

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

CISCO SYSTEMS INC.,

Petitioner

IPR2024-00194

U.S. Patent No. 9,204,099

**DECLARATION OF A.L. NARASIMHA REDDY, PH.D.,
UNDER 37 C.F.R. § 1.68 IN SUPPORT OF PETITION
FOR *INTER PARTES* REVIEW**

TABLE OF CONTENTS

I.	INTRODUCTION.....	4
II.	QUALIFICATIONS AND PROFESSIONAL EXPERIENCE.....	6
III.	LEVEL OF ORDINARY SKILL IN THE ART	10
IV.	RELEVANT LEGAL STANDARDS	12
V.	TECHNICAL BACKGROUND	13
	A. Client-Server and Peer-to-Peer Networking Models.....	13
VI.	OVERVIEW OF THE '099 PATENT.....	14
VII.	CLAIM CONSTRUCTION.....	17
VIII.	IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE....	18
IX.	THE COMBINATION OF SETTON AND BAKER RENDERS OBVIOUS CLAIMS 3 AND 10	19
	A. Summary of Setton.....	19
	B. Summary of Baker.....	28
	C. Reasons to Combine Setton with Baker	33
	D. Detailed Analysis of Claims.....	36
	1. Claim 3.....	36
	2. Claim 10.....	73
X.	THE COMBINATION OF SETTON, BAKER, AND WONG RENDERS OBVIOUS CLAIMS 4-6.....	76
	A. Summary of Wong	76
	B. Reasons to Combine Wong with Setton and Baker.....	79
	C. Detailed Analysis of Claims.....	84
	1. Claim 4.....	84
	2. Claim 5.....	89
	3. Claim 6.....	91
XI.	THE COMBINATION OF APOSTOLOPOULOS AND BAKER RENDERS OBVIOUS CLAIMS 3 AND 10.....	95

A.	Summary of Apostolopoulos	95
B.	Reasons to Combine Apostolopoulos with Baker	103
C.	Detailed Analysis of Claims	106
1.	Claim 3	106
2.	Claim 10	134
XII.	THE COMBINATION OF APOSTOLOPOULOS, BAKER, AND WONG RENDERS OBVIOUS CLAIMS 4-6	138
A.	Reasons to Combine Wong with Apostolopoulos and Baker	138
B.	Detailed Analysis of Claims	141
1.	Claim 4	141
2.	Claim 5	144
3.	Claim 6	146
XIII.	CONCLUSION	148

I, Narasimha Reddy, Ph.D., do hereby declare as follows:

I. INTRODUCTION

1. I am making this declaration at the request of Cisco Systems, Inc. in the matter of the *Inter Partes* Review of U.S. Patent No. 9,204,099 (“the ’099 patent”) to Brown.

2. I am being compensated for my work in this matter at my standard hourly rate. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this proceeding. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

3. I have been asked to provide my opinions regarding whether the subject matter of claims 3-6 and 10 (“the Challenged Claims”) of the ’099 patent would have been obvious to a person having ordinary skill in the art (“POSITA”) at the time of the alleged invention, in light of the prior art. It is my opinion that the Challenged Claims would have been obvious to a POSITA.

4. In the preparation of this declaration, I have studied:

- a. the ’099 patent, Ex.1001;
- b. the prosecution history of the ’099 patent (“’099 File History”), Ex.1002;
- c. U.S. Patent Publication No. 2010/0287238 to Setton, et al. (“Setton”), Ex.1005;

- d. U.S. Patent No. 8,581,958 to Baker, et al. (“Baker”), Ex.1006;
- e. U.S. Patent Publication No. 2005/0080894 to Apostolopoulos, et al. (“Apostolopoulos”), Ex.1007; and
- f. U.S. Patent Publication No. 2009/0232129 to Wong, et al. (“Wong”), Ex.1009.

5. In forming the opinions expressed below, I have considered: the documents listed above; the relevant legal standards, including the standard for obviousness; and my own knowledge and experience based upon my work in the field of network communications and security as described below, as well as portions of the following additional materials:

- a. U.S. Patent No. 7,692,683 to Kenoyer et al. (“Kenoyer”), Ex.1008;
- b. U.S. Patent No. 8,601,097 to Sartori (“Sartori”), Ex.1010;
- c. U.S. Patent No. 9,124,757 to Weber (“Weber”), Ex.1011;
- d. Andrew S. Tanenbaum, “Computer Networks,” 4th ed. (2002) (“Tanenbaum”), Ex.1012;
- e. U.S. 9,009,225 to Kjesbu et al. (“Kjesbu”), Ex.1013;
- f. U.S. 7,877,489 to Salesky et al., (“Salesky”), Ex.1014;

g. Michael Arrington, “Facebook Will Launch In-Browser Video Chat Next Week In Partnership with Skype,” TechCrunch, Jul. 1, 2011 (*archived* July 4, 2011) (“TechCrunch”), Ex.1015;

h. Ian Paul, “Facebook Video Calling vs. Google+ Hangouts,” PCWorld.com, July 7, 2011 (*archived* August 7, 2011) (PCWorld), Ex.1016;

i. U.S. Pub. No. 2006/0215765 to Hwang et al. (“Hwang”), Ex.1017;

j. U.S. Pub. No. 2009/0015661 to King et al. (“King”), Ex.1018;

k. U.S. Pub. No. 2010/0157017 to Pepperell (“Pepperell”), Ex.1019; and

l. U.S. Pub. No. 2010/0333004 to Kristiansen et al. (“Kristiansen”), Ex.1020.

6. Unless otherwise noted, all **emphasis** in any quoted material has been added. Claim terms are *italicized*.

II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

7. My complete qualifications and professional experience are described in my *Curriculum Vitae*, a copy of which can be found in Ex.1004. The following is a brief summary of my relevant qualifications and professional experience.

8. I am currently the Truchard Foundation Chair Professor of Electrical and Computer Engineering at Texas A&M University in College Station, Texas. I

have over 25 years of experience in a wide variety of technologies and industries relating to data and network communications, distributed systems, and audio and video teleconferencing technologies.

9. My academic credentials include a Bachelor of Technology degree in Electronics and Electrical Communications Engineering from the Indian Institute of Technology, Kharagpur, India, in August 1985. I then received a Master of Science and a Ph.D. in Computer Engineering from the University of Illinois at Urbana-Champaign in May 1987 and August 1990, respectively.

10. I have worked for over 25 years in the field of Electrical Engineering. My primary focus and research interest has been on computer networks, storage systems, multimedia systems, and computer architecture. I have authored and co-authored over a hundred technical papers and several book chapters related to several of these interests, including on such topics as multipath routing, route control, high-speed networks, network congestion, packet management, quality of service regulation, network security, network modeling, differentiated services, storage system enhancements, caching strategies, and multimedia system enhancements to name a few examples. I am listed as an inventor on five U.S. patents in the field of communication networks.

11. My employment history following my graduation from the University of Illinois at Urbana-Champaign began at the IBM Almaden Research Center in

San Jose, California in 1990. At IBM, I worked on projects related to disk arrays, multiprocessor communication, hierarchical storage systems and video servers.

12. In 1995, I joined the faculty of the department of Electrical Engineering at Texas A&M University initially as an Associate Professor and was later promoted to full, tenured Professor. At Texas A&M, during 2016-21, I was the Associate Agency Director for Strategic Initiatives and Centers for the Texas A&M Engineering Experiment Station (TEES), which engages in engineering and technology-oriented research and educational collaborations. Further, I also served as Associate Dean for Research in the College of Engineering during 2016-21. Currently, I am the Truchard Foundation Chair Professor in Electrical and Computer Engineering.

13. At Texas A&M, I have taught dozens of courses related to computer networking and communications, as well as computer architecture, multimedia systems and networks, topics in networking security, multimedia storage and delivery, as well as networking for multimedia applications. I have done research on various topics of network security including anomaly detection, botnet detection, building mechanisms in routers and switches to improve security, spammer detection in social networks and improving security of cyberphysical systems. I have also served on various committees for the benefit of the scientific community and the Texas A&M University community.

14. I am a member of a number of professional societies, including the Institute of Electrical and Electronic Engineers (IEEE), where I have been elevated to an IEEE Fellow, and the Association for Computing Machinery (ACM). I have been responsible for chairing or co-chairing numerous conferences and programs, as well as presenting research at major IEEE and ACM conferences. For example, I served as program co-chair for the 2008 5th International Conference on Broadband Communications, Networks and Systems, panels co-chair for the 2008 3rd International Conference on Communication Systems Software & Middleware, and panel chair of the IEEE Conference of High Performance Computer Architecture.

15. My presentations include a Keynote speech at International Conference on Information Technology-New Generations in 2013, a Keynote speech at IEEE International Symposium on Computers and Communications 2010, and several invited talks including at Georgia Tech (2013), COMSNETS Conference (2013), Int. Conf. on Networking and Communications (2012), Samsung (2011), Korea University (2011), Aijou University (2011), Catedra Series at University of Carlos III, Madrid (2009), Thomson Research, Paris (2009), Telefonica Research, Barcelona (2009), and a Distinguished Seminar at IBM Austin Research Lab (2008).

16. I have received multiple awards in the field of networks and computer architecture. I received the NSF Career Award from 1996-2000. I received an outstanding professor award by the IEEE student branch at Texas A&M during 1997-1998, an outstanding faculty award by the department of Electrical and Computer Engineering during 2003-2004, a Distinguished Achievement award for teaching from the former students' association of Texas A&M University, and a citation "for one of the most influential papers from the 1st ACM Multimedia conference."

17. In my career in industry and in academia, I have been continuously researching and teaching multimedia courses that cover details of video servers, video conferencing, and multicasting for more than 25 years. These details include audio and video compression, internet-based software development, and user interface design.

III. LEVEL OF ORDINARY SKILL IN THE ART

18. I understand there are multiple factors relevant to determining the level of ordinary skill in the pertinent art, including (1) the levels of education and experience of persons working in the field at the time of the invention; (2) the sophistication of the technology; (3) the types of problems encountered in the field; and (4) the prior art solutions to those problems.

19. A POSITA in the field of the '099 patent, as of its earliest alleged priority date of February 1, 2012, would have been someone knowledgeable and familiar with video conferencing techniques and mixed topology networks in early 2012. A POSITA would have had a working knowledge of mesh networks, as well as star-based or tree-based networks, and hybrid networks that include both mesh- and non-mesh-connected nodes. Such a POSITA would have a bachelor's degree computer science, computer engineering, electrical engineering, or equivalent training, and approximately three years of experience working in the field of video teleconferencing systems. A POSITA would have had a working knowledge of audio and video compression, internet-based software development, and user interface design. Additional work experience can substitute for specific educational background, and vice versa.

20. For purposes of this Declaration, in general, and unless otherwise noted, my statements and opinions, such as those regarding my own experience and what a POSITA would have understood or known generally (and specifically related to the references I consulted herein), reflect the knowledge that existed in the relevant field as of the priority date of the '099 patent.

IV. RELEVANT LEGAL STANDARDS

21. I am not an attorney. In preparing and expressing my opinions and considering the subject matter of the '099 patent, I am relying on certain basic legal principles that Cisco's counsel has explained to me.

22. I understand that prior art to the '099 patent includes patents and printed publications in the relevant art that predate the priority date of the '099 patent. For purposes of this Declaration, I am applying February 1, 2012, as the priority date of the '099 patent.

23. I have been informed by Cisco's counsel that a claimed invention is unpatentable under 35 U.S.C. § 103 if the differences between the claimed invention and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a POSITA. I have also been informed by Cisco's counsel that the obviousness analysis considers factual inquiries, including the level of ordinary skill in the art, the scope and content of the prior art, and the differences between the prior art and the claimed subject matter.

24. I have been further informed by Cisco's counsel that there are several recognized rationales for combining references or modifying a reference to show obviousness. These rationales include: (a) combining prior art elements according to known methods to yield predictable results; (b) simple substitution of one

known element for another to obtain predictable results; (c) use of a known technique to improve a similar device (method, or product) in the same way; (d) applying a known technique to a known device (method, or product) ready for improvement to yield predictable results; (e) choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; and (f) some teaching, suggestion, or motivation in the prior art that would have led a POSITA to modify the prior art or to combine prior art teachings to arrive at the claimed invention.

V. TECHNICAL BACKGROUND

A. Client-Server and Peer-to-Peer Networking Models

25. Internet services are deployed using a variety of networking techniques. Two common networking models are the client-server model, and the peer-to-peer (P2P) model. *See* Tanenbaum, §1.1.1, “Business Applications,” pp. 4-5 (client-server model), §1.1.2, “Home Applications,” pp. 7-8 (P2P model). In the client-server model, a networked computer system, which may be local or remote, hosts content and/or provides computing services for clients to access over the Internet. The networked computer system may be a single server or a cluster of servers. A client uses software that sends requests to a network address of the server to retrieve files, perform computation, or both. Tanenbaum, §1.1.1, pp. 4-5.

26. The P2P model is a decentralized network architecture where each participant, or “peer,” acts as both a client and a server, sharing the burden of providing network resources such as computational power, data storage, and bandwidth. *See* Setton, [0015]. Unlike traditional client-server models where a centralized server is responsible for managing and distributing data and resources, in a P2P network, each peer can operate both as a supplier and a consumer of resources. *See* Tanenbaum, §1.1.2, pp. 7-8. Each peer in the network can directly initiate a session with other peers, and data can be stored and accessed across multiple nodes. P2P networks can self-organize and dynamically adjust to the entry and exit of peers, making them scalable and robust against single points of failure.

27. In P2P, all peers need not be equal, strictly speaking. P2P peers may be supported by a central server, or may designate a peer as a “super-peer” that takes on additional responsibilities for the P2P. *See, e.g.,* Tanenbaum, §1.1.2, pp. 7-8 (describing server-supported P2P service, Napster).

VI. OVERVIEW OF THE '099 PATENT

28. The '099 patent proposes a videoconferencing system in which “a filmstrip containing simultaneous images or icons representing the participants are displayed on one or more display screens.” '099 patent, Abstract. The '099 patent notes that some videoconferencing systems have unsatisfactory ways of displaying many parties in the conference:

In a multiparty video call there are multiple video images (usually at least one from each end-point). If there are more images than video monitors, the images will be rendered at a small scale or viewed one at a time by some multiplexing technique. It is often difficult to view individual images at full scale and quickly sort between multiple images.

'099 patent, 1:24-29.

