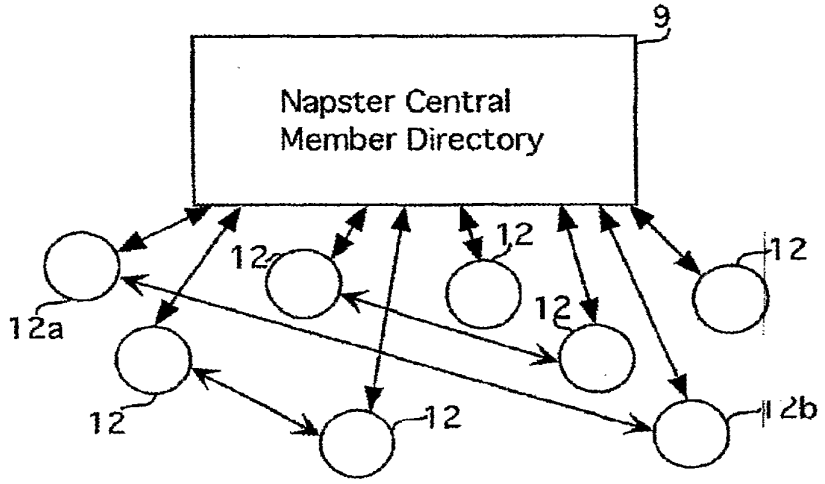
waiting a predefined period for an acknowledgement of the third response message, and if no acknowledgement is received, sending a firewall status on message to the first node over the first channel.

Aylo Exhibit 1018
Page 1778 of 2308



PRIOR ART

Fig. 1



PRIOR ART

Fig. 2

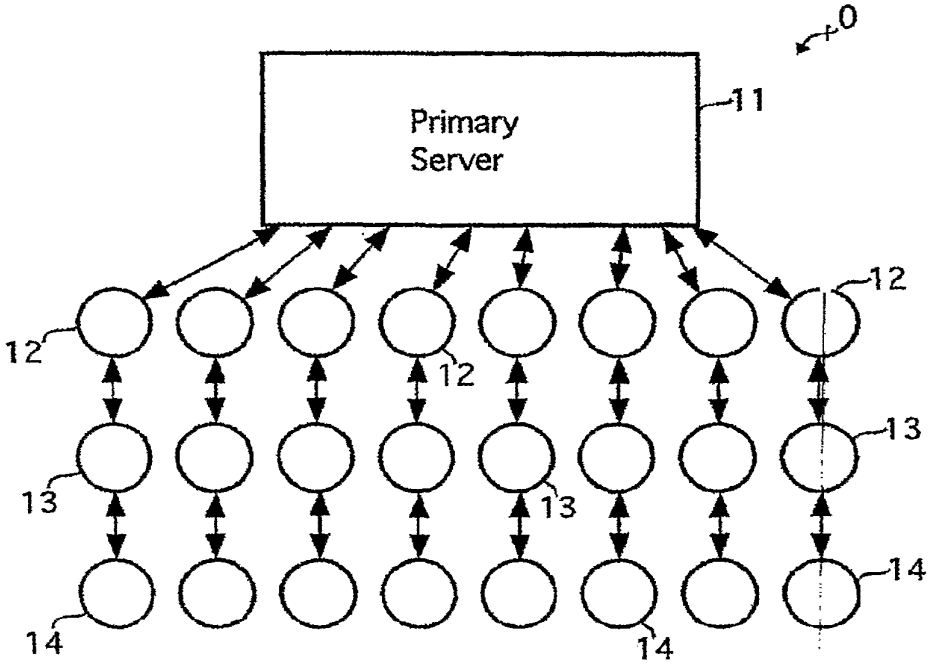


Fig. 3

"Start" [illegible]

[illegible]

Corresponds to [illegible] application [illegible]

31

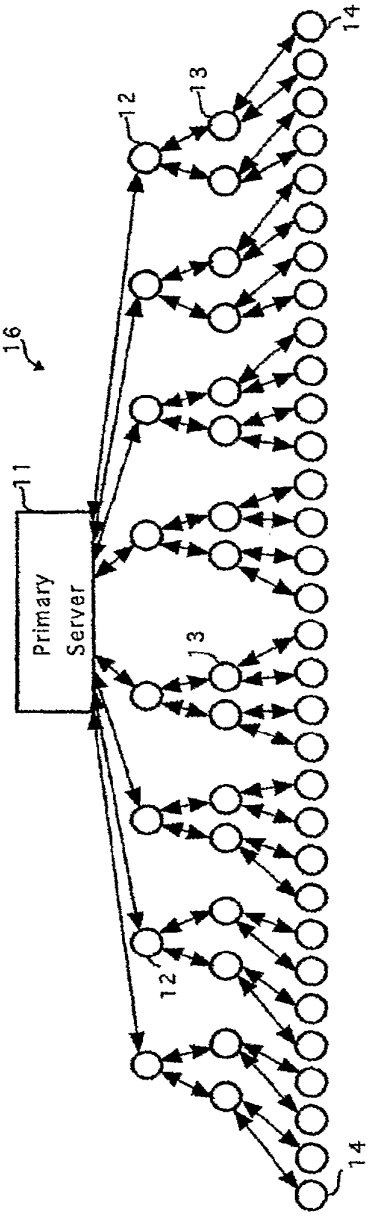


Fig. 4

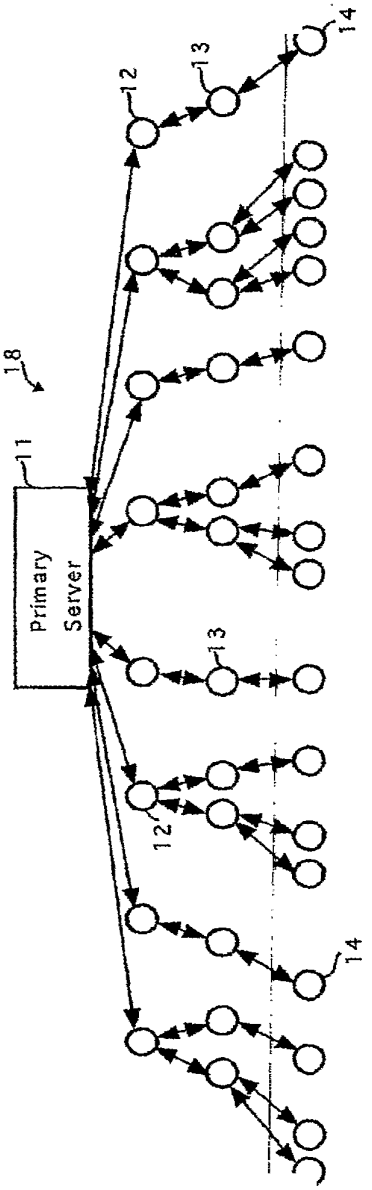


Fig. 5

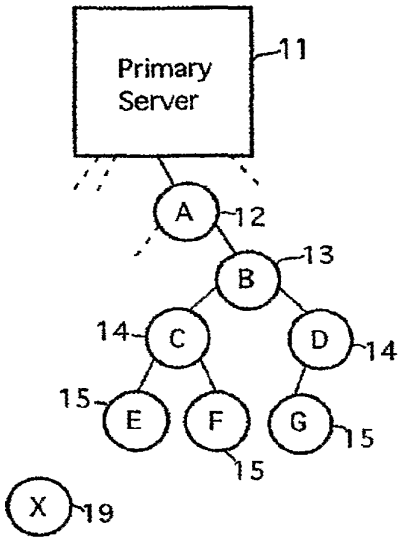


Fig. 6

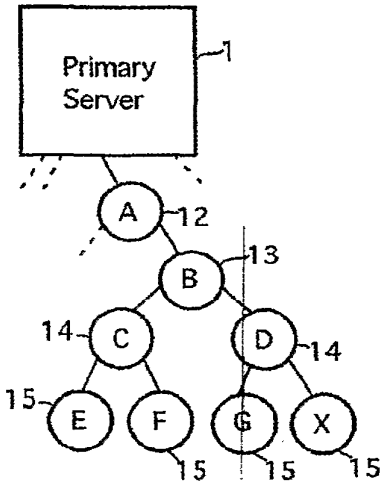


Fig. 7

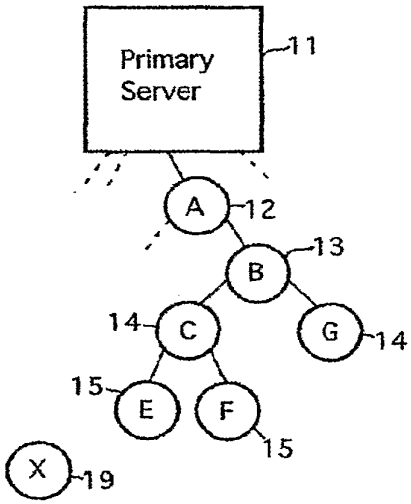


Fig. 8

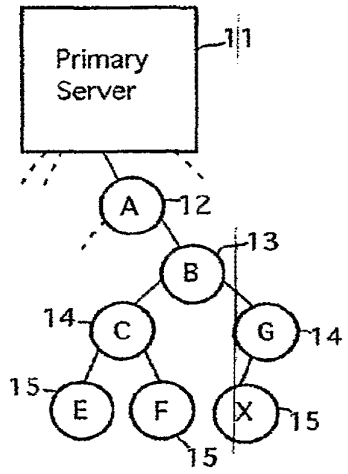


Fig. 9

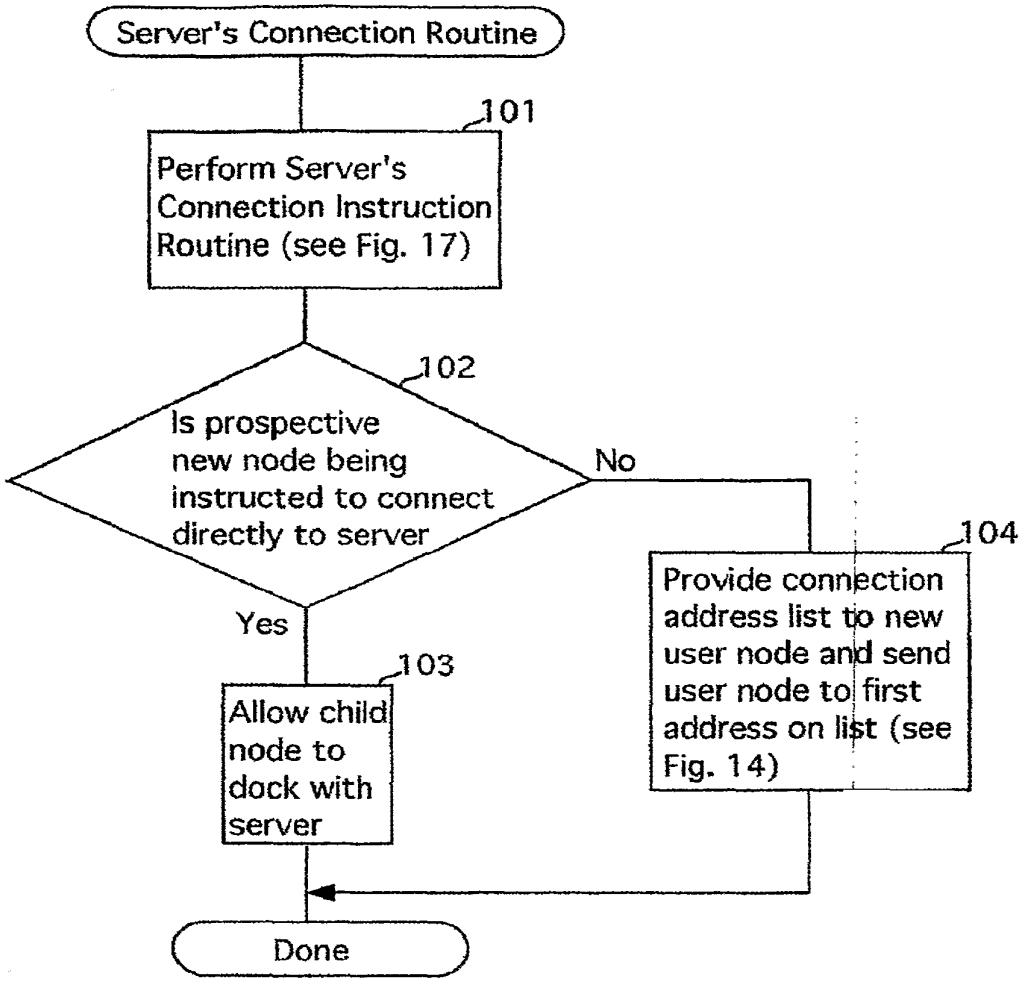

Server's Connection Routine
101
Perform Server's Connection Instruction Routine (see Fig. 17)
102
Is prospective new node being instructed to connect directly to server
No
Yes
103
Allow child node to dock with server
104
Provide connection address list to new user node and send user node to first address on list (see Fig. 14)
Done

Fig. 10

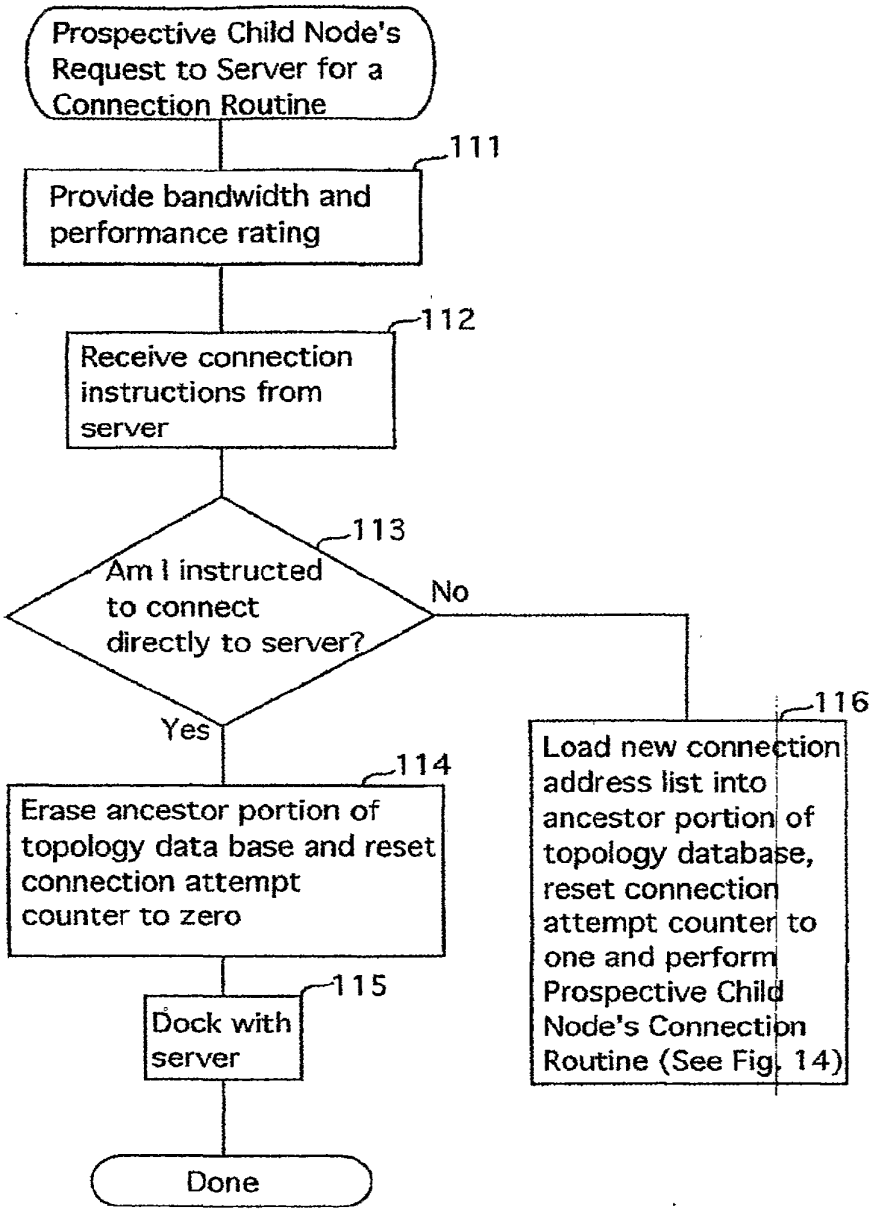

Prospective Child Node's Request to Server for a Connection Routine
111
Provide bandwidth and performance rating
112
Receive connection instructions from server
113
Am I instructed to connect directly to server?
No
Yes
114
Erase ancestor portion of topology data base and reset connection attempt counter to zero
115
Dock with server
Done
116
Load new connection address list into ancestor portion of topology database, reset connection attempt counter to one and perform Prospective Child Node's Connection Routine (See Fig. 14)

Fig. 11

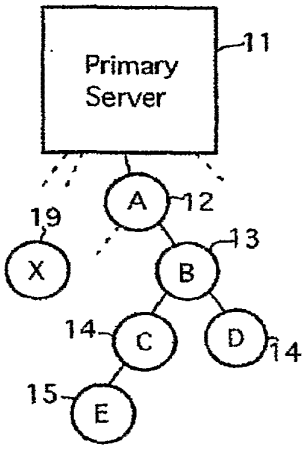


Fig. 12A

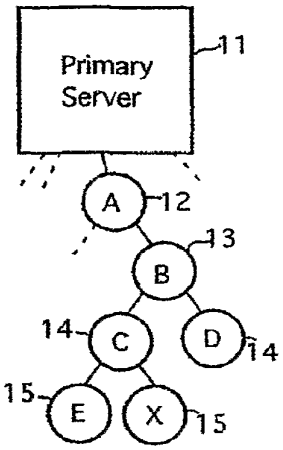


Fig. 12B

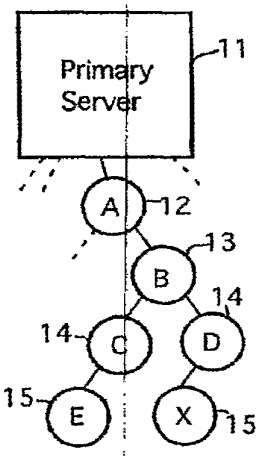


Fig. 12C

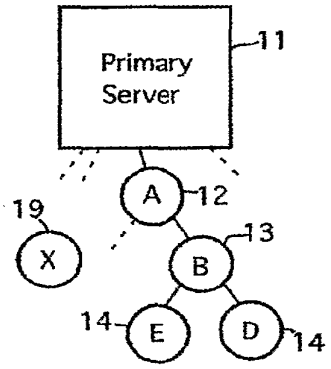


Fig. 12D

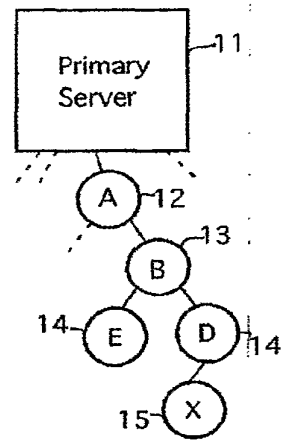


Fig. 12E

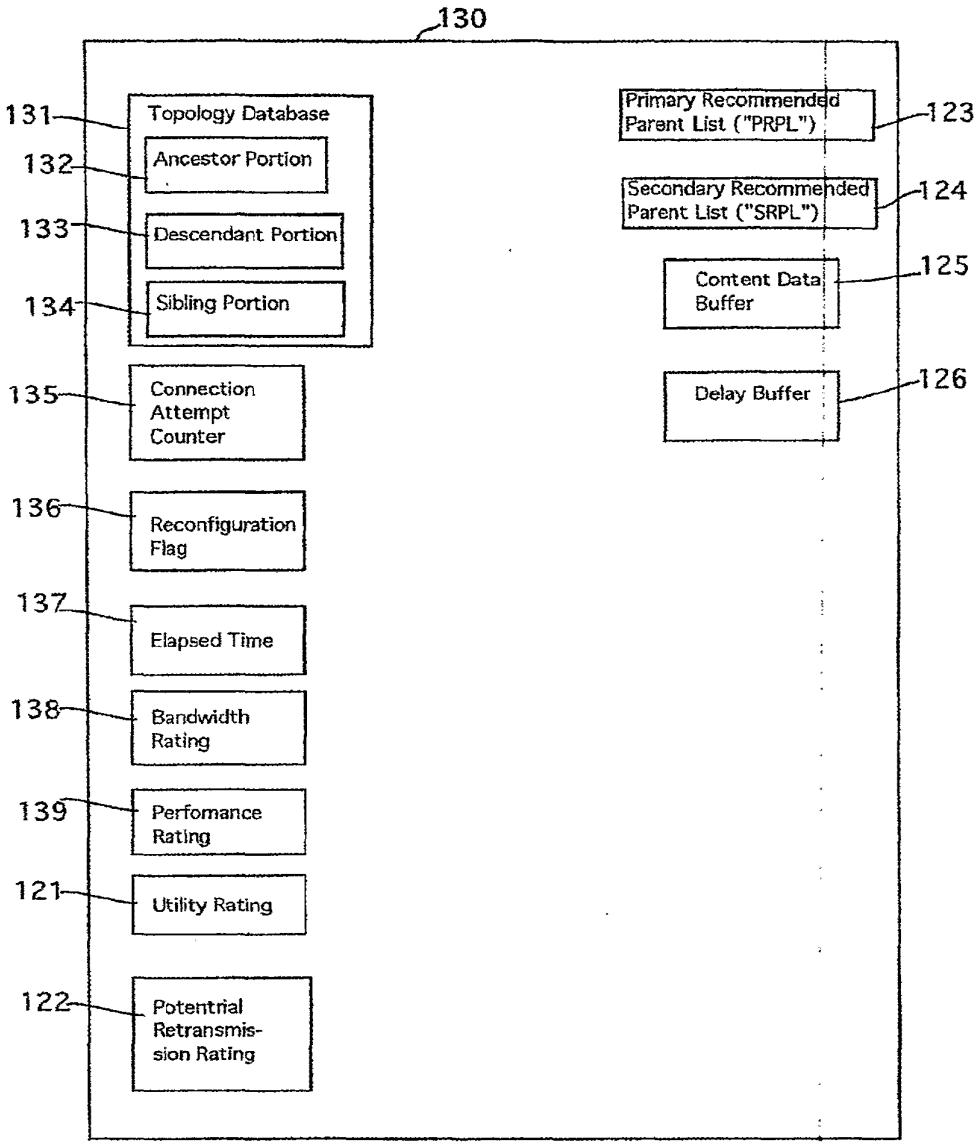

130
131
Topology Database
132
Ancestor Portion
133
Descendant Portion
134
Sibling Portion
135
Connection Attempt Counter
136
Reconfiguration Flag
137
Elapsed Time
138
Bandwidth Rating
139
Perfomance Rating
121
Utility Rating
122
Potentrial Retransmis-sion Rating
Primary Recommended Parent List ("PRPL")
123
Secondary Recommended Parent List ("SRPL")
124
Content Data Buffer
125
Delay Buffer
126

Fig. 13

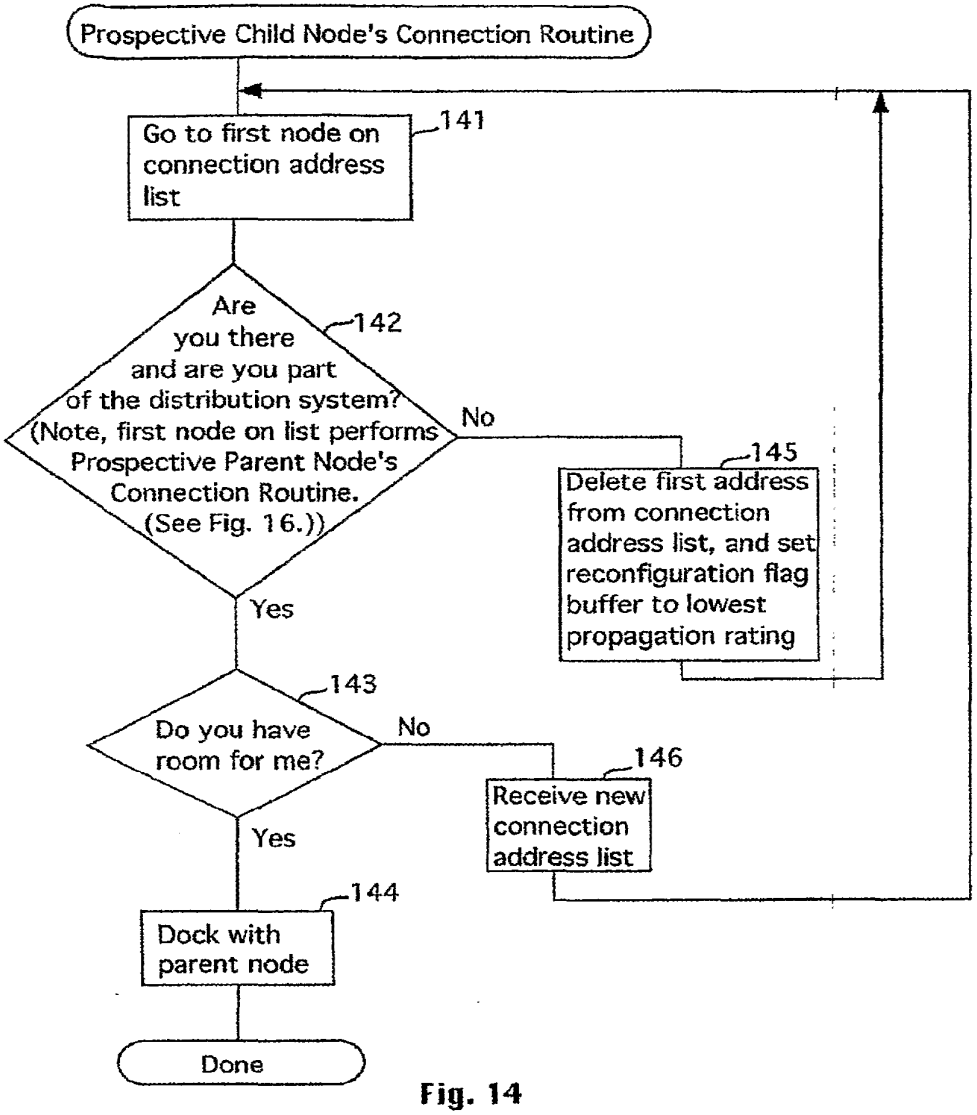

Prospective Child Node's Connection Routine
Go to first node on connection address list
141
Are you there and are you part of the distribution system? (Note, first node on list performs Prospective Parent Node's Connection Routine. (See Fig. 16.))
142
No
145
Delete first address from connection address list, and set reconfiguration flag buffer to lowest propagation rating
Yes
Do you have room for me?
143
No
146
Receive new connection address list
Yes
144
Dock with parent node
Done

Fig. 14

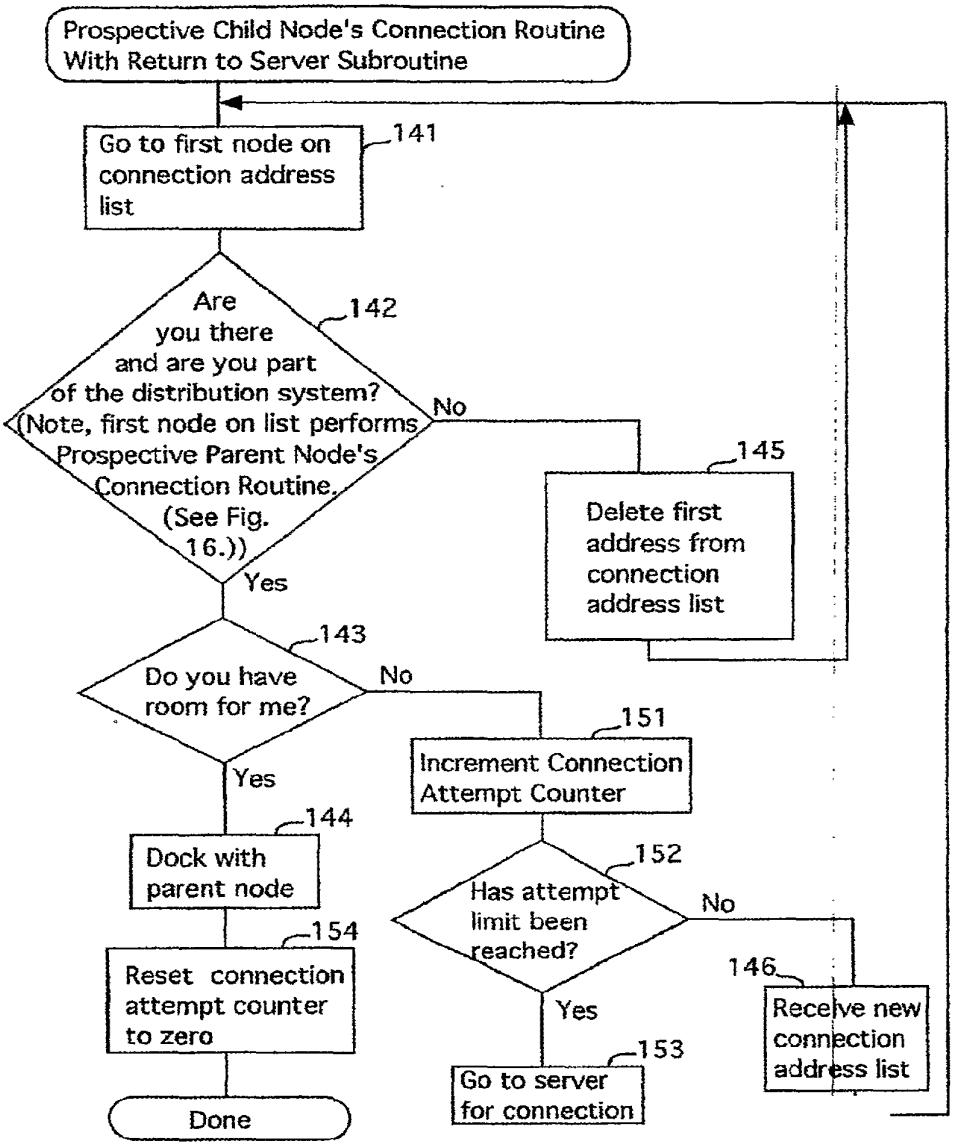

Prospective Child Node's Connection Routine With Return to Server Subroutine
Go to first node on connection address list
141
Are you there and are you part of the distribution system? (Note, first node on list performs Prospective Parent Node's Connection Routine. (See Fig. 16.))
142
No
Yes
Delete first address from connection address list
145
Do you have room for me?
143
No
Yes
Increment Connection Attempt Counter
151
Dock with parent node
144
Has attempt limit been reached?
152
No
Yes
Reset connection attempt counter to zero
154
Receive new connection address list
146
Go to server for connection
153
Done

Fig. 15

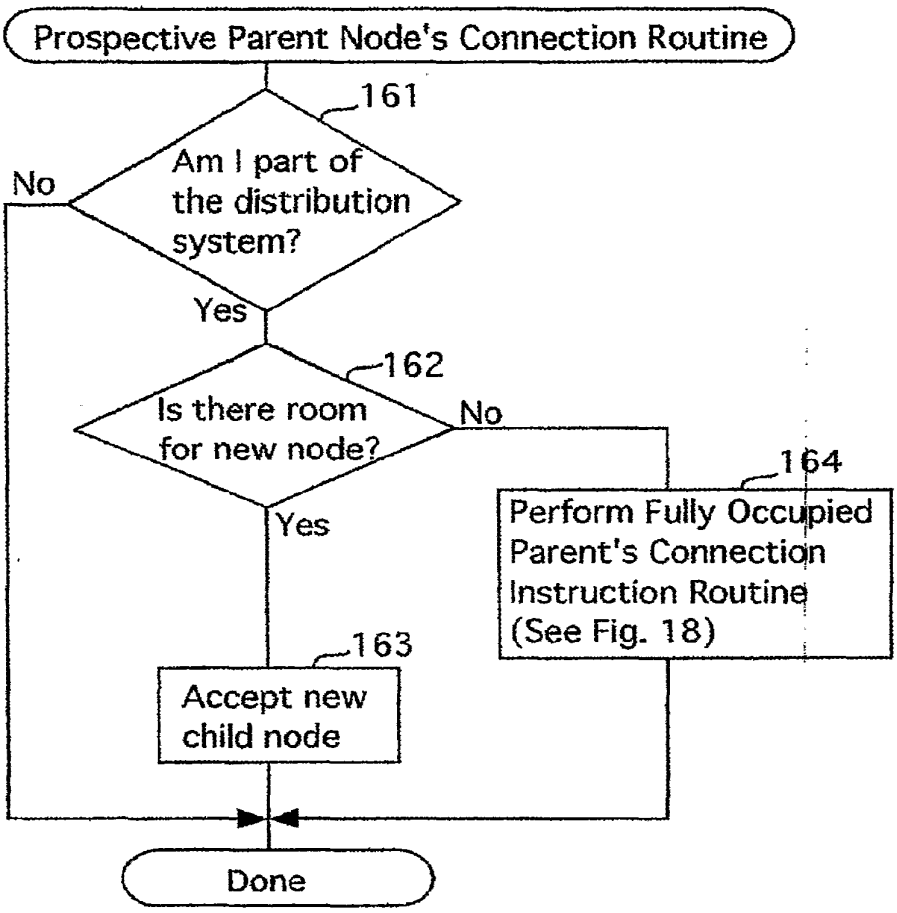

Prospective Parent Node's Connection Routine
161
Am I part of the distribution system?
No
Yes
162
Is there room for new node?
No
Yes
164
Perform Fully Occupied Parent's Connection Instruction Routine (See Fig. 18)
163
Accept new child node
Done

Fig. 16

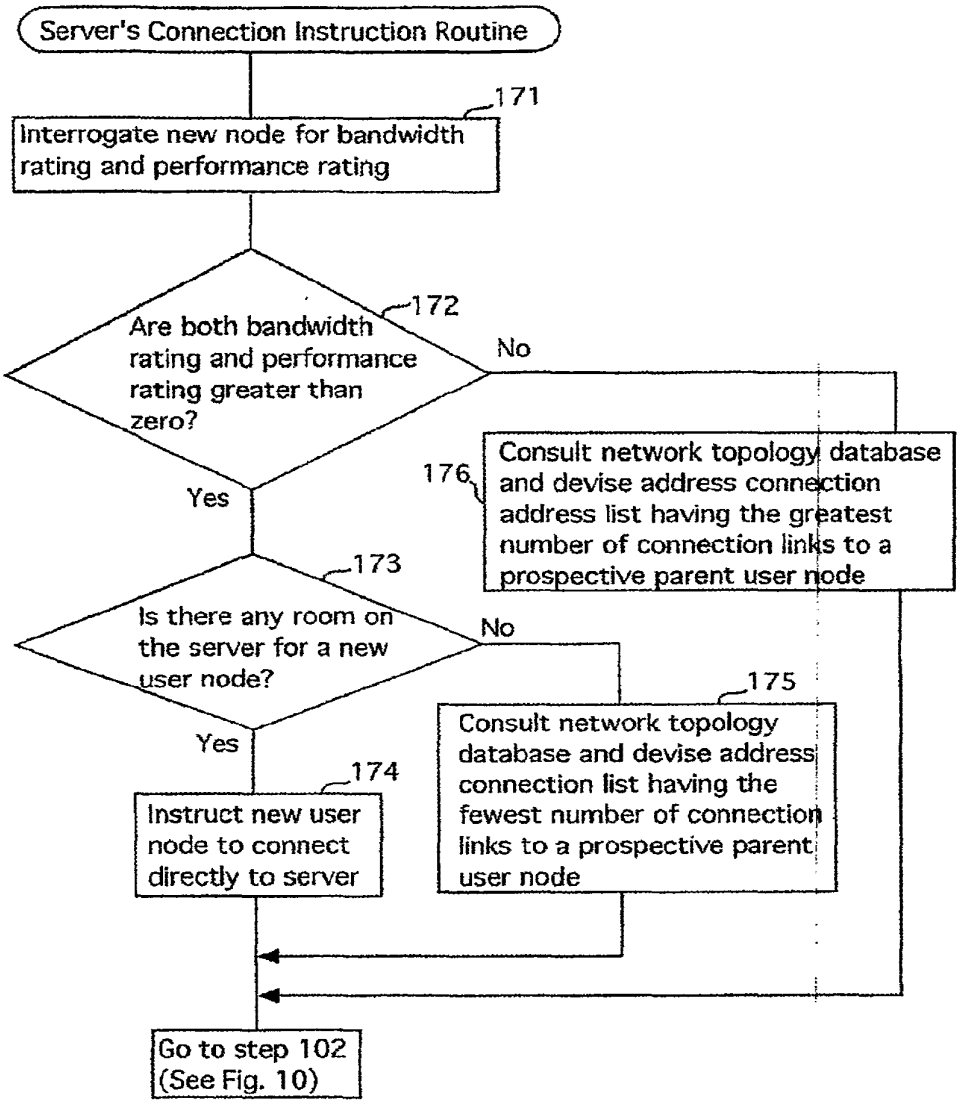

Server's Connection Instruction Routine
171
Interrogate new node for bandwidth rating and performance rating
172
Are both bandwidth rating and performance rating greater than zero?
No
Yes
176
Consult network topology database and devise address connection address list having the greatest number of connection links to a prospective parent user node
173
Is there any room on the server for a new user node?
No
Yes
175
Consult network topology database and devise address connection list having the fewest number of connection links to a prospective parent user node
174
Instruct new user node to connect directly to server
Go to step 102 (See Fig. 10)

Fig. 17

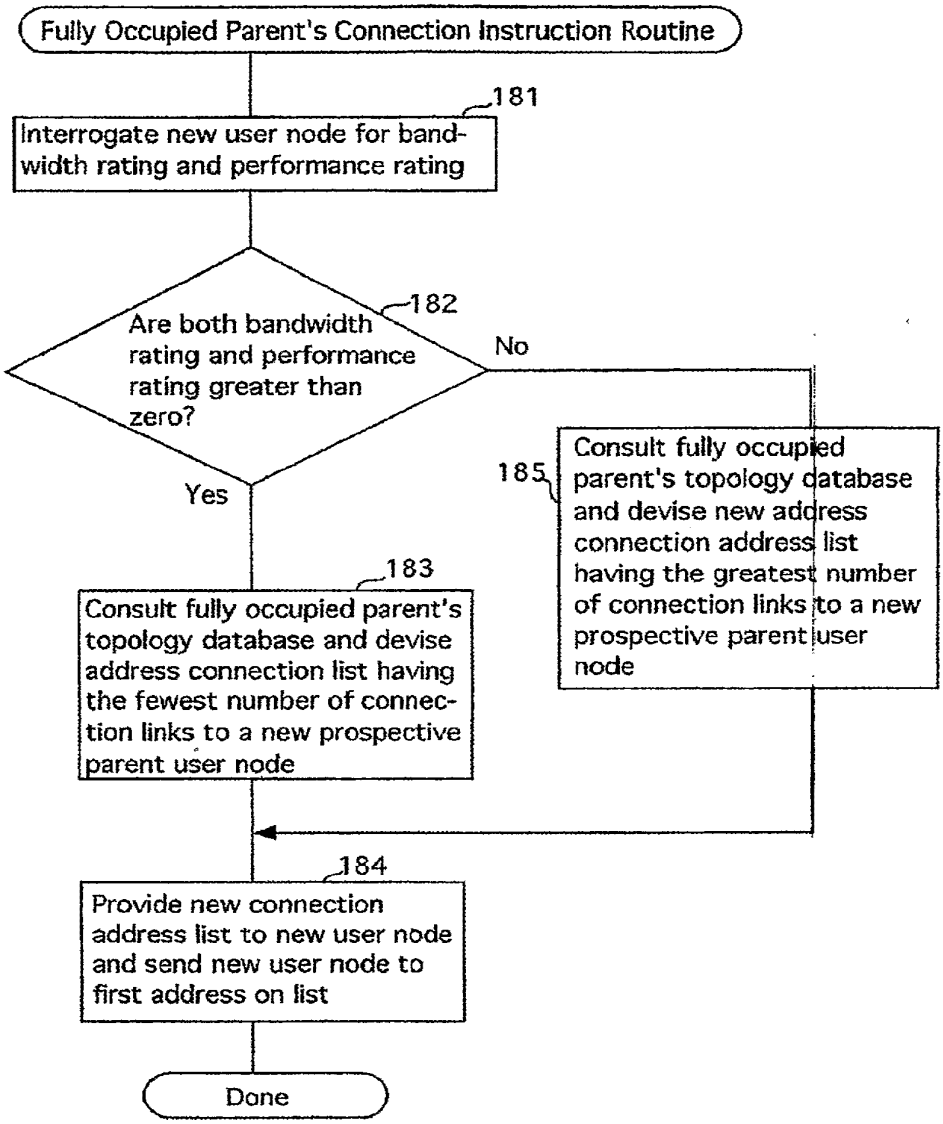

Fully Occupied Parent's Connection Instruction Routine
181
Interrogate new user node for bandwidth rating and performance rating
182
Are both bandwidth rating and performance rating greater than zero?
No
Yes
185
Consult fully occupied parent's topology database and devise new address connection address list having the greatest number of connection links to a new prospective parent user node
183
Consult fully occupied parent's topology database and devise address connection list having the fewest number of connection links to a new prospective parent user node
184
Provide new connection address list to new user node and send new user node to first address on list
Done

Fig. 18

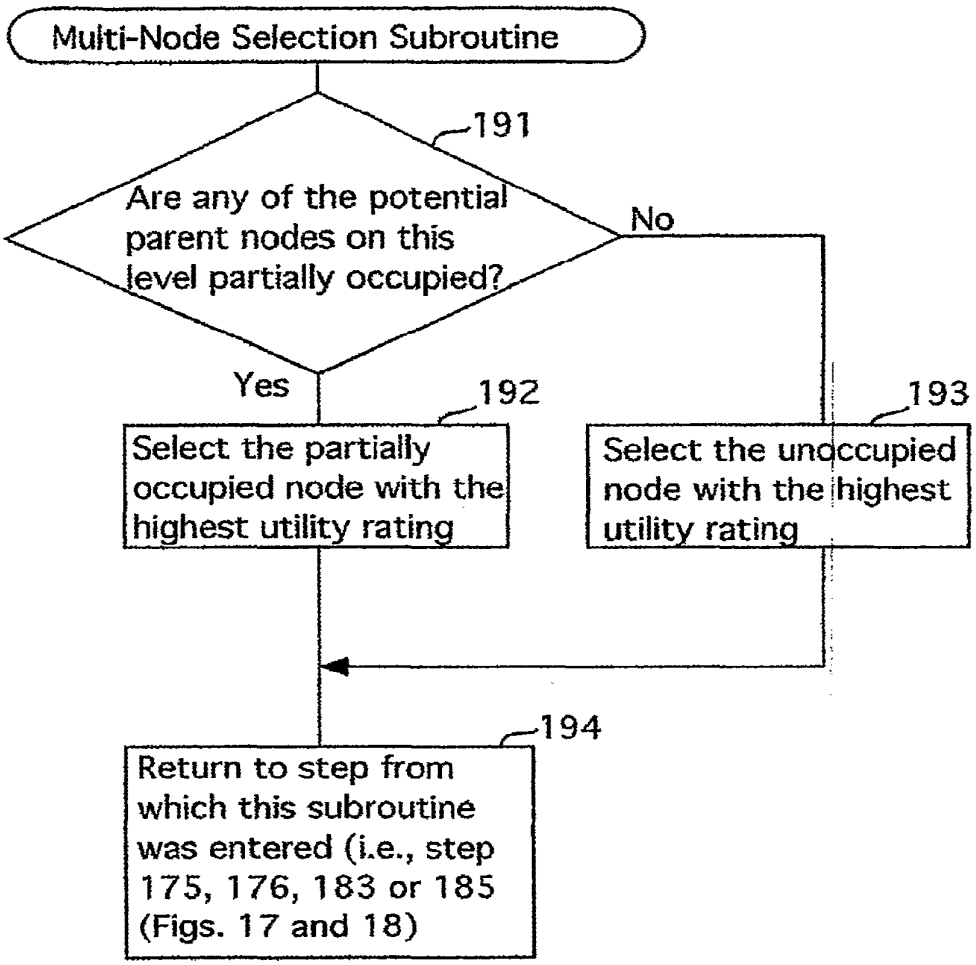

Multi-Node Selection Subroutine
191
Are any of the potential parent nodes on this level partially occupied?
No
Yes
192
Select the partially occupied node with the highest utility rating
193
Select the unoccupied node with the highest utility rating
194
Return to step from which this subroutine was entered (i.e., step 175, 176, 183 or 185 (Figs. 17 and 18)

Fig. 19

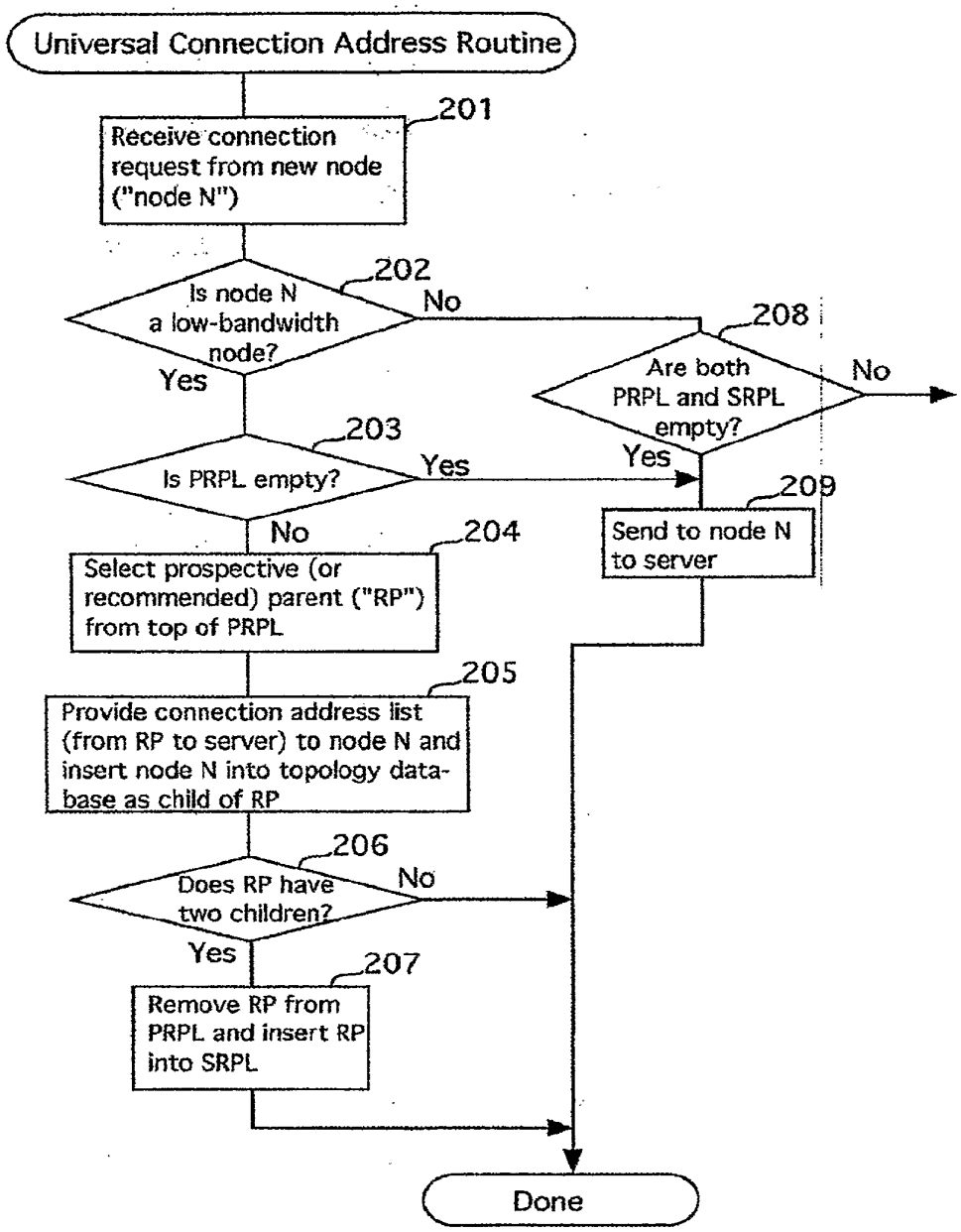

Universal Connection Address Routine
201
Receive connection request from new node ("node N")
202
Is node N a low-bandwidth node?
No
Yes
208
Are both PRPL and SRPL empty?
No
Yes
203
Is PRPL empty?
Yes
No
209
Send to node N to server
204
Select prospective (or recommended) parent ("RP") from top of PRPL
205
Provide connection address list (from RP to server) to node N and insert node N into topology data-base as child of RP
206
Does RP have two children?
No
Yes
207
Remove RP from PRPL and insert RP into SRPL
Done

Fig. 20A

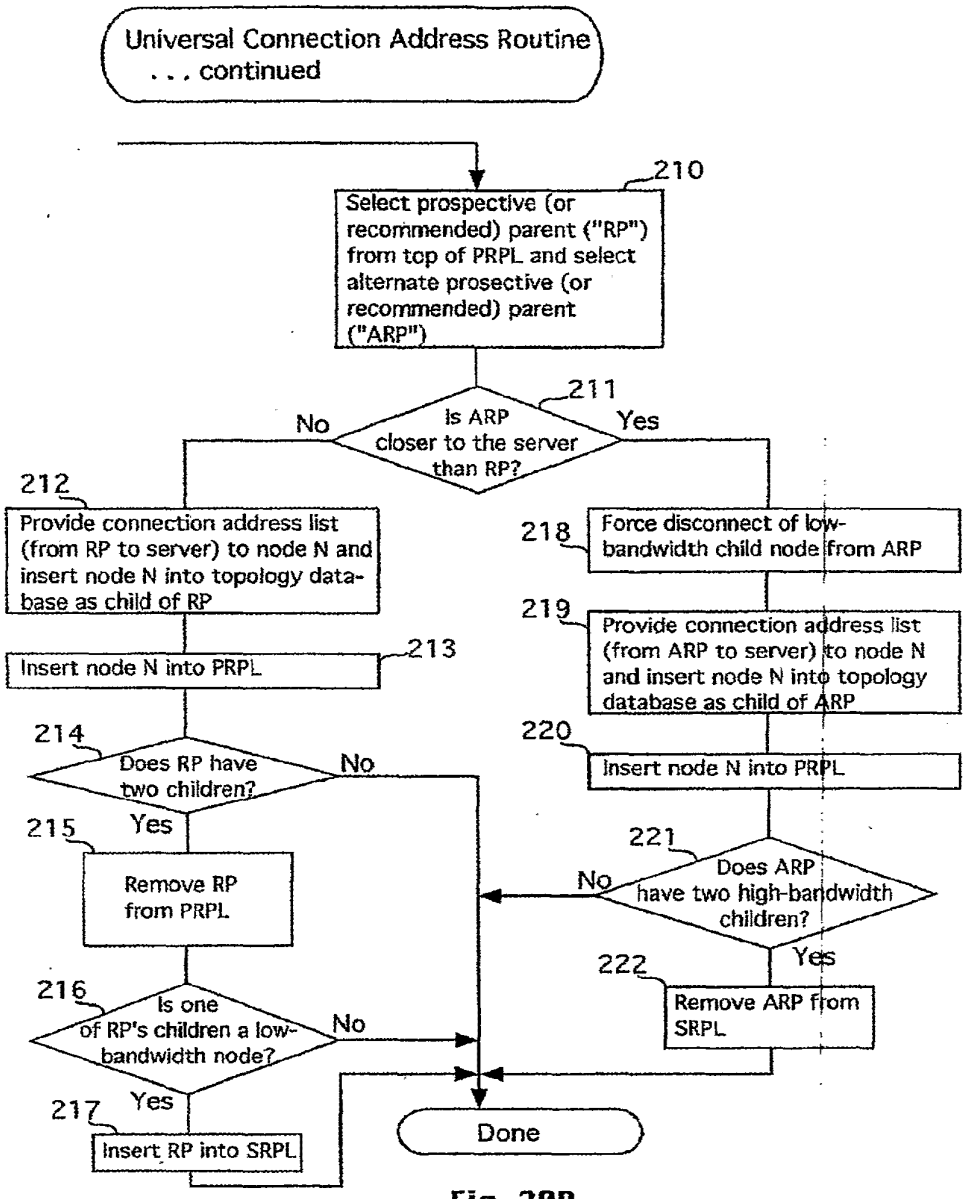

Universal Connection Address Routine
. . . continued
210
Select prospective (or recommended) parent ("RP") from top of PRPL and select alternate prosective (or recommended) parent ("ARP")
211
Is ARP closer to the server than RP?
No
Yes
212
Provide connection address list (from RP to server) to node N and insert node N into topology data-base as child of RP
213
Insert node N into PRPL
214
Does RP have two children?
No
Yes
215
Remove RP from PRPL
216
Is one of RP's children a low-bandwidth node?
No
Yes
217
Insert RP into SRPL
218
Force disconnect of low-bandwidth child node from ARP
219
Provide connection address list (from ARP to server) to node N and insert node N into topology database as child of ARP
220
Insert node N into PRPL
221
Does ARP have two high-bandwidth children?
No
Yes
222
Remove ARP from SRPL
Done

Fig. 20B

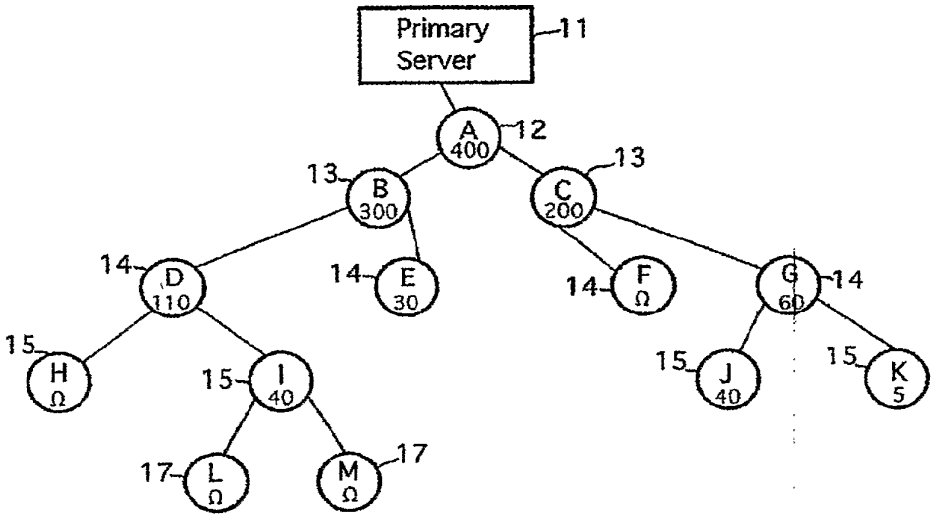

Primary Server
11
A
400
12
B
300
13
C
200
13
D
110
14
E
30
14
F
Ω
14
G
60
14
H
Ω
15
I
40
15
J
40
15
K
5
15
L
Ω
17
M
Ω
17

Fig. 21

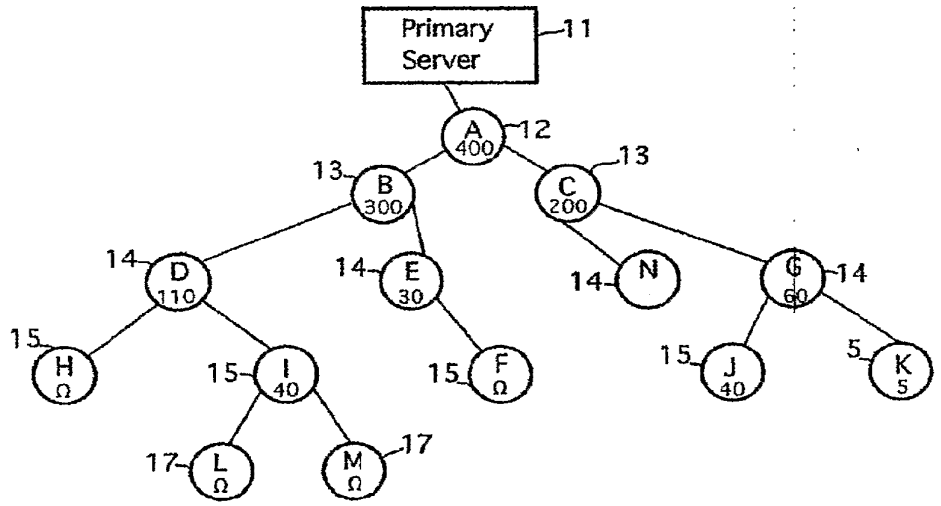

Primary Server
11
A
400
12
B
300
13
C
200
13
D
110
14
E
30
14
N
14
G
60
14
H
Ω
15
I
40
15
F
Ω
15
J
40
15
K
5
5
L
Ω
17
M
Ω
17

Fig. 22

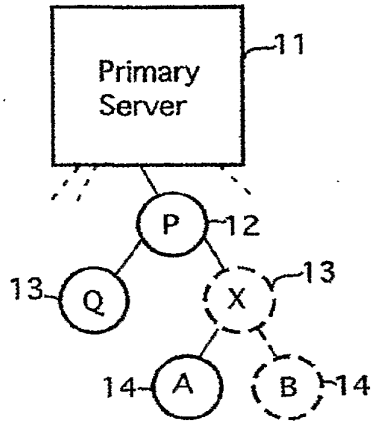


Fig. 23

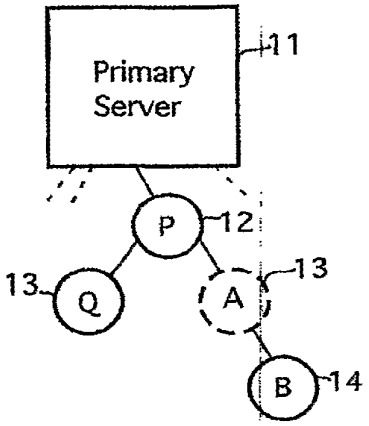


Fig. 24

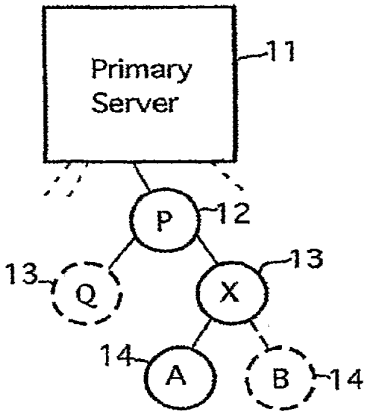


Fig. 25

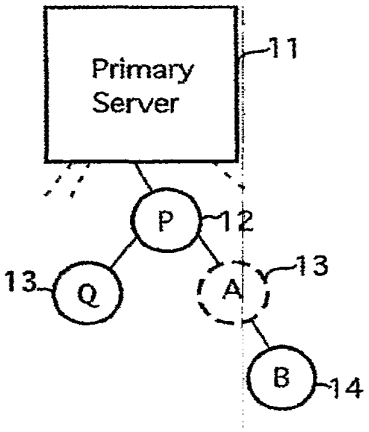


Fig. 26

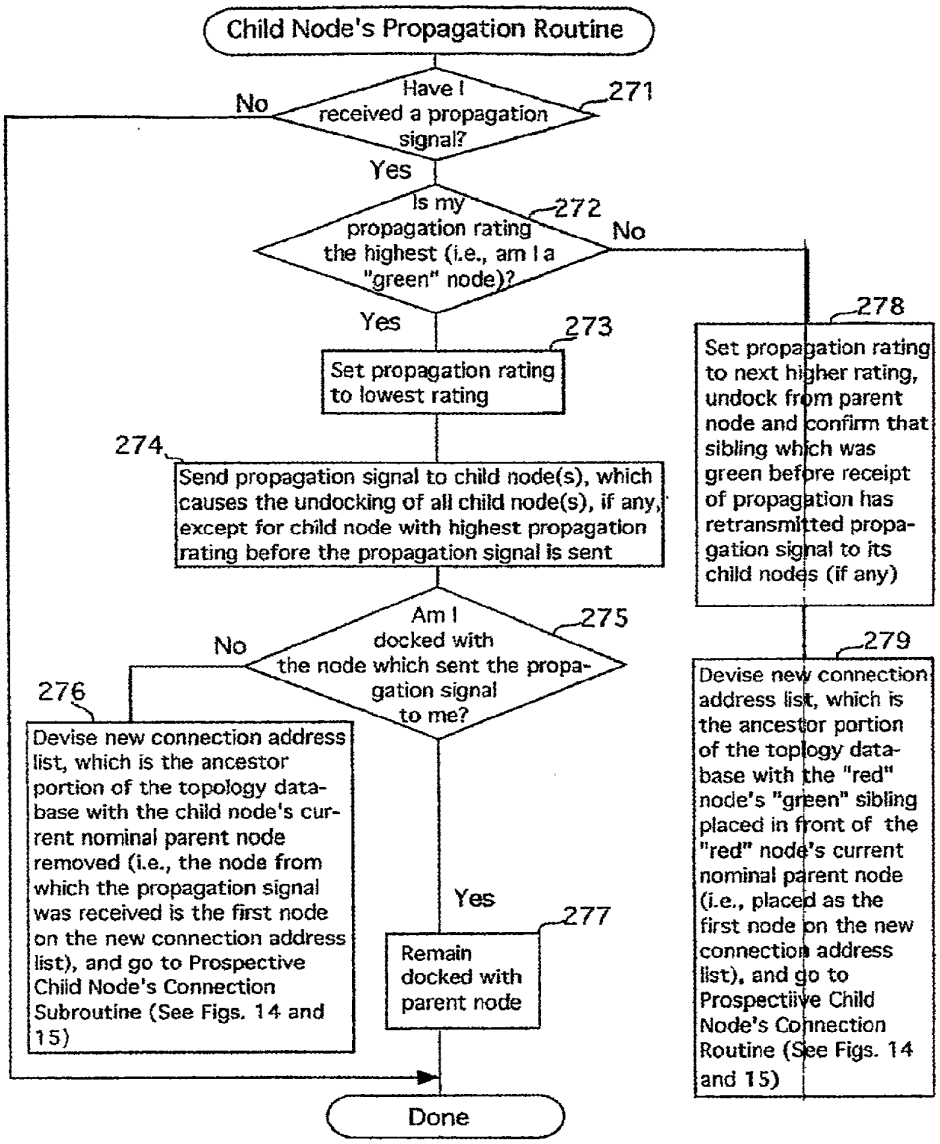


Fig. 27

Aylo Exhibit 1018
Page 1797 of 2308

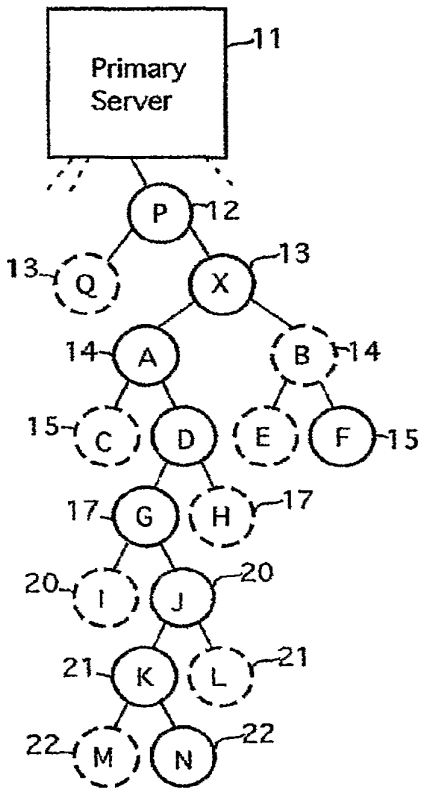

Primary Server
11
P
12
13
Q
X
13
14
A
B
14
15
C
D
E
F
15
17
G
H
17
20
I
J
20
21
K
L
21
22
M
N
22

Fig. 28

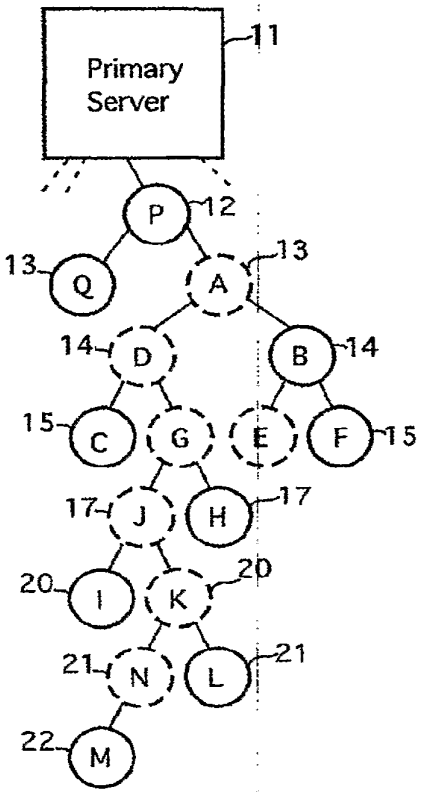

Primary Server
11
P
12
13
Q
A
13
14
D
B
14
15
C
G
E
F
15
17
J
H
17
20
I
K
20
21
N
L
21
22
M

Fig. 29

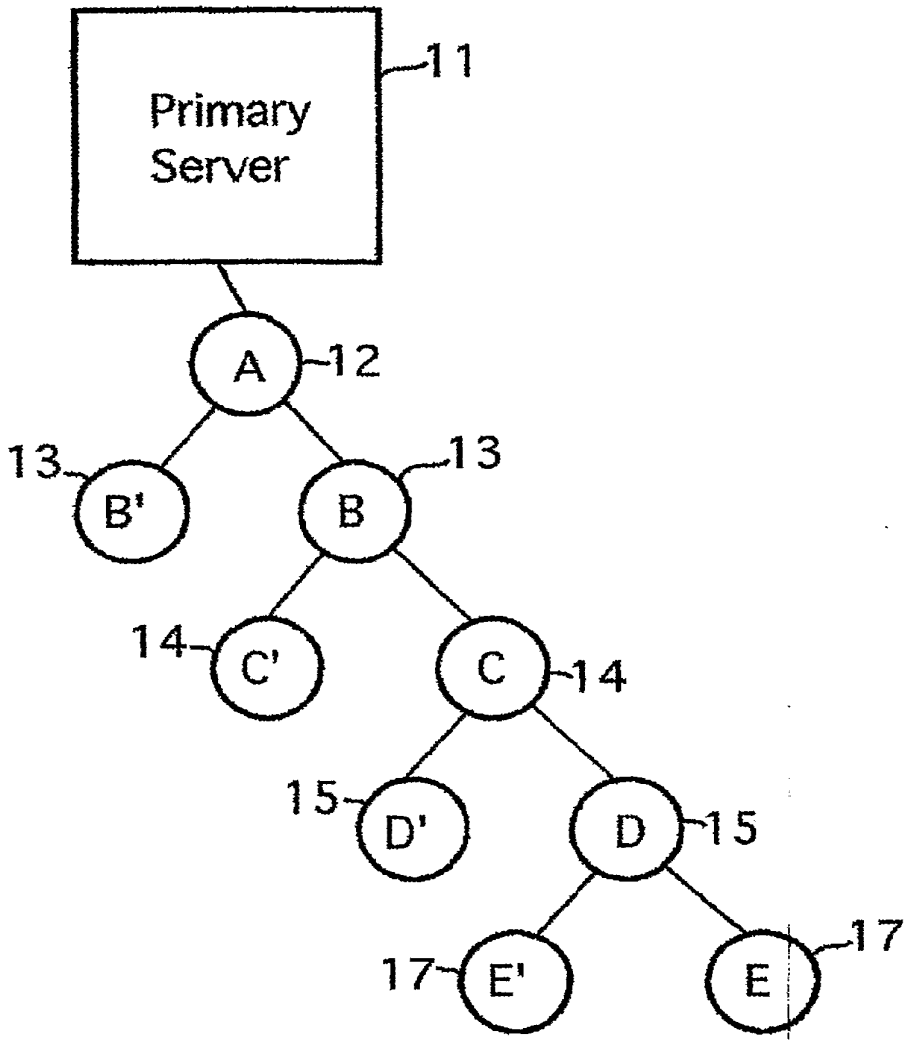


Fig. 30

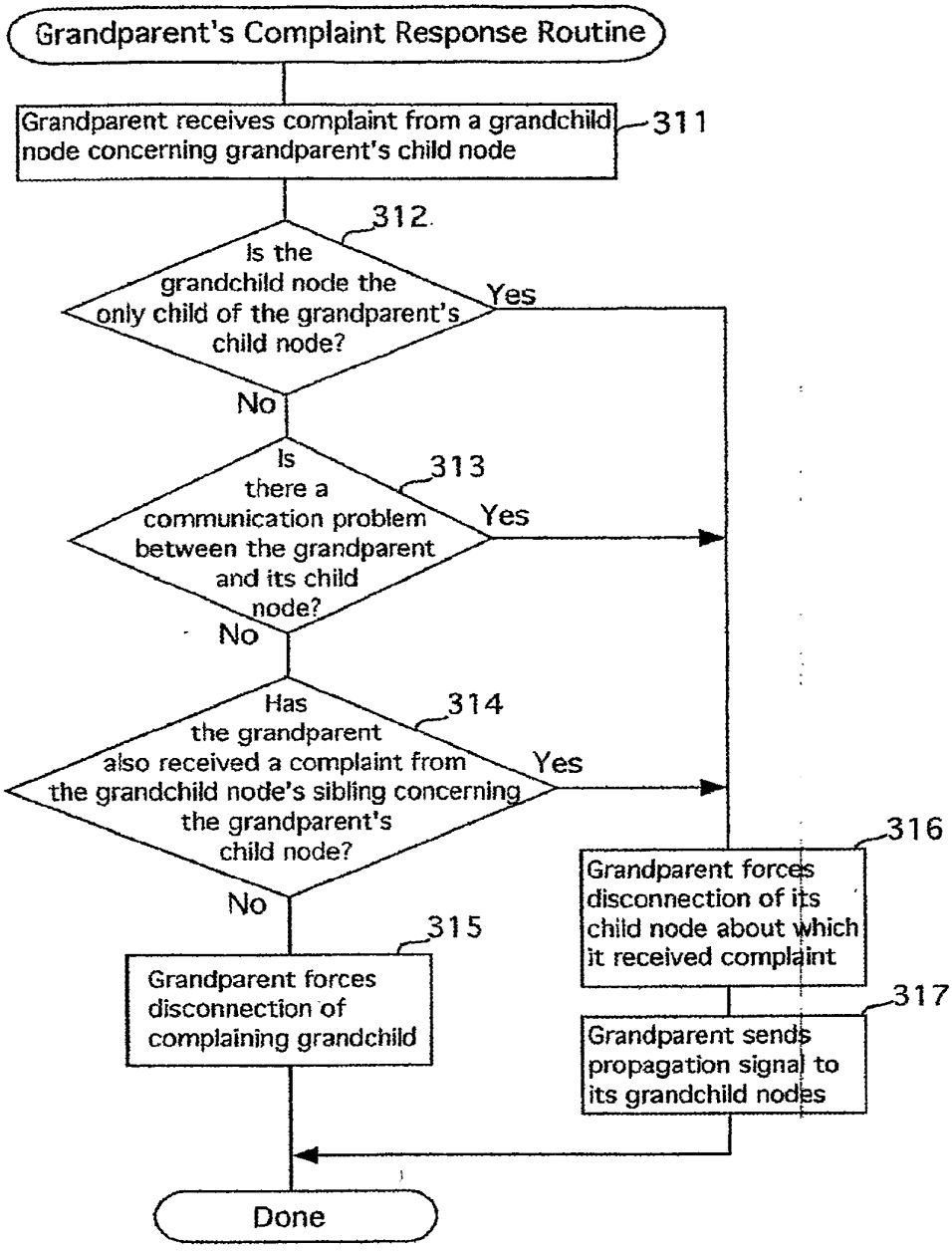

Grandparent's Complaint Response Routine
Grandparent receives complaint from a grandchild node concerning grandparent's child node
311
312
Is the grandchild node the only child of the grandparent's child node?
Yes
No
313
Is there a communication problem between the grandparent and its child node?
Yes
No
314
Has the grandparent also received a complaint from the grandchild node's sibling concerning the grandparent's child node?
Yes
No
315
Grandparent forces disconnection of complaining grandchild
316
Grandparent forces disconnection of its child node about which it received complaint
317
Grandparent sends propagation signal to its grandchild nodes
Done

Fig. 31

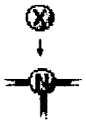


Fig. 32

Fig. 33A

Fig. 33B

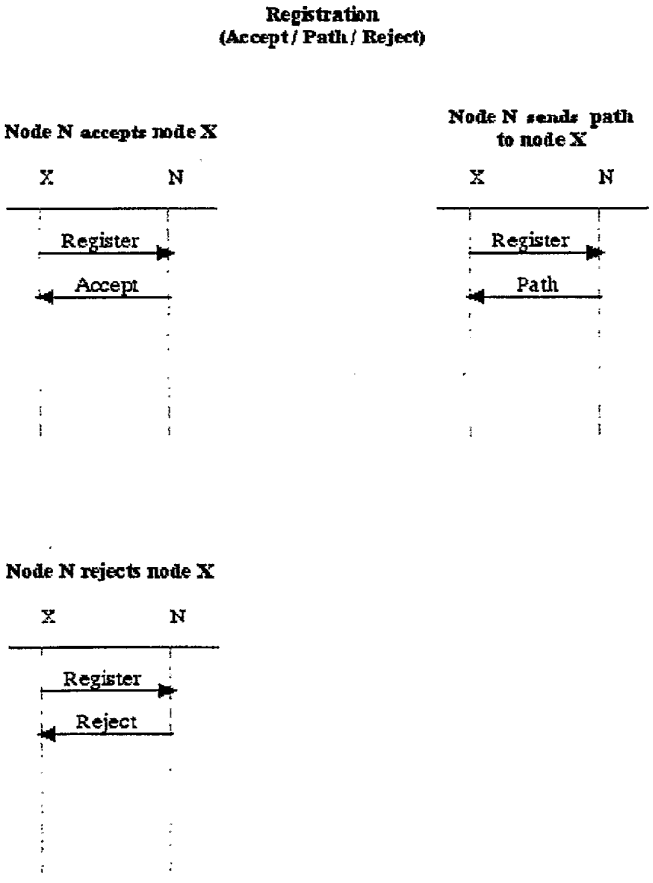


Fig. 33C

Fig. 34A

Fig. 34B

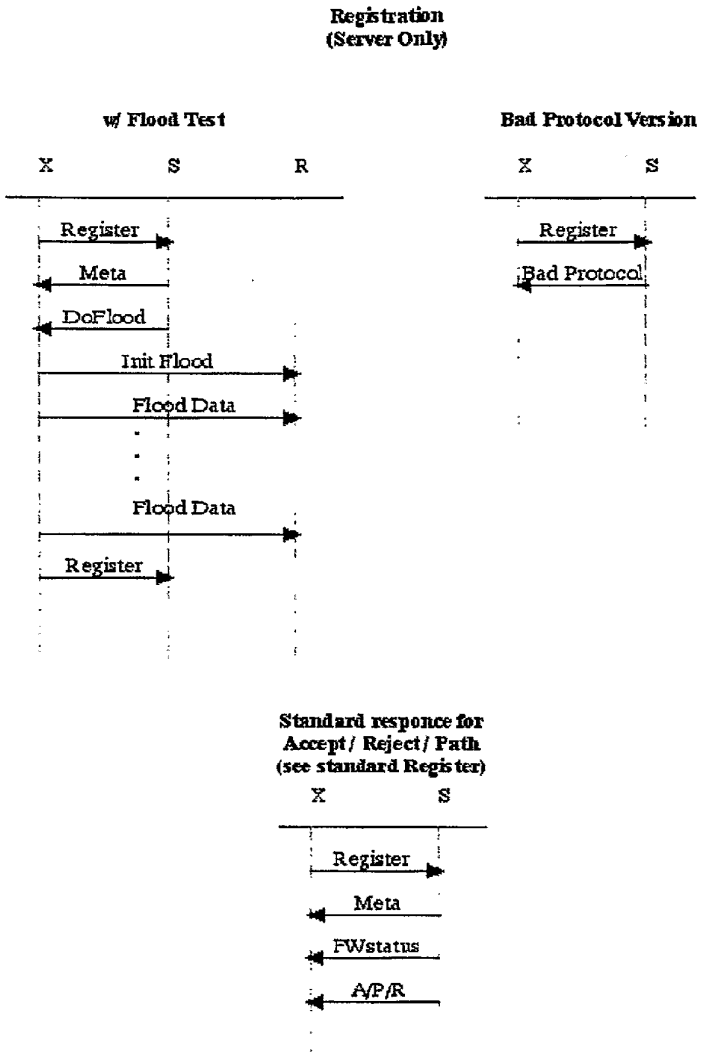


Fig. 34C

Fig. 35A

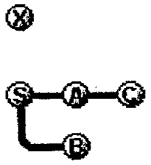
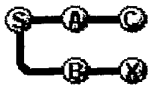


Fig. 35B



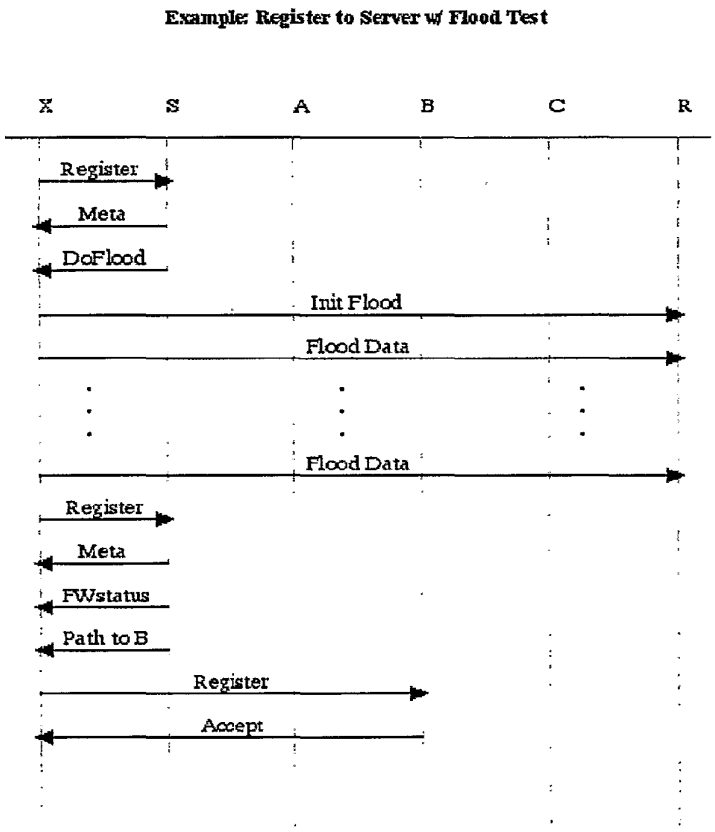


Fig. 36

Fig. 37



Fig. 38

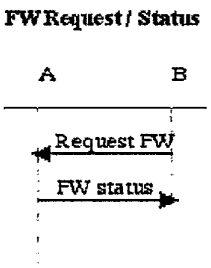


Fig. 39A



Fig. 39B

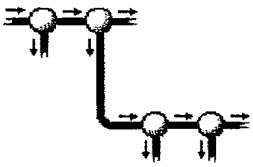


Fig. 40

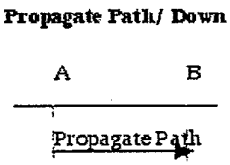


Fig. 41A



Fig. 41B

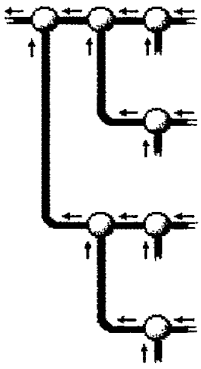


Fig. 42

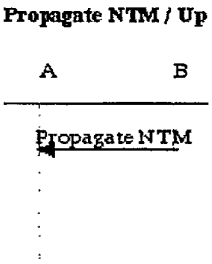


Fig. 43

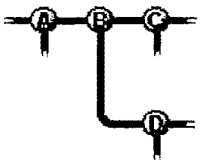


Fig. 44

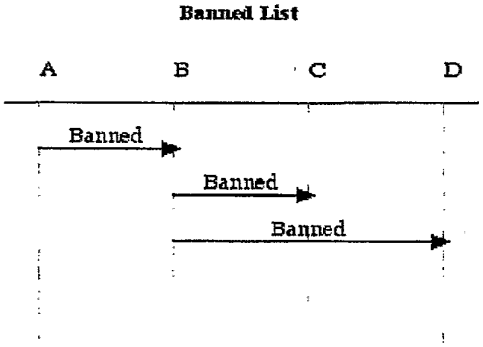


Fig. 43

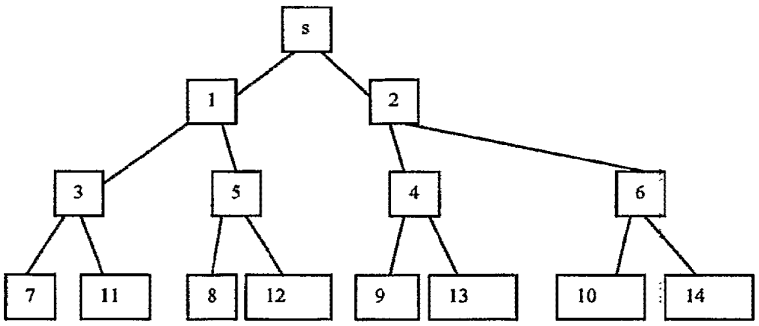


Fig. 46A

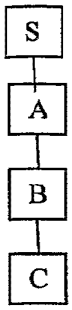


Fig. 46B

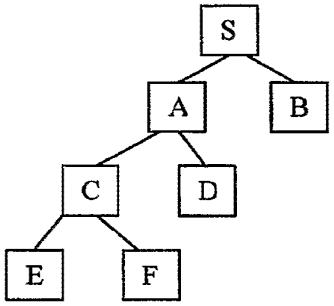
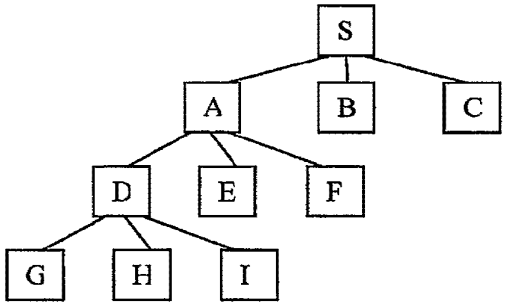


Fig. 46C



PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-092752
(43)Date of publication of application : 06.04.2001

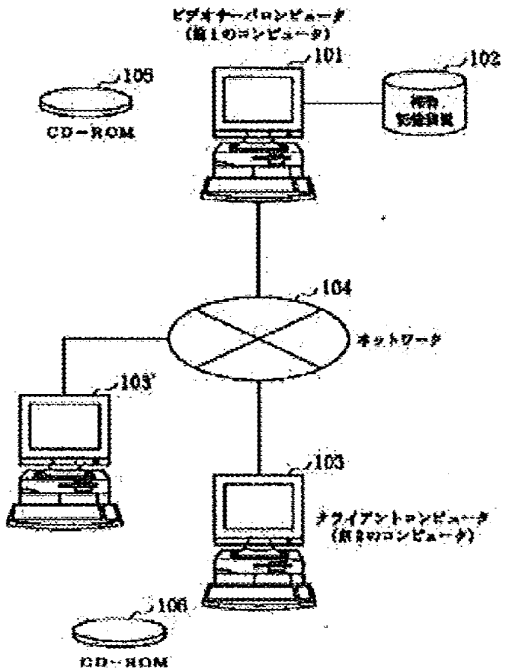
(51)Int.Cl. G06F 13/00
G06F 12/00
H04L 12/56
H04N 7/24

(21)Application number : 11-271422
(22)Date of filing : 24.09.1999

(71)Applicant : HITACHI INFORMATION SYSTEMS LTD
(72)Inventor : SHIOMI YOSHIHIRO
AIKAWA NOBUYUKI
NIGO MASAKI

(54) SYSTEM FOR DISTRIBUTING PICTURE DATA AND RECORDING MEDIUM USED FOR THE SAME

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a picture data distribution system capable of efficiently distributing picture data with appropriate quality in accordance with changes of the network connection environment of a client computer from moment to moment and a recording medium on which its program is recorded.
SOLUTION: This system has a 1st computer (image server computer 101) capable of distributing the picture data and one or more 2nd computers (client computer 103) receiving the picture data, and the 1st computer 101 receives a line situation notification message from the 2nd computer 103 and transmits the picture data (picture data extracted among a plurality of pieces of picture data which are preliminarily stored and have different compressibility or picture data compressed with optimum compressibility every time) which are subjected to compressibility being optimum to the communication speed of a line to which the 2nd computer is connected on the basis of the message.



