Attorney's Docket No: 25832.1369 (L0700C3)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In Re Application of:

Carbunar et al.

Application No.: 15/179,765

Filed: June 10, 2016

For: DISTRIBUTED CONTENT MANAGEMENT

EXAMINER: MONTOYA, OSCHTA I.

ART UNIT: 2421

CONFIRMATION NO.: 3643

E-Filed via EFS-Web
Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

CERTIFICATE OF TRANSMISSION
I hereby certify that this correspondence is being submitted electronically via EFS-Web on the date shown below.

/Hang Gao/ January 30, 2018
Hang Gao Date

APPEAL BRIEF

Pursuant to 37 C.F.R. § 41.37, Appellant submits the following Appeal Brief for consideration by the Board of Patent Appeals and Interferences (hereinafter "Board"). Appellant submits there are no fees due at this time of filing a brief in support of an appeal, as set forth in 37 C.F.R. § 41.20(b)(2)(i). Please charge any additional amounts due or credit any overpayment to deposit account no. 50-1358.

TABLE OF CONTENTS

I. REAL PARTY IN INTEREST

The real party in interest is the assignee of the full interest of the invention, Google Inc. located at 1600 Amphitheatre Parkway, Mountain View, CA, 94043.

II. RELATED APPEALS AND INTERFERENCES

To the best of Appellant's knowledge, there are no appeals or interferences related to the present appeal that will directly affect, be directly affected by, or have a bearing on the Board's decision in the instant appeal.

III. STATUS OF CLAIMS

Claims 1-20 are currently pending in the above-referenced application. Claims 1-9, 11, 16, 17, 19, and 20 were rejected in the Final Office Action mailed on September 20, 2017 (hereinafter "the Final Office Action"), and are presented for appeal. Claims 10, 12-15, and 18 were objected to for containing allowable subject matter and would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claim. No claims are canceled. A copy of the claims as they stand on appeal is set forth in the Claims Appendix.

IV. STATUS OF AMENDMENTS

No amendments to the application were submitted after final rejection.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

In an exemplary implementation of independent claim 1, a method includes receiving, by a processing apparatus at a first content source, a request for content (Specification, paragraphs [0017]-[0018] and [0074]; Figure 5, block 502; Figure 6, block 601). The method further includes in response to receiving the request, determining that the content is not available from the first content source (Specification, paragraphs [0017]-[0018] and [0076]; Figure 5, block

507; Figure 6, block 603). The method further includes in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source (Specification, paragraphs [0017]-[0018], [0052]-[0069], and [0074]-[0076]; Figure 6, block 511; Figure 6, block 607). The method further includes in response to determining that the second content source cost is less than the third content source cost, fetching the content from the second content source to the first content source, wherein the first content source, the second content source, and the third content source each maintain a different subset of content available from a master content source (Specification, paragraphs [0017]-[0018], [0075], and [0076]; Figure 6, block 505; Figure 6, block 607).

In an exemplary implementation of independent claim 9, a non-transitory computer-readable medium comprising instructions that, when executed by a processing apparatus, cause the processing apparatus to receive, by the processing apparatus, a request for content to be delivered from a first content source (Specification, paragraphs [0014], [0017]-[0018] and [0074]; Figure 2; Figure 5, block 502; Figure 6, block 60). The processing apparatus is further to: determine that the content is not available from the first content source in response to the receipt of the request (Specification, paragraphs [0017]-[0018] and [0076]; Figure 5, block 507; Figure 6, block 603). The processing apparatus is further to: in response to the determination that the content is not available from the first content source, determine that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source (Specification, paragraphs [0017]-[0018], [0052]-[0069], and [0074]-[0076]; Figure 6, block 511; Figure 6, block 607). The processing apparatus is further to: in response to the

determination that the second content source cost is less than the third content source cost, fetch the content from the second content source to the first content source, wherein the first content source, the second content source, and the third content source each maintain a different subset of content available from a master content source (Specification, paragraphs [0017]-[0018], [0075], and [0076]; Figure 6, block 505; Figure 6, block 607).