29. The '099 patent's videoconferencing system facilitates selection and promotion of different video feeds to a larger size:

In the case where multiple live video feeds are presented to one or more video monitors, but where there are more images than monitors embodiments of the invention permit easy selection and promotion to full scale of single images. Promotion of two images to $\frac{1}{2}$ scale and three images to $\frac{1}{3}$ scale etc. is permitted.

'099 patent, 3:34-39.

30. Further, the '099 patent describes a challenge of making features available to all of the participants in a mixed topology network, where some nodes are mesh connected, and others are non-mesh connected:

Another problem arises in mixed topology networks where a number of endpoints are connected in a mesh connection, and at least one endpoint is connected to a mesh-connected endpoint with a non-mesh connection. In

this case, **it would be desirable to permit the non-mesh connected endpoint to enjoy the same features as the mesh-connected endpoints.**

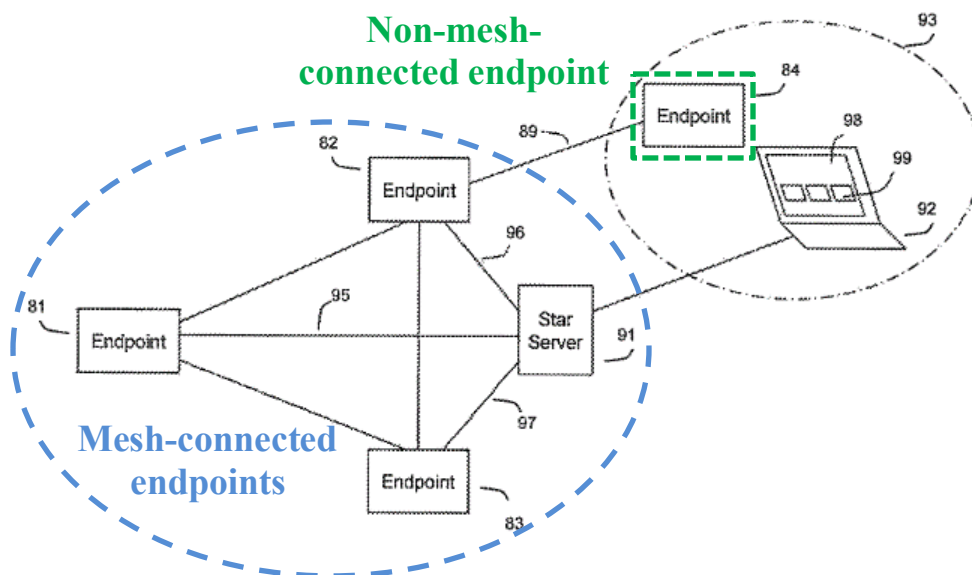
'099 patent, 4:5-11.

31. The '099 patent proposes a mechanism whereby “non-mesh endpoint(s) user(s) can use the filmstrip user interface described above to exercise the same control over the their [*sic*] video presentation as would a mesh connected user.” '099 patent, 15:46-49. In one example, the '099 patent proposes using a web browser at the non-mesh connected endpoints to enable display control features:

A web page including the filmstrip is hosted on a Star Server and displayed at the star endpoint location. The Star Server receives video from all mesh-connected endpoints constructing the filmstrip on the web page. **User interaction with the filmstrip received by the Star Server results in the appropriate video stream being transmitted to the star endpoint.**

'099 patent, 15:50-57.

32. Fig. 9 of the '099 patent illustrates a multiparty conference with a mixed topology network, including mesh-connected endpoints, and non-mesh-connected endpoints.



'099 patent, Fig. 9 (annotated)

33. In the conference in Fig. 9, a separate Star Server 91 connects the computer 92 co-located with the star (non-mesh) endpoint 84. '099 patent, 16:10-12. However, the '099 patent explains that the star server “could be a **program** running on a computer within one of the mesh endpoints e.g. 11 in FIG. 1.” '099 patent, 17:3-5.

34. As I will explain below, these concepts were well-known as of the priority date of the '099 patent.

VII. CLAIM CONSTRUCTION

35. It is my understanding that in order to properly evaluate the '099 patent, the terms of the claims must first be interpreted. It is my understanding that for the purposes of this *inter partes* review, the claims are to be construed under the so-called *Phillips* standard, under which claim terms are given their ordinary

and customary meaning as would have been understood by a POSITA in light of the specification and prosecution history, unless the inventor has set forth a special meaning for a term. I have also been informed that claim terms only need to be construed to the extent necessary to resolve the obviousness inquiry.

36. I have reviewed the entirety of the '099 patent, as well as its prosecution history. It is my opinion that for purposes of applying the asserted prior art to evaluate the patentability of the Challenged Claims, no terms require requires express construction.

VIII. IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE

37. The discussion in this Declaration provides a detailed analysis of how the asserted prior art teaches each limitation of the Challenged Claims.

38. As part of my analysis, I have considered, and discuss in detail, the scope and content of the prior art and any differences between the alleged invention and the prior art.

39. It is my opinion that the alleged invention recited in the Challenged Claims would have been obvious in view of the teachings of the asserted prior art and the knowledge of a POSITA.

IX. THE COMBINATION OF SETTON AND BAKER RENDERS OBVIOUS CLAIMS 3 AND 10

A. Summary of Setton

40. Setton relates to “[m]ultiway peer-to-peer media streaming.” Setton, Abstract. Setton and the ’099 patent are from the same field of endeavor of video teleconferencing. *See* ’099 patent, Abstract, 15:42-45; Setton, [0014], [0019].

Setton is relevant to the particular problem of the ’099 patent, which is video teleconferencing in mixed topology networks. *See* ’099 patent, 15:46-49; Setton, [0014], [0019], [0018], [0016], [0026].

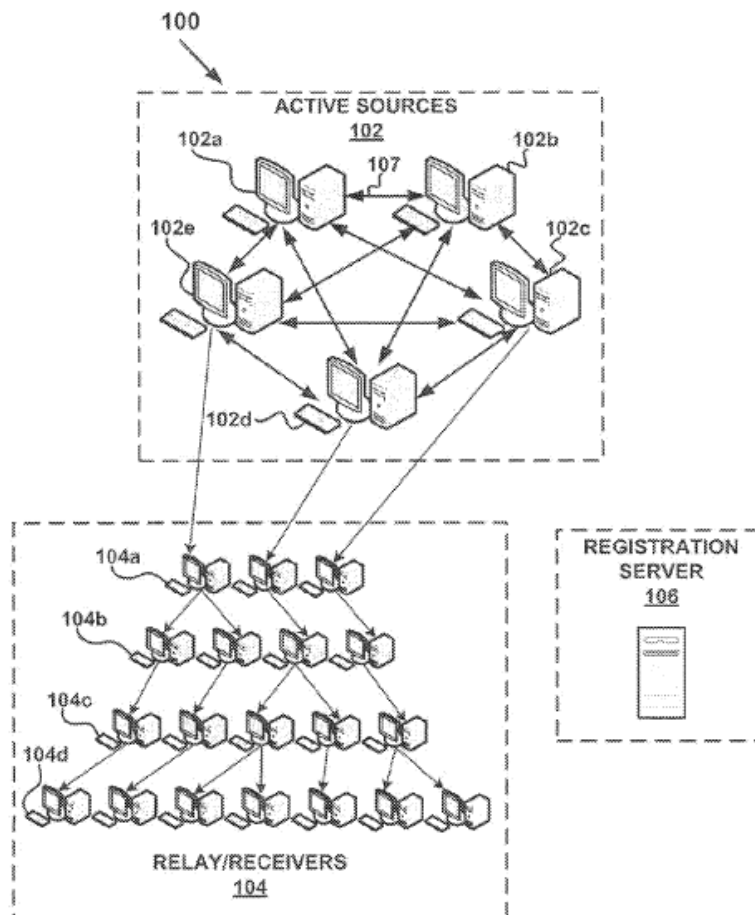
41. Setton’s multiway peer-to-peer network enables video teleconferencing between multiple participants over a network such as the Internet.

In embodiments in accordance with the present invention, **a peer-to peer network** is defined to have at least two portions. The first portion of the peer-to-peer network is comprised of **active sources**, and the second portion of the peer-to-peer network is comprised of **relay/receivers**. In embodiments in accordance with the present invention, the active sources aggregate the media streams using a **peer to-peer** data-driven streaming protocol. The aggregated stream is transmitted to all the relay/receiver peers via **application layer multicast tree(s)**. ... Embodiments in accordance with the present invention enable, for example, **a panel discussion between several distributed speakers** addressing a distributed audience, **over a**

network such as the Internet, at low cost, and in a scalable fashion. There are a variety of events and communities where a large number of people at distributed geographical locations may want to interact.

Setton, [0014]; *see also*, [0019].

42. Setton's multiway peer-to-peer network integrates a first portion of "active sources 102" with a second portion of "relay/receivers 104," as illustrated below in Fig. 1.

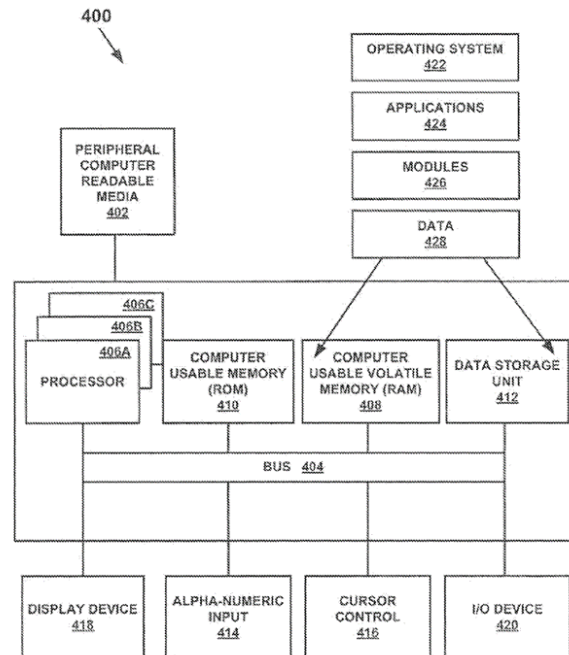


Setton, Fig. 1

In the embodiment depicted in FIG. 1, the peer-to-peer network is comprised of **a first portion comprised of active sources 102 and a second portion comprised of relay/receivers 104.**

Setton, [0016].

43. As shown in Fig. 1 above, each active source 102 and relay/receiver 104 is represented as a computer. The active sources 102 include an IP address and may “generate content (e.g., a video stream, an audio stream, etc.).” Setton, [0016]; *see* [0014] (describing a “list of the number of active sources and of their internet protocol (IP) addresses.”). Setton explains that the computer system illustrated in Fig. 4, for example could be any of the peers, including the active sources and relay/receivers. Setton, [0032].



Setton, Fig. 4

44. As shown in Fig. 4 above, the computer system includes one or more processors, a data storage unit, computer usable member (ROM), computer usable volatile memory (RAM), a display device, and various peripherals (e.g., keyboard). Setton, [0032]-[0035].

45. In some cases, a participant who does not wish to generate content may nevertheless be defined as an active source to gain a faster connection:

...a listener to a teleconference who does not wish to speak (i.e., generate content), but who wishes to be ensured prompt delivery of generated content, can indicate a preference to be defined as an active source. **In such an example, even though the peer is not generating content, the peer is still defined as an**

active source and, as such, receives generated content without considerable delay.

Setton, [0017].

46. The relay/receivers 104 are also associated with IP addresses and may or may not generate content. Setton, [0014]. For example, in most cases, the relay/receivers 104 may only be observers viewing the content generated by the active sources 102 without generating content themselves. However, the relay/receivers 104 may, at some point during the teleconferencing call, change roles and start generating content.

Registration server 106 also assists peers, in a similar fashion, when they transition between being a member of active sources 102 and relay/receivers 104. An example of a peer transitioning from one portion of the peer-to-peer network to another is a speaker (typically an active source peer) going silent for a long period of time (typically a relay/receiver peer). **Another example of a peer transitioning from one portion of the peer-to-peer network to another is a member of the audience (typically a relay/receiver peer) joining the discussion (typically an active source peer).**

Setton, [0026].

There are a variety of events and communities where a large number of people at distributed geographical locations may want to interact. Examples include a panel

discussion, lecture, town hall meeting, community event, religious ceremony, or presidential debate. **In these examples, typically only a small fraction of the large number of participants are active at any point in time, while the others are observing (in some cases the roles may change over time)**. Additionally, embodiments in accordance with the present invention are able to use different peer-to-peer multicast networks and protocols to transmit the different media streams to the peers and accommodate the peers' different quality of service or delay constraints.

Setton, [0014].

47. The active sources 102 may use a “mesh-based data aggregation peer-to-peer protocol” in which each active source (e.g. 102e) is connected to all of the other active sources (e.g. 102a-d) directly.

Referring again to FIG. 1, in the architecture depicted in FIG. 1, **the active sources 102 use a mesh-based data aggregation peer-to-peer protocol**, while the relay/receivers 104 use a tree-based video multicast protocol.

Setton, [0018].

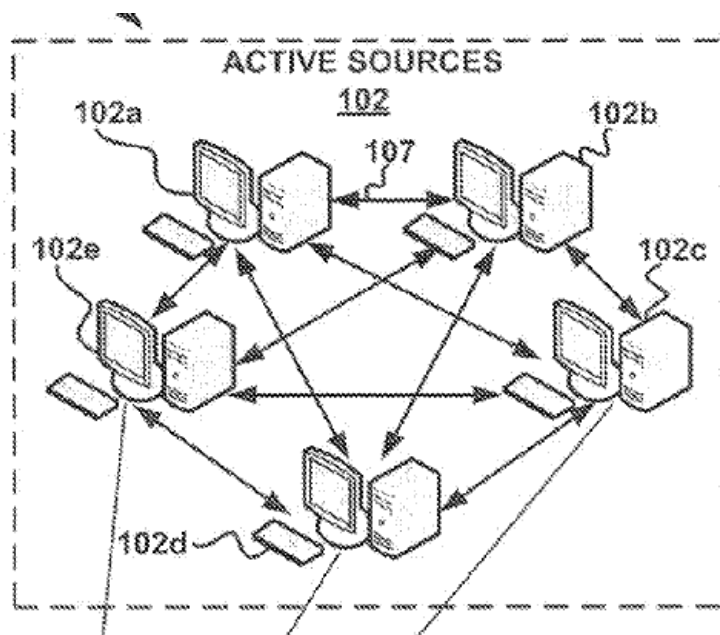
More specifically, in embodiments in accordance with the present invention, each active source (e.g., peers 102a, 102b, 102c, 102d, and 102e) which produces content, is responsible for aggregating all other content with its own.