Aylo Exhibit 1018
Page 1810 of 2308

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－92752
（P2001－92752A）
(43)公開日　平成13年４月６日(2001.4.6)

(51)Int.Cl.[7]	識別記号	ＦＩ		テーマコード（参考）
Ｇ０６Ｆ　13/00	３５４	Ｇ０６Ｆ　13/00	３５４Ｄ	５Ｂ０８２
12/00	５４５	12/00	５４５Ｍ	５Ｂ０８９
Ｈ０４Ｌ　12/56		Ｈ０４Ｌ　11/20	１０２Ａ	５Ｃ０５９
Ｈ０４Ｎ　7/24		Ｈ０４Ｎ　7/13	Ｚ	５Ｋ０３０

審査請求　未請求　請求項の数５　ＯＬ　（全 11 頁）

(21)出願番号　特願平11－271422

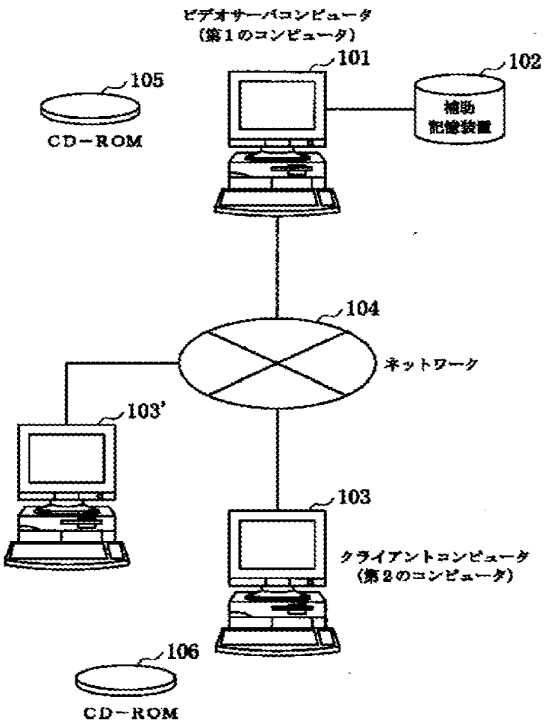
(22)出願日　平成11年９月24日(1999.9.24)

(71)出願人　000152985
株式会社日立情報システムズ
東京都渋谷区道玄坂１丁目16番５号
(72)発明者　塩見　芳弘
東京都渋谷区道玄坂一丁目16番５号　株式会社日立情報システムズ内
(72)発明者　相川　伸之
東京都渋谷区道玄坂一丁目16番５号　株式会社日立情報システムズ内
(74)代理人　100077274
弁理士　磯村　雅俊　（外１名）

最終頁に続く

(54)【発明の名称】　画像データ配信システムおよびそれに用いる記録媒体

(57)【要約】
【課題】　クライアントコンピュータのネットワーク接続環境の時々刻々の変化に応じて、適切な品質の画像データを効率よく配信することが可能な画像データ配信システムおよびそのためのプログラムを記録した記録媒体の提供。
【解決手段】　画像データを配信可能な第１のコンピュータ（画像サーバコンピュータ１０１）とその画像データを受信する１以上の第２のコンピュータ（クライアントコンピュータ１０３）を有し、第１のコンピュータ１０１は、第２のコンピュータ１０３から回線状況通知メッセージを受信し、そのメッセージに基づいて、第２のコンピュータが接続されている回線の通信速度に最適な圧縮を施した画像データ（予め記憶している圧縮率の異なる複数の画像データから抽出した画像データかまたはその都度最適な圧縮率で圧縮した画像データ）を送信する。



1

【特許請求の範囲】

【請求項１】　画像データをストリーミング配信可能な第１のコンピュータとその画像データを受信する１以上の第２のコンピュータを有する画像データ配信システムにおいて、

前記第１のコンピュータは、

前記第２のコンピュータから回線状況通知メッセージに基づいて、該第２のコンピュータが接続されている回線の通信速度に最適な圧縮を施した画像データを送信する最適画像配信手段を有することを特徴とする画像データ配信システム。

【請求項２】　前記最適画像配信手段は、

送信時間を計測するタイマと、

圧縮率の異なる複数の画像データおよび該画像データ毎に該画像データの内容を識別可能な内部識別アドレスと送信の際の送信タイミングを規定するタイムスタンプとの対応データを保持する記憶手段と、

前記第２のコンピュータからの回線状況通知メッセージに基づいて前記記憶手段から最適な圧縮率の画像データを抽出するとともに、該抽出した画像データの、前記タイマで計測された送信時間に対応する前記内部識別アドレス以降の画像データ内容を配信する手段とを有することを特徴とする請求項１記載の画像データ配信システム。

【請求項３】　前記最適画像送出手段は、

１つの画像データを保持する記憶手段と、

前記第２のコンピュータから所定の時間間隔で受信する回線状況通知メッセージに基づいて、前記記憶手段に保持されている１つの画像データから、該第２のコンピュータが接続されている回線に最適な圧縮率の画像データを動的に作成して配信する手段とを有することを特徴とする請求項１記載の画像データ配信システム。

【請求項４】　前記第１のコンピュータは、画像サーバコンピュータであり、前記第２のコンピュータは、クライアントコンピュータであることを特徴とする請求項１または請求項２記載の画像データ配信システム。

【請求項５】　請求項１または２記載の第１のコンピュータにおける各手段および第２のコンピュータにおける各手段を実現する処理をプログラムコード化して記憶したことを特徴とするコンピュータで読み取り可能な記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、ネットワーク上でビデオストリーミングデータを送信する場合におけるデータ作成および配信システムに関し、特にデータを要求するクライアントコンピュータのネットワーク接続環境の動的な変化に対応して最適なビデオデータ配信を実現するビデオデータ配信システムおよびそのためのプログラムを記録した記録媒体に関する。

2

【０００２】

【従来の技術】ネットワーク上における一般的なビデオデータ配信においては、ビデオサーバコンピュータはクライアントコンピュータのデータ送信要求に対し該当するデータのストリーミング配信を行なっている。通常のビデオデータは非常にサイズが大きいため、回線上に大きな帯域幅を必要としていた。従って、従来は、限られたネットワーク内において回線の帯域幅を考慮し適切な圧縮を行うことにより、ビデオデータ配信が行われていた。

【０００３】しかし、近年のインターネット技術の普及に伴い、インターネットを利用したビデオ配信システムが多数考案されている。インターネットを利用した場合、インターネットに接続され世界中に点在する膨大な数のコンピュータがクライアントコンピュータとなり得る。そのため、クライアントコンピュータのネットワーク接続環境はクライアントコンピュータ毎に異なることとなる。

【０００４】従って、例えば高品質（低圧縮率）のビデオデータを配信した場合、帯域幅の狭い通信回線に接続しているクライアントコンピュータは、配信された全てのビデオデータを受信しきれなくなる。この理由により、インターネット上でビデオ配信を行なう場合には、できるだけ多くのクライアントコンピュータで受信可能となるように、帯域幅の狭い通信回線に合わせた低品質（高圧縮率）のビデオデータを配信するようにしていた。この環境においては、高品質のビデオデータの受信可能なクライアントコンピュータであっても品質の低いビデオデータを受信することに甘んじなければならなかった。

【０００５】この問題点に関するものとして、ビデオサーバが、予め記憶装置に格納してある中間形態の動画データを事前に高品質および低品質の複数のデータを準備し、さらにビデオデータを要求するクライアントコンピュータが接続されている回線の通信速度を取得する手段を具備し、取得した通信速度により最適なデータを配信する仕組みが考案されている。

【０００６】

【発明が解決しようとする課題】しかしながら、一般に、クライアントコンピュータのネットワーク接続環境すなわち帯域幅は回線の混雑の度合等により刻々と変化するものであるが、上記従来技術においては、動画データの配信要求のあったクライアント端末のネットワーク性能（伝送速度）を検知し、それに応じた圧縮率の動画データを送信するものであり、ネットワーク性能の時々刻々の変化に応じて送信ビデオデータの品質（データ圧縮率）を変化させるものとはなっていない。

【０００７】本発明の目的は、クライアントコンピュータのネットワーク接続環境すなわち帯域幅の時々刻々の変化に応じて、適切な品質の画像データを効率よく配信

Aylo Exhibit 1018
Page 1812 of 2308

することが可能な画像データ配信システムおよびそのためのプログラムを記録した記録媒体を提供することである。

【０００８】

【課題を解決するための手段】本発明は、上記目的を達成するために、画像データをストリーミング配信可能な第１のコンピュータ（好ましくは画像サーバコンピュータ）とその画像データを受信する１以上の第２のコンピュータ（好ましくはクライアントコンピュータ）を有する画像データ配信システムにおいて、第１のコンピュータは、第２のコンピュータから回線状況通知メッセージに基づいて、第２のコンピュータが接続されている回線の通信速度に最適な圧縮を施した画像データを送信する最適画像配信手段を有することを特徴としている。

【０００９】なお、前記最適画像配信手段は、（ａ）送信時間を計測するタイマと、圧縮率の異なる複数の画像データおよび該画像データ毎に該画像データの内容を識別可能な内部識別アドレスと送信の際の送信タイミングを規定するタイムスタンプとの対応データを保持する記憶手段と、第２のコンピュータからの回線状況通知メッセージに基づいて記憶手段から最適な圧縮率の画像データを抽出するとともに、該抽出した画像データの、タイマで計測された送信時間に対応する内部識別アドレス以降の画像データ内容を配信する手段とを有するか、あるいは、（ｂ）１つの画像データを保持する記憶手段と、第２のコンピュータから所定の時間間隔で受信する回線状況通知メッセージに基づいて、記憶手段に保持されている１つの画像データから、第２のコンピュータが接続されている回線に最適な圧縮率の画像データを動的に作成して配信する手段とを有している。

【００１０】また、本発明の記録媒体は、前記第１のコンピュータ（画像サーバコンピュータ）における各手段および第２のコンピュータ（クライアントコンピュータ）における各手段を実現する処理をプログラムコード化してコンピュータ読み取り可能なＣＤ－ＲＯＭ，ＤＶＤ，ＦＤなどの記録媒体に記録したものである。

【００１１】

【発明の実施の形態】以下、本発明の好ましい実施の形態を説明する。画像データを配信する機能を有するネットワークに接続されている画像配信サーバを具備し、該画像配信サーバは画像データを格納可能な記憶装置を具備し、画像配信サーバにアクセス可能な不特定多数のクライアントコンピュータを具備する構成とする。

【００１２】配信するビデオデータ作成において、ビデオサーバコンピュータは、オリジナルビデオデータを基に、高品質から低品質まで複数のビデオデータを記憶装置に作成し、個々のビデオデータに対し一定時間間隔のタイムスタンプを付加する機能を具備する。

【００１３】また、ビデオデータを受信するクライアントコンピュータでは、ビデオデータ受信中に受信パケットの状態を監視し、損失パケットが一定の基準を超えた場合、クライアント側の通信回線に十分な帯域が確保されていないと判定し、また、一定の基準時間の間に損失パケットが無い場合は、クライアント側の通信回線に十分な帯域が確保されていると判定し、その判定結果を配信中のビデオサーバコンピュータへ送信する手段を具備する。

【００１４】上述の判定結果を受信したビデオサーバは、その判定結果に適切な品質のビデオデータファイルを選択し、現在配信中のビデオデータファイルの切り換えを行う。その際、切り換え前後の２つのビデオデータに付加されているタイムスタンプを参照し、配信を中止する時点でのタイムスタンプと同様のタイムスタンプを持つ時点より切り換え後のビデオデータの配信を開始する機能を具備する。

【００１５】また、上記目的を達成するために、本発明は、ビデオデータを配信する機能を有するネットワークに接続されているビデオサーバコンピュータを具備し、さらにその記憶装置に記憶されているオリジナルビデオデータを送信時に所定の圧縮率で圧縮しながら配信する機能を具備する構成とする。

【００１６】ビデオデータを受信する際において、クライアントコンピュータは、ビデオデータ受信中に受信パケットの状態を監視し、損失パケットが一定の基準を超えた場合クライアント側の通信回線に十分な帯域が確保されていないと判定し、また、一定の基準の時間、損失パケットが無い場合、クライアント側の通信回線に十分な帯域が確保されていると判定し、その判定結果を配信中のビデオサーバコンピュータへ送信する手段を特徴とする機能を具備する。

【００１７】さらに、上述の判定結果を受信したビデオサーバコンピュータは、その判定結果に従い、動的に圧縮率に変更し、その変更された圧縮率で圧縮されたビデオデータを配信する機能を具備する。

【００１８】上記構成を採用することにより、配信されるビデオデータは、受信するクライアントコンピュータのネットワーク接続環境に従い配信するビデオデータの品質を動的に変化させることができる。これにより、クライアントコンピュータは常に最適化されたビデオデータを見掛け上切れ目なく受信することが可能となる。

【００１９】また、本発明の記録媒体は、上記の如きクライアントコンピュータが行う処理やビデオサーバコンピュータが行う処理をプログラムコード化してコンピュータ読み取り可能なＣＤ－ＲＯＭ，ＤＶＤ，ＦＤなどの記録媒体に記録したものである。

【００２０】（実施例の説明）以下、本発明の実施例として、２つの実施例を説明する。図１は、第１の実施例および第２の実施例における共通のハードウエア構成を示すブロック図である。本発明は、圧縮して配信することが可能な全てのデータ（特に画像データ）に対して適

Aylo Exhibit 1018
Page 1813 of 2308

用可能であるが、本実施例では、ビデオデータ（動画データ）を例にとって説明する。

【００２１】図１において、１０１はビデオデータの作成および配信を行なうビデオサーバコンピュータ、１０２はビデオデータを格納する補助記憶装置、１０３および１０３’はクライアントコンピュータ、１０４はビデオサーバコンピュータ１０１とクライアントコンピュータ１０３が接続されるネットワーク環境である。以下の説明では、１つのサーバコンピュータと１つ以上のクライアントコンピュータからなるクライアント・サーバシステムの場合を用いて説明するが、第１のコンピュータから第２のコンピュータへの２つのコンピュータ間でのデータの配信にも適用可能であることはいうまでもない。

【００２２】（第１の実施例）まず、第１の実施例について図２～図６を用いて説明する。図２は、記憶装置に格納する圧縮ビデオデータファイルのファイル構造の説明図である。図２において、２０３は圧縮ビデオデータファイルのヘッダ情報、２０４は本実施例で取り上げる画像データファイル、２０５はＥＯＦ（ＥＯＦ：End of File）を示す。また、２０１は本発明の特徴である圧縮ビデオデータファイルのタイムスタンプであり、２０２は、タイムスタンプ２０１と対応する圧縮ビデオデータファイルの内部識別アドレスである。なお、本実施例で取り上げる圧縮率の異なる他の画像ファイル（圧縮画像ファイルＢおよび圧縮画像ファイルＣ）構造も図２と同様である。内部識別アドレス２０２は、ファイル毎にタイムスタンプ１０２に対応して設けられるものであるが、ファイルの品質（圧縮率）が異なればそれに伴って異なる値を有する。

【００２３】図３は、同一画像ファイルから生成された圧縮率の異なる３種類の圧縮ビデオデータファイルのヘッダ情報の説明図である。同図（ａ）は低圧縮画像ファイルＡのヘッダ情報を、同図（ｂ）は中圧縮画像ファイルＢのヘッダ情報を、同図（ｃ）は高圧縮画像ファイルＣのヘッダ情報を示している。３０１はタイムスタンプ情報、３０２は、インデックスに対応する図２に示すファイルの内部識別アドレスである。

【００２４】図４は、圧縮率の異なる３つの画像ファイルとタイムスタンプと内部識別アドレスの関係を説明するための図である。同図（ａ）は、低圧縮画像ファイルＡとタイムスタンプと内部識別アドレスの関係を、同図（ｂ）は中圧縮画像ファイルＢとタイムスタンプと内部識別アドレスの関係を、同図（ｃ）は高圧縮画像ファイルＣとタイムスタンプと内部識別アドレスの関係を、それぞれ示している。本例では、説明を簡単にするために、同図（ａ）に示す低圧縮画像ファイルＡを１／２に圧縮したものを中圧縮ファイルＢとして同図（ｂ）に示し、さらにこれを１／２に圧縮したものを高圧縮画像ファイルＣとして同図（ｃ）に示すとともに、タイムスタンプ，内部識別アドレスと圧縮画像ファイルＡ～Ｃの内容の関係がわかりやすいように画像ファイル中に「い」、「ろ」、「は」・・・などを便宜的に示した。

【００２５】図５は、ビデオサーバコンピュータ１０１に保持される、通信速度と該通信速度に最適化された圧縮画像ファイルとの対応図である。図５における通信速度を表わす記号（ｍ、ｎ）の大小関係は、ｍ＜ｎとする。同図に示すように、通信速度が低速の場合は（～ｍ）、低品質である高圧縮画像ファイルＣを、通信速度が高速の場合は（ｎ～）、高品質である低圧縮画像ファイルＡを配信するようにする。

【００２６】図６および図７は、第１の実施例の動作を説明するためのフローチャートである。まず、ビデオデータを配信するビデオサーバコンピュータ１０１の動作を、図６を用いて詳細に説明する。

【００２７】図６において、クライアントコンピュータ１０３からのデータ要求に従い、該当する記憶装置内のビデオデータ（システムには初期配信ファイルが設定済とする。本実施例では、初期配信ファイルは「中圧縮画像ファイルＢ」とする）の配信を開始し、かつ配信時間の計測を開始する（ステップ６０１）。

【００２８】クライアントコンピュータ１０３から低品質データの受信要求（後述する図７のステップ７０６参照）があった場合（ステップ６０３）、図５の対応図に従い、現在送信中のファイル（ファイルＢ）より１段階品質の劣る「高圧縮画像ファイルＣ」を送信ファイルとして確定する（ステップ６０４、６０５）。

【００２９】次に現在時点の配信時間を測定する（ステップ６０６）。本実施例では、この時の測定時間が４ｍｓであったと仮定し、図３に示すヘッダ情報より、「ファイルＢ」の４ｍＳの開始内部識別アドレスは「iiii」と確定する（ステップ６０７）。すなわち、画像「い」、「ろ」は配信済みとする。ビデオサーバコンピュータ１０１は即座に配信中の「中圧縮画像ファイルＢ」の配信を中止し、内部識別アドレス「iiii」から「高圧縮画像ファイルＣ」の画像「は」、「に」、・・・の配信を開始する（ステップ６０８）。

【００３０】逆に、クライアントコンピュータ１０３より高品質データの受信要求（後述する図７のステップ７０７参照）があった場合（ステップ６０３：Ｙ）、図５の対応図に従い、現在配信中の「中圧縮画像ファイルＢ」より１段階品質の高い「低圧縮画像ファイルＡ」を次送信ファイルとして確定する（ステップ６０４、６０５）。以下の動作は同様である。すなわち、ビデオサーバコンピュータ１０１は即座に配信中の「中圧縮画像ファイルＢ」の配信を中止し、内部識別アドレス「iiii」から「低圧縮画像ファイルＡ」の画像「は」、「に」、・・・の配信を開始する（ステップ６０８）。

【００３１】なお、現時点での配信データが「高圧縮画像ファイルＣ」であった場合にさらに低品質の高圧縮の

Aylo Exhibit 1018
Page 1814 of 2308

画像ファイルが要求された場合、図５より１段階品質の劣るビデオデータは存在しないことがわかる。また、現時点での配信データが「低圧縮画像ファイルＡ」であった場合にさらに高品質の低圧縮画像ファイルが要求された場合にも、図５より１段階品質の高いビデオデータは存在しないことがわかる。このような場合は、データの存在しない旨のメッセージを送信し、現時点でのデータ配信を続行する（ステップ６０９）。

【００３２】また、ビデオサーバコンピュータ１０１は、ビデオデータの終端すなわち図２に示す「ＥＯＦ」まで配信を行った場合（ステップ６０２：Ｙ）、データの配信を終える。

【００３３】次に、ビデオデータを受信するクライアントコンピュータ１０３の動作を、図７を用いて詳細に説明する。図７において、先ず、ビデオサーバコンピュータ１０１に対するビデオデータ要求により、ビデオサーバから送信されたデータの受信を開始し、同時に損失パケット監視タイマをリセットする（ステップ７０１）。

【００３４】損失パケット数がシステムに既設定済の数を超過した場合（ステップ７０２：Ｙ）、クライアントコンピュータ側の通信回線に十分な帯域が確保されていないと判断し、ビデオサーバに低品質データの要求メッセージを送信する（ステップ７０６）。損失パケット数がシステムに既設定済の数を超過しない場合（ステップ７０２：Ｎ）、損失パケットの発生を常に監視し（ステップ７０３）、損失パケットが発生した場合（ステップ７０３：Ｙ）、損失パケット監視タイマをリセットする（ステップ７０４）。

【００３５】この損失パケット監視タイマが、システムに既設定済の一定時刻を超過した場合、すなわち損失パケットが一定時間発生しなかった場合（ステップ７０３：Ｎ，７０５：Ｙ）、クライアントコンピュータ１０３側の通信回線に十分な帯域が確保されており現在受信中のビデオデータよりも高品質なデータ（低圧縮のデータ）の受信が可能と判断し、ビデオサーバコンピュータ１０１に高品質データの要求メッセージを送信する（ステップ７０７）。ステップ７０８では、ビデオサーバコンピュータ１０１からの配信データに従い今までと異なる品質の配信データを受信し、同時に損失パケット監視タイマをリセットする。

【００３６】ここで、図６のステップ６０９に示すデータ無しのメッセージを受信した場合、そのメッセージに従い、低品質データ無しの場合はステップ７０６の低品質データ要求メッセージ送信を、また、高品質データ無しの場合はステップ７０７の高品質データ要求メッセージ送信を中止する。なお、クライアントコンピュータ１０３はビデオデータの終端をしめす「ＥＯＦ」まで受信した場合（ステップ７０９：Ｙ）、受信動作を終了する。

【００３７】最後に、ビデオデータフォーマットについて詳述する。本実施例における配信ビデオデータのフォーマットは、図２に示すようにヘッダ情報部分（２０３）とデータ部分（２０４）とから構成され、ヘッダ部分には、図３に示すような、タイムスタンプ毎のアドレスが保持されている。ビデオサーバによりビデオデータを圧縮する際に、ヘッダ部分を生成し付加する。なお、図２および図３における２０１と３０１、２０２と３０２は同値である。

【００３８】以上、第１の実施例は、オリジナルなビデオデータを基に高品質（低圧縮）から低品質（高圧縮）まで、品質の異なる複数のビデオデータをビデオサーバの記憶装置に記憶しておき、クライアントコンピュータからの配信要求に対し、クライアントコンピュータにおける受信状態に応じて、最適な品質のビデオデータを動的に切替えながら配信するものである。

【００３９】（第２の実施例）次に説明する第２の実施例は、ビデオサーバに保持するビデオデータは、オリジナルなもの１種類のみとし、サーバから配信するビデオデータの圧縮を、クライアントコンピュータにおける受信状態に応じて動的に最適化するものである。

【００４０】以下、本第２実施例のビデオサーバ側の処理を、図８を用いて詳細に説明する。同図において、まず、クライアントコンピュータからのデータ要求に従い、該当する記憶装置内のオリジナルビデオデータをビデオサーバに取り出し（ステップ８０１）、圧縮を行いながら（システムには初期圧縮レートが設定済とする。）送信する（ステップ８０３）。クライアントコンピュータ１０３より低品質データ要求または高品質データ要求のメッセージ受信があった場合（ステップ８０４）、圧縮レートを変更する（ステップ８０５）。

【００４１】なお、ビデオサーバコンピュータ１０１はビデオデータの終端まで配信した場合（ステップ８０２）、配信動作を終了する。

【００４２】ビデオデータを受信するクライアントコンピュータの動作は、第１の実施例と同様のため詳細な説明は省略する。

【００４３】なお、上記の如き、ビデオサーバコンピュータが行う各処理手順およびクライアントコンピュータが行う各処理手順を、ＣＤ－ＲＯＭ，ＤＶＤ，あるいはＦＤなどの記録媒体に一緒にまたは別々に記録して市場に流通させることができる。利用者は、この記録媒体をビデオサーバコンピュータとクライアントコンピュータにインストールすることにより、本発明を簡単に利用することが可能になる。

【００４４】

【発明の効果】以上詳細に説明したように、本発明によれば、ビデオサーバにより配信されるビデオデータは、受信するクライアントコンピュータのネットワーク接続環境に従い動的に変化させることが実現できる。これにより、クライアントコンピュータは常に最適化されたビ

Aylo Exhibit 1018
Page 1815 of 2308

デオデータを見掛け上切れ目なく受信することが可能となる。

【図面の簡単な説明】

【図１】本発明の第１の実施例および第２の実施例における共通のハードウエア構成を示すブロック図である。

【図２】記憶装置に格納する圧縮ビデオデータファイルのファイル構造の説明図である。

【図３】同一画像ファイルから生成された圧縮率の異なる３種類の圧縮ビデオデータファイルのヘッダ情報の説明図である。

【図４】圧縮率の異なる３つの画像ファイルとタイムスタンプと内部識別アドレスの関係を説明するための図である。

【図５】ビデオサーバコンピュータに保持される、通信速度と該通信速度に最適化された圧縮画像ファイルとの対応図である。

【図６】第１の実施例におけるビデオサーバコンピュータの動作を説明するためのフローチャートである。

【図７】第１の実施例におけるクライアントコンピュータの動作を説明するためのフローチャートである。

【図８】第２の実施例におけるビデオサーバコンピュータの動作を説明するためのフローチャートである。

【符号の説明】

１０１：ビデオデータの作成および配信を行なうビデオサーバコンピュータ（第１のコンピュータ）

１０２：ビデオデータを格納する記憶装置

１０３，１０３’：ネットワークと接続されるクライアントコンピュータ（第２のコンピュータ）

１０４：ビデオサーバコンピュータとクライアントコンピュータが接続されるネットワーク環境

１０５：プログラム記録媒体（ＣＤ－ＲＯＭ）

１０６：プログラム記録媒体（ＣＤ－ＲＯＭ）

２０１：圧縮ファイルのタイムスタンプ

２０２：圧縮ファイルのタイムスタンプに対応した内部識別アドレス

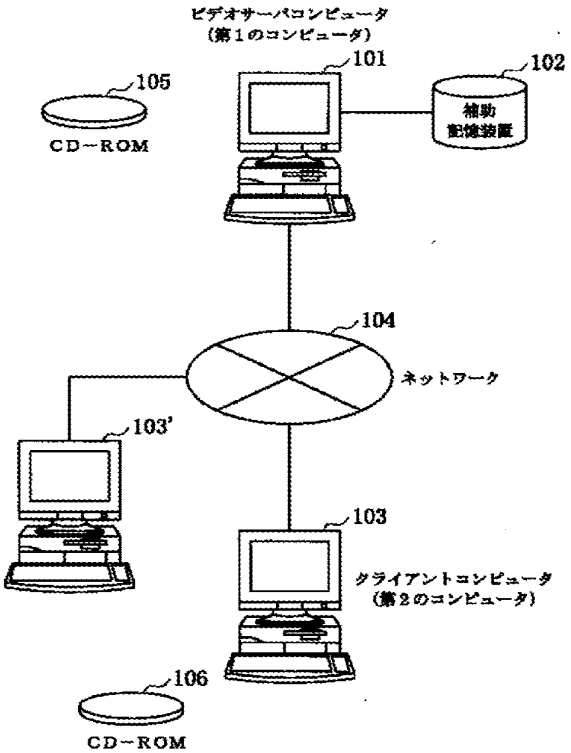
２０３：圧縮ファイルのファイルのヘッダ情報

２０４：本実施例で取り上げるファイルＡのファイルフォーマット

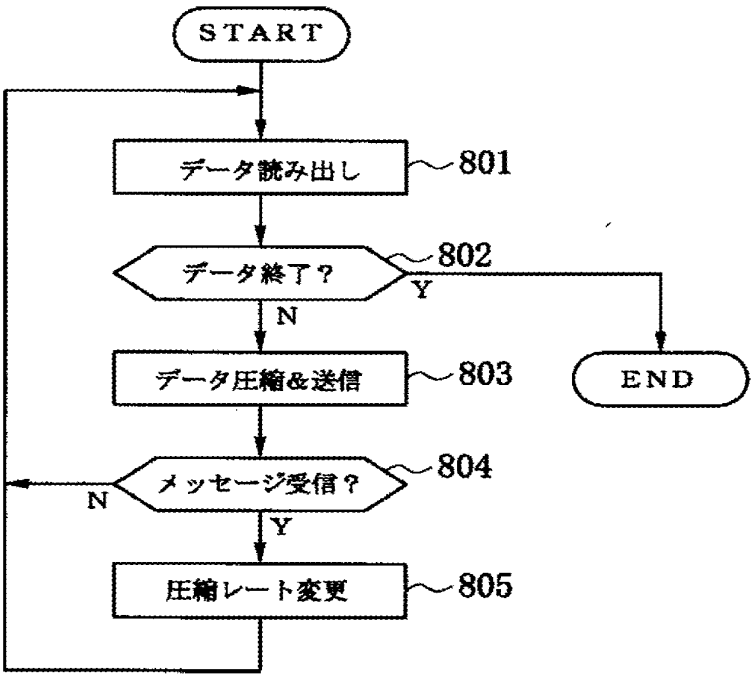
３０１：圧縮ファイルヘッダのタイムスタンプ情報

３０２：圧縮ファイルヘッダのタイムスタンプ情報に対応する内部識別アドレス

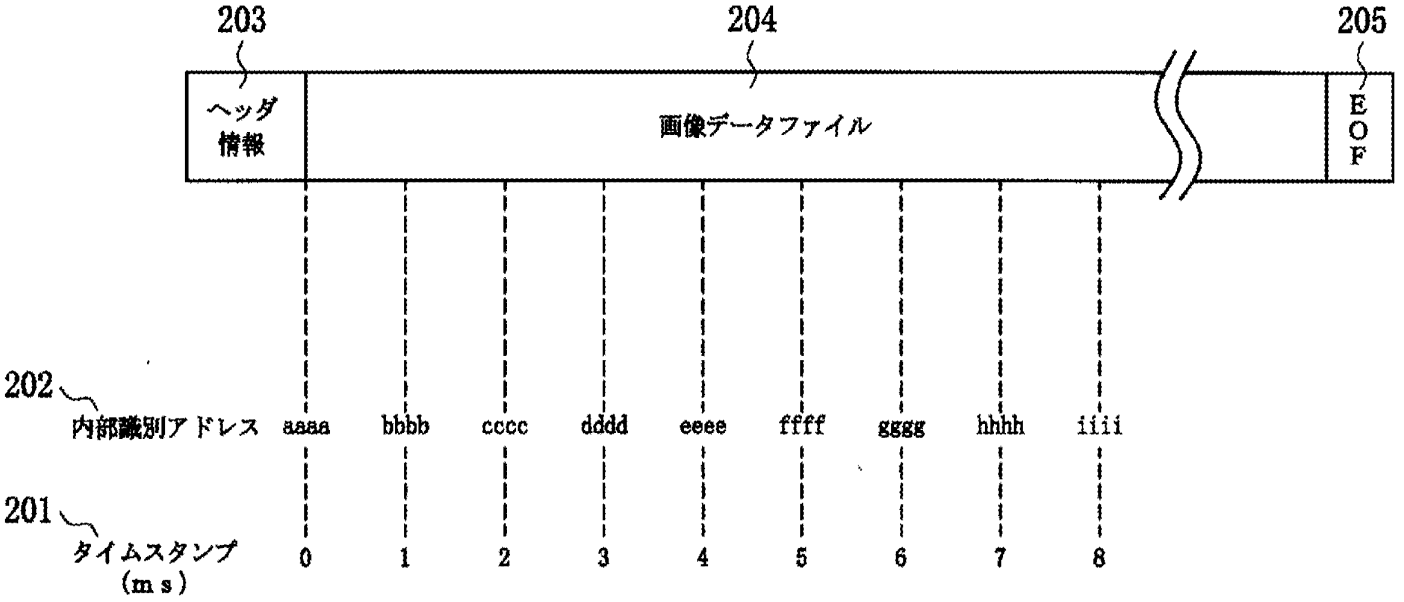
【図１】



【図８】



【図２】



Aylo Exhibit 1018
Page 1817 of 2308

【図３】

ヘッダ情報の構造

301　302

タイムスタンプ	内部識別アドレス
0	a a a a
1	b b b b
2	c c c c
3	d d d d
4	e e e e
5	f f f f
6	g g g g
7	h h h h
8	i i i i
:	:

低圧縮
画像ファイル（A）

（a）

301　302

タイムスタンプ	内部識別アドレス
0	a a a a
1	c c c c
2	e e e e
3	g g g g
4	i i i i
5	k k k k
6	m m m m
7	o o o o
8	q q q q
:	:

中圧縮
画像ファイル（B）

（b）

301　302

タイムスタンプ	内部識別アドレス
0	a a a a
1	e e e e
2	i i i i
3	m m m m
4	q q q q
5	u u u u
6	y y y y
7	:
8	:
:	:

高圧縮
画像ファイル（C）

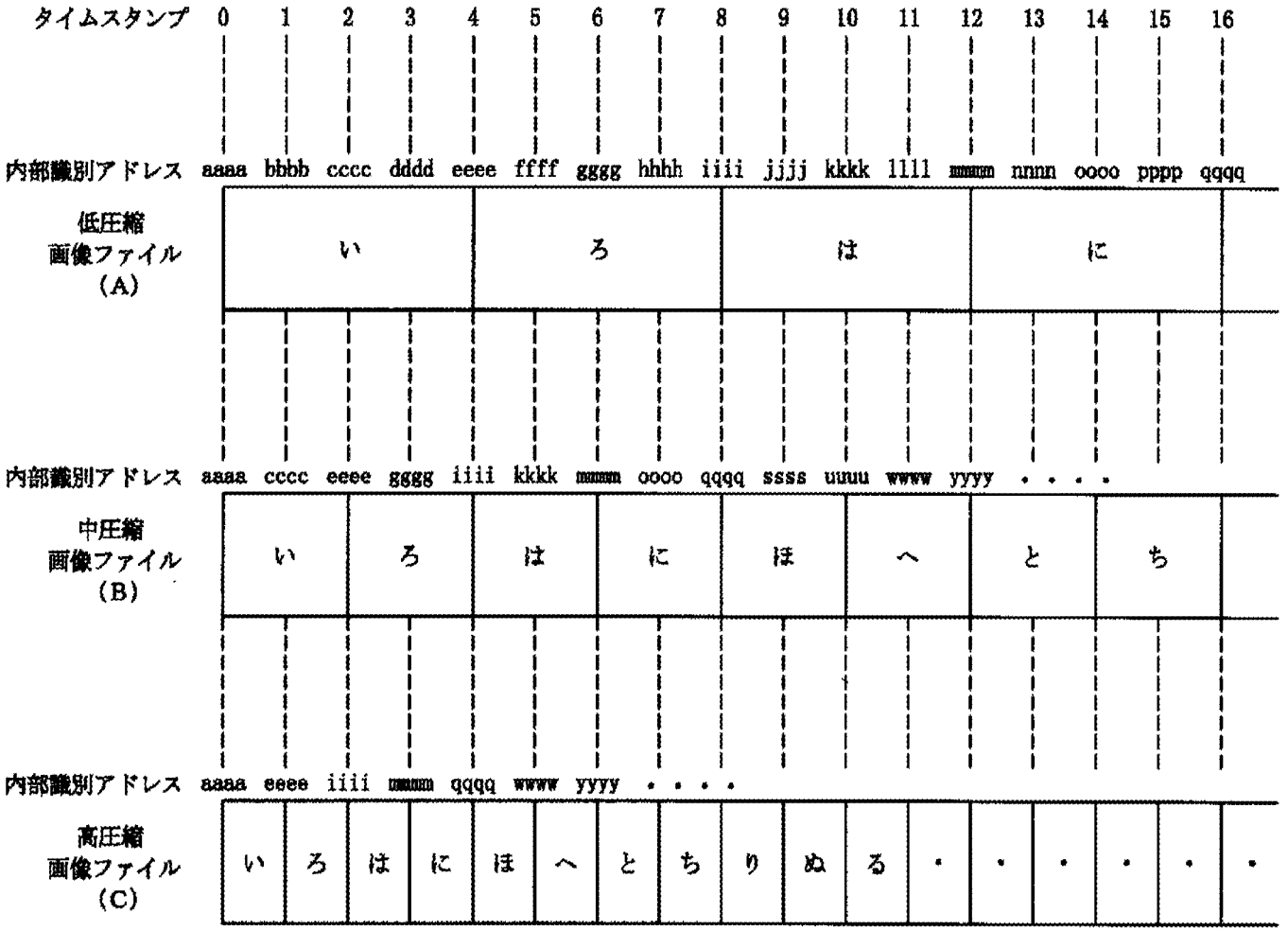
（c）

【図５】

通信速度と対応圧縮ファイル

項番	通信速度（ｂｐｓ）	ファイル名
1	～m	圧縮画像ファイルC
2	m～n	圧縮画像ファイルB
3	n～	圧縮画像ファイルA

通信速度の大小関係　m＜n



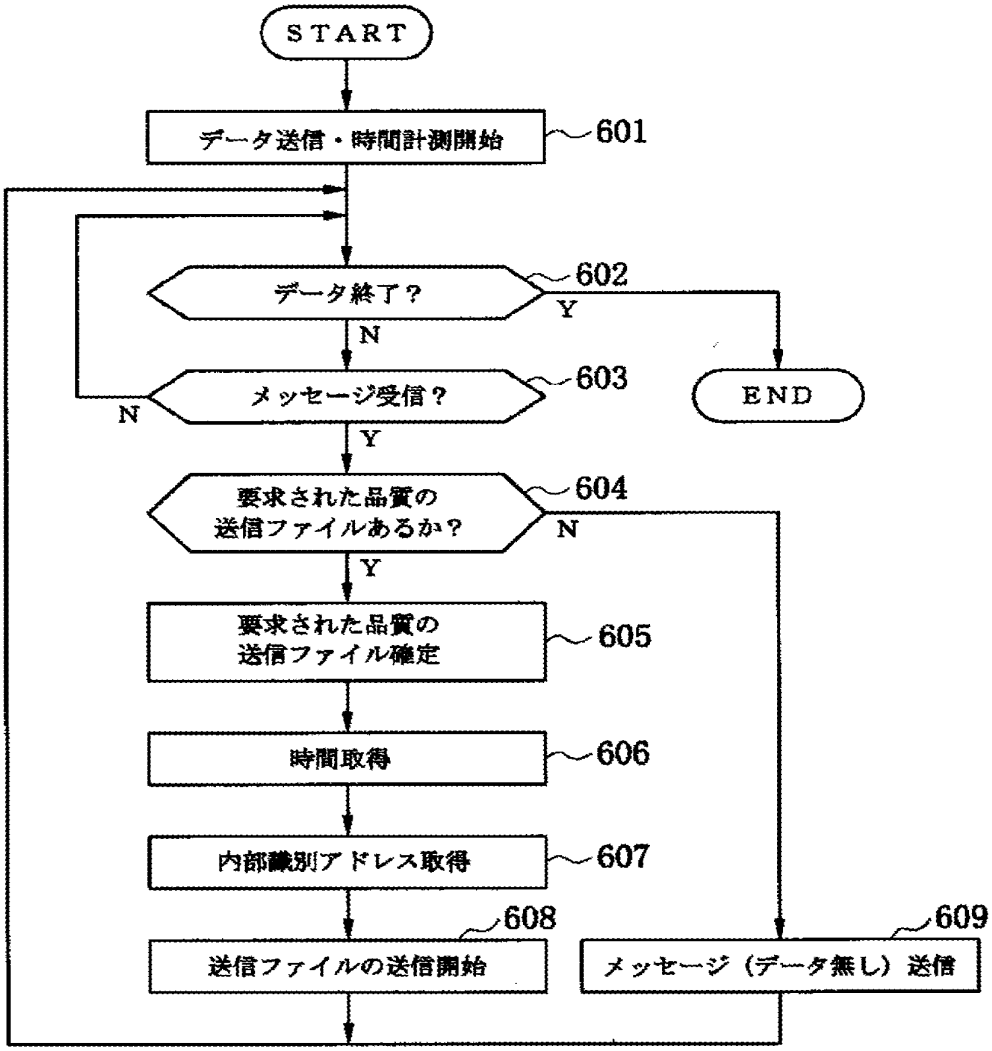
【図４】

特開２００１－９２７５２

Aylo Exhibit 1018
Page 1819 of 2308

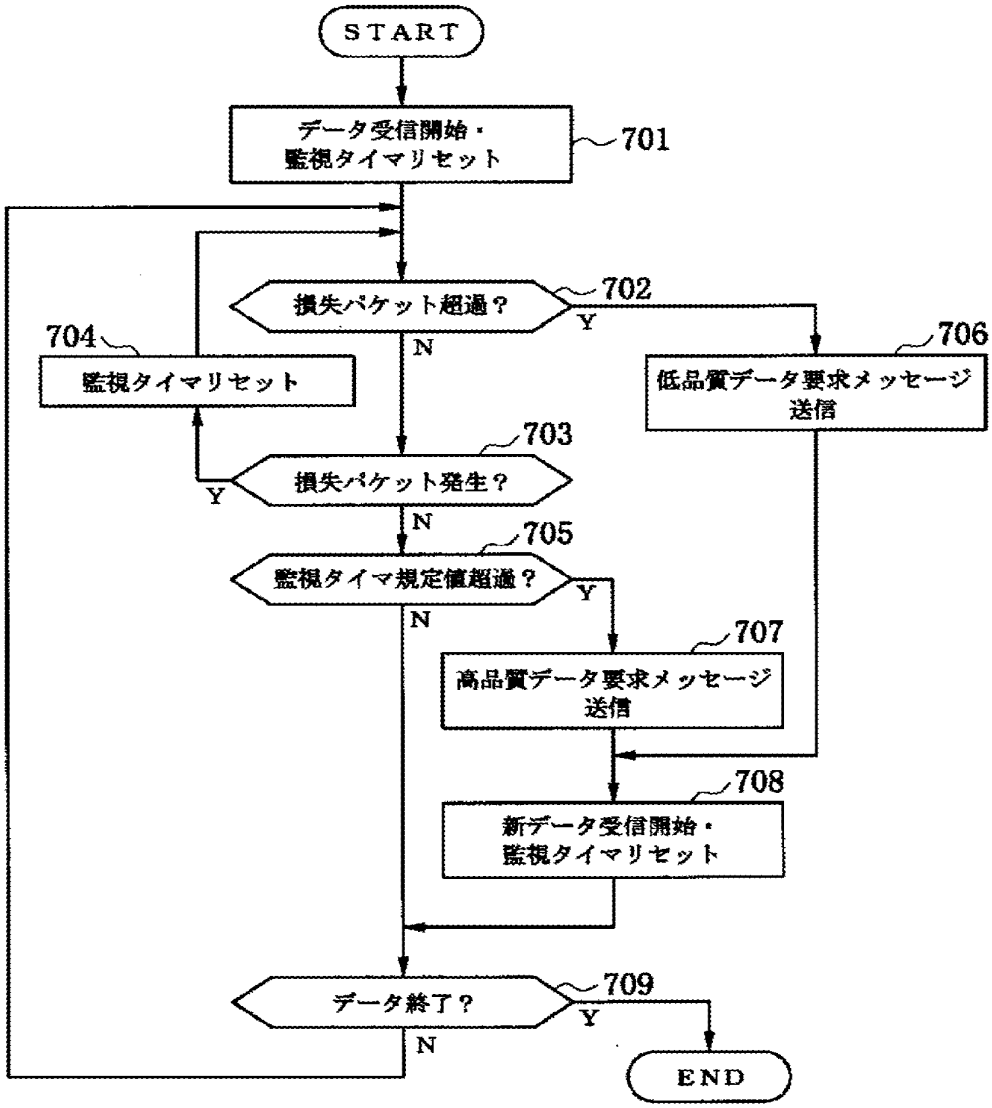
【図６】

（ビデオサーバコンピュータの処理）



【図７】

(クライアントコンピュータの処理)



フロントページの続き

(72)発明者　二郷　正樹
東京都渋谷区道玄坂一丁目16番５号　株式
会社日立情報システムズ内

Ｆターム(参考)　5B082 GA01 HA05
5B089 GA11 JA33 JB03 JB22 KA05
KA12 KB06 KB11 KC28 KC53
KC59 KC60 KH28
5C059 KK34 SS09 TA60 TA71 TC21
UA02 UA38
5K030 GA08 HB02 LA07 LC09 LD13
MB13

Aylo Exhibit 1018
Page 1821 of 2308

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2004-054930
(43)Date of publication of application : 19.02.2004

(51)Int.Cl. G06F 13/42
G06F 13/38
H04N 7/173

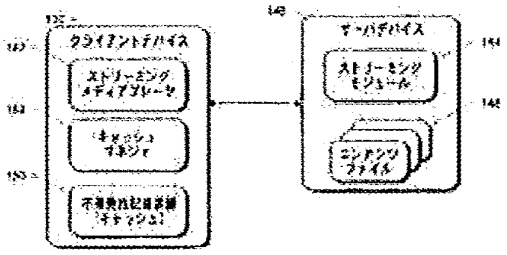
(21)Application number : 2003-180339
(22)Date of filing : 24.06.2003

(71)Applicant : MICROSOFT CORP
(72)Inventor : KLEMETS ANDERS E
BATTERBERRY TROY D
OLIVEIRA EDUARDO P

(30)Priority
Priority number : 2002 179770 Priority date : 24.06.2002 Priority country : US

(54) COMPUTER READABLE MEDIUM, SYSTEM, AND METHOD FOR CLIENT-SIDE CACHING OF STREAMING MEDIA CONTENT

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a technique for client-side caching of a streaming media content.
SOLUTION: Various functions with respect to streaming media content are made available to users. Such functions includes one or more functions for streaming the media content at a rate independent of the encoded bit rate of the content to allow streaming of the content to continue even when a user has selected various shuttle control options (e.g., pause, stop, fast forward, seek, rewind, etc.), a function for allowing the streaming content to be recorded for playback at a later time, and a function for allowing the streaming content to be played back while shifting time.



(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2004-54930
(P2004-54930A)
(43)公開日　平成16年2月19日(2004.2.19)

(51)Int.Cl.[7]	F I			テーマコード（参考）
G06F 13/42	G06F	13/42	320C	5B077
G06F 13/38	G06F	13/38	340C	5C064
H04N 7/173	H04N	7/173	640Z	

審査請求 未請求　請求項の数 77　OL　（全 35 頁）

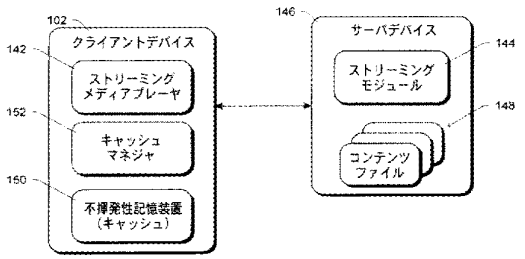
(21)出願番号　特願2003-180339(P2003-180339)
(22)出願日　平成15年6月24日(2003.6.24)
(31)優先権主張番号　10/179,770
(32)優先日　平成14年6月24日(2002.6.24)
(33)優先権主張国　米国(US)

(71)出願人　391055933
マイクロソフト　コーポレイション
ＭＩＣＲＯＳＯＦＴ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国　ワシントン州　９８０５２－６３９９　レッドモンド　ワン　マイクロソフト　ウェイ　（番地なし）
(74)代理人　100077481
弁理士　谷　義一
(74)代理人　100088915
弁理士　阿部　和夫
(72)発明者　アンダース　イー．クレメッツ
アメリカ合衆国　９８１２２　ワシントン州　シアトル　２８　アベニュー　５２３

最終頁に続く

(54)【発明の名称】ストリーミングメディアコンテンツのクライアント側でのキャッシングのためのコンピュータ可読媒体、システムおよび方法

(57)【要約】
【課題】ストリーミングメディアコンテンツのクライアント側でのキャッシングの技術を提供する。
【解決手段】ストリーミングメディアコンテンツに関する各種の機能をユーザが利用できるようにする。そのような機能には、メディアコンテンツをそのコンテンツの符号化ビットレートに依存しないレートでストリーミングすることにより、ユーザが各種のシャトルコントロールオプション（例えば一時停止、停止、早送り、シーク、巻き戻しなど）を選択した際でも、コンテンツのストリーミングを継続することを可能にする、また後の時間に再生するためにストリーミングコンテンツを記録することを可能にする、また時間をずらしてストリーミングコンテンツを再生することを可能にする、のいずれか１つまたは複数が含まれる。
【選択図】　図２



【特許請求の範囲】
【請求項１】
クライアントデバイスの１つまたは複数のプロセッサによって実行されるときに、前記１つまたは複数のプロセッサにより、
サーバデバイスから、ストリーミングメディアコンテンツを該ストリーミングメディアコンテンツの符号化ビットレートに依存しないレートで受信させる
複数の命令が格納されていることを特徴とする１つまたは複数のコンピュータ可読媒体。
【請求項２】
前記複数の命令は、さらに前記１つまたは複数のプロセッサに、
前記受信したストリーミングメディアコンテンツを不揮発性の記憶装置に保存させる
ことを特徴とする請求項１に記載の１つまたは複数のコンピュータ可読媒体。
【請求項３】
前記サーバデバイスはプロキシサーバを含むことを特徴とする請求項１に記載の１つまたは複数のコンピュータ可読媒体。
【請求項４】
前記複数の命令はさらに前記１つまたは複数のプロセッサに、
前記ストリーミングメディアコンテンツについて要求されたレートを指示する速度要求を前記サーバデバイスに送信させ、
前記ストリーミングメディアコンテンツの前記レートの指示を前記サーバデバイスから受信させる
ことを特徴とする請求項１に記載の１つまたは複数のコンピュータ可読媒体。
【請求項５】
前記要求されたレートは値ｘで表され、前記速度要求は、
プラグマ：速度＝ｘ
を含むストリーミング速度メッセージヘッダであることを特徴とする請求項４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項６】
前記要求されたレートは値ｘで表され、前記速度要求は、
速度＝ｘ
を含むストリーミング速度メッセージヘッダであることを特徴とする請求項４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項７】
サーバデバイスで実施される方法であって、
ストリーミングメディアコンテンツに対する要求をクライアントデバイスから受信することと、
前記メディアコンテンツを前記コンテンツの符号化ビットレートに依存しないレートで前記クライアントデバイスにストリーミングすることと
を含むことを特徴とする方法。
【請求項８】
前記ストリーミングメディアコンテンツに対する要求をユーザから受信することと、
前記ユーザの要求に応答して、前記サーバデバイスに前記ストリーミングメディアコンテンツを要求することと
をさらに含むことを特徴とする請求項７に記載の方法。
【請求項９】
前記受信したストリーミングメディアコンテンツをキャッシュすること
をさらに含むことを特徴とする請求項７に記載の方法。
【請求項１０】
前記サーバデバイスは、プロキシサーバを含むことを特徴とする請求項７に記載の方法。
【請求項１１】
前記ストリーミングメディアコンテンツに要求されたレートを指示する速度要求を前記サ

ーバデバイスに送信することと、
前記ストリーミングメディアコンテンツの前記レートの指示を前記サーバデバイスから受信することと
をさらに含むことを特徴とする請求項７に記載の方法。
【請求項１２】
前記要求されたレートは値ｘで表され、前記速度要求は、
プラグマ：速度＝ｘ
を含むストリーミング速度メッセージヘッダであることを特徴とする請求項１１に記載の方法。
【請求項１３】
前記要求されたレートは値ｘで表され、前記速度要求は、速度＝ｘ
を含むストリーミング速度メッセージヘッダであることを特徴とする請求項１１に記載の方法。
【請求項１４】
クライアントデバイスの１つまたは複数のプロセッサによって実行されると、前記１つまたは複数のプロセッサに、
ストリーミングメディアコンテンツの再生を少なくとも一時的に停止する要求を受信させ、
前記クライアントデバイスにおける前記ストリーミングメディアコンテンツの再生を停止させ、
前記ストリーミングメディアコンテンツの再生を停止させる要求に関わらず、前記ストリーミングメディアコンテンツの受信を継続させる
複数の命令が格納されていることを特徴とする１つまたは複数のコンピュータ可読媒体。
【請求項１５】
前記要求は停止コマンドを含むことを特徴とする請求項１４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項１６】
前記要求は一時停止コマンドを含むことを特徴とする請求項１４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項１７】
前記複数の命令はさらに、前記１つまたは複数のプロセッサに前記受信したストリーミングメディアコンテンツをキャッシュさせることを特徴とする請求項１４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項１８】
サーバからストリーミングメディアコンテンツを受信することと、
ユーザからシャトルコントロールオプションを受信することと、
前記サーバに前記シャトルコントロールオプションを指示せずに、前記シャトルコントロールオプションに従って前記ストリーミングメディアコンテンツの再生を変更することと
を含むことを特徴とする方法。
【請求項１９】
前記シャトルコントロールオプションは停止コマンドを含むことを特徴とする請求項１８に記載の方法。
【請求項２０】
前記シャトルコントロールオプションは一時停止コマンドを含むことを特徴とする請求項１８に記載の方法。
【請求項２１】
前記シャトルコントロールオプションは巻き戻しコマンドを含むことを特徴とする請求項１８に記載の方法。
【請求項２２】
前記シャトルコントロールオプションは早送りコマンドを含むことを特徴とする請求項１

８に記載の方法。
【請求項２３】
前記サーバから受信すると前記ストリーミングメディアコンテンツをキャッシュすることをさらに含むことを特徴とする請求項１８に記載の方法。
【請求項２４】
クライアントデバイスの１つまたは複数のプロセッサによって実行されると、前記１つまたは複数のプロセッサに、
サーバデバイスからストリーミングメディアコンテンツを受信させ、
前記受信したストリーミングメディアコンテンツを前記クライアントデバイスの不揮発性記憶装置に保存させる
複数の命令が格納されていることを特徴とする１つまたは複数のコンピュータ可読媒体。
【請求項２５】
前記サーバデバイスはプロキシサーバを含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項２６】
前記不揮発性記憶装置はハードディスクドライブを含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項２７】
前記不揮発性記憶装置は、光ディスクおよび光ディスクドライブを含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項２８】
前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、コンテンツをキャッシュすることが許される旨の指示を前記サーバデバイスから受信した場合にのみ、前記１つまたは複数のプロセッサに、その受信したストリーミングメディアコンテンツを保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項２９】
前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、ストリーミングメディアコンテンツのキャッシングが前記サーバデバイスによってサポートされることを示すサポートヘッダを含むメッセージを前記サーバデバイスから受信した場合にのみ、前記１つまたは複数のプロセッサに、その受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項３０】
前記サポートヘッダは、ストリーミングメディアコンテンツのクライアントでのキャッシングが前記サーバデバイスによってサポートされることを示す機能名を含み、前記機能名は、
ｃｏｍ．ｍｉｃｒｏｓｏｆｔ．ｗｍ．ｆａｓｔｃａｃｈｅ
を含むことを特徴とする請求項２９に記載の１つまたは複数のコンピュータ可読媒体。
【請求項３１】
前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、キャッシュ制御指示に関連のストリーミングメディアコンテンツをキャッシュすることが前記クライアントデバイスに許されることを示すキャッシュ制御指示を含むメッセージを前記サーバデバイスから受信した場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。
【請求項３２】
前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを

前記不揮発性記憶装置に保存させる前記命令は、前記ストリーミングメディアコンテンツのキャッシュが許されることが、前記ストリーミングメディアコンテンツの識別子によって示されている場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３３】

前記識別子はＵＲＬ（Ｕｎｉｆｏｒｍ　ＲｅｓｏｕｒｃｅＬｏｃａｔｏｒ）を含むことを特徴とする請求項３２に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３４】

前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしていない場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３５】

前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、前記クライアントデバイスのメディアプレーヤがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしていない場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３６】

前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる前記命令は、前記ストリーミングメディアコンテンツのタイプがキャッシュ可能タイプである場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存させる命令を含むことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３７】

前記複数の命令はさらに前記１つまたは複数のプロセッサに、
前記ストリーミングメディアコンテンツに対する事後の要求を受信させ、
前記事後の要求に応答して、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置から取得させる
ことを特徴とする請求項２４に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３８】

前記複数の命令はさらに前記１つまたは複数のプロセッサに、
前記ストリーミングメディアコンテンツの有効期限が切れているかどうかを調べさせ、
前記ストリーミングメディアコンテンツの有効期限が切れていない場合にのみ、前記事後の要求に応答して、前記不揮発性記憶装置から前記ストリーミングメディアコンテンツを取得させる
ことを特徴とする請求項３７に記載の１つまたは複数のコンピュータ可読媒体。

【請求項３９】

前記複数の命令はさらに前記１つまたは複数のプロセッサに、
前記ストリーミングメディアコンテンツの有効期限が切れているかどうかを調べさせ、
前記ストリーミングメディアコンテンツの有効期限が切れている場合は、前記ストリーミングメディアコンテンツを再確認させる
ことを特徴とする請求項３７に記載の１つまたは複数のコンピュータ可読媒体。

【請求項４０】

前記複数の命令はさらに前記１つまたは複数のプロセッサに、

Aylo Exhibit 1018
Page 1827 of 2308

前記不揮発性記憶装置からの前記ストリーミングメディアコンテンツの再生に係わるロギング情報を含むメッセージを生成させ、前記メッセージは、データ構造のメッセージ本体部分に前記ロギングが含まれることを示すロギング情報ヘッダを含むことを特徴とする請求項３７に記載の１つまたは複数のコンピュータ可読媒体。

【請求項４１】

前記ロギング情報ヘッダは、

Ｃｏｎｔｅｎｔ－Ｔｙｐｅ：ａｐｐｌｉｃａｔｉｏｎ／ｘ－ｗｍｓ－ｓｅｎｄｅｖｅｎｔ

を含むことを特徴とする請求項４０に記載の１つまたは複数のコンピュータ可読媒体。

【請求項４２】

不揮発性記憶装置と、

システム上で再生するためにストリーミングメディアコンテンツをサーバデバイスから受信するように構成され、さらに、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されたストリーミングメディアプレーヤコンポーネントとを備えることを特徴とするシステム。

【請求項４３】

前記ストリーミングメディアプレーヤコンポーネントはさらに、ストリーミングメディアコンテンツのキャッシングが前記サーバデバイスによってサポートされることを示すサポートヘッダを含むメッセージを前記サーバデバイスから受信した場合にのみ、前記受信したストリーミングメディアコンテンツを前記不揮発性記憶装置に保存するように構成されることを特徴とする請求項４２に記載のシステム。

【請求項４４】

前記サポートヘッダは、ストリーミングメディアコンテンツのクライアントでのキャッシングが前記サーバデバイスによってサポートされることを示す機能名を含み、前記機能名は、

ｃｏｍ．ｍｉｃｒｏｓｏｆｔ．ｗｍ．ｆａｓｔｃａｃｈｅ

を含むことを特徴とする請求項４３に記載のシステム。

【請求項４５】

前記ストリーミングメディアプレーヤコンポーネントはさらに、前記コンテンツのキャッシングが許されることを示す指示を前記サーバデバイスから受信した場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されることを特徴とする請求項４２に記載のシステム。

【請求項４６】

前記ストリーミングメディアプレーヤコンポーネントはさらに、キャッシュ制御指示に関連のストリーミングメディアコンテンツをキャッシュすることが前記システムに許されることを示すキャッシュ制御指示を含むメッセージを前記サーバデバイスから受信した場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されることを特徴とする請求項４２に記載のシステム。

【請求項４７】

前記ストリーミングメディアプレーヤコンポーネントはさらに、前記ストリーミングメディアコンテンツのキャッシングが許されることが、前記ストリーミングメディアコンテンツの識別子によって示されている場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されることを特徴とする請求項４２に記載のシステム。

【請求項４８】

前記識別子はＵＲＬ（Ｕｎｉｆｏｒｍ　Ｒｅｓｏｕｒｃｅ　Ｌｏｃａｔｏｒ）を含むことを特徴とする請求項４７に記載のシステム。

【請求項４９】

前記ストリーミングメディアプレーヤコンポーネントはさらに、ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしていない場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成

Aylo Exhibit 1018
Page 1828 of 2308

されることを特徴とする請求項４２に記載のシステム。
【請求項５０】
前記ストリーミングメディアプレーヤコンポーネントはさらに、前記ストリーミングメディアプレーヤがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしていない場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されることを特徴とする請求項４２に記載のシステム。
【請求項５１】
前記ストリーミングメディアプレーヤコンポーネントはさらに、前記ストリーミングメディアコンテンツのタイプがキャッシュ可能タイプである場合にのみ、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶するように構成されることを特徴とする請求項４２に記載のシステム。
【請求項５２】
前記ストリーミングメディアプレーヤコンポーネントはさらに、
前記ストリーミングメディアコンテンツに対する事後の要求を受信し、
前記事後の要求に応答して、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置から取得する
ように構成されることを特徴とする請求項４２に記載のシステム。
【請求項５３】
前記ストリーミングメディアプレーヤコンポーネントはさらに、
前記ストリーミングメディアコンテンツの有効期限が切れているかどうかを調べ、
前記ストリーミングメディアコンテンツの有効期限が切れていない場合にのみ、前記事後の要求に応答して、前記ストリーミングメディアコンテンツを前記不揮発性記憶装置から取得する
ように構成されることを特徴とする請求項５２に記載のシステム。
【請求項５４】
前記ストリーミングメディアプレーヤコンポーネントはさらに、
前記ストリーミングメディアコンテンツの有効期限が切れているかどうかを調べ、
前記ストリーミングメディアコンテンツの有効期限が切れている場合は、前記ストリーミングメディアコンテンツを再確認する
ように構成されることを特徴とする請求項５２に記載のシステム。
【請求項５５】
前記ストリーミングメディアプレーヤコンポーネントはさらに、
前記不揮発性記憶装置からの前記ストリーミングメディアコンテンツの再生に係わるロギング情報を含むメッセージを生成し、前記メッセージは、データ構造のメッセージ本体部分に前記ロギング情報が含まれることを示すロギング情報ヘッダを含む
ように構成されることを特徴とする請求項５２に記載のシステム。
【請求項５６】
前記ロギング情報ヘッダは、
Ｃｏｎｔｅｎｔ－Ｔｙｐｅ：ａｐｐｌｉｃａｔｉｏｎ／ｘ－ｗｍｓ－ｓｅｎｄｅｖｅｎｔ
を含むことを特徴とする請求項５５に記載のシステム。
【請求項５７】
クライアントデバイスの１つまたは複数のプロセッサによって実行されると、前記１つまたは複数のプロセッサに、
サーバデバイスからストリーミングメディアコンテンツを受信させ、
後の時間に再生するために、前記受信したストリーミングメディアコンテンツを前記クライアントデバイスでキャッシュさせる
複数の命令が格納されていることを特徴とする１つまたは複数のコンピュータ可読媒体。
【請求項５８】
前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツをキャッシュさせる前記命令は、コンテンツのキャッシングが許される旨の指示を前記サー

Aylo Exhibit 1018
Page 1829 of 2308

バデバイスから受信した場合にのみ、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツを保存させる命令を含むことを特徴とする請求項５７に記載の１つまたは複数のコンピュータ可読媒体。
【請求項５９】
前記後の時間は、前記ストリーミングメディアコンテンツすべてを前記サーバデバイスから受信した後の時間を含むことを特徴とする請求項５７に記載の１つまたは複数のコンピュータ可読媒体。
【請求項６０】
クライアントデバイスの１つまたは複数のプロセッサによって実行されると、前記１つまたは複数のプロセッサに、
サーバデバイスからストリーミングメディアコンテンツを受信させ、
前記ストリーミングメディアコンテンツの再生の時間をずらすことを可能にさせる
複数の命令が格納されていることを特徴とする１つまたは複数のコンピュータ可読媒体。
【請求項６１】
前記複数の命令はさらに、前記１つまたは複数のプロセッサに、前記受信したストリーミングメディアコンテンツをキャッシュさせることを特徴とする請求項６０に記載の１つまたは複数のコンピュータ可読媒体。
【請求項６２】
不揮発性記憶装置と、
ストリーミングメディアコンテンツの符号化ビットレートに依存しないレートで前記ストリーミングメディアコンテンツをサーバデバイスから受信し、
前記ストリーミングメディアコンテンツを前記不揮発性記憶装置に記憶し、
前記ストリーミングメディアコンテンツの後の時間の再生を可能にし、
前記ストリーミングメディアコンテンツの時間をずらした再生を可能にし、
受け取ったシャトルコントロールオプションを前記サーバデバイスに指示することなく、
前記シャトルコントロールオプションに従って、前記ストリーミングメディアコンテンツの再生を変更し、
ストリーミングメディアコンテンツの再生を少なくとも一時的に停止する要求を受信し、
前記システムにおける前記ストリーミングメディアコンテンツの再生を停止し、
前記ストリーミングメディアコンテンツの再生を停止させる要求に関わらず、前記ストリーミングメディアコンテンツの受信を継続する
ように構成されたストリーミングメディア再生コンポーネントと
を備えることを特徴とするシステム。
【請求項６３】
データ構造が格納されたコンピュータ可読媒体であって、
データ構造タイプを識別するデータを含む第１の部分と、
１つまたは複数のヘッダを識別するデータを含む第２の部分であって、前記１つまたは複数のヘッダは、ストリーミングメディアコンテンツのクライアントでのキャッシングが、前記ストリーミングメディアコンテンツのソースによってサポートされるかどうかを示すサポートヘッダを含む第２の部分と、
前記データ構造中の前記第２の部分の終わりを識別するデータを含む第３の部分と
を備えることを特徴とするコンピュータ可読媒体。
【請求項６４】
前記サポートヘッダは、ストリーミングメディアコンテンツのクライアントでのキャッシングが前記ソースによってサポートされることを示す機能名を含み、前記機能名は、
ｃｏｍ．ｍｉｃｒｏｓｏｆｔ．ｗｍ．ｆａｓｔｃａｃｈｅ
を含むことを特徴とする請求項６３に記載のコンピュータ可読媒体。
【請求項６５】
データ構造が格納されたコンピュータ可読媒体であって、
データ構造タイプを識別するデータを含む第１の部分と、

キャッシュ制御指示に関連のストリーミングメディアコンテンツをクライアントデバイスがキャッシュできるかどうかを示すキャッシュ制御指示を識別するデータを含む第２の部分と、
前記データ構造中の前記第２の部分の終わりを識別するデータを含む第３の部分と
を備えることを特徴とするコンピュータ可読媒体。
【請求項６６】
前記キャッシュ制御指示はさらに、プロキシデバイスが前記ストリーミングメディアコンテンツをキャッシュできるかどうかを示すことを特徴とする請求項６５に記載のコンピュータ可読媒体。
【請求項６７】
前記キャッシュ制御指示はさらに、ストリーミングメディアコンテンツのキャッシングをサポートしないキャッシングデバイスが前記ストリーミングメディアコンテンツをキャッシュできるかどうかを示すことを特徴とする請求項６５に記載のコンピュータ可読媒体。
【請求項６８】
前記クライアントデバイスがストリーミングメディアコンテンツをキャッシュできることを示すのに、キャッシュ制御指示を使用しないことを特徴とする請求項６５に記載のコンピュータ可読媒体。
【請求項６９】
前記クライアントデバイスのストリーミングメディアプレーヤが前記ストリーミングメディアコンテンツをキャッシュできることを示すキャッシュ制御指示は、
ｎｏ－ｃａｃｈｅ，　ｕｓｅｒ－ｐｕｂｌｉｃ
を含むことを特徴とする請求項６５に記載のコンピュータ可読媒体。
【請求項７０】
前記クライアントデバイスのストリーミングメディアプレーヤが前記ストリーミングメディアコンテンツをキャッシュできないことを示すキャッシュ制御指示は、
ｎｏ－ｕｓｅｒ－ｃａｃｈｅ
を含むことを特徴とする請求項６５に記載のコンピュータ可読媒体。
【請求項７１】
データ構造が格納されたコンピュータ可読媒体であって、
データ構造タイプを識別するデータを含む第１の部分と、
１つまたは複数のヘッダを識別するデータを含む第２の部分であって、前記１つまたは複数のヘッダは、ストリーミングメディアコンテンツをストリーミングする速度を示すストリーミング速度ヘッダを含み、前記ストリーミングメディアコンテンツをストリーミングする速度は、前記ストリーミングメディアコンテンツの符号化ビットレートに依存しない第２の部分と、
前記データ構造中の前記第２の部分の終わりを識別するデータを含む第３の部分と
を備えることを特徴とするコンピュータ可読媒体。
【請求項７２】
前記ストリーミングメディアコンテンツをストリーミングする速度は値ｘによって表され、前記ストリーミング速度ヘッダは、
プラグマ：速度＝ｘ
を含むことを特徴とする請求項７１に記載のコンピュータ可読媒体。
【請求項７３】
前記ストリーミングメディアコンテンツをストリーミングする速度は値ｘによって表され、前記ストリーミング速度ヘッダは、
速度＝ｘ
を含むことを特徴とする請求項７１に記載のコンピュータ可読媒体。
【請求項７４】
前記ストリーミングメディアコンテンツをストリーミングする速度は、前記ストリーミングメディアコンテンツを受信するクライアントデバイスによって要求された速度であるこ

とを特徴とする請求項７１に記載のコンピュータ可読媒体。
【請求項７５】
データ構造が格納されたコンピュータ可読媒体であって，
データ構造タイプを識別するデータを含む第１の部分と，
１つまたは複数のヘッダを識別するデータを含む第２の部分であって、前記１つまたは複数のヘッダは、前記データ構造のメッセージ本体部分に、クライアントデバイスキャッシュからのストリーミングメディアコンテンツの再生に係わるロギング情報が含まれることを示すロギング情報ヘッダを含む第２の部分と，
前記データ構造中の前記第２の部分の終わりを識別するデータを含む第３の部分と
を備えることを特徴とするコンピュータ可読媒体。
【請求項７６】
前記ロギング情報ヘッダは、
Ｃｏｎｔｅｎｔ－Ｔｙｐｅ：ａｐｐｌｉｃａｔｉｏｎ／ｘ－ｗｍｓ－ｓｅｎｄｅｖｅｎｔ
を含むことを特徴とする請求項７５に記載のコンピュータ可読媒体。
【請求項７７】
前記複数の命令が、さらに前記１つまたは複数のプロセッサに、
前記ストリーミングメディアコンテンツに対する要求をユーザから受信させ、前記ユーザの要求に応答して前記サーバデバイスに対して前記ストリーミングメディアコンテンツを要求させる
ことを特徴とする請求項１に記載の１つまたは複数のコンピュータ可読媒体。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、ストリーミングメディアに関し、詳細には、ストリーミングメディアコンテンツのクライアント側でのキャッシングに関する。
【０００２】
【従来の技術】
オーディオ、ビデオ、および／またはテキストメディアコンテンツのストリーミングなどのコンテンツストリーミングが次第に普及しつつある。一般に、用語「ストリーミング」とは、メディアを表すデータがネットワークを通じてクライアントコンピュータに提供され、クライアントコンピュータは、「ファイル」全体が配信されるのを待つのではなく、それをネットワークサーバから受信しながらストリーミングコンテンツをレンダリングすることを意味する。
【０００３】
ストリーミングメディアコンテンツの利用性が高まることにより、以前にはインターネットまたは他のコンピュータネットワークを介しては入手できなかった様々な情報コンテンツの入手が可能になっている。ライブコンテンツは、そのようなコンテンツの顕著な一例である。ストリーミングメディアコンテンツを使用すると、注目すべきイベントのオーディオ、ビデオ、またはオーディオ／ビジュアルの対象を、その進行と同時にインターネットを通じて放送することができる。同様に、テレビ局またはラジオ局は、インターネットを通じて各自のライブコンテンツを伝送することができる。
【０００４】
現在、ストリーミングメディアコンテンツは、「ジャストインタイム（ちょうど間に合う）配信」の概念を使用してストリーミングされている。コンテンツは、クライアントで再生するためにそのコンテンツの符号化ビットレートでクライアントに配信される。ストリームメディアコンテンツには、（例えば、再送する必要のある損失データやその他のネットワークの不整合を考慮に入れて）いくらかのバッファリングを行う。しかし、現在のクライアントシステムは通例、クライアントで行うバッファリング量を最小に抑え、それによりクライアントにおけるメモリ要件を低減し、またバッファリングによる始動待ち時間も短縮することを試みている。

【０００５】
【発明が解決しようとする課題】
しかし、現在のストリーミングメディアコンテンツおよびバッファリングシステムは、他タイプのコンテンツ再生システムに備わる機能を持たないことにより損をしている。例えばＶＣＲ（ビデオカセットレコーダ）で録画したテレビ番組はユーザによって巻き戻すことができ、すでに見た部分を容易に繰り返し見ることができ、あるいはユーザの必要に応じて後の時間に見ることができる。これに対し、現在のストリーミングメディアコンテンツシナリオでは、そのような動作をユーザによって行うことができない。ストリームメディアコンテンツを巻き戻すには、コンテンツのストリーミングを停止し、新しい位置で再開する（そしてしたがってその新しい位置から開始するストリームを再度バッファリングする）必要があり、通例ユーザは、ストリーミングメディアコンテンツを後の時間に見ることができない。すなわち、コンテンツは受信と同時にユーザに対して再生され、ユーザがコンテンツを後の時間に見たい場合はその後の時間にコンテンツをストリーミングしなければならない。
【０００６】
【課題を解決するための手段】
本発明による下記のストリーミングメディアコンテンツのクライアント側でのキャッシングは、上記の課題およびその他の課題を解決することができる。
【０００７】
以下、本発明によるストリーミングメディアコンテンツのクライアント側でのキャッシングについて説明する。
【０００８】
本発明の一実施形態によれば、ストリーミングメディアコンテンツを、コンテンツの符号化ビットレートに依存しないレートでストリーミングすることができる。
【０００９】
本発明の他の実施形態によれば、ユーザが様々なシャトルコントロールオプション（例えば一時停止、停止、早送り、シーク、巻き戻しなど）を選択した場合でも、ストリーミングメディアコンテンツのストリーミングを継続することができる。
【００１０】
本発明のさらに他の実施形態によれば、後の時間に再生するためにストリーミングメディアコンテンツを記録することができる。
【００１１】
本発明のさらに別の実施形態によれば、ストリーミングメディアコンテンツを時間をずらして再生する（ｔｉｍｅ－ｓｈｉｆｔ）ことができる。
【００１２】
追加的な実施形態によれば、サーバデバイスから受信したストリーミングメディアコンテンツをクライアントデバイスにキャッシュ（キャッシュメモリに格納）することができる。そしてクライアントデバイスのストリーミングメディアコンテンツに対する事後の要求は、サーバデバイスからそのストリーミングメディアコンテンツを再度ストリーミングするのではなく、キャッシュされたストリーミングメディアコンテンツを使用して満たすことができる。
【００１３】
本明細書の記述を通じて、同様の構成要素および／または特徴を参照する際は同じ参照符号を使用することとする。
【００１４】
【発明の実施の形態】
図面を参照して、本発明によるストリーミングメディアコンテンツのクライアント側でのキャッシングの実施の形態を以下に詳述する。ストリーミングメディアコンテンツをキャッシュすることにより、従来は利用できなかった様々な機能をユーザが利用できるようにすることができる。そのような機能の例には、コンテンツの符号化ビットレートに依存し