In an exemplary implementation of independent claim 17, a system includes a memory and a processing apparatus, coupled to the storage, configured to execute instructions to receive a request for content from a client device (Specification, paragraphs [0017]-[0018] and [0074]; Figure 2, Figure 5, block 502; Figure 6, block 601). The processing apparatus is further to: in response to the receipt of the request, determine that the content is not stored by the storage (Specification, paragraphs [0017]-[0018] and [0076]; Figure 5, block 507; Figure 6, block 603). The processing apparatus is further to: in response to the determination that the content is not stored by the storage, determine that a second server cost associated with retrieving the content from a second server is less than a third server cost associated with retrieving the content from a third server, wherein the second server cost is determined based on a network impact to fetch the content from the second serve to the first server, and wherein the third server cost is determined based on a network impact to fetch the content from the third server to the first server (Specification, paragraphs [0017]-[0018], [0052]-[0069], and [0074]-[0076]; Figure 6, block 511; Figure 6, block 607). The processing apparatus is further to: in response to the determination that the second server cost is less than the third server cost, fetch, by the first server, the content from the second server, wherein the first server, the second server, and the third server each maintain a different subset of content available from a master server (Specification, paragraphs [0017]-[0018], [0075], and [0076]; Figure 6, block 505; Figure 6, block 607).

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

The issues involved in this Appeal are as follows:

A. Whether claims 1, 3-4, 8-9, 11, 16-17, and 19 are unpatentable under 35 U.S.C. § 103(a) over U.S. Patent Publication No. 2004/0143850 to Costa et al. (hereinafter "Costa") in view of U.S. Patent No. 8,028,319 to Scholl et al. (hereinafter "Scholl").

B. Whether claim 2 is unpatentable under 35 U.S.C. § 103(a) over Costa and Scholl and further in view of U.S. Patent Publication No. 2004/0103437 to Allegrezza et al. (hereinafter "Allegrezza") and further in view of .U.S. Patent No. 8,087,056 to Ryu (hereinafter "Ryu").

C. Whether claims 5 and 14 are unpatentable under 35 U.S.C. § 103(a) over Costa and Scholl and further in view of U.S. Patent Publication No. 2011/0107030 to Borst et al. (hereinafter "Borst").

VII. ARGUMENT

A. Claims 1, 3-4, 8-9, 11, 16-17, and 19 Are Patentable Over Costa and Scholl

1. Claims 1, 9, and 17 and corresponding dependent claims are patentable over Costa and Scholl because Costa and Scholl fail to teach or suggest every limitation of these claims.

Claim 1 recites:

> A method comprising:
> receiving, by a processing apparatus at a first content source, a request for content;
> in response to receiving the request, determining that the content is not available from the first content source;
> **in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is**

determined based on a network impact to fetch the content from the third content source to the first content source; and

in response to determining that the second content source cost is less than the third content source cost, fetching the content from the second content source to the first content source, wherein the first content source, the second content source, and the third content source each maintain a different subset of content available from a master content source. (Emphasis added.)

Appellant respectfully submits that the cited references do not teach or suggest at least the above-emphasized features as recited by claim 1. Appellant submits that independent claim 1 is a representative claim of the group of claims 1, 3-4, 8-9, 11, 16-17, and 19 and that claims 1, 3-4, 8-9, 11, 16-17, and 19 stand or fall together with respect to independent claim 1.

The Final Office Action asserts that Costa teaches the recitation of "in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source." (Final Office Action, page 3.) In particular, the Final Office Action refers to Figure 11 and paragraphs 85, 92-98, and 100-103 of Costa as allegedly teaching the above-quoted features of claim 1. Figure 11 of Costa is directed to "a block diagram of an embodiment of the video distribution architecture." (Costa, paragraph [0085].) In the relied upon portions of Costa, Costa states:

> A database manager 356 manages the storage of the main content 352 and the entries in the main program list 354. The database manager 356 also serves to determine how to distribute videos to the central offices to: (a) reduce, or ideally minimize, content redundancy; (b) reduce, or ideally minimize, a link distance between a serving point and a customer; and (c) reduce, or ideally eliminate, network congestion in the communication network 226 and communication medium 335. The database manager 356 distributes videos based on popularity, demographics, geography, and a random number generator.
>
> As indicated by block 360, the method comprises providing all central offices with a statistical mix of video content programs. As indicated by block 362, 364 and 366, this act includes distributing a first set, a second set, and a third set of videos from

the video servers 204 and 206 to the central offices for local storage at the central offices. The herein-disclosed teachings for communicating multiple video data streams without congestion may be applied in the acts indicated by blocks 362, 364 and 366.