In one embodiment, such aggregation is accomplished using **a mesh-based data aggregation peer-to-peer protocol**. In one embodiment, the content is, for example, a video stream and the corresponding audio stream generated by an attendee to a video teleconference.

Setton, [0019].

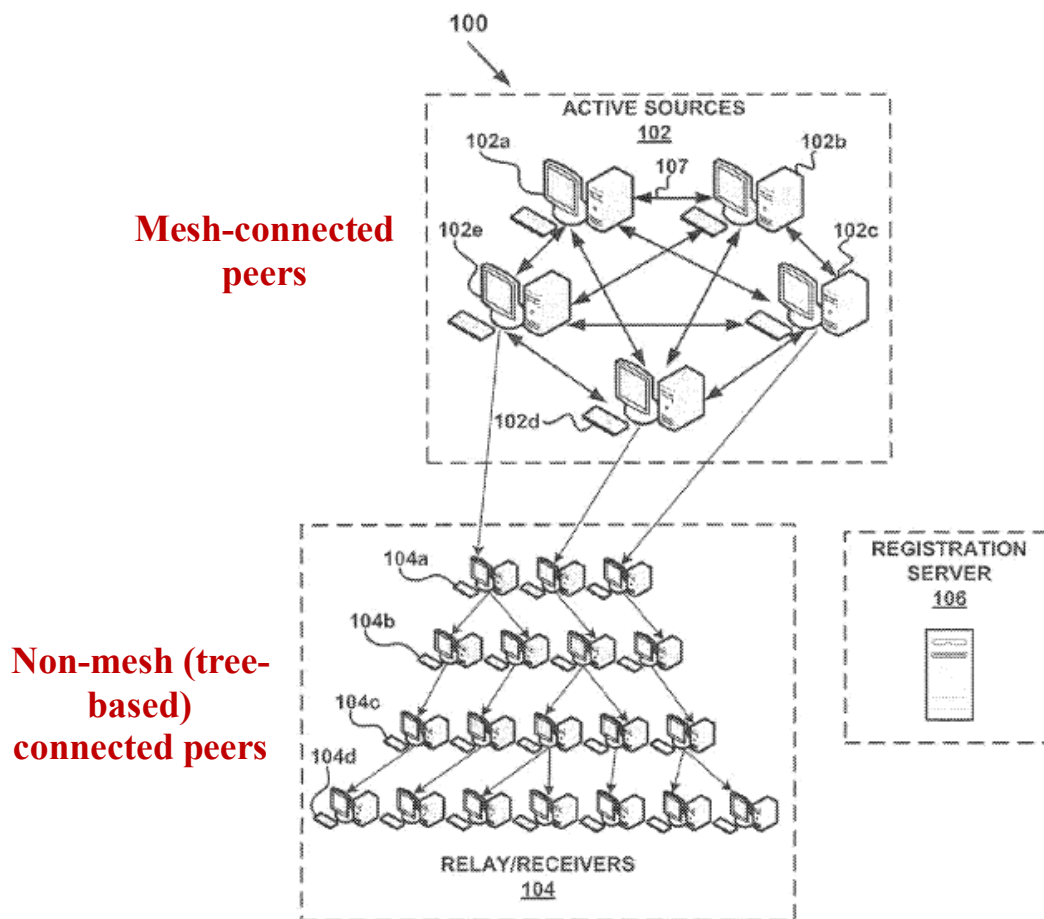
More specifically, in FIG. 1, **the bi-directional arrows, typically shown as 107, denote packet exchanges between the active sources (e.g., peers 102a, 102b, 102c, 102d, and 102e).**

Setton, [0023].



Setton, Fig. 1 (cropped)

48. Unlike active sources 102, the relay/receivers 104 are not directly connected to each other or each active source 102. As an example, relay/receiver 104a is connected only to one active source: active source 102e.



Setton, Fig. 1 (annotated)

49. Setton explains that the relay receiver 104a, and the other relay/receivers 104 connected to it, receive the aggregated video stream via the active source 102e:

50. In this configuration, relay/receiver 104a receives aggregated video/audio data generated by all active sources 102 from a single active source 102e. Additional relay/receivers 104b-d may be connected to the first relay/receiver 104a in an application layer multicast tree. For example, relay/receiver 104b may receive the aggregated video/audio data from relay/receiver 104a, relay/receiver 104c may receive the aggregated video/audio data from relay/receiver 104b, and relay/receiver 104d may receive the aggregated video/audio data from relay/receiver 104c (thereby forming a “root” of the “tree”):

...[I]n embodiments in accordance with the present invention, the aggregated media stream is transmitted to the relay/receivers 104 (e.g., 104a, 104b, 104c, and 104d) via a tree-based peer-to-peer video multicast protocol. Such a tree is depicted in FIG.1. In this embodiment, **the roots of the application layer multicast tree are selected from the active sources (102c, 102d, and 102e)** which aggregate the media stream as described above.

Setton, [0022].

In embodiments in accordance with the present invention, the active sources aggregate the media streams using a peer-to-peer data-driven streaming protocol. **The aggregated stream is transmitted to all the relay/receiver peers** via application layer multicast

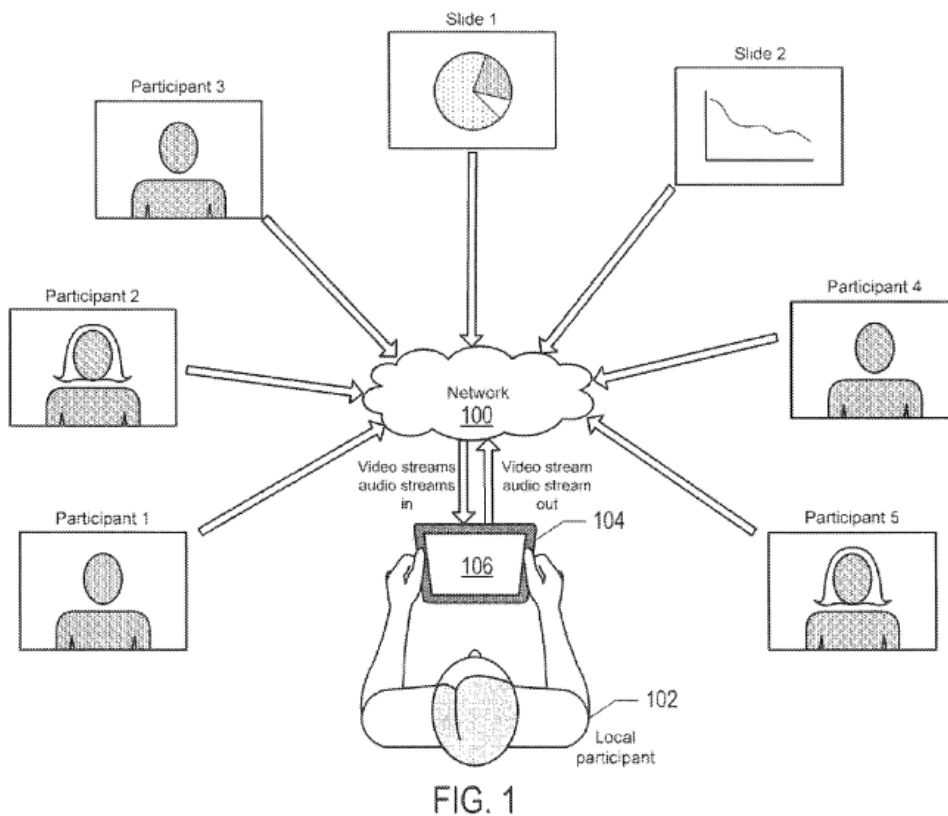
tree(s). **The roots of the tree are chosen among the active sources.**

Setton, [0014].

B. Summary of Baker

51. Baker relates to “[m]ethods and systems for using portable electronic devices in video conferences.” Baker, Abstract. Baker and the ’099 patent are from the same field of endeavor of video teleconferencing. *See* ’099 patent, 4:5-11, 14:46-49; Baker, 8:27-32, 40-44. Baker also addresses the particular problem of the ’099 patent, which includes displays and user interfaces for video teleconferencing systems. *See* ’099 patent, 4:5-11, 14:46-49; Baker, 8:40-41, 55-57.

52. In Baker, a portable electronic device “receives each remote participant's audio data stream and at least one video stream over a network and arranges the video streams in a data structure that describes the location of each video stream's associated viewing area within a virtual meeting space.” Baker, Abstract. The video and audio streams from each participant may be aggregated and transmitted to the device via a network such as the Internet, as shown in Fig. 1 below.



Baker, Fig. 1

The video conference is conducted over a network 100 that can be a local area network, a wide area network, a metropolitan area network, a global area network, or **the Internet**. In the example of FIG. 1, the local participant 102 uses a PED 104 to participate in the video conference with five other remotely situated participants. The PED 104 can be a tablet computer or a smartphone that sends video and audio streams generated by the local participant during the video conference to the network 100. **The video and audio streams generated by each of the remote participants are sent to the network 100 and received by the PED 104. The separate video**

streams are to be presented as separate viewing areas on a touchscreen/display 106 of the PED 104. The separate viewing areas are identified as participants 1-5.

Baker, 2:31-49.

53. Each participant's video stream is presented in a separate viewing area. Baker, 2:46-48. Baker describes several different ways of presenting the video streams. For example, a full size video stream or content shared by a participant may be located in the main display area and a "strip of miniaturized remote participant viewing areas" may be located along the edge of the main display area, as shown in Figs. 10A and 10B below. Baker, 8:40-44.

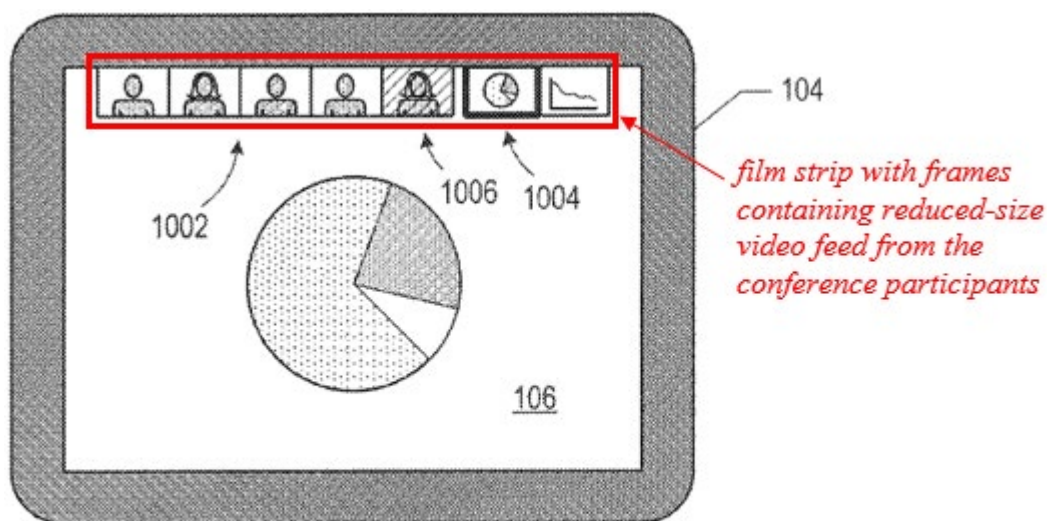


FIG. 10A

Baker, Fig. 10A (annotated)

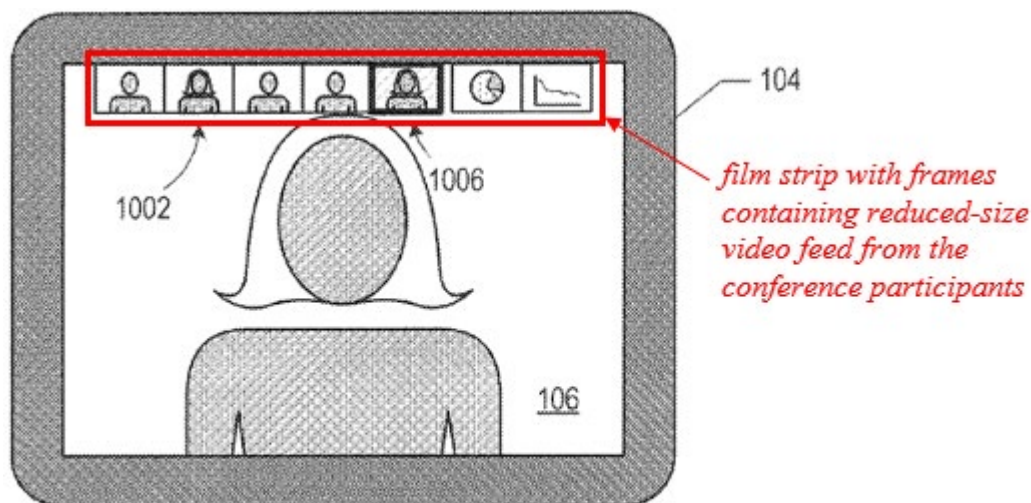


FIG. 10B

Baker, Fig. 10B (annotated)

Alternatively, **a strip of miniaturized remote participant viewing areas are located along an edge of the display** with the speaking participant(s) highlighted can be included in the display 106 so the local participant knows which way to pan, tilt, or scroll the display 106 to see the speaking participant. FIG. 10A shows an example strip 1002 of reduced size images of the live remote participant viewing areas and two content viewing areas located along the top of the display. The participant viewing areas can be arranged in the same manner the participant viewing areas are arranged in the virtual meeting space 500. In the example of FIG. 10A, the border of the reduced viewing area 1004 of slide 1 is highlighted to indicate that slide 1 is the image shown on the display 106.

Baker, 8:40-52.

54. The miniaturized video streams in the strip may have a lower bit-rate or quality than the video stream in the main display area.

Also, the video of non-speaking participants can be sent with a lower bit-rate/quality, based on the assumption that the participants are less likely to be looking at the non-speaking participants.

Baker, 9:11-14.

The local participant 102 may select to have the video data streams of non-speaking participants in the virtual meeting space, to be decoded at a slower frame rate to save CPU cycles and battery power or use still images of the non-speaking participants that are periodically updated. In addition, the video streams in the strip of miniaturized images 1002 shown in FIG. 10 can be decoded at a slower frame or can be still images that are periodically updated.

Baker, 9:19-26.