ないレートでメディアコンテンツをストリーミングすることにより、ユーザが各種のシャトル（ｓｈｕｔｔｌｅ）コントロールオプション（例えば一時停止、停止、早送り、シーク、巻き戻しなど）を選択した際でもコンテンツのストリーミングを継続することが可能になる、後の時間に再生するためにストリーミングコンテンツを記録することが可能になる、ストリーミングコンテンツを時間をずらして再生することが可能になることが含まれる。ストリーミングメディアコンテンツのクライアント側でのキャッシングの様々な実施形態は、これらの機能のうち異なる機能、あるいはこれらの機能の異なる機能の組合せを含むことができる。

【００１５】

図１に、ストリーミングメディアコンテンツのクライアント側でのキャッシングを用いることができる例示的なネットワーク環境１００を示す。環境１００では、複数の（ｘ個の）クライアントコンピューティングデバイス１０２（１）、１０２（２）、．．．、１０２（ｘ）が、ネットワーク１０６を介して複数（ｙ個）のオリジンサーバコンピューティングデバイス１０４（１）、１０４（２）、．．．、１０４（ｙ）に結合されている。ネットワーク１０６は、各種の従来のネットワークプロトコル（公開プロトコルおよび／または独自仕様のプロトコルを含む）のうち任意のプロトコルを用いる、各種の従来のネットワークトポロジおよびタイプ（光ネットワーク、配線および／または無線ネットワークを含む）の任意のネットワークを表すものとする。ネットワーク１０６は、例えば、インターネット、また可能性としては１つまたは複数のローカルエリアネットワーク（ＬＡＮ）および／または広域ネットワーク（ＷＡＮ）の少なくとも一部を含むことができる。

【００１６】

オリジンサーバデバイス１０４から利用可能なストリーミングメディアコンテンツを要求する際に、その要求は、クライアントデバイス１０２からネットワーク１０６を介してサーバデバイス１０４に送られる。オリジンサーバデバイス１０４はその要求を受信し、ネットワーク１０６を介して要求元のクライアントデバイス１０２にその要求されたコンテンツを返送する。１つまたは複数のプロキシサーバ（図示せず）がネットワーク１０６の一部であってよく、クライアントデバイス１０２からの要求およびクライアントデバイス１０２への応答は、実際のサーバデバイス１０４ではなく、そのようなプロキシサーバに送信し、そこから受信してもよい。クライアントデバイス１０２にメディアコンテンツをストリーミングするデバイス（オリジンサーバ、プロキシサーバ、あるいはその他のデバイス）はいずれも、そのストリームメディアコンテンツのソースデバイスと呼ばれる。

【００１７】

コンピューティングデバイス１０２および１０４はそれぞれ、各種の従来のコンピューティングデバイスのいずれであってもよく、これにはデスクトップＰＣ（パーソナルコンピュータ、パソコン）、ノートブック型または携帯型コンピュータ、ワークステーション、メインフレームコンピュータ（大型コンピュータ）、インターネット機器、ゲームコンソール、ハンドヘルド（手持ちサイズ）ＰＣ、携帯電話、携帯情報端末（ＰＤＡ）、これらの組合せなどが含まれる。デバイス１０２および１０４の１つまたは複数は、同じタイプのデバイスでも、あるいは異なるタイプのデバイスであってもよい。

【００１８】

サーバデバイス１０４は、各種のデータをクライアント１０２にストリーミング可能な状態にすることができる。用語「ストリーミング」は、メディア（マスコミ媒体）を表すデータがネットワークを通じてクライアントデバイスに提供され、そのコンテンツが完全に配信される前にコンテンツの再生を開始できることを指し示すのに使用される。データは、公開で利用可能としても、あるいは限定にしてもよい（例えば、特定のユーザだけに制限する、適切な料金を支払った場合のみ利用可能とするなど）。データは、オーディオ、ビデオ、テキスト、アニメーションなど、１つまたは複数の各種タイプのコンテンツのいずれであってもよい。さらに、データは「オンデマンド」式（例えば予め記録されており、サイズが分かっているもの）であっても、あるいは「ブロードキャスト」（例えば、演奏と同時に記録され、記録後すぐにストリーミングが可能な状態とされるコンサートのデ

ジタル表現など、予めサイズが分からないもの）してもよい。

【００１９】

図２に、例示的なクライアントデバイスおよびサーバデバイスを示す。クライアントデバイス１０２は、ソースサーバデバイス１４６のストリーミングモジュール１４４にアクセスするように構成されたストリーミングメディアプレーヤ１４２を含む。ソースサーバデバイス１４６は、例えば、図１のオリジンサーバデバイス１０４か、あるいは別のデバイス（例えばプロキシデバイス）である。ソースサーバデバイス１４６はまた（例えばプレーヤ１４２へのユーザ入力に基づいて）メディアプレーヤ１４２によりその中から選択を行うことができる１つまたは複数のストリーミングメディアコンテンツファイル１４８、および選択されてプレーヤ１４２にストリーミングされたコンテンツファイルを含む。ストリーミングメディアコンテンツは、受信されるとすぐに不揮発性の記憶装置１５０に保存することにより、クライアントデバイスでキャッシュされることができる。図２には示されていないが、１つまたは複数の追加的なデバイス（例えばファイアウォール、ルータ、ゲートウェイ、ブリッジ、複数のプロキシサーバなど）が、クライアントデバイス１０２とサーバデバイス１４６の間に位置してもよい。説明を簡潔にするために図２には単一のクライアント１０２とサーバ１４６しか図示していないが、サーバ１４６には複数のクライアント１０２がアクセスすることができ、また単一のクライアント１０２が複数のサーバ１４６にアクセスできることに留意されたい。

【００２０】

不揮発性の記憶装置１５０は、電源障害が生じた際でもその状態を維持するという点で不揮発性である。したがって、不揮発性記憶装置１５０に保存されたストリーミングメディアコンテンツは、クライアント１０２に電力損失が生じた際（例えばクライアント１０２の電源を切った際）でも消えずに存続する。デバイス１５０としては、磁気ディスクおよびディスクドライブ（例えば従来のハードディスク）、光ディスクおよびディスクドライブ、フラッシュメモリなど、各種の異なる不揮発性ストレージのタイプを使用することができる。

【００２１】

デバイス１０２と１４６間の通信は、各種の異なる従来のプロトコルを使用して行うことができる。一実装では、デバイス１０２と１４６間の通信は、バージョン１．１（ＨＴＴＰ１．１）など、あるバージョンのハイパーテキスト伝送プロトコル（ＨＴＴＰ）を使用して行われる。別の実装では、デバイス１０２と１４６間の通信は、リアルタイムストリーミングプロトコル（ＲＴＳＰ）を使用して行われる。あるいは、セッション開始プロトコル（ＳＩＰ）、シンプルオブジェクトアクセスプロトコル（ＳＯＡＰ）など、他のプロトコルを使用してもよい。

【００２２】

また、ストリーミングメディアコンテンツは、各種の異なるストリーミングメディアフォーマットの任意のものに従って記憶し、ストリーミングすることができる。例示的な一実装では、メディアは、ＡＳＦフォーマット（Ａｄｖａｎｃｅｄ Ｓｙｓｔｅｍｓ Ｆｏｒｍａｔあるいは Ａｄｖａｎｃｅｄ Ｓｔｒｅａｍｉｎｇ Ｆｏｒｍａｔ）に基づいてストリーミングされる。ＡＳＦに関するさらなる情報は、ワシントン州レドモンドのマイクロソフト（登録商標）社から得ることができる。これと同じ技術は、ＭＰＥＧ（Ｍｏｖｉｎｇ Ｐｉｃｔｕｒｅｓ Ｅｘｐｅｒｔｓ Ｇｒｏｕｐ）１、ＭＰＥＧ２、ＭＰＥＧ４、Ｑｕｉｃｋｔｉｍｅ（クイックタイム）など、他のフォーマットにも適用することができる。

【００２３】

図３に、ストリーミング媒体データを通信する際に使用できる例示的なメッセージフォーマットを示す。ＨＴＴＰ１．１メッセージやＲＴＳＰメッセージなどのメッセージのデータ構造１８０は、スタートラインフィールドまたは部分１８２、１つまたは複数のヘッダフィールドまたは部分１８４、空のラインフィールドまたは部分１８６、および任意選択のメッセージ本体フィールドまたは部分１８８を含む。スタートライン部分１８２は、メ

ッセージまたはデータ構造タイプを識別するデータを含み、これはリクエストライン（例えばＨＴＴＰ１．１のＧＥＴ要求や、ＲＴＳＰのＧＥＴ＿ＰＡＲＡＭＥＴＥＲ要求など）であっても、ステータスライン（例えばＨＴＴＰ１．１　２００のＯＫ応答やＲＴＳＰ１．０　２００のＯＫ応答など）であってもよい。メッセージに関する情報を表すデータを含む１つまたは複数のヘッダ１８４も含まれる。空ライン部分１８６は、ヘッダ１８４の終わりを識別するのに使用する。メッセージ本体部分１８８には任意選択で追加的なデータを含めることができる。この説明では、各種の指示をヘッダ１８４に含めるが、それらの指示はあるいはメッセージ本体１８８に位置してもよい。

【００２４】

図２のデバイス１０２と１４６間では、データ構造１８０を備えるデータを使用して、制御情報（例えばメディアコンテンツのストリーミングをセットアップする）や、データ（例えばストリーミングメディアコンテンツ）が適宜通信される。したがって、これらのメッセージは、ストリーミングされるメディアコンテンツに対応するか、または関連付けられる。

【００２５】

図２を参照すると、クライアントデバイス１０２のキャッシュマネジャ１５２は、オリジンサーバ１４６からストリーミングメディアプレーヤ１４２によって受信されたストリーミングメディアコンテンツのキャッシングを管理する。ストリーミングメディアプレーヤ１４２は、コンテンツを受信し、そのコンテンツをキャッシュすべきであると判断した際は、そのコンテンツを記憶するためにキャッシュマネジャ１５２に転送する。ストリーミングメディアプレーヤ１４２はこの判定を独自に行っても、あるいはキャッシュマネジャ１５２から受け取る情報（例えばキャッシュに利用可能な空間があるかどうかなど）に部分的に基づいて判定してもよい。

【００２６】

ストリーミングメディアプレーヤ１４２によるストリーミングメディアコンテンツのキャッシングは、ストリーミングメディアコンテンツのバッファリングとは異なることに留意されたい。キャッシングとバッファリングには様々な違いがあるが、通例その違いには次の１つまたは複数が含まれる：キャッシュされたコンテンツは、揮発性メモリではなく不揮発性の記憶装置に記憶される；キャッシュされたコンテンツは、すでに再生されたコンテンツを含むことができる；キャッシュされたコンテンツは、（例えば「ジャストインタイム」配信のように）直ちに再生されるのではなく、長時間再生されないコンテンツを含むことができる。

【００２７】

図２には異なるストリーミングメディアコンテンツを異なるファイル１４８として示しているが、これに代えて、１つのストリーミングメディアコンテンツを複数のファイルとして記憶してもよい（あるいはブロードキャストコンテンツの場合はファイルがなくともよい）。「１つの」コンテンツを定義する方式は、実装と媒体のタイプによって異なる可能性がある。例えば、音楽を扱うオーディオおよび／またはビデオコンテンツの場合は、１つの歌を１つのコンテンツとすることができる。コンテンツは、固有の境界（例えば異なる歌）に沿って複数部分に区分するか、あるいはその他の任意の方式で（例えばコンテンツを５分ごとに１コンテンツとするなど）区分することができる。

【００２８】

クライアントデバイス１０２の一部である従来のコンポーネントを任意選択で強化して、キャッシュ管理機能を行わせることができる。例示的な一実装では、マイクロソフト（登録商標）社のインターネットエクスプローラブラウザプログラムがキャッシュ管理機能を含み、ストリーミングメディアプレーヤ１４２はこの機能を強化し、使用することができる。

【００２９】

各メディアコンテンツは、ともに単一のファイルとして記憶することができるにも関わらず複数のストリームを含むことができる。そのようなストリームはそれぞれ、特定タイプ

Aylo Exhibit 1018
Page 1836 of 2308

のメディア（例えばオーディオ、ビデオ、テキストなど）を通例は特定の符号化ビットレートで表す。符号化ビットレートは、再生のためにメディアコンテンツを符号化するレート（処理速度）である。符号化ビットレートは、ユーザによって知覚されるコンテンツの再生速度とは無関係であることに留意されたい（例えば、通例は、通常のコンテンツストリームと、ユーザがその通常ストリームの再生速度の２倍と知覚する早送りのコンテンツストリームは、どちらも同じ符号化ビットレートである）。

【００３０】

同一タイプのメディアの複数のバージョン（例えば複数のオーディオバージョン、複数のビデオバージョンなど）をメディアコンテンツに含め、メディアプレーヤ１４２による再生のためにそれらのストリームの異なる組合せを選択できるようにすることができる。これらの異なるバージョンはそれぞれ、通例は異なるビットレートで符号化される（通例はビットレートが高いほどコンテンツの品質が高くなる）。ストリームのどの組合せをストリーミングメディアコンテンツに含めるかは各種の方式で選択することができ、これには、ユーザの嗜好（例えばより高い品質のコンテンツを望むか、あるいはより低い品質のコンテンツを望むか）、利用可能なネットワーク帯域幅、特定コンテンツに望まれるビットレート（例えば、あるビットレートをコンテンツ作成者または配信者が設定し、ストリーミングメディアコンテンツの識別子に含めて、ユーザがその識別子を選択するとストリーミングメディアプレーヤ１４２が（任意選択でユーザの嗜好に基づいて）その求められるビットレートに可能な限り近いストリームを選択するなど）などがある。

【００３１】

コンテンツをキャッシュする際は、キャッシュマネジャ１５２が、ストリーミングメディアコンテンツとしてサーバデバイス１４６から受信した（ストリーミングメディアプレーヤ１４２によって要求された）特定のストリームを不揮発性記憶装置１５０に格納する。同じメディアコンテンツについての異なるストリームの組合せをキャッシュマネジャ１５２によってキャッシュすることができる。あるいは、キャッシュマネジャ１５２は、特定のメディアコンテンツについてのすべてのストリームをサーバデバイス１４６から取得し、そのストリームすべてをキャッシュすることが可能だが、メディアプレーヤ１４２へは要求されたストリームだけを再生することができる。

【００３２】

複数のコンテンツをいっしょに再生リストにまとめることもでき、この再生リストは、それぞれがストリーミングされる特定のコンテンツである，１つまたは複数の項目からなるリストである。これらの異なるコンテンツを（例えばユーザまたはその他の者によって）選択していっしょに再生リストにまとめ、ユーザが単に再生リストを選択するだけでレンダリングのためにそれらのコンテンツすべてを選択できるようにすることができる。例として、ユーザは自分の好きな歌から２０曲を選んで再生リストの一部とし、その後は再生リストの再生を選択することによりそれらの歌を再生させることができる。

【００３３】

クライアント１０２でストリーミングメディアコンテンツをキャッシュすることにより、ストリーミングメディアプレーヤ１４２が各種の機能、およびしたがってユーザ体験の機能強化を提供することが可能になる。これらには、コンテンツの符号化ビットレートに依存しないレートでメディアコンテンツをストリーミングすることにより、ユーザが各種のシャトルコントロールオプション（例えば一時停止、停止、早送り、シーク、巻き戻しなど）を選択した際でもコンテンツのストリーミングを継続することが可能になる、後の時間に再生するためにストリーミングコンテンツを記録することが可能になる，時間をずらしてストリーミングコンテンツを再生することが可能になることが含まれる。

【００３４】

クライアント１０２でストリーミングメディアコンテンツをキャッシュすることにより、符号化ビットレートに依存しないレートでサーバデバイス１４６からクライアントデバイス１０２にメディアコンテンツをストリーミングすることが可能になる。ストリーミングレートが符号化ビットレートと同じになる状況が生じる可能性があるが、ストリーミング

レートは、符号化ビットレートよりも高く、または低くすることができる。ストリーミングメディアプレーヤ１４２によって受信されたストリーミングメディアコンテンツはキャッシュメモリに記憶され、コンテンツを再生する際は、メディアプレーヤ１４２は、そのコンテンツを再生のためにキャッシュメモリから取り出し、それと同時にサーバデバイス１４６からストリーミングされてくるコンテンツの続きの部分をキャッシュメモリに入れ続ける。ストリーミングレートが符号化ビットレートを下回る例示的な状況では、メディアプレーヤ１４２は、キャッシュされたコンテンツの量、符号化ビットレート、およびストリーミングレートに基づいて、ストリーミングされるコンテンツの再生に補足的な遅延を課す。例えば符号化ビットレートが３００ｋｂｐｓでストリーミングレートが１５０ｋｂｐｓである場合、メディアプレーヤ１４２は、十分なコンテンツが記憶され、ユーザが気づくような中断を生じずにコンテンツを始めから終わりまで完全に再生できるようになるまでコンテンツ再生の開始を遅らせる。

【００３５】

一実装では、メディアプレーヤ１４２は、ストリーミングモジュール１４４にストリーミングコンテンツを要求する際に、ストリーミングモジュール１４４に速度の要求も伝達する。この速度要求は、メディアプレーヤ１４２がメディアコンテンツをストリーミングさせたいレート（速度）を指示する。ストリーミングモジュール１４４は要求されたレート（速度）を受け入れ、続いてその要求されたレート（速度）でメディアコンテンツをメディアプレーヤ１４２にストリーミングするか、あるいはそのレート（速度）を変更することができる。例えば、ストリーミングモジュール１４４は、要求されたレートでコンテンツをストリーミングできない場合がある（例えばサーバポリシー（サーバの方針）、サーバのハードウェアおよび／またはソフトウェアの限界、サーバの現在の負荷などの原因による）。ストリーミングモジュール１４４は、どれだけのストリーミングレートになるかについての指示をメディアプレーヤ１４２に送信する（あるいは、ストリーミングモジュール１４４が特に指示しない場合は、ストリーミングレートはメディアプレーヤ１４２が要求するレートでよい）。さらに、ストリーミングモジュール１４４は、（新たな速度要求をストリーミングモジュール１４４に送信することにより）コンテンツのストリーミング中にストリーミングレートの変更を試みることができ、ストリーミングモジュール１４４は、（新たな速度指示をメディアプレーヤ１４２に送信することにより）コンテンツのストリーミング中にストリーミングレートを変更することができる。

【００３６】

メディアプレーヤ１４２は、自身が求めるストリーミングレート、すなわちそれがストリーミングモジュール１４４に要求するレートを各種の異なる要因に基づいて決定することができる。この決定は、コンテンツ単位で行うことができる。一実装では、メディアプレーヤ１４２は、メディアプレーヤ１４２とストリーミングモジュール１４４間の現在利用可能な帯域幅の割り出しを試みる。これは各種の従来の方法で割り出しすることができ、この方法には、デバイス１０２と１４６間でテストメッセージを送信する、デバイス１０２と１４６間の接続の現在および過去の振る舞いを監視する、ストリーミングモジュール１４４から利用可能な帯域幅の指示を受け取るなどがある。現在利用可能な帯域幅を知ると、メディアプレーヤ１４２は、まず現在利用可能な帯域幅よりも小さい特定の量であるストリーミングレートを要求する。この特定の量は、固定された量（例えば常に５０ｋｂｐｓ）でも、動的な量（例えば現在利用可能な帯域幅の１５％、あるいは現在利用可能な帯域幅の５％～２５％の間など）であってもよい。

【００３７】

メディアプレーヤ１４２からストリーミングモジュール１４４に送信されるストリーミング速度要求は、メディアプレーヤ１４２がいろいろな任意の方法のいずれかでデータをストリーミングさせたいレートを示すことができる。一実装では、このレートは、コンテンツの符号化ビットレートに比例する速度係数で表される（例えば、３．２の速度係数は、ストリーミングレートがコンテンツの符号化ビットレートに比べて３．２倍速いことを意味し、０．５の速度係数は、ストリーミングレートがコンテンツの符号化ビットレートの

半分であることを意味する）。別の実装では、速度は、単に所望のストリーミングビットレート（例えば３００ｋｂｐｓ、５００ｋｂｐｓ、２０ｋｂｐｓなど）を指定することによって示される。

【００３８】

速度要求の指示は、各種の異なる方式でメディアプレーヤ１４２からストリーミングモジュール１４４に送信することができる。ＲＴＳＰを使用してデバイス１０２と１４６間の通信を行う一実装では、ストリーミングメディアプレーヤ１４２からストリーミングモジュール１４４に送信されるメッセージに次のＲＴＳＰヘッダを（例えば図３のヘッダ１８４として）付加する。

速度：ｘ

ｘは、コンテンツの符号化ビットレートに比例する要求された速度係数を表す。値ｘは、整数でも実数でもよい（任意選択で小数点以下の桁数を制限することができ、通例は小数点第１位または第２位までに制限するが、他の値を使用することもできる）。このヘッダは様々なメッセージに付加することができ、一実装では、ストリーミングメディアプレーヤ１４２からストリーミングモジュール１４４に送信される、メディアコンテンツのストリーミングを要求するメッセージに含める。

【００３９】

ストリーミングモジュール１４４がその要求に応答する際には、ＲＴＳＰ　ＰＬＡＹ応答メッセージに類似のヘッダ（速度：ｙ）を入れてストリーミングメディアプレーヤ１４２に返送する。ｙは、ストリーミングモジュール１４４が要求されたコンテンツをストリーミングする（コンテンツの符号化ビットレートに比例する）速度を表す（ｙは、値ｘと同じでも同じでなくともよい）。値ｘと同様に、値ｙは、整数または実数とすることができる（小数点以下の桁数は任意選択で制限してもよく、通例は小数点第１位または第２位までに制限するが、他の値も使用することができる）。

【００４０】

ＨＴＴＰを使用してデバイス１０２と１４６間の通信を行う別の一実装では、ストリーミングメディアプレーヤ１４２からストリーミングモジュール１４４に送信されるメッセージに（例えば図３のヘッダ１８４として）次のＨＴＴＰヘッダを付加する。

プラグマ：速度＝ｘ

ｘは、コンテンツの符号化ビットレートに比例する要求された速度係数を表す。プラグマ（ｐｒｏｇｍａ）ヘッダは、実装に固有の指示を含めるためにＨＴＴＰで使用される一般ヘッダフィールドである。上述のＲＴＳＰスピードヘッダと同様に、このヘッダは様々なメッセージに付加することができ、一実装では、ストリーミングメディアプレーヤ１４２からストリーミングモジュール１４４に送信される、メディアコンテンツのストリーミングを要求するメッセージに含められる。

【００４１】

ストリーミングモジュール１４４は、この要求に応答する際には、ＨＴＴＰ　ＧＥＴ応答メッセージに同様のヘッダ（プラグマ：速度＝ｙ）を入れてストリーミングメディアプレーヤ１４２に返送する。ｙは、ストリーミングモジュール１４４が要求のコンテンツをストリーミングする（コンテンツの符号化ビットレートに比例する）速度を表す（ｙは、値ｘと同じでも同じでなくともよい）。値ｘと同様に、値ｙは整数でも実数でもよい（小数点以下の桁数を任意選択で制限することができ、通例は小数点第１位または第２位までに制限するが、他の値を使用してもよい）。

【００４２】

コンテンツのストリーミングレートは、符号化ビットレートを参照して識別することができるが、この識別は単にそれが容易な基準点であるために行うことに留意されたい。ストリーミングレートはなお符号化ビットレートには依存せず、実質的にどのような速度係数でも要求および／または選択することができ、その速度係数を識別する方式は符号化ビットレートを基準とする。

【００４３】

Aylo Exhibit 1018
Page 1839 of 2308

図４に、サーバからクライアントにメディアコンテンツをストリーミングする例示的プロセス２００を表すフローチャートを示す。プロセス２００は、図２のデバイス１０２および１４６などのクライアントデバイスおよびサーバデバイスによって実施され、またソフトウェア、ファームウェア、ハードウェア、またはそれらの組合せで行うことができる。

【００４４】

初めに、コンテンツの符号化ビットレートに依存しない、メディアコンテンツをストリーミングするときのストリーミングレートをクライアントデバイスとサーバデバイス間でネゴシエートする（動作２０２）。このネゴシエーションは、例えば、上述のようにクライアントによる要求速度の提出と、ストリーミング速度の戻し通知とすることができる。あるいは、このネゴシエーションは他の形を取ることもでき、サーバデバイス（または何らかの他のデバイス）によって開始し、クライアントデバイスがそのストリーミングレートに対して所望の変更を要求する、クライアントデバイスまたはサーバデバイスによって変更可能なデフォルトのレートを使用する、ストリーミングメディアコンテンツの識別子によって示され、クライアントデバイスまたはサーバデバイスによって変更することが可能なレートを使用する、などがある。ストリーミングレートをネゴシエートすると、ネゴシエートされたストリーミングレートでサーバデバイスからクライアントデバイスにメディアコンテンツをストリーミングする（動作２０４）。ストリーミングレートは、任意選択であって、後にコンテンツのストリーミング中に再度ネゴシエートすることができる（これを開始するのはクライアントデバイス、サーバデバイス、あるいは何らかの他のデバイスである）。

【００４５】

図２に戻ると、通例、ストリーミングモジュール１４４は、ストリーミングメディアプレーヤ１４２に対してコンテンツファイル１４８全体をストリーミングメディアコンテンツとして転送することはない。通例は、ファイルヘッダや索引付け情報などの特定の情報は、ストリーミングメディアコンテンツの一部としてストリーミングメディアプレーヤ１４２に対して転送されない。この情報の厳密な性質は、ストリーミングモジュール１４４の設計方式と、メディアコンテンツのストリーミングに使用するフォーマットとに基づいて異なる可能性がある。そのような情報がクライアントデバイス１０２に転送されない場合は、ストリーミングメディアプレーヤ１４２（あるいはキャッシュマネジャ１５２）がその情報を生成し、その情報をキャッシュ１５０に追加する。したがってこの適切な情報はキャッシュ内コンテンツの一部として取得可能になり、キャッシュからストリーミングメディアコンテンツを再生する際にストリーミングメディアプレーヤ１４２によって使用することができる。

【００４６】

ストリーミングメディアプレーヤ１４２は、ストリーミングメディアコンテンツの再生要求を受け取ると、キャッシュマネジャ１５２を介して、要求されたコンテンツがキャッシュ１５０内にあるかどうかを調べる。コンテンツがキャッシュ１５０にある場合（これを「キャッシュヒット」と呼ぶ）、ストリーミングメディアプレーヤ１４２は、サーバデバイス１４６ではなくキャッシュ１５０から要求されたストリーミングメディアコンテンツを取得する。ただし、コンテンツがキャッシュ１５０にない場合（これを「キャッシュミス」と呼ぶ）、ストリーミングメディアプレーヤ１４２は、要求されたストリーミングメディアコンテンツをサーバデバイス１４６から取得する。キャッシュが一杯になり、それ以上のストリーミングメディアコンテンツを入れるだけの記憶空間がなくなった場合には、新たなコンテンツを入れるスペースを作るためにすでにキャッシュされているコンテンツをキャッシュから削除（これをしばしばキャッシュから追い出すという）できることに留意されたい。各種ある従来のキャッシュ置換アルゴリズムを使用して、キャッシュから追い出すコンテンツを決定することができる。

【００４７】

状況によっては、コンテンツファイル１４８に、符号化ビットレートが異なる複数の異なるストリーム（例えば、５６ｋｂｐｓで符号化された低品質のストリームと、３００ｋｂ

ｐｓで符号化された高品質のストリームなど）が含まれる可能性がある。このようなファイルをしばしばマルチプルビットレートファイルと呼ぶ。そのような状況では、通例は１つのストリーム（例えば低品質のストリームまたは高品質のストリーム）だけをクライアントデバイス１０２にストリーミングし、そのストリームをキャッシュ１５０にキャッシュイングする。それと符号化ビットレートが同じである同一コンテンツに対する事後の要求は、キャッシュ１５０から満たすことができる（すなわちキャッシュヒットになる）が、別の符号化ビットレートに対する要求はキャッシュ１５０からは満たすことができない（キャッシュミスとなる）。あるいは、ストリーミングメディアプレーヤ１４２が、ストリーミングモジュール１４４からストリーミングされた符号化ビットレートをすべて備え、それによりそれらのレートすべてをキャッシュ１５０に記憶しておいてもよい。

【００４８】

ストリーミングモジュール１４４は任意選択で、ストリーミングメディアコンテンツの満了時間（有効期限）を含んでもよい。この満了時間は、比例する時間（例えばコンテンツを送信してから５分後）であっても、固定された時間（例えば特定の日時）であってもよい。ストリーミングメディアコンテンツの満了時間は、各種の異なる方式で示すことができる。一実装では、ＨＴＴＰのｍａｘ－ａｇｅキャッシュ制御指示を使用して満了時間を指示する。別の実装では、ＨＴＴＰまたはＲＴＳＰのＥｘｐｉｒｅｓヘッダを使用して満了時間を指示する。

【００４９】

期限が満了すると、ストリーミングメディアプレーヤ１４２は、キャッシュからコンテンツを再生する前にコンテンツを再確認する。この再確認は、通例、ストリーミングメディアプレーヤ１４２でコンテンツ要求（例えばコンテンツの再生を求めるユーザの要求）を受信したときに行っても、あるいは他のときに行ってもよい（例えば、そのコンテンツが要求されているかどうかに関係なく、ストリーミングメディアプレーヤ１４２またはキャッシュマネジャ１５２がキャッシュを監視して、満了時間が経過するとその期限切れを検出するということが可能である）。コンテンツの期限が満了すると、ストリーミングメディアプレーヤ１４２は、そのコンテンツを記述する新しい情報（それによりストリーミングメディアプレーヤ１４２はそのコンテンツに変化があるかどうかを判定することができる）と、新しい満了時間を回復する。

【００５０】

コンテンツが変化していない場合、ストリーミングメディアプレーヤ１４２は、単にそのコンテンツの満了時間を新しい期限日に更新することができ、キャッシュ内のコンテンツには変更を行う必要がない。ただし、コンテンツが変化している場合は、ストリーミングメディアプレーヤ１４２は、そのコンテンツをキャッシュから取得するのではなく、要求されたそのコンテンツをストリーミングモジュール１４４からストリーミングする。あるいは、ストリーミングメディアプレーヤ１４２は、コンテンツの何が変化したのかを判断し、変化した部分だけをストリーミングモジュール１４４からストリーミングし、残りの部分はキャッシュから得ることを試みることもできる。

【００５１】

ストリーミングメディアプレーヤ１４２は、記述オペレーションを使用して、サーバデバイス１４６から取得できる特定のコンテンツを記述した情報を得ることができる。ストリーミングモジュール１４４は、記述要求に応答して、ヘッダ情報（例えば、コンテンツについての書誌学的な（ｂｉｂｌｉｏｇｒａｐｈｉｃ）情報、コンテンツのストリームの特徴、そのコンテンツのインデックスを再生成するのに必要な情報などを含む）、コンテンツに関連のメタデータ、コンテンツの一部として取得可能なストリームのリストなど、識別されるメディアコンテンツを記述する各種の情報を返送する。返送された正確な情報は、ストリーミングモジュール１４４の実装方式と、メディアコンテンツをストリーミングするのに使用する特定のフォーマットとに応じて異なる可能性がある。この情報を使用してコンテンツを再確認することができる。

【００５２】

Aylo Exhibit 1018
Page 1841 of 2308

一実装では、特定のコンテンツストリームのヘッダ情報は、コンテンツが変更された場合にコンテンツ作成者によって変更されるＧＵＩＤ（Ｇｌｏｂａｌｌｙ Ｕｎｉｑｕｅ ＩＤ：グローバル一意識別子）を含む。この値は、「ＭＭＳタイプ」属性として知られる。コンテンツの変更はこのＧＵＩＤ値を調べることによって検出することができ、値が変化していなければコンテンツは変化しておらず、ＧＵＩＤ値が変化していればコンテンツが変化していることになる。

【００５３】

あるいは、別の情報を調べてコンテンツが変化しているかどうかを判定することができる。一実装では、コンテンツ記述情報は、最後に修正された日付の指示を含む（例えば図３のヘッダ１８４とすることができる）。新たに受け取った最後に修正された日付が、以前に受け取った最後に修正された日付と異なる場合、ストリーミングメディアプレーヤ１４２は、コンテンツが変化していると判断する。調べることが可能なこの他の情報は、コンテンツのハッシュ（ｈａｓｈ）である。従来の各種のハッシングアルゴリズムを使用してコンテンツのハッシュ値を生成することができ、一般にはコンテンツを変更するとそのコンテンツのハッシュ値が変化する。ハッシュ値が変化している場合、ストリーミングメディアプレーヤ１４２は、コンテンツが変更されたと判断する。

【００５４】

上述のように、ストリーミングメディアコンテンツは、複数コンテンツからなる再生リストとすることができる。メディアコンテンツの再生リストがキャッシュされている場合、ストリーミングメディアプレーヤ１４２は、再生リストの各メディアコンテンツを必要に応じて再確認する。キャッシュされた時点から変化しているコンテンツは、キャッシュから取得するのではなく、ストリーミングモジュール１４４からストリーミングする。また、再生リストのメディアコンテンツのすべてがクライアントデバイス１０２でキャッシュされてはいない状況が生じる場合がある。そのような状況では、キャッシュされていないコンテンツは、キャッシュから取得するのではなくストリーミングモジュール１４４からストリーミングする。

【００５５】

また、再生リストに関しては、ストリーミングメディアプレーヤ１４２が、再生リスト中の次のコンテンツについての準備が整ったことをストリーミングモジュール１４４に指示する状況が生じる可能性がある。この指示は、「次のエントリのストリーム（ｓｔｒｅａｍ ｎｅｘｔ ｅｎｔｒｙ）」のコマンドをストリーミングモジュール１４４に通信することによって、あるいは、単にそのコンテンツがキャッシュから取得できる場合でも、現在のコンテンツをストリーミングモジュール１４４からストリーミングさせることによって、行うことができる（ストリーミングモジュール１４４は、現在のコンテンツの最後に達すると次のコンテンツのストリーミングを開始する時間であることを知る）。

【００５６】

ストリーミングメディアコンテンツをキャッシングすることにより、ユーザが各種のシャトルコントロールオプション（例えば停止、一時停止、早送り、シーク、巻き戻しなど）を選択した場合でもコンテンツのストリーミングを継続することも可能になる。ストリーミングメディアプレーヤ１４２は、プレーヤ１４２のユーザにシャトルコントロールオプションが利用できるようにし、それによりユーザが必要に応じてストリーミングモジュール１４４の再生を制御できるようにする。ユーザが特定のシャトルコントロールコマンドを入力した際、ストリーミングメディアプレーヤ１４２の振る舞いは、入力されたコマンドに応じて異なる。ただし、多くの場合、ストリーミングメディアプレーヤ１４２は、ストリーミングモジュール１４４からのメディアコンテンツのストリーミングを継続する。

【００５７】

ユーザが停止コマンドを入力した場合、ストリーミングメディアプレーヤ１４２は、ストリーミングメディアコンテンツの再生を中止するが、ストリーミングモジュール１４４からのコンテンツのストリーミングとキャッシュを継続する。同様に、ユーザが一時停止コマンドを入力した場合、ストリーミングメディアプレーヤ１４２は、ストリーミングメデ

Aylo Exhibit 1018
Page 1842 of 2308

ィアコンテンツの再生を一時停止するが、ストリーミングモジュール１４４からのコンテンツのストリーミングとキャッシングを継続する。さらに、ユーザが巻き戻しコマンドを入力した場合は、ストリーミングメディアプレーヤ１４２は、ストリーミングメディアコンテンツの巻き戻しを開始するが、引き続きストリーミングモジュール１４４からコンテンツをストリーミングし、キャッシュする。さらに、ユーザが早送りコマンドを入力した場合は、ストリーミングメディアプレーヤ１４２は、すでにキャッシュされたコンテンツ部分を通じてストリーミングメディアコンテンツの早送りを開始し、必要な範囲までコンテンツをストリーミングし、キャッシュし続ける。

【００５８】

図５に、サーバからクライアントにメディアコンテンツをストリーミングし、同時にシャトルコントロールコマンドを受け取る例示的プロセス２３０を説明するフローチャートを示す。プロセス２３０は、図２のクライアントデバイス１０２のストリーミングメディアプレーヤ１４２など、クライアントデバイスのストリーミングメディアプレーヤによって実施され、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組合せで実行することができる。

【００５９】

初めに、ストリーミングメディアコンテンツの再生に係わるシャトルコントロール要求をストリーミングメディアプレーヤで受け取る（動作２３２）。ストリーミングメディアプレーヤは、要求に従ってストリーミングメディアコンテンツの再生を変える（動作２３４）が、ストリーミングメディアコンテンツの受信を継続する（動作２３６）。ストリーミングメディアプレーヤは、ユーザアクションを詳述する以後のログ提出を除いてはシャトルコントロールコマンドをサーバデバイスに知らせる必要はなく、通例は指示しないことに留意されたい。

【００６０】

また、ユーザがシャトルコントロールコマンドを使用してストリーミングメディアコンテンツの再生をナビゲート（操縦）または制御する能力は、特定のコンテンツ（あるいは特定コンテンツの一部）に対しては制限されることもある。例えば、ユーザは、ストリーミングメディアコンテンツの特定の部分（例えば広告）では早送りできない場合がある。そのような制限の指示は、ストリーミングモジュール１４４からストリーミングメディアプレーヤによって受信され、ストリーミングメディアコンテンツとともにキャッシュされる。したがって、キャッシュからストリーミングメディアコンテンツを再生する際には、そのような制限も実施される。

【００６１】

ストリーミングメディアコンテンツをキャッシュすることにより、さらに、後の時間に再生するためにコンテンツを記録することが可能になる。コンテンツの少なくとも一部は、コンテンツがストリーミングされ、キャッシュされるのと同時にユーザに対して再生することができ、あるいは、コンテンツのどの部分も後の時間まで再生しないことも可能である（例えば、コンテンツのキャッシングが完了してからしばらく経ってから再生するなど）。例えば、ユーザは、複数のストリーミングメディアコンテンツをキャッシュする要求をストリーミングメディアプレーヤ１４２に入力し、後の時間（例えば同じ日の後の時間、翌日、翌週、翌月など）にそれらのコンテンツを再生することができる。そのような要求は、例えば具体的な「コンテンツのキャッシュ」要求であるか、あるいは単に再生要求である。ストリーミングメディアプレーヤ１４２は、そのような要求を受信すると、ストリーミングモジュール１４４と通信して、ストリーミングメディアプレーヤ１４２がユーザのためにそのコンテンツを再生する場合と同じ方式で、要求されたコンテンツをストリーミングさせる。ただし、コンテンツをクライアントデバイス１０２で受信した場合は、ストリーミングメディアプレーヤ１４２は、そのコンテンツを不揮発性記憶装置１５０にキャッシュし、（ユーザから要求されない限りは）そのコンテンツをその時間には再生しない。ユーザが後にそのコンテンツの再生を要求すると、ストリーミングメディアプレーヤ１４２は、そのコンテンツをキャッシュメモリから入手する（ただし、コンテンツの期

Aylo Exhibit 1018
Page 1843 of 2308

限が切れていないことを前提とする。または、コンテンツの期限が切れている場合はコンテンツが変更されていないことを前提とする）。

【００６２】

ストリーミングメディアコンテンツをキャッシュすることにより、さらに、時間をずらしたストリーミングメディアコンテンツの再生が可能になる。時間をずらした再生では、ユーザは、ストリーミングと同時にストリーミングコンテンツをナビゲートし、制御することができる（例えば一時停止、巻き戻し、早送りなどのコマンドの入力により）。例えば、ユーザはストリーミングメディアコンテンツの再生を特定の位置で一時停止し、数分後にその位置から再生を再開することができ、再生が一時停止されていた間にさらなるコンテンツがクライアントデバイス１０２にストリーミングされていても、再生はなおユーザが再生を一時停止した位置から再開する。別の例として、ユーザは、コンテンツのある部分をもう一度（あるいはさらに複数回）見るために、ストリーミング中にストリーミングメディアコンテンツを巻き戻すことができる。ユーザが例えば３分だけコンテンツを巻き戻した場合、ユーザはその３分間をもう一度見ることができ、その３分間分先行するコンテンツが再度再生されていた間にさらなるコンテンツがクライアントデバイス１０２にストリーミングされていても、再生は、ユーザが巻き戻しコマンドを入力したコンテンツ中の時点から（再生にユーザが気づくような中断を生じずに）継続する。

【００６３】

図６に、後に再生するためにストリーミングメディアコンテンツを記録する例示的プロセス２６０を説明するフローチャートを示す。プロセス２６０は、図２のクライアントデバイス１０２のストリーミングメディアプレーヤ１４２など、クライアントデバイスのストリーミングメディアプレーヤによって実行され、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組合せで行うことができる。

【００６４】

初めに、サーバからストリーミングメディアコンテンツを受信する（動作２６２）。受信したストリーミングメディアコンテンツは、クライアントデバイスのキャッシュにローカルに記憶する（動作２６４）。ストリーミングメディアコンテンツを再生する要求を後の時間に受け取る（動作２６６）が、これはメディアコンテンツのストリーミングが完了した後でも、完了しない時点でもよい。再生要求に応答して、ローカルに保存されたストリーミングメディアコンテンツを再生する（動作２６８）。

【００６５】

また、ストリーミングメディアプレーヤ１４２がストリーミングメディアコンテンツを再生する際には、通例、ストリーミングメディアプレーヤ１４２からストリーミングモジュール１４４に各種のロギング情報が返送される。ストリーミングメディアプレーヤ１４２がストリーミングモジュール１４４ではなくクライアントデバイス１０２のキャッシュからストリーミングメディアコンテンツを取得する状況でも、このロギング情報をストリーミングモジュール１４４に返送しなければならない。ロギング情報は、ストリーミングメディアプレーヤ１４２がコンテンツの再生に費やした時間量、複数回再生されたコンテンツ部分、スキップされたコンテンツ部分、コンテンツの再生が一時停止されたか、もしそうであればコンテンツ中のどの点で一時停止が行われたか、ストリームコンテンツを受信したネットワーク接続に伴う問題など、各種の情報を含むことができる。

【００６６】

ロギング情報は、各種の異なる方式でサーバデバイス１４６に通信することができる。一実装では、通例のヘッダを使用してメッセージを送信して、メッセージ本体（図３の本体１８８）にロギング情報が含まれることを知らせる。

Ｃｏｎｔｅｎｔ－Ｔｙｐｅ：ａｐｐｌｉｃａｔｉｏｎ／ｘ－ｗｍｓ－ｓｅｎｄｅｖｅｎｔ

【００６７】

ただし、この代わりに異なるヘッダを送信してもよい。このメッセージは、例えば、ＲＴＳＰのＳＥＴ＿ＰＡＲＡＭＥＴＥＲメッセージ、またはＨＴＴＰのＰＯＳＴメッセージなどである。

Aylo Exhibit 1018
Page 1844 of 2308

【００６８】
すべてのストリーミングメディアコンテンツがクライアントデバイス１０２でキャッシュ可能であるとは限らないことに留意されたい。ストリーミングメディアコンテンツをクライアントデバイス１０２でキャッシュしない様々な状況がありうる。こうした状況には、例えば、サーバデバイスがキャッシングをサポートしない状況、コンテンツ作成者または配信者が特定のコンテンツをキャッシュすべきでないことを知らせている状況、クライアントデバイス１０２のユーザによってキャッシングが使用禁止にされている状況などがある。さらに、一部の実施形態では、ブロードキャストコンテンツなど特定のコンテンツはキャッシュするべきでないとしている（他の実施形態ではブロードキャストコンテンツをキャッシュすることができる）。ストリーミングメディアプレーヤ１４２は、特定のコンテンツをキャッシュすべきでないと判定した場合、そのコンテンツをキャッシュしない。

【００６９】
一実装では、ストリーミングメディアプレーヤ１４２がストリーミングモジュール１４４からコンテンツをストリーミングするように要求すると、ストリーミングモジュール１４４は、要求されたコンテンツのキャッシングがストリーミングモジュール１４４によってサポートされるかどうかの通知を返送する。この指示はメッセージヘッダ（例えば図３のヘッダ１８４）に含めることができ、「サポート（ｓｕｐｐｏｒｔｅｄ：支援）」ヘッダと称する。サポートヘッダ中の指示の一例を次に示す。
ｃｏｍ．ｍｉｃｒｏｓｏｆｔ．ｗｍ．ｆａｓｔｃａｃｈｅ

【００７０】
ただし、これに代えて他のヘッダを使用することもできる。このサポートヘッダがあるということは、ストリーミングモジュール１４４がクライアントデバイス１０２のキャッシングをサポート（支援）するということを示しており、それでストリーミングモジュール１４４は、例えば、メディアコンテンツを符号化ビットレートに依存しないレートでストリーミングし、キャッシュされたコンテンツについてのロギング情報を受信することなどを可能にする。このサポートヘッダがない場合は、ストリーミングメディアプレーヤ１４２は、メッセージに関連のストリーミングメディアコンテンツをキャッシュしない。

【００７１】
さらに、特定の部分のストリーミングメディアコンテンツのキャッシングを制御するのに、各種のキャッシュ制御指示を使用することができる。キャッシュ制御ヘッダは、キャッシュマネジャ１５２（また、キャッシュメモリを含むプロキシサーバなどの任意の仲介デバイス）が、関連のストリーミングメディアコンテンツのキャッシングをどのように処理するかを記述する情報を含む。キャッシュ制御ヘッダは、そのメッセージを受け取るすべてのキャッシング機構が従うべき指示を指定するのに使用される。ここでは述べないが、追加的なキャッシュ制御指示（例えば従来のＨＴＴＰまたはＲＴＳＰのキャッシュ制御指示）をキャッシュ制御ヘッダに含めてもよい。

【００７２】
使用するプロトコルに応じて、ストリーミングメディアプレーヤだけが関連のコンテンツをキャッシュできることを示すキャッシュ制御指示を含めてもよいし、含めなくともよい。ストリーミング媒体および非ストリーミング媒体（例えばＨＴＴＰ）両方に向けて設計されたプロトコルには、通例、ストリーミングメディアプレーヤ（一般的なＨＴＴＰコンポーネント（例えばウェブブラウザ）やプロキシデバイスではない）が関連のコンテンツをキャッシュできることを示す指示を使用する。ただし、ストリーミング媒体用に特に設計されたプロトコル（例えばＲＴＳＰ）には、ストリーミングメディアプレーヤだけが関連のコンテンツをキャッシュできるという指示は必要ない（したがって通例は使用しない）。

【００７３】
下記の表Ｉに、例示的なキャッシュ制御指示と、ストリーミング媒体用に特に設計されたプロトコル（例えばＲＴＳＰ）とでキャッシングが可能かどうかを示す。表ＩＩは、例示的なキャッシュ制御指示と、ストリーミング媒体向けでないプロトコル（例えばＨＴＴＰ

）でキャッシングが可能かどうかを示す。

【００７４】

【表１】

表１

キャッシュ制御指示	ストリーミングメディアプレーヤによるキャッシング
<none>	可能
"no-user-cache"	不可能
"no-cache"	不可能

【００７５】

【表２】

表１１

キャッシュ制御指示	一般的なHTTPキャッシュ／プロキシデバイスでのキャッシング	ストリーミングメディアプレーヤによるキャッシング
<none>	可能	可能
"no-cache, user-public"	不可能	可能
"no-cache"	不可能	不可能

【００７６】

表ＩＩで、「ｕｓｅｒ－ｐｕｂｌｉｃ」の指示は、ストリーミングメディアプレーヤ１４２によって、関連のコンテンツをキャッシュできるものと理解されたく、また「ｎｏ－ｃａｃｈｅ」の指示を無効にするということに留意されたい。ただし、「ｕｓｅｒ－ｐｕｂｌｉｃ」の指示が含まれない場合は、「ｎｏ－ｃａｃｈｅ」の指示により、関連のコンテンツをキャッシュできないことがストリーミングメディアプレーヤ１４２に示される。さらに、一般的なＨＴＴＰキャッシュまたはプロキシデバイスは、「ｕｓｅｒ－ｐｕｂｌｉｃ」の指示を理解せず、したがって単に「ｎｏ－ｃａｃｈｅ」指示に従い、コンテンツをキャッシュしない。

【００７７】

特定の部位のコンテンツをキャッシュするかどうかを判定する際には、各種の他の要因も絡む可能性がある。一実装では、ＵＲＬ（Ｕｎｉｆｏｒｍ　Ｒｅｓｏｕｒｃｅ　Ｌｏｃａｔｏｒ）などのストリーミングメディアコンテンツ識別子が、コンテンツでのキャッシュが可能である、あるいは可能でないことを示す指示を含むことができる。例えば、値「？ＷＭＣａｃｈｅ＝０」を使用して、コンテンツのキャッシングを行わないことを示し、値「？ＷＭＣａｃｈｅ＝１」を使用してコンテンツのキャッシングを行うことを示すことができる。別の実装では、ストリーミングメディアプレーヤ１４２は、ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを明確に使用禁止にすることを可能にし、それに応じて、ストリーミングメディアコンテンツがストリーミングメディアプレーヤ１４２によってキャッシュされないこととなる。さらに別の実装では、ストリーミングメディアプレーヤ１４２は、どのストリーミングメディアコンテンツもキャッシュしないことを決定することができる（例えば、ストリーミングメディアプレーヤ１４２は

、コンテンツをキャッシングするのに十分な記憶空間がない、あるいはコンテンツをそのキャッシュから再生した回数が閾値回数を下回るため、以降受信するストリーミングメディアコンテンツをキャッシュしないと決定することなどができる）。さらに別の実装では、特定タイプのストリーミングメディアコンテンツのみをキャッシュすることができる（例えばオンデマンドコンテンツはキャッシュ可能であるが、ブロードキャストコンテンツはキャッシュ可能でないなど）。

【００７８】

また、ストリーミングメディアプレーヤ１４２が、あるストリーミングメディアコンテンツをキャッシュすると決定した場合、そのキャッシングのために任意選択で各種の制約をストリーミングメディアプレーヤに課すことができる。例えば、そのコンテンツをキャッシュする場合にはストリームの間引き（ｔｈｉｎｎｉｎｇ）（例えば、帯域幅を節減するためにサーバデバイスからストリーミングするフレーム数を減らすなど）を要求しないようにストリーミングメディアプレーヤ１４２を設定することができる。

【００７９】

図７に、クライアントデバイスでストリーミングメディアコンテンツをキャッシュする例示的プロセス３００を説明するフローチャートを示す。プロセス３００は、図２のクライアントデバイス１０２のストリーミングメディアプレーヤ１４２などクライアントデバイスのストリーミングメディアプレーヤによって実施され、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組合せで実行することができる。

【００８０】

初めに、サーバデバイスからストリーミングメディアコンテンツを受信する（動作３０２）。プロセス３００は、そのストリーミングメディアコンテンツのキャッシングが可能であるかどうか（動作３０４）に基づいて進行する。キャッシングが可能である場合はコンテンツを不揮発性の記憶装置に保存し（動作３０６）、ストリーミングメディアコンテンツを再生が可能な状態にする（動作３０８）。動作３０８でそのストリーミングメディアコンテンツをキャッシュした場合には、そのコンテンツのストリーミングと同時に、かつ／または後の時間に再生可能な状態にできることに留意されたい。ただしキャッシングが可能でない場合は、単にストリーミングメディアコンテンツを再生可能な状態とし（動作３０８）、すなわちキャッシュしない場合には、そのコンテンツは再度ストリーミングしなければ後の時間に再生することができない。

【００８１】

図８Ａおよび８Ｂは、クライアントデバイスでストリーミングメディアコンテンツをキャッシュするかどうかを判定する例示的プロセス３３０を説明するフローチャートである。プロセス３３０は、図２のクライアントデバイス１０２のストリーミングメディアプレーヤ１４２など、クライアントデバイスのストリーミングメディアプレーヤによって実施され、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組合せで実行することができる。

【００８２】

初めに、ストリーミングメディアプレーヤは、サーバデバイスから取得可能なストリーミングメディアコンテンツにアクセスすることを試み（図８Ａの動作３３６）、サーバデバイスからそのストリーミングメディアコンテンツのヘッダ情報を取得する（動作３３８）。ストリーミングメディアプレーヤは次いで、ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしているかどうかを調べる（動作３４０）。ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしている場合は、クライアントでストリーミングメディアコンテンツをキャッシュメモリに入れない（動作３４２）。ただし、ユーザがストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしていない場合は、ストリーミングメディアプレーヤは、ストリーミングメディアコンテンツの識別子にそのコンテンツがキャッシングしないと示されているかどうかを調べる（動作３４４）。キャッシングしないと示されている場合は、ストリーミングメディアコンテンツをクライアントでキャッシュしない

（動作３４２）。

【００８３】

ただし、ストリーミングメディアコンテンツの識別子にそのコンテンツがキャッシングしないと示されていない場合は、ストリーミングメディアプレーヤは、そのストリーミングメディアプレーヤが、ストリーミングメディアコンテンツのクライアントでのキャッシングを使用禁止にしているかどうかを調べる（動作３４６）。使用禁止にしている場合には、ストリーミングメディアコンテンツをクライアントでキャッシュメモリに入れない（動作３４２）。使用禁止でない場合には、ストリーミングメディアプレーヤは、ストリーミングメディアコンテンツ識別子がコンテンツビットレートの変更子を含むかどうかを調べる（動作３４８）。

【００８４】

コンテンツビットレートの変更子が含まれる場合には、ストリーミングメディアプレーヤは、そのコンテンツビットレート変更子の値を超えない最高のビットレートを有するストリームのセットを選択する（図８Ｂの動作３５２）。コンテンツビットレートの変更子が含まれない場合には、ストリーミングメディアプレーヤは、リンク帯域幅とも称するクライアントとサーバ間の帯域幅に基づいてストリームのセットを選択する（図８Ｂの動作３５０）。

【００８５】

ストリームを選択すると、ストリーミングメディアプレーヤは、ストリーミングメディアコンテンツの識別子がそのコンテンツをキャッシュすべきと示しているかどうかに基づいて進行する（動作３５４）。キャッシュすべきと示されていない場合には、ストリーミングメディアプレーヤは、リンク帯域幅が、選択したストリームのビットレートの合計の１．５倍を超えるか、または選択したストリームのビットレートの合計の０．９倍未満であるかどうかを調べる（動作３５６）。あるいは、動作３５６では１．５および０．９以外の値を使用してもよい。リンク帯域幅が、選択したストリームのビットレート合計の１．５倍を超えないか、または選択したストリームのビットレート合計の０．９倍未満でない場合には、クライアントでストリーミングメディアコンテンツをキャッシュしない（図８Ａの動作３４２）。ただし、リンク帯域幅が、選択したストリームのビットレート合計の１．５倍を超えるか、または、選択したストリームのビットレート合計の０．９倍未満である場合には、ストリーミングメディアプレーヤは、選択したプロトコルがストリーミングメディアコンテンツのクライアントでのキャッシングをサポートするかどうかを調べる（動作３５８）。例示的な一実装では、ＴＣＰ（伝送制御プロトコル）によるデータ配信を使用したＨＴＴＰ１．１およびＲＴＳＰで、ストリーミングメディアコンテンツのクライアントでのキャッシングがサポートされる。

【００８６】

プロトコルがストリーミングメディアコンテンツのクライアントでのキャッシングをサポートしない場合には、ストリーミングメディアコンテンツをクライアントでキャッシュしない（図８Ａの動作３４２）。プロトコルがストリーミングメディアコンテンツのクライアントでのキャッシングをサポートする場合、またはストリーミングメディアコンテンツの識別子がコンテンツをキャッシュすべきであると示す（動作３５４）場合には、ストリーミングメディアプレーヤは、サーバがクライアントでのキャッシングを許可するかどうかを調べる（動作３６０）。サーバがクライアントでのキャッシングを許可しない場合には、ストリーミングメディアコンテンツをクライアントでキャッシュしない（図８Ａの動作３４２）。ただしサーバでクライアントでのキャッシングが可能であるとしている場合には、ストリーミングメディアコンテンツを入れるだけの空間がキャッシュメモリにあるかどうかを調べる（動作３６２）。ストリーミングメディアコンテンツをキャッシュするのに必要な空間量は、例えば図８Ａの動作３３８で受信するヘッダ情報で識別することができる。必要な場合にはすでにキャッシュされているコンテンツを任意選択でキャッシュメモリから追い出して十分な空間を作成することもできる。キャッシュメモリに十分な空間がない場合には、ストリーミングメディアコンテンツをクライアントでキャッシュしな

い（図８Ａの動作３４２）。ただしキャッシュするのに十分な空間がある場合には、ストリーミングメディアコンテンツをクライアントでキャッシュする（動作３６４）。

【００８７】

ストリーミングメディアコンテンツがブロードキャストコンテンツである場合など、状況によっては、コンテンツのサイズが分からない場合には，そのためそのコンテンツのすべてに十分な空間がキャッシュにあるかどうかの判定（動作３６２）を行わない場合があることに留意されたい。そのような状況では、単にキャッシュにそれ以上コンテンツを入れるだけのメモリ空間がなくなるまでコンテンツをキャッシュして、空間がなくなった時点でストリーミングメディアコンテンツのキャッシングをやめるようにしてもよい。あるいは、図８Ａの動作３３８で受信するヘッダ情報に入れるなどにより、ストリーミングメディアコンテンツのおよそのサイズをストリーミングメディアプレーヤが取得できるようにしてもよい（例えば、予測されるブロードキャスト継続時間や、ブロードキャストストリーミングメディアコンテンツの符号化ビットレートなどからブロードキャストコンテンツのおよそのサイズを容易に判定することができる）。

【００８８】

図８Ａおよび８Ｂのプロセス３３０は、クライアントデバイスにストリーミングメディアコンテンツをキャッシュするかどうかを判定する例示的なプロセスであり、プロセス３３０には色々な修正を加えてよいことにも留意されたい。例えば、ブロードキャストコンテンツをキャッシュしない実施形態では、コンテンツがブロードキャストコンテンツであるかどうかを調べ、ブロードキャストコンテンツである場合にはコンテンツをキャッシュしない（動作３４２）ように、追加的な動作をプロセス３３０に含めることができる。別の実施形態では，コンテンツ識別子がコンテンツをキャッシュすべきと指示しているかどうかを調べることができ、キャッシュすべきとの場合は、その指示が他のデフォルト設定に優先することができる（例えば、デフォルトではブロードキャストコンテンツをキャッシュしないが，キャッシュが可能というコンテンツ識別子中の指示によってそれを無効にすることができる）。さらに別の実施形態では、ユーザはクライアントでのキャッシングを使用禁止にすることが許されず、したがって動作３４０をプロセス３３０に含める必要がない。さらに別の実施形態では、コンテンツをキャッシュすべきかどうかをコンテンツ識別子で示すことができず、したがって動作３４４をプロセス３３０に含める必要がない。

【００８９】

本明細書には各種のプロセスをフローチャートの形で例示している。これらのプロセスに伴う動作は、フローチャートに示す順序で行っても、あるいは別の順序で行ってもよいことに留意されたい。例えば図８Ａおよび８Ｂでは、動作を図の順序で行っても、あるいは動作３６０または動作３６２を動作３４０より前に行っても，動作３４６を動作３４４の前に行ってもよい。

【００９０】

図９に、本明細書に記載された本発明に係わる技術を実施するのに使用できる例示的な一般的コンピューティング環境４００を示す。コンピュータ環境４００はコンピューティング環境の一例に過ぎず、このコンピュータおよびネットワークアーキテクチャの使用または機能性の範囲について何らの制限を示唆するものではない。またコンピュータ環境４００は、例示的コンピュータ環境４００に示す構成要素の任意の１つまたは組合せに関連する依存関係または必要性を有するものとも解釈すべきでない。

【００９１】

コンピュータ環境４００は，コンピュータ４０２の形態の汎用コンピューティングデバイスを含む。コンピュータ４０２は、例えば図１のクライアント１０２またはサーバ１０４、または図２のクライアント１０２またはサーバ１４６である。コンピュータ４０２の構成要素には、これらに限定しないが、１つまたは複数のプロセッサまたは処理装置４０４、システムメモリ４０６、およびプロセッサ４０４を含む各種のシステムコンポーネントをシステムメモリ４０６に結合するシステムバス４０８が含まれる。

【００９２】

システムバス４０８は、各種のバスアーキテクチャの任意のものを使用したメモリバスまたはメモリコントローラ、周辺バス、アクセラレーテッドグラフィックポート（ＡＧＰ）、およびプロセッサバスまたはローカルバスを含む数タイプのバス構造の任意の１つまたは複数を表す。例として、このようなアーキテクチャには、ＩＳＡ（Ｉｎｄｕｓｔｒｙ Ｓｔａｎｄａｒｄ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、ＭＣＡ（Ｍｉｃｒｏ Ｃｈａｎｎｅｌ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、ＥＩＳＡ（Ｅｎｈａｎｃｅｄ ＩＳＡ）バス、ＶＥＳＡ（Ｖｉｄｅｏ Ｅｌｅｃｔｒｏｎｉｃｓ Ｓｔａｎｄａｒｄｓ Ａｓｓｏｃｉａｔｉｏｎ）バス、およびメザニンバスとも称されるＰＣＩ（Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔｓ）バスが含まれる。

【００９３】

コンピュータ４０２は通例各種のコンピュータ可読媒体を含む。そのような媒体は、コンピュータ４０２がアクセスできる任意の利用可能媒体でよく、揮発性および不揮発性の媒体、取り外し可能および取り外し不能媒体が含まれる。

【００９４】

システムメモリ４０６は、ランダムアクセスメモリ（ＲＡＭ）４１０などの揮発性メモリの形態と、および／または読み取り専用メモリ（ＲＯＭ）４１２など不揮発性メモリの形態のコンピュータ可読媒体を含む。起動時などにコンピュータ４０２内の要素間の情報転送を助ける基本ルーチンを含む基本入出力システム（ＢＩＯＳ）４１４は、ＲＯＭ４１２に記憶される。ＲＡＭ４１０は通例、処理装置４０４から即座にアクセス可能な、かつ／または現在処理装置４０４によって操作中のデータおよび／またはプログラムモジュールを含む。

【００９５】

コンピュータ４０２は、この他の取り外し可能／取り外し不能、揮発性／不揮発性のコンピュータ記憶媒体も含むことができる。例として、図９には、取り外し不能、不揮発性の磁気媒体（図示せず）の読み取りおよび書き込みを行うハードディスクドライブ４１６、取り外し可能、不揮発性の磁気ディスク４２０（例えば「フロッピー（登録商標）ディスク」）の読み取りおよび書き込みを行う磁気ディスクドライブ４１８、およびＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、あるいは他の光媒体などの取り外し可能、不揮発性の光ディスク４２４の読み取りおよび書き込みを行う光ディスクドライブ４２２を示す。ハードディスクドライブ４１６、磁気ディスクドライブ４１８、および光ディスクドライブ４２２はそれぞれ、１つまたは複数のデータメディアインタフェース４２５によってシステムバス４０８に接続される。あるいは、ハードディスクドライブ４１６、磁気ディスクドライブ４１８、および光ディスクドライブ４２２は、１つまたは複数のインタフェース（図示せず）によってシステムバス４０８に接続することもできる。

【００９６】

これらのディスクドライブとそれに関連のコンピュータ可読媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、およびコンピュータ４０２のその他のデータの不揮発性の記憶を提供する。この例示としてハードディスク４１６、取り外し可能磁気ディスク４２０、および取り外し可能光ディスク４２４を図示しているが、これ以外に、磁気カセットまたはその他の磁気記憶装置、フラッシュメモリカード、ＣＤ－ＲＯＭ、デジタル多用途ディスク（ＤＶＤ）またはその他の光ストレージ、ランダムアクセスメモリ（ＲＡＭ）、読み取り専用メモリ（ＲＯＭ）、電気的に消去可能なプログラム可能読み取り専用メモリ（ＥＥＰＲＯＭ）など、データを記憶することができ、コンピュータによるアクセスが可能な他のタイプのコンピュータ可読媒体をこの例示的コンピューティングシステムおよび環境の実装に利用できることも理解されたい。

【００９７】

ハードディスク４１６、磁気ディスク４２０、光ディスク４２４、ＲＯＭ４１２、および／またはＲＡＭ４１０には任意数のプログラムモジュールを格納することができ、この例には、オペレーティングシステム４２６、１つまたは複数のアプリケーションプログラム４２８、その他のプログラムモジュール４３０、およびプログラムデータ４３２が含まれ

る。そのようなオペレーティングシステム４２６、１つまたは複数のアプリケーションプログラム４２８、他のプログラムモジュール４３０、およびプログラムデータ４３２（またはこれらの何らかの組合せ）はそれぞれ、分散ファイルシステムをサポートする常駐コンポーネントのすべてまたは一部を実装することができる。

【００９８】

ユーザは、キーボード４３４やポインティングデバイス４３６（例えば「マウス」）などの入力装置を介して、コンピュータ４０２にコマンドと情報を入力することができる。この他の入力装置４３８（具体的には図示せず）には、マイクロフォン、ジョイスティック、ゲームパッド、衛星受信アンテナ、シリアルポート、スキャナなどがある。これらおよびその他の入力装置は、システムバス４０８に結合された入出力インタフェース４４０を介して処理装置４０４に接続されるが、パラレルポート、ゲームポート、あるいはユニバーサルシリアルバス（ＵＳＢ）など、他のインタフェースおよびバス構造によって接続することもできる。

【００９９】

モニタ４４２またはその他のタイプの表示装置もビデオアダプタ４４４などのインタフェースを介してシステムバス４０８に接続することができる。モニタ４４２以外の他の出力周辺装置には、入出力インタフェース４４０を介してコンピュータ４０２に接続できるスピーカ（図示せず）やプリンタ４４６などのコンポーネントが含まれる。

【０１００】

コンピュータ４０２は、遠隔コンピュータ装置４４８など１つまたは複数の遠隔コンピュータへの論理結合を使用するネットワーク環境で動作することができる。例として、遠隔コンピュータ装置４４８は、パーソナルコンピュータ、携帯型コンピュータ、サーバ、ルータ、ネットワークコンピュータ、ピアデバイス、あるいはその他の一般的なネットワークノードなどである。遠隔コンピュータ装置４４８は、ここでコンピュータ４０２との関連で挙げた要素および機能の多くまたはすべてを含むことが可能な携帯型コンピュータとして図示している。

【０１０１】

コンピュータ４０２と遠隔コンピュータ４４８間の論理接続は、ローカルエリアネットワーク（ＬＡＮ）４５０および一般的な広域ネットワーク（ＷＡＮ）４５２として図示している。このようなネットワーク環境は、オフィス、企業内のコンピュータネットワーク、イントラネット、およびインターネットに一般的に見られる。

【０１０２】

ＬＡＮネットワーク環境で実装する場合、コンピュータ４０２は、ネットワークインタフェースまたはアダプタ４５４を介してローカルネットワーク４５０に接続される。ＷＡＮネットワーク環境で実装する場合、コンピュータ４０２は通例、広域ネットワーク４５２を通じて通信を確立するためのモデム４５６またはその他の手段を含む。モデム４５６は、コンピュータ４０２の内部にあっても外部にあってもよく、入出力インタフェース４４０またはその他の適切な機構を介してシステムバス４０８に接続することができる。図のネットワーク接続は例示的なものであり、コンピュータ４０２と４４８の間に通信リンクを確立することのできる他の手段を用いてもよいことを理解されたい。

【０１０３】

コンピューティング環境４００で示すようにネットワークされた環境では、コンピュータ４０２との関連で図に示すプログラムモジュールまたはその一部は、遠隔のメモリ記憶装置に格納することができる。例として、リモートアプリケーションプログラム４５８が遠隔コンピュータ４４８のメモリ装置に常駐している。例示のために、アプリケーションプログラムおよびオペレーティングシステムなどのその他の実行可能なプログラムコンポーネントをここでは別個のブロックとして表しているが、そのようなプログラムおよびコンポーネントは、様々な時にコンピューティングデバイス４０２の別々のストレージコンポーネント（記憶構成要素）に常駐し、コンピュータのデータプロセッサによって実行されることが分かる。

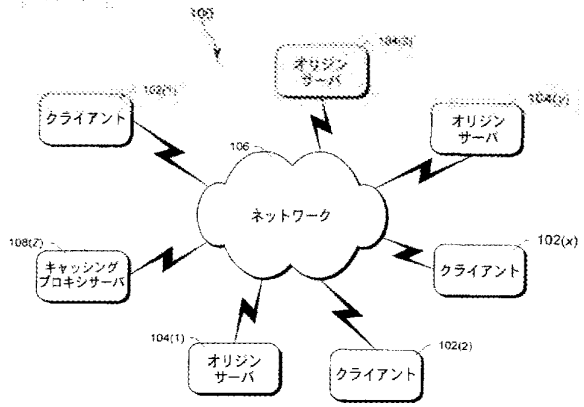
Aylo Exhibit 1018
Page 1851 of 2308

【０１０４】
ここでは各種のモジュールおよび技術を、１つまたは複数のコンピュータまたは他のデバイスによって実行されるプログラムモジュールなどのコンピュータ実行可能命令の一般的な文脈で説明することができる。一般に、プログラムモジュールには、特定のタスクを実行するか、特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、データ構造などが含まれる。通例、プログラムモジュールの機能は、各種の実施形態で必要に応じて組み合わせても分散させてもよい。
【０１０５】
こうしたモジュールおよび技術の実装は、何らかの形態のコンピュータ可読媒体に格納しても、そうしたコンピュータ可読媒体を通じて送信してもよい。コンピュータ可読媒体は、コンピュータによるアクセスが可能な任意の利用可能媒体でよい。例として、これに限定しないが、コンピュータ可読媒体は、「コンピュータ記憶媒体」と「通信媒体」を含むことができる。
【０１０６】
「コンピュータ記憶媒体」には、コンピュータ可読命令、データ構造、プログラムモジュール、あるいは他のデータなどの情報を記憶するための任意の方法または技術で実行される、揮発性および不揮発性、取り外し可能および取り外し不能の媒体が含まれる。コンピュータ記憶媒体には、これらに限定しないが、ＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリまたはその他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多用途ディスク（ＤＶＤ）またはその他の光ストレージ、磁気カセット、磁気テープ、磁気ディスクストレージまたはその他の磁気記憶装置、あるいは所望の情報の記憶に使用することができ、コンピュータによってアクセスできる任意の他の媒体が含まれる。
【０１０７】
「通信媒体」は通例、コンピュータ可読命令、データ構造、プログラムモジュール、またはその他のデータを搬送波やその他の移送機構などの変調データ信号で具現化する。通信媒体には、任意の情報伝達媒体も含まれる。用語「変調データ信号」とは、信号中に情報を符号化するような形でその特性の１つまたは複数を設定または変化させた信号を意味する。例として、通信媒体には、これらに限定しないが、配線ネットワークや直接配線接続などの配線媒体と、音響、ＲＦ（無線周波数）、赤外線、その他の無線媒体などの無線媒体が含まれる。上記の媒体の組合せもコンピュータ可読媒体の範囲に含まれる。
【０１０８】
上記の説明では、構造的特徴および／または方法論的な動作に固有の術語を使用しているが、頭記の特許請求の範囲に定義する本発明はここに記載する特定の特徴または動作に限定されないことを理解されたい。それら特定の特徴および動作は、本発明を実施する例示的形態として開示する。
【図面の簡単な説明】
【図１】ストリーミングメディアコンテンツのクライアント側でのキャッシングを用いることができる例示的なネットワーク環境を示す図である。
【図２】例示的なクライアントデバイスおよびサーバデバイスを示す図である。
【図３】ストリーミングメディアデータを通信する際に使用できる例示的なメッセージフォーマットを示す図である。
【図４】サーバからクライアントにメディアコンテンツをストリーミングする例示的プロセスを示すフローチャートである。
【図５】サーバからクライアントにメディアコンテンツをストリーミングするのと同時に、シャトルコントロールコマンドを受け付ける例示的プロセスを示すフローチャートである。
【図６】後の再生のためにストリーミングメディアコンテンツを記録する例示的プロセスを示すフローチャートである。
【図７】クライアントデバイスにストリーミングメディアコンテンツをキャッシュする例示的プロセスを示すフローチャートである。

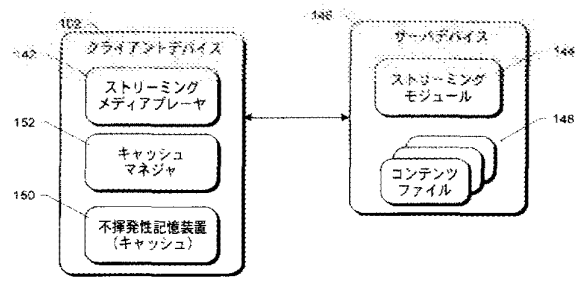
【図８Ａ】ストリーミングメディアコンテンツをクライアントデバイスにキャッシュするかどうかを判定する例示的プロセスを示すフローチャートである。
【図８Ｂ】ストリーミングメディアコンテンツをクライアントデバイスにキャッシュするかどうかを判定する例示的プロセスを示すフローチャートである。
【図９】本発明に係わる技術を実施するのに使用できる例示的な一般的コンピュータ環境を示す図である。
【符号の説明】
１００　ネットワーク環境
１０２（１）、１０２（２）、１０２（ｘ）　クライアントデバイス
１０４（１）、１０４（２）、１０４（ｘ）　オリジンサーバデバイス
１０６　ネットワーク
１０８　キャッシングプロキシサーバ
１４２　ストリーミングメディアプレーヤ
１４４　ストリーミングモジュール
１４６　ソースサーバデバイス
１４８　コンテンツファイル
１５０　不揮発性記憶装置（キャッシュメモリ）
１５２　キャッシュマネジャ
１８０　データ構造
１８２　スタートライン部分
１８４　ヘッダ
１８６　空のライン
１８８　メッセージ本体
４００　コンピュータ環境
４０２　コンピュータ
４０４　処理装置
４０６　システムメモリ
４０８　システムバス
４１０　ＲＡＭ
４１２　ＲＯＭ
４１４　ＢＩＯＳ
４１６　ハードディスクドライブ
４１８　磁気ディスクドライブ
４２０　磁気ディスク
４２２　光ディスクドライブ
４２４　光ディスク
４２５　データメディアインタフェース
４２６　オペレーティングシステム
４２８　アプリケーションプログラム
４３０　プログラムモジュール
４３２　プログラムデータ
４３４　キーボード
４３６　ポインティングデバイス（マウス）
４３８　他の入力装置（他のデバイス）
４４０　入出力インターフェース
４４２　モニタ
４４４　ビデオアダプタ
４４６　プリンタ
４４８　遠隔コンピュータ装置
４５０　ＬＡＮ

４５２　ＷＡＮ（インターネット）
４５６　モデム
４５８　リモートアプリケーションプログラム

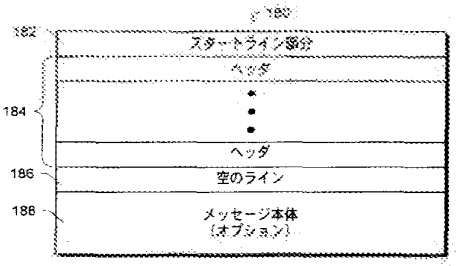
【図１】



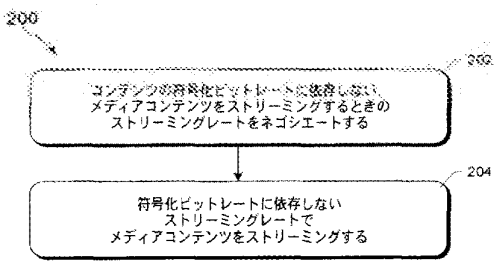
【図２】



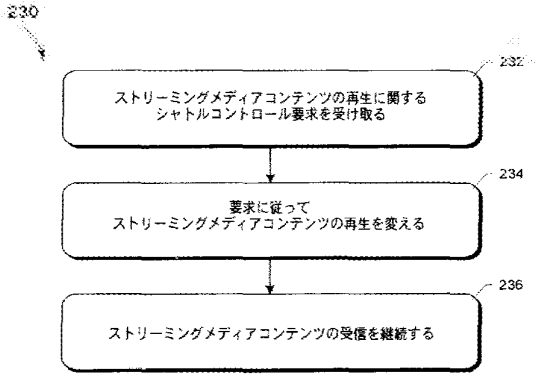
【図３】



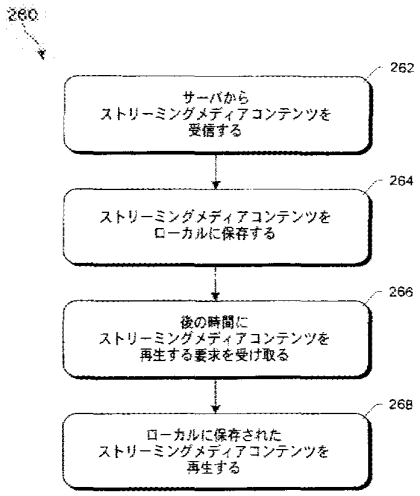
【図４】



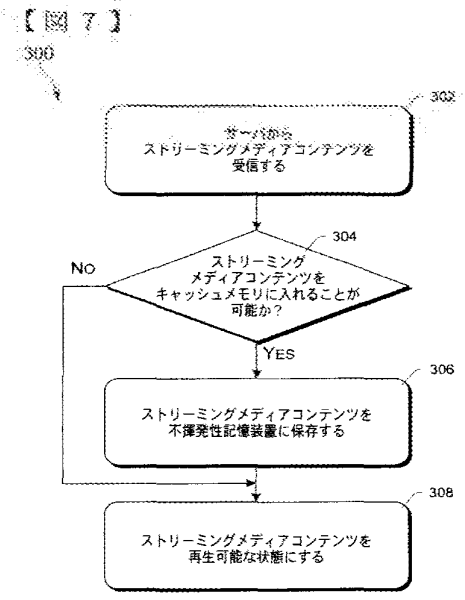
【図５】


230
232
ストリーミングメディアコンテンツの再生に関する
シャトルコントロール要求を受け取る
234
要求に従って
ストリーミングメディアコンテンツの再生を変える
236
ストリーミングメディアコンテンツの受信を継続する

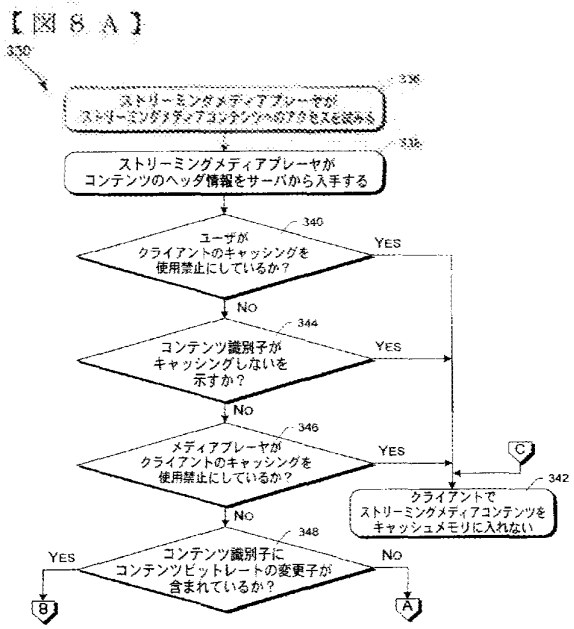
【図６】


260
262
サーバから
ストリーミングメディアコンテンツを
受信する
264
ストリーミングメディアコンテンツを
ローカルに保存する
266
後の時間に
ストリーミングメディアコンテンツを
再生する要求を受け取る
268
ローカルに保存された
ストリーミングメディアコンテンツを
再生する

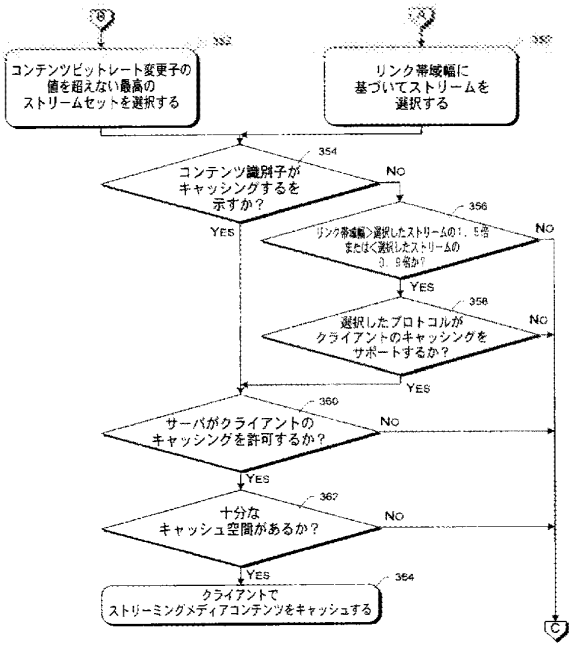
【図７】


300
302
サーバから
ストリーミングメディアコンテンツを
受信する
304
ストリーミング
メディアコンテンツを
キャッシュメモリに入れることが
可能か？
NO
YES
306
ストリーミングメディアコンテンツを
不揮発性記憶装置に保存する
308
ストリーミングメディアコンテンツを
再生可能な状態にする

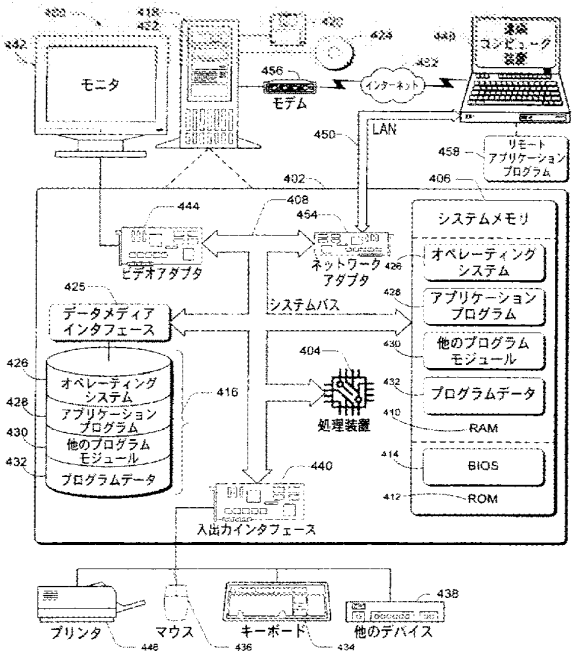
【図８Ａ】


330
336
ストリーミングメディアプレーヤが
ストリーミングメディアコンテンツへのアクセスを試みる
338
ストリーミングメディアプレーヤが
コンテンツのヘッダ情報をサーバから入手する
340
ユーザが
クライアントのキャッシングを
使用禁止にしているか？
YES
NO
344
コンテンツ識別子が
キャッシングしないを
示すか？
YES
NO
346
メディアプレーヤが
クライアントのキャッシングを
使用禁止にしているか？
YES
C
NO
342
クライアントで
ストリーミングメディアコンテンツを
キャッシュメモリに入れない
348
コンテンツ識別子に
コンテンツビットレートの変更子が
含まれているか？
YES
NO
B
A

【図８Ｂ】


B
352
コンテンツビットレート変更子の値を超えない最高のストリームセットを選択する
B
350
リンク帯域幅に基づいてストリームを選択する
354
コンテンツ識別子がキャッシングするを示すか？
NO
YES
356
リンク帯域幅>選択したストリームの1．5倍またはく選択したストリームの3．9倍か？
NO
YES
358
選択したプロトコルがクライアントのキャッシングをサポートするか？
NO
YES
360
サーバがクライアントのキャッシングを許可するか？
NO
YES
362
十分なキャッシュ空間があるか？
NO
YES
364
クライアントでストリーミングメディアコンテンツをキャッシュする
C

【図９】


モニタ
モデム
インターネット
コンピュータ
LAN
リモートアプリケーションプログラム
システムメモリ
オペレーティングシステム
アプリケーションプログラム
他のプログラムモジュール
プログラムデータ
RAM
BIOS
ROM
ビデオアダプタ
ネットワークアダプタ
システムバス
データメディアインタフェース
処理装置
入出力インタフェース
プリンタ
マウス
キーボード
他のデバイス

フロントページの続き

(72)発明者 トロイ ディー. バッターベリー
アメリカ合衆国 ９８０３３ ワシントン州 カークランド ノースイースト ８２ ランス １２２１６
(72)発明者 エドアルド ビー. オリベイラ
アメリカ合衆国 ９８０５２ ワシントン州 レッドモンド レッドーウッド ロード ノースイースト ９２００ アパートメント ディー３２５

Ｆターム(参考) 5B077 AA43 BD02
5C064 BA01 BB05 BC10 BC16 BC20 BC23 BC25 BD01 BD02 BD07
BD08 BD09

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2011-004225
(43)Date of publication of application : 06.01.2011

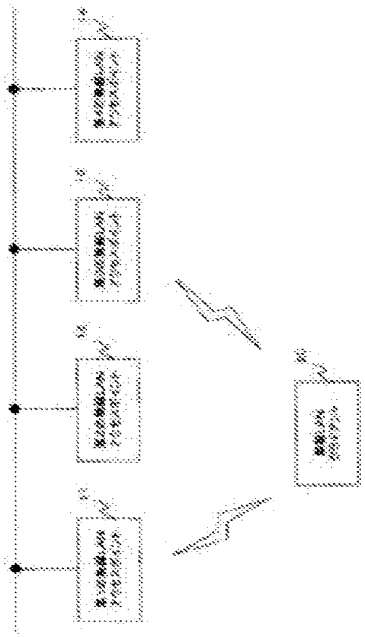
(51)Int.Cl. *H04W 48/20 (2009.01)*
H04W 84/12 (2009.01)
H04W 48/16 (2009.01)

(21)Application number : 2009-146288
(22)Date of filing : 19.06.2009

(71)Applicant : NEC INFRONTIA CORP
(72)Inventor : HODOSHIMA TAKESHI

(54) SYSTEM, METHOD AND PROGRAM OF HIGH-SPEED ROAMING OF WIRELESS LAN

(57)Abstract:
PROBLEM TO BE SOLVED: To attain high-speed roaming in a wireless LAN network.
SOLUTION: In a LAN system including: a plurality of wireless LAN access points; and a wireless LAN client to be connected to the wireless LAN access points, the wireless LAN access points are the ones existing on the same local area network, used channels and BSSIDS of the wireless LAN access points set to the same ESSID are recognized as wireless LAN access point information, held as data, and the wireless LAN client utilizes the wireless LAN access point information when the above connection is performed.