Each of the first set of videos is distributed to at least one but not all of the central offices. Each of the second set of videos is distributed to all of the central offices. Each of the third set of videos is distributed to at least one of the central offices.

The second set of videos comprises those videos (e.g. movies, television programs, advertisements) that are popular or anticipated to be popular in all serving areas (i.e. those that will have a relatively high order rate from each central office). **Thus, each of the second set of videos is distributed to all of the central offices for storage locally.** For purposes of illustration and example, the second set comprises Video1, Video4, and Video10.

The third set of videos comprises those videos (e.g. movies, television programs, targeted advertisements) that are popular or anticipated to be popular in one or more serving areas, but not all serving areas. **The videos in the third set that are distributed to a particular central office may be based on at least one demographic of viewers served by the particular central office. Examples of the at least one demographic include, but are not limited to, income, ethnicity, age, and gender. Thus, each central office may locally store videos/advertisements appropriately tailored and/or relevant to the viewers in its neighborhood.** For purposes of illustration and example, the third set comprises Video2 and Video3 that are relevant to customers of the central office 224, and Video5 and Video6 that are relevant to customers of the central office 326.

The first set of videos comprises videos (e.g. movies and television programs) that are not as popular or anticipated to be as popular as those in the second and third sets. **The videos in the first set are randomly distributed for storage by the central offices.** Each central office is randomly allocated a subset of the videos in the first set. The number of videos allocated to a central office may be commensurate with storage availability in its mass storage device, and an overall rate of orders placed by its customers. For purposes of illustration and example, the first set comprises Video7, Video8 and Video12. The central office 326 is randomly allocated Video7 and Video12. The central office 224 is randomly allocated Video8.

Optionally, all of the videos in the main content 352 may be distributed in block 360. Alternatively, the least popular of the videos in the main content 352 may not be distributed to any of the central offices. For purposes of

> illustration and example, Video9 and Video11 in the main content 352 are not distributed for local storage by any of the central offices. (Costa, paragraphs [0092]-[0098].)

Thus, Costa teaches *randomly* distributing a first set of videos for storage, distributing a second set of videos to all of the central offices for storage locally, and distributing a third set of videos to a particular central office based on at least one demographic of viewers served by the particular central office.

Costco further states:

> As indicated by block 372, the method comprises receiving an on-demand order for a selected video. The on-demand order is placed by a customer of one of the central offices. The on-demand order may be received by the central office from the customer via a digital subscriber line. For purposes of illustration and example, consider the order being placed by the customer CPE1,1 of the central office 224.
>
> As indicated by block 374, the method comprises determining where the central office can access the selected video. The selected video is accessible from either the central office's mass storage device, another central office, or the video servers 204 and 206**. If the selected video is stored locally, the selected video is retrieved from the local mass storage device and communicated to the customer (block 376). If the selected video is stored by another central office linked to the central office, the selected video is downloaded from the other central office via the communication medium 335, and communicated to the customer (block 380). Otherwise, the selected video is downloaded from one of the video servers 204 and 206 via the communication network 226, and communicated to the customer (block 382).**
>
> The herein-disclosed teachings for communicating multiple video data streams without congestion may be applied to the acts in blocks 380 and 382 to ensure that the selected video is downloaded without congestion in the communication network 226 and the communication medium 335, respectively. Thus, the video-on-demand order may be inhibited if the selected video cannot be downloaded without congestion. Optionally, if the selected video is accessible at another central office, and a congestion-free download of the selected video from the other central office cannot be ensured, the selected video may be downloaded from one of the video servers 204 and 206.
>
> Returning to the above example, if the customer CPE1,1, orders any of Video1, Video2, Video3, Video4, Video8 or Video10, the central office 224 retrieves the video(s) from the

> mass storage device 236 and communicates the video(s) to the customer. If the customer CPE1,1 orders any of Video5, Video6, Video7 or Video12, the central office 224 downloads the video(s) from the central office 326 (which retrieves the video(s) from its mass storage device 346) and communicates the video(s) to the customer. If the customer CPE1,1 orders any of Video9 or Video11, the central office 224 downloads the video(s) from one of the video servers 204 and 206 and communicates the video(s) to the customer. (Costa, paragraphs [0100]-[0103]; emphasis added.)