55. The user of the device can select via scroll, touch, or tap gestures, a miniaturized video stream from the strip to move to the main display area, thereby receiving a higher bit-rate or quality video stream from the chosen participant:

The reduced viewing area 1006 of participant 5 is **highlighted** to indicate that participant 5 is the only participant speaking. The local participant can then move the display 106 to view participant 5, as shown in FIG. 10, using at least one of the methods described above

with to FIGS. 8A-8D. Alternatively, **the local participant can scroll with his finger along the strip 1002 to participant 5's viewing area to move participant 5's viewing area into the display 106, or the local participant can touch or tap on the highlighted image 1006 of participant 5 to bring up participant 5's viewing area in the display 106.**

Baker, 8:52-62.

C. Reasons to Combine Setton with Baker

56. A POSITA would have been motivated to combine Baker's multi-participant display and display control techniques with Setton's multi-way peer-to-peer media streaming system. The incorporation of Baker's features into Setton's system would be simply combining prior art elements (Baker's "strip of miniaturized remote participant viewing areas" and "touch or tap" control interface with Setton's peer-to-peer media streaming system) according to known methods to yield predictable results. The methods for combining Setton and Baker would include well-known, routine software application programming to specify the formats of (e.g., data type, data rate), how the video streams are presented on screen, and to provide user interface objects, triggers, and corresponding actions that would modify the presentation of the video streams to enlarge at least one of the video streams. The user interface objects may include buttons, scroll bars, regions on the screen that can be tapped or clicked. The software would define the

actions or responses to the user interface actions that trigger the actions. Example actions would include enlarging a selected (e.g., clicked) thumbnail or miniaturized stream, or returning a full size stream to a miniaturized stream.

57. This type of software programming, including defining user interface visual objects and their associated actions, would have been within the skill set of a POSITA. The results would have been predictable, since the additional software features simply modify how the content is displayed (e.g., the size of each video feed on the screen), and when (e.g., in response to a touch or tap). Configuring a display and user interface for a video conference would have been an essential part of developing videoconference applications and functionality.

58. Additionally, the combination of Baker and Setton would be applying a known technique (Baker's video strip and display control features) to a known device (Setton's peer-to-peer media streaming system) ready for improvement to yield predictable results. Setton describes a network topology that allows both endpoints to connect and participate in a video conference via mesh and non-mesh connections. Setton discloses that the video streams are aggregated from mesh-connected nodes (active sources 102) and transmitted to non-mesh connected nodes (relay/receivers 104), but does not go into great detail about the format or arrangement of the video conferencing content displayed by the nodes. Setton, [0019], [0034].

59. Baker, however, discloses one implementation of Setton's mesh/non-mesh video conferencing system. In Baker, multiple video conferencing streams are aggregated, just like in Setton, and displayed in a single strip of multiple video streams, with some of the video streams having a reduced size. Baker, Figs. 10A, 10B, 8:40-62. In this way, Baker's features would further define Setton's video stream display and interfaces in a way that furthers Setton's purposes, such as reducing network load, accommodating network constraints of different participants. Further, providing Baker's strip of miniaturized streams instead of multiple full sized streams may reduce processing time and improve communication stability, thereby reducing delay. As explained above, incorporating these functionalities of Baker into Setton's system would involve routine software programming that would have been well within the skill set of a POSITA. Thus, the results would have been predictable.

60. A POSITA would be motivated to provide the film strip of reduced-size video feeds to provide flexibility for users who are constrained by network conditions but who do not need to be as active in the conference as other, such as an audience member. *See, e.g.*, Setton [0017], [0026]. In addition, a film strip advantageously allows participants to simultaneously focus on the speaking participant, while also monitoring all participants at once for various purposes, such as documenting attendance or monitoring engagement, making the video

conferencing system disclosed more suitable for a wide variety of applications, including education, corporate environments, or social settings.

61. A POSITA would have had a reasonable expectation of success by combining Setton and Baker. Incorporating Baker's user interface elements (strip of miniaturized images and touch or tap functionality) into Setton's peer-to-peer network, including the peers, would involve straight-forward computer programming techniques that a POSITA would be capable of completing. Setton provides limited detail on how to display the aggregated video streams and control them. However, a POSITA would have had this requisite background knowledge, and Baker simply provides implementation details that Setton expects. The implementation details of Baker include well-known video layouts and UI objects (e.g., buttons, clickable areas) that specify how the video streams are served and presented on screen. Baker, Figs. 10A, 10B, 8:40-62. These UI objects were common and familiar to both application developers and users. Thus, a POSITA would have a reasonable expectation of success in implementing the features of Baker into Setton's peer-to-peer network.

D. Detailed Analysis of Claims

1. Claim 3

a. [3.0] *A teleconferencing system comprising:*

62. To the extent the preamble is limiting, Setton discloses it.

63. Setton teaches a “multiway peer-to-peer media streaming” network that is used for “panel discussions” and “video teleconferenc[ing].” Setton, Abstract, [0014], [0019]. Distributed speakers access the network from peers, which include “active sources” and “relay/receivers,” over the Internet:

In embodiments in accordance with the present invention, **a peer-to peer network** is defined to have at least two portions. The first portion of the peer-to-peer network is comprised of **active sources**, and the second portion of the peer-to-peer network is comprised of **relay/receivers**. In embodiments in accordance with the present invention, the active sources aggregate the media streams using a peer to-peer data-driven streaming protocol. The aggregated stream is transmitted to all the relay/receiver peers via application layer multicast tree(s). ... Embodiments in accordance with the present invention enable, for example, **a panel discussion between several distributed speakers** addressing a distributed audience, **over a network such as the Internet**, at low cost, and in a scalable fashion. There are a variety of events and communities where a large number of people at distributed geographical locations may want to interact.

Setton, [0014].

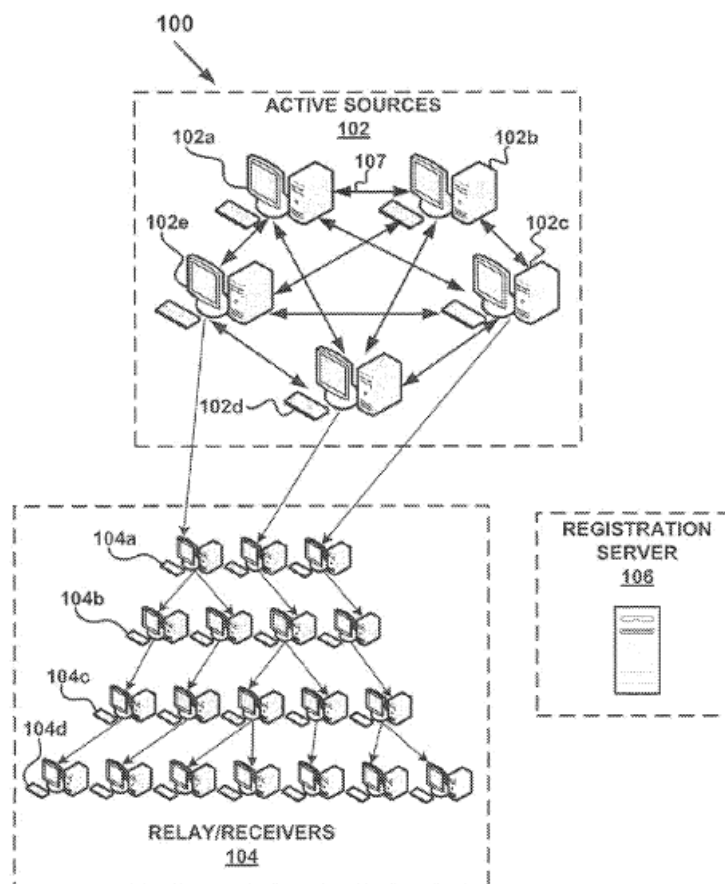
64. Setton's "multiway peer-to-peer media streaming" network, including the "active sources" and "relay/receivers" peers, discloses the claimed "teleconferencing system."

- b. [3.1.1] *a plurality of endpoints, at least some of which are connected in a mesh configuration and at least one of which is connected in a non-mesh configuration to a selected one of the mesh-connected endpoints, and*

65. Setton discloses this limitation.

66. **First**, Setton's network is used by "active sources 102" and "relay/receivers 104." As explained in more detail below, Setton's active sources 102 are *endpoints ... connected in a mesh configuration*. The relay/receivers 104 are *endpoints ... in a non-mesh configuration*.

67. Setton's Fig. 1, shows the peer-to-peer network system that includes a first portion of active sources 102 and a second portion of relay/receivers 104.



Setton, Fig. 1

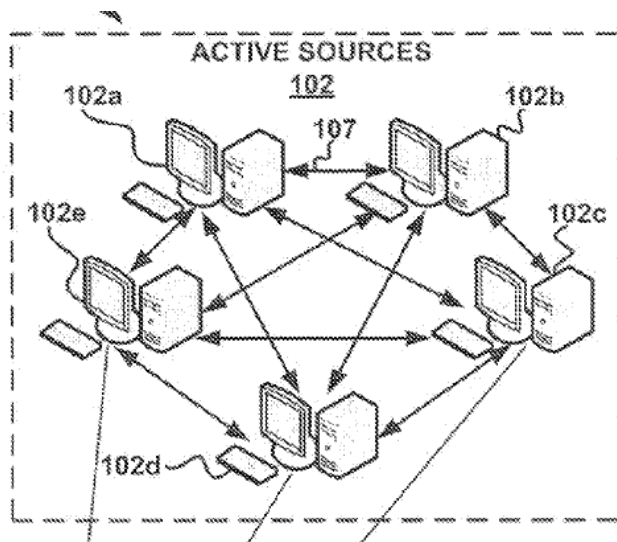
68. The active sources 102 are computers, as seen in Fig. 1, that each have an IP address and “generate content (e.g., a video stream, an audio stream, etc.).” Setton, [0016]; *see also* [0014] (describing a “list of the number of active sources and of their internet protocol (IP) addresses.”). Similarly, the relay/receivers 104 are also computers with IP addresses that can generate video and audio content. *See* Setton, [0026] (“Another example of a peer transitioning from one portion of the peer-to-peer network to another is a member of the audience (typically a relay/receiver peer) joining the discussion (typically an active source peer).”).

69. The active sources 102 and relay/receivers 104 are the claimed *plurality of endpoints*. Setton's active sources 102 and relay/receivers 104 are consistent with the endpoints of the '099 patent, which it describes as including "video cameras," "monitors," and/or "display controller[s]" that have a "network address (e.g., IP address or URL) used for call set-up." '099 patent, 5:4-17. Further, although the '099 patent sometimes refers to computer devices separate from the endpoints (*see* Fig. 1), the '099 patent also explains that endpoints may comprise functionality implemented in a computer:

Computer 92 could be any computer for example a laptop of smart mobile phone (iPhone or iPad). Furthermore
Endpoint 84 function may run on that same computer 92.

'099 patent, 16:19-21, Fig. 9.

70. ***Second***, active sources 102 (*endpoints*) are a "first portion of the peer-to-peer network" that uses "a **mesh-based data aggregation peer-to-peer protocol**." Setton, [0014], [0018]. This arrangement is depicted in Fig. 1, below:



Setton, Fig. 1 (cropped)

71. Active sources 102 use this mesh-based protocol to exchange video and audio video teleconferencing content:

More specifically, in embodiments in accordance with the present invention, each **active source (e.g., peers 102a, 102b, 102c, 102d, and 102e) which produces content**, is responsible for aggregating all other content with its own. In one embodiment, such aggregation is accomplished using **a mesh-based data aggregation peer-to-peer protocol**. In one embodiment, the **content is**, for example, a **video stream and the corresponding audio stream** generated by an attendee to a **video teleconference**.

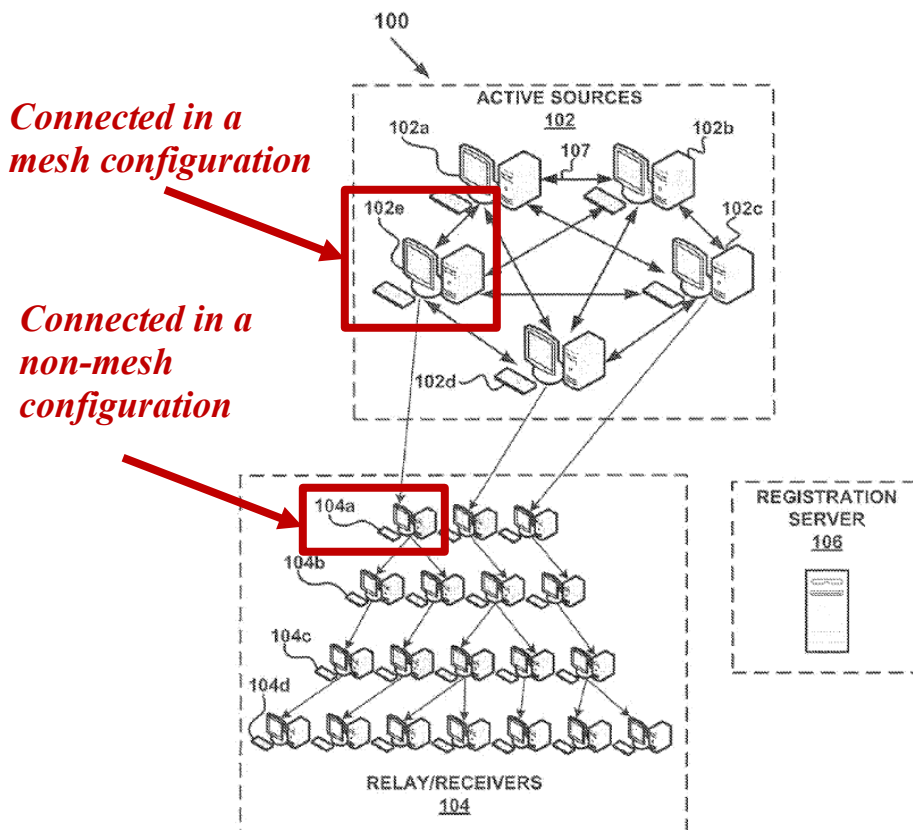
Setton, [0019].

72. In addition, Setton explains that active sources 102 that use its mesh-based protocol communicate with each other directly. *See* Setton, [0023] (“in FIG.

1, the bi-directional arrows, typically shown as 107, denote packet exchanges between the active sources (e.g., peers 102a, 102b, 102c, 102d, and 102e).”).

73. Setton’s description of how its active sources 102 communicate is consistent with the ’099 patent’s description of the “mesh configuration,” which is a configuration in which “each endpoint is connected directly to each other endpoint in the conference.” ’099 patent, 2:34-37. Thus, the active sources 102 are *endpoints ... connected in a mesh configuration*.