Aylo Exhibit 1018
Page 1858 of 2308

(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2011-4225
(P2011-4225A)
(43)公開日　平成23年1月6日(2011.1.6)

(51)Int.Cl.		F I			テーマコード（参考）
H04W 48/20	(2009.01)	H04Q	7/00	416	5K067
H04W 84/12	(2009.01)	H04Q	7/00	630	
H04W 48/16	(2009.01)	H04Q	7/00	404	

審査請求 未請求　請求項の数 15　OL　（全 11 頁）

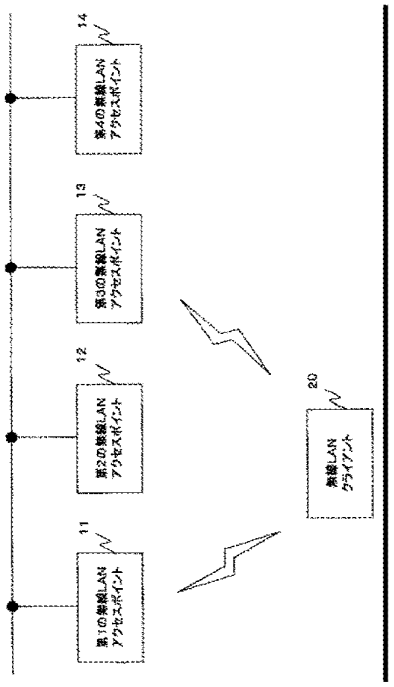
(21)出願番号　特願2009-146288 (P2009-146288)
(22)出願日　平成21年6月19日 (2009.6.19)

(71)出願人　000227205
ＮＥＣインフロンティア株式会社
神奈川県川崎市高津区北見方２丁目６番１号
(74)代理人　100130029
弁理士　永井　道雄
(74)代理人　100166338
弁理士　関口　正夫
(74)代理人　100152054
弁理士　仲野　孝雅
(72)発明者　程島　健
神奈川県川崎市高津区北見方２丁目６番１号　ＮＥＣインフロンティア株式会社内

Ｆターム(参考)　5K067 AA15 EE02 EE10 HH22 HH23 JJ72

(54)【発明の名称】無線ＬＡＮの高速ローミングシステム、高速ローミング方法及び高速ローミングプログラム

(57)【要約】
【課題】無線ＬＡＮネットワークにおいて高速ローミングを実現する。
【解決手段】複数の無線ＬＡＮアクセスポイントと、当該無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントと、を有する無線ＬＡＮシステムにおいて、前記無線ＬＡＮアクセスポイントは同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として認識しデータとして保持し、前記無線ＬＡＮクライアントは前記接続を行う際に前記無線ＬＡＮアクセスポイント情報を利用する。
【選択図】図１



【特許請求の範囲】
【請求項1】
複数の無線ＬＡＮアクセスポイントと、当該無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントと、を有する無線ＬＡＮシステムにおける接続方法において、
前記無線ＬＡＮアクセスポイントは同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として認識しデータとして保持し、前記無線ＬＡＮクライアントは前記接続を行う際に前記無線ＬＡＮアクセスポイント情報を利用することを特徴とする接続方法。
【請求項2】
前記無線ＬＡＮアクセスポイントは、前記無線ＬＡＮクライアントと接続された際に、自らが保持している前記無線ＬＡＮアクセスポイント情報を無線ＬＡＮクライアントに受け渡すことを特徴とする請求項１に記載の接続方法。
【請求項3】
前記無線ＬＡＮクライアントは最初に前記無線ＬＡＮアクセスポイントに接続を試みる際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、当該スキャンをしたチャンネルの中から最良の無線ＬＡＮアクセスポイントを選択し接続することを特徴とする請求項１又は２に記載の接続方法。
【請求項4】
前記最低限のチャンネルのスキャンは、前記無線ＬＡＮクライアントが、前記無線ＬＡＮアクセスポイント情報の中から自分の接続可能なＥＳＳＩＤを持つ無線ＬＡＮアクセスポイントのチャンネルを抽出し、当該抽出したチャンネルのみを前記最低限のチャンネルと判断することにより行われることを特徴とする請求項１乃至３の何れか１項に記載の接続方法。
【請求項5】
複数の無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントが行う接続方法であって、
前記無線ＬＡＮアクセスポイントに最初に接続を試みる際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、接続された際に前記無線ＬＡＮアクセスポイントから、当該無線ＬＡＮアクセスポイントと同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として受け取り、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、その中から最良の無線ＬＡＮアクセスポイントを選択し接続することを特徴とする接続方法。
【請求項6】
複数の無線ＬＡＮアクセスポイントと、当該無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントと、を有する無線ＬＡＮ接続システムにおいて、
前記無線ＬＡＮアクセスポイントは同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として認識しデータとして保持し、前記無線ＬＡＮクライアントは前記接続を

Aylo Exhibit 1018
Page 1860 of 2308

行う際に前記無線ＬＡＮアクセスポイント情報を利用することを特徴とする接続システム。

【請求項7】

前記無線ＬＡＮアクセスポイントは、前記無線ＬＡＮクライアントと接続された際に、自らが保持している前記無線ＬＡＮアクセスポイント情報を無線ＬＡＮクライアントに受け渡すことを特徴とする請求項６に記載の接続システム。

【請求項8】

前記無線ＬＡＮクライアントは最初に前記無線ＬＡＮアクセスポイントに接続を試みる際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、当該スキャンをしたチャンネルの中から最良の無線ＬＡＮアクセスポイントを選択し接続することを特徴とする請求項６又は７に記載の接続システム。

【請求項9】

前記最低限のチャンネルのスキャンは、前記無線ＬＡＮクライアントが、前記無線ＬＡＮアクセスポイント情報の中から自分の接続可能なＥＳＳＩＤを持つ無線ＬＡＮアクセスポイントのチャンネルを抽出し、当該抽出したチャンネルのみを前記最低限のチャンネルと判断することにより行われることを特徴とする請求項６乃至８の何れか１項に記載の接続システム。

【請求項10】

複数の無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントにおいて、

前記無線ＬＡＮアクセスポイントに最初に接続を試みる際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、接続された際に前記無線ＬＡＮアクセスポイントから、当該無線ＬＡＮアクセスポイントと同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として受け取り、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、その中から最良の無線ＬＡＮアクセスポイントを選択し接続することを特徴とする無線ＬＡＮクライアント。

【請求項11】

複数の無線ＬＡＮアクセスポイントと、当該無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントと、を有するシステムに搭載される無線ＬＡＮ接続プログラムにおいて、

前記無線ＬＡＮアクセスポイントは同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として認識しデータとして保持し、前記無線ＬＡＮクライアントは前記接続を行う際に前記無線ＬＡＮアクセスポイント情報を利用するシステムとしてコンピュータを機能させることを特徴とする接続プログラム。

【請求項12】

前記無線ＬＡＮアクセスポイントは、前記無線ＬＡＮクライアントと接続された際に、自らが保持している前記無線ＬＡＮアクセスポイント情報を無線ＬＡＮクライアントに受け渡すことを特徴とする請求項１１に記載の接続プログラム。

【請求項13】

前記無線ＬＡＮクライアントは最初に前記無線ＬＡＮアクセスポイントに接続を試みる

際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、当該スキャンをしたチャンネルの中から最良の無線ＬＡＮアクセスポイントを選択し接続することを特徴とする請求項１１又は１２に記載の接続プログラム。

【請求項14】

前記最低限のチャンネルのスキャンは、前記無線ＬＡＮクライアントが、前記無線ＬＡＮアクセスポイント情報の中から自分の接続可能なＥＳＳＩＤを持つ無線ＬＡＮアクセスポイントのチャンネルを抽出し、当該抽出したチャンネルのみを前記最低限のチャンネルと判断することにより行われることを特徴とする請求項１１乃至１３の何れか１項に記載の接続プログラム。

【請求項15】

複数の無線ＬＡＮアクセスポイントと接続される無線ＬＡＮクライアントに搭載される接続プログラムにおいて、

前記無線ＬＡＮアクセスポイントに最初に接続を試みる際は前記ローカルエリアネットワーク上に存在する前記無線ＬＡＮアクセスポイントの全チャンネルをスキャンし、当該スキャンをした中から最良の条件の無線ＬＡＮアクセスポイントを選択し接続し、接続された際に前記無線ＬＡＮアクセスポイントから、当該無線ＬＡＮアクセスポイントと同一のローカルエリアネットワーク上に存在する無線ＬＡＮアクセスポイントであって同一のＥＳＳＩＤに設定されている無線ＬＡＮアクセスポイントの使用チャンネル及びＢＳＳＩＤを同一ネットワークの無線ＬＡＮアクセスポイント情報として受け取り、前記無線ＬＡＮクライアントが前記無線ＬＡＮアクセスポイントと圏外から再接続を試みる際、或いは、ローミングする際には、前記受け取った無線ＬＡＮアクセスポイント情報を基に、最低限のチャンネルをスキャンし、その中から最良の無線ＬＡＮアクセスポイントを選択し接続する無線ＬＡＮクライアントとしてコンピュータを機能させることを特徴とする接続プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、無線ＬＡＮに関し、より詳細には無線ＬＡＮのローミング又は再接続に関する。

【背景技術】

【0002】

近年無線ＬＡＮの機器の低価格化等の影響もあり、無線ＬＡＮが一般に広く普及している。また、無線ＬＡＮを構築する際に１つのローカルエリアネットワーク内に多数の無線ＬＡＮアクセスポイントを敷設するといったことも一般的に行われている。

【0003】

このような環境下において無線ＬＡＮクライアントは、１つのローカルエリアネットワーク内の全ての無線ＬＡＮアクセスポイントを対象として、接続可能な無線ＬＡＮアクセスポイントを検索する。

【0004】

そして、無線ＬＡＮクライアントは、検索の結果に基づいて最良と判断した無線ＬＡＮアクセスポイントに対して接続動作をする（例えば特許文献１参照）。

【図面の簡単な説明】

【0016】

【図1】本発明の実施形態における無線ＬＡＮネットワークの基本的構成の例を表す図である。

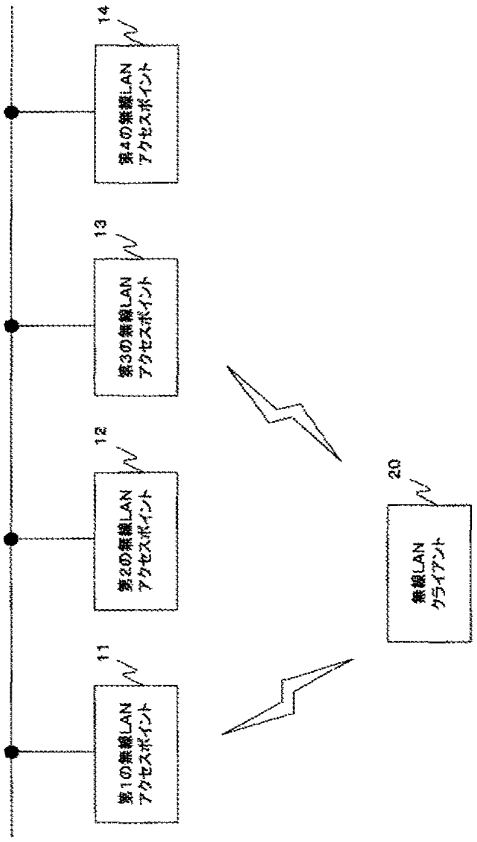
【図2】本発明の実施形態における各機器の設定の例を表す図である。

【図3】本発明の実施形態における無線ＬＡＮアクセスポイント情報の例を表す図である

Aylo Exhibit 1018
Page 1862 of 2308

。
【図4】本発明の実施形態における無線ＬＡＮクライアントの無線ＬＡＮアクセスポイント接続動作を表すフローチャートである。
【符号の説明】
【0045】
１１　第１の無線ＬＡＮアクセスポイント
１２　第２の無線ＬＡＮアクセスポイント
１３　第３の無線ＬＡＮアクセスポイント
１４　第４の無線ＬＡＮアクセスポイント
２０　無線ＬＡＮクライアント

【図1】


第1の無線LAN
アクセスポイント
11
第2の無線LAN
アクセスポイント
12
第3の無線LAN
アクセスポイント
13
第4の無線LAN
アクセスポイント
14
無線LAN
クライアント
20

【図2】

各機器の設定			
SYSTEM NAME	BSSID	ESSID	CH
AP11	11111−11111	Office1	1
AP12	22222−22222	Office1	4
AP13	33333−33333	Office1	7
AP14	44444−44444	Office2	14
CL20	−	Office1	−

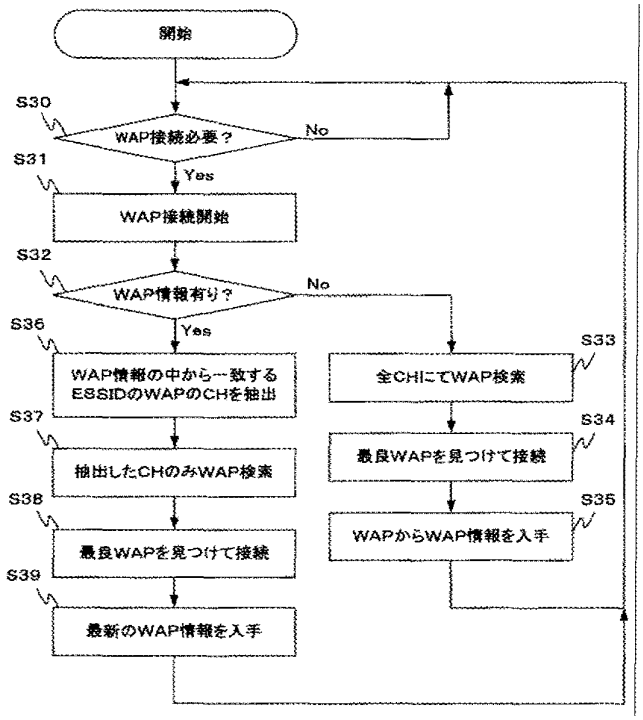
【図3】

無線LANアクセスポイント情報			
SYSTEM NAME	BSSID	ESSID	CH
AP11	11111−11111	Office1	1
AP12	22222−22222	Office1	4
AP13	33333−33333	Office1	7
AP14	44444−44444	Office2	14

特開2011-4225(P2011-4225A)

Aylo Exhibit 1018
Page 1866 of 2308

【図4】


開始
S30
WAP接続必要？
No
Yes
S31
WAP接続開始
S32
WAP情報有り？
No
Yes
S36
WAP情報の中から一致するESSIDのWAPのCHを抽出
S37
抽出したCHのみWAP検索
S38
最良WAPを見つけて接続
S39
最新のWAP情報を入手
S33
全CHにてWAP検索
S34
最良WAPを見つけて接続
S35
WAPからWAP情報を入手

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-201343
(43)Date of publication of application : 18.07.2000

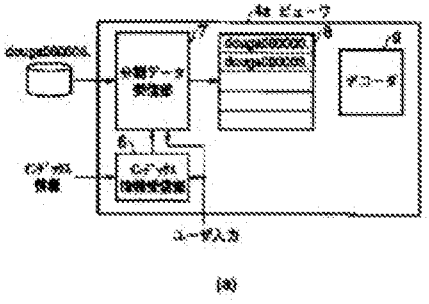
(51)Int.Cl. H04N 7/173

(21)Application number : 11-000497
(22)Date of filing : 05.01.1999

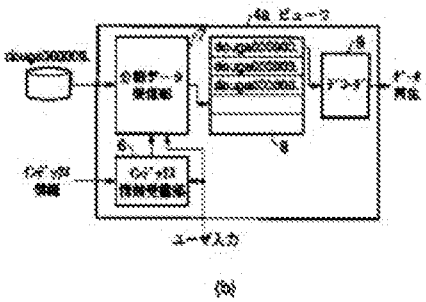
(71)Applicant : TOSHIBA CORP
(72)Inventor : CHIJITANI MASAHIDE

(54) MOVING IMAGE DATA REPRODUCING DEVICE, MOVING IMAGE DATA DIVIDING DEVICE AND RECORDING MEDIUM READABLE BY COMPUTER IN WHICH PROGRAM IS RECORDED

(57)Abstract:
PROBLEM TO BE SOLVED: To enable random access to moving image data without changing a server function.
SOLUTION: A moving image data reproducing device with an index information receiving means 6 to receive index information to be generated by relating plural pieces of divided data to be obtained by deviding the moving image data to its reproduction time information by requesting the information to a server, a divided data receiving means 7 to receive the divided data related to reproduction time information at the reproduction time information to which starting of reproducing is instructed and afterward by successively requesting the information to the server by referring to the index information received by the index information receiving means 6, a storage means 9 to store the divided data received by the divided data receiving means 7 and a reproducing means to successively reproduce the divided data stored in the storage means 9 according to the index information received by the index information receiving means 6 is provided.



(a)



(b)

Aylo Exhibit 1018
Page 1869 of 2308

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（Ａ）　(11)特許出願公開番号
特開2000－201343
（P2000－201343A）
(43)公開日　平成12年７月18日(2000.7.18)

(51)Int.Cl.⁷	識別記号	ＦＩ		テーマコード（参考）
Ｈ０４Ｎ　7/173	６３０	Ｈ０４Ｎ　7/173	６３０	５Ｃ０６４

審査請求　未請求　請求項の数13　ＯＬ　（全 13 頁）

(21)出願番号　特願平11－497

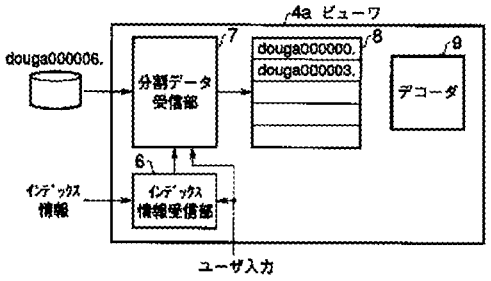
(22)出願日　平成11年１月５日(1999.1.5)

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者　千々谷　眞英
東京都府中市東芝町１番地　株式会社東芝府中工場内
(74)代理人　100058479
弁理士　鈴江　武彦　（外６名）

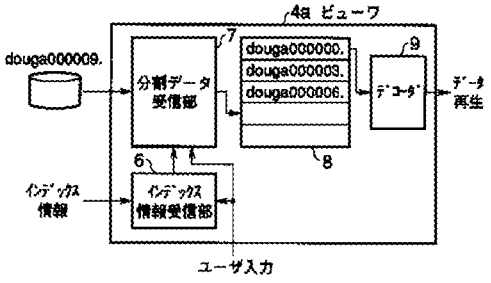
Ｆターム(参考)　5C064 BA07 BB10 BC16 BC20 BC27 BD16

(54)【発明の名称】　動画像データ再生装置及び動画像データ分割装置並びにプログラムを記録したコンピュータ読み取り可能な記録媒体

(57)【要約】
【課題】サーバの機能を変更することなく、動画像データのランダムアクセスを可能にする。
【解決手段】動画像データが分割されて得られる複数の分割データとその再生時刻情報とを関連付けて生成されるインデックス情報を、サーバに要求して受信するインデックス情報受信手段と、当該インデックス情報受信手段に受信されたインデックス情報を参照し、再生開始を指示された再生時刻情報以降の再生時刻情報に関連付けられた分割データをサーバに順次要求して受信する分割データ受信手段と、当該分割データ受信手段に受信された分割データを記憶する記憶手段と、インデックス情報受信手段に受信されたインデックス情報にしたがって、記憶手段に記憶されている分割データを順次再生させる再生手段とを具備した動画像データ再生装置を提供する。



(a)



(b)

1

【特許請求の範囲】

【請求項１】　動画像データが分割されて得られる複数の分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報を、サーバに要求して受信するインデックス情報受信手段と、

当該インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられた分割データを、サーバに順次要求して受信する分割データ受信手段と、

当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、

前記記憶手段に記憶されている分割データを順次再生させる再生手段とを具備したことを特徴とする動画像データ再生装置。

【請求項２】　請求項１記載の動画像データ再生装置において、

前記分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定手段を付加し、

前記分割データ受信手段は、当該時間推定手段によって推定された時間とほぼ同じ時間分の分割データが前記記憶手段に記憶されていない場合に、前記インデックス情報にしたがってサーバに分割データを要求し、

前記再生手段は、前記時間推定手段によって推定された時間とほぼ同じ時間分の分割データが前記記憶手段に記憶された場合に、この記憶手段に記憶されている分割データの再生を開始することを特徴とする動画像データ再生装置。

【請求項３】　動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複数の種別が生成されている場合に、この複数種別の各分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報を、サーバに要求して受信するインデックス情報受信手段と、

前記分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定手段と、

当該時間推定手段によって推定された時間に基づいて、要求する分割データの種別を選択する分割データ選択手段と、

前記インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられ、かつ前記分割データ選択手段によって選択されている種別の分割データを、サーバに順次要求して受信する分割データ受信手段と、

当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、

前記記憶手段に記憶されている分割データを順次再生させる再生手段とを具備したことを特徴とする動画像データ再生装置。

2

【請求項４】　請求項３記載の動画像データ再生装置において、

前記分割データ受信手段は、前記時間推定手段に推定された時間とほぼ同じ時間分の分割データが前記記憶手段に記憶されていない場合に、前記データ選択手段によって選択されている種別の分割データを、前記インデックス情報にしたがって前記サーバに要求し、

前記再生手段は、前記時間推定手段によって推定された時間とほぼ同じ時間分の分割データが前記記憶手段に記憶された場合に、この記憶手段に記憶されている分割データの再生を開始することを特徴とする動画像データ再生装置。

【請求項５】　動画像データを複数に分割して複数の分割データを生成し、かつ前記複数の分割データとこの分割データの再生時刻情報とを関連付けたインデックス情報を生成する動画像データ分割装置。

【請求項６】　動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複数の種別を生成し、この複数の種別の各分割データをそれぞれの再生時刻情報に関連付けたインデックス情報を生成する動画像データ分割装置。

【請求項７】　コンピュータに、

動画像データが分割されて得られる複数の分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報を、サーバに要求して受信するインデックス情報受信機能と、

当該インデックス情報受信機能によって受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられた分割データを、サーバに順次要求して受信する分割データ受信機能と、

当該分割データ受信機能によって受信された分割データを記憶する記憶機能と、

前記記憶機能によって記憶されている分割データを順次再生させる再生機能とを実現させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項８】　請求項７記載のコンピュータ読み取り可能な記録媒体において、

前記分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定機能を付加し、

前記分割データ受信機能は、当該時間推定機能によって推定された時間とほぼ同じ時間分の分割データが前記記憶機能によって記憶されていない場合に、前記インデックス情報にしたがってサーバに分割データを要求し、

前記再生機能は、前記時間推定機能によって推定された時間とほぼ同じ時間分の分割データが前記記憶機能によって記憶された場合に、この記憶機能によって記憶され

Aylo Exhibit 1018
Page 1871 of 2308

ている分割データの再生を開始することを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項９】　コンピュータに、
動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複数の種別が生成されている場合に、この複数種別の各分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報をサーバに要求して受信するインデックス情報受信機能と、
前記分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定機能と、
当該時間推定機能によって推定された時間に基づいて、要求する分割データの種別を選択する分割データ選択機能と、
前記インデックス情報受信機能によって受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられ、かつ前記分割データ選択機能によって選択されている種別の分割データを、サーバに順次要求して受信する分割データ受信機能と、
当該分割データ受信機能によって受信された分割データを記憶する記憶機能と、
前記記憶機能によって記憶されている分割データを順次再生させる再生機能とを実現させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項１０】　請求項９記載のコンピュータ読み取り可能な記録媒体において、
前記分割データ受信機能は、前記時間推定機能によって推定された時間とほぼ同じ時間分の分割データが前記記憶機能に記憶されていない場合に、前記データ選択機能によって選択されている種別の分割データを、前記インデックス情報にしたがって前記サーバに要求し、
前記再生機能は、前記時間推定機能によって推定された時間とほぼ同じ時間分の分割データが前記記憶機能によって記憶された場合に、この記憶機能によって記憶されている分割データの再生を開始することを特徴とするコンピュータ読み取り可能な記録媒体。

【請求項１１】　コンピュータに、
動画像データを複数に分割して複数の分割データを生成し、かつ前記複数の分割データとこの分割データの再生時刻情報とを関連付けたインデックス情報を生成する機能を実現させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項１２】　コンピュータに、
動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複数の種別を生成し、この複数の種別の各分割データをそれぞれの再生時刻情報に関連付けたインデックス情報を生成する機能を実現させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項１３】　サーバに記憶されている動画像データをクライアントに配信する動画像データ配信システムにおいて、
前記サーバは、
前記動画像データが分割されて得られる複数の分割データと、この各分割データとそれぞれの再生時刻情報とを関連付けたインデックス情報とを保持し、
前記クライアントは、
前記インデックス情報を前記サーバに要求して受信するインデックス情報受信手段と、
当該インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられた分割データを前記サーバに順次要求して受信する分割データ受信手段と、
当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、
前記記憶手段に記憶されている分割データを順次再生させる再生手段とを備えることを特徴とする動画像データ配信システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、サーバの構成及び機能を変更することなく、動画像データのランダムアクセスを可能にするための動画像データ再生装置及び動画像データ分割装置並びにプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【０００２】

【従来の技術】従来の動画像データ配信システムは、主にクライアントとＷＷＷサーバとから構成される。この従来の動画像データ配信システムでは、まずクライアントからＷＷＷサーバに再生を求める動画像データ（動画コンテンツ）の伝送要求が送信される。

【０００３】ＷＷＷサーバは各種データを保持しており、このデータの中から要求された動画像データを選択し送信する。

【０００４】送信された動画像データは、伝送要求を送信したクライアントに受信され、ストリーム再生がなされる。

【０００５】このような従来の動画像データ配信システムにおいて、クライアント及びＷＷＷサーバがインターネットを介して接続されている場合には、一般的にＨＴＴＰプロトコルにより動画像データの伝送がなされる。

【０００６】ここで、例えば動画像データの再生を求めるユーザの中には、動画像データの途中位置からの再生（以下、「ランダムアクセス」という）を求める者もいる。

【０００７】従来の動画像データ配信システムによってランダムアクセスを実現するためには、ランダムアクセスを可能とする専用のＨＴＴＰサーバをこの動画像デー

Aylo Exhibit 1018
Page 1872 of 2308

タ配信システムに適用する必要がある。

【０００８】また、別の手法として、ランダムアクセスを可能とする専用のプログラムであるＣＧＩ（Common Gateway Interface）をWWWサーバに実装させる必要がある。

【０００９】この専用のＨＴＴＰサーバや、ＣＧＩを実装したWWWサーバは、動画像データを再生するクライアントから独自の要求を受信し、動画像データを指定された位置からＨＴＴＰプロトコルによって送信する機能を持つ。

【００１０】

【発明が解決しようとする課題】しかしながら、上記のように、ランダムアクセスを可能とするために専用のＨＴＴＰサーバを用いたり、サーバにＣＧＩを実装する方法においては、従来の動画像データ配信システムのWWWサーバに手を加える必要があり、現状のインフラのままでは実現することができない。

【００１１】したがって、ランダムアクセスの実現、導入が困難であり、作業が複雑である。

【００１２】また、この従来の方法では、ランダムアクセスのための処理が主にサーバでなされるため、サーバの負荷が増加するという問題が生じる。特に、動画像データ配信システムが複数のクライアントを有する場合には、その傾向が強くなる。

【００１３】本発明は、以上のような実状に鑑みてなされたもので、サーバの構成や機能に手を加えることなく動画像データのランダムアクセスを可能とするための動画像データ再生装置及び動画像データ分割装置並びにプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを目的とする。

【００１４】

【課題を解決するための手段】本発明の骨子は、動画像データを複数の分割して分割データとし、この分割データにそれぞれの再生時間を示す再生時刻情報を関連づけてインデックス情報とし、このインデックス情報にしたがって分割データを順次再生する点にある。

【００１５】以下、本発明を実現するにあたって講じた具体的手段について説明する。

【００１６】第１の発明は、動画像データが分割されて得られる複数の分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報をサーバに要求して受信するインデックス情報受信手段と、当該インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられた分割データをサーバに順次要求して受信する分割データ受信手段と、当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、インデックス情報受信手段に受信されたインデックス情報の再生時刻情報にしたがって、記憶手段に記憶されている分割データを順次再生させる再生手段とを具備した動画像データ再生装置である。

【００１７】この第１の発明の動画像データ再生装置においては、再生開始を指示された再生時刻情報以降の分割データが順次受信されストリーム再生される。

【００１８】また、この動画像データ再生装置を適用する場合には、サーバは自己が保持するデータのうち要求されたデータを送信する機能を有していればよい。

【００１９】したがって、第１の発明においては、サーバに新しい構成や機能を付加することなく容易にランダムアクセスが実現される。

【００２０】また、ランダムアクセスを実現させるための処理が、サーバではなく動画像データ再生装置（クライアント）側によって実行されるため、サーバの負荷の増加を防止することができる。

【００２１】第２の発明は、第１の発明の動画像データ再生装置において、分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定手段を付加し、分割データ受信手段は、当該時間推定手段によって推定された時間とほぼ同じ時間分の分割データが記憶手段に記憶されていない場合に、前記インデックス情報にしたがってサーバに分割データを要求し、再生手段は、時間推定手段によって推定された時間とほぼ同じ時間分の分割データが記憶手段に記憶された場合に、この記憶手段に記憶されている分割データの順次再生を開始する動画像データ再生装置である。

【００２２】この第２の発明の動画像データ再生装置においては、分割データを要求してから受信するまでの間に再生されると推測される分割データが、記憶手段に保持される。

【００２３】したがって、この第２の発明においては、第１の発明と同様の作用効果に加えて、必要最低限の量の分割データが格納されしだい、動画像データの再生がなされるため、再生が開始されるまでの時間を最適化することができる。

【００２４】これにより、再生開始までの時間を短縮させることができる。

【００２５】また、次の分割データが要求されてから受信されるまでの間に必要な分割データは確保済みであるため、動画像データの再生が途切れることを防止することができる。

【００２６】さらに、分割データをサーバに要求してから受信するまでの時間が逐次測定され、この時間にしたがって再生可能な状態で記憶しておく分割データの量が決定される。

【００２７】ゆえに、回線の動的な負荷変動にも対応することができ、適切な量の分割データが確保される。

【００２８】したがって、スムーズなストリーム再生が可能となる。

【００２９】第３の発明は、動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複

Aylo Exhibit 1018
Page 1873 of 2308

数の種別が生成されている場合に、この複数種別の各分割データとそれぞれの再生時刻情報とを関連付けて生成されるインデックス情報をサーバに要求して受信するインデックス情報受信手段と、分割データの送信をサーバに要求してから、この分割データを受信するまでの時間を推定する時間推定手段と、当該時間推定手段によって推定された時間に基づいて、要求する分割データの種別を選択する分割データ選択手段と、インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられ、かつ分割データ選択手段によって選択されている種別の分割データをサーバに順次要求して受信する分割データ受信手段と、当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、インデックス情報受信手段に受信されたインデックス情報の再生時刻情報にしたがって、記憶手段に記憶されている分割データを順次再生させる再生手段とを具備した動画像データ再生装置である。

【００３０】この第３の発明の動画像データ再生装置においては、伝送速度の異なる複数種別の分割データのうち、回線の負荷変動を考慮して適した伝送速度の分割データが選択され受信される。

【００３１】したがって、回線の伝送速度が遅い場合には、高速伝送可能な分割データにより伝送がなされるため、再生すべきデータが未伝送となることを防止することができる。

【００３２】また、第１の発明と同様の作用により、サーバを変更することなく、ランダムアクセスが可能である。

【００３３】第４の発明は、第３の発明の動画像データ再生装置において、分割データ受信手段は、時間推定手段に推定された時間とほぼ同じ時間分の分割データが記憶手段に記憶されていない場合に、データ選択手段によって選択されている種別の分割データを、インデックス情報にしたがってサーバに要求し、再生手段は、時間推定手段によって推定された時間とほぼ同じ時間分の分割データが記憶手段に記憶された場合に、この記憶手段に記憶されている分割データの順次再生を開始する動画像データ再生装置である。

【００３４】この第４の発明の動画像データ再生装置においては、第３の発明と同様の作用効果に加えて、分割データをサーバに要求してから受信するまでの間の再生に必要と推測される分の分割データが記憶手段に保持される。

【００３５】したがって、先の第２の発明と同様の効果を得ることができる。

【００３６】第５の発明は、動画像データを複数に分割して複数の分割データを生成し、かつ前記複数の分割データとこの分割データの再生時刻情報とを関連付けたインデックス情報を生成する動画像データ分割装置である。

【００３７】この第５の発明の動画像データ分割装置を用いて複数の分割データ及びインデックス情報を作成してサーバに備え、第１又は第２の発明の動画像データ再生装置を用いることで、サーバの構成、機能に手を加えることなくランダムアクセスが可能となる。

【００３８】また、この動画像データ分割装置を用いることで、複数の分割データ及びインデックス情報を容易に作成できる。

【００３９】第６の発明は、動画像データが分割されて得られる複数の分割データについて伝送速度が異なる複数の種別を生成し、この複数の種別の各分割データをそれぞれの再生時刻情報に関連付けたインデックス情報を生成する動画像データ分割装置である。

【００４０】この第６の発明の動画像データ分割装置を用いて複数種別の各分割データ及びインデックス情報を作成してサーバに備え、第３又は第４の発明の動画像データ再生装置を用いることで、サーバの構成、機能に手を加えることなくランダムアクセスが可能となり、伝送にかかる時間を調節することができる。

【００４１】また、この動画像データ分割装置を用いることで、複数種別の分割データ及びインデックス情報を容易に作成できる。

【００４２】第７乃至第１２の発明は、それぞれ上記第１乃至第６の発明と同様の機能をコンピュータに実現させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【００４３】この第７乃至第１２の発明のようなプログラムを記録した記録媒体を用いることによって、上述した機能を有していない装置や、計算機においても、簡単に上述した機能を付加することができる。

【００４４】第１３の発明は、サーバに記憶されている動画像データをクライアントに配信する動画像データ配信システムにおいて、サーバは、動画像データが分割されて得られる複数の分割データと、この各分割データとそれぞれの再生時刻情報とを関連付けたインデックス情報とを保持し、クライアントは、インデックス情報をサーバに要求して受信するインデックス情報受信手段と、当該インデックス情報受信手段に受信されたインデックス情報にしたがって、再生開始を指示された再生時刻情報及びその後の再生時刻情報に関連付けられた分割データをサーバに順次要求して受信する分割データ受信手段と、当該分割データ受信手段によって受信された分割データを記憶する記憶手段と、インデックス情報受信手段に受信されたインデックス情報の再生時刻情報にしたがって、記憶手段に記憶されている分割データを順次再生させる再生手段とを備える動画像データ配信システムである。

【００４５】この第１３の発明の動画像データ配信システムにおいては、クライアントが先の第１の発明の動画

Aylo Exhibit 1018
Page 1874 of 2308

像データ再生装置と同様に機能する。

【００４６】これにより、第１の発明と同様の作用効果を得ることができる。

【００４７】なお、第２乃至第１２の発明も、この第１３の発明と同様に、動画像データ配信システムに適用可能である。

【００４８】

【発明の実施の形態】以下、図面を参照しながら本発明の実施の形態について説明する。

【００４９】（第１の実施の形態）本実施の形態においては、動画像データ（動画コンテンツ）を複数に分割した分割データと、この複数の分割データとそれぞれの再生時間情報を関連付けたインデックス情報とを用いることで、サーバを変更することなく動画像データのランダムアクセスを可能とする動画像データ配信システムについて説明する。

【００５０】図１は、本実施の形態に係る動画像データ配信システムの構成を示すブロック図である。

【００５１】この動画像データ配信システム１は、動画像データから分割データとインデックス情報とを生成するエンコーダ２と、内部の記憶装置３ａに各種データをファイル形式で記憶するサーバ３と、動画像データ再生用のビューワ４ａを有するクライアント４とから構成されている。

【００５２】この動画像データ配信システム１では、動画像データがスプライスポイントでファイル分割されており、ビューワ４ａがインデックス情報に基づいて好きな位置の分割データをサーバ３に要求し、受信する。

【００５３】以下、この動画像データ配信システム１の各構成要素について詳しく説明する。

【００５４】図２は、本実施の形態に係る動画像データ配信システム１に備えられるエンコーダ２の機能を示すブロック図である。

【００５５】このエンコーダ２は、動画エンコード時に、ビデオ等の動画素材からなる動画像データ（動画コンテンツ）を所定のスプライスポイント毎に分割して複数の分割データを生成する。また生成した各分割データには異なるファイル名を付けて出力する。この図２では、分割データのファイル名の例として「douga000000」、「douga000003」、「douga000006」等が示されている。

【００５６】なお、この各分割データは、ビューワ４ａ内の受信ページバッファの１ページ内に収まるサイズで生成される。

【００５７】さらに、このエンコーダ２は、各分割データのファイル名と、それぞれの再生時刻の情報とを関係付けてインデックス情報５を生成し、ファイル形式で出力する。

【００５８】図３は、本実施の形態に係る動画像データ配信システム１に備えられるサーバ３の構成を示すブロック図である。

【００５９】サーバ３は、主に記憶装置３ａと、サーバ本体３ｂとから構成される。

【００６０】サーバ本体３ｂは、記憶装置３ａに様々なデータを蓄積する機能と、クライアント４からの要求にしたがって記憶装置３ａの内容を読み出して送信する機能を有する。

【００６１】記憶装置３ａは、各種のデータを記憶し、ここではエンコーダ２によって生成された複数の分割データ及びインデックス情報５も記憶している。

【００６２】クライアント４は、図１で示すように、主にクライアント本体４ｂと、動画像データ再生装置として機能するビューワ４ａとから構成されており、クライアント本体４ｂ上でビューワ４ａが稼働する。

【００６３】図４は、本実施の形態に係る動画像データ配信システム１に備えられるビューワ４ａの機能を示すブロック図である。

【００６４】なお、図４（ａ）は再生開始前のバッファリング中の状態を示し、３ページ分の分割データがページバッファに格納されるまでの状態を示している。

【００６５】一方、図４（ｂ）は再生中の状態を示しており、１ページ分の分割データがデコーダに読み出されると、再びサーバ３から分割データが読み出される状態を示している。

【００６６】このビューワ４ａは、動画像データを再生する装置であり、インデックス情報受信部６、分割データ受信部７、ページバッファ８、デコーダ９から構成される。

【００６７】インデックス情報受信部６は、再生を求める動画像データの指定をユーザから入力する。また、指定された動画像データに関するインデックス情報５をサーバ３に要求し、この要求の応答としてサーバ３からインデックス情報５を受信して保持する。

【００６８】分割データ受信部７は、動画像データの再生開始位置となる再生時刻情報をユーザから入力する。また、インデックス情報受信部６のインデックス情報５を参照し、ユーザから入力された再生時刻情報以降の再生時刻情報に関連付けられているファイルの送信をサーバ３に要求する。この要求は、ページバッファ８に格納された分割データの量が一定量（ここでは３つとする）になるまで順次出力する。

【００６９】そして、この分割データ受信部７は、この要求の応答として分割データを順次受信し、ページバッファ８に出力する。

【００７０】ページバッファ８は、複数のページからなり、各ページ毎に一つの分割データをバッファリングする。また、各ページの内容を再生時刻順に並び替えて記憶する。さらに、デコーダ９によって読み出されたページ内の分割データを破棄し、次に入力される分割データを新規に確保する。

Aylo Exhibit 1018
Page 1875 of 2308

【００７１】デコーダ９は、ページバッファ８内に一定量の分割データが記憶されると、デコード処理を開始し、その後順次デコード処理を続ける。これにより滞りなく動画像データの再生がなされる。

【００７２】次に、上記のような構成を有する動画像データ配信システム１の動作について以下に説明する。

【００７３】まず、エンコーダ２によって動画像データが分割され、分割データ及びインデックス情報が生成され、サーバ３に記憶される。

【００７４】次に、再生を求める動画像データの指定及び再生を開始させる位置が、ユーザによってビューワ４ａに入力される。

【００７５】次に、ユーザに指定された動画像データから生成されたインデックス情報５が、ビューワ４ａのインデックス情報受信部６からサーバ３に要求され、受信される。

【００７６】この受信されたインデックス情報５には、動画像データの再生時刻情報に対して分割データのファイル名が記述されている。

【００７７】分割データ受信部７では、受信されたインデックス情報５に基づいて、再生を開始したい位置以降に対応付けられている分割データがサーバ３に要求され、受信される。

【００７８】受信された分割データは、ページバッファ８の各ページに格納され、再生される時刻順に並び替えられる。

【００７９】また、ページバッファ８において一定量のバッファリングがなされると、デコーダ９によってストリーム再生が開始される。

【００８０】ページバッファ８においては、分割データが１ページ分再生される度にそのデータが破棄されてページが空けられる。

【００８１】そして、分割データ受信部７において、未受信の次の分割データが、インデックス情報５に基づいて検索され、サーバ３に要求、受信され、再生が繰り返される以上説明したように、本実施の形態に係る動画像データ配信システム１においては、動画像データがスプライスポイントにおいてファイル分割されており、その状態がインデックス情報５に記述されている。

【００８２】また、ビューワ４ａが、このインデックス情報５を要求し、続いてこのインデックス情報５に記されている好きな位置から分割データを順次要求・受信し、再生する。

【００８３】したがって、ユーザの望む位置からの分割データが順次再生されるため、動画像データのランダムアクセスが可能となる。

【００８４】また、本実施の形態においては、サーバ３の機能はファイルを保持し、またこの保持されているファイルの中から要求されたファイルを送信するのみである。したがって、この動画像データ配信システム１を導入する場合にはサーバ３に新しい機能や作用を追加する必要がない。

【００８５】ゆえに、クライアント４側のビューワ４ａを備え、サーバ３に分割データとインデックス情報５とを格納するのみで、容易にランダムアクセスを実現することができる。

【００８６】さらに、ランダムアクセスを実現させるための処理が、サーバ３ではなくビューワ４ａによって実行されるため、クライアント４が複数存在してもサーバ３の負荷の増加を防止することができる。

【００８７】なお、本実施の形態に係る動画像データ配信システム１では、エンコーダ２の機能をクライアント４に実装させ、クライアント４がネットワークを介して分割データ及びインデックス情報５をサーバ３に格納するとしてもよい。

【００８８】また、本実施の形態においては、ビューワ４ａの機能をブラウザ等の機能として実現可能であり、サーバ３はWWWサーバとすることができる。

【００８９】これにより、ブラウザで参照する動画コンテンツのランダムアクセスが可能となる。

【００９０】さらに、本実施の形態に係る動画像データ配信システム１では、クライアント４の台数を複数にすることができる。

【００９１】（第２の実施の形態）本実施の形態においては、第１の実施の形態に係る動画像データ配信システム１に備えられるビューワ４ａを変形したビューワについて説明する。

【００９２】先の第１の実施の形態においては、ビューワ４ａがサーバ３に分割データを要求してから受信するまでの時間がネットワークの回線の負荷に依存する。また、この回線の負荷は、様々な要因により時々刻々と変化する。

【００９３】先のビューワ４ａでは、伝送に要する時間や、回線の負荷による伝送速度の変化を吸収するために、動画像データの再生開始時に一定量分（しきい値を設定する）の分割データをバッファリングし、以後順次分割データを読み出している。

【００９４】この手法では、ビューワ４ａにバッファリングされる分割データの量を大きくすればその分だけ回線の負荷による遅延に対応可能となるが、その分再生開始までの待ち時間が長くなる。

【００９５】この再生開始までの待ち時間は、回線の負荷による遅延を吸収できる範囲で最小化されることが望ましい。したがって、バッファリングされる分割データの量も、回線負荷による伝送時間の変化に対応可能な範囲で最小化されることが望ましい。

【００９６】本実施の形態においては、分割データを要求してから受信するまでの時間を推定する。また、この推定された時間に基づいて、ページバッファに記憶しておく分割データの量を変更し、バッファリングされる分

Aylo Exhibit 1018
Page 1876 of 2308

割データの量を動的に最適化するビューワに関して説明する。

【００９７】図５は、本実施の形態に係るビューワの構成を示すブロック図であり、図４と同一の部分には同一の符号を付してその説明を省略し、ここでは異なる部分についてのみ詳しく説明する。

【００９８】このビューワ１０は、第１の実施の形態において説明したビューワ４ａの分割データ受信部７の機能を変更した分割データ受信部１１を備え、新規の構成として時間推定部１２を付加している。また、このビューワ１０は、第１の実施の形態と同様の動画像データ配信システム１に設置可能である。

【００９９】時間推定部１２は、分割データ受信部１１がサーバ３に分割データを要求し、この分割データを受信するまでの時間（以下、「要求受信時間」という）を測定する。

【０１００】なお、この測定された要求受信時間を厳密に用いると、バッファリングの際に余裕がなくなる。したがって、時間推定部１２は多少余分を持たせるために若干多めの値を要求受信時間として推定する。例えば実際の要求受信時間が０．５秒であってとしても、要求受信時間を０．７秒と推定し、設定する。

【０１０１】この時間推定部１２において、要求受信時間は、再生開始以降には分割データが要求されて受信される度に再推定され、更新される。

【０１０２】分割データ受信部１１は、先の第１の実施の形態で説明した分割データ受信部７と同様の処理を行うが、サーバ３から分割データを読み出す際の条件が異なっている。

【０１０３】再生開始前においては、この分割データ受信部１１は、ページバッファ８内の分割データの再生時間の合計が要求受信時間を超えるまで、サーバ３から分割データを受信する。

【０１０４】再生中においては、現在再生中の分割データに関連付けされている再生時刻情報に現状の要求受信時間を加算し、この加算値に該当する再生時刻情報に関連付けされた分割データを要求し、受信する。すなわち、例えば要求受信時間が０．７秒と設定されているとすると、０．７秒後に再生されることになる分割データを要求し、受信する。

【０１０５】上記のような構成を有するビューワ１０の動作について以下に説明する。

【０１０６】図６は、本実施の形態に係るビューワ１０の再生開始前までの動作を示すフローチャートである。

【０１０７】このビューワ１０においては、まずインデックス情報５がサーバ３から受信される（ｓ１）。

【０１０８】次に、このインデックス情報５とユーザの指定内容に基づいて、再生開始位置の分割データが要求される（ｓ２）。

【０１０９】次に、要求された分割データが受信され（ｓ３）、この際に余分を持たせた要求受信時間が推定される（ｓ４）。

【０１１０】次に、インデックス情報５を参照して、次の分割データが要求され（ｓ５）、受信される（ｓ６）。

【０１１１】そして、要求受信時間分の分割データが、ページバッファ８に受信されたか否かが判定され（ｓ７）、受信されている場合には、再生が開始される。

【０１１２】受信されていない場合には、再度次の分割データが要求され、受信される。

【０１１３】図７は、本実施の形態に係るビューワ１０の再生開始後の動作を示すフローチャートである。

【０１１４】再生開始後においては、ページバッファ８に空きがあり（ｔ１）、かつ要求受信時間分の分割データがページバッファ８にない場合に（ｔ２）、次の分割データが要求される（ｔ３）。

【０１１５】なお、この間にもページバッファ８内の分割データはデコードされ再生されている（ｔ４、ｔ９）。

【０１１６】次に、要求された分割データが受信され（ｔ５）、この分割データが最後の分割データか否かが判定される（ｔ６）。

【０１１７】最後の分割データの場合にはデコード処理及び再生が継続され（ｔ７）、処理が終了される。

【０１１８】最後の分割データでない場合には、余分を持たせた要求受信時間が推定され（ｔ８）、デコード処理の実行により動画像データが再生されつつ（ｔ９）、上記の処理が繰り返される。

【０１１９】以上説明したように、本実施の形態に係るビューワ１０においては、サーバに分割データを要求してから受信するまでの時間に多少の余分を加えた時間分の分割データがページバッファ８に蓄積される。

【０１２０】したがって、必要最小限の分割データのみが先行確保されるため、再生開始までの時間を最適化することができる。

【０１２１】（第３の実施の形態）動画像データの再生中に、ネットワークの負荷により回線の伝送速度が分割データの伝送速度未満となると、動画像データがビューワにおいて途切れなく再生されない場合がある。

【０１２２】本実施の形態においては、このような場合に、再生の途中からでも高速伝送可能な分割データに切り換えて伝送を行い、これによりスムーズな再生を維持する動画像データ配信システムについて説明する。

【０１２３】本実施の形態に係る動画像データ配信システムは、図１と同様の構成を有するが、エンコーダ及びビューワに特別の機能が付加されている。

【０１２４】本システムのエンコーダは、元の動画像データが同一であり、スプライスポイントも同一であるが同一の条件における伝送速度が異なる複数種別の分割データ及びインデックス情報を生成する。

Aylo Exhibit 1018
Page 1877 of 2308

【０１２５】また、本システムのビューワは、要求受信時間に基づいて分割データの種別を選択する機能を有する。

【０１２６】以下、図１の場合と異なる機能を有するエンコーダ、及びビューワについて詳しく説明する。

【０１２７】図８は、本実施の形態に係る動画像データ配信システムに備えられるエンコーダの機能を示すブロック図である。

【０１２８】このエンコーダ１３は、エンコード時に、１つの動画像データから伝送速度の異なる複数種別の分割データを生成し、出力する。

【０１２９】なお、以下においては伝送速度の異なる２つの種別の分割データが生成される場合を例として説明するが、３つ以上の種別が生成される場合も同様である。また、この２つの種別のうち、一方を標準分割データとし、この標準分割データよりも伝送速度の速いもう一方を高速分割データとして説明する。

【０１３０】さらに、このエンコーダ１３は、生成された標準分割データのファイル名と高速分割データのファイル名を再生時刻情報に関連付けたインデックス情報１４を生成し、ファイル形式で出力する。

【０１３１】すなわち、このインデックス情報１４には、各再生時刻情報に複数列の分割データのファイル名が関連付けられており、この各列は異なる伝送速度になるようにエンコードされた分割データ列である。

【０１３２】サーバは、標準分割データ、高速分割データ、インデックス情報をファイルとして記憶し、クライアントのビューワから要求されたファイルを送信する。

【０１３３】ビューワは、第２の実施の形態に係る図５のビューワ１０と同様の構成を有するが、分割データ受信部１１に分割データ選択機能が付加されている点が異なる。

【０１３４】この分割データ選択機能は、標準分割データを要求してから受信するまでの時間が、時間推定部１２で推定された要求受信時間よりも大幅に大きい場合に、ネットワークの伝送速度が分割データの伝送速度よりも遅いと判断する。この場合、分割データ選択機能は、インデックス情報１４中の分割データの参照を、標準分割データ列から高速分割データ列に切り換える。

【０１３５】一方、高速分割データを要求してから受信するまでの時間が、時間推定部１２で推定された要求受信時間よりもある程度小さくなった場合には、ネットワークの伝送速度が分割データの伝送速度よりも速くなったと判断し、インデックス情報１４中の分割データの参照を、高速分割データから伝送速度は遅いが画質のよい標準分割データ列に切り換える。

【０１３６】図９は、このビューワにおいて再生開始後に実行される処理を示すフローチャートである。

【０１３７】このビューワでは、動画像データがデコードされ、再生されるが（ｕ１）、それと並列に要求伝送時間分の分割データがページバッファ８に格納されているか否かが判定される（ｕ２）。

【０１３８】ここで、要求伝送時間分の分割データがページバッファ８に格納されていない場合には、この要求伝送時間が大きすぎるか否かが分割データ選択機能によって判定される（ｕ３）。

【０１３９】要求伝送時間が大きすぎる場合には、インデックス情報１４の分割データ列のうち伝送速度の速い方が選択される（ｕ４）。

【０１４０】一方、要求伝送時間が大きすぎない場合には、要求伝送時間が小さすぎないか否かが判定される（ｕ５）。

【０１４１】要求伝送時間が小さすぎる場合には、インデックス情報１４の分割データ列のうち伝送速度の遅い方が選択される（ｕ６）。

【０１４２】そして、この分割データ選択機能によって選択された種別における次の分割データが、サーバに要求される（ｕ７）。

【０１４３】この間にも、動画像データのデコード処理が実行され、再生がなされている（ｕ８）。

【０１４４】要求された分割データが受信されると（ｕ９）、余分を持たせた要求受信時間が推定される（ｕ１０）。

【０１４５】そして、要求する分割データがなくなり（ｕ１１）、かつ再生する分割データもなくなると（ｕ１２）、処理が終了する。

【０１４６】要求する分割データが存在する場合には、上記の処理が繰り返される。

【０１４７】以上説明したように、本実施の形態に係る動画像データ配信システムにおいては、異なる伝送速度の分割データが生成され、分割データを要求してから受信するまでの時間が長い場合には、ビューワは伝送速度の速い分割データを読み出す。

【０１４８】また、分割データの種別の切り替えは各分割データの読み出し毎に行うことができるため、動画像データの再生開始後においても伝送速度を変更することができ、ネットワークの状況変更に柔軟に対応することができる。

【０１４９】さらに、伝送速度に余裕のある場合には画質のよい分割データに変更することができる。

【０１５０】ゆえに、クライアントのビューワで動画像データを途切れなく、良好に再生することができる。

【０１５１】なお、上記の各実施の形態において説明したビューワ、エンコーダの機能は、コンピュータに実行させることのできるプログラムとして、例えば磁気ディスク（フロッピーディスク、ハードディスク等）、光ディスク（ＣＤ－ＲＯＭ、ＤＶＤ等）、半導体メモリなどの記録媒体に書き込んで適用したり、通信媒体により伝送して各種装置に適用することも可能である。このビューワやエンコーダの機能を実現するコンピュータは、記

Aylo Exhibit 1018
Page 1878 of 2308

録媒体に記録されたプログラムを読み込み、プログラムによって動作が制御されることにより、上述した処理を実行する。

【０１５２】

【発明の効果】以上詳記したように本発明では、動画像データを分解した複数の分割データと、この各分割データとそれぞれの再生時刻情報とを関連付けたインデックス情報をサーバに備える。

【０１５３】そして、動画像データ再生装置は、まずインデックス情報を受信し、再生を開始したい位置の時間再生情報以降に関連付けられている分割データを順次読み出す。

【０１５４】これにより、サーバの構成、機能に手を加えることなく、ランダムアクセスが可能になる。

【０１５５】また、このランダムアクセスのための処理がサーバではなく動画像データ再生装置によって実行されるため、ランダムアクセス実行によるサーバの負荷の増加を防止することができる。

【０１５６】さらに、本発明では、動画像データ再生装置に蓄える分割データの量を、分割データを要求してからこの分割データを受信するまでの時間分とすることで、再生開始までの待ち時間を最適化することができる。

【０１５７】さらに、本発明では、伝送速度の異なる複数種別の分割データを用意しておき、ネットワークの状況によって受信する分割データの種別を切り換えることで、ネットワークにおける負荷の影響を受けることなくスムーズな再生が可能となる。

【図面の簡単な説明】

【図１】本発明の第１の実施の形態に係る動画像データ配信システムの構成を示すブロック図。

＊【図２】同実施の形態におけるエンコーダの機能を示すブロック図。

【図３】同実施の形態に係る動画像データ配信システムに備えられるサーバの構成を示すブロック図。

【図４】同実施の形態に係る動画像データ配信システムに備えられるビューワの機能を示ブロック図。

【図５】本発明の第２の実施の形態に係るビューワの構成を示すブロック図。

【図６】同実施の形態に係るビューワの再生開始前までの動作を示すフローチャート。

【図７】同実施の形態に係るビューワの再生開始後の動作を示すフローチャート。

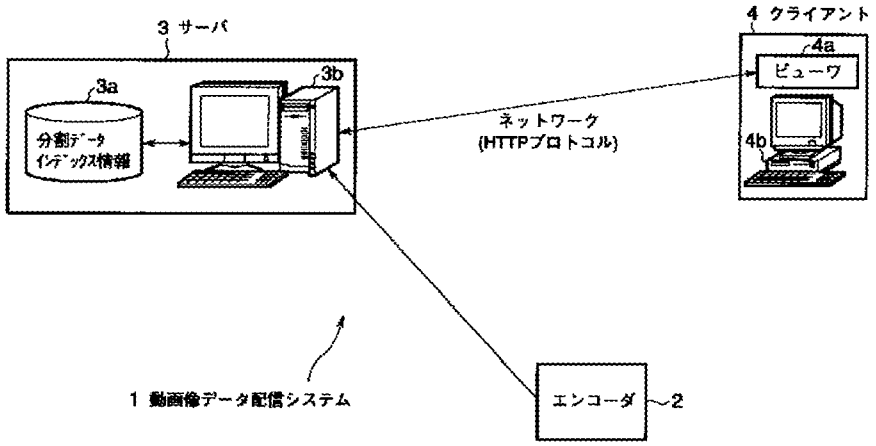
【図８】本発明の第３の実施の形態に係る動画像データ配信システムに備えられるエンコーダの機能を示すブロック図。

【図９】同実施の形態に係る動画像データ配信システムに備えられるビューワの再生開始後の動作を示すフローチャート。

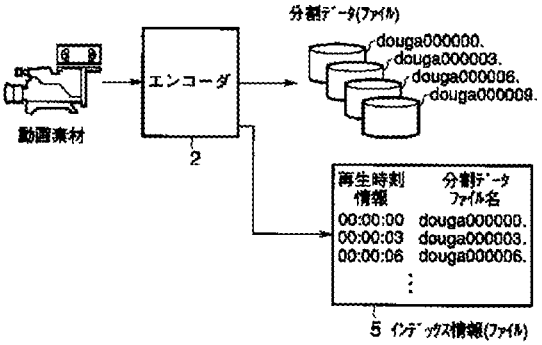
【符号の説明】

１…動画像データ配信システム
２、１３…エンコーダ
３…サーバ
４…クライアント
４ａ、１０…ビューワ
５、１４…インデックス情報
６…インデックス情報受信部
７、１１…分割データ受信部
８…ページバッファ
９…デコーダ
１２…時間推定部

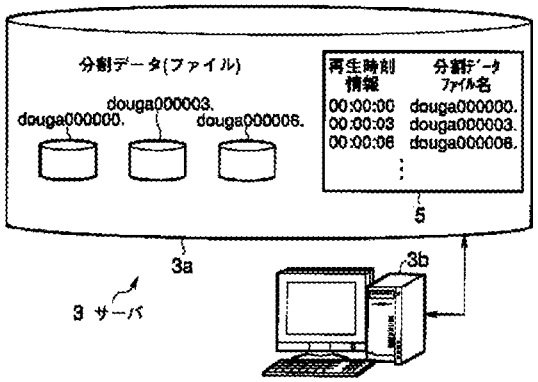
【図１】



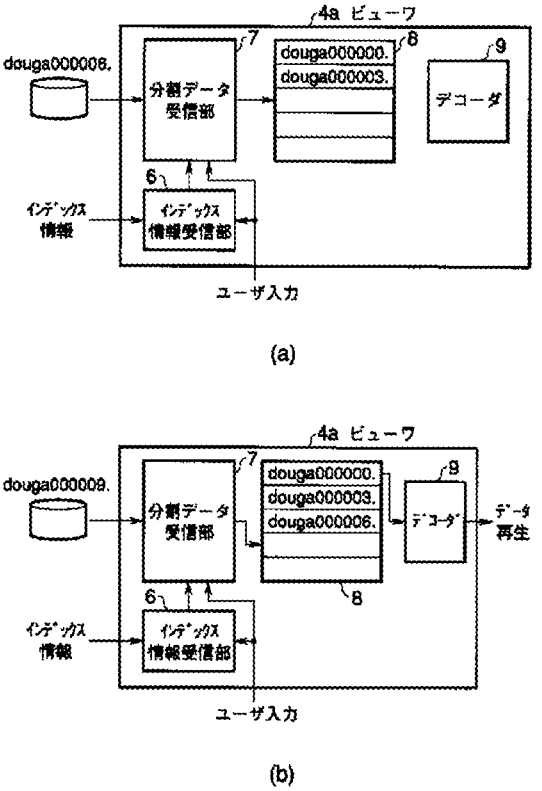
【図２】


分割データ(ファイル)
douga000000.
douga000003.
douga000006.
douga000009.
エンコーダ
動画素材
2
再生時刻情報
分割データファイル名
00:00:00 douga000000.
00:00:03 douga000003.
00:00:06 douga000006.
5 インデックス情報(ファイル)

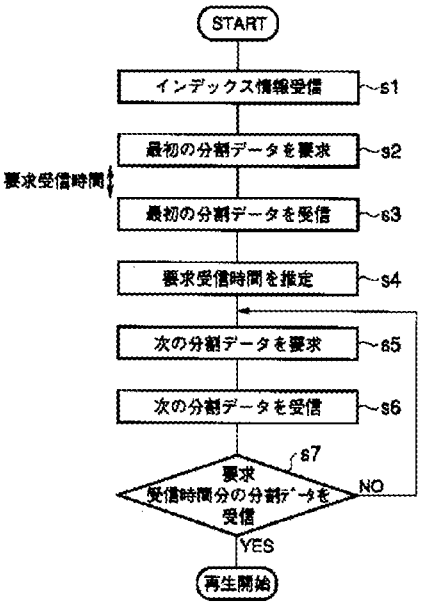
【図３】


分割データ(ファイル)
douga000000.
douga000003.
douga000006.
再生時刻情報
分割データファイル名
00:00:00 douga000000.
00:00:03 douga000003.
00:00:06 douga000006.
5
3a
3b
3 サーバ

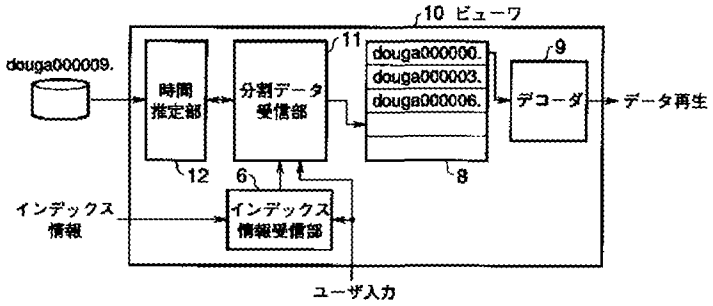
【図４】


4a ビューワ
douga000006.
分割データ受信部
7
douga000000.
douga000003.
8
デコーダ
9
6
インデックス情報
インデックス情報受信部
ユーザ入力
(a)
4a ビューワ
douga000009.
分割データ受信部
7
douga000000.
douga000003.
douga000006.
8
デコーダ
9
データ再生
6
インデックス情報
インデックス情報受信部
ユーザ入力
(b)

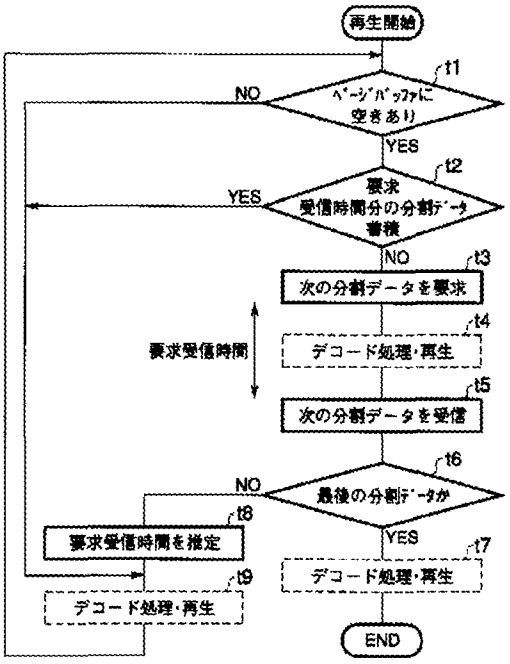
【図６】


START
インデックス情報受信
s1
最初の分割データを要求
s2
要求受信時間
最初の分割データを受信
s3
要求受信時間を推定
s4
次の分割データを要求
s5
次の分割データを受信
s6
s7
要求受信時間分の分割データを受信
NO
YES
再生開始

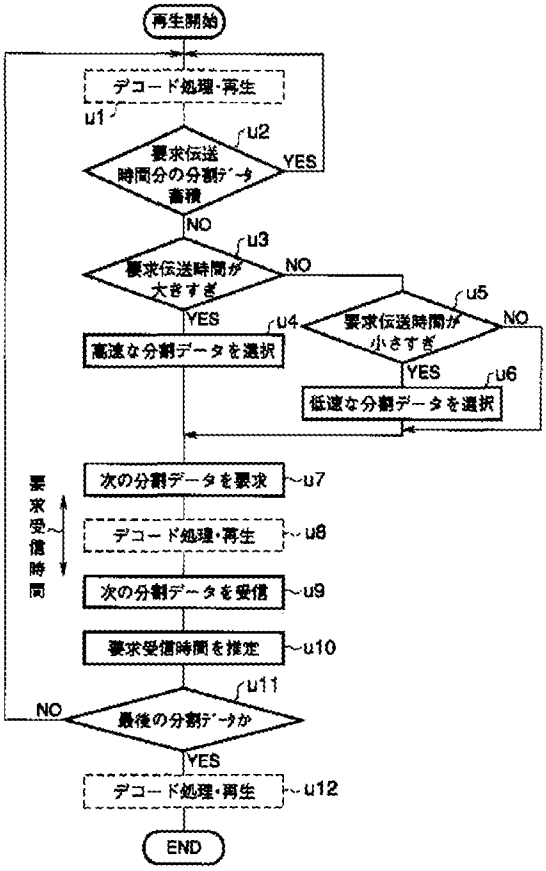
【図５】


10 ビューワ
douga000009.
時間推定部
分割データ受信部
11
douga000000.
douga000003.
douga000006.
9
デコーダ
データ再生
12
6
8
インデックス情報
インデックス情報受信部
ユーザ入力

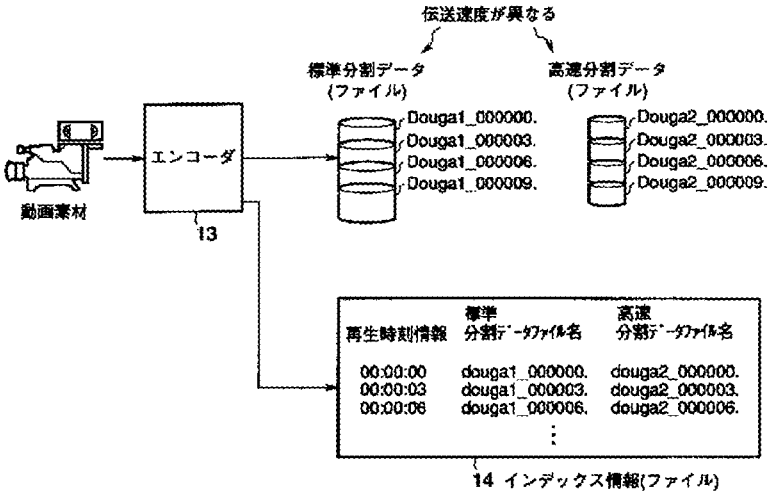
【図７】


再生開始
t1
NO
ページバッファに空きあり
YES
t2
YES
要求受信時間分の分割データ蓄積
NO
t3
次の分割データを要求
要求受信時間
t4
デコード処理・再生
t5
次の分割データを受信
t6
NO
最後の分割データか
YES
t8
要求受信時間を推定
t7
デコード処理・再生
t9
デコード処理・再生
END

【図９】


再生開始
デコード処理・再生
u1
u2
要求伝送時間分の分割データ蓄積
YES
NO
u3
要求伝送時間が大きすぎ
NO
YES
u4
高速な分割データを選択
u5
要求伝送時間が小さすぎ
NO
YES
u6
低速な分割データを選択
要求受信時間
次の分割データを要求
u7
デコード処理・再生
u8
次の分割データを受信
u9
要求受信時間を推定
u10
u11
NO
最後の分割データか
YES
デコード処理・再生
u12
END

【図８】


伝送速度が異なる
標準分割データ
(ファイル)
高速分割データ
(ファイル)
Douga1_000000.
Douga1_000003.
Douga1_000006.
Douga1_000009.
Douga2_000000.
Douga2_000003.
Douga2_000006.
Douga2_000009.
エンコーダ
動画素材
13
再生時刻情報
標準
分割データファイル名
高速
分割データファイル名
00:00:00
00:00:03
00:00:06
douga1_000000.
douga1_000003.
douga1_000006.
douga2_000000.
douga2_000003.
douga2_000006.
14 インデックス情報(ファイル)

PTO/SB/08a (01-10)
Approved for use through 07/31/2012. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	11116783
Filing Date	2005-04-28
First Named Inventor	R. Drew Major
Art Unit	2444
Examiner Name	Donabed, Ninos J.
Attorney Docket Number	066.0141 (P2004-04-01.1)

U.S.PATENTS Remove

Examiner Initial*	Cite No	Patent Number	Kind Code[1]	Issue Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	7523181	B2	2009-04-21	Swildens et al.	
	2	6976090	B2	2005-12-13	Ben-Shaul et al.	
	3	7240100	B1	2007-07-03	Wein et al.	
	4	6850965	B2	2005-02-01	Allen	
	5	7334044	B1	2008-02-19	Allen	
	6	7760801	B2	2010-07-20	Ghanbari et al.	
	7	7099954	B2	2006-08-29	Li et al.	
	8	7054911	B1	2006-05-30	Lango et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	9	7349358	B2	2008-03-25	Hennessey et al.	
	10	7797439	B2	2010-09-14	Cherkasova et al.	
	11	7477688	B1	2009-01-13	Zhang et al.	
	12	7054774	B2	2006-05-30	Batterberry et al.	
	13	6760772	B2	2004-07-06	Zou et al.	
	14	5168356		1992-12-01	Acampora et al.	
	15	7546355	B2	2009-06-09	Kalnitsky	
	16	7174385	B2	2007-02-06	Li	
	17	7376747	B2	2008-05-20	Hartop	
	18	7719985	B2	2010-05-18	Lee et al.	
	19	8683066	B2	2014-03-25	Hurst et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	20	5768527		1998-06-16	Zhu et al.	
	21	7274740	B2	2007-09-25	van Beek et al.	
	22	6185736	B1	2001-02-06	Ueno	
	23	6845107	B1	2005-01-18	Kitazawa et al.	
	24	6874015	B2	2005-03-29	Kaminsky et al.	
	25	7093001	B2	2006-08-15	Yang et al.	
	26	7599307	B2	2009-10-06	Seckni et al.	
	27	7788395	B2	2010-08-31	Bowra et al.	
	28	7536469	B2	2009-05-19	Chou et al.	
	29	7412531	B2	2008-08-12	Lango et al.	
	30	7295520	B2	2007-11-13	Lee et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	31	6795863	B1	2004-09-21	Doty, Jr.	
	32	7609652	B2	2009-10-27	Kellerer et al.	
	33	7325073	B2	2008-01-29	Shao et al.	
	34	6374289	B2	2002-04-16	Delaney et al.	
	35	7194549	B1	2007-03-20	Lee et al.	
	36	7054365	B2	2006-05-30	Kim et al.	
	37	6122660		2000-09-19	Baransky et al.	
	38	7391717	B2	2008-06-24	Klemets et al.	
	39	7260640	B1	2007-08-21	Kramer et al.	
	40	7310678	B2	2007-12-18	Gunaseelan et al.	
	41	7593333	B2	2009-09-22	Li et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	42	6618752	B1	2003-09-09	Moore et al.	
	43	6721723	B1	2004-04-13	Gibson et al.	
	44	6859839	B1	2005-02-22	Zahorjan et al.	
	45	7779135	B2	2010-08-17	Hudson et al.	
	46	7369610		2008-05-06	Xu et al.	
	47	7330908		2008-02-12	Jungck	
	48	7577750		2009-08-18	Shen et al.	
	49	7328243		2008-02-05	Yeager et al.	
	50	7116894		2006-10-03	Chatterton	
	51	7653735		2010-01-26	Mandato et al.	
	52	8135852	B2	2012-03-13	Nilsson et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	53	5841432		1998-11-24	Carmel et al.	
	54	5404446		1995-04-04	Bowater et al.	
	55	6449719	B1	2002-09-10	Baker	
	56	7817985	B2	2010-10-19	Moon	
	57	8209429	B2	2012-06-26	Jacobs et al.	
	58	7733830	B2	2010-06-08	Curcio et al.	
	59	7660906	B1	2010-02-09	Armour	
	60	8036265	B1	2011-10-11	Reynolds et al.	
	61	5267334		1993-11-30	Normille et al.	
	62	8370514	B2	2013-02-05	Hurst et al.	
	63	8612624	B2	2013-12-17	Frueck et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	64	6654790	B2	2003-11-25	Ogle et al.	
	65	6675199	B1	2004-01-06	Mohammed et al.	
	66	6697072	B2	2004-02-24	Russell et al.	
	67	6792449	B2	2004-09-14	Colville et al.	
	68	6801947	B1	2004-10-05	Li	
	69	6885471	B1	2005-04-26	Minowa et al.	
	70	7031700	B1	2006-04-18	Weaver et al.	
	71	7046805	B2	2006-05-16	Fitzhardinge et al.	
	72	7111044	B2	2006-09-19	Lee	
	73	7124164	B1	2006-10-17	Chemtob	
	74	7176957	B2	2007-02-13	Ivashin et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	75	7177642	B2	2007-02-13	Sanchez Herrero et al.	
	76	7190670	B2	2007-03-13	Varsa et al.	
	77	7313236	B2	2007-12-25	Amini et al.	
	78	7529541	B2	2009-05-05	Cho et al.	
	79	6552227	B2	2003-04-22	Mendelovici et al.	
	80	7555464	B2	2009-06-30	Candelore	
	81	7558472	B2	2009-07-07	Locket et al.	
	82	7631039	B2	2009-12-08	Eisenberg	
	83	7761609	B1	2010-07-20	Srinivasan et al.	
	84	7873040	B2	2011-01-18	Karlsgodt	
	85	7818444	B2	2010-10-19	Brueck et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	86	8402156	B2	2013-03-19	Brueck et al	

If you wish to add additional U.S. Patent citation information please click the Add button. Add

U.S.PATENT APPLICATION PUBLICATIONS Remove

Examiner Initial*	Cite No	Publication Number	Kind Code[1]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear
	1	20040083283	A1	2004-04-29	Sundaram et al.	
	2	20080219151	A1	2008-09-11	Ma et al.	
	3	20020152318	A1	2002-10-17	Menon et al.	
	4	20060080718	A1	2006-04-13	Gray et al.	
	5	20050120107	A1	2005-06-02	Kagan et al.	
	6	20040103444	A1	2004-05-27	Weinberg et al.	
	7	20050188051	A1	2005-08-25	Sneh	
	8	20030154239	A1	2003-08-14	Davis et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	9	20080133766	A1	2008-06-05	Luo	
	10	20070079325	A1	2007-04-05	de Heer	
	11	20030014684	A1	2003-01-16	Kashyap	
	12	20060206246	A1	2006-09-14	Walker	
	13	20050076136	A1	2005-04-07	Cho et al.	
	14	20030018966	A1	2003-01-23	Cook et al.	
	15	20040030599	A1	2004-02-12	Sie et al.	
	16	20050262257	A1	2005-11-24	Major et al.	
	17	20030140159	A1	2003-07-24	Campbell et al.	
	18	20010047423	A1	2001-11-29	Shao et al.	
	19	20030021166	A1	2003-01-30	Soloff	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	20	20020176418	A1	2002-11-28	Hunt et al.	
	21	20020169926	A1	2002-11-14	Pinckney, III et al.	
	22	20020178330	A1	2002-11-28	Schlowsky-Fischer et al.	
	23	20050084166	A1	2005-04-21	Bonch et al.	
	24	20020091840	A1	2002-07-11	Pulier et al.	
	25	20040117427	A1	2004-06-17	Allen et al.	
	26	20040030547	A1	2004-02-12	Leaning et al.	
	27	20060165166	A1	2006-07-27	Chou et al.	
	28	20020097750	A1	2002-07-25	Gunaseelan et al.	
	29	20020152317	A1	2002-10-17	Wang et al.	
	30	20080037527	A1	2008-02-14	Chan et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	31	20040168052	A1	2004-08-26	Clisham et al.	
	32	20050066063	A1	2005-03-24	Grigorovitch et al.	
	33	20050204046	A1	2005-09-15	Watanabe	
	34	20060059223	A1	2006-03-16	Klemets et al.	
	35	20060168295	A1	2006-07-27	Batterberry et al.	
	36	20080281803	A1	2008-11-13	Gentric	
	37	20050185578	A1	2005-08-25	Padmanabhan et al.	
	38	20070280255	A1	2007-12-06	Tsang et al.	
	39	20100098103	A1	2010-04-22	Xiong et al.	
	40	20090210549	A1	2009-08-20	Hudson et al.	
	41	20090055547	A1	2009-02-26	Hudson et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	42	20080205291	A1	2008-08-28	Li et al.	
	43	20070094405	A1	2007-04-26	Zhang	
	44	20060184688	A1	2006-08-17	Ganguly et al.	
	45	20080263180	A1	2008-10-23	Hurst et al.	
	46	20090055471	A1	2009-02-26	Kozat et al.	
	47	20040260701	A1	2004-12-23	Lehikoinen et al.	
	48	20060168290	A1	2006-07-27	Doron	
	49	20070030833	A1	2007-02-08	Pirzada et al.	
	50	20070067480	A1	2007-03-22	Beek et al.	
	51	20060133809	A1	2006-06-22	Chow et al.	
	52	20040049780	A1	2004-03-11	Gee	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	53	20030065803	A1	2003-04-03	Heuvelman	
	54	20020161911	A1	2002-10-31	Pinckney, III et al.	
	55	20060236219	A1	2006-10-19	Grigorovitch et al.	
	56	20080046939	A1	2008-02-21	Lu et al.	
	57	20060130118	A1	2006-06-15	Damm	
	58	20070204310	A1	2007-08-30	Hua et al.	
	59	20020176418	A1	2002-11-28	Hunt et al.	
	60	20050223087	A1	2005-10-06	Van Der Stok	
	61	20040010613	A1	2004-01-15	Apostolopoulos et al.	
	62	20030204602	A1	2003-10-30	Hudson et al.	
	63	20010013128	A1	2001-08-09	Hagai et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	64	20050055425		2005-03-10	Lango et al.	
	65	20050015509		2005-01-20	Sitaraman, Ramesh K.	
	66	20020131496		2002-09-19	Vasudevan et al.	
	67	20020161898		2002-10-31	Hartop et al.	
	68	20040220926		2004-11-04	Lamkin et al.	
	69	20100158101	A1	2010-06-24	Wu et al.	
	70	20090132721	A1	2009-05-21	Soroushian et al.	
	71	20090132599	A1	2009-05-21	Soroushian et al.	
	72	20090055417	A1	2009-02-26	Hannuksela	
	73	20090049186	A1	2009-02-19	Agnihotri et al.	
	74	20090006538	A1	2009-01-01	Risney, JR. et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	75	20060277564	A1	2006-12-07	Jarman	
	76	20040153458	A1	2004-08-05	Noble et al.	
	77	20020087634	A1	2002-07-04	Ogle et al.	
	78	20020118809	A1	2002-08-29	Eisenberg	
	79	20020122491	A1	2002-09-05	Karczewicz et al.	
	80	20020133547	A1	2002-09-19	Lin	
	81	20020136406	A1	2002-09-26	Fitzhardinge et al.	
	82	20020138619	A1	2002-09-26	Ramaley et al.	
	83	20020146102	A1	2002-10-10	Lang	
	84	20020178138	A1	2002-11-28	Ender et al.	
	85	20020184391	A1	2002-12-05	Phillips	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	86	20020194608	A1	2002-12-19	Goldhor	
	87	20030007464	A1	2003-01-09	Balani	
	88	20030037103	A1	2003-02-20	Salmi et al.	
	89	20030078972	A1	2003-04-24	Tapissier et al.	
	90	20030220972	A1	2003-11-27	Montet et al.	
	91	20040098748	A1	2004-05-20	Bo et al.	
	92	20050009520	A1	2005-01-13	Herrero et al.	
	93	20050024487	A1	2005-02-03	Chen	
	94	20050050152	A1	2005-03-03	Penner et al.	
	95	20050066063	A1	2005-03-24	Grigorovitch et al.	
	96	20050204385	A1	2005-09-15	Sull et al.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	97	20050254508	A1	2005-11-17	Aksu et al.	
	98	20060047779	A1	2006-03-02	Deshpande	
	99	20060218264	A1	2006-09-28	Ogawa et al.	
	100	20060242315	A1	2006-10-26	Nichols	
	101	20060270404	A1	2006-11-30	Tuohino et al.	
	102	20060282540	A1	2006-12-14	Tanimoto	
	103	20060288099	A1	2006-12-21	Jefferson et al.	
	104	20070037599	A1	2007-02-15	Tillet et al.	
	105	20070078768	A1	2007-04-05	Dawson	
	106	20080060029	A1	2008-03-06	Park et al.	
	107	20080091838	A1	2008-04-17	Miceli	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	108	20140207966	A1	2014-07-24	Hurst et al.	