As shown above, in Costa, when a selected video is not stored by another central office, the selected video is downloaded from one of the video servers 204 and 206. Costa is silent as to any determination that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, much less making such a determination "in response to determining that the content is not available from the first content source," as recited in claim 1.

Accordingly, Costa neither teaches nor suggests "in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source," as recited by claim 1.

Further, as acknowledged by the Final Office Action, "Costa is silent about in response to determining that the second content source cost is less than the third content source cost, fetching, to the first content source." (Final Office Action, page 4.) However, the Final Office Action cites Scholl for such teachings. (Final Office Action, pages 2 and 4.)

Appellant respectfully disagrees and submits that Scholl does not cure the deficiencies of Costa. Scholl is directed to "[t]echniques for delivering video on demand (VOD) channel content via streaming from a first edge aggregation device." (Scholl, abstract.) The Final Office Action states as follows:

> Applicant argues that Scholl does not teach "in response to determining that the second content source cost is less than the third content source cost, fetching the content from the second

> content source to the first content source". To this matter the examiner respectfully disagrees. Scholl teaches that [sic] an index is search[ed] for viewers watching the content and then ***according to some rules a determination is made***. Finally the content is routed based on the determination (steps 230-248 – figure 2), meeting the claim language. (Final Office Action, page 2; emphasis added.)

In the Advisory Action of December 18, 2017, that was mailed as a result of the Appellant's response to the Final Office Action, the Office asserted that "Scholl clearly teaches that a determination is made (steps 230-248 figure 2)." (Advisory Action, page 2). The Office further asserted that "Scholl teaches that [sic] an index is search for viewers watching the content and then according to some rules a determination is made" and that "[f]inally the content is routed based on the determination (steps 230-248 – figure 2), meeting the claim language." *Id.*

However, making a determination "*according to some rules*" does not teach or suggest "in response to determining that the second content source cost is less than the third content source cost, fetching the content from the second content source to the first content source," as recited in claim 1. In the portions of Scholl referred to in the Final Office Action and the Advisory Action, Scholl describes a method including the steps of "search[ing] an index for another viewer having the VOD channel content," "instruct[ing] the edge aggregation device to forward the VOD channel content to the requestor," determining that "the requestor is below the edge aggregation device," determining that "the other viewer and the requestor are below a first metro aggregation switch," determining that "the other viewer is below a first metro aggregation switch and the requestor is below a second metro aggregation switch," and "subsequently rerout[ing] the VOD channel content to the requestor." (Scholl, FIG. 2, steps 230-248).

Appellant respectfully submits that Scholl neither teaches nor suggests at least the features of "in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source" and "in response to determining that the second content source cost is less than the third content source cost, fetching

the content from the second content source to the first content source," as recited by claim 1.

Accordingly, Costa and Scholl, alone or in hypothetical combination, fail to teach or suggest all recitations of independent claim 1. Similar, although not identical, language is also included in independent claims 9 and 17. For reasons similar to those given above in regards to claim 1, Appellant respectfully submits that independent claims 9 and 17 and their dependent claims are also not unpatentable over Costa and Scholl. Thus, Appellant respectfully requests the Board reverse the rejection of claims 1, 3-4, 8-9, 11, 16-17, and 19 under 35 U.S.C. § 103(a).

B. Claim 2 Is Patentable Over Costa, Scholl, Allegrezza, and Ryu

1. Claim 2 is patentable over Costa, Scholl, Allegrezza, and Ryu because Costa, Scholl, Allegrezza, and Ryu fail to teach or suggest every limitation of this claim.

Claim 2 recites, *inter alia*:

> determining that there is not sufficient memory to cache the content at the first content source; and
>
> selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items.

Appellant respectfully submits that the cited references do not teach or suggest the features of "determining that there is not sufficient memory to cache the content at the first content source" and "selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items," as recited by claim 2.