74. **Third**, relay/receivers 104 (*endpoints*) utilize a “**tree-based peer-to-peer video multicast protocol**.” Setton, [0022]. As depicted in Fig. 1, below, unlike active sources 102, the relay/receivers 104 are not directly connected to each other or each active source 102; they are *connected in a non-mesh configuration*:



Setton, Fig. 1 (annotated)

75. Setton further describes how the “tree-based” relay/receiver peers communicate media streams from the active sources 102 down through the tree:

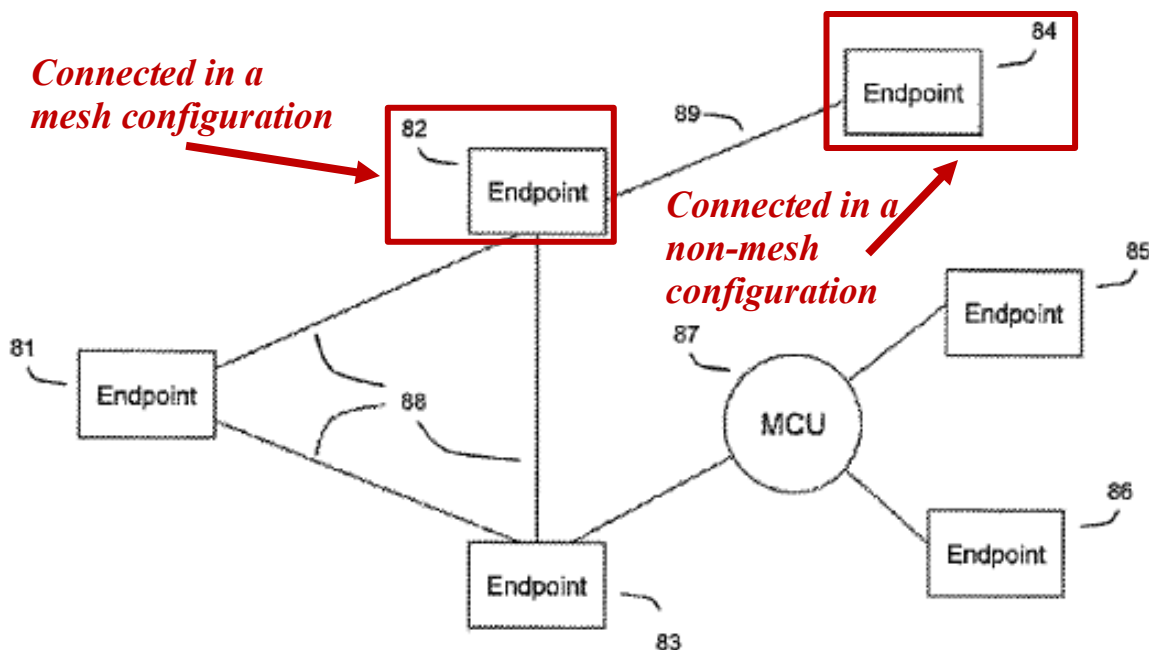
In embodiments in accordance with the present invention, the **active sources aggregate the media streams** using a peer-to-peer data-driven streaming protocol. **The aggregated stream is transmitted to all the relay/receiver peers** via application layer **multicast tree(s)**. The roots of the tree are chosen among the **active sources**.

Setton, [0014].

[I]n embodiments in accordance with the present invention, the **aggregated media stream is transmitted to the relay/receivers 104 (e.g., 104a, 104b, 104c., and 104d) via a tree-based peer-to-peer video multicast protocol**. Such a tree is depicted in FIG.1. In this embodiment, **the roots of the application layer multicast tree are selected from the active sources (102c, 102d, and 102e)** which aggregate the media stream as described above.

Setton, [0023].

76. Setton's description of the relay-receivers' 104 tree-based connections is consistent with the '099 patent's description of an endpoint connected in a non-mesh configuration, which is one that is not directly connected to each endpoint of a mesh configuration. '099 patent, 15:46-16:35. An endpoint connected in a non-mesh configuration is illustrated in Fig. 8 of the '099 patent, below, which shows endpoint 84 (non-mesh connected) is connected only to the endpoint 82 (mesh-connected):



'099 patent, Fig. 8 (annotated)

77. In the configuration depicted in Setton's Fig. 1, relay/receiver 104a (*endpoint ... connected in a non-mesh configuration*) is connected to only one active source: active source 102e. Relay/receiver 104a receives video/audio data generated by active sources 102 (*endpoints connected in a mesh configuration*) from a single active source 102e (connected in a non-mesh configuration *to a selected one of the mesh-connected endpoints*).

78. **Fourth**, Setton explains that an active source 102 is “selected” as a “root[] of the application layer multicast tree” to deliver the aggregated media stream to the relay/receivers 104:

the aggregated media stream is transmitted to the relay/receivers 104 (e.g., 104a, 104b, 104c, and 104d) via a tree-based peer-to-peer video multicast protocol. Such a

tree is depicted in FIG. 1. In this embodiment, **the roots of the application layer multicast tree are selected** from the active sources (102c, 102d, and **102e**) which aggregate the media stream as described above.

Setton, [0023].

79. Active source 102e was therefore *selected* to provide the connection point between the multicast tree for relay/receiver 104a and the active sources 102. Thus, Setton discloses that the relay/receiver 104a is connected in a non-mesh configuration *to a selected one of the mesh-connected endpoints*.

80. Thus, Setton discloses this limitation.

c. **[3.1.2] *which are configured such that a film strip with frames containing reduced-size video feeds from the conference participants is displayed at the mesh endpoints;***

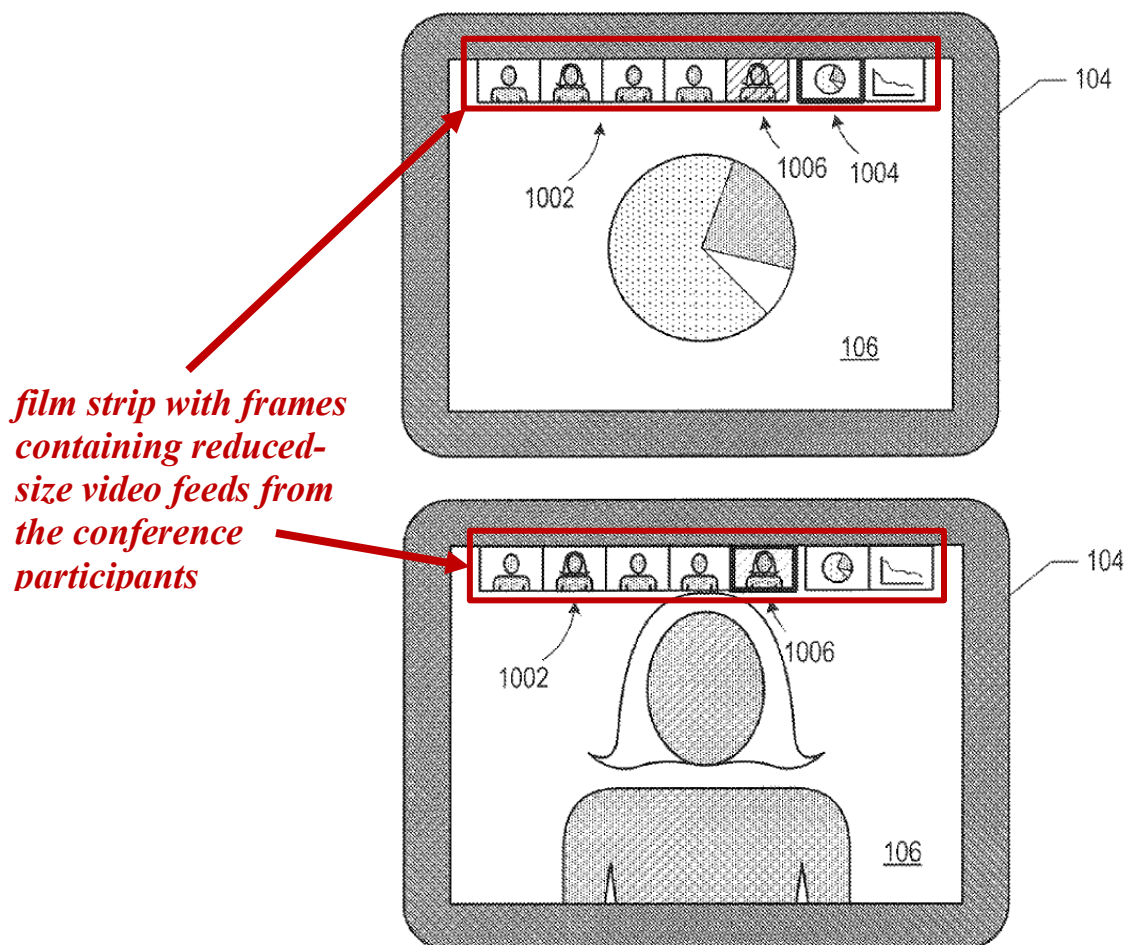
81. The combination of Setton and Baker renders this limitation obvious.

82. ***First***, Setton's systems and methods relate to the video streaming of multi-way video communication. For instance, Setton's system is described "with respect to the streaming of video data." Setton, [0014]. Setton explains that "embodiments in accordance with the present invention provide a real-time multi-way video communication protocol based on peer-to-peer technologies." Setton, [0014]. Further, Setton describes aggregating video streams for attendees in a video teleconference (*video feeds from the conference participants*):

In one embodiment, such aggregation is accomplished using a **mesh-based data aggregation** peer-to-peer protocol. In one embodiment, the content is, for example, **a video stream and the corresponding audio stream generated by an attendee to a video teleconference.**

Setton, [0019].

83. *Second*, Baker describes displaying “video streams” of participants in a teleconference as “a strip of miniaturized images.” Baker, 9:19-26. Baker explains that a screen display 106 on a portable electronic device (“PED”) 104 viewed by a local participant 102. Baker, 2:31-48. The display 106 includes “an example strip 1002.” Baker, 8:45-47.



Baker, Figs. 10A and 10B (annotated)

84. Strip 1002 displays “miniaturized remote participant viewing areas” (reduced-size video feeds from the conference participants) along the edge of the display:

Alternatively, a strip of miniaturized remote participant viewing areas are located along an edge of the display with the speaking participant(s) highlighted can be included in the display 106 so the local participant knows which way to pan, title, or scroll the display 106 to see the speaking participant. FIG. 10A shows **an example**

strip 1002 of reduced size images of the live remote participant viewing areas and two content viewing areas located along the top of the display.

Baker, 8:40-47.

85. The miniaturized video of non-speaker participants “can be sent with a lower bit-rate/quality” to save on processing power:

Also, **the video of non-speaking participants** can be sent with a **lower bit-rate/quality**, based on the assumption that the participants are less likely to be looking at the non-speaking participants.

Baker, 9:11-14.

The local participant 102 may select to have the **video data streams of non-speaking participants** in the virtual meeting space, to be **decoded at a slower frame rate** to save CPU cycles and battery power or use still images of the non-speaking participants that are periodically updated. In addition, the **video streams in the strip of miniaturized images 1002 shown in FIG. 10 can be decoded at a slower frame** or can be still images that are periodically updated.

Baker, 9:19-26.

86. Baker’s “strip of miniaturized remote participant viewing areas” displayed on PED’s display is the claimed *film strip with frames containing reduced-size video feeds from the conference participants*.

87. Third, it would have been obvious to present Setton's video streams of conference participants (*video feeds from the conference participants*) in the form of Baker's strip of miniaturized images (*film strip with frames containing reduced-size video feeds from the conference participants*). One of Setton's objectives is to "accommodate the peers' different quality of service or delay constraints." Setton, [0014], [0028]. Setton also states that a video communication system

...should maintain very **low latency** communication between the different speakers to enable them to converse naturally, and present the audience with a synchronized multiplexed stream at **a high enough quality to create a good viewing experience**.

Setton, [0004].

88. Displaying some remote participants with reduced-size video feeds while others are displayed at a larger size, or full sized, was generally known in the art. For example, U.S. Publication No. 2006/0215765 to Hwang, et al. ("Hwang", Ex.1017) describes "present[ing] decoded bitstreams in a '5+1' split screen format, in which video content for five sources are each presented at the same resolution (e.g. 104×80 pixels), and video content for one source is presented at a larger resolution (e.g. 208×160 pixels)." Hwang, [0060] (referring to Fig. 9); *see also* Figs. 17A-18B. Similarly, U.S. Publication No. 2010/0157017 to Pepperell

(“Pepperell”, Ex.1019) describes several pane layouts for video conferencing terminals including smaller “continuous presence (CP) panes overlaid near the base of full screen panes.” Pepperell, [0047] (referring to Fig. 4); *see also*, Fig. 9; U.S. Publication No. 2009/0015661 to King et al. (“King,” Ex.1018), Fig. 10.

89. The '099 patent itself describes videoconferencing displays that have a combination of large viewing areas and reduced-size viewing areas as “typical prior art screen layouts.” *See* '099 patent, Figs. 3b-3c, 4:42. A POSITA also would have found it obvious to combine Baker’s miniaturized images with Setton’s videoconferencing system because showing miniaturized images of videoconferencing participants was a well-known approach for organizing videoconferencing displays.

90. In order to maintain low latency and/or high frame rate, it would have been obvious to a POSITA to stream and display Baker’s “strip of miniaturized remote participant viewing areas” in a reduced size (*a film strip with frames containing reduced-size video feeds from the conference participants is displayed*) to optimize bandwidth, on the display of Setton’s active sources 102 and relay/receivers 104 (*mesh-connected and non-mesh connected endpoints*). *See* Section IX.C.

91. The combination of Setton and Baker renders this limitation obvious.

d. **[3.2] *a proxy at the selected mesh-connected endpoint for the non-mesh connected endpoint and connected to***

each of the mesh-connected endpoints to receive video feeds from the mesh-connected endpoints for the non-mesh connected endpoint;

92. Setton renders this limitation obvious.

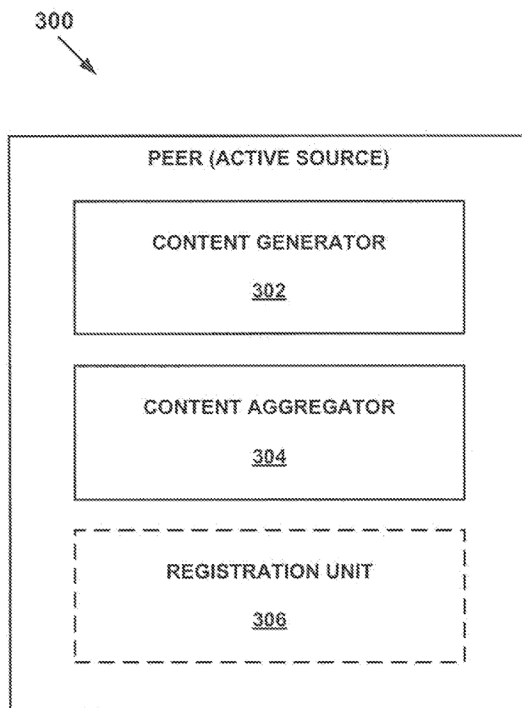
93. *First*, Setton teaches that at least one of the active sources 102 (*mesh-connected endpoint*) includes “a content aggregator 304 for aggregating the content generated by the peer with content generated by other peers” (*a proxy*).