If you wish to add additional U.S. Published Application citation information please click the Add button. Add

FOREIGN PATENT DOCUMENTS Remove

Examiner Initial*	Cite No	Foreign Document Number[3]	Country Code[2] i	Kind Code[4]	Publication Date	Name of Patentee or Applicant of cited Document	Pages,Columns,Lines where Relevant Passages or Relevant Figures Appear	T[5]
	1	2006010113	WO	A2	2006-01-26	Network Foundation Technologies, LLC		☐
	2	200192752	JP		2001-04-06	Hitachi Information Systems Ltd.		☒
	3	2003009581	WO	A1	2003-01-30	British Telecommunications Public Limited Company		☐
	4	2004036824	WO	A1	2004-04-29	Nokia corporation		☐
	5	2004054930	JP		2004-02-19	Microsoft Corporation		☒
	6	2001067264	WO	A1	2001-09-13	Streamcenter, Inc.		☐
	7	2003003760	WO	A2	2003-01-09	Burst.com		☐
	8	2003027876	WO	A1	2003-04-03	Reynolds et al.		☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	9	2004025405	WO	A2	2004-03-25	Thomson Licensing S.A.		☐
	10	1298931	EP	A2	2003-02-04	Oplayo		☐
	11	1202487	EP	A2	2001-10-31	Kabushiki Kaisha Toshiba		☐
	12	1395014	EP	A1	2002-08-27	Matsushita Electric Industrial Co.		☐
	13	2367219	GB	A	2000-09-20	Vintage Global		☐
	14	2011004225	JP	A	2011-01-06	NEC Infronta Corp.		☒
	15	2000201343	JP	A	2000-07-18	Toshiba Corp.		☒

If you wish to add additional Foreign Patent Document citation information please click the Add button Add

NON-PATENT LITERATURE DOCUMENTS Remove

Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.	T[5]
	1	USPTO "International Search Report" mailed December 12, 2008; International Appln. No. PCT/US2008/061035, filed April 21, 2008.	☐
	2	AUSTRALIAN GOVERNMENT "Examiner's First Report" dated October 17, 2011; Australian Patent Appln. No. 2011213730.	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	3	KOREAN INTELLECTUAL PROPERTY OFFICE "Official Notice of Preliminary Rejection" issued July 28, 2011; Korean Patent Appln. No. 10-2006-7025274.	☒
	4	JAPAN PATENT OFFICE "Notice of Rejection Ground" mailed April 26, 2011; Japanese Patent Appln. No. 2007-511070.	☒
	5	FUJISAWA, HIROSHI ET AL. "Implementation of Efficient Access Mechanism for Multiple Mirror-Servers" IPSJ SIG Technical Report, vol. 2004, no. 9 (2004-DPS-116), January 30, 2004, Information Processing Society of Japan, pp. 37-42	☒
	6	LIU, JIANGCHUAN ET AL. “Opportunities and Challenged of Peer-to-Peer Internet Video Broadcast,” School of Computing Science, Simon Fraser University, British Columbia, Canada	☐
	7	USPTO INTERNATIONAL SEARCHING AUTHORITY “International Search Report and Written Opinion,” mailed November 5, 2008; International Appln. No. PCT/US2008/009281, filed August 1, 2008.	☐
	8	ZHANG, XINYAN ET AL. "CoolStreaming/DONet: A Data-Driven Overlay Network for Peer-to-Peer Live Media Streaming" IEEE 2005.	☐
	9	GUO, YANG "DirectStream: A Directory-Based Peer-To-Peer Video Streaming Service" LexisNexis, Elsevier B.V. 2007.	☐
	10	LIU, JIANGCHUAN ET AL. "Adaptive Video Multicast Over the Internet" IEEE Computer Society, 2003.	☐
	11	REJAIE, REZA ET AL. "Architectural Considerations for Playback of Quality Adaptive Video Over the Internet" University of Southern California, Information Sciences Institute, 1998.)	☐
	12	ROY, SUMIT ET AL. "A System Architecture for Managing Mobile Streaming Media Services" Streaming Media Systems Group, Hewlett-Packard Laboratories, 2003.	☐
	13	XU, DONGYAN ET AL. "On Peer-to-Peer Media Streaming" Department of Computer Sciences, Purdue University, 2002.	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	14	KOZAMERNIK, FRANC "Media Streaming Over the Internet - An Over of Delivery Technologies" EBU Technical Review, October 2002.	☐
	15	LIENHART, RAINER ET AL. "Challenges in Distributed Video Management and Delivery" Intel Corporation, EECS Dept., UC Berkeley, 2000-2002.	☐
	16	JAPAN PATENT OFFICE "Final Office Action" mailed February 28, 2012 in Patent Application No. 2007-511070 filed on October 26, 2006.	☒
	17	JAPAN PATENT OFFICE "Interrogation" mailed November 06, 2012 in Patent Application No. 2007-511070 filed on October 26, 2006.	☒
	18	CANADIAN INTELLECTUAL PROPERTY OFFICE "Office Action" mailed September 09, 2013 in Patent Application No. 2,564,861 filed on October 30, 2006.	☐
	19	USPTO "Office Action" mailed September 13, 2013 in Patent Application No. 13/757,571 filed on February 01, 2013.	☐
	20	USPTO "Notice of Allowance" mailed June 24, 2014 in Patent Application No. 13/757.571, filed on February 01, 2013.	☐
	21	EUROPEAN PATENT OFFICE "Extended Search Report" dated July 10, 2014 in Patent Application No. 12154559.4 filed on September 20, 2002.	☐
	22	NGUYEN, Thinh, "Multiple Sender Distributed Video Streaming" in IEEE Transactions on Multimedia, Volume 6, No. 2, Published April 2, 2004.	☐
	23	WEBLIO, The Meaning of Performance Factor --- English-Japanese Weblio Dictionary, [online], February 24, 2012; retrieved from the internet - URL:http://ejje, weblio.jp/content/performance+factor	☐
	24	MASATO TSURU ET AL., Recent Evolution of the Internet Measurement and Inference Techniques, IEICE Technical Report, vol. 103, no. 123 (IN2003-16 to 23), IEICE, June 12, 2003, pp. 37 to 42, ISSN: 0913-05685.	☐

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)		
	Application Number	11116783
	Filing Date	2005-04-28
	First Named Inventor	R. Drew Major
	Art Unit	2444
	Examiner Name	Donabed, Ninos J.
	Attorney Docket Number	066.0141 (P2004-04-01.1)

	25	TAKESHI YOSHIMURA ET AL., Mobile Streaming Media CDN Enabled by Dynamic SMIL, WWW2002, May 7-11, 2002; retrieved from the Internet at http://www2002.org/CDROM/refereed/515/	☐
	26	CANADIAN INTELLECTUAL PROPERTY OFFICE, Office Action, mailed October 15, 2012 for Patent Application No. 2,564,861.	☐
	27	CLEMENT, B., Move Networks Closes $11.3 Million on First Round VC Funding, Page One PR, Move Networks, Inc. Press Releases, February 7, 2007, http://www.move.tv/press/press20070201.html.	☐
	28	MOVE NETWORKS, INC., The Next Generation Video Publishing System, April 11, 2007; http://www.movenetworks.com/wp-content/uploads/move-networks-publishing-system.pdf.	☐
	29	U.S. Patent and Trademark Office, Non-Final Office Action, dated August 7, 2014 for U.S. Patent Application No. 14/106,051.	☐

If you wish to add additional non-patent literature document citation information please click the Add button Add

EXAMINER SIGNATURE

Examiner Signature		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. [2] Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). [3] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [4] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [5] Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Not for submission under 37 CFR 1.99)

Application Number	11116783
Filing Date	2005-04-28
First Named Inventor	R. Drew Major
Art Unit	2444
Examiner Name	Donabed, Ninos J.
Attorney Docket Number	066.0141 (P2004-04-01.1)

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

☐ That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

☒ See attached certification statement.

☐ Fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☐ None

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/BRETT A. CARLSON/	Date (YYYY-MM-DD)	2014-08-20
Name/Print	Brett A. Carlson	Registration Number	39928

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these record s.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 298 931 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.04.2003 Bulletin 2003/14

(51) Int Cl.7: **H04N 7/24**

(21) Application number: **02396145.1**

(22) Date of filing: **20.09.2002**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **20.09.2001 FI 20011856**

(71) Applicant: **Oplayo Oy**
00180 Helsinki (FI)

(72) Inventor: **Ala-Honkola, Pekka**
00180 Helsinki (FI)

(74) Representative: **Savela, Antti-Jussi**
Patent Agency Compatent Ltd.
Hämeentie 29, 4th. Fl.
00500 Helsinki (FI)

(54) Adaptive media stream

(57) This invention relates to media streams transmitted in a network, such as the Internet. The idea is to change the source of the content of a media stream when an available transmission speed rate changes. Each source contains essentially the same information (such as video and voice), but the suitability of each source for transmission has been adjusted to a certain speed rate.

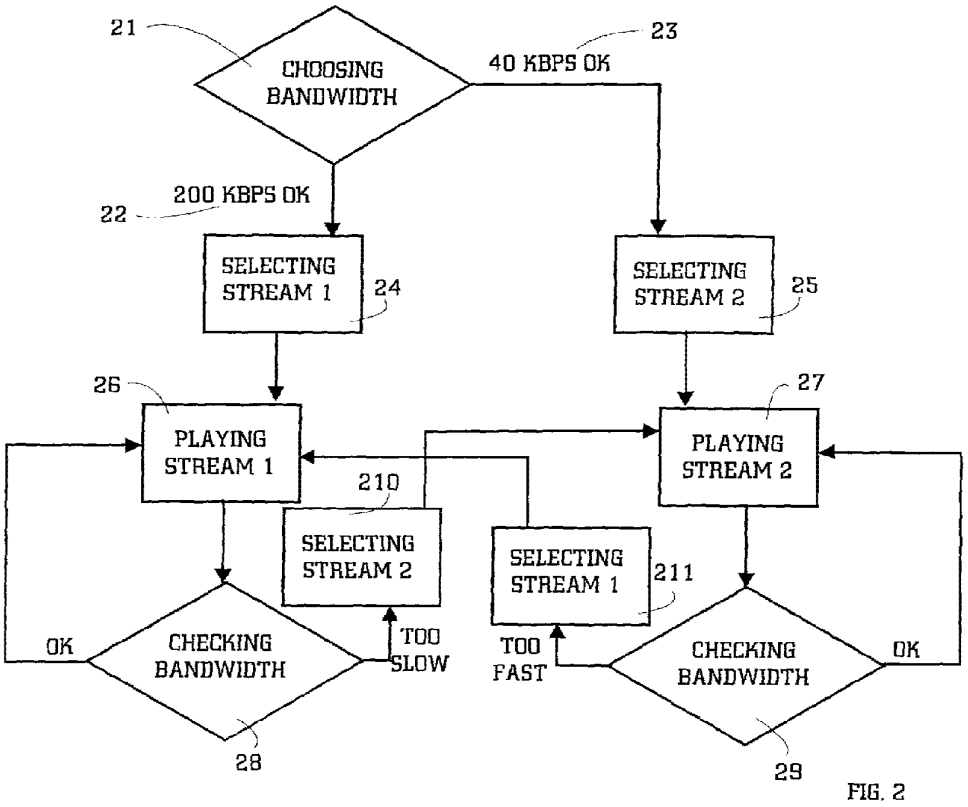


FIG. 2

Printed by Jouve, 75001 PARIS (FR)

EP 1 298 931 A2

Description

FIELD OF THE INVENTION

[0001] This invention relates to media streams transmitted in a network, such as the Internet. Especially, the invention relates to media streams concerning video pictures and voice.

BACKGROUND OF THE INVENTION

[0002] When media streams that contain continuous data, for example video pictures and voice, are transmitted in a network, such as the Internet, the speeds of the media streams have to be adjusted with the speed of the transmission connection available at the time. Often, in the case of video pictures, the adjustment is made by changing the visual quality of the pictures. The speed of a transmission connection in the Internet depends strongly on a user's Internet terminal. Some terminals support technically higher speed rates than others (usually older terminals). Thus, speed adjusted media streams must be offered for a variety of terminals.

[0003] Transmission speed may also vary with time. For example, when congestion happens in the network, the transmission speed may decrease considerably. After the congestion, the transmission speed returns to the normal speed rate.

[0004] Nowadays, most of the content of the Internet is in WWW servers, which use the HTTP protocol for transmitting desired contents to users. Originally, the HTTP protocol was not designed for continuous media streams (video, voice). The problem of the known solutions is that when a media stream has to be adjusted for a changed transmission condition, i.e. to be an adaptive media stream, a special software or protocol must be used, instead of the HTTP protocol. Real Server® (RealNetworks Inc, Seattle, Washington), Quicktime® Server (Apple Computers, Cupertino California), and Windows Media Server® (Microsoft, Redmond, Washington) are commercial server software, which support adaptive media streams. These software packages use special protocols, such as RTP (Real-Time Transport Protocol) and RTSP (Real Time Streaming Protocol), which are designed for the adaptive media streams.

[0005] However, using these software packages usually requires a user to invest and maintain parallel devices and software in addition to the WWW server that is used to serve other Internet content, such as web pages. User communities such as corporations require support staff which must be educated to use these systems. For a service provider this means in practice that a special media stream server must be maintained. The goal of this invention is to eliminate these drawbacks. This is achieved in a way described in the claims.

SUMMARY OF THE INVENTION

[0006] The following definitions should be kept in mind when reading these specifications:

Connection: A transport layer virtual circuit established between two programs for the purpose of communication. A connection can have varying transmission speed rates.

HTTP transport: An application layer transmission of data from a file on a server computer to a user terminal using HTTP 1.1 protocol. One HTTP 1.1 connection can consist of one or several HTTP 1.1 transports.

Stream: A continuous transmission of data over for example HTTP transport

[0007] An important aspect of the invention is to change the source of an adaptive media stream when an available connection speed rate changes. Each source contains essentially the same information (such as video and voice), but the suitability of each source for transmission has been adjusted to a certain connection speed rate. The coding of the source information can be different among the sources, which affects the speed rate. The picture information can be black-and-white instead of a color picture. The picture size can be smaller or larger. The video can be with or without voice. When the source is changed, also a new HTTP transmission for the new source is started for a new stream containing the information of the new source. In other words, the adaptive media stream, i.e. the adaptive stream, comprises at least one media stream for transmitting information.

[0008] In an arrangement according to the invention a source file is formed. The source file contains a plurality of sources intended for different media streams, according to such parameters as bandwidth, computing capability of the receiving terminal, etc. At the beginning of the transmission the optimal transmission speed rate at the time is chosen according to an information transfer condition. The source in the source file, which corresponds to the available speed rate, is selected to be the source of the adaptive media stream. When the available transmission speed rate changes, either to be faster or slower, a command for changing the source of the adaptive media stream, is sent from the user's terminal to the service provider's server, i.e. the source server, wherein the file of the service is. Preferably commands are sent using an HTTP protocol. When the available transmission speed rate changes again, a new command for changing the source is sent from the user's terminal to the source server. The commands are sent when it is possible to use a higher speed rate or when a lower speed rate must be used.

[0009] In one aspect, the invention thus provides a method for providing an adaptive media stream between a sending terminal and a user's terminal wherein

Aylo Exhibit 1018
Page 1909 of 2308

the adaptive media stream comprises at least one media stream and the method comprises the steps of choosing a transmission speed rate for the media stream, selecting a suitable data source for the media stream among sources containing essentially the same information content, each source intended for a certain information transfer condition and playing the media stream. During such playing, checking the suitability of the actual transmission speed rate, and continuing playing the media stream if the transmission speed is still suitable or selecting a new data source, which is suitable at the moment as a response for the selection of the new data source, playing a new media stream. Those selection actions continue until the played stream stops.

[0010] Optionally, the step of choosing comprises the phase of reading a header field of a file, which contains at least two data sources, the header indicating the starting points of the sources and to which transmission speed rate each source is indented. Yet another option calls for the selecting step to include the phase of starting to read from the starting point of the selected source.

[0011] Other aspect of the invention provide for the step of monitoring to include the phase of monitoring a filling degree of a buffer for incoming data in the user's terminal, e.g. if the buffer is becoming empty or full, then selecting to continue to play the present media stream or to start playing another media stream. The information of the selection of the new data source may transmitted from the user's terminal to the sending terminal using a GET method of an HTTP protocol with an optional RANGE field. When a new stream is played, an optional phases of terminating the old media stream after starting to play the new media stream and creating the new media stream before playing the new media stream is also contemplated.

[0012] If a new stream is selected, according to another aspect of the invention, the step of reading timing marks in said adaptive media stream for defining the right point to start playing the new media stream is also performed. Further, the step of selecting the initial speed may comprise the phase of reading a header field of files, where each contains one data source, the header indicating the starting points of the source and to which transmission speed rate the source is indented.

[0013] An arrangement for providing an adaptive media stream between a sending terminal and a user's terminal, is also contemplated, wherein the adaptive media stream comprises at least one media stream and the arrangement comprises:

a) Means in the client terminal for choosing a transmission speed rate for the media stream,
b) A data structure in the sending terminal comprising at least two data sources containing essentially the same information, each source intended for a certain transmission speed rate,
c) Means in the client terminal and the sending terminal for selecting a suitable data source from among the sources for the media stream,
d) Means in the user's terminal for checking the suitability of the transmission speed rate when the media stream is played,
e) Means in the user's terminal for reselecting a suitable data source from among the sources for a new media stream.

[0014] An arrangement as described above may further comprise means for playing the data structure in a way that the playing is started from the right point of the data structure. Preferably, means a) comprises means for reading a message from the user's terminal, which defines alternative transmission speed rates. Further preferably, each source comprises timing marks, which indicate the corresponding point in the other sources. More preferably, the arrangement further comprises a GET method of an HTTP protocol with an optional RANGE field for transmitting the information of the reselection of the suitable data source from the user's terminal to the sending terminal. In another preferable embodiment, means d) comprises means for monitoring a filling degree of a buffer for incoming data in the user's terminal. It also preferable that the data structure is a file containing the data sources. Alternatively, the data structure may be a number of files, each of them containing one data source.

[0015] Means c) may further comprise means for reading a header of the data structure, which indicates the starting points of the sources and to which transmission speed rate each source is intended. Means c) may further comprise means for reading headers of the files, which indicate the starting point of the source and to which transmission speed rate the source is intended.

[0016] In yet another aspect, the invention teaches a method for storing adaptive media stream data comprising the steps of storing a plurality of media streams, each representing a substantially similar information content, and each adapted for a different set of information transfer conditions; and providing a plurality of pointers into a first of said media streams, said pointers associated with a source location in one of said media streams and pointing to at least one target location in a second of said media streams; wherein the information content of said target location is substantially a continuation of the information content prior to said source location.

[0017] Preferably, said plurality of media streams is contained in a single file. Said pointers are preferably embedded within said adaptive media stream data, and comprise target locations pointer comprises a byte offset into said single file. Alternatively, said target location pointer comprises an offset into said second source stream, within said single file. Another alternative teaches that said target locations pointers comprise a pointer to another file containing said second media stream or a portion thereof. Optionally in this case, said target location pointer further comprises a byte offset into said

Aylo Exhibit 1018
Page 1910 of 2308

other file.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] In the following the invention is described in more detail by means of Figs. **1 - 3** in the attached drawings where.

FIG. **1** illustrates an example of the structure of a file according to the invention,
FIG. **2** illustrates an example of a flow chart describing the function of the inventive arrangement,
FIG. **3** illustrates an example of an arrangement according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0019] FIG. **1** describes an example of a file structure according to the invention. The file is in the service provider's server (source server) from where users can stream it. For example the file contains a video in two different formats. At the beginning of the file there exists a header **1**, which contains starting bytes for two different media sources - the sources for the media stream. Source 1, called stream 1 contains video data in a resolution of 320*240, for a speed rate 200 kbps (kilobit per second) **2.** The starting byte **4** for that video data is byte 2000 in the file. Source 2, called stream 2 contains video data in a resolution of 160*120 optimized for a speed rate 40 kbps **3.** The starting byte of that video data **5** in the file is at byte 800000. If the network and a user's terminal support the transmission speed rate 200 kbps then stream 1 is sent to the user - otherwise stream 2 is sent. In this embodiment, all data feeds are assumed to support communication at a speed of at least 40 kbps. It should be noted that stream 1 and 2 are sources i.e. the parts of the file structures in this example, and a media stream is a continuous transmission of data of a source over a transmission protocol. The transmission protocol used can be any protocol, in this text the HTTP protocol is used.

[0020] Stream 1 is sent, i.e. streamed, to the user. Stream 1 (and the media stream for it as well) contains timing marks **6, 7, 8, 9.** The timing marks are embedded in stream 1 at a various locations, and correlate the information in the two separate resolution streams. Thus the timing marks of stream 1 in the example of Fig. 1 indicate the byte of the file where the stream is running at the time and the corresponding byte **6A, 7A, 8A, 9A,** in stream 2. If for some reason, the network can not transmit stream 1 with 200 kbps, or the user's terminal can not receive at this speed rate, the transmission of the video is changed to stream 2. For example, when the need to change the speed rate to a lower one is noticed and the most recently received (or alternatively the next) timing mark is, let's say, the mark **7** indicating time 1000 ms, it tells the corresponding byte 805000 in stream 2 to where the source of the transmission of the video transfers **11.**

[0021] Similarly, stream 2 (and the media stream for it) contains timing marks **10**, **11**, **12**, **13**, which point to the location of corresponding information in stream 1, as shown by the corresponding byte **10A, 11A**, **12A**, **13A**, in stream 1. When the capability of transmitting 200 kbps is noticed during the transmission at 40 kbps rate, the transmission of the video is changed to stream 1. For example, when the availability of the higher speed rate is noticed and the next timing mark is the mark **13** indicating time 3000 ms, the corresponding byte 77000, located in stream 1, points to the location of corresponding information in stream 1 so the source of transmission is shifted to stream 1, using data stored in location **9** onwards.

[0022] FIG. **2** shows an example of a flow chart, which illustrates a function and method of the invention. FIG. **3** shows an example of the inventive arrangement. When a user visits a home page of a service provider, he/she may find something he wants to order, such as a video show. The video player client running in the user terminal sends **34** a request message from the terminal **33** to the service provider's server **31** through the network **32**, such as the Internet, to get the show using a GET method, which is one of the methods defined in the HTTP1.1 protocol. Let the request be:

GET http://www.serviceprovider.com/video/music/song1.video. **HTTP/1.1**

[0023] Often, the connection speed rate depends on the available (and optional) connection and processing speeds of the terminal. At the beginning, the first stream which is sent works like a test stream. The test stream, i.e. the first stream preferably forms the first part of the file structure, which preferably contains data for the lowest bandwidth. Thus the first stream over the HTTP transport also contains a list of alternative speed rates. The player in the client terminal chooses **21** one of the supported speed rate from the list. (It should be noted that the server can be informed of the capability of a terminal or the available bandwidth using other techniques and the specific method of learning the link capabilities is a matter of technical choice to one skilled in the art.) In this example, the user's terminal supports speed rates of 200 kbps and 40 kbps. Let's assume that the desired file **36** is the file in FIG. **1**. If the speed rate of 200 kbps **22** is available, stream 1 is selected **24**. Otherwise, the speed rate of 40 kbps **23** is chosen and stream 2 is selected **25.** (Notice that the higher speed rate means the higher bandwidth.)

[0024] In the selection of the speed rate of 200 kbps the server starts to play **26** stream 1 beginning at byte (2000) of the file. The data and the timing marks are sent **35** to the user's terminal during the play. The terminal checks **28** frequently that the currently available bandwidth is sufficient for the selected stream. If the bandwidth is sufficient, playing stream 1 is continued. If the bandwidth becomes unavailable for stream 1, stream 2 is selected **210** and stream 2 starts to play **27**, **311** be-

Aylo Exhibit 1018
Page 1911 of 2308

ginning at the byte of the file defined by the latest timing mark. The bandwidth of stream 2 is checked **29** similarly as the bandwidth of stream 1, and when the bandwidth is again available for the speed rate of 200 kbps, stream 1 is selected **211**.

[0025] In the selection of the speed rate of 40 kbps the server starts to play **25** stream 2 from the right byte (800000) of the file. The data and the timing marks are sent **35** to the user's terminal during the play. The terminal checks **29** frequently that the bandwidth is okay for the selected stream. If the bandwidth is still restricted to 40 kbps, playing stream 2 is continued. If the bandwidth rise to a level sufficient to support 200 kbps transmission, stream 1 is selected **211** and stream 1 starts to play **26**, and is sent **311** to the terminal **33** beginning at the byte of the file defined by the latest timing mark. The bandwidth of stream 1 is checked **28** similarly as the bandwidth of stream 2, and when the bandwidth is again too slow for the speed rate of 200 kbps, stream 2 is selected **210**.

[0026] The checking of the bandwidth of the streams can be made, for example, measuring a filling degree of a buffer for incoming data connection in the user's terminal. When the buffer is becoming empty, the required bandwidth for current stream is too high compared to what is available through the network connection, and the speed rate should be dropped. When the buffer is becoming full, bandwidth for current stream is lower than what is available through the network connection, and the speed rate can be raised. After the measurement, the suitable stream source is selected **210, 211** and a GET command with an additional field RANGE is sent **310** from the user's terminal to the server for changing the stream. For example, let the GET command be GET http://www.serviceprovider.com/video/music/song3video RANGE: bytes= 805000- when the transmission speed rate changes from 200 kbps to 40 kbps.

[0027] It is worth noting that preferably, when a stream is changed, a new HTTP transport (and naturally a new media stream) is started and the old media stream is terminated after the new media stream starts playing. From the view of the user the desired file, such as a video show, seems to continue without any break because the new stream starts playing from the byte indicated by the timing mark, thus the information continuity is maintained, despite the media stream switch. So, when the source is changed, also a new HTTP transport is created for the new stream, i.e. media stream, instance and for the new source. Thus the adaptive media stream, i. e. the adaptive stream, is made up of one media stream or several consecutive media streams. Since the consecutive media streams carry essentially the same data, a user gets a visual impression of continuous shows. In other words, the adaptive media stream comprises at least one media stream for transmitting information. The HTTP 1.1 protocol provides a feature called Persistent Connections. Persistent Connection means that several requests (such as GET), can be sent via one TCP connection. Thus one connection may carry several consecutive streams. However, it should be noted that several TCP connections may be needed or used for transmitting consecutive streams.

[0028] Furthermore, it may be said that the adaptive media stream comprises several (or one if the speed rate remains the same during the whole transmission) instances (separate streams) of the same object (the adaptive media stream containing information such as video). Naturally, there are more ways to model the adaptive media stream, but common for these models are that transmission of information of the adaptive stream, (the stream of a continuous data, such as video over one or several connections) happens in one or several consecutive media streams over a transport protocol, each media stream adapted for certain conditions, and certain links or pointers periodically connect or cross connect the media streams in essentially content coherent fashion. Thus varying transmission conditions may dictate how many suitable media streams are sent forming together the adaptive stream.

[0029] For creating the inventive arrangement an initial selection module **39** (i.e. means) for choosing an original transmission speed rate and selecting a suitable source in a service provider's server is needed. In a user's terminal another estimation module **37** (i.e. means for checking transmission rate) is needed, for checking the suitability of the transmission speed rate as reselection module **38** (i.e. means) for reselecting a suitable source are needed.

[0030] Although, the processing and network connection speed of a user's terminal is mostly the decisive element when selecting a suitable transmission speed rate, it may happen that the network itself becomes congested. In congestion, the available speed rate decreases and the transmission speed rate must be lowered. On the other hand, if the network allows a higher speed, the transmission speed rate may be increased.

[0031] Using the inventive arrangement, an adaptive media stream can be served from with a common HTTP server. Therefore, there is no need for the use of a special media stream server, or parallel devices and software. Remarkable cost savings are achieved, since usually service providers and users support the HTTP protocol. Also the reliability of the system increases because of fewer sources of faults.

[0032] It is clear that inventive arrangement can be implemented in many ways. For example, the time period between two timing marks can be different than in the example of this text. It can be any suitable period depending on the features of a solution, such as 2 or 4 seconds. The format of the inventive file can be any suitable format as well, such as MPEG or another.

[0033] The inventive file may comprise more than two different streams (containing the same content), from which the most suitable stream is selected at any given time. Therefore, timing marks must contain a number of references for the corresponding byte in the other data

Aylo Exhibit 1018
Page 1912 of 2308

sources.

[0034] All the sources do not have to be in the same file, but each source can be a separate file. However, the use of this kind of structure is not as efficient when attempting to use as little processing power as possible compared to the use of one file structure. In the structure of several files, timing marks must indicate the right file with the corresponding byte.

[0035] Although, a sending terminal have been described to be a server in the examples above, it is clear that the sending terminal can be another terminal capable of sending information.

[0036] Similarly other mechanisms are known and suitable to measure and select the most appropriate source. Such methods may be server based or user terminal based, or even utilize intermediate network nodes to estimate network congestion and bandwidth, and similar measurements that are indicative of total node to node capacity, as well as specific node capabilities, and the selection of sources (and media streams) may be carried out in various ways that will provide the best end-user experience according to varying conditions.

[0037] According to the above-mentioned matters, it is evident that the invention is not restricted to the solutions described in this text, but it can be modified for various solutions, in the scope of the inventive idea.

Claims

1. A method for providing an adaptive media stream between a sending terminal and a user's terminal **characterized in that** the adaptive media stream comprises at least one media stream and the method comprises the steps of:

    a) choosing a transmission speed rate for the media stream,
    b) selecting a suitable data source for the media stream among sources containing essentially the same information, each source intended for a certain transmission speed rate,
    c) playing the media stream,
    d) checking the suitability of the transmission speed rate,
    e) either continuing playing the media stream if the transmission speed is still suitable or,
    f) selecting a new data source, which is suitable at the moment,
    g) as a response for the selection of the new data source, playing a new media stream,
    h) repeating steps d) to g) until the stream, which is played, stops.

2. A method according to claim 1 **characterized in that** step b) comprises the phase of reading a header field of a file, which contains at least two data sources, the header indicating the starting points of the sources and to which transmission speed rate each source is intended.

3. A method according to claim 2 **characterized in that** step c) comprises the phase of starting to read from the starting point of the selected source.

4. A method according to claim 1 or 3 **characterized in that** step d) comprises the phase of monitoring a filling degree of a buffer for incoming data in the user's terminal.

5. A method according to claim 4 **characterized in that** if the buffer is becoming empty or full, then step f) is chosen, else step e) is chosen.

6. A method according to claim 1 or 5 **characterized in that** the information of the selection of the new data source is transmitted from the user's terminal to the sending terminal using a GET method of an HTTP protocol with an optional RANGE field.

7. A method according to claim 1 **characterized in that** step g) comprises phases of terminating the old media stream after starting to play the new media stream and creating the new media stream before playing the new media stream.

8. A method according to claim 1 or 7, **characterized in that** step g) further comprises the phase of reading timing marks in the data source for defining the right point to start playing the new media stream.

9. A method according to claim 1 **characterized in that** step b) comprises the phase of reading a header field of files, where each contains one data source, the header indicating the starting points of the source and to which transmission speed rate the source is intended.

10. An arrangement for providing an adaptive media stream between a sending terminal and a user's terminal, **characterized in that** the adaptive media stream comprises at least one media stream and the arrangement comprises:

    a) means in the sending terminal for choosing a transmission speed rate for the media stream,
    b) a data structure in the sending terminal comprising at least two data sources containing essentially the same information, each source intended for a certain transmission speed rate,
    c) means in the sending terminal for selecting a suitable data source from among the sources for the media stream,
    d) means in the user's terminal for checking the suitability of the transmission speed rate when the media stream is played,

Aylo Exhibit 1018
Page 1913 of 2308

e) means in the user's terminal for reselecting a suitable data source from among the sources for a new media stream.

11. An arrangement according to claim 10, **characterized in that** the arrangement further comprises means for playing the data structure in a way that the playing is started from the right point of the data structure.

12. An arrangement according to claim 11, **characterized in that** means a) comprises means for reading a message from the user's terminal, which defines alternative transmission speed rates.

13. An arrangement according to claim 11, **characterized in that** each source comprises timing marks, which indicate the corresponding point in the other sources.

14. An arrangement according to claim 13, **characterized in that** the arrangement further comprises a GET method of an HTTP protocol with an optional RANGE field for transmitting the information of the reselection of the suitable data source from the user's terminal to the sending terminal.

15. An arrangement according to claim 11, **characterized in that** means d) comprises means for monitoring a filling degree of a buffer for incoming data in the user's terminal.

16. An arrangement according to claim 11, **characterized in that** the data structure is a file containing the data sources.

17. An arrangement according to claim 16, **characterized in that** means c) comprises means for reading a header of the data structure, which indicates the starting points of the sources and to which transmission speed rate each source is intended.

18. An arrangement according to claim 11, **characterized in that** the data structure is a number of files, each of them containing one data source.

19. An arrangement according to claim 18, **characterized in that** means c) comprises means for reading headers of the files, which indicate the starting point of the source and to which transmission speed rate the source is intended.

Aylo Exhibit 1018
Page 1914 of 2308

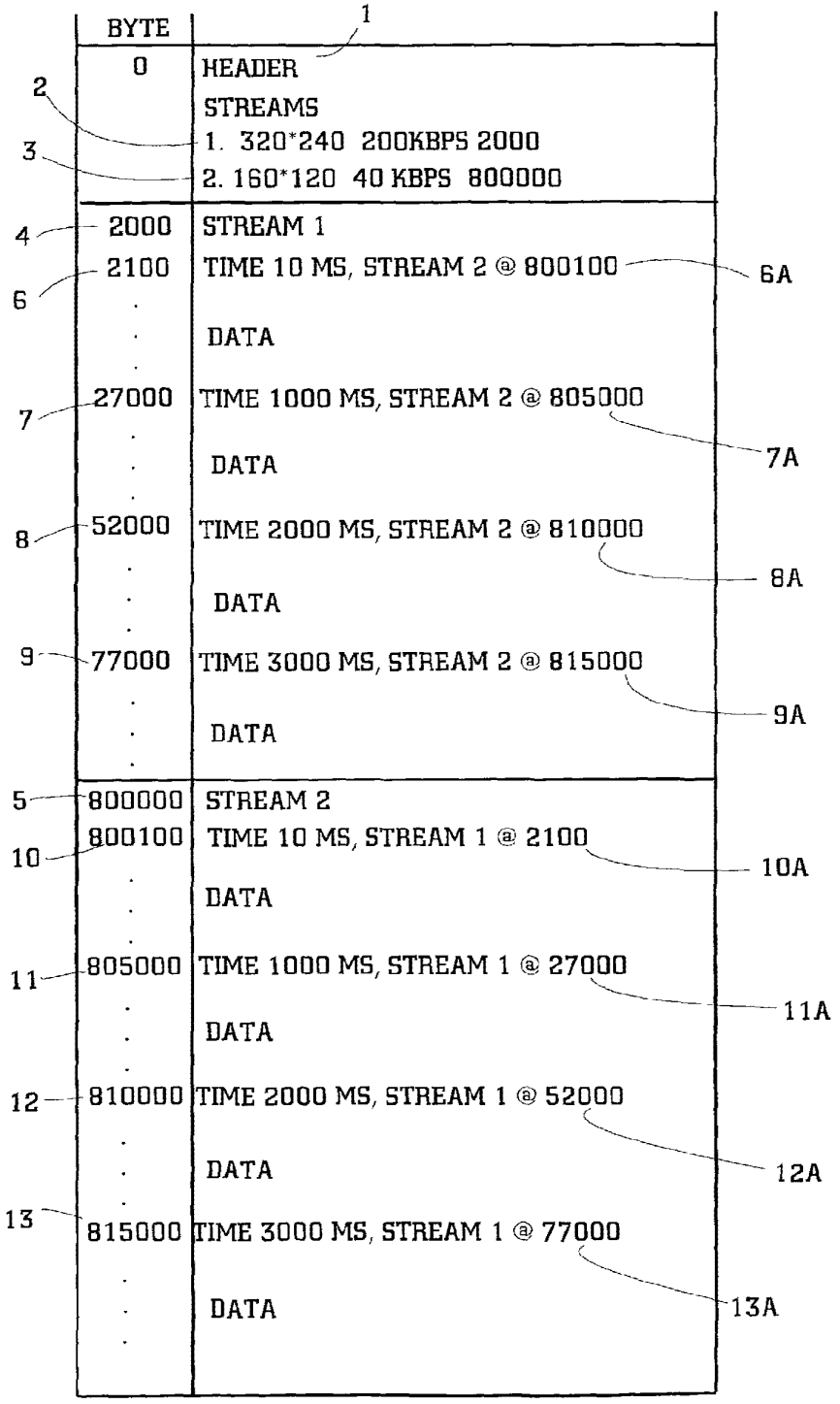


FIG. 1

Aylo Exhibit 1018
Page 1915 of 2308

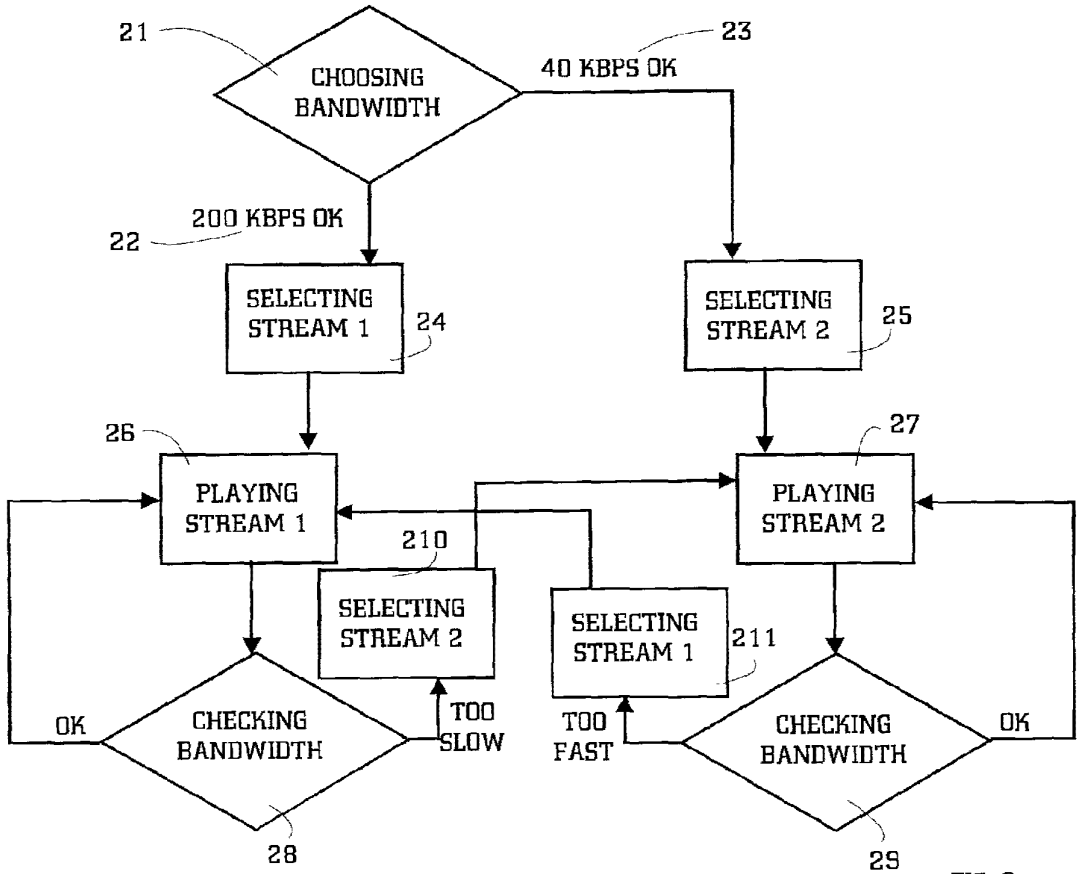


FIG. 2

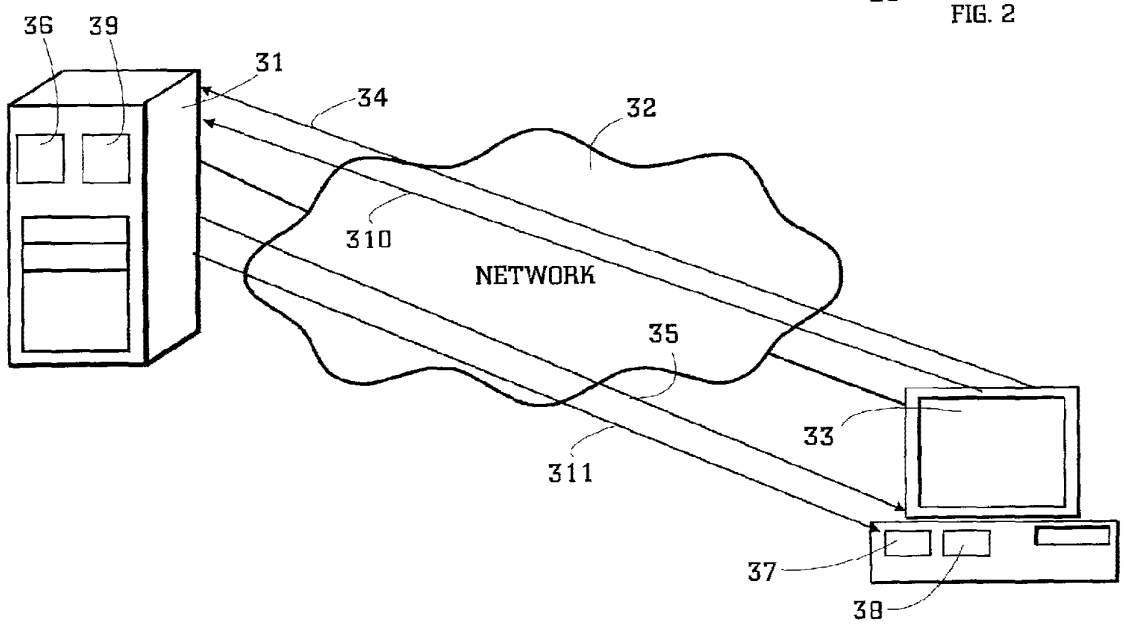


FIG. 3

Aylo Exhibit 1018
Page 1916 of 2308

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 202 487 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.05.2002 Bulletin 2002/18

(51) Int Cl.[7]: **H04L 1/00**

(21) Application number: **01309276.2**

(22) Date of filing: **31.10.2001**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **31.10.2000 JP 2000333916**
29.06.2001 JP 2001200156

(71) Applicant: **Kabushiki Kaisha Toshiba**
Tokyo 105-8001 (JP)

(72) Inventors:
- **Nagai, Takeshi, c/o Kabushiki Kaisha Toshiba**
  Minato-ku, Tokyo 105-8001 (JP)
- **Kikuchi, Yoshihiro,**
  c/o Kabushiki Kaisha Toshiba
  Minato-ku, Tokyo 105-8001 (JP)
- **Masuda, Tadaaki, c/o Kabushiki Kaisha Toshiba**
  Minato-ku, Tokyo 105-8001 (JP)

(74) Representative: **Midgley, Jonathan Lee**
Marks & Clerk
57-60 Lincoln's Inn Fields
GB-London WC2A 3LS (GB)

(54) Data transmission apparatus and method

(57) A data transmission apparatus comprises a network condition determination device for determining a transmission condition of the transmission channel in the transmitter (103) based on at least Jitter information or a packet loss rate obtained from a receiver (104), and a video encoder (102) including means for changing at least one of a bit rate of transmission data and a error resilience level according to the transmission condition.

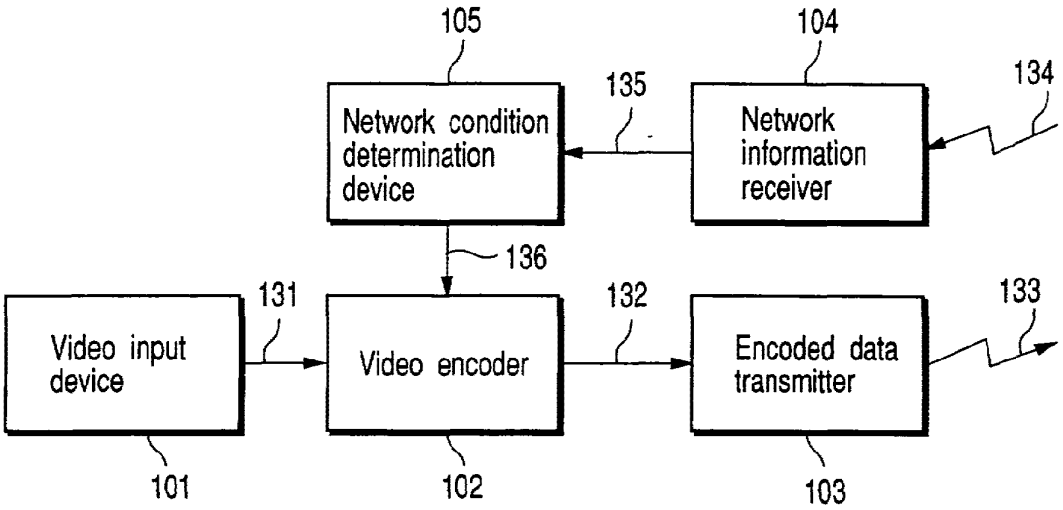


FIG. 1

EP 1 202 487 A2

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] The present invention relates to a data transmission method for transmitting encoded moving/still picture by using a wired communication network such as ISDN (Integrated Services Digital Network) or a wireless communication network such as PHS (Personal handy-phone System) or satellite communication and a data transmission apparatus to which such method is applied.

[0002] In recent years, with progress of digital encoding technology of a variety of information such as image and a broadband network technology, the development of applications using these technologies becomes active, and a system of transmitting compression-encoded image by utilizing a communication network is developed.

[0003] For example, an input video signal is encoded at a video encoder, the obtained encoded data is provided to an encoded data transmitter, and transmission data is transmitted to a transmission channel. In this way, a video is encoded and transmitted. In recent years, with the spread of Internet/Intranet, there has been increased an application or system of transmitting/receiving a packet of data. Providing data as a packet is very effective means for efficiently sharing a bandwidth of a communication channel by a plurality of users.

[0004] In the meantime, as a protocol for transmitting/receiving packet data via Internet/Intranet, there exists TCP/IP (Transmission Control Protocol/Internet Protocol) or UDP/IP (User Datagram Protocol/Internet Protocol) and the like.

[0005] Among them, TCP/IP is an Internet standard protocol. This protocol is applied as the world's best known protocol to Intranet or LAN as well as Internet. TCP works in a transport layer that is a fourth layer of an OSI basic reference model, and IP works in a network layer that is a third layer.

[0006] In addition, UDP/IP is one of the TCP/IP suite transport layer protocols. TCP is a connection type protocol. UDP is a disconnection type protocol. As with TCP, a host application is identified by using a port number. This protocol is often used to carry short data for monitor control of a network or real time oriented data such as audio data or video data.

[0007] TCP/IP incorporates a frame such as retransmission. Thus, this protocol is strong against errors or the like. Even if a large amount of time is required, this protocol is effective to download type applications if an attempt is made to validly receive data. However, this protocol is not powerful for applications requiring real time properties.

[0008] In contrast, although UDP/IP does not have a retransmission frame, a delay associated with retransmission does not occur. Thus, this protocol is very effective to applications requiring real time processing.

[0009] Transmission of moving picture is exemplified as a typical example of application requiring real time processing. However, in the case of general video communication, video data has a very large amount of data, and overflows a network bandwidth in most cases. In this case, a technique for encoding a video data signal, and transmitting a reduced amount of data is used. Techniques such as motion compensation, discrete cosine transform (DCT), sub-band encoding, pyramid encoding, and variable-length encoding or a combination of these techniques are developed as compression-encoding of a video signal.

[0010] As an international standard system for moving picture encoding, there are ISO MPEG-1, MPEG-2, ITU-T H. 261, H. 262, and H. 263. In addition, as an international standard system for multiplying an encoded data stream having a speech/audio signal compressed or other data, there are an ISO MPEG system, ITU-T H. 221, and H. 223.

[0011] Internet or the like is encompassed via an infinite number of networks. Usually, no one knows which network is used or what state is. In addition, an amount of data flowing a network changes from time to time. Thus, a system of judging what amount of data can be communicated in a real time is required.

[0012] Therefore, with further advancement from real time applications utilizing UDP/IP, there have been increased in number applications using a packet format called RTP (Real-time Transport Protocol) for adding and transmitting time information or the like to a packet.

[0013] This RTP is a protocol for transferring audio data signal or video data signal ruled in RFC 1889 in real time. Usually, data is carried on UDP (User Datagram Protocol). This protocol is assumed to be applied to a multimedia system such as video conference, and data can be exchanged in a real time interactive scheme. However, the protocol has no sound quality or picture quality assurance function. Sequential number or time stamp data is assigned into an RTP header, thereby supporting a read time operation.

[0014] RFC is an abbreviation of Request for Comments, and denotes a document such as technical proposal or comments disclosed by TETF (Internet Engineering Task Force). Many de-fact standards such as a variety of protocols of TCP/IP suite are described in RFC.

[0015] By utilizing this RTP, time information and a packet number are added to a packet, so that a receiver can display a voice or image by using correct time information, can determine a packet or the like whose order is replaced in a network, and can detect that a packet is lost by referring to a packet number.

[0016] Moreover, the RTP provides with a system (RTCP) for notifying network state information such as jitter or packet loss rate from a transmitter to a receiver.

[0017] However, how this RTCP information is utilized depends on applications, and is not determined in standards.

[0018] In addition, in the case of an image, a bandwidth corresponding to video transmission cannot be al-

Aylo Exhibit 1018
Page 1918 of 2308

located in a network bandwidth, thus making it necessary to compress and deliver a video signal in an encoding system such as MPEG, as described previously. Although this encoding system is effective in a decreased amount of data, the system becomes very fragile to a packet loss or error entry by supplying data to unstable Internet. This is because a moving picture encoding system transmits only a difference between a current frame and a previous frame. Thus, partial data missing causes a very serious problem. In the case of using the UDP or RTP, data retransmission is not basically carried out, thus making it necessary to counteract this problem.

[0019] There generally exist two modes in picture encoding, that is, one is an inter-frame encoding mode for transmitting a difference between the current frame and the previous frame and an intra-frame encoding mode for carrying out encoding in one frame. In general, intra-frame encoded pictures are provided at a proper timing, and inter-frame encoded pictures are provided between them. An interval between the frames encoded in an intra-frame encoding mode is called a GOP (Group of Picture) interval.

[0020] When an encoded picture is missing during transmission to the decoder, the decoded picture corresponding to the missing encoded picture is destroyed. Thereafter, the intra-encoded pictures following the missing encoded picture are decoded based on the destroyed picture. Thus, all the subsequent decoding pictures are affected by the destroyed picture, and decoding cannot be carried out correctly. Therefore, an intra-frame encoded picture is inserted in the way of a stream of frames to cut off a propagation of missing error and recover adequate decoding.

[0021] In a conventional technology, it can determine that errors occur in a network. However, how to utilize the determination result has been unknown. Also, a system for notifying the network information to the transmitter in an unique format is conceived. However, the system is based on the specification of associated applications and lacks a general-purpose use. The system receives transmission data using an information receiver, and controls a video encoder within a range of the information received from the network via a transmission channel.

[0022] A conventional network condition determination is conceived on assumption of network congestion. However, since Internet is popularized in mobile environment, it is required to consider an error or the like in wireless environment. Current technique does not assume countermeasures against such error.

[0023] Further, in the case of dealing with the error, a case in which a network condition changes, and an error rate is not constant is not considered. If a short GOP interval is set in order to counteract a case in which errors frequently occur, the intra-frame encoding increases resulting in impairing the encoding efficiency. When errors occur frequently qualitatively, it may be required to establish such state in which such errors occur frequently. However, in the case of a network in which an error hardly occurs in a normal state, and an error occurs only at one moment, this setting is very wasteful. Conversely, when a long GOP interval is set, it is troublesome that a great influence occurs when an error occurs.

[0024] Thus, a real time transmission using RTP is noted. As described above, by utilizing the RTP that is a protocol for transferring audio data signal or video data signal in real time, time information and a packet number are added to a packet, whereby the receiver can display a voice or image by using correct time information, can determine a packet or the like whose order is replaced in a network, or can detect that a packet is lost by referring to a packet number.

[0025] Moreover, at the RTP, there is provided a system (RTCP) for notifying network state information such as jitter or packet loss rate from the transmitter or receiver.

[0026] It is an object of the present invention to provide a data transmission apparatus and method capable of adjusting a transmission rate by using the RTP characteristics in accordance with a transmission state of a transmission channel based on notification of network state information such as jitter or packet loss rate, obtained from the receiver in the transmitter or capable of carrying out controls such as changing error resilience, thereby making it possible to efficiently transmit data to the maximum, and fully utilize transmission that requires real time properties.

[0027] According to an aspect of the invention, there is provided a data transmission apparatus that transmits transmission data from a transmitter to a receiver via a transmission channel, comprising: an estimation device configured to estimate a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and a controller configured to change at least one of a bit rate of transmission data and a error resilience level according to the estimated transmission condition.

[0028] According to another aspect of the invention, there is provided a data transmission method comprising: estimating a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and changing at least one of a bit rate of transmission data and an error resilience level according to the estimated transmission condition.

[0029] According to another aspect of the invention, there is provided a data transmission program stored on a computer readable medium, comprising: instruction means for instructing a computer to estimate a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and instruction means for instructing the computer to change at least

Aylo Exhibit 1018
Page 1919 of 2308

one of a bit rate of transmission data and a error resilience level according to the estimated transmission condition.

[0030] This summary of the invention does not necessarily describe all necessary features so that the invention may also be a sub-combination of these described features.

[0031] The invention can be more fully understood from the following detailed description when taken in conjunction with the accompanying drawings, in which:

FIG. 1 is a view showing a basic configuration of a data transmission apparatus according to a first embodiment of the present invention;
FIG. 2 is a view showing another configuration in the present embodiment;
FIG. 3 is a view showing a configuration of a network condition determination device in the present embodiment;
FIG. 4 is a view showing a configuration of a video encoder in the present embodiment;
FIG. 5 is a view showing a configuration of an encoded parameter determining device in the present embodiment;
FIG. 6 is a view showing a basic configuration of a data transmission apparatus in a second embodiment of the present invention;
FIG. 7 is a view showing a configuration of an encoded data switch device in the present embodiment;
FIG. 8 is a view illustrating a timing of encoded data switching at an encoded data switch device in the present embodiment;
FIG. 9 is a view illustrating a timing of switching encoded data with its different frame interval at the encoded data switch device in the present embodiment;
FIG. 10 is a flow chart showing a basic example of the network condition determination method at the network condition determination device in the first embodiment of the present invention;
FIG. 11 is a flow chart showing a basic example when a delay is forced to occur by using the network condition determination method at the network condition determination device in the first embodiment of the present invention;
FIG. 12 is a flow chart showing a basic example considering a time direction change with the network condition determination method at the network condition determination device in the present embodiment;
FIG. 13 is a view showing an exemplary configuration according to a third embodiment of the present invention;
FIG. 14 is a flow chart showing an example of a GOP interval calculation method at a GOP interval calculator in the present embodiment;
FIG. 15 is a flow chart showing another example of a GOP interval calculation method at a GOP interval calculator in the present embodiment; and
FIG. 16 is a flow chart showing still another example of a GOP interval calculation method at a GOP interval calculator in the present embodiment.

[0032] Hereinafter, embodiments of the present invention will be described with reference to the accompanying drawings.

[0033] The present invention utilizes RTP or RTCP that is a protocol for transferring audio data signal or video data signal in real time. The transmitter performs controls such as adjusting a transmission rate or changing error resilience in accordance with a transmission state of a transmission channel based on notification of network state information acquired from the receiver such as jitter (Jitter information) or packet loss rate. In this manner, data transmission can be carried out efficiently to the maximum. Hereinafter, embodiments of the present invention will be described in detail.

(First Embodiment)

[0034] FIG. 1 shows a basic configuration of an image transmission apparatus according to a first embodiment of the present invention. In FIG. 1, a video signal 131 input from an image input device 101 is processed to be encoded. That is, an encoded data transmitter 103 processes encoded data 132 that has been encoded by a video decoder 102 in the form which conforms to a network, and transmits the encoded data to a receiver. A network information receiver 104 receives network information 134 delivered from the receiver, and outputs the received information to a network condition determination device 105. This network condition determination device 105 has a function that determines a network condition from network information 135 output from the network information receiver 104, and notifies the result as network condition information 136 to the video encoder 102.

[0035] The video encoder 102 encodes the video signal 131 using the network condition information 136. This encoding is carried out as follows. That is, in the case of RTP or RTCP, a packet loss rate or Jitter information is provided to the network condition determination device 105. The network condition determination device 105 determines that, if a packet loss rate is not "0" or if a delay time is longer than a predetermined value, any load is applied to a network, and a presumed amount of data cannot be supplied. Then, the network condition determination device 105 is configured so as to carry out processing so as to issue an instruction (network condition information 136) for lowering a bit rate to the video encoder 102, and to lowly set a target bit rate in encoding.

[0036] According to the thus configured apparatus, when the video signal 131 is inputted from the image input device 101, the inputted video signal 131 is encod-

Aylo Exhibit 1018
Page 1920 of 2308

ed by the video encoder 102. Encoded data 132 encoded by the video encoder 102 is inputted to the encoded data transmitter 103.

[0037] The encoded data transmitter 103 processes the encoded data in the form suitable to a network, and receives the processed data to a receiver. The network information receiver 104 receives the network information 134 delivered from the receiver, and outputs the received information to the network condition determination device 105.

[0038] The network condition determination device 105 determines a network condition from the network information 135 outputted from a network information receiver, and notifies the result as the network condition information 136 to the video encoder 102.

[0039] The video encoder 102 encodes the video signal 131 using the thus notified network condition information 136.

[0040] FIG. 10 shows a very simple example of a method for determining a network condition at the network condition determination device 105 by way of showing a flow chart.

[0041] According to this flow chart, at the step S1001, a packet loss rate of network information is checked to see if the rate is "0" or not. As the result of this check, if the packet loss is not "0", an effective bit rate is calculated from the packet loss rate in accordance with the step S1002. This bit rate can be calculated by formula (1) shown below.

$$b' = b \times (1 - r) \quad (1)$$

where "b" denotes a current bit rate, b' denotes a new bit rate, and "r" denotes a packet loss rate.

[0042] Next, at the step S1003, a parameter (encoding parameter) obtained at the step S1002 is notified to the video encoder 102. Here, the above parameter is used to determine a rate control or error resilience level for encode processing at the image encoder 102. This parameter is also used for adjustments such as adjusting and controlling the encode processing speed of the input image signal 131 to this parameter, thereby changing a bit rate, a frame interval, or an error resilience level.

[0043] On the other hand, at the step S1001, when the packet loss rate is "0", no parameter change occurs at the step S1004. This determination method is provided as a mere example. It is possible to employ any other method for obtaining a bit rate based on a case in which the criterion of the step S1001 is determined by thresholds without being set to "0" or others or based on a formula other than the above formula (1).

[0044] Therefore, according to the first embodiment, even when a transmission network bandwidth is unknown or when the bandwidth changes halfway, it is possible to reset a parameter or the like based on the receiver's network information, and then, encode and transmit an image in accordance with the parameter suitable to a current network. In this manner, even in the case of a network congestion, although a bit rate is automatically lowered, and an image quality is lowered, it is possible to communicate an image without an occurrence of a phenomenon that an image is destroyed due to such packet loss. This method is very effective in image transmission in real time.

[0045] The first embodiment can be carried out by modifying it as follows. For example, as shown in FIG. 11, a method other than that shown in FIG. 10 is employed as a method for determining a network condition at the network condition determination device 105. That is, in the flow chart shown in FIG. 10, a parameter is set at only a packet loss rate. In contrast, in the method shown in FIG. 11, even when no packet loss occurs at the step S1104, a delay quantity is checked. As a result of this check, if a delay occurs in excess of a predetermined value, it is determined that a network is congested. When it is determined that a network is congested, a bit rate is calculated, and a new parameter is set by using formula (2) below, for example, at the step S1105.

$$b' = b \times \{(TN-TS)+(d-dth)\}/(TN-TS) \quad (2)$$

where "b" denotes a delay time, b' denotes a new bit rate, TN denotes a current time, TS denotes a start time, "d" denotes a delay time, and "dth" denotes delay time thresholds.

[0046] FIG. 2 shows an example when the network condition determination device 105 is provided with a function for generating the network condition information 136 so as to provide the information to the image encoder 102 by extending the configuration shown in FIG. 1.

[0047] According to a configuration shown in FIG. 2, the network condition determination device 105 is provided with a function for acquiring a current parameter at the current image encoder 102 and a function for generating the network condition information 136 on what parameter should be set next from the thus acquired current parameter and network information 135, thereby providing the parameter to the image encoder 102.

[0048] In the case of this construction, the network condition determination device 105 can recognize a state of a video parameter, i.e., a state of the encode processing currently done at the video encoder 102, thus making it possible for the network condition determination device to calculate and grasp a bit rate or the like. This construction is advantageous in the case of determining a network condition by comparing the network information 135 with parameter information 137. Thus, a system is realized which is capable of changing a parameter so as to set an optimal bit rate in accordance with the network condition, and providing the parameter to the video encoder 102, thereby adjusting the encode processing.

[0049] In this way, according to the present invention,

Aylo Exhibit 1018
Page 1921 of 2308

transmission is adjusted in accordance with a transmission state of a transmission channel or error resilience is changed. The transmission state is adjusted by changing a bit rate or changing a frame interval. With respect to a change in quality of a transmission channel, the error resilience is changed according to the transmission state of the transmission channel (for example, the error resilience is changed by changing an interval of an MPEG4 synchronous signal). Therefore, data transmission can be carried out efficiently to the maximum, thereby enabling data transmission that makes it possible to fully utilize transmission requiring real time processing.

[0050] FIG. 3 shows a configuration when the configuration shown in FIG. 2 is further developed. This network condition determination device 105 comprises a network information storage device 301 that stores past network information and an encoded parameter information storage device 302 that stores encoded parameter information supplied to the current video encoder 102. A network condition determination device 303 recognizes the past network information and the current encoded parameter from these elements, decides an elapse of time in network condition from these elements to determine an optimal parameter assigned to the video encoder 102, and provides the parameter to the video encoder 102.

[0051] With this construction, it is possible to determine a network condition more accurately by referring to an elapse of time relevant to network information. FIG. 12 shows a flow chart of a network condition determination method in the case of using this system. Processing at the network condition determination device 303 will be described in accordance with this flow chart. First, the network condition determination device 303 determines whether or not packet loss is "0" at the step S1201. As a result of this determination, when the packet loss is present, it is determined by the step S1202 whether or not the past packet loss occurs. As a result, when the past packet loss occurs, the loss rate is compared with the current rate. If the current packet loss rate is higher than the past packet loss rate, it can be determined that the past correction has been invalid.

[0052] It is understood from this determination that the bandwidth overflow of transmission data such as network congestion is not troublesome, but a possibility that a noise or the like occurs over a channel resulting in destroying is high. Therefore, the network condition determination device 303 does not change a bit rate at the step S1204, and carries out parameter setting to ensure reliable error correction, and notifies the fact to the video encoder 102 at the step S1205.

[0053] If the current packet loss rate is lower than the past packet loss rate at the step S1203, the network condition determination device 303 determines that the past change works efficiently, and re-calculates a valid bit rate from the packet loss rate at the step S1206. When the past packet loss does not occur at the step S1202, the last loss is determined as a channel noise, and processing goes to the step S1204.

[0054] On the other hand, when no packet loss occurs at the step S1201, the network condition determination device 303 checks a delay quantity at the step S1207. As a result, when the delay quantity exceeds thresholds "Th", a valid bit rate is calculated from the delay quantity at the step S1208, and the calculation result is notified to the video encoder 102 at the step S1205.

[0055] Lastly, when the delay quantity does not exceeds the thresholds "Th", the network condition determination device 303 determines that successful communication is established, and does not change a parameter at the step S1209.

[0056] By causing the network condition determination device 303 to carry out such processing, the network condition determination device 105 can have a function capable of determining whether a packet loss occurs on a transmission channel due to network congestion or whether a noise is generated, and the loss occurs due to the impairment of a channel state such as radio equipment.

[0057] This function is provided as a mere example. For example, when the past packet loss does not occur at the step S1202, a control for lowering a bit rate may be made by determining network congestion instead of determination of noise. If the current packet loss rate is improved more remarkably than the past packet loss rate at the step S1203, similar parameter change may be made after checking how the past change has been made. In this way, in this system, a variety of determination systems can be set according to a network.

[0058] In the foregoing, although there has been shown an example in which the video encoder 102 carries out encode processing based on encoded parameters determined and assigned at the network condition determination device 105, a configuration for the video encoder 102 to determine a parameter can be provided without being limited to the network condition determination device 105. An example is shown in FIG. 4.

[0059] FIG. 4 is a block diagram depicting an inclusive configuration when the video encoder 102 determines an encoded parameter. In the case of this example, the video encoder 102 comprises an encoded parameter determination device 401 and a signal processor 402.

[0060] In the case of this configuration, the network condition determination device 105 is configured to output the network condition information 136 from the network condition determination device 303, and then, provide the information to the video encoder 102.

[0061] In such a configuration, the network condition determination device 105 provides the network condition information 136 delivered from the network condition determination device 303 to the video encoder 102. Then, the video encoder 102 first inputs the network condition information 136 to an encoded parameter determination device 401. Then, the encoded parameter determination device 401 recognizes a network condi-

Aylo Exhibit 1018
Page 1922 of 2308

tion from this network condition information 136, and generates an encoded parameter 431 in the form suitable to the network condition.

[0062] The thus generated encoded parameter 431 is inputted to a signal processor 402. Then, this signal processor 402 encodes the video signal 131 inputted from the video input device 101 by using this encoded parameter 431.

[0063] Encoding information 432 such as the number of encoded bits after encoding is inputted to the encoded parameter determination device 401, and is used for the next encoded parameter determination. Encoded data 132 encoded at the signal processor 402 is output to the encoded data transmitter 103.

[0064] FIG. 5 is a block diagram depicting a configuration including a intra-frame encoding determination device 501 that forcibly sets intra-frame encoding at the encoded parameter determination device 401 shown in FIG. 4. As shown in the figure, the encoded parameter determination device 401 comprises an intra-frame encoding/determination device 501 and an encoded parameter determination device 502.

[0065] In the case of this configuration, the network condition information 136 outputted from the network condition determination device 105 is input to the encoded parameter determination device 502 possessed by the encoded parameter determination device 401. Then, encoded parameter information 531 corresponding to the network condition information 136 is output from the encoded parameter determination device 502.

[0066] The encoded parameter information 531 is input to the intra-frame encoding determination device 501, and it is determined whether inter-frame encoding is carried out as is or intra-frame encoding is forcibly carried out. When it is determined as intra-frame encoding, the encoding parameter information 531 is updated so as to carry out intra-frame encoding, and is output as the encoded,parameter information 431 to the signal processor 402 of the video encoder 102.

[0067] By doing this, even if the past data is not transmitted correctly to a receiver, when the data can be received correctly after parameter change, a correct image can be reproduced.

[0068] It is possible to configure a video encoder capable of determining a network condition, which has functions of the video encoder 102 and network condition determination device 105, thereby making it possible to provide the previously described function in one block.

[0069] The encoded parameter determination device 401 or network condition determination device 105 makes determination concerning the next encoded parameter setting. However, if a picture quality is changed suddenly due to temporary instability of network, an unclear picture may be produced. In order to suppress such a phenomenon, it is possible to incorporate a mechanism in which a change in encoded parameter is suppressed within its predetermined range. In the above described circumstance, this mechanism functions efficiently.

[0070] Although a bit rate or an error resilience parameter may be set as its calculated value, it is possible to employ a method of predetermining some patterns, and then, selecting the closest pattern from among them. Further, instead of the closest pattern, it is possible to make a selection of the closest bit rate which is lower than the calculated bit rate. This makes it possible to prevent the generation of a video image with its unexpected image quality due to a combination of a variety of parameters, and makes it possible to prepare a combination of encoded parameters that have been tested to some extent.

[0071] In the foregoing example, when the packet loss rate is not "0" or when a delay exceeds the thresholds, there has been made a control such that parameters such as bit rate or error resilience are changed, and the changed is reset to that suitable to the current state.

[0072] However, after recovery of a network condition, when no packet loss occurs or when a small delay occurs, it is possible to make a control for increasing a bit rate or decreasing an error resilience parameter.

[0073] The foregoing example will be described below according to a second embodiment.

(Second Embodiment)

[0074] FIG. 6 is a view showing a basic configuration of an image transmission apparatus according to the second embodiment of the present invention. The image transmission apparatus shown in FIG. 6 comprises a plurality of media 601, an encoded data switch device 602, an encoded data selector 603, an encoded data transmitter 103, a network information receiver 104, and a network condition determination device 105.

[0075] The storage media 601 are used to store encoded data 631 on contents that have been encoded in advance. The storage media 601 each store encoded data on the same contents encoded by different encoded parameters, respectively.

[0076] The encoded data switch device 602 selects one of a plurality of storage media 601, and provides encoding data 631 stored in the thus selected storage medium 601 to the encoded data transmitter 103. A selection of the storage media is carried out in accordance with switch information 632 from the encoded data selector 603.

[0077] The network information receiver 104 has a function that receives network information 134 such as packet loss rate or Jitter information delivered from a receiver or a network, and provides the received information as network information 135 to the network condition determination device 105.

[0078] The network condition determination device 105 determines a network condition from the input network information 135, and outputs the result as network condition information 136 to the encoded data selector

Aylo Exhibit 1018
Page 1923 of 2308

603. The encoded data selector 603 has a function for estimating, from this network condition information 136, which of the encoded data on encoding parameters is optimal to the current network, and then, outputting to the encoded data switch device 602 the switch information 632 used for switching in order to select encoded data caused by the estimated optimal encoding parameter as an output.

[0079] The encoded data switch device 602 selects one item of encoded data from among the input encoded data 631, and outputs encoded data 132.

[0080] With such a configuration, the encoded data 631 encoded in advance and store in the storage medium 601 or the like is input to an encoding data switch device 602 when the data is reproduced from the storage medium 601. That is, respective encoded data encoded by the differential encoding parameters are maintained in a plurality of storage media 601. As a result of reproducing these items of data, the encoded data encoded by the different encoding parameters, respectively, is inputted. The encoded data switch device 602 selects one from among these items of data, and outputs it to the encoded data transmitter 103.

[0081] This selection is carried out in accordance with switch information 632 from the encoded data selector 603.

[0082] On the other hand, network information 134 such as packet loss rate or Jitter information from a receiver is received at the network information receiver 104, and then, the received information is delivered to the network condition determination device 105. The network condition determination device 105 determines a network condition from the inputted network information 135, and outputs the result as network condition information 136 to an encoded data selector 603.

[0083] The encoded data selector 603 estimates, from the network condition information 136, which of the encoded data of the encoding parameters is optimal to the current network, and outputs the switch information 632 to the encoded data switch device 602.

[0084] The encoded data switch device 602 selects one item of the encoded data from among the input encoded data 631, and outputs encoded data 132.

[0085] The present embodiment will be described by way of a specific example.

[0086] Now, assume that there exist five storage media 601a to 601e each having stored therein five items of encoded data encoded at 384 kbps, 128 kbps, 64 kbps, 32 kbps, and 16 kbps.

[0087] That is, for example, a storage medium having stored therein encoded data on contents encoded at 384 kbps is defined as a storage medium 601a; a storage medium having stored therein encoded data on contents encoded at 128 kbps is defined as a storage medium 601b; a storage medium having stored therein encoded data on contents encoded at 64 kbps is defined as a storage medium 601c; a storage medium having stored therein encoded data on contents encoded at 32 kbps is defined as a storage medium 601d; and a storage medium having stored therein encoded data on contents encoded at 16 kbps is defined as a storage medium 601e.

[0088] In the case of using a transmission channel having its quality capable of being transmitted at 384 kbps, first, reproduced encoded data is selected by the storage medium 601a that is a storage medium having stored therein encoded data on contents encoded at 384 kbps; transmission is started at 384 kbps, and reception of network information from the receiver starts at the same time.

[0089] Although network information is received at a predetermined interval, if a problem occurs with a network based on the received information, a valid bit rate is calculated at a stage at which a packet loss or delay occurs. The encoded data selector 603 selects encoded data on a bit rate close to the calculated bit rate, and the encoded data switch device 602 switches and transmits the selected data.

[0090] For example, the calculated bit rate is 140 kbps, a value close to this bit rate is 128 kbps. In this case, the above switch device selects and switches encoded data reproduced from the storage medium 601b having stored therein encoded data on contents encoded at 128 kbps. In addition, if the calculated bit rate is 100 kbps, a value close to a bit rate lower than the calculated bit rate is 64 kbps. In this case, the switch device selects and switches encoded data reproduced from the storage medium 601c having stored therein encoded data on contents encoded at 64 kbps.

[0091] In this way, in the present embodiment, contents are prepared after encoded in advance for type of transmission bit rate, and a current optimal transmission rate is obtained based on network information such as packet loss rate or Jitter information from a receiver that is a function possessed by RTP. A storage medium having stored therein encoded data obtained at the bit rate corresponding to this transmission rate is selected by the switch 602, and reproduced encoded data output of the storage medium is transmitted. In this manner, there is no need for a server to carry out encode processing in real time, making it possible to reduce a load on the server. This is very effective in distributing a pre-produced video such as broadcast in particular.