The Office Action acknowledges that "Costa and Scholl are silent about determining that there is not sufficient memory to cache the content at the first content source; and selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty

associated with the eviction, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items." (Final Office Action, pages 7-8). However, the Final Office Action refers to paragraphs [0056]-[0057] and [0067] of Allegrezza as teaching "determining that there is not sufficient memory to cache the content at the first content source; and selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received from the content and the items". (*Id.*) The relied upon portion of Allegrezza states:

> **At decision 310, if the demand is less than a second threshold, Threshold B, then at step 340 the content file is moved to a more remote server so as to improve availability and response time for higher-ranked content files, and decrease the load on the inter-server communications links.** Again, as above, the system operator may specify that a file is not to be moved, or is not to be moved for a certain time, or until after a certain date, etc., to allow the system operator to position certain content file in expectation of future demand.
>
> **Therefore, when possible, higher demand files are moved closer to the customer, and lower demand files are moved further from the customer.**
>
> . . .
>
> It is generally desirable to manage the use of the various servers so that servers closest to the subscribers are completely filled with relevant content. By so doing, the response time to the subscriber will be optimal. In addition, if a subscriber requests a content, and there is a noticeable delay while the content is being downloaded from the content library, then the subscriber may be dissatisfied with the apparently slow response, may believe that this is the response that will be provided for the entire content duration, may reconsider the request, and may even cancel the request, thus causing a loss of revenue to the system operator. However, if the content is already on the server closest to the subscriber, then the content can be delivered immediately, thus reducing the likelihood that a subscriber will cancel, or will be able to cancel, the request. (Allegrezza, paragraphs [0056]-[0057] and [0067]; emphasis added.).

As shown above, Allegrezza, merely describes that "if the demand is less than a second threshold, … then at step 340 the content file is moved to *a more remote server*" which is **not** "determining that there is not sufficient memory to cache the content at the first content source"

as set forth in claim 2. Allegrezza does not teach or suggest the features of "determining that there is not sufficient memory to cache the content at the first content source" and "selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items," as recited by claim 2.

Appellant respectfully submits that Ryu does not cure the deficiencies of the combination of Costa, Scholl, and Allegrezza. Referring to col. 5 ll. 45-55 of Ryu, the Final Office Action asserts that "Ryu discloses the network penalty is based on sizes of the content and the items." (Final Office Action, page 8.) In the relied upon portion of Ryu, Ryu states:

> Further, a reference data value is determined in consideration of the size of the data of content, and content stored in the COD server 10 is classified into high-capacity content and low-capacity content based on the reference data value, so that the high-capacity content, for which the transmission time is expected to be long, may be classified as content for one-to-N transmission, and the low-capacity content, for which the transmission time is expected to be short, may be classified as content for one-to-one transmission. Further, a reference bandwidth value is determined in consideration of the value of the total bandwidth of the COD server 10, content requested by the respective COD clients 20 (20 a, 20 b, 20 c, and 20 d) is classified as content for one-to-one transmission or content for one-to-N transmission in real time using the reference bandwidth value, and the value of currently available bandwidth is compared with the reference bandwidth value. (Ryu, col. 5 ll. 45-61.)

As shown above, Ryu merely mentions classifying stored content into high-capacity content and low-capacity content. Appellant respectfully submits that Ryu neither teaches nor suggests at least the features of "determining that there is not sufficient memory to cache the content at the first content source" and "selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items," as recited by claim 2.

Accordingly, the combination of Costa, Scholl, Allegrezza, and Ryu does not teach or

suggest the limitations recited by claim 2. As such, Appellant respectfully submits that claim 2 is not unpatentable over Costa, Scholl, Allegrezza, and Ryu. Additionally, Appellant submits that neither Allegrezza nor Ryu cures the deficiencies of the combination of Costa and Scholl with respect to independent claims 1, 9, and 17 as described above. Thus, Appellant respectfully requests the rejection of claim 2 under 35 U.S.C. §103(a) be reversed.

C. Claims 5 and 14 Are Patentable Over Costa, Scholl, and Borst

1. Claims 5 and 14 are patentable over Costa, Scholl, and Borst because Costa, Scholl, and Borst fail to teach or suggest every limitation of these claims.

Claim 5 recites:

> The method of claim 1, wherein the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source.

Appellant respectfully submits that the cited references do not teach or suggest "wherein the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source," as recited by claim 5.

The Office Action acknowledges that "Costa and Scholl are silent about the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source." (Final Office Action, page 8). However, the Final Office Action cites to paragraph [0018] of Borst as teaching the above-recited limitation. (*Id.*) The relied upon portion of Borst states:

> Various embodiments described herein provide relatively lightweight cooperative content placement algorithms adapted to maximize the traffic volume served from a cache and thus minimize the bandwidth cost. The bandwidth cost need not be actual monetary expenses, but could also represent some QoS metric reflecting the congestion levels on the various network links, like link weights in Open Shortest Path First (OSPF) for example. (Borst, paragraph [0018]).