Specifically, Setton discloses:

With reference now to FIG. 3, a schematic block diagram 300 of a peer (**active source**) for use in the present peer-to-peer media streaming technology in accordance with one embodiment of the present invention. As shown in FIG.3, the peer includes a content generator 302 for generating content of the type described above. Additionally, **the peer includes a content aggregator 304 for aggregating the content generated by the peer with content generated by other peers in the manner as was described above.** ... In embodiments in accordance with the present invention, block diagram 300 may also include a feature, not shown, for providing content distribution and which establishes a bridge between active sources 102 and relay/receivers 104.

Setton, [0021].

94. The content aggregator 304 is illustrated in FIG. 3 as part of the “PEER (ACTIVE SOURCE).”



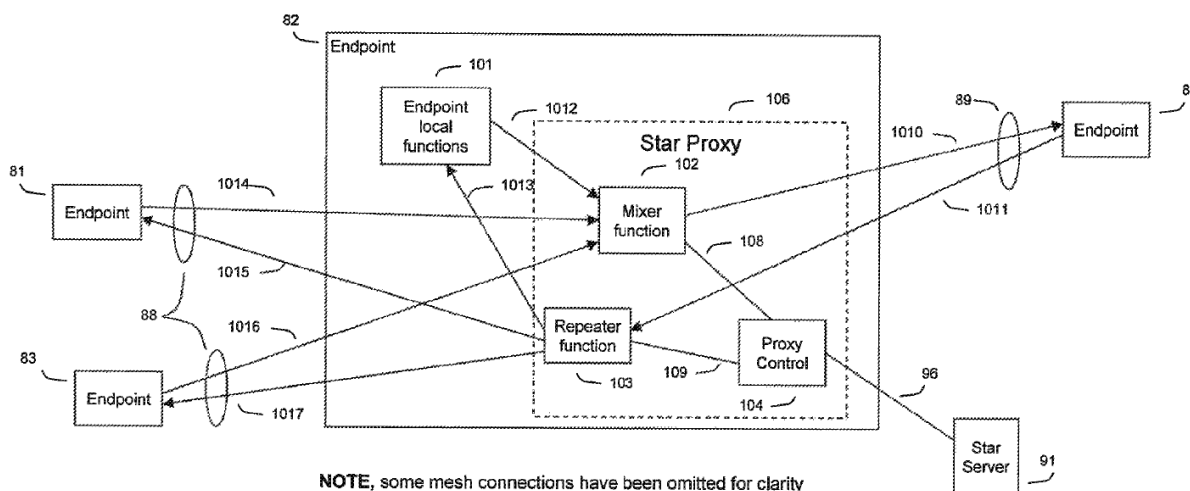
Setton, Fig. 3

95. Setton further explains that any or all of the active sources 102, including active source 102e (*the selected mesh-connected endpoint*) aggregate media streams:

More specifically, in embodiments in accordance with the present invention, **each active source** (e.g., peers 102a, 102b, 102c, 102d, and **102e**) which produces content, is responsible for **aggregating all other content with its own**. In one embodiment, such aggregation is accomplished **using a mesh-based data aggregation peer-to-peer protocol**.

Setton, [0019].

96. Setton's content aggregation feature is consistent with the '099 patent's description of a *proxy*. The '099 patent explains that a proxy performs "Mixer and Repeater" functions, and may be "similar to an embedded [Multiparty Control Unit] MCU." '099 patent, 16:36-38. An MCU "**consolidates all the video feeds** from the end points and presents appropriate video images to each of the end points of all the video camera and collaboration material video feeds in the conference." '099 patent, 1:36-39. This is depicted in the '099 patent's Fig. 10:



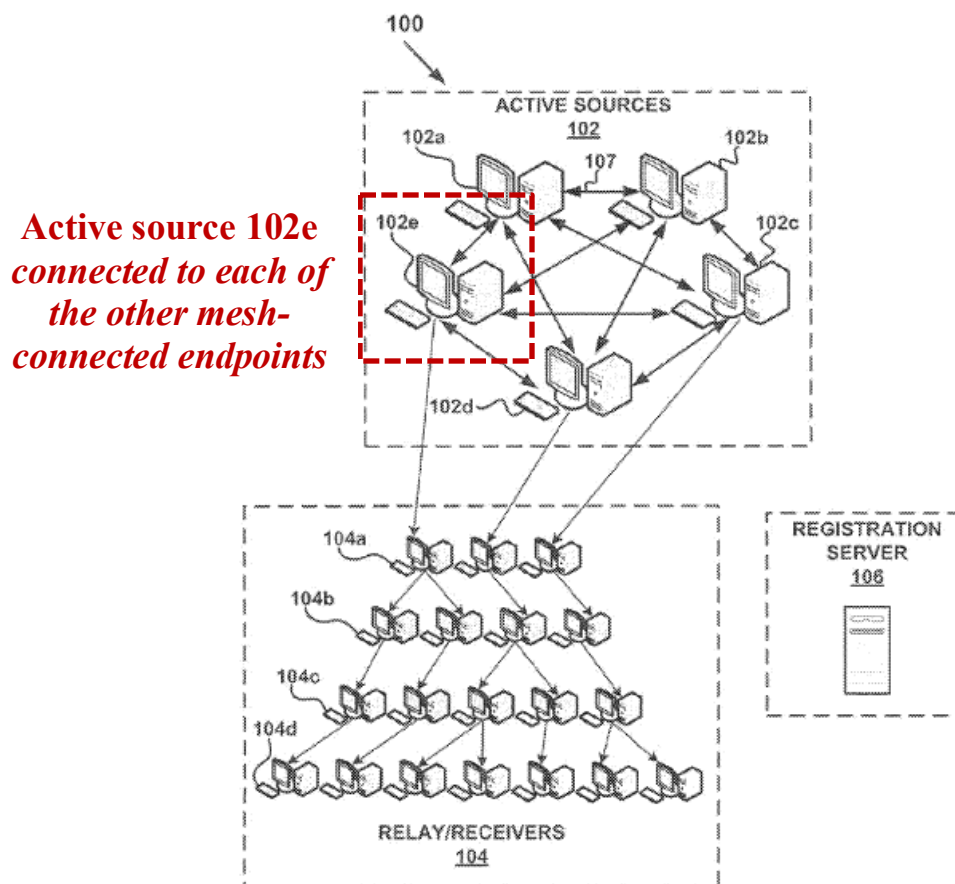
'099 patent, Fig. 10

97. To the extent Setton does not explicitly or implicitly disclose the claimed *proxy* at the active source 102e (*at the selected mesh-connected endpoint*), Setton at least renders this obvious. As shown in Fig. 1 of Setton below, active source 102e is connected to at least one of the relay/receivers 104, which Setton explains receive the aggregated streams from the active sources 102. *See* Setton, [0023]. The content aggregator 304 "aggregat[es] the content generated by the peer

with content generated by other peers.” Setton, [0021]. A POSITA would have understood that the active source 102e, which distributes aggregated content to relay/receiver 104a (see arrow in Fig. 1 below) includes a content aggregator that aggregates content from the other active sources 102 for distribution to the relay/receivers 104, or at the very least would have found this obvious.

98. Implementing the content aggregator 304 on the active source 102e would provide a direct and efficient communication pathway for the aggregated streams. Thus, to the extent Setton does not specifically describe the active source 102e as including or operating the content aggregator 304 and content distribution feature, it would have been obvious to a POSITA for active source 102e (*selected mesh-connected endpoint*) to include or operate the content aggregator 304 and content distribution feature of block diagram 300 (*proxy at the selected mesh-connected endpoint*).

99. *Second*, active source 102e is shown in Fig. 1 to “establish[] a bridge between active sources 102 and relay/receivers 104” (*proxy...for the non-mesh connected endpoint*) and it is also shown to be *connected to each* of the other active sources 102 (*mesh-connected endpoints*).



Setton, Fig. 1

100. Setton further explains that the aggregated media streams from the active sources are to be delivered to the relay/receiver peers (... *for the non-mesh connected endpoint*):

In embodiments in accordance with the present invention, the **active sources aggregate the media streams** using a peer to-peer data-driven streaming protocol. The **aggregated stream is transmitted to all the relay/receiver peers via application layer multicast tree(s)**. The roots of the tree are chosen among the active sources. ... In so doing, embodiments in accordance with

the present invention provide a **real-time multi-way video communication** protocol based on peer-to-peer technologies. Embodiments in accordance with the present invention enable, for example, a **panel discussion between several distributed speakers addressing a distributed audience**, over a network such as the Internet, at low cost, and in a scalable fashion.

Setton, [0014].

101. Setton specifically describes that a media stream can be aggregated by active source 102e (*the selected mesh-connected endpoint*) and delivered to relay/receiver 104a (*the non-mesh connected endpoint*):

embodiments in accordance with the present invention then deliver **the aggregated media stream from the first portion of the peer-to-peer network (i.e., active sources 102) to the second portion of the peer-to-peer network (i.e., relay/receivers 104)**. More specifically, in FIG. 1, the bi-directional arrows, typically shown as 107, denote packet exchanges between the active sources (e.g., peers 102a, 102b, 102c, 102d, and **102e**). Referring again to 208 of FIG. 2, in embodiments in accordance with the present invention, the aggregated media stream is **transmitted to the relay/receivers 104** (e.g., **104a**, 104b, 104c, and 104d) via a tree-based peer-to-peer video multicast protocol.

Setton, [0022].

102. Setton explains that a peer, such as the active source 102e, can include a “content distribution” feature in addition to the content aggregator in order to distribute the aggregated streams to the other peers:

In embodiments in accordance with the present invention, block diagram 300 may also include a feature, not shown, for providing content distribution and which establishes a bridge between active sources 102 and relay/receivers 104.

Setton, [0021].

103. Thus, to the extent Setton does not explicitly or implicitly disclose that a *proxy* running on the active source 102e aggregates a media stream to be delivered to the relay/receiver 104a (*proxy ... for the non-mesh connected endpoint and connected to each of the mesh-connected endpoints*), Setton renders this obvious.

104. Third, the active sources executing the content aggregator 304 (*proxy*) aggregate the media streams received from the other active sources 102 (*proxy...receive[s] video feeds from the mesh-connected endpoints*) and transmit the aggregated stream to the relay/receiving peers (*... for the non-mesh connected endpoint*). Setton explains that the bi-directional arrows shown in Fig. 1 above “denote packet exchanges between the active sources (e.g., peers 102a, 102b, 102c,

102d, and 102e).” Setton, [0023]. Each active source 102 generates an aggregated media stream for the attendees of a video teleconference:

Referring still to FIG. 1 and also to 206 of FIG. 2, embodiments in accordance with the present invention utilize the first portion (i.e., active sources 102) of the peer-to-peer network to generate an aggregated media stream, wherein **the aggregated media stream is comprised of a plurality of media streams**. More specifically, in embodiments in accordance with the present invention, each active source (e.g., peers 102a, 102b, 102c, 102d, and 102e) which produces content, **is responsible for aggregating all other content with its own**. In one embodiment, such aggregation is accomplished using a mesh-based data aggregation peer-to-peer protocol. In one embodiment, the content is, for example, a **video stream and the corresponding audio stream** generated by an attendee to a video teleconference.

Setton, [0019].

105. Accordingly, to the extent Setton does not disclose that the *proxy...receive[s] video feeds from the mesh-connected endpoints*, Setton renders this obvious.

106. Thus, Setton renders this limitation obvious.

- e. **[3.3] *a server associated with the proxy for serving the film strip with frames containing reduced-size video feeds from the mesh-connected conference participants to the non-mesh connected endpoint; and***

107. The combination of Setton and Baker renders this limitation obvious.

108. As discussed in [3.1.2], Baker discloses “a strip of miniaturized remote participant viewing areas” (*film strip with frames containing reduced-size video feeds from conference participants*).

109. As discussed in [3.2], the content aggregator 304 (*proxy*) of Setton’s active source 102e (*selected mesh-connected endpoint*) aggregates the media video streams from other active sources 102 (*mesh-connected endpoints*) and transmits the aggregated stream to all relay/receiving peers (*serv[es] the video media stream from the mesh-connected conference participants to the non-mesh connected endpoint*). A POSITA would have understood that active sources 102 (*mesh-connected endpoints*) host conference participants (*mesh-connected conference participants*). See Setton, [0014].

110. Setton explains that the peers, such as the active sources, may include, in addition to the content aggregator (*associated with the proxy*), a feature “for providing content distribution which establishes a bridge between active sources 102 and relay/receivers 104.” Setton, [0021]. The content distribution feature performing the transmitting/distribution of the aggregated stream is the *server*.

111. Further, it would have been obvious to a POSITA for active source 102e to also include and run software (*server*) to perform content distribution of Baker's film strip from the active sources 102 to the relay/receivers (*...for serving the film strip with frames containing reduced-size video feeds from the mesh-connected conference participants to the non-mesh connected endpoint*). For instance, Setton contemplates that any of the active sources 102 can incorporate registration server functionality:

Although **registration server 106** is shown as a separate and distinct device in FIG. 1, other embodiments in accordance with the present invention incorporate the functionality of registration server 106 into one or more members of the peer-to-peer network. For example, embodiments in accordance with the present invention **incorporate the functionality of server 106 into any member or members of active sources 102** (e.g., peers 102a, 102b, 102c, 102d, and **102e**)....

Setton, [0027], [0021].

112. Setton further explains that its peer-to-peer media streaming technology can operate in any computer system, including server devices:

It is appreciated that system 400 of FIG. 4 is exemplary only and that the present technology for peer-to-peer media streaming **can operate on or within a number of different computer systems** including general purpose

networked computer systems, embedded computer systems, routers, switches, **server devices**, client devices, various intermediate devices/nodes, stand alone computer systems, and the like.

Setton, [0032].