[0092] Now, a modified example of the second embodiment will be described here. FIG. 7 is a block diagram depicting a method for carrying out control for a switch timing of encoded data to be inputted at the encoded data switch device 602.

[0093] Encoded data 631 is inputted to the switching position detecting device 701 and a switch device 702. Switch information 632 is inputted to the switching position detecting device 701. If there occurs a need to execute switching from the current encoded data to the other encoded data according to the switch information 632, the switching position detecting device 701 analyzes encoded data, and detects a switching enable posi-

Aylo Exhibit 1018
Page 1924 of 2308

tion.

[0094] This means that a search is made for a frame encoded by intra-frame encoding (I-Picture).

[0095] When the switching enable position is set, an instruction for switching encoded data is supplied to the switch device 702 by using the switching instruction information 731. FIG. 8 shows an example of this instruction. Assume that encoded data A is currently selected, and there occurs a need to execute switching to encoded data B according to the switch information 632.

[0096] At this stage, assume that a frame is processed at the time of "switching instruction" shown in FIG. 8. In such a case, if encoded data is switched, an unmatched image occurs. This is because a difference from the previous image is encoded in intra-frame encoding (P-Picture). Therefore, if encoded data is switched at the time of P-Picture, an attempt is made to reproduce the next picture (picture P11 of encoded data (B)) by using immediately preceding picture (decoded picture 111 of encoding data (A)). This provides a problem that a correct decode picture cannot be obtained, and a picture is destroyed. Thus, there occurs a need to execute switching at a timing of a picture of intra-frame encoding (I-Picture) undergoing encoding by that frame only. This timing is defined as a timing at a position marked with "execute switching".

[0097] As in FIG. 9, a case in which the number of frames or time position differs depending on encoded data is considered. In this case as well, a switching enable position of encoded data at a switching destination is detected, and switching of encoded data is executed when the switching enable position is set.

[0098] At the switching position detecting device 701 shown in FIG. 7, when a need to execute switching according to the switch information 632 occurs, it is possible to employ a method for temporarily stopping an output 132 to the encoded data transmitter 103, and then, outputting new encoded data to the encoded data transmitter 103 at a stage at which the switching position is set.

[0099] This method is effective when it is necessary to suppress a buffer overflow or the like by switching encoded data. When an underflow is problematic, a stuffing bit is forcibly inserted, thereby making it possible to counteract such an underflow.

(Third Embodiment)

[0100] Now, a third embodiment according to the present invention will be described here.

[0101] The present embodiment describes an exemplary configuration of an encoding parameter determination device 401 of the video encoder 102 shown in FIG. 4 according to the first embodiment when a time interval (GOP interval) between frames encoded in a intra-frame encoded mode is adjusted based on the network condition information 136 outputted from the network state determination device 105. Here, a description will be given by focusing on differences from the first embodiment.

[0102] FIG. 13 shows an exemplary configuration of the encoding parameter determination device 401. This device comprises a GOP interval calculator (GOP interval calculator) 1501 and an encoding parameter determination device 502.

[0103] The network condition information 136 outputted from the network condition determination device 105 (including at least packet loss rate information) is input to an encoding parameter determination device 502. The parameter determination device 502 inputs packet loss rate information 1531 of the network condition information 136 to the GOP interval calculator 1501. The GOP interval calculator 1501 calculates a GOP interval from the packet loss rate information 1531. The GOP interval information 1532 is notified to the encoding parameter determination device 401. The encoding parameter determination device 401 outputs encoding parameter information 431 that contains the inputted GOP interval information 1532. When internally generated parameter information exists, the encoding parameter determination device 401 outputs GOP interval information 1532 and encoding parameter information 431 that contains the internally generated parameter information.

[0104] FIG. 14 shows an example of a method for determining a GOP interval at the GOP interval calculator.

[0105] At the step S1601, it is determined whether or not a packet loss occurs from the network condition information 136. When such a packet loss occurs, the GOP interval is estimated by formula (3) below, for example, at the step S1602.

$$Gop = (TN - TL)/\{(FN - FL) \times r\} \quad (3)$$

where "gop" denotes a GOP interval, "r" denotes a packet loss rate, FN denotes a total number of current frames, FL denotes a total number of frames during past calculation, TN denotes a time, and TL denotes a time during past calculation.

[0106] The GOP interval obtained at the step S1602 is notified to the signal processor 402 at the step S1603, and encoding is carried out based on the notified value.

[0107] On the other hand, when no packet loss has occurred at the step S1601, the default GOP thresholds are read out at the step S1604. At the step S1603, this value is notified to the signal processor 402.

[0108] According to the present embodiment, it is possible to change dynamically the GOP interval to a value suitable to a network. In this manner, when a small number of errors occur, the GOP interval is increased, thereby making it possible to reduce wasteful intra-frame encoded modes in number. In contrast, when a large number of errors occur, the GOP interval is decreased, and a frame in the intra-frame encoded mode appears quickly, thereby making it possible to acceler-

Aylo Exhibit 1018
Page 1925 of 2308

ate recovery. In the present embodiment, it is possible to efficiently select an intra-frame encoded mode.

[0109] FIG. 15 shows another example of a method for determining a GOP interval at the GOP interval calculator in accordance with a flow chart.

[0110] In this example, even when packet loss is fewer than that in the system shown in FIG. 14, the network condition is estimated from the past history of packet loss at the step S1704, and an optimal GOP interval is calculated. This technique makes it possible to estimate whether a network is completely free of an error or an error may occurs by checking the past history instead of resetting the GOP interval to the default value by determining that no packet loss occurs at one moment. In this manner, it is possible to determine the network condition more accurately.

[0111] FIG. 16 shows still another example of a method for determining a GOP interval at the GOP internal calculator in accordance with a flow chart.

[0112] In this example, the GOP interval is determined by a unique calculation system instead of determination based on the presence/absence of a packet loss. This system can be used to determine the GOP interval from the current packet loss rate in accordance with a single formula or to determine the most suitable GOP interval by checking the past history.

[0113] The following formula (4) shows an example of calculation formula for determining the GOP interval.

$$Gop = [(TN - TL)/\{(FN - FL) \times r\}] \times \alpha \qquad (4)$$

where "gop" denotes a GOP interval, "r" denotes a packet loss rate, α denotes a sensitivity coefficient, FB denotes a total number of current frames, FL denotes a total number of frames during past calculation, TN denotes a current time, and TL denotes a time during past calculation.

[0114] In this example, the GOP interval calculated from an actual value is multiplied by a sensitivity coefficient, thereby making it possible to meet a requirement for accelerating recovery by sacrificing efficiency or a requirement when one does not take care even if an effect of error is somewhat extended. For example, if the sensitivity coefficient α is set to be smaller than 1, a GOP interval which is smaller than that calculated from an actually measured value is output. In this manner, although the efficiency is sacrificed, when an error occurs, it is possible to estimate faster recovery.

[0115] The present embodiment is not limited to formulas (3) and (4). For example, in these formulas, a value is calculated by using an increment during past calculation and subsequent, it is possible to calculate the value by taking a margin such as the n preceding increment. In this manner, it is possible to cope with a gradual change hidden in a change in detailed state or to eliminate a temporal change to some extent. Similarly, algorithms are not limited to those shown in the present embodiment.

[0116] The foregoing embodiment has described an exemplary configuration when a time interval (GOP interval) is adjusted, the time interval being between the frames encoded in the intra-frame encoded mode based on the network condition information 136 at the encoding parameter determination device 401 of the video encoder 102 shown in FIG. 4 according to the first embodiment. Alternatively, there can be provided a configuration in which the network condition determination device 105 according to the first embodiment (FIGS. 1, 2, or 3) determines a time interval (GOP interval) encoded in the intra-frame encoded mode based on information from the network information or information from the network information receiver 104 and information from the video encoder 102, and provides the encoded time interval to the image encoder 102.

[0117] As described in the first embodiment, it is possible to configure a video encoder capable of determining a network condition, the video encoder having functions of the video encoder 102 and network condition determination device 105, thereby providing the previously described functions in one block.

[0118] In the foregoing, although a variety of embodiments have been described, in short, the present invention utilizes characteristics of RTP, and comprises a system (RTCP) for notifying network state information such as jitter or packet loss rate from a transmitter or receiver. At the transmitter, a bit rate of transmission data at the transmitter is adjusted according to a transmission state of a transmission channel based on the notification of network state information acquired from the receiver in the transmitter such as jitter or packet loss rate so that control such as change of the error resilience level can be carried out. Therefore, according to the above described present invention, data can be transmitted efficiently to the maximum so as to achieve data transmission that makes it possible to fully utilize transmission requiring real time properties as well.

[0119] The techniques described in the embodiments of the present invention can be distributed to be stored as a computer executable program in a recording medium such as a magnetic disk (such as a flexible disk or a hard disk); an optical disk (such as CD-ROM, CD-R, CD-RW, DVD, or MO), and a recording medium such as a semiconductor memory. Further, the above program can be distributed by means of transmission via a network.

[0120] According to the present invention, it is possible to determine a network condition, and then, optimally set an encoding parameter. In addition, when a packet loss caused by congestion and a loss of a channel error coexist in a network, it is possible to determine the loss, and then, to set a parameter.

[0121] Therefore, according to the present invention, there can be provided a data transmission apparatus and method in which data can be transmitted efficiently to the maximum so as to achieve data transmission that

Aylo Exhibit 1018
Page 1926 of 2308

makes it possible to fully utilize transmission requiring real time properties.

Claims

1. A data transmission apparatus for transmitting transmission data from a transmitter to a receiver via a transmission channel, **characterized by** comprising:

   means (105) for estimating a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and
   means (102) for changing at least one of a bit rate of transmission data and an error resilience level according to the estimated transmission condition.

2. An apparatus according to claim 1, **characterized in that** the transmitter includes means (102) for generating encoded data obtained by encoding a single content at different transmission rates, and means (603) for selecting optimum encoded data having optimum bit rate from the encoded data of the encoded data generating means (102) to adjust a bit rate of the transmission data in the transmitter according to the estimated transmission condition.

3. A data transmission method of transmitting transmission data from a transmitter to a receiver via a transmission channel, **characterized by** comprising:

   generating encoded data obtained by encoding a single content at different transmission rates;
   estimating a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and
   selecting optimum encoded data having an optimum bit rate from the encoded data to adjust a bit rate of the transmission data in the transmitter according to the estimated transmission condition.

4. A data transmission method **characterized by** comprising:

   estimating a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and
   changing at least one of a bit rate of transmission data and an error resilience level according to the estimated transmission condition.

5. A data transmitting apparatus that transmits transmission data from a transmitter to a receiver via a transmission channel, **characterized by** comprising:

   means (102) for generating plural encoded data obtained by encoding a single content at different bit rates, in the transmitter;
   means (105) for estimating a condition of the transmission channel based on Jitter information or packet loss rate; and
   means (603) for selecting encoded data having an optimum transmission rate from the plural encoded data according to the estimated transmission condition.

6. A method of transmitting transmission data from a transmitter to a receiver via a transmission channel, the method **characterized by** comprising:

   storing plural encoded data obtained by encoding a content at different bit rates;
   estimating a condition of the transmission channel based on Jitter information or packet loss rate;
   selecting encoded data having optimum transmission rate from the plural encoded data to adjust a bit rate of the transmission data in the transmitter.

7. A data transmission program stored on a computer readable medium, the program **characterized by** comprising:

   instruction means for instructing a computer to estimate a transmission condition of the transmission channel in the transmitter based on at least Jitter information or a packet loss rate obtained from the receiver; and
   instruction means for instructing the computer to change at least one of a bit rate of transmission data and a error resilience level according to the estimated transmission condition.

8. A data transmission apparatus, **characterized by** comprising:

   receiver means (104) for receiving network information regarding a condition of a network;
   determination means (105) for determining a network condition based on the network information from the receiver means;
   encoder means (102) for encoding a video data signal to generate encoded video data and control a bit rate or an error resilience level of the encoded video data according to the network condition information; and
   transmitter means (103) for transmitting the en-

Aylo Exhibit 1018
Page 1927 of 2308

coded video data to the network.

9. An apparatus according to claim 6, **characterized in that** the encoder means (102) is supplied with an encoding parameter as the network condition information and generates encoded data whose bit rate is adjusted according to the encoding parameter,

the determination means (105) includes first storage means (302) for storing encoding parameter information obtained by the encoder means, second storage means (301) storing the network information input by the network information receiver, and determination means (303) for determine the encoding parameter by deciding the network condition based on network information input by the network information receiver, past encoding parameter information from the first storage means and past network information from the second storage means.

10. An apparatus according to claim 8, **characterized in that** the encoder means (102) includes parameter determination means (401) for determining the encoding parameter from the network condition information input from the receiver means, the encoder means (102) encoding the video data signal to generate the encoded video data whose bit rate is adjusted in correspondence with the encoding parameter determined by the determination means (401).

11. An apparatus according to claim 10, **characterized by** including encoding determination means (501) determine whether a next frame is to be encoded forcibly in an intra-frame encoding mode, when the parameter determination means (401) determines the encoding parameter using the network condition information.

12. An apparatus according to claim 8 or 9, **characterized in that** the determination means (105) includes computation means for computing a GPO interval from the network information input from the receiver means (104) to generate network condition information including GOP interval information indicating the GOP interval obtained by the computation means, and the encoder means (102) encodes the video data signal according to the GOP interval information included in the network condition information.

13. An apparatus according to claim 8 or 10, **characterized in that** the encoder means (102) includes GOP interval computation means for computing a GOP interval based on the network information input from the receiver means (104) and encodes the video data signal according to the GOP interval information indicating the GOP interval computed by the GOP interval computation means.

14. An apparatus according to claim 12 or 13, **characterized in that** the determination means (105) determines at least a packet loss rate as the network condition, and the GOP interval computation means computes the GOP interval based on at least the packet loss rate.

15. An apparatus according to claim 8, **characterized by** including switching means (602) for switching among plural encoded data encoded at different transmission rates, respectively, to select encoded data having a transmission rate corresponding to switching information, and selector means (603) for selecting a transmission rate of encoded data to be transmitted from the transmitter based on the network condition information outputted from the determination means (105) and output the selected transmission rate as the switching information to the switching means.

16. An apparatus according to claim 15, **characterized in that** the switching means (602) includes a switching position detector for detecting a switching position at which one of the encoded data can be switched to the other, and when an encoded data switching signal is generated by the switching information output from the selector means (603), the switching position detector detects the switching position, to switch the encoded data at the switching position.

Aylo Exhibit 1018
Page 1928 of 2308

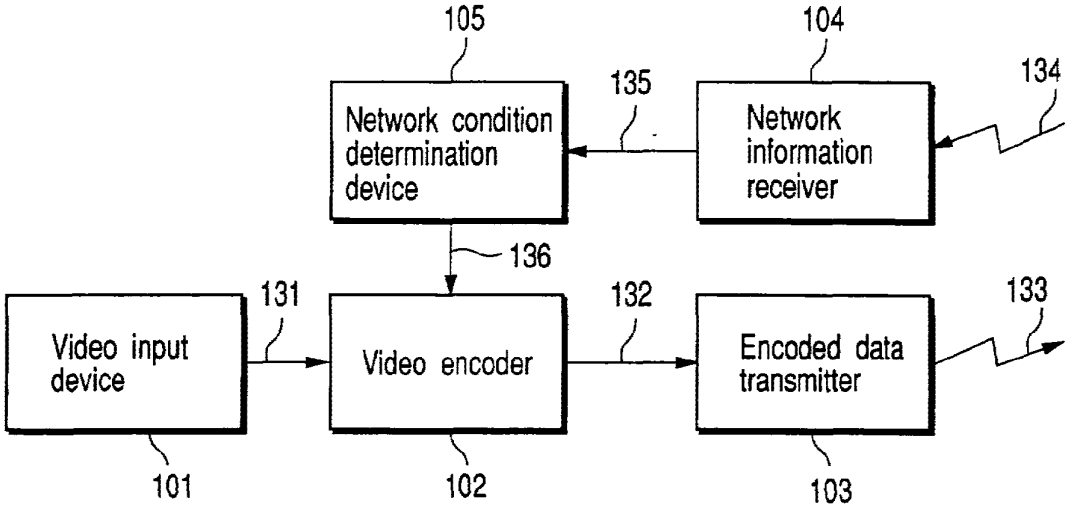


FIG. 1

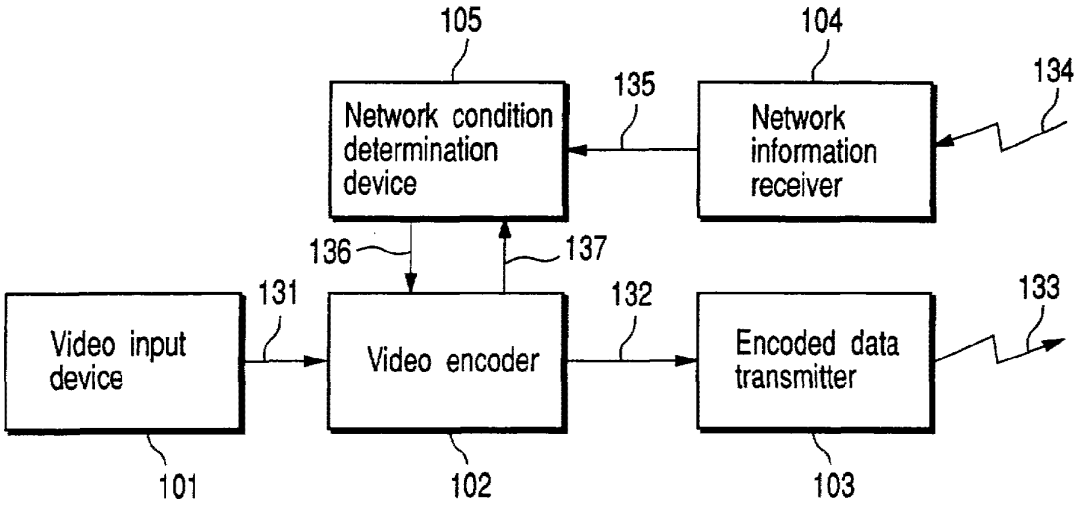


FIG. 2

Aylo Exhibit 1018
Page 1929 of 2308

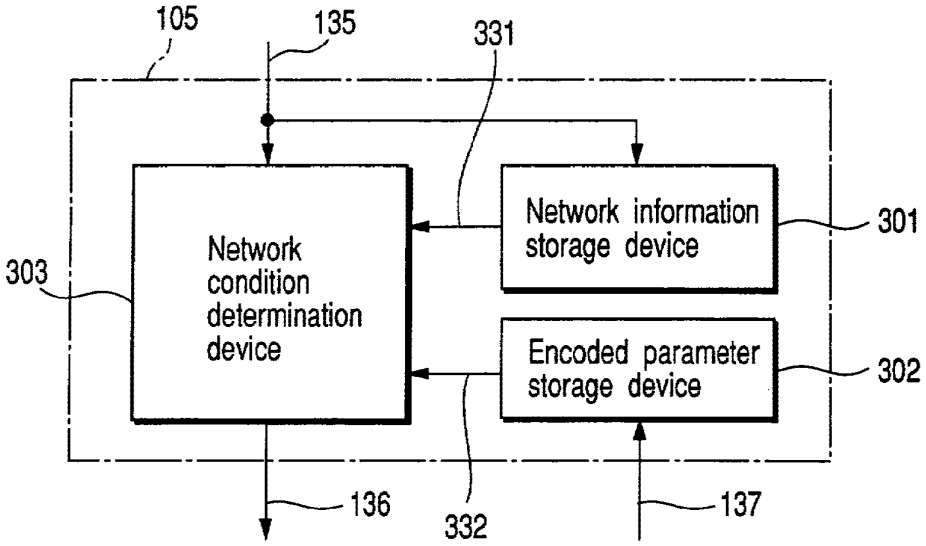


FIG. 3

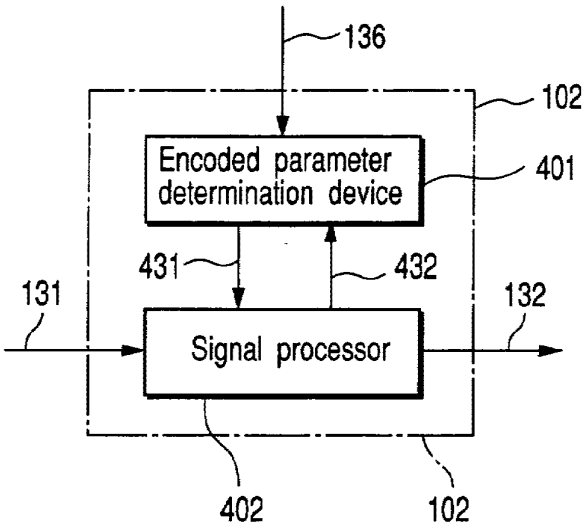


FIG. 4

Aylo Exhibit 1018
Page 1930 of 2308

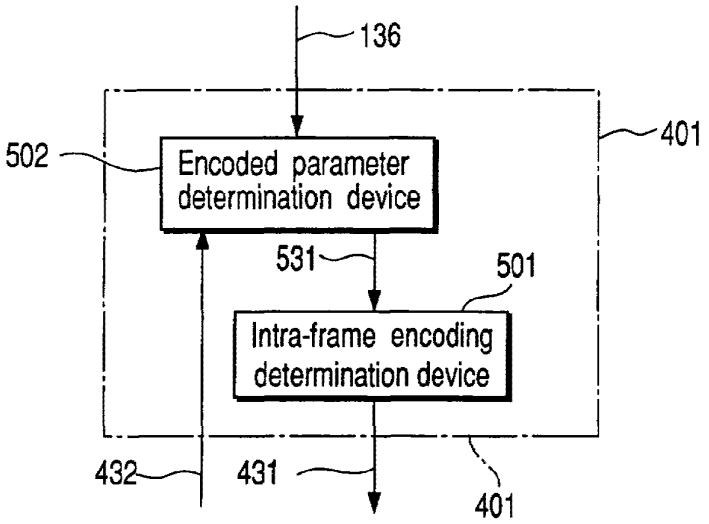


FIG. 5

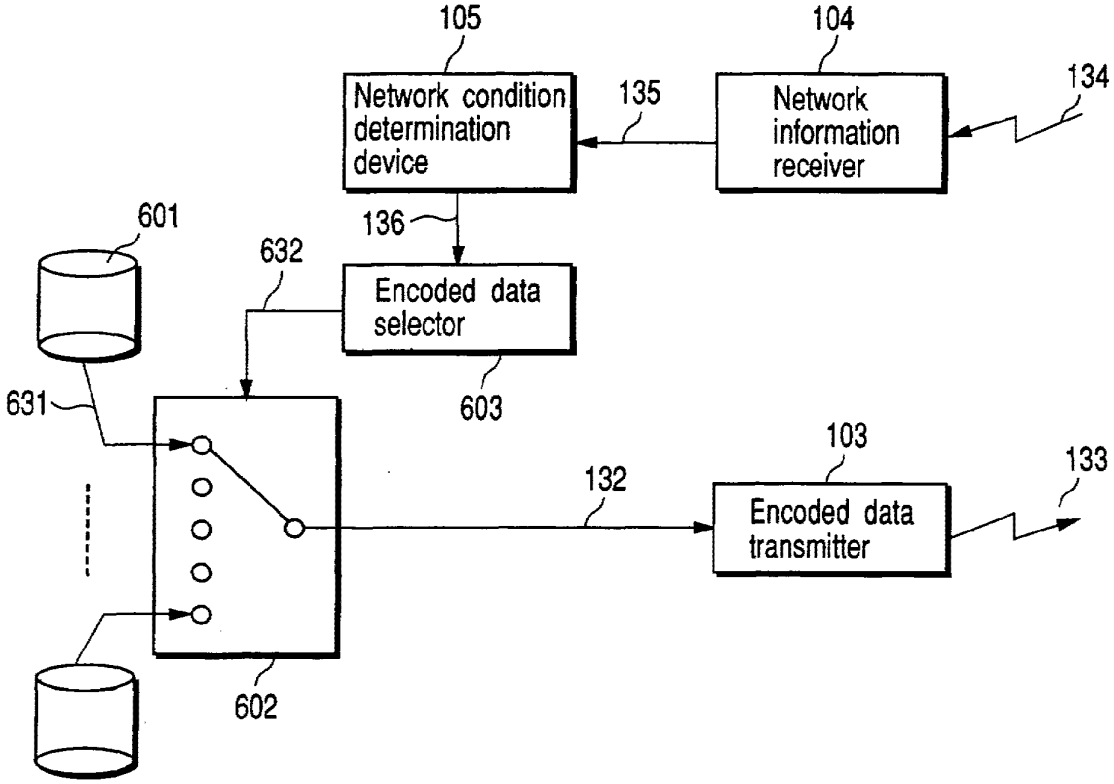


FIG. 6

Aylo Exhibit 1018
Page 1931 of 2308

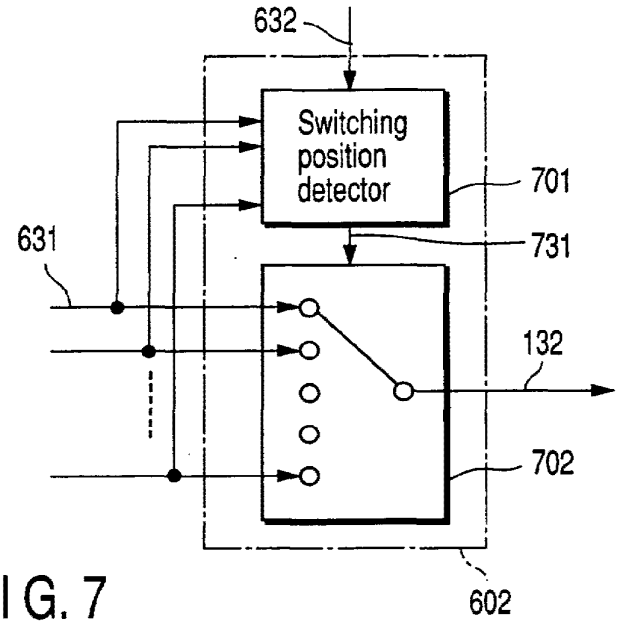


FIG. 7

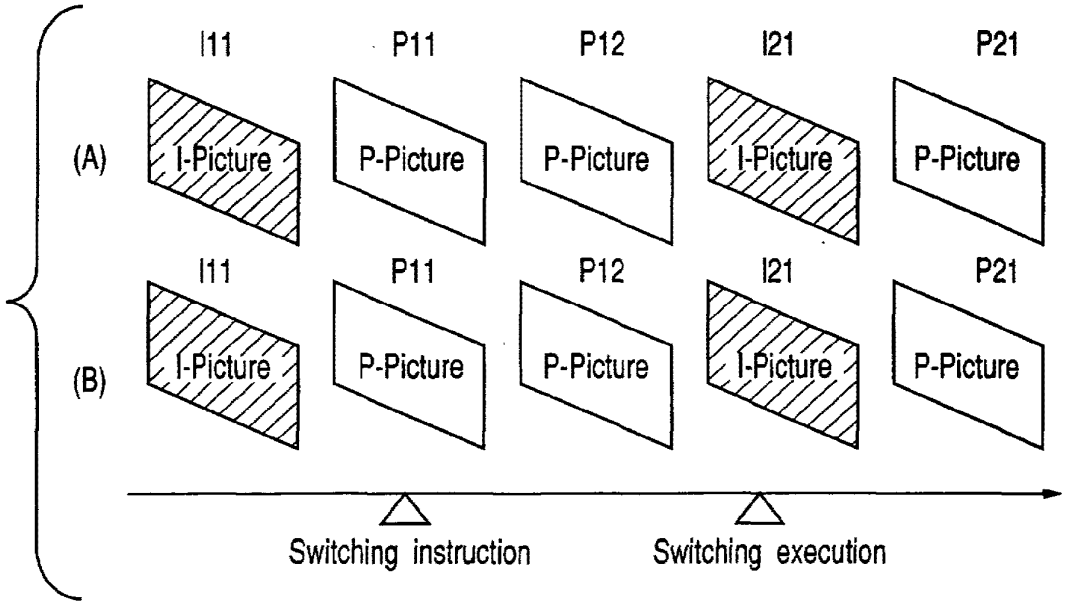


FIG. 8

Aylo Exhibit 1018
Page 1932 of 2308

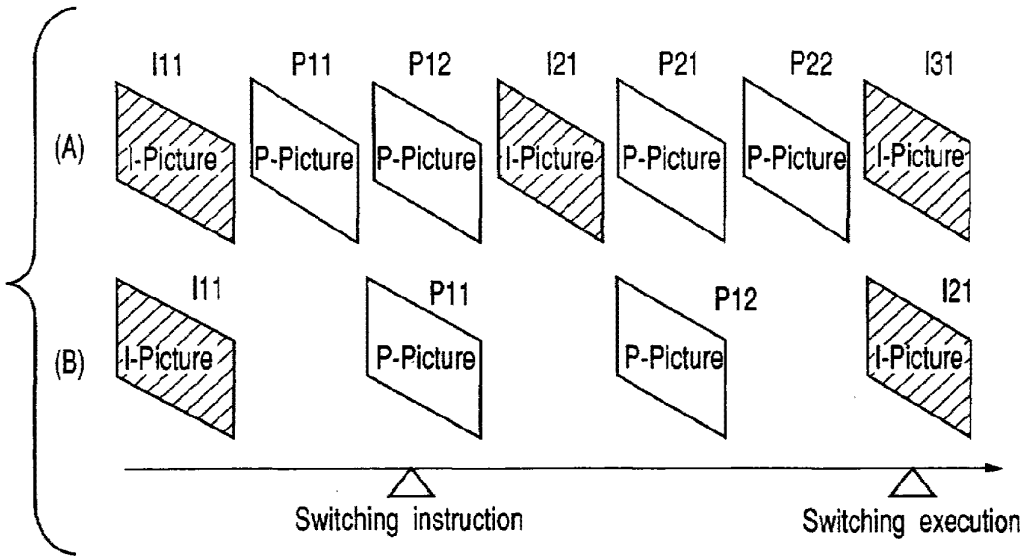


FIG. 9

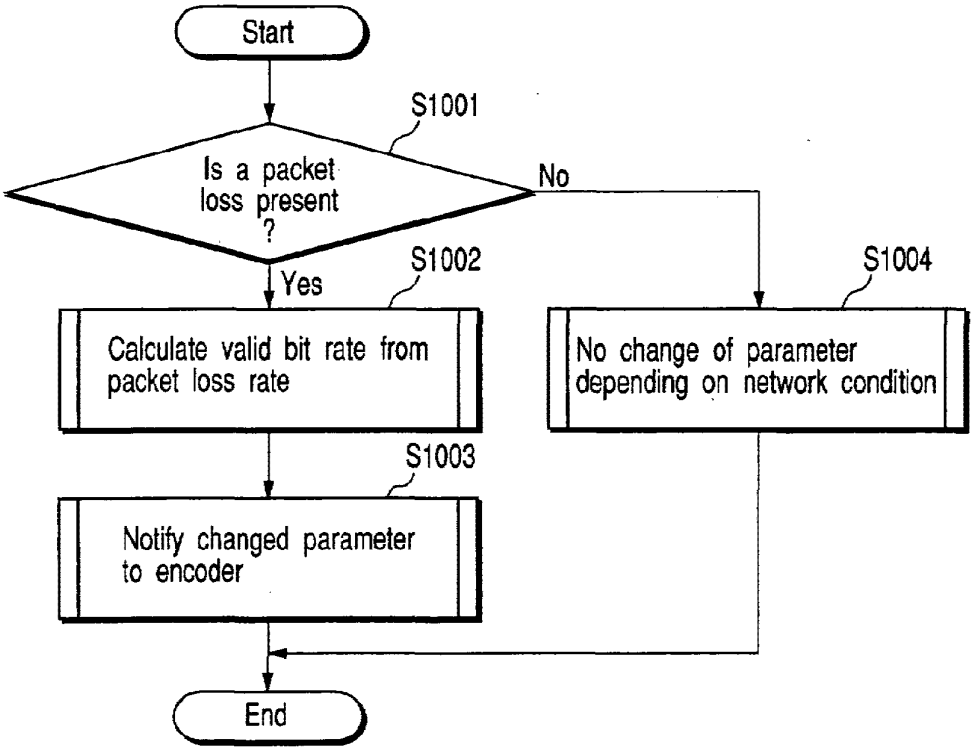


FIG. 10

Aylo Exhibit 1018
Page 1933 of 2308

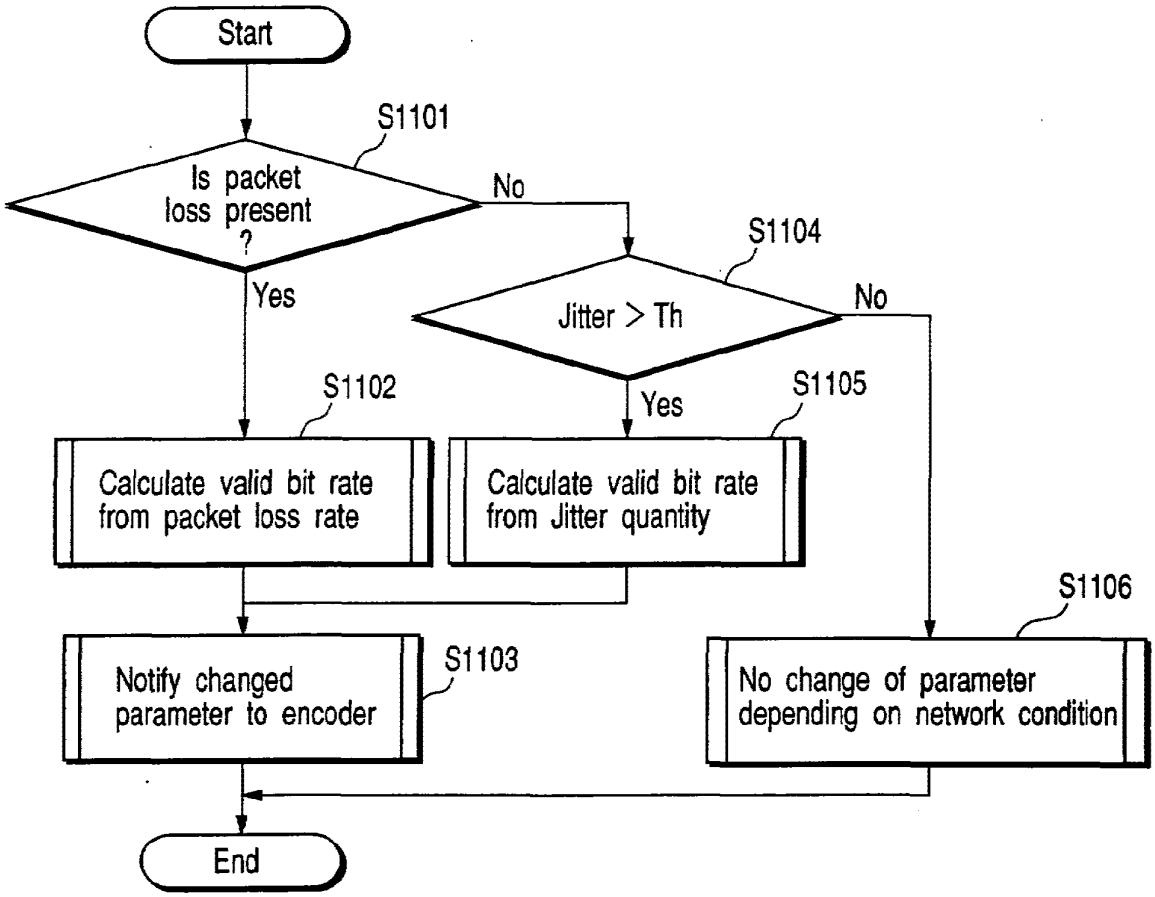


FIG. 11

Aylo Exhibit 1018
Page 1934 of 2308

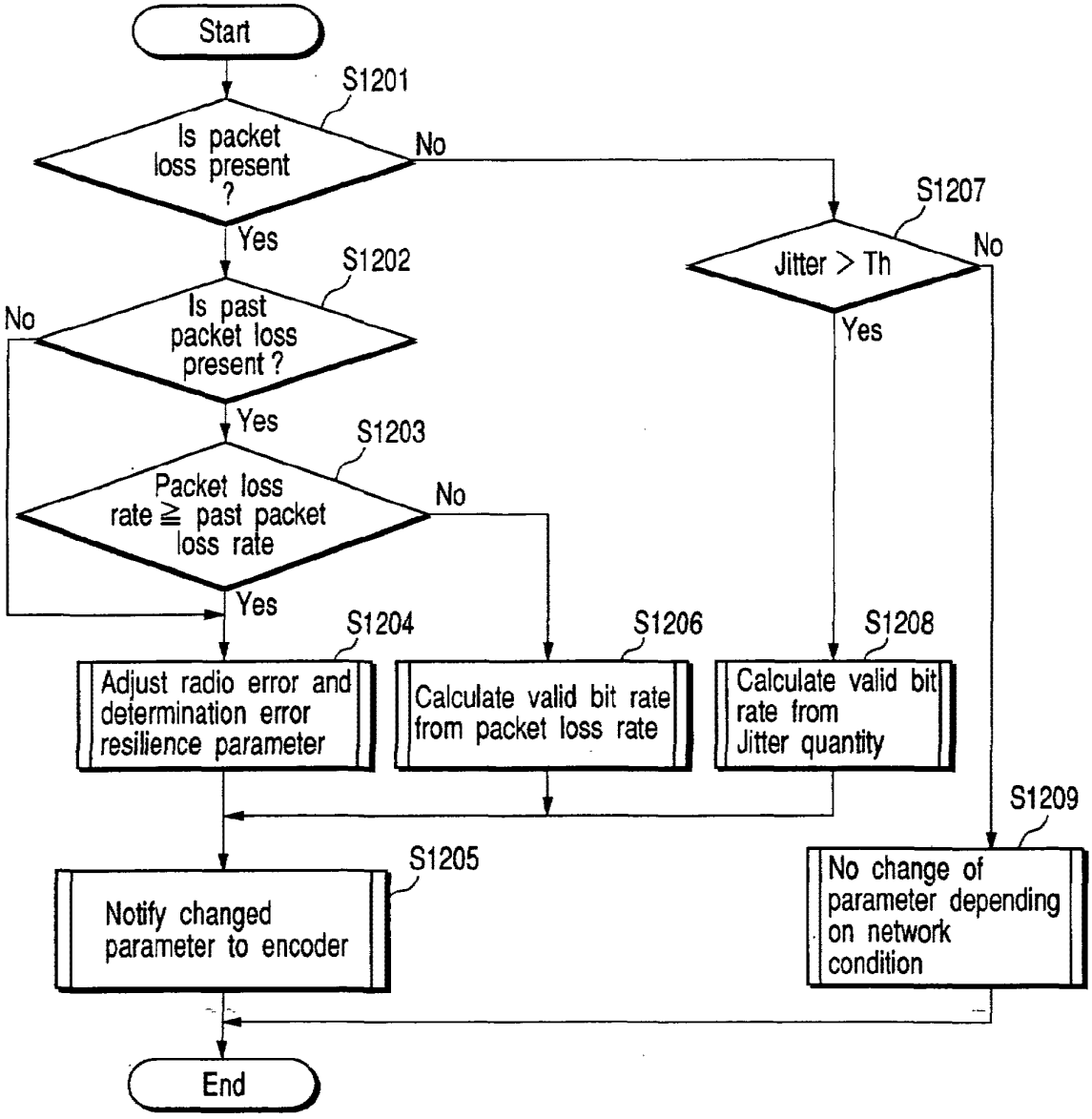


FIG. 12

Aylo Exhibit 1018
Page 1935 of 2308

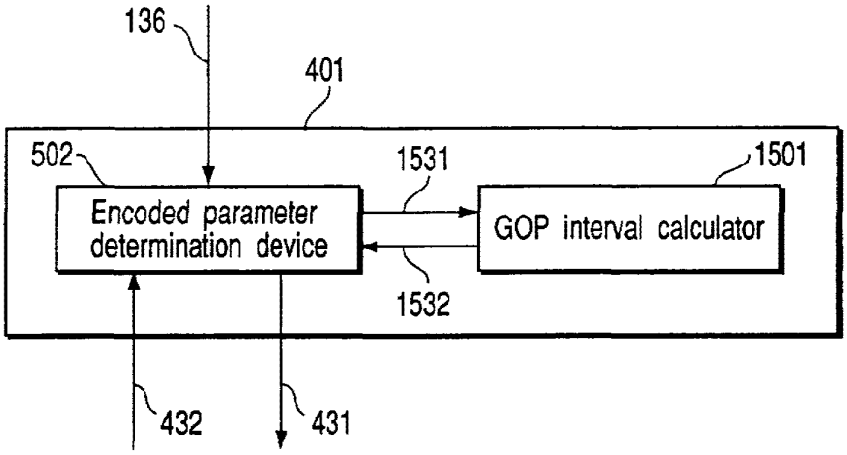


FIG. 13

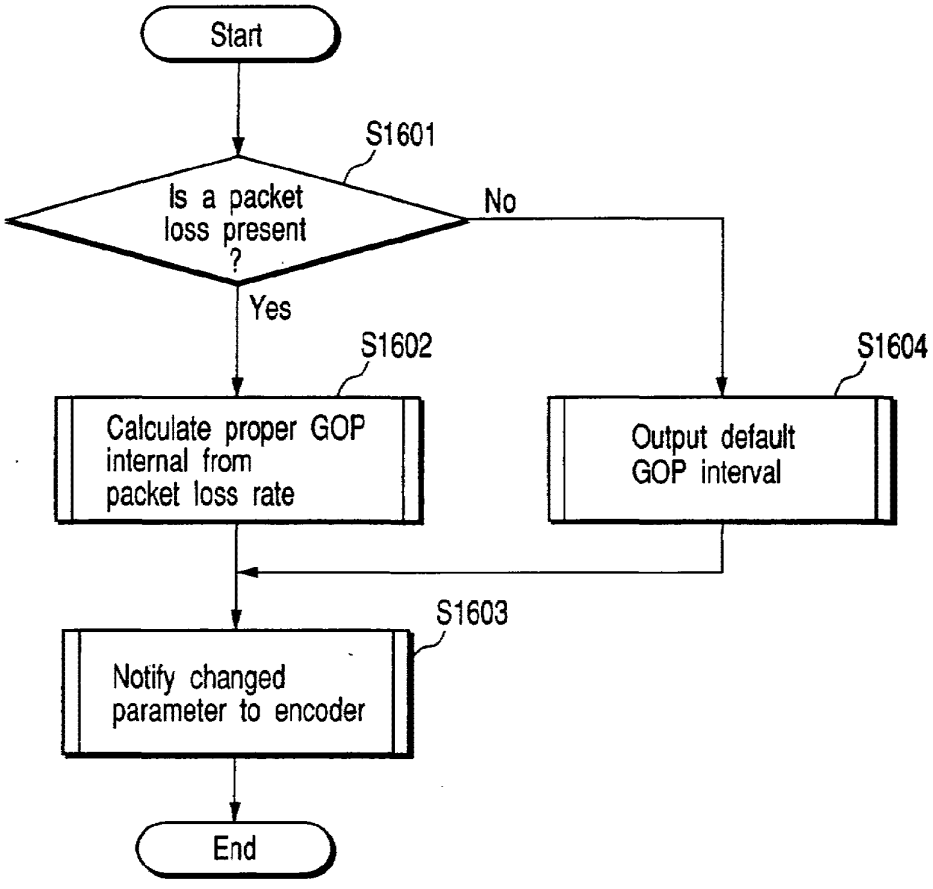


FIG. 14

Aylo Exhibit 1018
Page 1936 of 2308

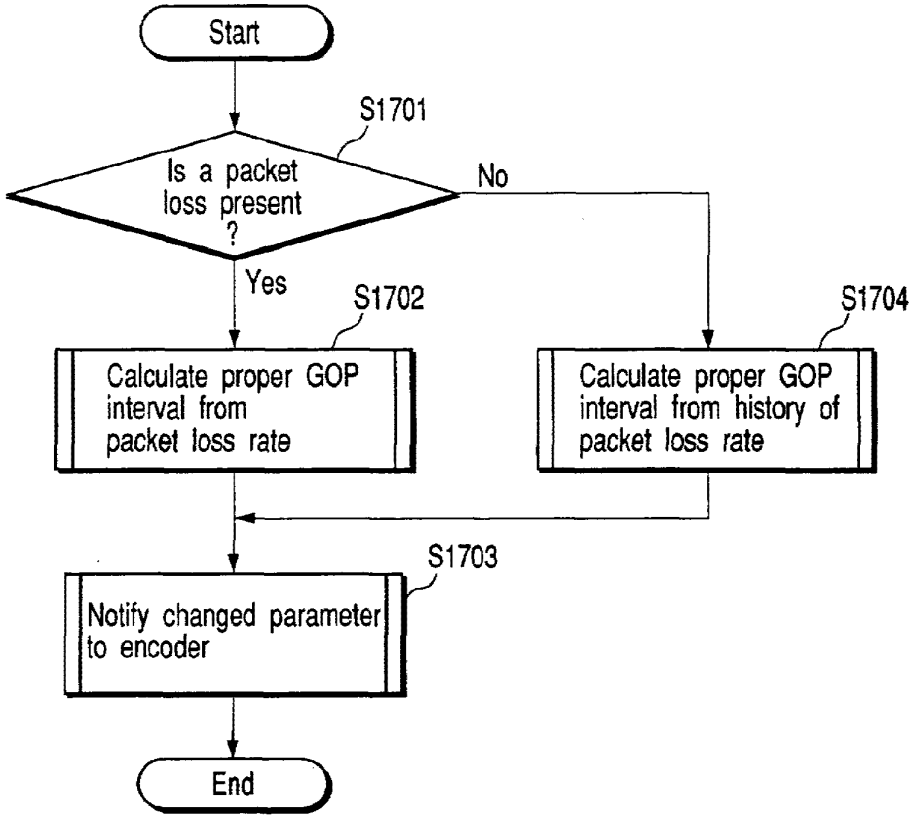


FIG. 15

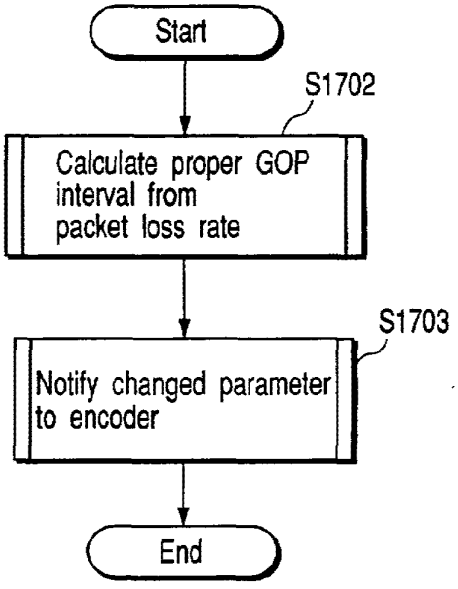


FIG. 16

Aylo Exhibit 1018
Page 1937 of 2308

(12) UK Patent Application (19) GB (11) 2 367 219 (13) A

(43) Date of A Publication 27.03.2002

(21) Application No 0023062.3

(22) Date of Filing 20.09.2000

(71) Applicant(s)
Vintage Global
(Incorporated in the British Virgin Islands)
c/o Ernst & Young Trust Corp (BVI) Ltd, P.O Box 3340, Road Town, Tortola, British Virgin Islands

(72) Inventor(s)
Benjamin Neeman
Yossi Rubner
Nehemia Davidson
Uri Salomon

(74) Agent and/or Address for Service
Geoffrey L Melnick
c/o Solo Photographic, 12 Catherine Street, ST ALBANS, Herts, AL3 5BX, United Kingdom

(51) INT CL[7]
H03M 7/30 , H04L 1/00

(52) UK CL (Edition T)
H4P PDCFM

(56) Documents Cited
EP 1049074 A1 **WO 99/00984 A1**
WO 98/09387 A1 **WO 00/72601 A2**
US 5847760 A

(58) Field of Search
UK CL (Edition S) **H4P PDCFM PDCFP PDCFX PF**
INT CL[7] **H03M 7/30 , H04L 1/00 , H04N 7/26 7/32**
Online: WPI, EPODOC, JAPIO

(54) Abstract Title
Streaming of media file data over a dynamically variable bandwidth channel

(57) A method of transmitting data streams over a variable bandwidth channel, comprising modifying the quantity of data to be injected into the channel to accord with a measured capacity or bandwidth of the channel. An input stream is passed through a plurality of parallel compressors 40.1-40.4 which each produce an output stream of a different rate to the other streams. Each stream preferably comprises key or reference frames and delta frames. The reference frames are synchronized between the streams so that a data unit in one stream contains the same information as the corresponding data unit in another stream, but is compressed to a different extent. A data stream for transmission along the channel is produced by selecting data units from more than one compressor output stream so as to produce a stream which makes optimum use of the available channel bandwidth. In one embodiment the duration of the data units are one of a group comprising fixed predetermined durations and duration determined by changes in the channel bandwidth. The method is useful in sending multimedia data streams such as audio and video over a channel such as a cellular telephone network or the internet.

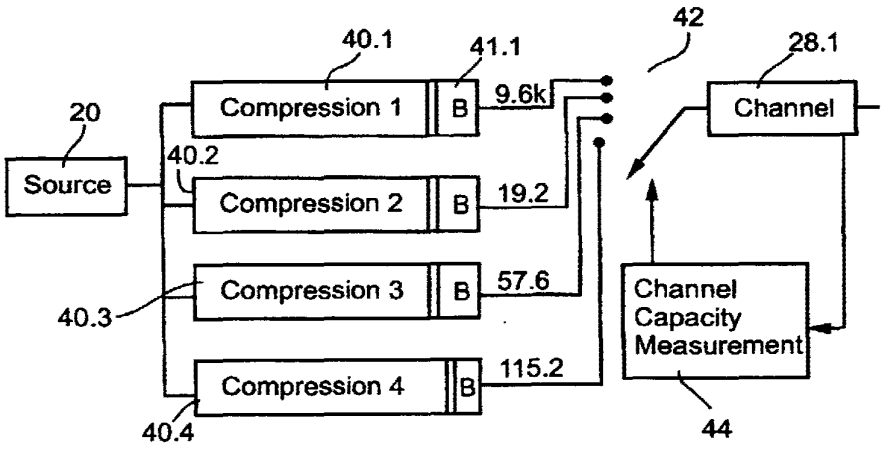


Fig. 4

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

GB 2 367 219 A

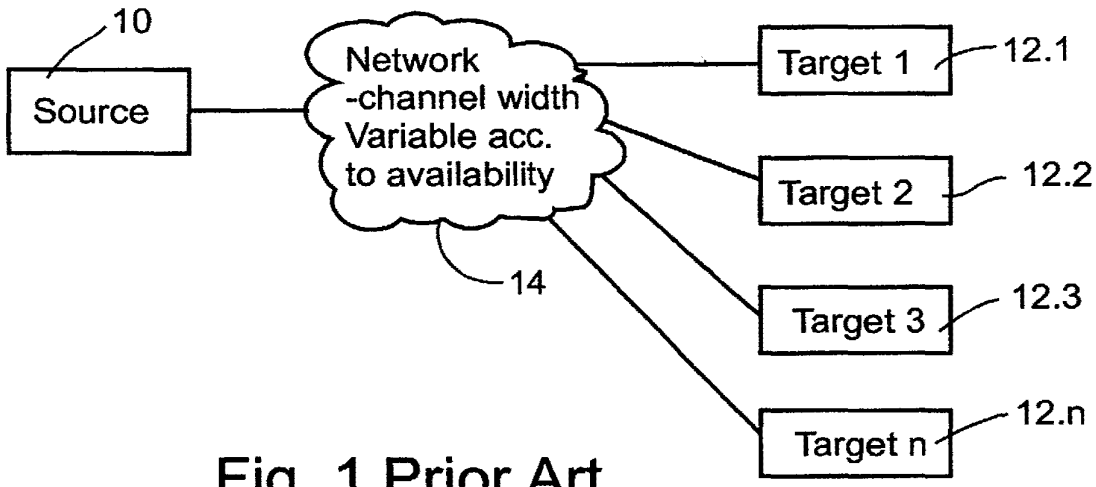


Fig. 1 Prior Art

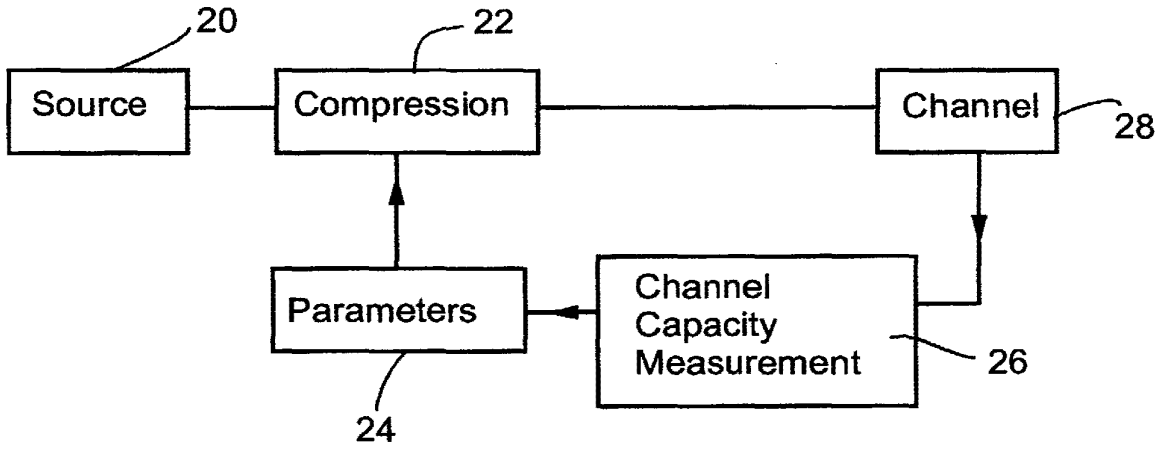


Fig. 2

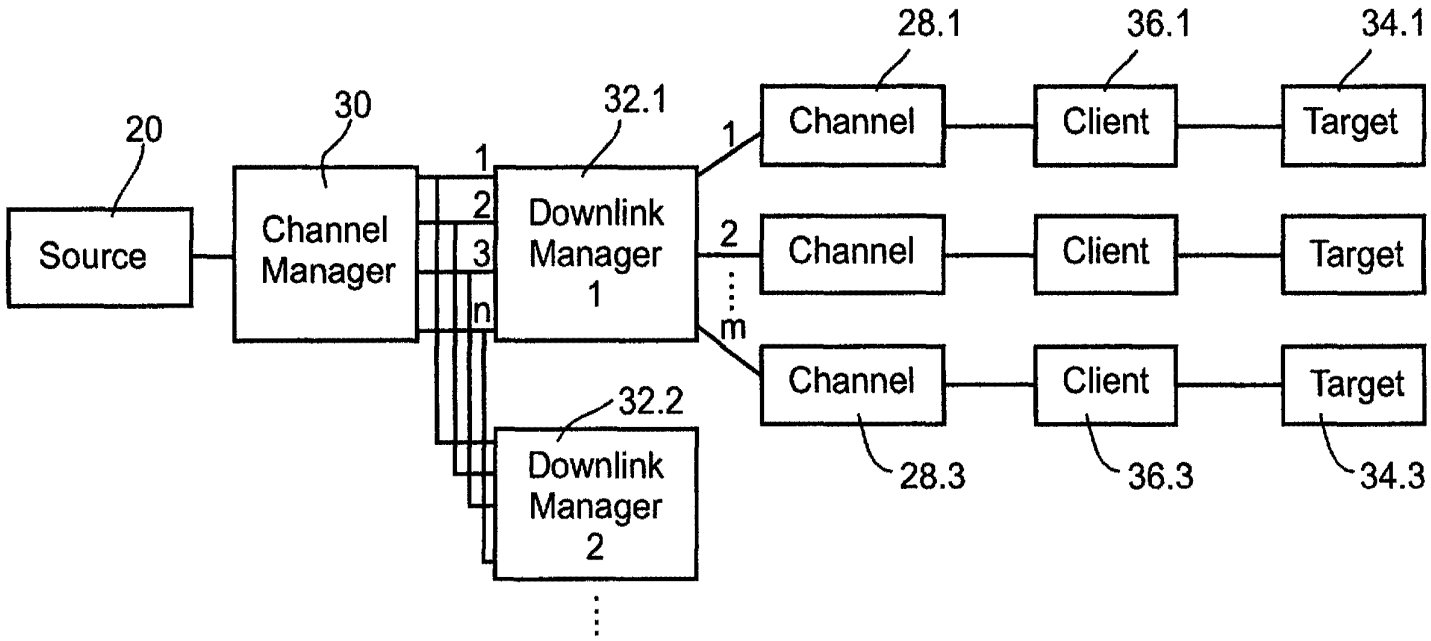

Source
20
Channel Manager
30
Downlink Manager 1
32.1
Downlink Manager 2
32.2
1
2
3
n
1
2
m
Channel
Client
Target
28.1
36.1
34.1
Channel
Client
Target
Channel
Client
Target
28.3
36.3
34.3

Fig. 3

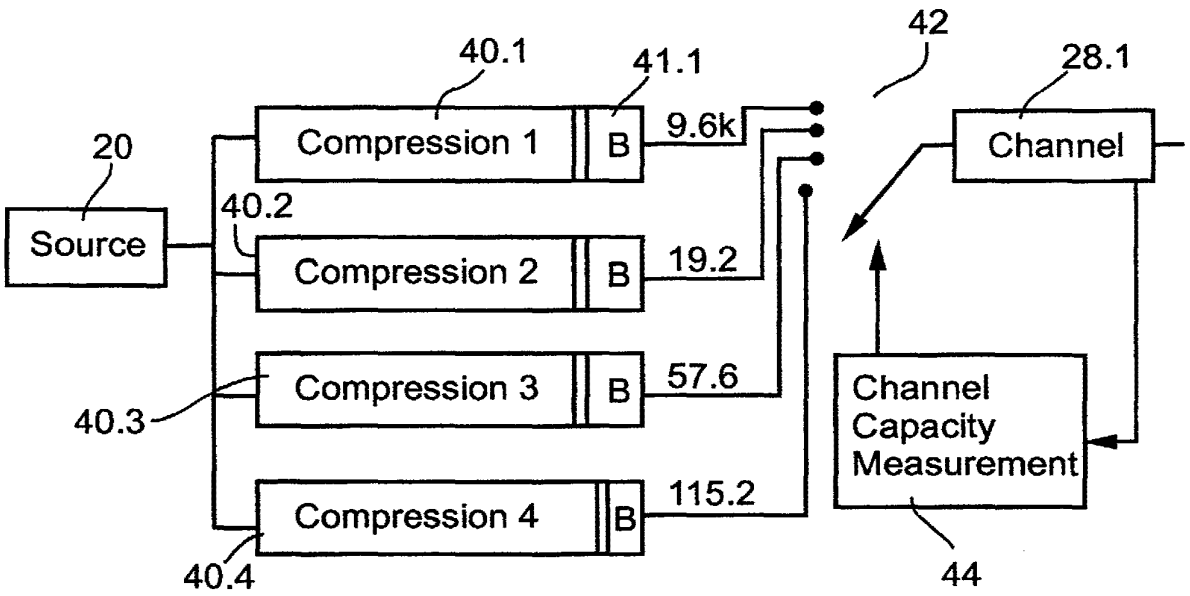

40.1
41.1
42
28.1
20
Compression 1
B
9.6k
Channel
40.2
Source
Compression 2
B
19.2
Compression 3
B
57.6
40.3
Channel
Capacity
Measurement
Compression 4
B
115.2
40.4
44

Fig. 4

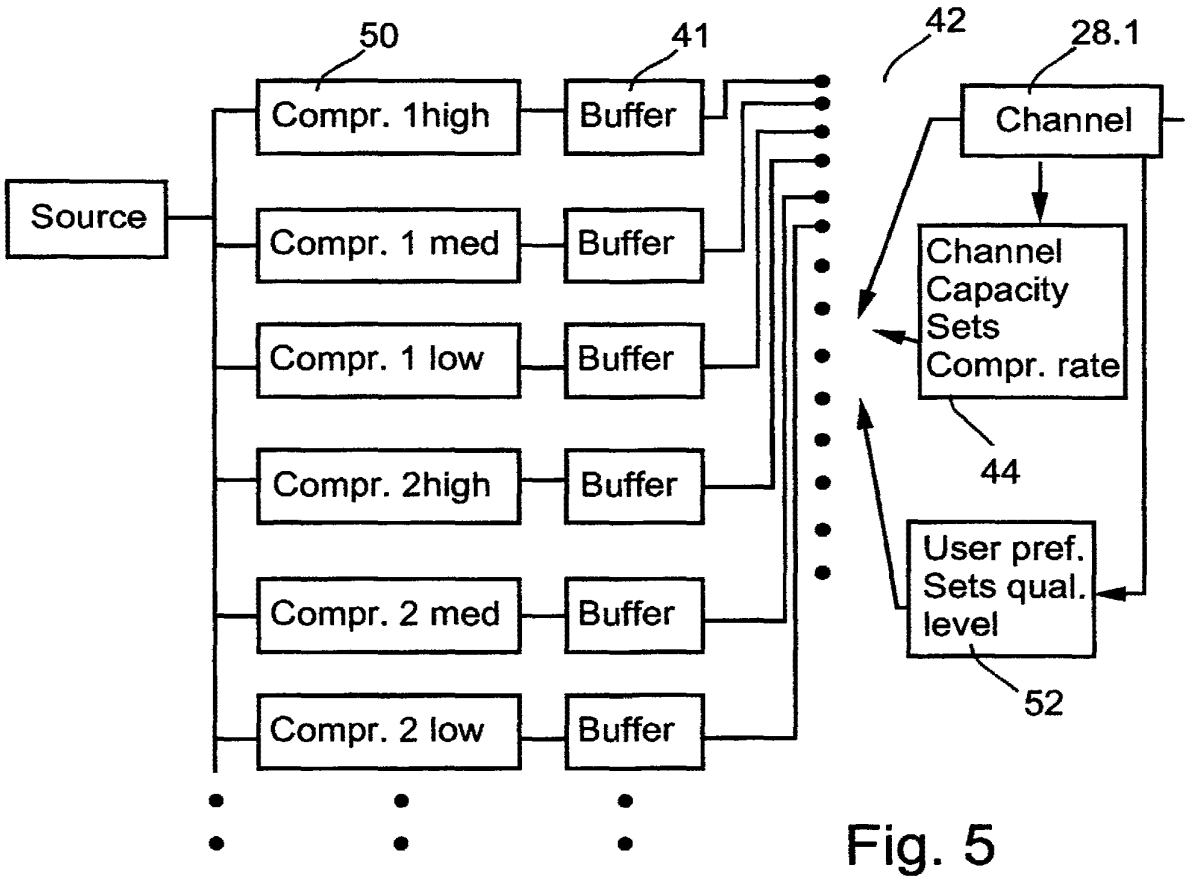

50
41
42
28.1
Compr. 1high
Buffer
Channel
Source
Compr. 1 med
Buffer
Channel
Capacity
Sets
Compr. rate
Compr. 1 low
Buffer
44
Compr. 2high
Buffer
User pref.
Sets qual.
level
Compr. 2 med
Buffer
52
Compr. 2 low
Buffer

Fig. 5

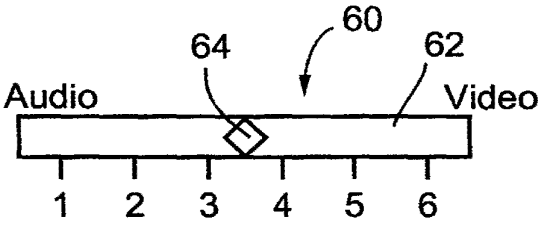


Fig. 6

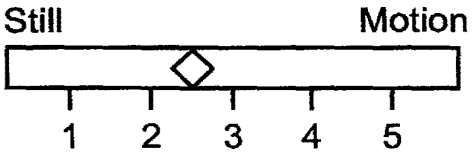


Fig. 7

00/20757 -UK. 9/12/00

Title of the Invention

Streaming of Media File Data Over a Dynamically Variable Bandwidth Channel

Field of the Invention

The present invention relates to streaming of media file data over a channel having a dynamically variable bandwidth, and more particularly but not exclusively to streaming of multimedia data such as video and /or audio data over a channel having a dynamically variable bandwidth.

Background of the Invention

Dynamically varying channel bandwidths are particularly encountered over the Internet and like networks and over radio and cellular telephone links. On the Internet the total bandwidth is fixed and the bandwidth available per channel tends to depend on the amount of usage, the greater the usage the smaller the bandwidth available per user. In radio and cellular links the available bandwidth depends on usage, but the overall capacity may also be dependent on environmental factors.

The Internet relies on packet switched communication, whereas radio and cellular telephone links may be either circuit-switched or packet switched. Either way, the same considerations of channel bandwidth broadly apply.

As the bandwidth changes, a number of approaches are currently in use. One of these is simply to buffer received video frames, to start to show the frames as soon as the buffer is full, and to continue to show the video until the buffer is empty, then to halt the display until the buffer is full again. Another method is to show the video at a real time rate, displaying all frames that arrive in time and dropping all frames that arrive too late.

In both of these methods, data is sent at the same high rate. In the first method, unacceptable delays in showing the data are experienced. In the second method, valuable channel bandwidth is wasted transmitting data that does not get used.

Whether the network is a cellular telephone network or the Internet or like wide area data network, the bandwidth of the data network is variable. In all cases the

larger the number of users the lower the bandwidth per user, and radio-based networks may also suffer from reduced bandwidth due to environmental conditions. Furthermore, in the case of cellular telephone networks, events such as handover between two base stations may also affect the available bandwidth.

In general, cellular networks are able to absorb changes in bandwidth without the user noticing a change in the quality of the voice signal. However, it is desirable to go beyond mere voice communication over cellular telephone networks. Already, low level data capability is being built into mobile telephones and it is envisaged that within the next few years, mobile telephones will be able to receive and display multimedia data. Multimedia data uses large amounts of bandwidth and generally requires to be displayed in real time. Thus bandwidth fluctuations will be immediately noticeable. Currently, however, transmission of video data over cellular phone channels is not being attempted on a commercial basis.

At present, mobile telephone systems are on their second generation and a third generation is on the drawing board but will not be available for some years. The problem is particularly associated with the design of advanced second generation equipment (generation 2.5) which, it is envisaged will be able to provide multimedia within the limitations of second generation capacity. It should be noted however that the same considerations apply to further generations of mobile telephones and also to other environments in which it is desired to transmit multimedia data, such as the Internet.

Generally, multimedia data is sent over a network as a datastream in compressed form. There are numerous algorithms for compressing data, and different forms of compression are appropriate for compressing different forms of data. Furthermore, a single form of compression may give rise to different levels of compression according to compression parameters that may be set. The greater the compression the lower will be the resulting data rate and a high level of compression may be used to overcome channel capacity constraints. A solution may involve selection of the compression parameters to meet a likely worst-case scenario. However, the solution is not complete as a high level of compression leads to low quality output and no advantage is taken of the availability for most of the transmission time of a higher bandwidth availability.

US Patent No. 6,014,694, Aharoni et al, discloses a system for adaptively transporting video over networks wherein the bandwidth varies with time. The

Aylo Exhibit 1018
Page 1944 of 2308

system comprises a video/audio codec that functions to compress, encode, decode and decompress video streams that are transmitted over networks having available bandwidths that vary with time and location. Depending on the channel bandwidth, the system adjusts the compression ration to accommodate a plurality of bandwidths ranging from 20kbps for POTS to several Mbps for switched LAN and ATM environments. Bandwidth adjustability is provided by offering a trade-off between video resolution, frame rate and individual frame quality. The system generates a video data stream comprising of key, P and B frames from a raw video source. Each frame type is further comprised of multiple levels of data representing varying degrees of quality. In addition, several video server platforms can be utilized in tandem to transmit video/audio information with each video server platform transmitting information for a single compression/resolution level.

P and B frames are delta frames that depend on previous frames to allow a complete picture to be generated. Thus it is not possible simply to mix groups of frames of different quality levels since the deltas between frames are different. The above US Patent solves the problem by taking advantage of what is known as a group of pictures (GOP) which is a self-contained unit such as a scene. Because they are self-contained, GOPs of different compression levels can be successfully mixed. However, GOP boundaries are content related and thus irregular. When GOP boundaries are widely spaced, dynamic following of bandwidth changes are highly inaccurate or generate delays for new users.

Summary of the Invention

Embodiments of the present invention attempt to solve the above problems by modifying the quantity of data to be injected into the channel to accord with a measured capacity or bandwidth of the channel. The quantity of data is preferably modified only by altering compression parameters, so that data fully usable in real time can be received from the channel. Thus a reduction in bandwidth of the channel leads to a drop in quality of the received signal and not to dropped frames or a reduction in performance.

According to a first aspect of the present invention there is thus a device for streaming of media file data to be sent over at least one channel having a dynamically variable bandwidth, comprising

an input for receiving media file data at a first rate,

a multiple stream creator for creating a plurality of streams of said media file data, each stream compressed to predetermined other rates, each stream comprising a series of autonomous compressed data units which are mutually compatible, and

a stream selector for dynamically matching a bandwidth of the channel to rates of said streams to send said data file as a combination of said autonomous units from said streams via said channel,

wherein said autonomous units have durations which are one of a group comprising fixed predetermined durations and durations determined by changes in said channel bandwidth.

Preferably, said stream selector is operable to select a combination of said streams in succession said combination being chosen to produce an average stream rate over a period of time which matches said channel bandwidth.

Preferably, said bandwidth of said channel is liable to vary dynamically and wherein said stream selector is operable to repeat said matching throughout said streaming operation.

Preferably, said bandwidth of said channel is liable to vary dynamically and wherein said stream selector is operable to repeat said matching throughout said streaming operation.

Preferably, said autonomous units match each other across said plurality of streams, thereby to enable switching between said streams.

Aylo Exhibit 1018
Page 1946 of 2308

Preferably, said data file is arranged in said streams in a compressed media format utilizing reference frames and delta frames and wherein each autonomous unit is self-contained in respect of reference frames.

In an embodiment, said channel is a cellular telephone channel. Alternatively, said channel is an Internet channel.

Preferably, the predetermined other rate is different for each stream. In another embodiment, the predetermined other rate is identical for a set of said streams, each stream in said set being assigned a different balance between quality and speed of said media.

A preferred device comprises a plurality of media types, said predetermined other rate being identical for a set of said streams, each stream in said set being assigned a different balance between quality of said different media types.

Preferably, streams in said set are selectable by a signal received from said channel.

Preferably, streams in said set are selectable by a signal received from said channel.

An embodiment comprises a detector for detecting a change in the bandwidth of said at least one channel and operable to request a corresponding change in the data rate, and wherein the device is operable, upon receipt of said request, to set a delay, and to issue a reference frame within said delay.

The device preferably issues said reference frame at the end of said delay.

The device alternatively issues said reference frame at the occurrence of a largest detected interframe change within the time delay.

According to a second aspect of the invention there is provided a multiple stream creator for receiving a data stream at a first data rate and converting said data stream into a plurality of data streams at a plurality of data rates, said creator comprising an input for receiving said data stream at said first data rate, and a plurality of outputs, each operable to receive said data stream at said first rate and to output said data stream at a predetermined rate individually set for each output as a series of autonomous units having a duration, and wherein said duration is one of a group comprising a fixed predetermined duration and a duration dynamically changeable in accordance with bandwidth conditions in a transmission channel.

Preferably said outputs each comprise data buffers.

Aylo Exhibit 1018
Page 1947 of 2308

The multiple stream creator preferably comprises a data stream processor for processing said data stream received at said first rate into said plurality of autonomous units, said autonomous units being synchronized blocks, thereby allowing said output streams to be mixed in said transmission channel to produce continuously usable output.

Preferably, said received data stream is in an uncompressed data format and said data is compressed into a compressed data format independently for each one of said output data streams.

According to a third aspect of the invention there is provided a method of sending a datastream over a variable bandwidth channel comprising modifying a quantity of data to be injected into the channel to accord with a measured capacity of the channel, said method comprising the steps of

compressing the datastream, using an algorithm having compression parameters, to produce a plurality of compressed datastreams each comprising reference frames and delta frames, each compressed datastream produced by using the algorithm together with different parameters, and

synchronizing reference frames between the compressed datastreams so that combinations of said compressed datastreams can be sent via said channel.

Preferably, said step of synchronizing reference frames is carried out at intervals determined by a dynamic status of said channel bandwidth.

Alternatively or additionally, said step of synchronizing reference frames is carried out at predetermined intervals.

Alternatively or additionally, said combination of said compressed datastreams is selected substantially dynamically to match the measured capacity of the channel.

Preferably, a change in bandwidth of a channel causes a request for commencement of a new autonomous unit, and wherein said request causes initiation of measurement of a time delay, and wherein a new autonomous unit is commenced within said time delay.

Preferably, the new autonomous unit is commenced at the end of said time delay.

Preferably, the step is provided of determining the largest delta frame during the time delay and sending a reference frame, to initiate said new autonomous unit, in place of said determined largest delta frame.

Aylo Exhibit 1018
Page 1948 of 2308

According to a fourth aspect of the present invention there is provided a tool for controlling receipt of media file data over a network, wherein said media file data is available at a plurality of quality levels, the tool comprising a quality selector to enable a user to select between said different quality levels for receipt.

Preferably, said media file comprises a plurality of data groups and each quality level is a different balance between quality levels of the plurality of data groups.

Preferably, said quality selector comprises a slider operable to be moved along a scale, the position of the slider on the scale indicating a quality level.

Brief Description of the Drawings

For a better understanding of the invention and to show how the same may be carried into effect, reference will now be made, purely by way of example, to the accompanying drawings, in which:

Fig. 1 is a generalized diagram showing a data source sending data to a series of targets over a variable bandwidth network,

Fig. 2 is a generalized block diagram showing a first embodiment of the present invention in which feedback concerning available bandwidth within a channel is used to modify parameters governing compression of the datastream,

Fig. 3 is a generalized block diagram showing a second embodiment of the present invention in which a series of compressed datastreams are continually created and are selected for use in individual channels according to measured parameters of the channel,

Fig. 4 is a generalized block diagram showing the channel manager and part of a downlink manager of Fig. 3 in greater detail,

Fig. 5 is a generalized block diagram showing a third embodiment of the present invention in which the user is able to set a quality level of the data that is being sent,

Fig. 6 is a generalized schematic diagram of a quality level selection tool for use by a user in selecting a quality level in the embodiment of Fig. 5, and

Fig. 7 is a generalized schematic diagram of another quality level selection tool for use with the device of Fig. 5.

Aylo Exhibit 1018
Page 1949 of 2308

Description of the Preferred Embodiments

Reference is now made to Fig. 1, which is a generalized diagram showing a data source 10 sending data to a series of targets 12.1..12.n over a variable bandwidth network 14. The network may be a cellular telephone network or the Internet or like wide area data network or a combination. In both cases, the bandwidth of the data network is variable. In all cases the larger the number of users the lower the bandwidth per user, and radio-based networks may also suffer from reduced bandwidth due to environmental conditions. Furthermore, in the case of cellular telephone networks, events such as handover between two base stations may also affect the available bandwidth.

Data is generally compressed before sending over the channel. The level of compression may be preselected to accord with known or likely channel capabilities, as described in the introduction. However such a solution has to concentrate on the lower margins of available bandwidth and thus fails to make full use of higher bandwidth levels when available. Conversely, solutions which send datastreams at higher data rates and do not take the lower bandwidth levels into account lead to unacceptable breaks and delays in performance at the receiving end.

Reference is now made to Fig. 2 which is a generalized block diagram showing a first embodiment of the present invention in which feedback concerning available bandwidth within a channel is used to modify parameters governing compression of a datastream.

A multimedia datastream is received from a source 20 such as a host or server. The datastream may be a video datastream and may be in uncompressed or barely compressed format, for example AVI or MPEG1. The datastream is passed to a compression unit 22 which compresses the data into a very low bit-rate stream using compression parameters. A typical format that may be used for the very low bit-rate stream is MPEG 4 which uses a system based on reference frames and delta frames to compress the image part of the video data. H.263 is another typical format that may be used. A video frame is selected as a reference frame and encoded in its entirety. Then a series of subsequent frames are selected, but for each of these frames, the only data that is encoded is that which has changed with respect to the previous frame. The encoded changes form what is known as the delta frame.

Whichever of the available compression formats are used, compression is governed by a series of parameters 24. This allows the compression level to be raised

Aylo Exhibit 1018
Page 1950 of 2308

or lowered depending on the data rate required and the quality of picture desired. A greater level of compression gives a lower quality of data at a lower data rate and a lower level of compression gives a higher quality of data at a higher data rate.

In the embodiment of Fig. 2 a channel capacity measurement 26 is received from a channel 28 over which the transmission occurs. The measurement may be received from a receiving client who simply reports the received data rate. If the received data rate is reported to be lower than the rate at which data is being sent then it is assumed that data is being produced too fast for the channel and the parameters 24 are altered to increase the compression level until the data output rate is the same as the channel rate. Conversely, if the data rate rises at the receiving end, the parameters 24 may be altered to reduce the compression level and increase the output data rate.

Reference is now made to Fig. 3, which is a generalized block diagram showing a second embodiment of the present invention in which a series of compressed datastreams are continually created and are selected for use in individual channels according to measured parameters of the channel. Parts that are the same as those mentioned in previous embodiments are given the same reference numerals and are not referred to again except as necessary for an understanding of the present embodiment.

In the embodiment of Fig. 3, an input datastream is received from a source 20 as before. The stream is passed to a channel manager 30, which compresses the input datastream. However, instead of compressing it once only, it compresses it into a plurality of output data streams, each one using different compression parameters so as to produce a plurality of streams each having a different data rate. The various streams are then passed to one or more downlink managers 32.1...32.n, each of which is able to manage a limited number of channels 28 for sending output data to a limited number of targets 34. Each target 34 is associated with a client 36, typically client software located at the target 34, which is able to send back, to the downlink manager 32, feedback data such as the received channel rate.

The downlink manager 32 simply obtains the received channel rate from the client 36 and mixes the various streams from the channel manager to produce an average stream rate which matches the channel rate as closely as possible. For example if the received channel rate is 39.5 kbits/sec and the downlink manager has available a 35 kbits/sec stream and a 45 kbits/sec stream then the downlink manager

Aylo Exhibit 1018
Page 1951 of 2308

will attempt to mix the two streams in suitable proportions to match the 39.5 kbits/sec channel. Thus the target 36 may receive a continuous stream of data, the quality of the channel influencing only the quality of the data and not the continuity.

It will be recalled that the compression systems in use include reference and delta frame style compression. Such compression systems, when used to produce streams at different data rates, do not produce delta frames which are interchangeable between the different data rates. Thus in order to create streams that are interchangeable, it is preferable to arrange all of the streams as a series of self-contained packets of given time duration. Each of the packets is self-contained in that it begins with a reference frame and hence does not depend on previously transmitted data. Thus the target is able to use consecutive packets from different streams without any data discontinuity.

The self-contained units are preferably of relatively short duration. The shorter the duration the better the system is able to follow dynamically changes in the bandwidth of the connection. On the other hand, since each unit must start with a reference or key frame, and such frames are very costly in bandwidth since they must be transmitted in full, the use of short self-contained units itself increases the required bandwidth. A compromise unit length of less than ten seconds, preferably 3, 5 or 7 seconds, is preferred.

In an alternative embodiment the length of the self-contained unit is itself allowed to vary dynamically in accordance with circumstances. When the downlink manager 32 detects a significant change in the channel bandwidth it issues a request for a new data rate from the channel manager 30. Alternatively, a new target may be requesting the data feed. The channel manager 30 initiates a time delay when a first request is received, the delay being less than ten seconds, as discussed above, and at the end of the delay starts new self-contained units at all data rates. Thus dynamic following of changes in bandwidth is provided whilst transmitting the minimum possible number of key or reference frames. If no request is received, then the system continues to use delta frames as long as it is practical.