As shown above, Borst merely mentions "some QoS metric reflecting the congestion levels on the various network linkes," which is **not** determining the second content source cost "based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source" or determining the third content source cost "based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source," as set forth in claim 5. Borst does not teach or suggest "wherein the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source," as recited by claim 5.

Accordingly, the combination of Costa, Scholl, and Borst does not teach or suggest the limitations recited by claim 5. As such, Appellant respectfully submits that claim 5 is not unpatentable over Costa, Scholl, and Borst. Claim 14 recites features similar to, although not identical, as to those of claim 5. As such, the combination of Costa, Scholl, and Borst does not teach or suggest the features of claim 14. Additionally, Appellant submits that Borst does not cure the deficiencies of the combination of Costa and Scholl with respect to independent claims 1, 9, and 17 as described above. Thus, Appellant respectfully requests the Board reverse the rejection of claims 5 and 14 under 35 U.S.C. §103(a) be reversed.

Based on the foregoing, Appellant respectfully submits that the Board should reverse the rejections of all pending claims and hold that all of the claims currently under review are allowable.

Respectfully submitted,

Dated: January 30, 2018

/Hang Gao/

Hang Gao
Reg. No. 63,093

VIII. CLAIMS APPENDIX

The claims involved in this appeal are presented below.

1. (Previously presented) A method comprising:

receiving, by a processing apparatus at a first content source, a request for content;

in response to receiving the request, determining that the content is not available from the first content source;

in response to determining that the content is not available from the first content source, determining that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source; and

in response to determining that the second content source cost is less than the third content source cost, fetching the content from the second content source to the first content source, wherein the first content source, the second content source, and the third content source each maintain a different subset of content available from a master content source.

2. (Previously presented) The method of claim 1, further comprising:

determining that there is not sufficient memory to cache the content at the first content source; and

selecting one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, and numbers of requests expected to be received for the content and the items.

3. (Previously presented) The method of claim 1, wherein the second content source cost is determined further based on traffic which is predicted to occur over a link in a network path from the second content source to the first content source while fetching of the content occurs, and wherein the third content source cost is determined further based on traffic which is predicted to occur over a link in a network path from the third content source to the first content source while fetching of the content occurs.

4. (Original) The method of claim 1, wherein the first content source comprises a first server, the second content source comprises a second server, the third content source comprises a third server, and the master content source comprises a master server.

5. (Previously presented) The method of claim 1, wherein the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source.

6. (Previously presented) The method of claim 1, wherein the second content source cost is determined further based on using historical traffic data to predict traffic for a target fetch time for the content over a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on using historical traffic data to predict traffic for a target fetch time for the content over a network path from the third content source to the first content source.

7. (Previously presented) The method of claim 1, wherein the second content source cost is determined further based on second predicted traffic for one or more specific time intervals during a day over a second network path from the second content source to the first content source, wherein the second predicted traffic is based on an analysis of repetitive traffic patterns over the second network path, wherein the third content source cost is determined further based on third predicted traffic for one or more specific time intervals during a day over a third

network path from the third content source to the first content source, and wherein the third predicted traffic is based on an analysis of repetitive traffic patterns over the third network path.

8. (Original) The method of claim 1, wherein the first content source is a first video home office (VHO), the second content source is a second VHO, and the third content source is a third VHO, wherein the master content source is a video service office (VSO).

9. (Previously presented) A non-transitory computer-readable medium comprising instructions that, when executed by a processing apparatus, cause the processing apparatus to:

receive, by the processing apparatus, a request for content to be delivered from a first content source;

in response to the receipt of the request, determine that the content is not available from the first content source;

in response to the determination that the content is not available from the first content source, determine that a second content source cost associated with retrieving the content from a second content source is less than a third content source cost associated with retrieving the content from a third content source, wherein the second content source cost is determined based on a network impact to fetch the content from the second content source to the first content source, and wherein the third content source cost is determined based on a network impact to fetch the content from the third content source to the first content source; and

in response to the determination that the second content source cost is less than the third content source cost, fetch the content from the second content source to the first content source, wherein the first content source, the second content source, and the third content source each maintain a different subset of content available from a master content source.