113. Accordingly, it would have been obvious to a POSITA for the active source 102e to incorporate content distribution software (*a server*) to handle the delivery of the aggregated streams to the other peers, such as the relay/receiver peers. The '099 patent states that its Star Server 91, while shown as a separate computer, “could be a **program running on a computer** within one of the mesh endpoints e.g. 11 in Fig. 1.” '099 patent, 17:3-5. Thus, implementing the content aggregator (server) as software is consistent with the '099 patent's *server*.

114. Further, it was well-known to a POSITA to run multiple software servers and functions on a single device such that active source 102e would be able to accomplish its already stated function of aggregating video media streams for delivery to the relay/receiver peers. Accordingly, it would have been obvious to a POSITA for the same active source 102e running the server software (*server*) to also operate the content aggregator 304 (...*associated with the proxy*). It was well-known for peers in a peer-to-peer network to perform various tasks including content distribution, and a POSITA would have easily recognized that implementing the server and proxy in one active source would have simplified the

infrastructure and allow for more flexible network configurations and reconfigurations. *See* Setton, [0027].

115. Finally, it was generally known in the art to reduce the size of images placed into a “film strip”-style display of remote conference participants. For instance, King describes scaling different video feeds to be displayed with different sizes:

the MCU 108 may have one or more scalers 411 (e.g., four scalers) and compositors 413 to scale received video images 555 and composite two or more of the video images 555 from the endpoints 103 into, for example, a composite video image 505 (e.g. which may include one or more video images 555 in, for example, a continuous presence).

King, [0049]; *see also* Figs. 5b, 10, [0051].

116. Hwang describes, referring to Fig. 9, down-sampling certain frames (e.g., Fig. 9, users 2-6) to produce a “reduced-resolution video” (e.g., 104 x 80), while at least one user viewing area has a higher resolution (e.g., 208 x 160):

As illustrated in FIG. 9, the destination EP 920 presents the decoded bitstreams in a “5+1” split screen format, in which video content for five sources are each presented at the same resolution (e.g. 104×80 pixels), and video content for one source is presented at a larger resolution (e.g. 208×160 pixels).

...

Reduction of video resolution is achieved by either down-sampling the input video pixels or cutting/cropping out some video pixels around the picture borders.

Hwang, [0060]-[0061].

117. A POSITA would have found it obvious to serve the aggregated stream to include a reduced-size strip of video streams and at least one larger video stream, as this would reduce network load and streaming requirements for the peers in the video conference.

118. Thus, the combination of Setton and Baker render this limitation obvious.

- f. **[3.4] *a control connection permitting a user located at the non-mesh connected endpoint to select a video feed on the film strip for promotion to full size and to initiate transfer of the full size video feed to the non-mesh connected endpoint.***

119. The combination of Setton and Baker render this limitation obvious.

120. The '099 patent describes "control connections" as including "software 'handles' [which] control the connection and operation of the Mixer and Repeater respectively using well-known MCU technology." '099 patent, 16:43-46.

121. As discussed in [3.3], the content aggregator 304 a (*proxy*) of Setton's active source 102e aggregates the media video streams from other active sources 102 for distribution and transmission to all relay/receiving peers (*serv[es]* the video

media stream *from the mesh-connected conference participates to the non-mesh connected endpoint*).

122. As also discussed in [3.3], Setton explains that any of the active sources 102, including active source 102e, may function as a registration server. Setton, [0027]. The registration server assists peers in joining a teleconferencing session and to transition from a relay/receiver peer to an active source peer:

When a new peer wants to join the session, the new peer **sends a request** to registration server 106 indicating whether or not it is an active source.... Registration server 106 also assists peers, in a similar fashion, **when they transition between being a member of active sources 102 and relay/receivers 104....** Another example of a peer transitioning from one portion of the peer-to-peer network to another is **a member of the audience (typically a relay/receiver peer) joining the discussion (typically an active source peer).**

Setton, [0026].

123. Thus, Setton explains that the registration server handles requests from a member of the audience using a relay/receiver peer (*user located at the non-mesh connected endpoint*) to change their membership to an active source, and that the registration server 106 may be incorporated into one of the active sources 102, such as active source 102e. It would have been obvious to a POSITA to facilitate requests from a relay/receiver peer to change to an active source peer

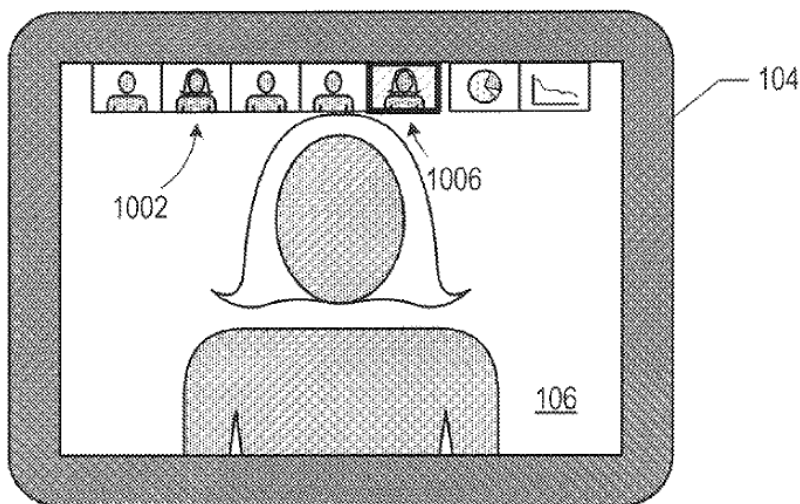
through a *control connection* between the relay/receiver peer (e.g., 104a) and the active source peer (e.g., 102e). This would allow “a listener to a teleconference who does not wish to speak (i.e., generate content), but who wishes to be ensured prompt delivery of generated content” to change from a relay/receiver 104 to an active source 102. Setton, [0017].

124. Baker discloses that a local participant can “touch or tap” a conference participant’s reduced-sized video feed on the strip (*select a video feed on the film strip*) to bring up that conference participants viewing area on the display (*for promotion to full size*).

Alternatively, **a strip of miniaturized remote participant viewing areas are located along an edge of the display with the speaking participant(s) highlighted can be included in the display 106** so the local participant knows which way to pan, title, or scroll the display 106 to see the speaking participant. FIG. 10A shows an example strip 1002 of reduced size images of the live remote participant viewing areas and two content viewing areas located along the top of the display. The participant viewing areas can be arranged in the same manner the participant viewing areas are arranged in the virtual meeting space 500. In the example of FIG. 10A, the border of the reduced viewing area 1004 of slide 1 is highlighted to indicate that slide 1 is the image shown on the display 106. **The reduced viewing area 1006 of**

participant 5 is highlighted to indicate that participant 5 is the only participant speaking. The local participant can then move the display 106 to view participant 5, as shown in FIG. 10, using at least one of the methods described above with to FIGS. 8A-8D. Alternatively, the **local participant can scroll with his finger along the strip 1002 to participant 5's viewing area to move participant 5's viewing area into the display 106, or the local participant can touch or tap on the highlighted image 1006 of participant 5 to bring up participant 5's viewing area in the display 106.**

Baker at 8:40-62.



Baker, Fig. 10B

125. It would have been obvious to a POSITA for the local participant touching or tapping Baker's strip to be facilitated by a *control connection* configured as software defining the relevant triggers (e.g., touch or tap on display)

and the responses (e.g., promote stream to main display area) for the triggers.

Sending touchscreen commands to control videoconference displays was known in the art. For instance, U.S. Publication No. 2010/0333004 to Kristiansen et al.

(“Kristiansen,” Ex.1020) describes using touch screen gestures to resize areas of a screen, and transmitting the selection as a command indicating a “layout configuration” to a Video Composing Server (VCS):

the user may resize the graphical objects. The resizing may be performed by applying **a gesture recognized by the computer 43**, e.g. by applying a pinching movement of two or more fingers while continuously applying pressure to **the touch screen** over a graphical object, as illustrated in FIG. 5 a.

...

when the user has modified the graphical representation on the touch screen by manipulating the graphical object(s) according to one or more of the embodiments above, **the computer 43 transmits a command** (or signal) **comprising a layout configuration** describing the modified graphical representation **to the VCS 26** via the port 47, the communication link 56 and a port 28 on the VCS 26. In response to the received layout configuration from the computer 43, the VCS 26 modifies the composition of the composite video signal according to

the layout configuration, and hence in accordance to the modified graphical representation on the touch screen.

Kristiansen, [0037], [0039].

126. In the combination of Setton and Baker, the act of touching or tapping controls what content is displayed on the local participant's display and, as a result, triggers the full size video feed to be transmitted to and displayed on the local participant's device. This type of software-based control connection is consistent with the '099 patent's description of a control connection. *See* '099 patent, 16:43-46. Baker's touch or tap features to enlarge a display would give attendees control over the organization of the display, and to select a video stream to enlarge in the main display area.

127. In Baker, all conference participants are displayed in the film strip. This is consistent with the '099 patent's display. *See* '099 patent, 5:42-47 ("Showing all sources from all distant end point(s) all of the time, even if they have also been promoted to a larger size, is a key differentiator from prior art endpoints (FIG. 3 described above) in which promotion means that a smaller video is removed on promotion."). Indeed, as depicted in Baker's Fig. 10B, the strip miniaturization of participant 5 remains in the strip even when participant 5 is selected and moved to the center of display 106.

128. *Third*, a POSITA would have found it obvious for the conference participant's viewing area to be displayed at full size when that video feed is

selected to be brought up on the display (*initiate transfer of the full size video feed*). Baker explains how a reduced viewing area can be moved to a main display area:

The reduced viewing area 1006 of participant 5 is highlighted to indicate that participant 5 is the only participant speaking. The local participant can then move the display 106 to view participant 5, as shown in FIG. 10.

Baker, 8:52-57.

Methods include determining that participant 3 is the only video stream that to be [*sic*] streamed to the PED 104. **Also, the video of non-speaking participants can be sent with a lower bit-rate/quality, based on the assumption that the participants are less likely to be looking at the non-speaking participants. ... In addition, the video streams in the strip of miniaturized images 1002 shown in FIG. 10 can be decoded at a slower frame or can be still images that are periodically updated.**

Baker, 9:9-26.

Alternatively, only the video streams of one or more speaking participants are received by the PED, or the **video streams of non-speaking participants are received with a lower bit-rate and quality than the video of streams of the one or more speaking participants.**

Baker, 9:45-49.

129. It would have been obvious to a POSITA to implement Baker's strip display and selection process into Setton's registration server, content aggregation, and content distribution functionalities. As explained above in [3.3], it would have been obvious for Setton's active source to provide "reduced-size video feeds" for the videos in the "film strip," since this would reduce network load and streaming requirements at the peers. It would have been obvious to incorporate Baker's display control features so that the user at the peer could select one video stream they prefer to be displayed in a larger size or full size.

130. Setton already contemplates similar control functionalities. For instance, Setton discloses aggregating and transmitting the aggregated video media stream from active sources 102 (*mesh-connected endpoints*) to relay/receivers 104 (*non-mesh connected endpoints*). Setton also contemplates that users of the videoconferencing technology can control their membership (*e.g.*, from relay/receiver to active source) via their respective peer. Setton, [0020]. Setton also explains that the peers can store and run applications, and include input devices such as keyboards to provide "command selections." Setton, Fig. 4, [0033].

131. *See* Section IX.C. A POSITA would have recognized the advantages of using Baker's miniaturized streams of non-speaking participants, and only displaying full size videos of speaking a participant. This would have been

beneficial in Setton's topology because Setton's active sources 102 have high streaming quality requirements which are not imposed on Setton's relay/receivers 104:

Typically, **the active source portion** of the peer-to-peer network is limited to a small subgroup of **peers whose constraints (e.g. delay requirements) are most stringent**. The relay/receiver portion of the peer-to-peer network is comprised of peers who do not suffer from stringent constraints.

Setton, [0014].

132. A POSITA would have understood that allowing a non-mesh participant to select a video feed to promote to full size would allow the participants flexibility to focus on a speaker or shared content, such as slide presentations, as needed. In addition, allowing a full size video helps to reduce visual complexity for larger meetings with many participants, and increases accessibility of the video conference platform by enabling participants to modify the display to meet their needs. For example, participants may view particular feeds in full size, such as the video feed of an interpreter. Thus, Setton and Baker render obvious using touch or tap controls to select a video stream on a miniaturized film strip to be promoted to full size in the main viewing area (*select*

a video feed on the film strip for promotion to full size and to initiate transfer of the full size video feed to the non-mesh connected endpoint).

133. Accordingly, the combination of Setton and Baker render this limitation obvious.

2. **Claim 10**

- a. ***[10.0] An endpoint for use as a mesh-connected endpoint in a teleconferencing system comprising a plurality of endpoints, at least some of which are connected in a mesh configuration and at least one of which is connected in a non-mesh configuration to a selected one of the mesh-connected endpoints, wherein a film strip with frames containing reduced-size video feeds from the conference participants is displayed at the mesh endpoints, the mesh-connected endpoint comprising:***

134. This limitation is substantively identical to limitations [3.0], [3.1.1], and [3.1.2] and is rendered obvious for the same reasons. Setton's active source 102e is the *endpoint for use as a mesh-connected endpoint*.

- b. ***[10.1] a proxy for connection to each of the mesh-connected endpoints to receive video feeds from the mesh-connected endpoints for the non-mesh connected endpoint;***

135. This limitation is substantively identical to limitation [3.2] and is rendered obvious for the same reasons.

- c. ***[10.2] a server associated with the proxy for serving the film strip with frames containing reduced-size video***

feeds from the mesh-connected conference participants to the non-mesh connected endpoint; and

136. This limitation is substantively identical to limitation [3.3] and is rendered obvious for the same reasons.

- d. [10.3] ***a proxy controller permitting a user located at the non-mesh connected endpoint to select a video feed on the film strip for promotion to full size and to initiate transfer of the full size video feed to the non-mesh connected endpoint.***

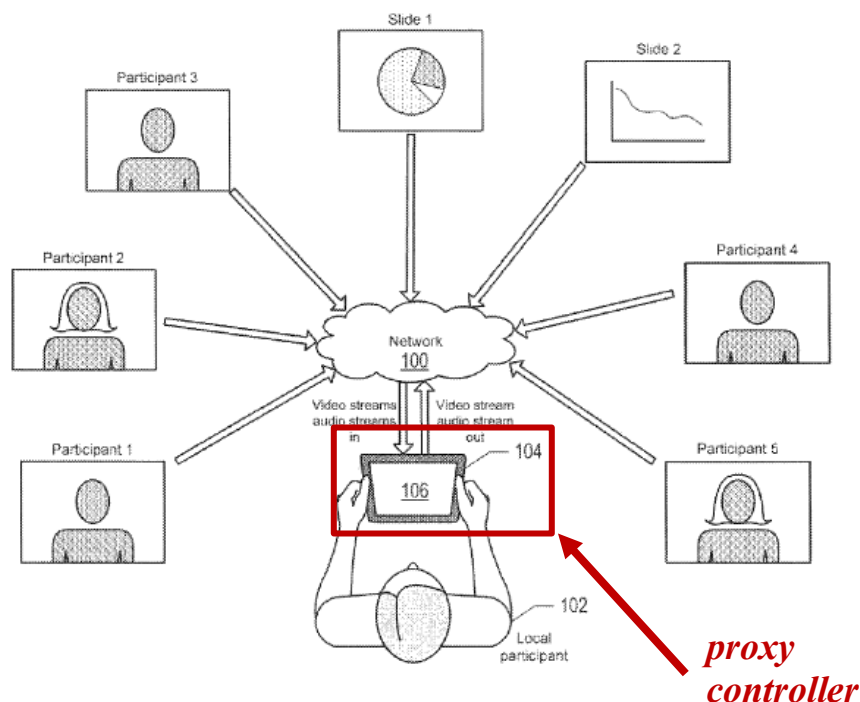
137. The combination of Setton and Baker renders this limitation obvious.