In a variation of the above, instead of waiting to the end of the time delay, the channel manager first of all looks ahead at the approaching data stream and checks whether there is a reference frame approaching during the time delay period in any case. If so, no further action is needed. If there is no reference frame then the channel manager looks ahead to find the biggest delta frame, representing the biggest

Aylo Exhibit 1018
Page 1952 of 2308

instantaneous change in the picture, over the delay time period. A reference frame is then inserted in place of the identified largest delta frame. Thus, the insertion of a full reference frame is optimized to the data.

Reference is now made to Fig. 4 which is a generalized block diagram showing the channel manager 30 and a part of a downlink manager 32 of Fig. 3 in greater detail. Parts that are the same as those mentioned in previous embodiments are given the same reference numerals and are not referred to again except as necessary for an understanding of the present embodiment.

As shown in Fig. 4, the channel manager 30 comprises a series of compression units 40.1, 40.2, 40.3, etc. Each of these compression units is preferably operative to compress an input signal received from source 20, using a different compression parameter or set of compression parameters to produce an output having a different data rate. Each compression unit is followed by a buffer 41. The buffer 41 may be used to arranged the compressed data into packets of fixed time duration.

The downlink manager 32 comprises a series of selector switches 42, one for each channel being managed. For the sake of simplicity, only one such switch 42 is shown in the figure. The switch 42 receives each one of the data streams produced by the source 20 and selects one of them at any given time to be sent to the channel 28. Selection is carried out based on a channel capacity measurement 44, which is received from the client 36.

Reference is now made to Fig. 5, which is a generalized block diagram showing a third embodiment of the present invention in which the user is able to set a quality level of the data that is being sent. Parts that are the same as those mentioned in previous embodiments are given the same reference numerals and are not referred to again except as necessary for an understanding of the present embodiment. As described above, compression parameters are altered in order to change the data rate. Generally speaking, there are numerous parameters that can be set in order to alter the data rate, and there are numerous combinations of these parameters that can give the same data rate. For example, parameters that may be altered include:

the extent of image compression within an individual frame,

the number of frames per second to be transmitted,

the proportion of reference frames to delta frames, and

a balance between compression levels of different parts of the data – for example video data includes both sound and image. It is possible to increase a level

Aylo Exhibit 1018
Page 1953 of 2308

of compression to the image and to decrease a level of compression to the sound or vice versa.

In Fig. 5, part of a channel manager 30 is shown as well as a switch 42 of a downlink manager. In the channel manager of the embodiment of Fig. 5, a series of data compressors 50 are provided for each required data rate. The data compressors are labeled high, medium and low, although in fact they are different compromises between compression parameters. For example "high" may indicate high image quality but low frame rate, whereas "low" may indicate low image quality but high frame rate. The selector switch 48 is the same as in the previous embodiment except that it now chooses channels based on two different items of information. One item of information is the received channel rate 44, which selects the data rate as before. The other item of information is a data quality request which is received from the target. This selects, within the given data rate, between the different quality levels.

Reference is now made to Fig. 6, which is a generalized schematic diagram of a quality level selection tool 60 for use by a user in selecting a quality level in the embodiment of Fig. 5. The selection tool 60 comprises a simple longitudinal bar 62 which may be marked with labels and numbers, and a slider 64 for moving along the bar 62. The bar shown in Fig, 6 indicates that the user may select different compromises between sound quality and image quality.

Reference is now made to Fig. 7, which is a generalized schematic diagram of another quality level selection tool 70 for use by a user in selecting a quality level in the embodiment of Fig. 5. The selection tool 70 is likewise in the form of a bar and slider but this time the choice is between high quality images at a low frame rate (marked "still") and low quality images at a high frame rate (marked "motion"). It will be appreciated that a low bandwidth "high" may yield a lower image quality than a high bandwidth "low" and vice versa.

Embodiments of the present invention thus provide a system and method in which the measured received data level of a channel is used to select a transmitted data rate so that full advantage may be taken of fluctuating channel capacity levels to send the highest possible quality multimedia data.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention

Aylo Exhibit 1018
Page 1954 of 2308

which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined by the appended claims and includes both combinations and subcombinations of the various features described hereinabove as well as variations and modifications thereof which would occur to persons skilled in the art upon reading the foregoing description.

Aylo Exhibit 1018
Page 1955 of 2308

Claims

1. A device for streaming of media file data to be sent over at least one channel having a dynamically variable bandwidth, comprising

an input for receiving media file data at a first rate,

a multiple stream creator for creating a plurality of streams of said media file data, each stream compressed to predetermined other rates, each stream comprising a series of autonomous compressed data units which are mutually compatible, and

a stream selector for dynamically matching a bandwidth of the channel to rates of said streams to send said data file as a combination of said autonomous units from said streams via said channel,

wherein said autonomous units have durations which are one of a group comprising fixed predetermined durations and durations determined by changes in said channel bandwidth.

2. A device according to claim 1, wherein said stream selector is operable to select a combination of said streams in succession said combination being chosen to produce an average stream rate over a period of time which matches said channel bandwidth.

3. A device according to claim 1, wherein said bandwidth of said channel is liable to vary dynamically and wherein said stream selector is operable to repeat said matching throughout said streaming operation.

4. A device according to claim 2, wherein said bandwidth of said channel is liable to vary dynamically and wherein said stream selector is operable to repeat said matching throughout said streaming operation.

5. A device according to claim 1, wherein said autonomous units match each other across said plurality of streams, thereby to enable switching between said streams.

6. A device according to claim 5, wherein said data file is arranged in said streams in a compressed media format utilizing reference frames and

Aylo Exhibit 1018
Page 1956 of 2308

delta frames and wherein each autonomous unit is self-contained in respect of reference frames.

7. A device according to claim 1, wherein said channel is a cellular telephone channel.

8. A device according to claim 1, wherein said channel is an Internet channel.

9. A device according to claim 1, said predetermined other rate being different for each stream.

10. A device according to claim 1, said predetermined other rate being identical for a set of said streams, each stream in said set being assigned a different balance between quality and speed of said media.

11. A device according to claim 1, said media file comprising a plurality of media types, said predetermined other rate being identical for a set of said streams, each stream in said set being assigned a different balance between quality of said different media types.

12. A device according to claim 10, said streams in said set being selectable by a signal received from said channel.

13. A device according to claim 11, said streams in said set being selectable by a signal received from said channel.

14. A device according to claim 1, comprising a detector for detecting a change in the bandwidth of said at least one channel and operable to request a corresponding change in the data rate, and wherein the device is operable, upon receipt of said request, to set a delay, and to issue a reference frame within said delay.

Aylo Exhibit 1018
Page 1957 of 2308

15. A device according to claim 14, operable to issue said reference frame at the end of said delay.

16. A device according to claim 14, operable to issue said reference frame at the occurrence of a largest detected interframe change within the time delay.

17. A multiple stream creator for receiving a data stream at a first data rate and converting said data stream into a plurality of data streams at a plurality of data rates, said creator comprising an input for receiving said data stream at said first data rate, and a plurality of outputs, each operable to receive said data stream at said first rate and to output said data stream at a predetermined rate individually set for each output as a series of autonomous units having a duration, and wherein said duration is one of a group comprising a fixed predetermined duration and a duration dynamically changeable in accordance with bandwidth conditions in a transmission channel.

18. A multiple stream creator according to claim 17, wherein said outputs each comprise data buffers.

19. A multiple stream creator according to claim 17, comprising a data stream processor for processing said data stream received at said first rate into said plurality of autonomous units, said autonomous units being synchronized blocks, thereby allowing said output streams to be mixed in said transmission channel to produce continuously usable output.

20. A multiple stream creator according to claim 17, wherein said received data stream is in an uncompressed data format and wherein said data is compressed into a compressed data format independently for each one of said output data streams.

21. A method of sending a datastream over a variable bandwidth channel comprising modifying a quantity of data to be injected into the channel to accord with a measured capacity of the channel, said method comprising the steps of

Aylo Exhibit 1018
Page 1958 of 2308

compressing the datastream, using an algorithm having compression parameters, to produce a plurality of compressed datastreams each comprising reference frames and delta frames, each compressed datastream produced by using the algorithm together with different parameters, and

synchronizing reference frames between the compressed datastreams so that combinations of said compressed datastreams can be sent via said channel.

22. A method according to claim 21, wherein said step of synchronizing reference frames is carried out at intervals determined by a dynamic status of said channel bandwidth.

23. A method according to claim 21, wherein said step of synchronizing reference frames is carried out at predetermined intervals.

24. A method according to claim 21, wherein said combination of said compressed datastreams is selected substantially dynamically to match the measured capacity of the channel.

25. A method according to claim 22, wherein a change in bandwidth of a channel causes a request for commencement of a new autonomous unit, and wherein said request causes initiation of measurement of a time delay, and wherein a new autonomous unit is commenced within said time delay.

26. A method according to claim 25, wherein said new autonomous unit is commenced at the end of said time delay.

27. A method according to claim 25, comprising the step of determining the largest delta frame during the time delay and sending a reference frame, to initiate said new autonomous unit, in place of said determined largest delta frame.

28. A tool for controlling receipt of media file data over a network, wherein said media file data is available at a plurality of quality levels, the

Aylo Exhibit 1018
Page 1959 of 2308

tool comprising a quality selector to enable a user to select between said different quality levels for receipt.

29. A tool according to claim 28, wherein said media file comprises a plurality of data groups and wherein each quality level is a different balance between quality levels of the plurality of data groups.

30. A tool according to claim 28, wherein said quality selector comprises a slider operable to be moved along a scale, the position of the slider on the scale indicating a quality level.

31. A device for streaming of media file data to be sent over at least one channel having a dynamically variable bandwidth, substantially as hereinbefore described with reference to the accompanying drawings.

32. A multiple stream creator for receiving a data stream at a first data rate and converting said data stream into a plurality of data streams at a plurality of data rates, substantially as hereinbefore described with reference to the accompanying drawings.

33. A method of sending a datastream over a variable bandwidth channel comprising modifying a quantity of data to be injected into the channel to accord with a measured capacity of the channel, substantially as hereinbefore described with reference to the accompanying drawings.

34. A tool for controlling receipt of media file data over a network, substantially as hereinbefore described with reference to the accompanying drawings.

Aylo Exhibit 1018
Page 1960 of 2308

PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS
The Patent Office
INVESTOR IN PEOPLE

Application No: GB 0023062.3 **Examiner:** John Cullen
Claims searched: 1-27, 31-33 **Date of search:** 6 November 2001

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.S): H4P (PDCFM, PDCFP, PDCFX, PF)

Int Cl (Ed.7): H03M 7/30; H04L 1/00; H04N 7/26, 7/32

Other: Online: WPI, EPODOC, JAPIO

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
A, E	EP 1049074 A1	(LUCENT) See Abstract, Figs. 2 and 4, and paragraphs 14-18 and 28-35.	---
A, E	WO 00/72601 A2	(BIGBAND NETWORKS) See Figs. 1-4, 6, 7 and 9, line 26 of p15 to line 24 of p17 and line 10 of p21 to line 2 of p22.	---
A	WO 99/00984 A1	(CITRIX) See Fig. 15 and line 13 of p30 to line 14 of p33.	---
A	WO 98/09387 A1	(ERICSSON) See Figs. 3A, 6 and 7, and line 35 of p14 to line 26 of p15.	---
A	US 5847760	(OPTIBASE) See whole document.	---

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

IDS CERTIFICATION STATEMENTS

I. THIS IDS IS BEING FILED UNDER 37 C.F.R. 1.97(b): (check one box)

☐ Within three months of the filing date of a national application other than a continued prosecution application under 1.53(d).

☐ Within three months of the date of entry of the national stage as set forth in 1.491 in an international application.

☐ Before the mailing date of a first Office Action on the merits. In the event that a first Office Action on the merits has been mailed, please consider this IDS under 37 C.F.R. 1.97(c) and see the statement under 1.97(e) **or** charge deposit account 50-2091 the fee specified by 37 C.F.R. 1.17(p) if such statement is not selected in Section IV.

☒ Before the mailing of a first Office Action after the filing of a request for continued examination under 1.114.

II. THIS IDS IS BEING FILED UNDER 37 C.F.R. 1.97(c):

☐ After the period specified in 37 C.F.R. 1.97(b) and before the mailing date of any of a final action under 1.113, a notice of allowance under 1.311 or an action that otherwise closes prosecution in the application, and accompanied by:

- ☐ a statement specified in 1.97(e) as subsequently provided in this IDS; or
- ☐ the fee set forth in 1.17(p).

III. THIS IDS IS BEING FILED UNDER 37 C.F.R. 1.97(d):

☐ After the period specified in 37 C.F.R. 1.97(c) and on or after the mailing date of any of a final action under 1.113, a notice of allowance under 1.311 or an action that otherwise closes prosecution in the application, but before or simultaneous with payment of the issue fee, and accompanied by a statement specified in 1.97(e) as subsequently provided in this IDS **and** the fee set forth in 1.17(p).

IV. Statement under 37 C.F.R. 1.97(e) (check only one box)

The undersigned hereby states that:

☐ Each item of information contained in the IDS was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of this IDS.

☐ No item of information contained in the IDS was cited in a communication from a foreign patent office in a counterpart foreign application, and, to knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the IDS was known to any individual designated in 37 C.F.R. 1.56(c) more than three months prior to the filing of this IDS.

V. Statement under 37 C.F.R. §1.704(d):

☐ The undersigned hereby states that each item of information contained in the information disclosure statement: (i) was first cited in any communication from a patent office in a counterpart foreign or international application or from the Office, and this communication was not received by any individual designated in § 1.56(c) more than thirty days prior to the filing of the information disclosure statement; or (ii) is a communication that was issued by a patent office in a counterpart foreign or international application or by the Office, and this communication was not received by any individual designated in § 1.56(c) more than thirty days prior to the filing of the information disclosure statement..

VI. CONCLUSION

It is respectfully submitted that the claims presently on file patently distinguish the present invention from each of the references cited in this IDS. The references cited in this IDS are being cited only in the interest of candor and without any admission that they constitute statutory prior art or contain matter which anticipates the invention or which would render the same obvious, either singly or in a combination, to a person of ordinary skill in the art. Furthermore the filing of this IDS should not be construed to be an admission that the information cited in this IDS is, or is considered to be, material to patentability as defined in 1.56(b). In addition, the filing of this IDS should not be construed as a representation that a search has been made.

If the Examiner has any questions concerning this IDS, the Examiner is requested to contact the undersigned at the telephone number associated with Customer No. 70560. If it is determined that this IDS has been filed under the wrong rule, the PTO is requested to consider this IDS under the proper rule, with a petition if necessary, and charge the appropriate fee to Deposit Account No. 50-2091.

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
BRIAN C. KUNZLER
8 EAST BROADWAY, SUITE 600
SALT LAKE CITY, UT 84111

RECEIVED
JAN 09 2009

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference 2082.2.4PCT	Date of mailing *(day month year)* 12 DEC 2008 **FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 08/61035 8532P003PCT	International filing date *(day month year)* 21 April 2008 (21.04.2008)
Applicant MOVE NETWORKS, INC.	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

   Filing of amendments and statement under Article 19:
   The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

   When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

   Where? Directly to the International Bureau of WIPO, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 740 14 35

   For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

   ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

   ☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

   Shortly after the expiration of **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

   The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established. These comments would also be made available to the public but not before the expiration of 30 months from the priority date.

   Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

   In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months.

   See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (January 2004) *(See notes on accompanying sheet)*

Date In to SV FF: 1-12-09
Docketed by: MFA
Reviewed by:
Sent Out by SV FF on:

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 2082.2.4PCT	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 08/61035	International filing date *(day/month/year)* 21 April 2008 (21.04.2008)	(Earliest) Priority Date *(day/month/year)* 19 April 2007 (19.04.2007)
Applicant MOVE NETWORKS, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 9 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**.

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 4

☐ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☒ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/61035

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 15/16 (2008.04)
USPC - 709/227
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)
USPC: 709/227

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 709/201, 203, 223, 227, 228, 238; 715/736, 744, 748, 968

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWest(PGPB,USPT,USOC,EPAB,JPAB); Google
Search Terms Used: content acquisition, resilient, networks, tcp, directive handling module, content acquisition module, directive server, network connection, remote host etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/0188051 A1 (SNEH) 25 August 2005 (25.08.2005), para [0005], para [0055]-[0097], para [0121]-[0136], para [0188]	1-36
Y	US 2003/0014684 A1 (KASHYAP) 16 January 2003 (16.01.2003), para [0101]	1-36
A	US 2006/0206246 A1 (WALKER) 14 September 2006 (14.09.2006)	1-36

☐ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
25 November 2008 (25.11.2008)	12 DEC 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:
BRIAN C. KUNZLER
8 EAST BROADWAY, SUITE 600
SALT LAKE CITY, UT 84111

PCT

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	12 DEC 2008
Applicant's or agent's file reference 2082.2.4PCT	FOR FURTHER ACTION See paragraph 2 below

International application No.	International filing date *(day month year)*	Priority date *(day month year)*
PCT/US 08/61035	21 April 2008 (21.04.2008)	19 April 2007 (19.04.2007)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G06F 15/16 (2008.04)
USPC - 709/227

Applicant MOVE NETWORKS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

3. For further details, see notes to Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer:
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	26 November 2008 (26.11.2008)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (April 2007)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 08/61035

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of:

   a. type of material

      ☐ a sequence listing

      ☐ table(s) related to the sequence listing

   b. format of material

      ☐ on paper

      ☐ in electronic form

   c. time of filing/furnishing

      ☐ contained in the international application as filed

      ☐ filed together with the international application in electronic form

      ☐ furnished subsequently to this Authority for the purposes of search

4. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table(s) relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 08/61035

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-36	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-36	NO
Industrial applicability (IA)	Claims	1-36	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-36 lack an inventive step under PCT Article 33(3) as being obvious over US 2005/0188051 A1 (Sneh) in view of US 2003/0014684 A1 (Kashyap).

Regarding claim 1, Sneh discloses an apparatus for resilient content acquisition (content, para [0005]), the apparatus comprising: a directive handling module configured to implement a directive (directives, para [0055]); wherein the directive defines requirements for establishing, locating, and maintaining a network connection (maintain download state, para [0097]); and a content acquisition module configured to acquire content over a network according to the directive (content acquisition, para [0081]). Sneh does not expressly disclose maintaining a network connection with a remote host. Kashyap teaches maintaining a network connection with a remote host (remote host, para [0101]). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to modify the method of Sneh to include maintaining a network connection with a remote host as taught by Kashyap because for the advantage of expanding the network beyond immediate systems for content acquisition.

Regarding claim 2, in the combination of Sneh and Kashyap, Sneh further discloses the content acquisition module is further configured to transmit a directive server query to a directive server (CA Server, para [0136]).

Regarding claim 3, in the combination of Sneh and Kashyap, Sneh further discloses the directive handling module is further configured to receive the directive from the directive server (directives, para [0055]).

Regarding claim 4, in the combination of Sneh and Kashyap, Sneh further discloses the content acquisition module is further configured to locate the content, authorize the content, retrieve the content, display the content, and report statistics in accordance with the directive (content acquisition, para [0081]).

Regarding claim 5, in the combination of Sneh and Kashyap, Sneh further discloses a directive server module configured to maintain a list of directive servers and to select a directive server for each directive request (CA Server, para [0136]).

Regarding claim 6, in the combination of Sneh and Kashyap, Sneh further discloses a server status module configured to maintain a status indicator for each of a plurality of content delivery network servers (status, para [0188]).

Regarding claim 7, in the combination of Sneh and Kashyap, Sneh further discloses a broadcast module configured to broadcast content availability over a network such that a second apparatus acquires the content from the apparatus over the network (content acquisition, para [0081]).

Regarding claim 8, in the combination of Sneh and Kashyap, Kashyap further discloses a discovery module configured to discover available content on the network such that content is acquired from a second apparatus (remote host, para [0101]).

Regarding claim 9, in the combination of Sneh and Kashyap, Kashyap further discloses the directive is further configured to identify a plurality of backup remote hosts (remote host, para [0101]).

(See Supplemental Box)

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
Box No. V — Reasoned statement
2. Citations & explanations:

Regarding claim 10, Sneh discloses a system for resilient content acquisition (content, para [0005]), the system comprising: a client module in communication with a network (modules, para [0121]-[0127]), the client module comprising: a directive handling module configured to implement a directive (directives, para [0055]); wherein the directive defines requirements for establishing, locating, and maintaining a network connection (maintain download state, para [0097]); a content acquisition module configured to acquire content over the network according to the directive (content acquisition, para [0081]); wherein the content acquisition module is further configured to communicate a directive server query with a directive server (CA Server, para [0136]); and wherein the directive server is configured to parse the content request and generate a directive in response to the content request (CA Server, para [0136]). Sneh does not expressly disclose maintaining a network connection with a remote host. Kashyap teaches maintaining a network connection with a remote host (remote host, para [0101]). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to modify the method of Sneh to include maintaining a network connection with a remote host as taught by Kashyap because for the advantage of expanding the network beyond immediate systems for content acquisition.

Regarding claim 11, in the combination of Sneh and Kashyap, Sneh further discloses the directive handling module is further configured to receive the directive from the directive server (directives, para [0055]).

Regarding claim 12, in the combination of Sneh and Kashyap, Sneh further discloses the content acquisition module is further configured to locate the content, authorize the content, retrieve the content, display the content, and report statistics in accordance with the directive (content acquisition, para [0081]).

Regarding claim 13, in the combination of Sneh and Kashyap, Sneh further discloses a directive server module configured to maintain a list of directive servers and to select a directive server for each directive request (CA Server, para [0136]).

Regarding claim 14, in the combination of Sneh and Kashyap, Sneh further discloses a server status module configured to maintain a status indicator for each of a plurality of content delivery network servers (status, para [0188]).

Regarding claim 15, in the combination of Sneh and Kashyap, Sneh further discloses a broadcast module configured to broadcast content availability over a local network such that a second apparatus acquires the content from the apparatus over the network (content acquisition, para [0081]).

Regarding claim 16, in the combination of Sneh and Kashyap, Kashyap further discloses a discovery module configured to discover available content on a network such that content is acquired from a second apparatus (remote host, para [0101]).

Regarding claims 17, 25 and 33, Sneh discloses a method, computer program product, and apparatus for resilient content acquisition (content, para [0005]), comprising: implementing a directive (directives, para [0055]); defining requirements for locating, establishing, and maintaining a network connection (maintain download state, para [0097]); and acquiring content over a network according to the directive (content acquisition, para [0081]). Sneh does not expressly disclose maintaining a network connection with a remote host. Kashyap teaches maintaining a network connection with a remote host (remote host, para [0101]). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to modify the method of Sneh to include maintaining a network connection with a remote host as taught by Kashyap for the advantage of expanding the network beyond immediate systems for content acquisition.

Regarding claims 18 and 26, in the combination of Sneh and Kashyap, Sneh further discloses transmitting a directive server query to a directive server (CA Server, para [0136]).

Regarding claims 19 and 27, in the combination of Sneh and Kashyap, Sneh further discloses receiving the directive from the directive server (CA Server, para [0136]).

Regarding claims 20 and 28, in the combination of Sneh and Kashyap, Sneh further discloses locating the content, authorizing the content, retrieving the content, displaying the content, and reporting statistics in accordance with the directive (content acquisition, para [0081]).

Regarding claims 21 and 29, in the combination of Sneh and Kashyap, Sneh further discloses maintaining a list of directive servers and selecting a directive server for each directive request (CA Server, para [0136]).

Regarding claims 22 and 30, in the combination of Sneh and Kashyap, Sneh further discloses maintaining a status indicator for each of a plurality of content delivery network servers (status, para [0188]).

Regarding claims 23 and 31, in the combination of Sneh and Kashyap, Kashyap further discloses broadcasting content availability over a local network such that a second apparatus acquires the content from the apparatus over the network (remote host, para [0101]).

Regarding claims 24 and 32, in the combination of Sneh and Kashyap, Kashyap further discloses discovering available content on a network such that content is acquired from a second apparatus (remote host, para [0101]).

(See Next Supplemental Box)

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

Continuation of:
Preceding Supplemental Box
Box No. V --- Reasoned statement
2. Citations & explanations:

Regarding claim 34, Sneh discloses an apparatus for resilient content delivery (content, para [0005]), the system comprising: a parsing module configured to receive a content request from a source and identify content information (modules, para [0121]-[0127]); a response module configured to generate a directive in response to the content information (directives, para [0055]), wherein the directive defines requirements for establishing, locating, and maintaining a network connection (maintain download state, para [0097]); and wherein the response module is further configured to communicate the directive with the source (content acquisition, para [0081]). Sneh does not expressly disclose maintaining a network connection with a remote host. Kashyap teaches maintaining a network connection with a remote host (remote host, para [0101]). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to modify the method of Sneh to include maintaining a network connection with a remote host as taught by Kashyap because for the advantage of expanding the network beyond immediate systems for content acquisition.

Regarding claim 35, in the combination of Sneh and Kashyap, Sneh further discloses the content information is selected from a group consisting of host, domain, content request source, path, and content name (content acquisition, para [0081]).

Regarding claim 36, in the combination of Sneh and Kashyap, Sneh further discloses the response module is further configured to update and communicate the directive with the source (CA Server, para [0136]).

Claims 1-36 have industrial applicability as defined by PCT Article 33(4) because the subject matter can be made or used in industry.

PATENT COOPERATION TREATY

RECEIVED
NOV 10 2008
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN
SUNNYVALE

From the INTERNATIONAL SEARCHING AUTHORITY

To:
LESTER J. VINCENT
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
1279 OAKMEAD PARKWAY
SUNNYVALE, CA 94085-4040

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

RECEIVED
NOV 14 2008
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN
SUNNYVALE

Date of mailing *(day/month/year)* 05 NOV 2008	
Applicant's or agent's file reference 8532P004PCT	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 08/09281	International filing date *(day/month/year)* 01 August 2008 (01.08.2008)
Applicant MOVE NETWORKS, INC.	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

Where? Directly to the International Bureau of WIPO, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 740 14 35

For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

Shortly after the expiration of **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established. These comments would also be made available to the public but not before the expiration of 30 months from the priority date.

Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months.

See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (January 2004) *(See notes on accompanying sheet)*

Entered into PIP by: dlm

DATE IN TO FOREIGN DOCKETING 11/11/08
DOCKETED BY
REVIEWED BY
DATE OUT

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
LESTER J. VINCENT
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
1279 OAKMEAD PARKWAY
SUNNYVALE, CA 94085-4040

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT AND THE WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY, OR THE DECLARATION

(PCT Rule 44.1)

	Date of mailing *(day/month/year)* 05 NOV 2008
Applicant's or agent's file reference 8532P004PCT	**FOR FURTHER ACTION** See paragraphs 1 and 4 below
International application No. PCT/US 08/09281	International filing date *(day/month/year)* 01 August 2008 (01.08.2008)
Applicant MOVE NETWORKS, INC.	

1. ☒ The applicant is hereby notified that the international search report and the written opinion of the International Searching Authority have been established and are transmitted herewith.

Filing of amendments and statement under Article 19:
The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

Where? Directly to the International Bureau of WIPO, 34 chemin des Colombettes 1211 Geneva 20, Switzerland, Facsimile No.: +41 22 740 14 35

For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect and the written opinion of the International Searching Authority are transmitted herewith.

3. ☐ **With regard to the protest** against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. **Reminders**

Shortly after the expiration of **18 months** from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90*bis*.1 and 90*bis*.3, respectively, before the completion of the technical preparations for international publication.

The applicant may submit comments on an informal basis on the written opinion of the International Searching Authority to the International Bureau. The International Bureau will send a copy of such comments to all designated Offices unless an international preliminary examination report has been or is to be established. These comments would also be made available to the public but not before the expiration of 30 months from the priority date.

Within **19 months** from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase **until 30 months** from the priority date (in some Offices even later); otherwise, the applicant must, **within 20 months** from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

In respect of other designated Offices, the time limit of **30 months** (or later) will apply even if no demand is filed within 19 months.

See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/220 (January 2004) *(See notes on accompanying sheet)*

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 8532P004PCT	FOR FURTHER ACTION	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. PCT/US 08/09281	International filing date *(day/month/year)* 01 August 2008 (01.08.2008)	(Earliest) Priority Date *(day/month/year)* 06 August 2007 (06.08.2007)
Applicant MOVE NETWORKS, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 9 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 1

☐ as suggested by the applicant.

☒ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 08/09281

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 15/16 (2008.04)
USPC - 709/231
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)
USPC: 709/231

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 709/206, 207,.219, 225, 230-235

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic Databases Searched: PubWEST (PGPB,USPT,USOC,EPAB,JPAB); GoogleScholar
Search Terms Used: streamlets, adaptive stream, multi-rate, multi-bitrate, rate-shifting, skip, ignore, scrub, streaming video/media, filter, block, simulcoding, timeline, live horizon, ratings, porn, mute etc.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/0262257 A1 (MAJOR et al.) 24 November 2005 (24.11.2005) (shared author) entire document, especially: Abstract, para [0045], [0047], [0073]	1-26
Y	Clement, B., Page One PR, "Move Networks Closes $11.3 Million In First Round VC Funding," Move Networks, Inc. Press Releases, 07 February 2007 (07.02.2007) [retrieved 12 October 2008 (12.10.2008)] Retrieved from the Internet. <URL: http://www.move.tv/press/press20070207.html> entire document, especially: pg 1, para 4-6	1-26
Y	US 2006/0277564 A1 (JARMAN) 07 December 2006 (07.12.2006) entire document, especially: Abstract, para [0010], [0015], [0068]	4-6, 12-13, 17-19, 22-24
A	Move Networks, "The Next Generation Video Publishing System," 11 April 2007 (11.04.2007) [retrieved 13 October 2008 (13.10.2008)]. Retrieved from the Internet. <URL: http://www.movenetworks.com/wp-content/uploads/move-networks-publishing-system.pdf>	1-26

☐ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
14 October 2008 (14.10.2008)	05 NOV 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:
LESTER J. VINCENT
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
1279 OAKMEAD PARKWAY
SUNNYVALE, CA 94085-4040

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	05 NOV 2008
Applicant's or agent's file reference 8532P004PCT	FOR FURTHER ACTION See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US 08/09281	01 August 2008 (01.08.2008)	06 August 2007 (06.08.2007)

International Patent Classification (IPC) or both national classification and IPC
IPC(8) - G06F 15/16 (2008.04)
USPC - 709/231

Applicant MOVE NETWORKS, INC.

1. This opinion contains indications relating to the following items:

- [X] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [ ] Box No. IV Lack of unity of invention
- [X] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

3. For further details, see notes to Form PCT/ISA/220.

Name and mailing address of the ISA/US	Date of completion of this opinion	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	14 October 2008 (14.10.2008)	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/237 (cover sheet) (April 2007)

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 06/09281

Box No. I Basis of this opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of:

   a. type of material

      ☐ a sequence listing

      ☐ table(s) related to the sequence listing

   b. format of material

      ☐ on paper

      ☐ in electronic form

   c. time of filing/furnishing

      ☐ contained in the international application as filed

      ☐ filed together with the international application in electronic form

      ☐ furnished subsequently to this Authority for the purposes of search

4. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table(s) relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE
INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/US 08/09281

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-26	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-26	NO
Industrial applicability (IA)	Claims	1-26	YES
	Claims	None	NO

2. Citations and explanations:

Claims 1-3, 7-11, 14-16, 20-21 and 25-26 lack inventive step under PCT Article 33(3) as being obvious over US 2005/0262257 A1 to Major et al. (hereinafter 'Major') in view of the article entitled "Move Networks Closes $11.3 Million in First Round VC Funding," (Clement).

As to claims 1, 9, 15, 21 and 26, Major teaches a system, method, apparatus and computer program product for multi-bitrate content streaming (adaptive-rate shifting of streaming content, see Abstract), wherein the media content comprises a plurality of streamlets (plurality of streamlets, see Abstract); and at least one data module configured to maintain multi-bitrate streamlet information (staging module is configured to stage the streamlets and arrange the streamlets for playback, see Abstract).
Major does not expressly teach a timeline module configured to maintain a programming lineup of media content available over a network. However, Clement teaches multi-bitrate content streaming (adaptive stream that upshifts and downshifts to match the current constraints, see pg 1, para 4) and a timeline module configured to maintain a programming lineup of media content available over a network (TiVo style navigation, see pg 1, para 6) and a client module configured to acquire content based upon the programming lineup provided by the timeline module (Move Media Player - Viewer Controls, TiVo style, see para 6).
It would have been obvious to one of ordinary skill in the art to combine the timeline module, as taught by Clement, with the Major system for multi-bitrate content streaming in order to enable the user to efficiently select and manage their desired content, thus increasing the ease of use for the end-user and increasing the revenue potential for the content providers.

As to claims 2 and 10, Clement teaches each streamlet comprises a portion of the media content encapsulated as an independent media object (Each streamlet 212 may comprise a portion of the content contained in stream 210, encapsulated as an independent media object, see para [0047]), and wherein each streamlet comprises a predetermined duration of time (in one embodiment, the content contained in each streamlet 212 has a duration of two seconds, see para [0047]).

As to claims 3 and 11 and 16, Major teaches multi-bitrate streamlet information is selected from the group consisting of start time, end time, live video, publisher data, encryption, duration, bitrate values, frame size, channels, codecs, sample rate, and frames parser (content file 200 may also be live or archived content, see para [0045] and Fig. 2a).

As to claims 7 and 25, Clement teaches a content management system configured to dynamically replace portions of the media content (shows can even be pre-empted and replaced by alternative programming, see pg 1, para 5).

As to claim 8, Major teaches an anchor module configured to associate the start of the timeline module with a point in time such that playback of the media content begins at a point corresponding to a live horizon (presentation at the current playback time index, see pg 1, para [0073]).

As to claims 14 and 20, Clement teaches a content management system configured to dynamically replace portions of the media content (shows can even be pre-empted and replaced by alternative programming, see pg 1, para 5).

(See Supplemental Box)

Supplemental Box

In case **the space in any of the preceding boxes is not sufficient.**
Continuation of:
Box No. V --- Reasoned Statement
2. Citations and Explanations:

Claims 4-6, 12-13, 17-19 and 22-24 lack inventive step under PCT Article 33(3) as being obvious over Major in view of Clement as applied above, and further in view of US 2006/0277564 A1 (Jarman).

As to claims 4, 12, 17, 22 and 23, the combination of Major and Clement teach the apparatus of claim 1, the system of claim 9, the method of claim 15 and the computer program product of claim 21, and Major further teaches streamlets providing content (a staging module configured to stage the streamlets and arrange the streamlets for playback on a content player , see para [0015]), but the combination of Major and Clement does not expressly teach a module configured to skip content in response to a rating system level. However, Jarman teaches a programming content distribution system which further comprising a client module configured to skip content in response to a rating system (selectively mutes context specific words as a function of program specific criteria, see Abstract).
It would have been obvious to one of ordinary skill in the art to combine the Jarman content filtering system with the content distribution system taught by Major and Clement because this would enable the user to instruct the more precisely control the desired viewing content; further, Clement suggests content filtering (complete editing and publishing control, see pg 1, para 4) such as that taught by the Jarman content filtering system.

As to claims 5 and 18, Jarman teaches the client module is configured to skip streamlets that do not conform to a selected rating level (allow play of only programs associate with selected program specific information; and blocking display of all programs not including the selected program specific information, see para [0010]).

As to claim 6, Jarman teaches the client module is configured to mute the audio of streamlets that do not conform to a selected rating level (mute words for an audio or audio/video presentation that may be objectionable in one context, see para [0068]).

As to claims 13, 19 and 24, the combination of Major and Clement teach the system of claim 9 and the method of claim 15, but does not expressly teach the client module is configured to mute the audio of streamlets that do not conform to a selected rating level.
However, Jarman teaches multi-bitrate content streaming (adaptive stream that upshifts and downshifts to match the current constraints, see pg 1, para 4) and a timeline module configured to maintain a programming lineup of media content available over a network (TiVo style navigation, see pg 1, para 6) and wherein the client module is configured to mute the audio of streamlets that do not conform to a selected rating level (electively mutes context specific words as a function of program specific criteria, see Abstract. It would have been obvious to combine the Jarman selective content filtering and muting system with the content distribution system taught by Major and Clement because this would enable the user to instruct the more precisely control the desired viewing content; further, Clement suggests content filtering (complete editing and publishing control, see pg 1, para 4).

Claims 1-26 have industrial applicability as defined by PCT Article 33(4) because the subject matter claimed can be made or used in industry.

NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under Article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the *PCT Applicant's Guide*, a publication of WIPO.

In these Notes, "Article," "Rule" and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions, respectively.

INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report and the written opinion of the International Searching Authority, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international publication. Furthermore, it should be emphasized that provisional protection is available in some States only (see *PCT Applicant's Guide*, Volume I/A, Annexes B1 and B2).

The attention of the applicant is drawn to the fact that amendments to the claims under Article 19 are not allowed where the International Searching Authority has declared, under Article 17(2), that no international search report would be established (see *PCT Applicant's Guide*, Volume I/A, paragraph 296).

What parts of the international application may be amended ?

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Preliminary Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

When ? Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

Where not to file the amendments ?

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below.

How ? Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Section 205(b)).

The amendments must be made in the language in which the international application is to be published.

What documents must/may accompany the amendments ?

Letter (Section 205(b)):

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.

Aylo Exhibit 1018
Page 1979 of 2308

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 March 2004 (25.03.2004)

PCT

(10) International Publication Number
WO 2004/025405 A2

(51) International Patent Classification[7]: G06F

(21) International Application Number:
PCT/US2003/028275

(22) International Filing Date:
10 September 2003 (10.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/409,645 10 September 2002 (10.09.2002) US

(71) Applicant *(for all designated States except US)*: **THOMSON LICENSING S.A.** [FR/FR]; 46, Quai A. Le Gallo, F-92648 Boulogne (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BOYCE, Jill, MacDonald** [US/US]; 3 Brandywine Court, Manalapan, NJ 07726 (US). **RAMASWAMY, Kumar** [IN/US]; 71 Sayre Drive, Princeton, NJ 08540 (US).

(74) Agents: **TRIPOLI, Joseph, S.** et al.; Thomson Licensing Inc., Suite #200, Two Independence Way, Princeton, NJ 08540 (US).

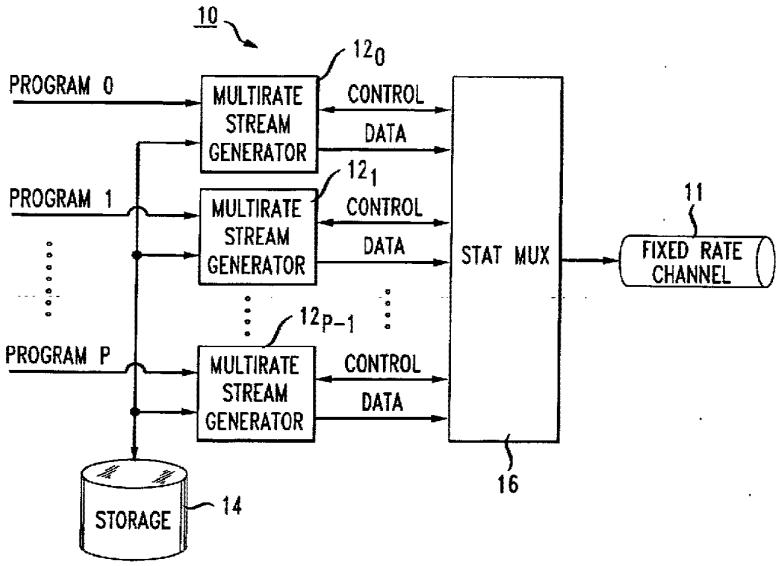
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: VIDEO ON DEMAND SERVER SYSTEM AND METHOD



(57) Abstract: A Video-on-Demand Server architecture (100, 200, 300) transmits a plurality of pre-coded programs having different bit rates across a fixed bandwidth channel (11). For each program, a generator (120-12P-1, 110, 210) generates a plurality of different bit rate representations for each program. Each generator also provides control information at each of a plurality of successive time windows T for each bit rate representation. The control information provides a bit rate and a quality measure during each time window T. The control information enables a statistical multiplexer (16, 160, 260) to select a bit rate representation for each program during each time window T to maximize the quality of the selected representations while not exceeding the total available channel capacity.

WO 2004/025405 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

VIDEO ON DEMAND SERVER SYSTEM AND METHOD

TECHNICAL FIELD

This invention relates to a technique for efficiently providing video on demand while maintaining bandwidth constraints of a fixed channel.

BACKGROUND ART

Present day Video-on-Demand (VoD) systems typically transmit Constant Bit Rate (CBR) video. Transmitting CBR video does not achieve the same efficiency as transmitting variable bit rate (VBR) video for equivalent long-term average bit rates. Transmitting multiple pre-recorded VBR-encoded video programs over a constant bandwidth channel has problems since the sum of the instantaneous bitrates required by each program can sometimes exceed the total available bitrate. On the other hand, broadcast applications employing real-time encoders use statistical multiplexing to exploit the instantaneous bitrate variations between multiple programs being carried on the same carrier or transponder. The encoders operate in real time to enable the system to constrain the total combined bit rate for all programs so as not to exceed the channel capacity over a given time window. Each video program has an associated complexity measure. A central controller dynamically adjusts the bit rate allocated to each video program based upon the relative complexities.

There now exist devices known as "video transraters" that operate to reduce the bit rate of video streams within a common compression standard, such as MPEG 2 for example. Present day transraters often experience difficulties when converting between constant and variable bit rates. Such transraters also can experience difficulties when trying to alter the bit rate for bit streams that have many scene changes, or a large number of I-Pictures.

Scalable video encoding, which permits dividing a video signal into a base layer and one or more enhancement layers, can also address bit rate issues. Several methods of scalable video encoding exist, including spatial, SNR, temporal, data partitioning, fine grain scalability (FGS), frequency scalability. The MPEG-2 and MPEG-4 video compression standards include several scalability methods. Using scalable encoding requires both the transmitter and receiver have the same ability to implement different coding algorithms, thus introducing additional complexity.

Thus, there exists a need for a technique for managing variable bit rate video on demand that obviates the disadvantages of the prior art.

BRIEF SUMMARY OF THE INVENTION

Briefly, in accordance with present principles, there is provided a method for transmitting a plurality of pre-coded programs having different bit rates across a fixed bandwidth channel. For each program, at least two, and preferably, a plurality of different bit rate representations are generated. Control information is provided at each of a plurality of successive time windows for each bit rate representation. The control information provides a bit rate and a quality measure during each time window. The control information enables selection during each time window of a bit rate representation for each program to maximize the quality of the selected representations while not exceeding the total available channel capacity in that time window.

BRIEF SUMMARY OF THE DRAWING

FIGURE 1 depicts a block schematic of first preferred embodiment of a Video-on-Demand server architecture in accordance with the present principles;

FIGURE 2 depicts a block schematic of second preferred embodiment of a Video-on-Demand server architecture in accordance with the present principles; and

FIGURE 3 depicts a block schematic of third preferred embodiment of a Video-on-Demand server architecture in accordance with the present principles.

DETAILED DESCRIPTION

Before proceeding to describe the technique of the present principles, the following terminology will prove helpful:

T is the time interval over which the system optimization is being contemplated

C is the total channel capacity available in time frame T

P is the total number of programs

$p \in (0, P\text{-}1)$, is the index of a particular program

$N[p]$ is the total number of representations of program p

$n[p] \in (0, N[p]-1)$ is the index of a particular representation of program p

$r[p, x]$ is the bit rate of representation x of program p during T

$q[p, x]$ is the quality of representation x of program p during T

FIGURE 1 depicts a first preferred embodiment of a Video-on-Demand server architecture 10 in accordance with the present principles for transmitting pre-encoded CBR and/or VBR audio/video programs over a fixed bandwidth channel 11 having a total channel capacity, C. The server architecture 10 includes a plurality of multirate stream generators 12_0, $12_1 \ldots 12_{P-1}$, where P is an integer greater than zero corresponding to the number of separate input streams. The stream rate generators 12_0-12_{P-1} each pre-code a corresponding one of the programs Program 0, Program 1...Program $P-1$ at a plurality of different bit rates. Stated another way, each multirate stream generator pre-codes the corresponding program to yield at least two or more different bit rate representations. A representation can include a particular stored encoded bit stream, or a combination of various portions of different encoded bit streams. A storage device 14, typically in the form of a single disc drive, a Redundant Array of Inexpensive Discs (RAID), or a plurality of RAIDs, serves to store the pre-coded representations generated by the multirate stream generators 12_0-12_{P-1}.

Each of the multirate stream generators 12_0-12_{P-1} generates control information for each time window T of for the corresponding one of Programs 0 to $P-1$, respectively. The control information for each time window T for each representation of each program includes an indication of the bit rate of that representation and a quality measure, such as the peak signal-to-noise ratio (PSNR) of the representation. A central statistical multiplexer ("stat mux") 16 receives the control information for each representation. At each successive time window T, the stat mux 16 selects a representation for each program to maximize the quality of the transmitted programs while maintaining the total bit rate at or below the total capacity C of the channel 11.

The lowest bit rate representation for each program should not exceed a prescribed value such that the sum of the lowest bit rates for all programs will not exceed the channel capacity for each interval T. In this way, at least one representation of each program can undergo transmission over the channel 11. Requiring the lowest bit rate representation for each program to have a peak bit rate of C/P or less can achieve this constraint. Other methods exist that meet the constraint $\sum_{p=0}^{P-1} r[p,0] \le C$ for all time windows T of all programs. Typically, advance knowledge does not exist as to when playback of particular programs will commence. Therefore,

enforcing capacity restraints becomes easiest by requiring that the bit rate limit of the lowest bit rate representation be the same for all time windows T of a particular program.

The process for the maximization of the overall combined quality can occur in several different ways. In all cases, the constraint $\sum_{p=0}^{P-1} r[p,n[p]] \leq C$ must remain met for all time windows T of all programs. A minimax approach can satisfy the constraint by choosing *n[p]* for each $p \in (0, P\text{-}1)$ to maximize the quality of the minimum quality program (i.e., minimizing the maximum distortion.)

The minimax approach can be implemented by sorting the control information for each representation of each program so that the quality and bit rate index increases monotonically increasing with the index. The change in bit rate for each step (referred to as the "delta (Δ)"bit rate) is then stored. The stat mux 16 begins with the lowest index for each program representation and computes the total capacity S. The stat mux 16 the selects the program representation at the lowest quality and checks if adding its delta bit rate to S exceeds C. If the addition of the representation does not exceed C, the index for that representation is incremented, and process is repeated. Once C is exceeded, the representation with the next to lowest quality is checked to see if adding its delta bit rate to S exceeds C. The process is repeated until no increment in any representation can be made without exceeding the channel capacity C. Alternatively, the total quality of all programs could be optimized by maximizing the sum of the individual program qualities which involves solving the following constrained optimization problem:

$$\max_{n[.]} \sum_{p=0}^{P-1} q[p,n[p]]; \text{ subject to } \sum_{p=0}^{P-1} r[p,n[p]] \leq C$$

It is also possible to optimize the product of the individual program qualities in the following manner:

$$\max_{n[.]} \prod_{p=0}^{P-1} q[p,n[p]]; \text{ subject to } \sum_{p=0}^{P-1} r[p,n[p]] \leq C$$

A weighted average can also be used, in order to provide different classes of service for different viewers.

The multirate stream generators 12_0-12_{P-1} can use several different methods, or a combination of methods to form multiple representations of each program. In all cases, a random access point, such as an intra-coded (I) frame should occur at the beginning of each time window T, for each representation, so that for each time window T a different representation can be

chosen without causing drift. Each representation can use either CBR or VBR coding as long as the capacity of the lowest bit rate representation meets the total channel capacity constraint.

In a first method, a multirate video encoder encodes each program at several different bit rates into several independent bit streams. Each different bit rate bit stream serves as a different representation. FIGURE 2 depicts an architecture 100 for carrying out this method. As seen in FIG. 2, a multirate video encoder 110 serves to encode a corresponding program to yield a plurality of different rate representations. A separate one of transport packetizers 112_0, $112_2 \ldots 112_{P-1}$ each packetizes a corresponding representation. The multirate encoding performed by the multirate encoder 110, and the packetization performed by the packetizers 112_0-112_{P-1} occurs once, not necessarily in real time, prior to storage in a storage device 140 similar in construction to the storage device 14 of FIG. 1.

Selection of the representations stored in storage device 140 for output in real time occurs in response to a signal received from a selector block 150. The selector block 150 identifies the stored representation for output in response to a request from the stat mux 160 for a particular representation. Upon receipt of the signal from the selector block 150, the storage device 140 supplies the selected packetized representations to the stat mux 160 for output.

In a second method, portions of the several different bit streams are combined to yield additional representations of a program. The presence of random access points at the same location in all bit streams of a program (as would naturally occur at scene changes) or the use of fixed group-of-pictures structures, permits the combination of compressed data from different bit rate streams at random access boundaries to form new representations. No need exists to store each representation independently, as long as the ability exists to generate each representation from the data that is stored. Consider the following example where T is one second, the data is coded at 30 fps, and I frames are inserted every 15 frames, yielding three different bit rate bit streams, designated as Bit stream 0, Bit stream 1, and Bit Stream 2, respectively. Each bit stream constitutes a separate representation. An additional representation can be formed which uses Bitstream 0 for the first 15 frames and Bitstream 1 for the second 15 frames, and so on.

Assuming alignment of the random access points with the transport packets, the server architecture 100 of FIG. 2 works well when the transport packetization occurs in advance. FIGURE 3 depicts a VoD server architecture 200 better suited when the random access points do not necessarily align with the transport packets. Referring to FIG. 3, VoD server architecture 200 includes multirate video encoder 210 for generating a plurality of different bit rate

representations for storage in a storage device 240 similar to the storage devices 14 and 140. The representations stored in the storage device 140 exist as un-packetized bit streams.

Selection of the representations stored in storage device 240 for output in real time occurs in response to a signal received from a selector block 250. The selector block 250, similar in nature to the selector block 150 of FIG. 2, identifies the stored representation for output in response to a request from a static mixer/multiplexer 260 for a particular representation. Upon receipt of the signal from the selector block 250, the storage device 240 supplies the selected packetized representations to the static mixer/multiplexer. A packetizer 212 packetizes the output stream of the static mixer/multiplexer 260 for output on the channel (not shown).

Representations also can be formed by switching between stored bit streams for non-reference pictures, such as for video codecs (not shown) that include pictures that are not used as reference pictures for predicting other pictures, such as MPEG-2 B frames or JVT non-stored pictures. Each non-reference picture can be selected from a different stored bitstream without affecting the quality of subsequently coded pictures, as the non-stored pictures are not used for prediction. Complete elimination of a non-reference picture in a representation can also occur. Each non-reference picture could be switched individually, or groups of non-reference pictures could be switched together, allowing for many possible total bit rate representations. The multirate stream signal generators 12_0-12_{P-1} of FIG. 1 can choose to limit the number of possible representations to send to the stat mux 16 of FIG. 1 to choose those with significant differences in quality and bit rate. It is not necessary for each representation to be independently stored in storage, but a table listing the location and length of each coded frame can be stored to simplify generation of the representation when needed. It is also possible to store multiple bit rate bit streams only for the non-stored pictures and only store a single version of the reference (I and P) pictures, which can significantly reduce storage requirements.

With the VoD server architectures 100 and 200 of FIGS. 2 and 3, respectively, the stat mux 160 and the static mixer/multiplexer 260 generate a bit stream corresponding to the selected representation. If individual pictures are not transport packet aligned, the architecture 200 of FIG. accomplishes transport packetization after the generation of the bit stream for a particular representation. Otherwise, the VoD server architecture 100 of FIG. 2 FIG. 2 will work well.

In accordance with another aspect of the present principles, scalable video coding can occur by the use of a scalable video encoder (not shown) to form a base layer and one or more enhancement layers. Any type of scalability that which employs the base layer for motion compensated reference picture prediction in order to avoid drift can be used, such as frequency

scalability, FGS, SNR scalability or temporal scalability. The lowest bit rate representation corresponds to the base layer. The peak bit rate of the base layer must be such that when the bit rates of all programs' base layers are summed, the channel capacity C is not exceeded for each time window T. Other higher bit rate representations will correspond to the base layer plus some portion of the enhancement layer. This approach requires that the video decoder support the type of scalability used. The VoD server architecture 100 of FIG. 2 is used upon inclusion of an entire enhancement layer in a particular representation. The VoD server architecture 200 of FIG. 3 is used if portions of an enhancement layer are used in a representation.

For systems which use encryption for contents right management, the VoD server architecture 100 of FIG. 2 does not require any decryption or encryption in real time, as encryption is added to the transport packetization process done in non-real time and entire encrypted transport packets are stored and transmitted. The VoD server architecture 200 of FIG. 3 requires that encryption be performed along with transport packetization in real-time.

VoD programs generally include both audio and video. As audio bit rates are generally significantly lower than video bit rates and are also generally at a fixed bit rate, there is less advantage to be gained by selecting from different pre-encoded audio bit streams by the stat mux 16 and 160 of FIGS 1 and 2, and the static mixer/multiplexer 260 of FIG. 3. A user likely will find the switching of audio quality during the viewing of a single program disturbing. The total channel capacity C can be considered to be totally consumed by the video channel; assuming that a single bit rate audio stream is also sent for each program. Alternatively, the above-described methods can also be applied to audio, where multiple representations of audio programs with different bit rates and qualities can be generated.

A video player (not shown) that works with the VoD server architectures of the present principles will contain a video decoder (not shown) and some storage for buffering. For the particular program being received, the data can arrive at a non-uniform data rate, but with the requirement that the data corresponding to time unit T will arrive any time within the T time window. The video player must have the capability of buffering and delaying for T. An upper limit on the maximum bit rate that can be transmitted for a given representation of a program for a time window T can be pre-determined to limit the decoder complexity and decoder buffer size. The decoding buffer is generally a requirement of a video decoder, e.g. MPEG-2 levels sets buffer size particular requirements, and the combination of T and the bitrate may be chosen to meet it.

Fast forward and fast reverse tracks for each program can also be stored at the VoD server. They can be stored at a single bit rate, or at several different bit rates, and allowed to adapt as is done for the normal play program.

Determination of the time window T length is based on several factors. T should be as large as possible to get the most statistical multiplexing gain, but small enough to constrain the delay at the video player for start-up and switching between normal play and trick play streams. T should be small enough to meet decoder requirements.

The foregoing describes a system and method for efficiently providing video on demand while maintaining bandwidth constraints of a fixed channel.

CLAIMS

1. A method for transmitting a plurality of pre-coded programs having different bit rates across a fixed bandwidth channel, comprising the steps of:

generating at least two different bit rate representations of each program;

providing control information at each of a plurality of successive time windows T for each representation of each program, the control information for each successive window indicating a bit rate and quality measure for a representation of a corresponding program; and

during each time window T, selecting a representation for each program such to maximize the quality of the selected representations while not exceeding a total available capacity for the channel.

2. The method according to claim 1 wherein the step of generating at least two different bit rates representation further comprises the step of generating for each program a lowest bit rate representation having a peak bit rate not greater than C/P where C is the total channel capacity in time T and P is the total number of programs.

3. The method according to claim 1 wherein the step of providing the control information further comprises the step of establishing the peak signal-to-noise ratio (PSNR) as the quality measure embodied in the control information.

4. The method according to claim 1 wherein the selecting step further comprises the step of selecting a representation for each program which meets the constraint $\sum_{p=0}^{P-1} r[p, n[p]] \leq C$ for all time windows wherein:

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p] - 1)$ is the index of a particular representation of program p; and

$r[p, x]$ is the bit rate of representation x of program p during T

5. The method according to claim 4 further comprising the step of choosing each program's representation n[p] $\in$ (0, N[P]-1) to maximize the quality of the program p that had the minimum quality.

6. The method according to claim 5 further comprising the steps of:

(a) sorting the quality information for with the bit rate and quality measure monotonically increasing with an index value;

(b) storing each bit rate increment (delta) and quality value for each index value;

(c) beginning with a lowest index value, computing total capacity S for program representations selected thus far for such index value;

(d) selecting a program representation at a lowest quality measure;

(e) checking whether the bit rate increment of the selected program at the lowest quality, when added to the representations selected thus far, exceeds total channel capacity, and if not

(f) incrementing the index value; and

(g) repeating steps (c)-(f).

7. The method according to claim 1 wherein the selecting step further comprises the step of selecting the representation for each program such to maximize a sum of individual program qualities by solving $\max_{n[.]} \sum_{p=0}^{P-1} q[p, n[p]]$; subject to $\sum_{p=0}^{P-1} r[p, n[p]] \leq C$

wherein ,

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p] - 1)$ is the index of a particular representation of program p;

$r[p, x]$ is the bit rate of representation x of program p during T; and

$q[p, x]$ is the quality of representation x of program p during T.

8. The method according to claim 1 wherein the selecting step further comprises the step of selecting the representation for each program such to maximize a product of individual program qualities by solving

$$\max_{n[.]} \prod_{p=0}^{P-1} q[p,n[p]]; \text{ subject to } \sum_{p=0}^{P-1} r[p,n[p]] \leq C$$

where,

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p]-1)$ is the index of a particular representation of program p;

$r[p, x]$ is the bit rate of representation x of program p during T; and

$q[p, x]$ is the quality of representation x of program p during T.

9. The method according to claim 8 further comprising the step of applying a weighted average to provide different classes of service for different viewers.

10. A system for transmitting a plurality of pre-coded programs having different bit rates across a fixed bandwidth channel, comprising the steps of:

means for generating at least two different bit rate representations of each program;

means providing control information at each of a plurality of successive time windows T for each representation of each program, the control information for each successive window indicating a bit rate and quality measure for a representation of a corresponding program; and

means for selecting during each time window T a representation for each program such to maximize the quality of the selected representations while not exceeding a total available capacity for the channel.

11. The system according to claim 10 wherein the generating means and control information providing means collectively comprise:

a plurality of multirate stream generators, each associated with a corresponding one of the plurality of pre-coded programs.

12. The system according to claim 10 wherein the generating means and control information providing means collectively comprise:

a multirate video encoder for encoding at least two bit rate representations of each pre-coded program.

13. The system according to claim 10 wherein the generating means and control information providing means collectively comprise:

a multirate video encoder for encoding at least two bit rate representations of each pre-coded program; and

a plurality of transport packetizers, each serving to packetize the bit rate presentations for each pre-coded program.

14. The system according to claim 10 wherein the selecting means includes a static multiplexer.

15. The system according to claim 12 wherein the selecting means comprises:

a static multiplexer; and

a transport packetizer for packetizing the selecting representation.

16. The system according to claim 10 wherein the selecting means generates for each program a lowest bit rate representation having a peak bit rate not greater than C/P where C is the total channel capacity in time T and P is the total number of programs.

17. The system according to claim 10 wherein control information providing means establishes quality measure in accordance with a peak signal-to-noise ratio (PSNR).

18. The system according to claim 10 wherein the selecting means selects a representation for each program which meets the constraint $\sum_{p=0}^{P-1} r[p, n[p]] \leq C$ for all time windows where:

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p] - 1)$ is the index of a particular representation of program p; and

$r[p, x]$ is the bit rate of representation x of program p during T.

19. The system according to claim 18 wherein the selecting means chooses each program's representation n[p] $\in$ (0, N[P]-1) to maximize the quality of the program p that had the minimum quality.

20. The system according to claim 10 wherein the selecting means selects the representation for each program such to maximize a sum of individual program qualities by solving:

$$\max_{n[.]} \sum_{p=0}^{P-1} q[p,n[p]]; \quad \text{subject to} \quad \sum_{p=0}^{P-1} r[p,n[p]] \le C$$

where,

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p]-1)$ is the index of a particular representation of program p;

$r[p, x]$ is the bit rate of representation x of program p during T; and

$q[p, x]$ is the quality of representation x of program p during T.

21. The system according to claim 10 wherein the selecting means selects the representation for each program such to maximize a product of individual program qualities by solving:

$$\max_{n[.]} \prod_{p=0}^{P-1} q[p,n[p]]; \quad \text{subject to} \quad \sum_{p=0}^{P-1} r[p,n[p]] \le C$$

where,

C is the total channel capacity available in time frame T;

P is the total number of programs;

$p \in (0, P\text{-}1)$, is the index of a particular program;

$N[p]$ is the total number of representations of program p;

$n[p] \in (0, N[p]-1)$ is the index of a particular representation of program p;

$r[p, x]$ is the bit rate of representation x of program p during T; and

$q[p, x]$ is the quality of representation x of program p during T.

22. The system according to claim 10 wherein a weighted average is applied to provide different classes of service for different viewers.

ABSTRACT OF THE DISCLOSURE

A Video-on-Demand Server architecture (100, 200, 300) transmits a plurality of pre-coded programs having different bit rates across a fixed bandwidth channel (11). For each program, a generator (12_0-12_{P-1}, 110, 210) generates a plurality of different bit rate representations for each program. Each generator also provides control information at each of a plurality of successive time windows T for each bit rate representation. The control information provides a bit rate and a quality measure during each time window T. The control information enables a statistical multiplexer (16, 160, 260) to select a bit rate representation for each program during each time window T to maximize the quality of the selected representations while not exceeding the total available channel capacity.

FIG. 1

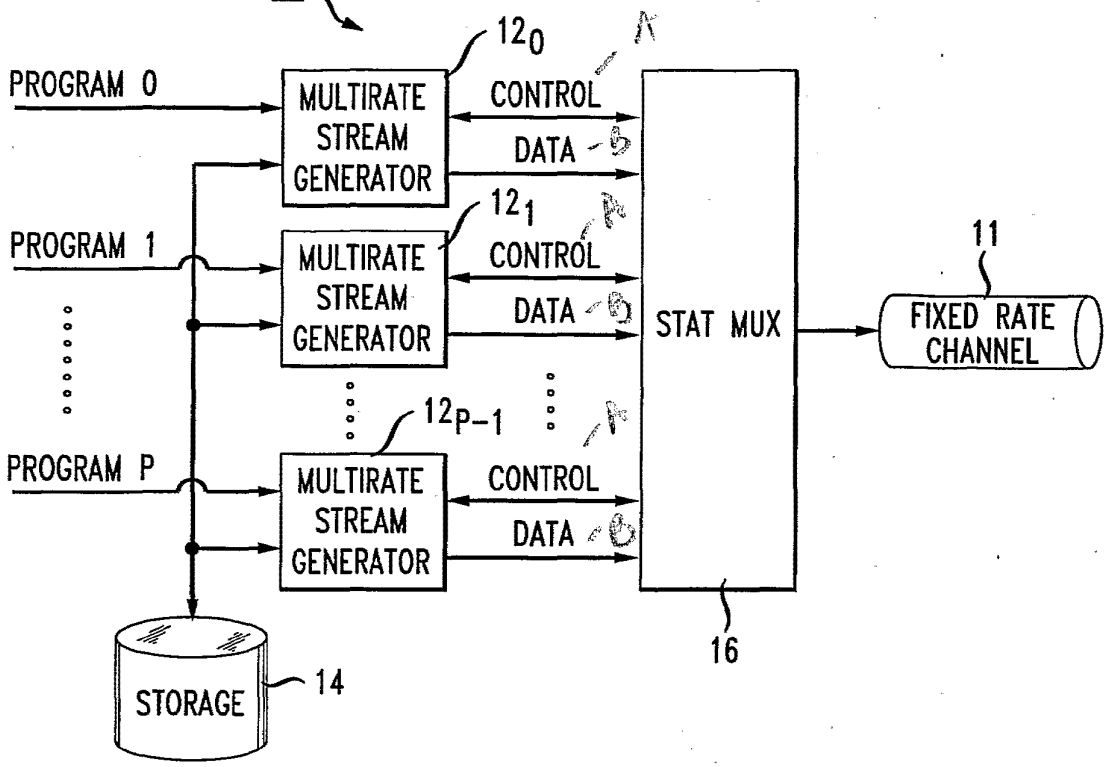

10
PROGRAM 0
PROGRAM 1
PROGRAM P
MULTIRATE STREAM GENERATOR
MULTIRATE STREAM GENERATOR
MULTIRATE STREAM GENERATOR
12_0
12_1
12_{P-1}
CONTROL
DATA
CONTROL
DATA
CONTROL
DATA
STAT MUX
11
FIXED RATE CHANNEL
16
STORAGE
14

FIG. 2

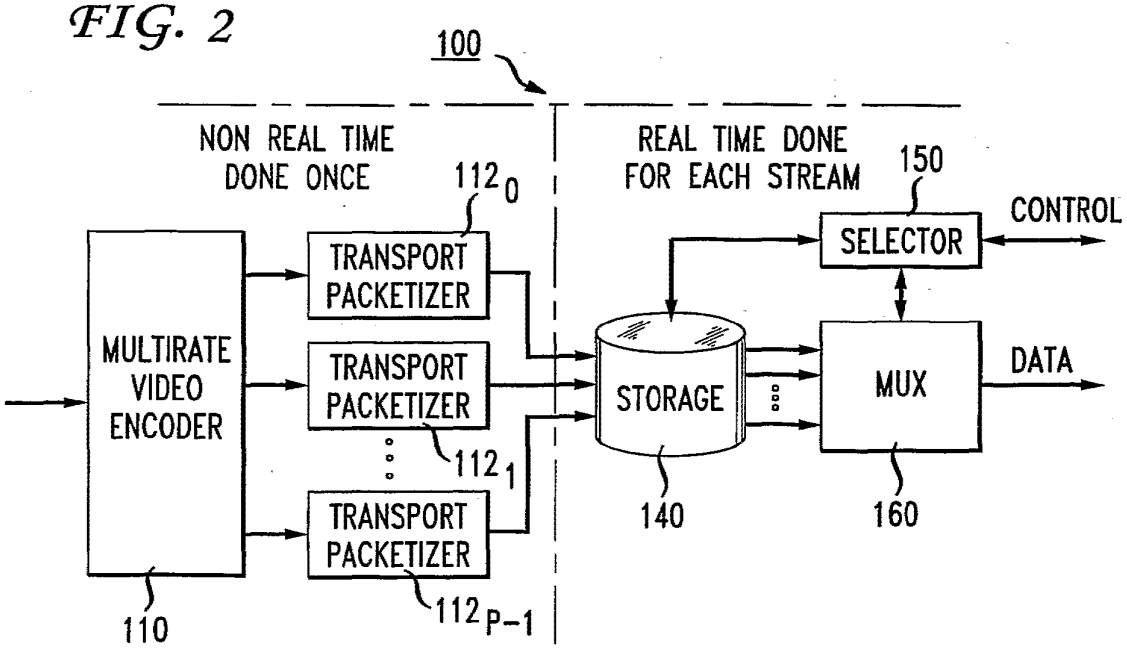
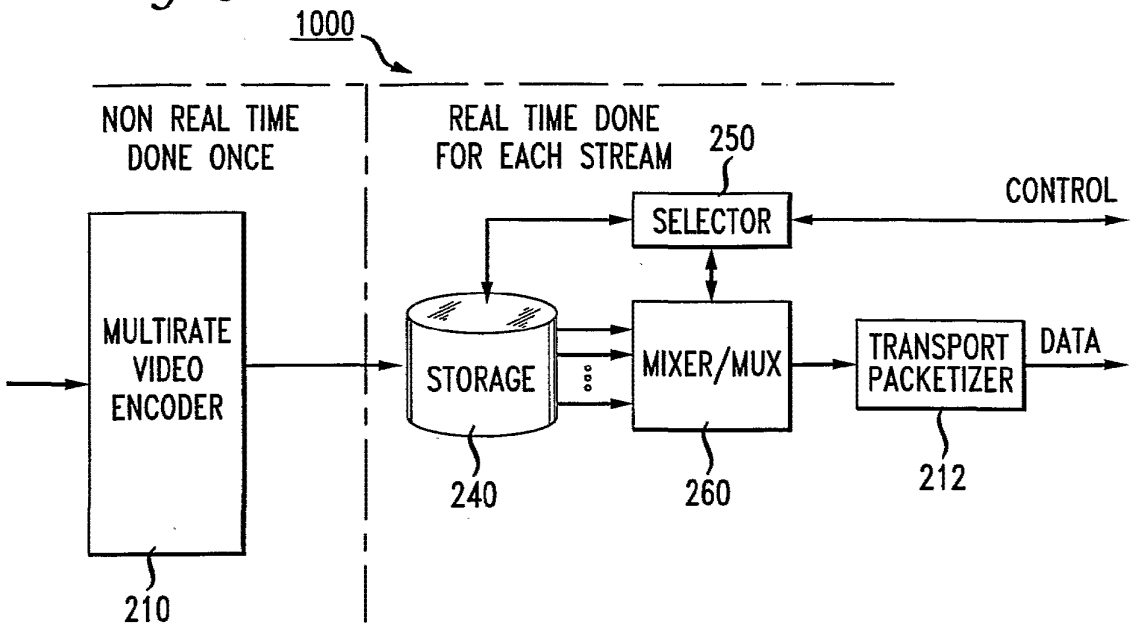

100
NON REAL TIME DONE ONCE
REAL TIME DONE FOR EACH STREAM
MULTIRATE VIDEO ENCODER
110
TRANSPORT PACKETIZER
TRANSPORT PACKETIZER
TRANSPORT PACKETIZER
112_0
112_1
112_{P-1}
STORAGE
140
150
SELECTOR
CONTROL
MUX
160
DATA

FIG. 3


1000
NON REAL TIME
DONE ONCE
REAL TIME DONE
FOR EACH STREAM
250
SELECTOR
CONTROL
MULTIRATE
VIDEO
ENCODER
STORAGE
MIXER/MUX
TRANSPORT
PACKETIZER
DATA
210
240
260
212

PCT
New International Application
Inventory of Unscannable or Missing Items

Serial Number

Check This Column if Item Is Present	Item	Check This Column if Item Is Missing on Filing
✓	Return Receipt Postcard	
	Check (amount $ _____)	✓
✓	PCT EASY Diskette	
	DNA Diskette	
	Exhibit	
	Express Mail Label or Envelope	
✓	Applicant Supplied Priority Document	
	Other (specify)	XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX
XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX	Cover Letter	
XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX XXXXXXXXXXXX	Other (specify)	

Winword/PTO \ Automation\Unscanned inventory.wpt sheet

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 April 2003 (03.04.2003)

PCT

(10) International Publication Number
WO 03/027876 A1

(51) International Patent Classification[7]: G06F 15/16

(21) International Application Number: PCT/US02/30874

(22) International Filing Date:
26 September 2002 (26.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/325,483 26 September 2001 (26.09.2001) US

(71) Applicants and
(72) Inventors: **REYNOLDS, Jodie, Lynn** [US/US]; 728 King Way, Folsom, CA 95630 (US). **INGRAHAM, Robert, Walter** [US/US]; 127 Honey Cook Circle, Folsom, CA 95630 (US).

(74) Agents: **GALLENSON, Mavis, S.** et al.; Ladas & Parry, Suite 2100, 5670 Wilshire Boulevard, Los Angeles, CA 90036-5679 (US).

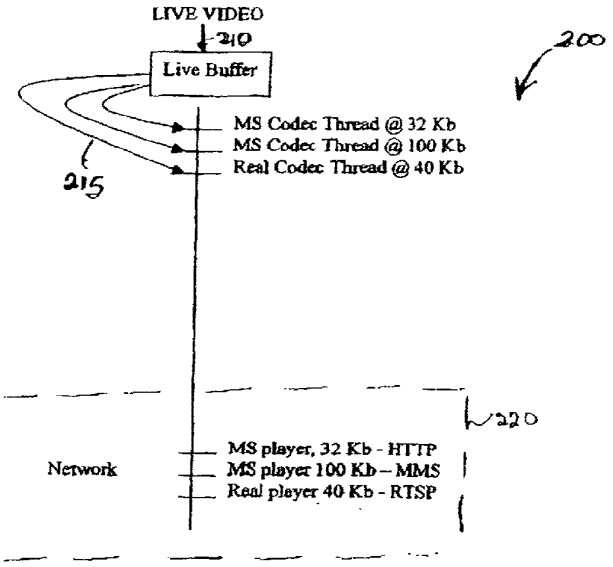
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR COMMUNICATING MEDIA SIGNALS



(57) Abstract: A media streaming system for streaming media signals is provided. The media streaming system takes a library of separate and distinct CODECs (145) that are provided as a searchable CODEC library and used in determining specific characteristics in the media signal to identify similar sections of the signal. The media streaming system uses a computer implemented intelligence system, such as an intelligence mechanism to learn and capture the unique charateristics of a signal as the signal is being streamed (150). The media streaming system also compresses and decompresses the media signal as the signals are streamed from a source media (100) to a destination device (130, 135, 140).

WO 03/027876 A1

SYSTEM AND METHOD FOR COMMUNICATING MEDIA SIGNALS

RELATED APPLICATION

This application is a non-provisional application of "Reynolds et al. a provisional application entitled "System and Method for Communicating Media Signals" attorney docket number ID-PAT-001PR1, application number 60/325,483, filed on September 26, 2001.

FIELD

This disclosure relates to a system and method for communicating media signals between source and destination devices. More specifically, it relates to a system and method for compressing and decompressing streaming and static media signals for efficiently communicating those signals between source and destination devices using artificial intelligence mechanism.

BACKGROUND

The ability to efficiently communicate streaming and static media between remotely located devices is a significant need that has emerged exponentially with the advent of networked communications such as the Internet. This need has been recently addressed with substantial development resources on a worldwide scale.

The term "media" is herein intended to mean information that may be communicated in the form of a signal from a source device to a destination device for use by the destination device; and, where used herein, media is generally contemplated to comprise either streaming or static media signals. For the purpose of this disclosure, the term "use" as applied to a destination device's operation on media signals is intended to include playing (e.g. sounds, images, video), processing (e.g. telemetry data), or any other use or operation that is the intended purpose of the media signal.

The terms "streaming media" are herein intended to mean media signals that comprise information intended to be communicated to and used by a destination device in a temporal, streaming fashion. The term "streaming" as applied to streaming media signals is herein intended to include signals communicated and processed in a continuous manner over time, or signals that may be

signals communicated and processed in a continuous manner over time, or signals that may be communicated in a series of discrete packets, pieces, or blocks that are interrelated and may be thereafter used by the destination device in a continuous, interrelated fashion. Examples of streaming media signals for the purpose of this disclosure therefore include, without limitation, the following types of media: video, audio, audio combined with video, and data strings such as temporal telemetry. The terms "streaming media" are most typically used by reference to digitized forms of data representing the subject media.

The terms "static media" are herein intended to generally mean media that is not "streaming" as defined above. Static media signals are of the type that generally may be communicated and are intended to be used as a single packet, block, or piece. Static media therefore may include for example, without limitation the following: a discrete image, an individual and relatively temporally short video clip, a sound or sound bite, or a piece or block of information such as telemetry information. It is contemplated, however, that such a "single piece" of static media may be of sufficient magnitude to consist of a plurality of smaller pieces or sub-parts, such as for example regions or pixels of an overall image, individual frames that together form a video clip, digital bits that together comprise a sound, a group of sounds that comprise a sound bite, or bits of information that together comprise a larger block of information.

Streaming media generally includes data files that are significantly larger than static media files, and also often represent many more variables over the temporal communication of such files than experienced with most static media files. Therefore, the ability to efficiently compress streaming media for appropriate communication to destination devices for use is often a much more complex and difficult to achieve goal. Accordingly, much of this disclosure is provided by reference specifically to streaming media communication, and the present invention has been observed to provide significant benefits for such communication. However, where streaming media is specifically referenced herein with respect to this background, and further with respect to the many benefits of the present invention herein disclosed, static media is also further contemplated where appropriate according to one of ordinary skill.

Many different "type-specific" media systems have been in use for quite a long time for transmitting specific types (e.g. video, audio, image, voice, etc.) of streaming and static media signals between sources and remote destinations. Typical examples of such type-specific media systems include television transmission systems, telephone line systems, and radio transmission systems, and every television, telephone, and radio is therefore a receiving device for media. Accordingly, the needs

for efficient communication of streaming and static media touch upon many diverse communications industries, including for example the telephone, television, movie, music, and more recently interactive gaming industries.

Moreover, many medial communications systems, including the various long-standing type-specific systems, are also "format specific", wherein the subject media signals are communicated in a particular format such that the source, transmission channel, and destination device must be specifically compliant to work within that format. Examples of format specific media systems include for example encoded cable television systems that work only for certain types of media and only delivered in particular encoded formats from the cable carrier. Therefore, these systems, in hardware and software, are generally dedicated to only the type and format of media to be provided by the content provider.

Society's needs have outpaced the abilities of these dedicated, content-specific and format-specific systems. In particular, these dedicated systems are not structured to accommodate the ever increasing client demand, real-time, for specified streaming media. Still further, technology developments in the recently interconnected world has tempted the palate of society for the ability to pull, receive, push, and send multiple types of media in multiple formats using one device. Moreover, content providers need to be able to deliver many different media signals to many different types of devices in their clients' offices, living rooms, and hands. Individuals and corporations also desire to communicate with each other using various different formats and using various different respective devices.

Accordingly, a significant industry has emerged for delivering streaming and static media over the centralized network of the Internet. Content delivery companies are currently delivering a wide range of streaming media, from live horse racing and entertainment to medical telemetry and education, over the Internet, and in video and audio formats. According to one published report from DFC Intelligence, video streaming on the Internet grew 215 percent in 2000 to over 900 million total streams accessed. This includes broadband streams, which made up almost 29 percent of total accesses. This same report also estimates that as much as 15 percent of available stream inventory is now being exploited with in-stream advertising. In another report published by Internet researcher Jupiter Media Metrix, business spending alone on streaming video technology will balloon from one-hundred forty million (US$140M) US dollars in 2000 to nearly three billion (US$3B) US dollars by 2005 as companies turn to electronic interaction in communicating with employees, consumers and other businesses.

Aylo Exhibit 1018
Page 2003 of 2308

Still further, the population explosion and increasing number of people transmitting on these systems has severely impacted the available bandwidth for available information. Therefore, the ability to stream media efficiently, using limited bandwidth resources and limited available transmission speeds, is of increased societal importance.

Compression/Decompression Algorithms ("CODECS")

In view of the exponential demand for communicating the different types of media, various compression/decompression systems ("CODEC(s)") have been developed over many years, and have in particular become the recent topic of significant research and development. Specific types of CODECS and systems for managing the operation of CODECS with respect to communicating streaming and static media signals have been developed for specific types of media, including for example still-frame images such as graphics and photographs, and streaming media.

Image CODECS

Various different types of static media CODECS have been developed, and a wide variety of these CODECS are widely known and used. One specific type of static media that has been the topic of particular attention includes images (though a long series of interrelated image frames such as in video context is generally treated as streaming media due to more complex variables, e.g. size and temporal relationship between frames, that significantly impact appropriate compression/decompression needs). Examples of static media CODECing is therefore herein exemplified by reference to certain specific types of conventional image CODEC technologies and methods.

The two most common file formats for graphic images on the world wide web are known as "GIF" and "JPEG" formats, generally considered the respective standards for drawings (e.g. line art) and photographs, and are further described together with other image compression modalities for the purpose of further understanding as follows.

"JPEG" is an acronym for "Joint Photographic Experts Group", and is a graphic image file that complies with ISO standard 10918. Commonly used for photograph compression/decompression, a JPEG file is created by choosing from a range of compression qualities, or, as has also been described, by choosing from one of a suite of compression algorithms. In order to create a JPEG file, or convert

an image from another format to JPEG, the quality of image that is desired must be specified. In general, because the highest quality results in the largest file, a trade-off may then be made, as chosen by the user, between image quality and image size. The JPEG mode of compression generally includes 29 distinct coding processes although a JPEG implementer may not use them all. A JPEG image is typically given a name suffix ".jpg".

"GIF" is an acronym for "Graphics Interchange Format", and is generally considered the de facto standard form of drawing image compression/decompression for Internet communication. GIF formatting uses a compression algorithm known as the LZW algorithm, which was developed by Abraham Lempel, Jacob Ziv, and Terry Welch and made commercially available by Unisys Corporation (though in general such algorithm has been made publicly available without requiring fee-bearing licenses). More specifically, a "LZW" compression algorithm takes each input sequence of bits of a given length (e.g. 12 bits) and creates an entry in a table, sometimes called a "dictionary" or "codebook", for that particular bit pattern. The entry consists of the pattern itself and a shorter code. As input is read, any pattern that has been read before results in the substitution of the shorter code, effectively compressing the total amount of input to something smaller. Earlier approaches, known as LZ77 and LZ78, did not include the look-up table as part of the compressed file. However, the more recent LZW algorithm modality does include the table in the file, and the decoding program that decompresses the file for viewing is able to build the table itself using the algorithm as it processes the encoded input. The GIF format uses the 2D raster data type (associated with display screens using raster lines) and is encoded in binary.