10. (Previously presented) The non-transitory computer-readable medium of claim 9, wherein the instructions are further to cause the processing apparatus to:

determine that there is not sufficient memory to cache the content at the first content source; and

select one or more items to evict from a cache at the first content source to make available sufficient memory for the content, wherein the selection of the items to evict minimizes a

network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, numbers of requests expected to be received for the content and the items, and fetch costs associated with retrieval of the content and the items, wherein each of the fetch costs is based on a sum of link weights of links in a network path for fetching each of the content and the items, and wherein each of the link weights is based on traffic predicted on a link in the links of the network path.

11. (Previously presented) The non-transitory computer-readable medium of claim 9, wherein the instructions are further to cause the processing apparatus to:

determine a stream cost associated with streaming the content from a content source other than the first content source to fulfill the request for the content; and

in response to a determination that the stream cost is less than a cost to cache the content to the first content source, stream the content from the second content source.

12. (Previously presented) The non-transitory computer-readable medium of claim 9, wherein the second content source cost is determined further based on traffic which is predicted to occur over a most utilized link in a network path from the second content source to the first content source while the fetch of the content occurs, and wherein the third content source cost is determined further based on traffic which is predicted to occur over a most utilized link in a network path from the third content source to the first content source while the fetch of the content occurs.

13. (Previously presented) The non-transitory computer-readable medium of claim 12, wherein the second content source cost is determined further based on historical traffic data in a network path from the second content source to the first content source, and wherein the third content source cost is determined further based on historical traffic data in a network path from the third content source to the first content source.

14. (Previously presented) The non-transitory computer-readable medium of claim 12, wherein the second content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the second content source to the first content

source, and wherein the third content source cost is determined further based on a number of items simultaneously transferred over a link in a network path from the third content source to the first content source.

15. (Previously presented) The non-transitory computer-readable medium of claim 12, wherein the second content source cost is determined further based on second predicted traffic for one or more specific time intervals during a day over a second network path from the second content source to the first content source, wherein the second predicted traffic is based on an analysis of repetitive traffic patterns over the second network path, wherein the third content source cost is determined further based on third predicted traffic for one or more specific time intervals during a day over a third network path from the third content source to the first content source, and wherein the third predicted traffic is based on an analysis of repetitive traffic patterns over the third network path.

16. (Original) The non-transitory computer-readable medium of claim 9, wherein the first content source is a first distributed storage component, the second content source is a second distributed storage component, and the third content source is a third distributed storage component, wherein the master content source is a central repository.

17. (Previously presented) A system for a first server comprising:

a storage; and

a processing apparatus, coupled to the storage, configured to execute instructions to:

receive a request for content from a client device;

in response to the receipt of the request, determine that the content is not stored by the storage;

in response to the determination that the content is not stored by the storage, determine that a second server cost associated with retrieving the content from a second server is less than a third server cost associated with retrieving the content from a third server, wherein the second server cost is determined based on a network impact to fetch the content from the second serve to the first server, and wherein the third server cost is determined based on a network impact to fetch the content from the third server to the first server; and

in response to the determination that the second server cost is less than the third server cost, fetch, by the first server, the content from the second server, wherein the first server, the second server, and the third server each maintain a different subset of content available from a master server.

18. (Previously presented) The system of claim 17, wherein the processing apparatus is further to execute the instructions to:

determine that the storage is insufficient to cache the content at the first server; and

select one or more items to evict from a cache at the first server to make available sufficient space on the storage for the content, wherein the selection of the items to evict minimizes a network penalty associated with the eviction of the items, wherein the network penalty is based on sizes of the content and the items, numbers of requests expected to be received for the content and the items, and fetch costs associated with retrieval of the content and the items, wherein each of the fetch costs is based on a sum of link weights of links in a network path for fetching each of the content and the items, and wherein each of the link weights is based on traffic predicted on a link in the links of the network path.

19. (Previously presented) The system of claim 17, wherein the processing apparatus is further to execute the instructions to:

determine a stream cost associated with streaming the content from a server other than the first server to fulfill the request for the content; and

in response to a determination that the stream cost is less than a cost to cache the content to the first server, stream the content to fulfill the request for the content.

20. (Previously presented) The system of claim 17, wherein the second server cost is determined further based on historical traffic data and predicted traffic over a network path from the second server to the first server, and wherein the third server cost is determined further based on historical traffic data and predicted traffic over a network path from the third server to the first server.