Two versions of GIF formats include GIF 87a, and more recently GIF89a that allows for "animated GIF" file creation, or short sequences of images within a single GIF file that are played in sequence to present movement or change in the image (either in an endless loop or through a progression that reaches an end). GIF89A also allows for, and also for "interlaced GIF", which is a GIF image that arrives and is displayed by the receiver first as a fuzzy outline of an image that is gradually replaced by seven successive waves of bit streams that fill in the missing lines until full resolution is reached. Interlaced GIF allows, for example, a viewer using 14.4 Kbps and 28.8 Kbps modems to observe a briefer wait-time before certain information in a subject image may be processed, such as for example to make decisions (e.g. to click on the image to execute an operation such as a link).

By presenting waves of resolution filling image sequences, interlaced GIF is similar to "Progressive JPEG", which describes an image created using the JPEG suite of compression algorithms

that will "fade in" in successive waves. While the progressive JPEG is often observed to be more appealing way to deliver an image at modem connection speeds, users with faster connections may not likely notice a difference.

"PNG" or "Portable Network Graphics" format has been more recently developed for image compression and that, in time, has been publicized to replace the GIF format for Internet use (though not generally the JPEG format allowing size/quality trade-offs). This format has been developed for public consumption and development. Similar to GIF, PNG is considered a "lossless" compression format, and therefore all image information is restored when a compressed file is decompressed during viewing. However, PNG formatted files are generally intended to be from 10 to 30 percent more compressed than with a GIF format. Further aspects of PNG file formats are provided as follows: (i) color transparency may not be limited to only one color, but the degree of transparency may be controlled ("opacity"); (ii) "interlacing" of images is improved versus standard GIF; (iii) "gamma correction" is enabled, allowing for "tuning" of images in terms of color brightness required by specific display manufacturers; (iv) images can be saved using true color, palette, and gray-scale formats similar to GIF; and (v) "animation" is generally not supported, though PNG is generally considered extensible and therefore software may be layered to provide for such scriptable image animation.

"TIFF" is an acronym for "Tag Image File Format", and is a common format for exchanging raster graphics (or "bitmap") images between application programs, such as for example graphics used for scanner images. A TIFF file is usually given a name suffix of ".tif" or ".tiff", and had generally been developed in the mid-1980's with the support of Adobe Software, Microsoft, and Hewlett-Packard. TIFF files can be in any of several classes, including gray scale, color palette, or RGB full color, the descriptions and differences of which are further developed elsewhere herein this disclosure. TIFF files may also include files with JPEG, LZW, or CCITT Group 4 standard run-length image compression, which are also further described elsewhere herein. As one of the most common graphic image formats, TIFF files are typically used in desktop publishing, faxing, 3-D applications, and medical imaging applications.

Video CODECS

Video compression has been the topic of intense development for various applications, including, for example: pre-recorded video (e.g. "video-on-demand"), teleconferencing, and live video (e.g. broadcasts). "Desk-top" computers, wireless devices, conventional televisions, and high

definition televisions are examples of the different types of receiving devices that an efficient video compression system must serve.

In general, video CODEC algorithms operate on either or both of an individual, frame-by-frame basis, and/or on a "temporal compression" basis wherein each frame is the most common video compression algorithms in conventional use are based on several mathematic principles, including the following: Discrete Cosine Transforms ("DCT"), Wavelet Transforms and Pure Fractals.

"Discrete Cosine Transforms" or "DCT's" are by far the most popular transforms used for image compression applications . In general, DCT is a technique for representing waveform data as a weighted sum of cosines. The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain. The DCT helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). Reasons for its popularity include not only good performance in terms of energy compaction for typical images but also the availability of several fast algorithms. DCTs are used in two international image/video compression standards, JPEG and MPEG.

"Wavelet transforms" are generally mathematical algorithms that convert signal data into a set of mathematical expressions that can then be decoded by a destination receiver device, such as for example in a manner similar to Fourier transform. Wavelets have been observed to enhance recovery of weak signals from noise, and therefore images processed in this manner can be enhanced without significant blurring or muddling of details. For this reason, wavelet signal processing has been particularly applied to X-ray and magnetic-resonance images in medical applications. In Internet communications, wavelets have been used to compress images to a greater extent than is generally possible with other conventional methods. In some cases, the wavelet-compressed image can be as small as about 25 percent the size of a similar quality image using the more familiar JPEG format, which is discussed in further detail elsewhere in this disclosure. Thus, for example, a photograph that requires 200Kb and takes a minute to download in JPEG format may require only 50Kb and take only 15 seconds to download in wavelet-compressed format. A wavelet-compressed image file is often given a name suffix ".wif", and either the receiver (e.g. Internet browser on a computer receiver) must support these format specific files, or a plug-in program will be required to read such file.

Fractal image compression is a modern technique of lossy image coding that provides several improvements over existing Fourier series compression schemes. Edge depiction is improved since, when modeled as a step function, edges require a large number of Fourier series terms to properly depict. Other advantages of fractals include fast decoding time and scale independence. Fractal

compression is based on Mandelbrot sets which take advantage of a self similar, scaling dependent, statistical feature of nature (Mandelbrot, 1983). Fractal compression and decompression involves a clustering approach to find regions which show the same characteristics as a sample region independent of rotation and scale. The fractal image compresses images as recursive equations and instructions about how to reproduce them. The equations describe the image in terms of the relationships between its components. The reduction in storage need is due to the fact that fractal compression saves equations and instructions instead of a pixel representation of the image.

"MPEG" is an acronym for Moving Picture Experts Group and has come to be used synonymously with certain evolving video and audio compression standards promulgated therefrom. In general, to use MPEG video files, a personal computer is required with sufficient processor speed, internal memory, and hard disk space to handle and play the typically large MPEG file, usually given the name suffix ".mpg". A specified MPEG viewer or client software that plays MPEG files must be available on the client system, and generally can be downloaded shareware or versions of commercial MPEG players from various sites on the Web. The modes of operation for MPEG formatted media are herein described by reference to these sequentially evolved standards as follows.

More specifically, MPEG-1 standard was designed for coding progressive video generally at a transmission rate of about 1.5Mbps. This was generally designed for the specific application for Video-CD and CD-I media. MPEG-1 audio layer-3 ("MP3") has also evolved from early MPEG work. "MPEG-2" is a standard generally designed for coding interlaced images at transmission rates above 4 Mbps, and was generally intended for use with digital TV broadcast and digital versatile disk. Though it is generally observed that many MPEG-2 players can handle MPEG-1 data as well, the opposite is not generally observed to be true and MPEG-2 encoded video is generally incompatible with MPEG-1 players. Yet another progressive standard, "MPEG-3", has also been proposed for use with high definition television ("HDTV"), though in general MPEG-3 has merged with MPEG-2 which is generally believed to meet the HDTV requirements. Finally, an "MPEG-4" standard has also been most recently developed and is intended to provide a much more ambitious standard to address speech and video synthesis, fractal geometry, and computer visualization, and has further been disclosed to incorporate artificial intelligence in order to reconstruct images.

MPEG-1 and –2 standards define techniques for compressing digital video by factors varying from 25:1 to 50:1. This compression is achieved according to these standards generally using five different compression techniques: (i) discrete cosine transform (DCT), which is a frequency-based transform; (ii) "quantization", which is a technique for losing selective information, e.g. lossy

prediction", wherein some images are predicted from the pictures immediately preceding and following the image.

Further more detailed examples of commercially available video compression technologies include: Microsoft Media Player™ (available from Microsoft Corporation), RealPlayer™ or RealSystem G2™ (commercially available from Real Networks™), Apple's QuickTime™ (commercially available from Sorenson™); and "VDO". The Microsoft Media Player™ is generally believed to apply the MPEG standard of CODEC for compression/decompression, whereas the others have been alleged to use proprietary types of CODECS. Standard compression algorithms, such as MPEG4, have made their way into the hands of developers who are building embedded systems for enterprise streaming, security, and the like.

One example of a more recent effort to provide streaming video solutions over Wireless and IP networks has been publicized by a company named Emblaze Systems (LSE:BLZ). This company has disclosed certain technology that is intended for encoding and playback of live and on-demand video messages and content on any platform: PC's, PDA's, Video cell phones and Interactive TV. Emblaze Systems is believed to be formerly GEO Interactive Media Group. The following Published International Patent Applications disclose certain streaming media compression technologies that is believed to be related to Emblaze Systems to the extent that GEO Interactive Media Group is named as "Assignee": WO9731445 to Carmel et al.; and WO9910836 to Carmel. The disclosures of these references are herein incorporated in their entirety by reference thereto.

Another company that has published CODEC technology that is intended to improve communication of streaming media for wireless applications is Packetvideo™ Corporation, more specifically intending to communicate streaming video to cellular phones. In addition, they are believed to be promoting CODEC technology that is intended to track temporal scalability and signal error resistance in order to protect video and audio streams from the hazards of the wireless environment. U.S. Patent No. 6,167,092 to Lengwehasatit discloses further examples of certain streaming media compression/decompression technology that are believed to be associated with Packetvideo as the named "Assignee" on the face of this Patent reference. The disclosure of this patent reference is herein incorporated in its entirety by reference thereto.

Another prior reference discloses CODEC technology that is intended to provide a cost effective, continuously adaptive digital video system and method for compressing color video data for moving images. The method involves capturing an analog video frame and digitizing the image into a preferred source input format for compression using a combination of unique lossy and lossless digital

Another prior reference discloses CODEC technology that is intended to provide a cost effective, continuously adaptive digital video system and method for compressing color video data for moving images. The method involves capturing an analog video frame and digitizing the image into a preferred source input format for compression using a combination of unique lossy and lossless digital compression techniques including sub-band coding, wavelet transforms, motion detection, run length coding and variable length coding. The system includes encoder and decoder (CODEC) sections, generally disclosed to use a "Huffman" encoder, for compression and decompression of visual images to provide high compression that is intended to provide good to excellent video quality. The compressed video data provides a base video layer and additional layers of video data that are multiplexed with compressed digital audio to provide a data stream that can be packetized for distribution over inter or intranets, including wireless networks over local or wide areas. The CODEC system disclosed is intended to continuously adjust the compression of the digital images frame by frame in response to comparing the available bandwidth on the data channel to the available bandwidth on the channel for the previous frame to provide an output data stream commensurate with the available bandwidth of the network transmission channel and with the receiver resource capabilities of the client users. The compression may be further adjusted by adjustment of the frame rate of the output data stream.

Further more detailed examples of CODEC systems that are intended for use at least in part for streaming video communication are disclosed in the following U.S. Patent Nos.: 6,081,295 to Adolph et al.; 6,091,777 to Guetz et al.; 6,130,911 to Lei; 6,173,069 B1 to Daly et al.; 6,263,020 B1 to Gardos et al.; 6,272,177 to Murakami et al.; and 6,272,180 B1 to Lei. The disclosures of these references are herein incorporated in their entirety by reference thereto.

Most if not all prior streaming video compression methodologies look to the extremely complex mathematical tools within such CODECS, and subtle changes to them, to carry "one size fits all" video over public and private networks of all types, from ultra-low bandwidth networks such as that found in wireless networks, to satellite communications to ultra-high speed fiber optic installations. Among the various conventional methods of compression, there are generally user-definable parameters, including tradeoffs between image size, frame rate, color depth, contrast, brightness, perceived frame quality, buffer length, etc. Further, within the algorithms themselves there are numerous non-user definable qualities and weighted calculations. It is up to the developers to set these one time for one "general" interest, and then package and ship the product.

However, while the video streaming market continues to grow rapidly, the world has not chosen one standard for compression as no one algorithm is ideal for all video sources, destinations, or transmission modalities. While a first CODEC may be best for one type of signal, or for a first portion of a signal (e.g. frame or scene comprising a series of frames), another second CODEC may be best for another type of signal, or even another second portion of the same signal. Still further, one CODEC may be best suited for compression/decompression of a particular streaming signal among send, receive, and transmission devices in a communications network; another second CODEC may be better suited than the first for the same streaming media signal but for another set of communication device parameters. For example, some video streams may deliver color to handheld devices while other video streams can take advantage of the loss of pixels in a black and white transmission to a cellular phone to increase frame rate. Required sound quality, frame rates, clarity, and buffering tolerance all decidedly impact the compression algorithm of choice for optimized video and audio delivery across multiple platforms.

In fact, certain communication device parameters may be sufficiently transient during the streaming media transmission such that an initially appropriate CODEC for an initial set of parameters may be rendered less efficient than another CODEC due to changes in those parameters during the same streamed signal transmission. Examples of such transient parameters include, without limitation: available band width in the data transmission channel, available memory or processing power in either the send or receiving devices, and dedicated display resolution/window in the receiving device (e.g. minimizing windows on a screen). These problems are compounded exponentially by a vast number of iterations of different combinations of such factors that may differentiate one CODEC from another as being most efficient for compression, decompression, and delivery of a specific streaming media signal along a particular communications device system.

As CODEC systems are "format-specific", source and destination devices must be "pre-configured" to communicate media signals between each other according to common, specific compression/decompression modalities, else transcoders must be used. However, even if conventional transcoders are used, constraints in the communication system (e.g. source, transmission channel, destination device) are not generally considered and the communication may be significantly faulty. For the purpose of further illustration, Figures 1A and 1B show two different schematic representations of conventional methods for communicating media between source 110 -120 and destination devices 130 -140. These illustrations specifically exemplify streaming video communication, though other media forms may be represented by similar systems.

Aylo Exhibit 1018
Page 2011 of 2308

It has been observed that CODEC algorithms can generally be modified for a specific application and then perform better than a similar unmodified set over that limited example. However, this generally must be done either for a series of frames, or ideally, for each individual frame. Some DCT based algorithms have as many as two billion mathematical operations occurring for each frame at higher resolutions and lower perceived quality. This is entirely too much math for average machines, even commercial servers, to perform thirty to sixty times in a single second. This is the reason for the advent of the dedicated compression board or ASIC.

Audio CODECS

In addition to society's recent interests in improving video compression, audio compression has likewise been the topic of significant efforts also for various live or pre-recorded applications, including audio broadcast, music, transmission synchronized with video, live interactive voice (e.g. telephone). Any and all of these audio compression applications must be compatible with a wide range of client-side receiver/players, such as on a multitude of handheld or desk-top devices having widely varied capabilities and operating parameters.

Conventional audio CODECS generally comprise several different types, a few of which are herein briefly summarized for the purpose of illustration.

"Code Excited Linear Prediction" or "CELP" is a type of speech compression method using waveform CODECS that use "Analysis-by-Synthesis" or "AbS" within the excitation-filter framework for waveform matching of a target signal. CELP-based CODECS have recently evolved as the prevailing technique for high quality speech compression, and has been published to transmit compressed speech of toll-quality at data rates nearing as low as about 6 kbps. However, at least one publication discloses that the quality of CELP coded speech is reduced significantly for bit rates at or below 4 kbps.

"Vocoders" are speech CODECS that are not based on the waveform coding scheme, but rather use a quantized parametric description of the target input speech to synthesize the reconstructed output speech. Vocoders have been disclosed to deliver better speech quality at low bit-rates, such as about 4 kbps, and have been developed for such applications. Low bit rate vocoders use the periodic characteristics of voiced speech and the "noise-like" characteristics of stationary unvoiced speech for speech analysis, coding and synthesis. Some early versions of vocoders (e.g. federal standard 1015 LPC-10, use a time-domain analysis and synthesis method. However, most of the more recent versions,

which at least one publication labels "harmonic coders", utilize a harmonic spectral model for voiced speech segments.

Notwithstanding the previous description of certain specific speech compression techniques, a vast number of speech CODECs and standards have been developed by industry and managed by industry and nonprofit groups. Examples of such groups include without limitation the following, and their standards are often used as reference types of CODECS: the European Telecommunications Standards Institute ("ETSI"); the Institute of Electrical and Electronics Engineers ("IEEE"); and the International Telecommunication Union Telecommunications Standards Sector ("ITU-T"), formerly the "CCITT".

One more recently disclosed method and apparatus for hybrid coding of speech, specified at 4 Kbps, encodes sppech for communication to a decoder for reproduction of the speech where the speech signal is classified into three types: (i) steady state voiced or "harmonic"; (ii) stationary unvoiced; and (iii) "transitory" or "transition" speech. A particular type of coding scheme is used for each class. Harmonic coding is used for steady state voiced speech, "noise-like" coding is used for stationary unvoiced speech, and a special coding mode is used for transition speech, designed to capture the location, the structure, and the strength of the local time events that characterize the transition portions of the speech. The compression schemes are intended to be applied to the speech signal or to the LP residual signal.

Another recently disclosed method and arrangement for adding a new speech encoding method to an existing telecommunication system is also summarized as follows. A CODEC is introduced into a speech transmitting transceiver of a digital telecommunications system in order to use a "new" CODEC and an "old" CODEC in parallel in the system. A CODEC is selected by implementing a handshaking procedure between transceivers where a speech encoding method is implemented in all transceivers and previously used in the telecommunications system concerned. The handshaking is used at the beginning of each connection. At the beginning of a phone call and after handover, the method checks whether both parties can also use the new speech encoding. The handshaking messages have been selected so that their effect on the quality of speech is minimal, and yet so that the probability of identifying the messages is maximal.

Still another relatively recent reference discloses a tunable perceptual weighting filter for tandem CODECS intended for use in speech compression. Specific filter parameters are tuned to provide improved perfromance in tandeming contexts. More specifically, the parameters used are 10th

order LPC predictor coefficients. This system is specified to use "Low-Delay Excited Linear Predictive" CODECS, or "LD-CELP".

Further more detailed examples of streaming audio communications systems using CODECS such as according to the examples just described are provided in the following US Patent references: 6,144,935 to Chen et al.; 6,161,085 to Haavisto et al.; and 6,233,550 to Gersho. The disclosures of these references are herein incorporated in their entirety by reference thereto.

Artificial Intelligence ("AI") and Neural Networks with CODECS

Various systems and methods have been recently disclosed that are intended to integrate artificial intelligence ("AI") or neural networks with the compression and decompression of streaming media signals.

The terms "artificial intelligence" are herein intended to mean the simulation of human intelligence processes by computer systems, including learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction. Particular applications of AI include "expert systems", which are computer programs that simulate judgment and behavior of a human or organization that has expert knowledge and experience in a particular field. Typically, expert systems contain a knowledge base to each particular situation that is described to the program, and can be enhanced with additions to the knowledge base or to the set of rules.

The terms "neural network" are herein intended to mean a system of programs and data structures that approximates the operation of the human brain, usually involving a large number of processors operating in parallel, each with its own small sphere of knowledge and access to data in its local memory. Typically, a neural network is initially trained or fed large amounts of data and rules about data relationships, after which a program can tell the network how to behave in response to an external stimulus (e.g. input information). In making determinations, neural networks use several principles, including without limitation gradient-based training and fuzzy logic. Neural networks may be further described in terms of knowledge layers, generally with more complex networks having deeper layers. In "feedforward" neural network systems, learned relationships about data can "feed forward" to higher layers of knowledge. Neural networks can also learn temporal concepts and have been widely used in signal processing and time series analysis. Other published applications of neural

networks include oil exploration data analysis, weather prediction, the interpretation of nucleotide sequences in biology labs, and the exploration of models of thinking and consciousness.

The terms "fuzzy logic" are herein intended to mean an approach to computing based upon "degrees of truth" rather than "Boolean logic" which operates within only a true/false (or "binary", as 1 or 0) domain. Fuzzy logic was first advanced by Dr. Lotfi Zadeh of the University of California at Berkeley in the 1960's in relation to work on a problem of computer understanding of natural language, which is not easily translated into absolute Boolean logic terms. Fuzzy logic often does include the cases of 0 and 1 as extreme cases of truth, but also includes the various states of truth in between (e.g. a determination of that the state of being is at some threshold, such as 0.98, may assists in making a decision to assign a 1 with an acceptably low occurrence of error in an operation).

One example of a previously disclosed streaming media compression/decompression system intended to use with artificial intelligence through a neural network uses a Radon transform in order to compress data such as video data. Several previously disclosed AI and/or neural network systems are intended to use AI and/or neural networks for the purpose of error correction during use of certain specified lossless compression CODECS. For example, a learning system is employed to determine a difference between what was received by a receiver after compression and transmission and what is predicted to have been received at the transmission end. That difference is processed as learning to modify the tuning of the CODEC for an additional transmission.

Another example of a disclosed method and device is intended to extrapolate past signal-history data for insertion into missing data segments in order to conceal digital speech frame errors. The extrapolation method uses past-signal history that is stored in a buffer. The method is implemented with a device that is disclosed to utilize a finite-impulse response ("FIR"), multi-layer, feed-forward, artificial neural network that is trained by back-propagation for one-step extrapolation of speech compression algorithm ("SCA") parameters. Once a speech connection has been established, the speech compression algorithm device begins sending encoded speech frames. As the speech frames are received, they are decoded and converted back into speech signal voltages. During the normal decoding process, pre-processing of the required SCA parameters will occur and the results stored in the past-history buffer. If a speech frame is detected to be lost or in error, then extrapolation modules are executed and replacement SCA parameters are generated and sent as the parameters required by the SCA. In this way, the information transfer to the SCA is intended to be transparent, and the SCA processing continues as usual. This disclosure alleges that the listener will not normally notice that a

speech frame has been lost because of the smooth transition between the last-received, lost, and next-received speech frames.

Further more detailed examples of systems that are intended to use artificial intelligence and/or neural networks in systems for media compression and/or decompression, generally relating to media type-specific CODEC methods (e.g. speech, video), are variously disclosed in the following U.S. Patent References: 5,005,206 to Naillon et al.; 5,041,916 to Yoshida et al.; 5,184,218 to Gerdes; 5,369,503 to Burel et al.; 5,598,354 to Fang et al.; 5,692,098 to Kurdziel; 5,812,700 to Fang et al.; 5,872,864 to Imade et al.; 5,907,822 to Prieto, Jr.; and 6,216,267 to Mitchell. Still further examples are provided in the following Published International Patent Applications: WO 01/54285 to Rising; EPO 0372608 A1 to Naillon et al.. The disclosures of all these references cited in this paragraph are herein incorporated in their entirety by reference thereto.

Other disclosures of CODEC systems using feedback or other systems for operating CODECS for use in processing a variety of streaming media signals, but that are not believed to specifically use the labels "AI" or "neural networks", are disclosed in the following U.S. Patent Nos.: 6,072,825 to Betts et al.; 6,182,034 B1 to Malvar; 6,253,165 B1 to Malvar; 6,256,608 B1 to Malvar. The disclosures of these references are herein incorporated in their entirety by reference thereto.

Notwithstanding the significant advancements in CODEC algorithms themselves, and despite prior intended uses of AI and other feedback systems for operating CODECS in order to improve compression efficiencies in communication, there is still a need for significant improvement in the ability to efficiently provide a wide variety of streaming media signals to a wide variety of destination receiver devices over a wide variety of transmission channels with varied bandwidths and communication protocols.

There is still a need to incorporate AI and/or neural networks to apply an appropriate CODEC for communication of a streaming media signal based upon a variety of parameters, including without limitation one or more of the following: (a) the automated choosing of an appropriately optimized CODEC from a library of available CODECS of different types and operation, including in particular based upon an intelligent knowledge of the chosen CODEC's operation compared to the other CODEC's operation and/or against a standard, (b) a pre-trained and/or iteratively learned knowledge of the particular CODEC's operation within a given set of operating parameters representative of the existing situation; and (c) a tuning of the appropriate CODEC based upon an intelligent knowledge of

its operation with respect to either or both of the existing situation or a test situation with reference parameters.

In particular, there is still a need for such an intelligent CODEC system that bases an applied CODEC upon an existing situation that is defined by one or more of the following: parameters of the streaming media signal itself; parameters of the transmission channel capabilities and constraints; and parameters of the receiver device capabilities and constraints.

Still further, there is also still a need for such an intelligent CODEC system that operates based upon an intelligent knowledge with respect to all of these operations and situational parameters in order to optimize the appropriate compression, transmission, decompression, and playing of the subject streaming media signal.

Conventional Transcoders for Streaming Media

Also of recent interest in the field of streaming media communication is providing intercommunication between the wide array of "format-specific" encoding systems in present use. An existing field of various different format-specific systems and pre-encoded content has created a widely fragmented ability to process encoded content, resulting in a significant quagmire of compatibility issues between content providers and client users. If one client desires to see or hear streaming content from a particular source and that content must be put through a CODEC for compression, a compatible CODEC must be used on the client side for decompression to enjoy the signal. Unfortunately, source content is often married to only a few, and often only one, specific CODEC schemes. Therefore, if a client requests such encoded content (or if the source desires to push the encoded content to a particular client), one of two criteria must be met: (1) the client must download or otherwise possess the format-specified CODEC (decoder); or (2) the source media must be put through a "transcoder" in order to decode the source media from the first format into a second format that is compatible with the client's device/system. The term "transcoder" is herein intended to mean a system that converts a media signal from one encoded (i.e. compressed) format to another.

Various techniques for transcoding one media format into another have been previously disclosed. Figure 1C shows one illustrative example of the general process that is characteristic of many known transcoding techniques. More specifically, a request 159 is first received from a particular type of device or player for content that exists in an initial, uncompatible format. According to the specific example shown in Figure 1C, a request 159 from a Microsoft Media™ Player for Real™ Video Content is received. As the content is specifically requested, the content is decoded from the

initial format (e.g. Real-encoded format), and is then "re-encoded" into the appropriate format for the requesting player (e.g. Microsoft Media™ format). This re-encoded media is then served to the requesting client for decoding within the resident system of that player.

This conventional system has significant scalability limitations, in that simultaneous feeds on multiple channels for multiple clients must be supported by an equal number of transcoders. For example, Figure 1D shows a schematic implementation of the conventional transcoding technique just described as it manages four simultaneous stream requests from four Microsoft Media Players, wherein the requested content is initially encoded in Real™ format. The system architecture necessary to support the four encoders 151 - 154 and four decoders 155 - 158 requires significant computing resources. For example, it is believed that each encoder 151 - 154 provided in the example requires a computer having a 600 MHz (e.g. Pentium™ III) having 128 Mbytes of RAM available, or dual 400 MHz processors (e.g. Pentium II) with 256 Mb available RAM. It is further believed that each decoder 155 – 158 needs a 233 MHz machine (e.g. Pentium™ II) having 64 Mb of available RAM. So, four such streams requires the equivalent of a Quad 900 Xeon (available from Compaq, Hewlett Packard, Dell and IBM, estimated to cost at the time of this disclosure about $9K retail). This is for four simultaneous streams – society is presently demanding thousands upon thousands of simultaneous streams.

There is still a need for a transcoder system that efficiently converts multiple format-specifically encoded streaming media signals into multiple other formats using minimal computing resources and in a cost-efficient manner.

Parameters Affecting Media Communication

For the purpose of further illustrating the many variables that may impact the choice of an appropriate CODEC in order to communicate a particular streaming media signal to a desired target, the following is a brief summary of various different types of streaming video formats and processing systems. It is believed that these different systems each generally require different types of compression modalities (e.g. CODECS) in order to optimize communication and playing of streaming media signals in view of available transmission speeds and bandwidth, as well as receiver processing parameters.

Although certain specific types of communications formats and systems are further herein described in detail, the following Table 1 provides a summary of a significant cross-section of the

various different communications systems and transmission carriers currently available or disclosed in view of available speed or bandwidth.

TABLE 1: Data Rates of Various Communications Carrier systems

Technology	Speed	Physical Medium	Application
GSM mobile telephone service	9.6 to 14.4 Kbps	RF in space (wireless)	Mobile telephone for business and personal use
High-Speed Circuit-Switched Data service (HSCSD)	Up to 56 Kbps	RF in space (wireless)	Mobile telephone for business and personal use
Regular telephone service (POTS)	Up to 53 Kbps	Twisted pair	Home and small business access
Dedicated 56Kbps on Frame Relay	56 Kbps	Various	Business e-mail with fairly large file attachments
DS0	64 Kbps	All	The base signal on a channel in the set of Digital Signal levels
General Packet Radio System (GPRS)	56 to 114 Kbps	RF in space (wireless)	Mobile telephone for business and personal use
ISDN	BRI: 64 Kbps to 128 Kbps PRI: 23 (T-1) or 30 (E1) assignable 64-Kbps channels plus control channel; up to 1.544 Mbps (T-1) or 2.048 (E1)	BRI: Twisted-pair PRI: T-1 or E1 line	BRI: Faster home and small business access PRI: Medium and large enterprise access
IDSL	128 Kbps	Twisted-pair	Faster home and small business access
AppleTalk	230.4 Kbps	Twisted pair	Local area network for Apple devices; several networks can be bridged; non-Apple devices can also be connected
Enhanced Data GSM Environment (EDGE)	384 Kbps	RF in space (wireless)	Mobile telephone for business and personal use
Satellite	400 Kbps (DirecPC and others)	RF in space (wireless)	Faster home and small enterprise access
Frame relay	56 Kbps to 1.544 Mbps	Twisted-pair or coaxial cable	Large company backbone for LANs to ISP ISP to Internet infrastructure
DS1/T-1	1.544 Mbps	Twisted-pair, coaxial cable, or optical fiber	Large company to ISP ISP to Internet infrastructure
Universal Mobile Telecommunications Service (UMTS)	Up to 2 Mbps	RF in space (wireless)	Mobile telephone for business and personal use (available in 2002 or later)
E-carrier	2.048 Mbps	Twisted-pair, coaxial cable, or optical fiber	32-channel European equivalent of T-1
T-1C (DS1C)	3.152 Mbps	Twisted-pair, coaxial cable, or optical fiber	Large company to ISP ISP to Internet infrastructure
IBM Token Ring/802.5	4 Mbps (also 16 Mbps)	Twisted-pair, coaxial cable, or optical fiber	Second most commonly-used local area network after Ethernet
DS2/T-2	6.312 Mbps	Twisted-pair, coaxial cable, or optical fiber	Large company to ISP ISP to Internet infrastructure
Digital Subscriber Line (DSL)	512 Kbps to 8 Mbps	Twisted-pair (used as a digital, broadband medium)	Home, small business, and enterprise access using existing copper lines
E-2	8.448 Mbps	Twisted-pair, coaxial cable, or optical fiber	Carries four multiplexed E-1 signals
Cable modem	512 Kbps to 52 Mbps (see "Key and explanation" below)	Coaxial cable (usually uses Ethernet); in some systems, telephone used for upstream requests	Home, business, school access
Ethernet	10 Mbps	10BASE-T (twisted-pair); 10BASE-2 or -5 (coaxial cable); 10BASE-F (optical fiber)	Most popular business local area network (LAN)
IBM Token Ring/802.5	16 Mbps (also 4 Mbps)	Twisted-pair, coaxial cable, or optical fiber	Second most commonly-used local area network after Ethernet
E-3	34.368 Mbps	Twisted-pair or optical fiber	Carries 16 E-1 signals
DS3/T-3	44.736 Mbps	Coaxial cable	ISP to Internet infrastructure Smaller links within Internet infrastructure

Aylo Exhibit 1018
Page 2019 of 2308

OC-1	51.84 Mbps	Optical fiber	ISP to Internet infrastructure Smaller links within Internet infrastructure
High-Speed Serial Interface (HSSI)	Up to 53 Mbps	HSSI cable	Between router hardware and WAN lines Short-range (50 feet) interconnection between slower LAN devices and faster WAN lines
Fast Ethernet	100 Mbps	100BASE-T (twisted pair); 100BASE-T (twisted pair); 100BASE-T(optical fiber)	Workstations with 10 Mbps Ethernet cards can plug into a Fast Ethernet LAN
Fiber Distributed-Data Interface (FDDI)	100 Mbps	Optical fiber	Large, wide-range LAN usually in a large company or a larger ISP
T-3D (DS3D)	135 Mbps	Optical fiber	ISP to Internet infrastructure Smaller links within Internet infrastructure
E-4	139.264 Mbps	Optical fiber	Carries 4 E3 channels Up to 1,920 simultaneous voice conversations
OC-3/SDH	155.52 Mbps	Optical fiber	Large company backbone Internet backbone
E-5	565.148 Mbps	Optical fiber	Carries 4 E4 channels Up to 7,680 simultaneous voice conversations
OC-12/STM-4	622.08 Mbps	Optical fiber	Internet backbone
Gigabit Ethernet	1 Gbps	Optical fiber (and "copper" up to 100 meters)	Workstations/networks with 10/100 Mbps Ethernet plug into Gigabit Ethernet switches
OC-24	1.244 Gbps	Optical fiber	Internet backbone
SciNet	2.325 Gbps (15 OC-3 lines)	Optical fiber	Part of the vBNS backbone
OC-48/STM-16	2.488 Gbps	Optical fiber	Internet backbone
OC-192/STM-64	10 Gbps	Optical fiber	Backbone
OC-256	13.271 Gbps	Optical fiber	Backbone

Comments & Key for Table:

(i)The term "Kbps" as the abbreviation for "thousands of bits per second." In international English outside the U.S., the equivalent usage is "kbits s^{-1}" or "kbits/s".
(ii) Engineers use *data rate* rather than *speed*, but speed (as in "Why isn't my Web page getting here faster?") seems more meaningful for the less technically inclined.
(iii) Relative to data transmission, a related term, bandwidth or "capacity," means how wide the pipe is and how quickly the bits can be sent down the channels in the pipe. These "speeds" are aggregate speeds. That is, the data on the multiple signal channels within the carrier is usually allocated by channel for different uses or among different users.

Key: (i) "T" = T-carrier system in U.S., Canada, and Japan....(ii) "DS"= digital signal (that travels on the T-carrier or E-carrier)...(iii) "E" = Equivalent of "T" that uses all 8 bits per channel; used in countries other than U.S. Canada, and Japan....(iv) "OC" = optical carrier (Synchronous Optical Network) "STM" = Synchronous Transport Modules (see Synchronous Digital Heirarchy). (v) Only the most common technologies are shown. (vi) "Physical medium" is stated generally and doesn't specify the classes or numbers of pairs of twisted pair or whether optical fiber is single-mode or multimode. (vii) The effective distance of a technology is not shown. (viii) There are published standards for many of these technologies.

Cable modem note: The upper limit of 52 Mbps on a cable is to an ISP, not currently to an individual PC. Most of today's PCs are limited to an internal design that can accommodate no more than 10 Mbps (although the PCI bus itself carries data at a faster speed). The 52 Mbps cable channel is subdivided among individual users. Obviously, the faster the channel, the fewer channels an ISP will require and the lower the cost to support an individual user.

Internet Carrier Systems

Communication of streaming video via the Internet may take place over a variety of transmission modalities, including for example digital subscriber lines ("DSL"), "T1" lines, cable modem, plain old telephone service ("POTS") dial-up modem, and wireless carriers. While a description of the many different wireless transmission modalities is treated separately herein, a

summary of various of these other transmission modes is herein provided immediately below for the purpose of further illustration as follows.

The terms "POTS" or "plain old telephone service", or "dial-up", as applied to communications transmission channels, are herein interchangeably used. These terms are intended to mean "narrow-band" communication that generally connects end users in homes or small businesses to a telephone company office over copper wires that are wound around each other, or "twisted pair". Traditional phone service was created to let you exchange voice information with other phone users via an analog signal that represents an acoustic analog signal converted into an electrical equivalent in terms of volume (signal amplitude) and pitch (frequency of wave change). Since the telephone company's signaling is already set up for this analog wave transmission, it's easier for it to use that as the way to get information back and forth between your telephone and the telephone company. Therefore, dial-up modems are used to demodulate the analog signal and turn its values into the string of 0 and 1 values that is called digital information. Because analog transmission only uses a small portion of the available amount of information that could be transmitted over copper wires, the maximum amount of data that you can receive using ordinary modems is about 56 Kbps. The ability of your computer to receive information is constrained by the fact that the telephone company filters information that arrives as digital data, puts it into analog form for your telephone line, and requires your modem to change it back into digital. In other words, the analog transmission between your home or business and the phone company is a bandwidth bottleneck.

With "ISDN", or "Internet subscriber digital network", which some consider to be a limited precursor to DSL, incoming data rates up to about 128 Kbps may be achieved for some end user clients.

A "DSL" or "digital subscriber line" is generally defined as a "broadband" transmission carrier for communicating high-bandwidth communication over ordinary copper telephone lines. Many different types of DSL services have been disclosed, having generally varied data rates and intended applications. Though further discussion is herein provided about certain of these DSL types, the following Table 2 provides a summary of information for certain of these DSL types for the purpose of further developing an overview understanding:

TABLE 2: Types of known DSL services.

DSL Type	Description	Data Rate Downstream; Upstream	Distance Limit	Application
IDSL	ISDN Digital Subscriber Line	128 Kbps	18,000 feet on 24 gauge wire	Similar to the ISDN BRI service but data only (no voice on the same line)
CDSL	Consumer DSL from Rockwell	1 Mbps downstream; less upstream	18,000 feet on 24 gauge wire	Splitterless home and small business service; similar to DSL Lite
DSL Lite (same as G.Lite)	"Splitterless" DSL without the "truck roll"	From 1.544 Mbps to 6 Mbps downstream, depending on the subscribed service	18,000 feet on 24 gauge wire	The standard ADSL; sacrifices speed for not having to install a splitter at the user's home or business
G.Lite (same as DSL Lite)	"Splitterless" DSL without the "truck roll"	From 1.544 Mbps to 6 Mbps , depending on the subscribed service	18,000 feet on 24 gauge wire	The standard ADSL; sacrifices speed for not having to install a splitter at the user's home or business
HDSL	High bit-rate Digital Subscriber Line	1.544 Mbps duplex on two twisted-pair lines; 2.048 Mbps duplex on three twisted-pair lines	12,000 feet on 24 gauge wire	T1/E1 service between server and phone company or within a company; WAN, LAN, server access
SDSL	Symmetric DSL	1.544 Mbps duplex (U.S. and Canada); 2.048 Mbps (Europe) on a single duplex line downstream and upstream	12,000 feet on 24 gauge wire	Same as for HDSL but requiring only one line of twisted-pair
ADSL	Asymmetric Digital Subscriber Line	1.544 to 6.1 Mbps downstream; 16 to 640 Kbps upstream	1.544 Mbps at 18,000 feet; 2.048 Mbps at 16,000 feet; 6.312 Mpbs at 12,000 feet; 8.448 Mbps at 9,000 feet	Used for Internet and Web access, motion video, video on demand, remote LAN access
RADSL	Rate-Adaptive DSL from Westell	Adapted to the line, 640 Kbps to 2.2 Mbps downstream; 272 Kbps to 1.088 Mbps upstream	Not provided	Similar to ADSL
UDSL	Unidirectional DSL proposed by a company in Europe	Not known	Not known	Similar to HDSL
VDSL	Very high Digital Subscriber Line	12.9 to 52.8 Mbps downstream; 1.5 to 2.3 Mbps upstream; 1.6 Mbps to 2.3 Mbps downstream	4,500 feet at 12.96 Mbps; 3,000 feet at 25.82 Mbps; 1,000 feet at 51.84 Mbps	ATM networks; Fiber to the Neighborhood

Typically published data rates for DSL service, which may vary depending upon distance from the central office of the offering service company, includes rates up to 6.1 Mbps (theoretically published at 8.448 Mbps), which is believed to enable continuous transmission of motion video, audio, and 3-D effects. More typical individual connections provide from 512 Kbps to 1.544 Mbps downstream and about 128 Kbps upstream. A DSL line can carry both data and voice signals and the data part of the line is continuously connected. DSL has been anticipated in some publications to replace ISDN in many areas and to compete with cable modem for multimedia communication to homes and businesses. DSL operates purely within the digital domain and does not require change into analog form and back. Digital data is transmitted to destination computers directly as digital data and this allows the phone company to use a much wider bandwidth for forward transmission. Meanwhile, if

Aylo Exhibit 1018
Page 2022 of 2308

a client user chooses, the signal can be separated so that some of the bandwidth is used to transmit an analog signal so that a telephone and computer may be used on the same line and at the same time.

Most DSL technologies require that a signal splitter be installed at a home or business, requiring the expense of a phone company visit and installation. However, it is possible to manage the splitting remotely from the central office. This is known as splitterless DSL, "DSL Lite," G.Lite, or Universal ADSL (further defined below) and has recently been made a standard. Several modulation technologies are used by various kinds of DSL, although these are being standardized by the International Telecommunication Union (ITU). Different DSL modem makers are using either Discrete Multitone Technology (DMT) or Carrierless Amplitude Modulation (CAP). A third technology, known as Multiple Virtual Line (MVL), is another possibility.

A variety of parameters of DSL operation are variable and affect the effective data rates that can be achieved. DSL modems generally follow the data rate multiples established by North American and European standards. In general, the maximum range for DSL without a repeater is 5.5 km (18,000 feet). As distance decreases toward the telephone company office, the data rate increases. Another factor is the gauge of the copper wire. The heavier 24 gauge wire carries the same data rate farther than 26 gauge wire. For destination devices beyond the 5.5 kilometer range, DSL may still be provided, though only generally if the respective phone company provider has extended the local loop with optical fiber cable.

To interconnect multiple DSL users to a high-speed network as a "backbone", the telephone company uses a Digital Subscriber Line Access Multiplexer ("DSLAM"). Typically, the DSLAM connects to an asynchronous transfer mode ("ATM") network that can aggregate data transmission at gigabit data rates. At the other end of each transmission, a DSLAM demultiplexes the signals and forwards them to appropriate individual DSL connections.

"ADSL" or "Asymmetric Digital Subscriber Line" is the form of DSL that will become most familiar to home and small business users. ADSL is called "asymmetric" because most of its two-way or "duplex" bandwidth is devoted to the downstream direction, sending data to the user. Only a small portion of bandwidth is available for upstream or user-interaction messages. However, most Internet and especially graphics- or multi-media intensive Web data need lots of downstream bandwidth, but user requests and responses are small and require little upstream bandwidth. Using ADSL, up to 6.1 megabits per second of data can be sent downstream and up to 640 Kbps upstream. The high downstream bandwidth means that a telephone line may carry motion video, audio, and 3-D images to destination computers or television displays. In addition, a small portion of the downstream bandwidth

can be devoted to voice rather data, and phone conversations may be carried without requiring a separate line. Unlike a similar service over "cable" television lines, ADSL does not compete for bandwidth with neighbors in a given area. In many cases, your existing telephone lines will work with ADSL. In some areas, they may need upgrading.

"CDSL" or "Consumer DSL" is a trademarked version of DSL, to be made available by Rockwell Corporation, that is somewhat slower than ADSL (1 Mbps downstream, generally predicted to be lower upstream) but has the advantage that a "splitter" does not need to be installed at the user's end. Hardware may be required to carry CDSL by local phone companies to homes or businesses. CDSL uses its own carrier technology rather than DMT or CAP ADSL technology.

Various companies have worked with telephone companies in developing a standard and easier installation version of ADSL, called "G.Lite", that is believed to be under deployment at the time of this disclosure. "G.Lite" or "DSL Lite" (also known as "splitterless ADSL", and "Universal ADSL") is believed to be essentially a slower ADSL that doesn't require splitting of the line at the user end but manages to split it for the user remotely at the telephone company, which is believed to lower costs. G.Lite, officially ITU-T standard G-992.2, is published to provide a data rate from 1.544 Mbps to 6 Mpbs downstream and from about 128 Kbps to about 384 Kbps upstream. At least one publication has predicted G.Lite to become the most widely installed form of DSL.

"HDSL" or "High bit-rate DSL" is believed to be the earliest variation of DSL to be widely used for wideband digital transmission within a corporate site and between the telephone company and a customer. The main characteristic of HDSL is that it is symmetrical: an equal amount of bandwidth is available in both directions. For this reason, the maximum data rate is generally lower than for ADSL. HDSL can carry as much on a single wire of twisted-pair as can be carried on a T1 line in North America or an E1 line in Europe (up to about 2.32 Mbps).

"IDSL" or "ISDN DSL" is somewhat of a misnomer since it's really closer to ISDN data rates and service at about 128 Kbps than compared with the much higher rates generally associated with ADSL.

"RADSL" or "Rate-Adaptive DSL" is an ADSL technology to be made available from Westell company in which software is able to determine the rate at which signals can be transmitted on a given customer phone line and adjust the delivery rate accordingly. Westell's "FlexCap2™" version system uses RADSL to deliver from about 640 Kbps to about 2.2 Mbps downstream and from about 272 Kbps to about 1.088 Mbps upstream over an existing line.

"SDSL" or "Symmetric DSL" is similar to HDSL with a single twisted-pair line, carrying about 1.544 Mbps (U.S. and Canada) or about 2.048 Mbps (Europe) each direction on a duplex line. It's symmetric because the data rate is the same in both directions.

"UDSL" or "Unidirectional DSL" is a proposal from a European company, and is generally believed to provide a unidirectional version of HDSL.

"VDSL" or "Very high data rate DSL" is believed to be a technology under development that promises much higher data rates over relatively short distances, for example between about 51 and about 55 Mbps over lines up to about 1,000 feet or about 300 meters in length. At least one publication has predicted that VDSL may emerge somewhat after ADSL is widely deployed and co-exist with it. The transmission technology (CAP, DMT, or other) and its effectiveness in some environments is not yet determined. A number of standards organizations are working on it.

"x2/DSL" is modem from 3Com that supports 56 Kbps modem communication but is upgradeable through new software installation to ADSL when it becomes available in the user's area. At least one publication cites 3Com as describing this technology to be "the last modem you will ever need."

A "T1" transmission line is generally considered a "broadband" carrier and is defined as a type of "T-carrier" system, which is believed to have been first introduced by the Bell System in the U.S. in the 1960's as the first successful system that supported digitized voice transmission. The T-carrier system is entirely digital, using pulse code modulation and time-division multiplexing. Voice signals are typically sampled at about 8,000 times a second and each sample is digitized into an 8-bit word. With 24 channels digitized at the same time, a 192-bit frame, representing 8-bit words on each of 24 channels, is thus transmitted about 8,000 times a second. Each frame is separated from the next by a single bit, resulting in a 193-bit block. The T-1's published data rate of 1.544 Mbps generally represents the 192 bit frame, and the 1-bit signaling bit, multiplied by 8,000.

A T-1 system typically uses 4 wires and provides duplex capability, with two wires dedicated for receiving and two for sending at the same time. The T-1 digital stream includes 24, 64 Kbps channels that are multiplexed, wherein the standard 64 Kbps channel is based on the bandwidth required for a voice conversation. The four wires were originally a pair of twisted pair copper wires, but more recent systems provide coaxial cable, optical fiber, digital microwave, and other carrier technologies. The number and use of the channels may be varied from the standard guidelines.

The original transmission rate (1.544 Mbps) for T-1 lines is in common use today in Internet service provider ("ISP") connections to the Internet. Another level, the T-3 line, is published to

provide 44.736 Mbps, and is also commonly used by Internet service providers. Another commonly used service is "fractional T-1", which is the rental of some portion of the 24 channels in a T-1 line, with the other channels unused.

Display Capabilities/Constraints & Related Standards

Various different types of receiver display capabilities may also significantly impact the appropriate CODEC modality for efficiently communicating particular streaming media signals for display by the receiver. A brief summary of certain examples to illustrate such varied display parameters (e.g. resolution, clarity, color, depth, size, type/format-specific) is provided for a better understanding as follows.

One parameter that is highly variable between different types and makes of streaming media receiver devices, and therefore that may have significant impact on the appropriate CODEC to be used, is the range of colors that may be expressed by a display device, or "palette". A standard "browser-safe" palette, which may be accommodated by most software for Internet-based streaming media display, may include for example about 216 colors, though for web-based streaming media the computer display capability as well as the browser software capability must be understood.

With respect to computer display technology, a color is set for each individual pixel or addressable illumination element on the screen. Each pixel has a red, green, and blue (RGB) component. By specifying the amount of intensity for each of these components, a distinct color is given to that pixel. A "true color" display generally defines the color of a pixel on a display screen using a 24-bit value, allowing the possibility of up to 16,777,216 possible colors. The number of bits used to define a pixel's color shade is called the "bit-depth". True color is sometimes referred to as "24-bit color", though many modern color display systems offer a 32-bit color mode. An extra byte, called the "alpha channel", is typically used for control and special effects information. A "gray scale" (composed of discrete shades of gray) display setting is generally defined as having N bits of depth where N represents the saturation of black within the pixel. If N=1, the image is not called gray scale but instead monochrome, or black and white, as the bit can only be on or off and can contain no shading information.

Common computer resolutions include for example and without limitation the following:

Aylo Exhibit 1018
Page 2026 of 2308

(i) VGA or Video Graphics Array capable of displaying 640x480 pixels in 16 colors or 320x240 pixels in 256 colors in a 4:3 aspect ratio;

(ii) SVGA or Super Video Graphics Array capable of 800x600x6 bits/pixel (16 colors) or 650x480x8 bits/pixel (256 colors). SVGA was created by the Video Electronics Association (VESA); and

(iii) XGA (v1-4) or eXtended Graphics Array capable of 1024x768 pixels at 32,768 colors.

Additional standards have been added such as SXGA, defining pixel sizes above 1960x1440 and color depths of 32 bits/pixel and higher.

In the event that a larger range of colors (or palette) is used by a media signal than a particular display or browser can handle, most browsers are typically adapted to "dither" the colors, which is herein intended to mean that the browser will find colors within its palette that it can substitute for any color that is outside of its palette. To further illustrate the wide range of different system display capabilities, systems using WindowsTM (commercially available from Microsoft Corporation) and MacintoshTM (commercially available from Apple Corporation) based operating systems do not have identical palettes; within the usual 256 color palette, 216 are common to both types of browsers, whereas 40 are different and therefore require dithering by a browser operating within one of the systems if an image signal is communicated to that type of system in a format specified by the other.

Many different technologies also exist with respect to how a visual display is enabled from electronic information. The terms "VDT" or "Video Display Terminals" are generally used within the computer industry and are herein intended to be used interchangeably with simple references to "display". With respect to computer terminal use, VDT's comprise a computer output surface and projecting mechanism that shows text and graphic images to the computer user. VDT's may use a variety of specific display technologies, including for example cathode ray tubes ("CRTs"), liquid crystal displays ("LCDs"), light-emitting diodes ("LEDs"), gas plasma, or other image projection technology. The display is usually considered to include the screen or projection surface and the device that produces the information on the screen. In some computers, the display is packaged in a separate unit or "monitor", or the display may be fully integrated in a single unit with the computer processor.

With respect to LCD's in particular, this technology generally requires minimal volume and physical depth compared to other VDT's, and therefore is typically used in laptop computers and cellphone/PDA's. LCD's consume much less power than LED and gas–display VDT's because they work on the principle generally of blocking light rather than emitting it. An LCD may be either

"passive matrix" or "active matrix", which is also known as "thin film transistor" or "TFT" display. The passive matrix LCD has a grid of conductors with pixels located at each intersection in the grid. A current is sent across two conductors on the grid to control the light for any pixel. An active matrix has a transistor located at each pixel intersection, requiring less current to control the luminance of a pixel. For this reason, the current in an active matrix display can be switched on and off more frequently, improving the screen refresh time and therefore efficacy for higher speeds of streaming media (e.g. action video). Some passive matrix LCD's have dual scanning, in that they scan the grid twice with current in the same time as one scan in earlier versions; however, the active matrix is still generally considered to be the superior technology of the two. Reflective color display technology—the integration of color filters into passive-matrix display construction—is a low-power, low-cost alternative to active-matrix technology. Because they reflect ambient light, reflective LCDs deliver particularly high performance during use outside in daylight. Various different display technologies, and therefore transmission formats, have also been specifically developed for television viewing. Thus several different standards have evolved for television transmission, and their differences may significantly impact the nature and extent of compression desired (and therefore the choice of a particular CODEC) for communicating streaming media signals in television environs. These standards include in particular and without limitation: standard definition television ("SDTV"); and high definition television ("HDTV").

"SDTV" or "standard definition television" and "HDTV" or "high definition television" are the two categories of display formats for digital television ("DTV") transmissions, which are becoming the standard. These formats provide a picture quality similar to digital versatile disk ("DVD"), and are summarized relative to their similarities and differences as follows.

HDTV provides a higher quality display, with a vertical resolution display from about 720p to at least about 1080i and an aspect ratio (the width to height ratio of the screen) of generally 16:9, for a viewing experience similar to watching a movie. In comparison, SDTV has a range of lower resolutions and no defined aspect ratio. New television sets will be either HDTV-capable or SDTV-capable, with receivers that can convert the signal to their native display format. SDTV, in common with HDTV, using the MPEG-2 file compression method in a manner that generally reduces a digital signal from about 166 Mbps to about 3 Mbps. This allows broadcasters to transmit digital signals using existing cable, satellite, and terrestrial systems. MPEG-2 uses the lossy compression method, which means that the digital signal sent to the television is compressed and some data is lost, but this lost data may or may not affect how the human eye views the picture. Both the ATSC and DVB standards selected

Aylo Exhibit 1018
Page 2028 of 2308

MPEG-2 for video compression and transport. The MPEG-2 compression standard is elsewhere herein described in further detail.

Because a compressed SDTV digital signal is smaller than a compressed HDTV signal, broadcasters can transmit up to five SDTV programs simultaneously instead of just one HDTV program, otherwise known as "multicasting". Multicasting is an attractive feature because television stations can receive additional revenue from the additional advertising these extra programs provide. With today's analog television system, only one program at a time can be transmitted. Note that this use of the term "multicasting" is distinct from its use in streaming video where it involves using special addressing techniques.

When the United States decided to make the transition from analog television to DTV, the Federal Communications Commission decided to let broadcasters decide whether to broadcast SDTV or HDTV programs. Most have decided to broadcast SDTV programs in the daytime and to broadcast HDTV programs during prime time broadcasting. Both SDTV and HDTV are supported by the Digital Video Broadcasting (DTV) and Advanced Television Systems Committee (ATSC) set of standards.

HDTV as a television display technology provides picture quality similar to 35 mm. movies with sound quality similar to that of today's compact disc (further with respect to audio quality, HDTV receives, reproduces, and outputs Dolby Digital 5.1). Some television stations have begun transmitting HDTV broadcasts to users on a limited number of channels. HDTV generally uses digital rather than analog signal transmission. However, in Japan, the first analog HDTV program was broadcast on June 3, 1989. The first image to appear was the Statue of Liberty and the New York Harbor. It required a 20 Mhz channel, which is why analog HDTV broadcasting is not feasible in most countries.

HDTV provides a higher quality display than SDTV, with a vertical resolution display from 720p to 1080i. The *p* stands for *progressive scanning*, which means that each scan includes every line for a complete picture, and the *i* stands for *interlaced scanning* which means that each scan includes alternate lines for half a picture. These rates translate into a frame rate of up to 60 frames per second, twice that of conventional television. One of HDTV's most prominent features is its wider aspect ratio (the width to height ratio of the screen) of 16:9, a development based on a research-based belief that the viewer's experience is enhanced by screens that are wider. HDTV pixel numbers range from one to two million, compared to SDTV's range of 300,000 to one million. New television sets will be either HDTV-capable or SDTV-capable, with receivers that can convert the signal to their native display format.

In the United States, the FCC has assigned broadcast channels for DTV transmissions. In SDTV formats, DTV makes it possible to use the designated channels for multiple signals at current quality levels instead of single signals at HDTV levels, which would allow more programming with the same bandwidth usage. Commercial and public broadcast stations are currently deciding exactly how they will implement their use of HDTV.

Simulcast is the simultaneous transmission of the same television program in both an analog and a digital version using two different channels or frequencies. At the end of the DTV transition period, it is believed by that analog transmission will be substantially replaced such that current analog channels will be used solely for DTV. The extra channels that were used for digital broadcasting may for example then be auctioned and used for more television channels or other services such as datacasting. Simulcast is also used for the transmission of simultaneous television and Internet services, the transmission of analog and digital radio broadcasts, and the transmission of television programs in different screen formats such as the traditional format and the wide screen format. Simulcast broadcasting is used worldwide.

The transition to DTV is not an easy or inexpensive transition. For a television station to transmit DTV programming, it must build its DTV facilities, but a station must have revenue to build these facilities. Simulcast allows stations to continue receiving revenues from traditional analog programming and also gain extra revenues from the extra digital programming. Another obstacle in the transition to DTV is lack of interest among consumers. The need for special equipment is prohibiting viewers from seeing the difference between digital and analog programs, which is also slowing down public enthusiasm for DTV.

The equipment needed for operating DTV depends on whether terrestrial, cable, or satellite services are used as the transmission channel/carrier. In any event, and according to known or anticipated systems, it is generally believed that consumers will, at a minimum, have to purchase a converter to view DTV transmissions on their old television sets. In addition, consumers that use terrestrial services or antennas to receive television signals need an antenna equipped for digital signals. A consumer located in mountainous terrain in an ATSC-compliant country may not be able to receive terrestrial-based digital signals because of multipath effects. This is common even with today's analog television system. In DVB compliant countries, terrain does not affect the reception of digital signals. Satellite users are already enjoying DTV broadcasting, but a larger satellite dish might be needed to view HDTV programming.

A "set-top" box is herein defined as a device that enables a television set to become a user interface to the Internet and also enables an analog television set to receive and decode DTV broadcasts. DTV set-top boxes are sometimes called receivers. It is estimated that 35 million homes will use digital set-top boxes by the end of 2006, the estimated year ending the transition to DTV.

A typical digital set-top box contains one or more microprocessors for running the operating system, usually Linux or Windows CE, and for parsing the MPEG transport stream. A set-top box also includes RAM, an MPEG decoder chip, and more chips for audio decoding and processing. The contents of a set-top box depend on the DTV standard used. DVB-compliant set-top boxes contain parts to decode COFDM transmissions while ATSC-compliant set-top boxes contain parts to decode VSB transmissions. More sophisticated set-top boxes contain a hard drive for storing recorded television broadcasts, for storing downloaded software, and for other applications provided by the DTV service provider. Digital set-top boxes can be used for satellite and terrestrial DTV but are used mostly for cable television. A set-top box price ranges from $100 for basic features to over $1,000 for a more sophisticated box.

In the Internet realm, a set-top box often really functions as a specialized computer that can "talk to" the Internet - that is, it contains a Web browser (which is really a Hypertext Transfer Protocol client) and the Internet's main program, TCP/IP. The service to which the set-top box is attached may be through a telephone line as, for example, with Web TV or through a cable TV company like TCI.

To take advantage of Dolby Digital 5.1 channel for satellite broadcasts, a satellite receiver that provides a Dolby Digital output is necessary. For cable users, all digital set-top boxes are equipped with a Dolby Digital two-channel decoder. To use 5.1 channel sound, a 5.1 channel-compliant set-top box is needed or an external 5.1 channel decoder unit.

The most dramatic demonstration of digital television's benefits is through a high-end HDTV, because of the larger screen, wider aspect ratio and better resolution. Like most new technologies, however, HDTV is expensive. Nevertheless, less expensive digital TVs provide a markedly improved viewing experience over regular TV, and for those who choose to retain their old sets, even the addition of a set-top converter will deliver a discernibly improved picture and sound.

The FCC's schedule for transition to DTV proposes that everyone in the U.S. should have access to DTV by 2002 and that the switch to digital transmission must be completed either by 2006 or when 85% of the households in a specific area have purchased digital television sets or set-top converters.

Aylo Exhibit 1018
Page 2031 of 2308

In the early 1990s, European broadcasters, consumer equipment manufacturers, and regulatory bodies formed the European Launching Group (ELG) which launched a "DVB" or "Digital Video Broadcasting" project in order to introduce DTV throughout Europe. DVB is intended to provide an open system as opposed to a closed system. Closed systems are content-provider specific, not expandable, and optimized only for the system they were developed for. An open system, such as DVB, allows the subscriber to choose different content providers and allows integration of PCs and televisions. DVB systems are intended to be optimized for television, but as well as supporting home shopping and banking, private network broadcasting, and interactive viewing. DVB is intended to open the possibilities of providing crystal-clear television programming to television sets in buses, cars, trains, and even hand-held televisions. DVB is also promoted as being beneficial to content providers because they can offer their services anywhere DVB is supported regardless of geographic location. They can also expand their services easily and inexpensively and ensure restricted access to subscribers reducing lost revenues due to unauthorized viewing. Today, the DVB Project consists of over 220 organizations in more than 29 countries worldwide and DVB broadcast services are available in Europe, Africa, Asia, Australia, and parts of North and South America

Format-Specific Media

Various different formats for the streaming media signals themselves are also herein summarized by way of non-limiting example to also provide a further understanding of how CODECS may vary for a particular case.

"DVD" is an acronym for "digital versatile disc" is generally defined as a relatively recent optical disc technology that holds up to about 4.7 Gigabytes of information on one of its two sides, or enough for a movie about 133 minutes long on average. With two layers on each of its two sides, it may hold up to 17 Gigabytes of video, audio, or other information, compared to current CD-ROM discs of approximately the same physical size that hold about 600Mbytes (DVD holds more than about 28 times the information). DVD players are required to play DVD's, though they will also play regular CD-ROM discs. DVDs can be recorded in any of three general formats variously optimized for: (i) video (e.g. continuous movies); (ii) audio (e.g. long playing music); and (iii) or a mixture (e.g. interactive multimedia presentations). The DVD drive has a transfer rate somewhat faster than an 8-speed CD-ROM player. DVD format typically uses the MPEG-2 file and compression standard, which has about 4-times the resolution of MPEG-1 images and can be delivered at about 60 interlaced fields

per second where two fields constitute one image (MPEG-1 delvers about 30 non-interlaced frames per second. MPEG-2 and -1 standards are elsewhere herein defined in more detail. Audio quality on DVD is comparable to that of current audio compact discs.

"DVD-Video" is the name typically given for the DVD format designed for full-length movies and is a box that will work with a television set. "DVD-ROM" is a name given to the player that is believed by some to be the future replacement to CD-ROM drives in computers, as these newer drives are intended to play both regular CD-ROM discs as well as DVD-ROM discs. "DVD-RAM" is the name given to writeable versions of DVDs. "DVD-Audio" is the name typically given to players designed to replace the compact disc player.

"VHS" is an acronym for "Video Home System" and is generally defined as a magnetic videotape cartridge format, typically a half-inch wide, developed for home use with the ability to record and playback analog video and audio signals. VHS has become a popular format and the de facto standard for home movie distribution and reproduction mainly due to its pervasive presence and recordability. VHS stores signals as an analog format on a magnetic tape using technology similar to that of audiocassettes. The tapes are played back and recorded on using VHS video cassette recorders (VHS VCRs). VHS tapes store up to around two hours of video typically, although some VCRs are able to record to them at a slower speed allowing up to six or even eight hours of recording per tape.

The VHS format outputs a little over 200 lines of horizontal resolution. This compares to DVDs that output over 500 lines of horizontal resolution. Technically and perceptually, VHS is a format that has been surpassed by other formats, including for example DVD, S-VHS, Hi-8 and others. However, VHS remains a pervasive means for viewing video, and VHS tapes are still easily found across the country and around the world everywhere from movie rental stores to grocery stores making then easily accessible.

"CD" is an acronym for "compact disc" and is generally defined as small, portable, round medium for electronically recording, storing, and/or playing audio, video, text, and other information in digital form. Initially, CD's were only read-only; however, newer versions allow for recording as well (e.g. "CD-RW").

"Super audio compact disc" or "SACD" is a high resolution audio CD format that, together with DVD-Audio ("DVD-A"), are the two formats competing to replace the standard audio CD (though most of the industry generally is backing DVD-A, with a general exception believed to be Philips and Sony). SACD, like DVD-A, offers 5.1 channel surround sound in addition to 2-channel stereo. Both formats improve the complexity of sound by increasing the bit rate and the sample rate, and can be

played on existing CD players, although generally only at quality levels similar to those of traditional CDs. SACD uses Direct Stream Digital ("DSD") recording, which is published as being proprietary to Sony, that converts an analog waveform to a 1-bit signal for direct recording, instead of the pulse code modulation ("PCM") and filtering used by standard CDs. DSD uses lossless compression and a sampling rate of about 2.8 MHz to improve the complexity and realism of sound. SACD may also include additional information, such as text, graphics, and video clips.

Also for the purpose of further understanding, Internet-based communications also have particular protocols for communication that must be accommodated by a streaming media communications system using the Internet "superhighway". These protocols, in particular with respect to streaming media communication, are briefly summarized immediately below for the purpose of providing a more detailed understanding.

With respect to Internet communication, streaming media signals are generally communicated in digital format via data packets. The terms "packets" are herein intended to mean units of data that are routed between an origin and a destination via the Internet or any other packet-switched network. More specifically, when a file is sent, the protocol layer of the communications system (e.g. TCP layer of TCP/IP based system) divides the file into chunks of efficient size for routing. Each of these packets is separately numbered and includes the Internet address of the destination. The individual packets for a given file may travel different routs through the Internet. When they have all arrived, they are reassembled into the original file, for example by the TCP layer at the receiving end. A packet-switching scheme is an efficient way to handle transmissions on a connectionless network such as the Internet. An alternative scheme, circuit-switched, is used for networks allocated generally for voice connections. In circuit-switching, lines in the network are shared among many users as with packet-switching, but each connection generally requires the dedication of a particular path for the duration of the connection.

Wireless Communications & the WAP Gateway

Of equal importance to the contemporary age of the Internet, the age of wireless communications has significantly extended society's ability to interact outside of the fixed confines of the home and office, allowing our remote communications to break free from the umbilical cords of wires and cables. For example, in 2000, the number of mobile subscribers grew by close to 50%.

However, wireless communications systems, protocols, and enabling technologies have developed in a significantly fragmented, "format-specific" market on a world-wide scale. This is particularly true in comparing systems in wide use in the United States as compared to the rest of the world. Therefore, much effort has been expended in overcoming compatibility issues between format-specific systems and between the related wireless devices operating on different platforms. For the purpose of further understanding wireless communication as it is later related to the present invention, the following is a brief overview of significant technologies, systems, and protocols used in the wireless communication industry.

In general, the progression of wireless communications systems for cellular telephones is colloquially given the terms "1G", "2G", "2.5G", and "3G", representing respectively first generation, second generation, and so-on. Initial systems were purely analog, known as the 1G phones and systems. However, with rapid growth, available bandwidth for cellular phone use quickly eroded, giving way to digital signal processing in the 2G, which significantly widened the available bandwidth and ability for complex signal processing for advanced telecommunications. However, as demand progressed for wireless Internet access, so went the technology development from 2G phones (generally not Internet enabled), to 2.5G and 3G (progressively more enabled). As will be further developed immediately below, the systems, protocols, and enabling technologies thus have developed toward a concentrated focus in bringing the 2.5G and 3G modes to industry and consumers.

In general, there are four major digital wireless networks based upon 2G technology: time division multiple access ("TDMA"), code division multiple access ("CDMA"), Global System for Mobile communication ("GSM") and cellular digital packet data ("CDPD"). These are briefly herein described as follows.

Time division multiple access ("TDMA") is a technology used in digital cellular telephone communication that divides each cellular channel into three time slots in order to increase the amount of data that can be carried. TDMA is used by Digital-American Mobile Phone Service (D-AMPS), Global System for Mobile communication ("GSM"), and Personal Digital Cellular ("PDC). However, each of these systems implements TDMA in a somewhat different and incompatible way. An alternative multiplexing scheme to TDMA and FDMA (frequency division multiple access) is code division multiple access ("CDMA").

Code division multiple access ("CDMA") refers to any of several protocols used in 2G and 3G wireless communications. As the term implies, CDMA is a form of multiplexing that allows numerous signals to occupy a single transmission channel, optimizing the use of available bandwidth. the

technology is used in ultra-high-frequency (UHF) cellular telephone systems in the 800 MHz to 1.9 GHz bands. CDMA uses analog-to-digital conversion (ADC) in combination with spread spectrum technology. Audio input is first digitized into binary elements. The frequency of the transmitted signal is then made to vary according to a defined pattern (code), so it can be intercepted only by a receiver whose frequency response is programmed with the same code, so it follows exactly along with the transmitter frequency. There are trillions of possible frequency-sequencing codes, thus enhancing privacy and making cloning difficult. The CDMA channel is nominally 1.23 MHz wide. CDMA networks use a scheme called "soft handoff", which minimizes signal breakup as a handset passes from one cell to another. The combination of digital and spread spectrum modes supports several times as many signals per unit bandwidth as analog modes. CDMA is compatible with other cellular technologies; this allows for nationwide roaming.

The original CDMA, also known as CDMA One, was standardized in 1993 and is considered a 2G technology that is still common in cellular telephones in the U.S. One version of cdmaOne, IS-95A, is a protocol that employs a 1.25 MHz carrier and operates in RF bands at either 800 MHz or 1.9 GHz; this supports data rates of up to 14.4 Kbps. Another version, IS-95B, is capable of supporting speeds of up to 115 Kbps by bundling up to eight channels.

More recent CDMA varieties, CDMA2000 and wideband CDMA offer data speeds many times faster. CDMA2000, also known as IMT-CDMA Multi-Carrier or IS-136, is a CDMA version of the IMT-2000 standard developed by the International Telecommunications Union (ITU). The CDMA2000 standard is a 3G technology that is intended to support data communications at speeds ranging from 144Kbps to 2 Mbps. Companies that have developed versions of this standard include Ericsson and Qualcomm corporations. Wideband CDMA, or "WCDMA", is an ITU standard derived from CDMA that is also known as IMT-2000 direct spread. WCDMA is a 3G technology intended to support data rates of up to 2Mbps for local area access, or 384 Kbps for wide area access, and supports mobile/portable voice, images, data, and video communications at these speeds. WCDMA digitizes input signals and transmits the digitized output in coded, spread-spectrum mode over a 5 MHz wide carrier – a much broader range than the 200 KHz wide narrowband CDMA.

The Global System for Mobile communication ("GSM") is a digital mobile telephone system that is widely used in Europe and other parts of the world; this system uses a variation of "TDMA" (introduced immediately below) and is the most widely used of the three digital wireless telephone technologies (TDMA, GSM, and CDMA). GSM digitizes, compresses, and then sends data down a channel with two other streams of user data, each in its own time slot. It operates at either the 900 Mhz

or 1800 MHz frequency band. At the time of this disclosure, GSM is generally considered the wireless telephone standard in Europe, and has been published to have over 120 million users worldwide and is available in 120 countries. At least one company in the United States, American Personal Communications (SprintTM subsidiary), is using GSM as the technology for a broadband personal communications services ("PCS"). PCS are telecommunications services that bundle voice communications, numeric and text messaging, voice-mail and various other features into one device, service contract and bill. PCS are most often carried over digital cellular links. This service is planned to have more than 400 base stations for various compact mobile handsets that are being made by manufacturers such as Ericsson, Motorola, and Nokia corporations; these devices generally include a phone, text pager, and answering machine. GSM is part of an evolution of wireless mobile telecommunications that includes High-Speed Circuit-Switched Data (HCSD), General Packet Radio System (GPRS), Enhanced Data GSM Environment (EDGE), and Universal Mobile Telecommunications Service (UMTS).

Cellular Digital Packet Data ("CDPD") is a wireless standard providing two-way, 19.2 kbps packet data transmission over existing cellular telephone channels.

Several different protocols have also been put into use for communicating over the various wireless networks. Various specific such protocols are briefly introduced as follows.

"X.25" is a packet-based protocol, principally used at the time of this disclosure in Europe and adapted as a standard by the Consultative Committee for International telegraph and Telephone (CCITT). X.25 is a commonly used network protocol that allows computers on different public networks (e.g. CompuServe, Tymnet, or TCP/IP network) to communicate through an intermediary computer at the network layer level. X.25's protocols correspond closely to the data-link and physical-layer protocols defined in the Open Systems Interconnection ("OSI").

"OSI" is a model of network architecture and a suite of protocols (a protocol stack) to implement it, developed by ISO in 1978 as a framework for international standards in heterogeneous computer network architecture. The OSI architecture is split between seven layers, from lowest to highest: (1) physical layer; (2) data link layer; (3) network layer; (4) transport layer; (5) session layer; (6) presentation layer; and (7) application layer. Each layer uses the layer immediately below it and provides a service to the layer above. In some implementations, a layer may itself be composed of sub-layers.

General Packet Radio Services ("GPRS") is a packet-based wireless communication service that promises data rates from 56 to 114 Kbps and continuous connection to the Internet for mobile phone

Aylo Exhibit 1018
Page 2037 of 2308

and computer users. The higher data rates will allow users to take part in video conferences and interact with multimedia Web sites and similar applications using mobile handheld devices as well as notebook computers. GPRS is based on Global System for Mobile ("GSM") communication and will complement existing services such as circuit-switched cellular phone connections and the Short Message Service ("SMS"). SMS is a message service offered by the GSM digital cellular telephone system. Using SMS, a short alphanumeric message (160 alphanumeric characters) can be sent to a mobile phone to be displayed there, much like in an alphanumeric pager system. The message is buffered by the GSM network until the phone becomes active.

The packet-based service of GPRS is publicized to cost users less than circuit-switched services since communication channels are being used on a shared-use, as-packets-are-needed basis rather than dedicated only to one user at a time. It is also intended to make applications available to mobile users because the faster data rate means that middleware currently needed to adapt applications to the slower speed of wireless systems will no longer be needed. As GPRS becomes widely available, mobile users of a virtual private network ("VPN") will be able to access the private network continuously rather than through a dial-up connection. GPRS is also intended to complement "Bluetooth", a standard for replacing wired connections between devices with wireless radio connections. In addition to the Internet Protocol ("IP"), GPRS supports X.25 protocol. GPRS is also believed to be an evolutionary step toward Enhanced Data GSM Environment ("EDGE") and Universal Mobile Telephone Service ("UMTS").

Universal Mobile Telecommunications Service ("UMTS") is intended to be a 3G, broadband, packet-based transmission of text, digitized voice, video, and multimedia at data rates up to 2 Mbps. UMTS is also intended to offer a consistent set of services to mobile computer and phone users no matter where they are located in the world. This service is based upon the GSM communication standard, and is endorsed by major standards bodies and manufacturers, and is the planned standard for mobile users around the world by 2002. Once UMTS is fully implemented, computer and phone users can be constantly attached to the Internet as they travel.

Enhanced digital GSM enterprise ("EDGE")service is a faster version of the Global System for Mobile (GSM) wireless service, designed to deliver data at rates up to 384 Kbps and enable the delivery of multimedia and other broadband applications to mobile phone and computer users. The EDGE standard is built on the existing GSM standard, using the same time-division multiple access (TDMA) frame structure and existing cell arrangements. EDGE is expected to be commercially

available in 2001. It is regarded as an evolutionary standard on the way to Universal Mobile Telecommunications Service (UMTS).

Wireless Application Protocol ("WAP") is a specification for a set of communication protocols to standardize the way that wireless devices such as cellular telephones and radio transceivers, can be used for Internet access, including e-mail, the World Wide Web, newsgroups, and Internet Relay Chat ("IRC"). While Internet access has been possible prior to WAP, different manufactures have used "format-specific" technologies. WAP enables devices and service systems to intercooperate.

In most recent times, much effort has been expended to merge the fields of wireless communications and the Internet in order to bridge the gap of cords, wires, and cables that had before separated the "information superhighway" from reaching people on wireless devices. Such technology merger has developed for example within the home and office network setting itself, where wireless infrared and radio frequency communications systems have been developed for interfacing equipment within a "wireless" office or home. Another substantial effort has also been underway to communicate and share information with more remote wireless devices, such as cell phones and personal digital assistants ("PDA's").

PDA's are typically small, mobile devices that may be "hand-held" and usually contain limited processors and display screens for managing, storing, and displaying telephone books, calendars, calculator, and the like. Recently available PDA's have been made "wireless enabled", either by having wireless modems embedded within the PDA itself, or by coupling to wireless modem "plug-ins" such as a cell phone. Wireless enabled PDA's are also generally "Internet enabled" with limited "browser" capability allowing the PDA to communicate with server devices over the Internet. Examples of commercially available wireless "enabled" PDA's include the Palm VII (from Palm, Inc.), and the iPAQ™ (from Compaq, Inc.). These PDA's include a Windows CE™ operating system that provides the limited browser capability and screen display for content. These phones have processing capabilities from about 33 MHz to about 220 MHz and varied screen display capabilities, such as for example 320 x 240 pixel screen displays.

Similarly, cellular phones themselves have also been recently rendered "Internet enabled", also with limited browser capability and screens to display content. Examples of "Internet enabled" cellular phones include, for example: Sanyo SCP-4000™, Motorola *i1000plus*™, among a wide range of others; this wide field represents hundreds of different processing and display capabilities.

In either the case of the PDA or the cellular phone that is "Internet-enabled", compatibility with the Internet protocols of communication must be achieved. In general, wireless communications take

place over a wireless applications protocol ("WAP"), whereas communications over the Internet proceed according to one of several different protocols, the most common being Transmission Control Protocol/Internet Protocol ("TCP/IP"). Therefore, a WAP Gateway, as shown in Figure 1E, forms a bridge between the world of the Internet (or any other IP packet network) and the wireless phone/data network, which are fundamentally different in their underlying technologies. The gateway, in essence, does the interpretation between these two distinct entities, allowing the consumer to use their cell phone or hand held computing device (e.g. PDA) to access the Internet wirelessly.

However, streaming media that is formatted for transmission to higher power computing devices such as desk-top computers having significant display capabilities is not generally compatible for receipt and viewing on these devices that have severely limited processing and display functionality. Particular "format-specific" compression schemes have been developed for use specifically with only these devices, and only specific media content may be transmitted to these devices in those formats.

There is still a need for a streaming media communications system that is adapted to transmit a wide variety of streaming media signals in appropriate formats to be played by wireless devices such as cellular phones and PDA's having unique constraints, such as, for example, limited and variable processing, memory, and display capabilities.

SUMMARY OF THE INVENTION

The present invention addresses and overcomes the various limitations, inefficiencies, resource limitations, and incompatibilities of prior known methods for streaming media communication, and is provided in various beneficial modes, aspects, embodiments, and variations as follows.

The present invention according to one embodiment is a streaming media communications system that uses a computer implemented intelligence system, such as artificial intelligence, in a network system, such as a neural network, to communicate a streaming media signal between a transmission device and at least one destination device.

The present invention according to another embodiment is a system for communicating a streaming media signal between a transmission device and a plurality of destination devices each having different media signal processing capabilities.

The present invention according to another embodiment is a streaming media communications system that is adapted to communicate a streaming media signal from a single transmission device and

at least one destination device over a plurality of different transmission channels, each having a different transmission capability or constraints with respect to communicating the streaming media signal.

The invention according to another embodiment is a neural network incorporating an artificial intelligence implementation that is adapted to be trained in an adaptive learning process with respect to the ability of a streaming media compression system's ability to compress a streaming media signal at a source into a compressed representation of the streaming media signal, transmit the compressed representation over a transmission channel to a destination device, and decompress the compressed representation into a decompressed representation of the streaming media signal which is adapted to be played by the destination device.

The invention according to another embodiment is a system for compressing streaming media signals according to a CODEC that is used at least in part based upon at least one parameter affecting communication of the streaming media signals. According to one aspect of this mode, the CODEC is used according to at least one of the following parameters: a previously learned behavior of the CODEC with respect to another reference signal, a previously learned behavior of the CODEC with respect to a prior attempt at compressing or decompressing the same streaming media signal, a comparison of the CODEC's operation with respect to the streaming media signal against a reference algorithm compression of the streaming media signal, a learned constraint of the transmission channel, and a learned constraint of the destination device. In one beneficial embodiment, the CODEC is used based upon more than one of these parameters, and in still a further beneficial variation is used based upon all of the parameters.

The invention according to another embodiment is a system for compressing streaming media signals using a CODEC library that is adapted to store multiple CODECS of different types and operations, and that is adapted to be searched and accessed by a network system, such as a neural network, in order to provide an appropriate CODEC from the CODEC library for use in compressing the input streaming media signal into a compressed representation for transmission to a destination device.

The invention according to another embodiment is a CODEC operating system that is adapted to interface with a CODEC Library and also with a neural network in order to use the neural network in a process, such as an artificial intelligence process, to choose an appropriate CODEC from the CODEC library and use the chosen CODEC for compressing the streaming media signal into a compressed representation of the streaming media signal for transmission to a destination device.