138. As explained in [3.2], Setton's active sources 102 (*mesh-connected endpoint*) include "a content aggregator 304 for aggregating the content generated by the peer with content generated by other peers" (*proxy*).

139. As explained in [3.4], it would have been obvious to a POSITA that Baker's "touch or tap" functionality would be facilitated by a *control connection* between the active source 102e and the relay/receiver 104a because the act of touching or tapping controls what content is displayed on the local participant's display and, as a result, triggers the full size video feed to be transmitted to and displayed on the local participant's device. To the extent that *proxy controller* is interpreted to be the same as the *control connection*, Setton and Baker render this limitation obvious. See Section IX.C.

140. However, if the *proxy controller* of [10.3] is interpreted to be distinct from the *control connection*, Setton and Baker also render this feature obvious.

Baker discloses a “PED [portable electronic device] 104” which is the claimed *proxy controller*.



Baker, Fig. 1 (annotated)

141. The user of PED 104 selects which video stream to be displayed in full size on the display 106 using touch commands. Baker, 8:52-57. A POSITA would have understood that these touch commands cause the full size video stream to be transmitted from active source 102e (*mesh connected endpoint*) to the relay/receiver 104 (*non-mesh connected endpoint*) by active source 102e’s content aggregator (*proxy*). Accordingly, it would have been obvious to a POSITA that Baker’s “touch or tap” functionality would be implemented by the PED 104 (*proxy controller*), which facilitates touch inputs to control Baker’s graphical user interface elements.

142. Thus, the combination of Setton and Baker renders this limitation obvious.

X. THE COMBINATION OF SETTON, BAKER, AND WONG RENDERS OBVIOUS CLAIMS 4-6

A. Summary of Wong

143. Wong relates to “digital multimedia communications over telecommunications networks.” Wong, [0002]. Wong and the ’099 patent are from the same field of endeavor of video teleconferencing. *See* ’099 patent, 15:50-53; Wong, [0013], [0100], [0041]. Wong addresses the particular problem of the ’099 patent, which includes using a web browser interface to display and control a video teleconference. *See* ’099 patent, 15:50-53; Wong, [0013], [0041].

144. The multimedia services described by Wong include video streaming services:

...Video Share/CSI (Combined Circuit Switched and IMS) augmentation and enhancement, user experience enhancement, **Video Share casting, Video Share blogging**, video share customer service, interworking between various access technologies and methods, mobile to web services, live web portal (LWP), **video callback service, and the like.**

Wong, [0006].

145. Wong's video casting and sharing services include streaming video to a web browser:

[T]he extended video share casting service further comprises streaming video casting from a first group to a first video portal, linking the first video portal to a web-portal page, and **streaming the first video portal to a web-browser** through a flash proxy component.

Wong, [0010].

[A] method of streaming one or more **group video castings** to one or more video portals, linking the one or more video portals to a web server, and **streaming the one or more video portals to a web-browser** accessing the web server via a proxy of web-browser plug-in media.

Wong, [0013].

146. Automatically-downloaded plug-ins for the web browsers can show video casts once a thumbnail is selected for viewing:

A web page can show thumbnails of live video casts that a user can navigate to. The user can click on a thumbnail to view that particular blog or cast. **The web browser can automatically download a plug-in that can implement the multimedia communication to show the user the blog or the video cast.** The plug-in can use an Adobe Flash approach or an ActiveX approach, or more generally a software program or script that can **execute within the**

browser or in the PC and show the user blog or live cast.

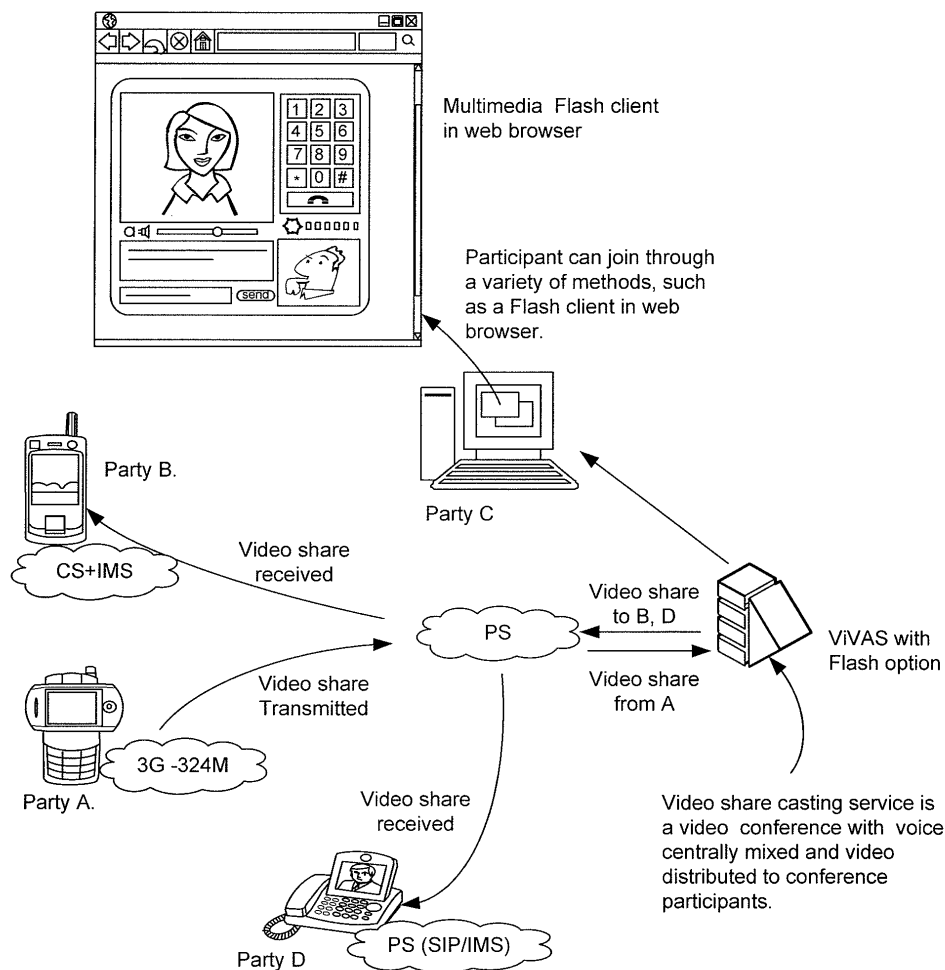
Wong, [0041].

147. The video share casting services include conference calls and chats:

The video share casting can be further extended to a multi-casting service from single casting service in a **conferencing call or chat**. For example, multiple users can broadcast their video, and other users can select the cast to view. On selection of another user who is not currently broadcasting his video, an avatar may be automatically played.

Wong, [0100].

148. As shown in Fig. 8 below, a video share casting service allows a user (e.g., Party C) to view a video stream such as a conference call via a web browser with a downloaded plug-in:



Wong, Fig. 8

B. Reasons to Combine Wong with Setton and Baker

149. The combination of Wong with Setton and Baker would have been combining prior art elements (Wong's web browser the peer-to-peer network of Setton and Baker) according to known methods (computer programming) to yield predictable results. *See* Section IX.C (describing reasons to combine Setton and Baker).

150. Web browsers, including the web browsers disclosed in Wong, were well-known and commonly used applications for accessing the internet and for participating in a variety of internet-based communications. *See* TechCrunch (Ex.1015); PCWorld (describing Google Hangouts and Facebook Skype services that use web browsers for video calls) (Ex.1016). Setton describes the peers in its network as internet-connected devices. Setton, [0014]. For instance, Setton describes that the active sources have “internet protocol (IP) addresses.” Setton, [0014]. It’s peers aggregate and distribute content, including “a video stream and the corresponding audio stream generated by an attendee to a video teleconference.” Setton, [0014]. The distributed audience (e.g., at relay/receivers 104) can be addressed “over a network such as the Internet.” Setton, [0019]. Further, Baker explains that a “computer-readable medium 214 can store software applications, including word processors, **browsers**, email. Instant Messaging, media players, and telephone software.” Baker, 3:30-34. Wong describes the use of web browsers for accessing and controlling video share casting services. Wong, [0013], [0041], [0106], [0107]. Using the web browser disclosed by Wong would have been a natural choice as a platform for a videoconferencing application. A POSITA would have recognized that the videoconferencing display of Setton and Baker could be implemented within a browser as evidenced by the Google

Hangouts and Facebook Skype services, and would have also recognized the benefits of doing so.

151. The combination would be achieved by routine computer programming techniques, including web application programming. A POSITA would have used standardized web app programming languages and practices to create the interface and functionality involved with Setton's and Baker's videoconferencing systems and methods.

152. The results of the combination would have been predictable. Browsers and web-based programming would have been familiar to a POSITA at the time, and web browsers provided ample functionality for videoconferencing developers to receive and display video feeds, and to allow users to control various aspects of the video conference, including the relative sizes of the video feeds on the display, how many video feeds are shown on the screen, and how to arrange them. User interface (UI) objects including buttons, scroll bars, and clickable areas were also well-known and commonly used in web-based applications.

153. Further, the combination would be applying a known technique (Wong's web browser) to a known device ready for improvement (the combined peer-to-peer network system of Setton and Baker) to yield predictable results. While Baker describes a miniaturized film strip and tap or touch functionality controlling it (Baker, Figs. 10A, 10B, 8:40-62), Setton and Baker provide limited

details on what specific type of application or platform is used. Wong specifically describes using a web browser to view and participate in video share casting (Wong, [0041]) and that video share casting includes video conferences:

The video share casting can be further extended to a multi-casting service from single casting service in a **conferencing call or chat**.

Wong, [0100].

154. Wong also explains that a viewer of a video conference can “select the cast to view.” Wong, [0100].

155. The results of the combination would have been predictable for the same reasons explained above.

156. Further, a POSITA would have been motivated to combine Wong with Setton and Baker. Web browsers were, and still are, a popular platform for application developers of all types, and in particular for communications application developers. Because browsers execute standardized web-based program code (e.g., HTML), a web application can be executed for a variety of operating systems and hardware configurations. This makes the service accessible to a wide array of users without the need for specialized, platform-specific software.

157. Finally, a POSITA would have had a reasonable expectation of success in combining Wong’s web browser-based video share casting technique

with Setton's peer-to-peer network and Baker's film strip. A POSITA would have had the skills and general knowledge (e.g., computer programming and user interface design) to implement a web browser-based videoconferencing application. Setton provides limited detail on what type of software (e.g., programming language) or platforms could be used to view and control the conference. Baker describes storing "software applications including...browsers." Baker, 3:29-34, but otherwise does not go into specifics on the software implementation. However, a POSITA would have had this requisite background knowledge, and Wong simply provides implementation details that Setton and Baker expect. Modifying Setton and Baker to serve Baker's film strip to a web browser, as explained in Wong, installed at Setton's peers would have been within a POSITA's skill set. Wong, [0013], [0041], [0100]. Further, a POSITA would have understood that using well-known and commonly-used web browsers to access the videoconferencing service would have make the service available to more people.

C. Detailed Analysis of Claims

1. Claim 4

- a. [4.0] *A teleconferencing system as claimed in claim 3, wherein the server is configured to serve the filmstrip to a browser co-located with the non-mesh connected endpoint.*

158. The combination of Setton, Baker, and Wong renders this claim obvious.

159. As discussed in [3.3], it would have been obvious to a POSITA that active source 102e would include and run server software (*server*) to distribute aggregated video streams to the video conference participants. *See* Setton, [0014], [0021].

160. *First*, it would have been obvious for the server software running on Setton's active source 102e (*server*) to serve Baker's strip of video media to a *browser* on Setton's relay/receiver 104 (*non-mesh connected endpoint*). Setton discloses that its system may be used to stream both video and web page-based data.

Although the embodiments of the present invention are described with respect to the streaming of **video data**, it should be understood that embodiments of the present invention are not limited to the streaming of video data. It should be appreciated that embodiments of the present invention also apply to other types of media, including without limitation: audio-based data, image-based data,

graphics data, video-based data, text-based data, **web page-based data**, and the like.

Setton, [0014].

161. Baker discloses that its strip of video feeds is generally displayed in a video conference application having a “virtual meeting space”:

The **video conference application** 228 provides various Software components for sending and receiving video and audio signals with one or more remote participants and creating, navigating, and **viewing portions of a virtual meeting space**, as described below.

Baker, 3:50-54.

162. It would have been obvious to a POSITA that the video conference application could operate within a *browser*, such that the film strip is served from the active source 102e’s content distribution feature (*server*) to the *browser*. It was common to access and control video conferences via internet browsers. For instance, U.S. Patent No. 9,009,225 to Kjesbu et al. (Ex.1013), discloses a “Video Conferencing Management Program” (VCMP) that “normally operates with an intuitive web interface requiring no additional installation on the user terminal other than a conventional web browser.” Kjesbu, 5:3-17 *See also*, U.S. Patent No. 7,877,489 to Salesky et al. (Ex.1014), 2:21-29.

163. By the priority date of the ’099 patent, there were multiple browser-based video conferencing services available. Facebook’s Messenger Skype service,

a partnership between Facebook and Skype, permitted users to access a video call through a web browser. TechCrunch (Facebook's Skype-powered service "will be an in browser experience") (Ex.1015); PCWorld (Ex.1016). Similarly, Google Hangouts was accessed through a web browser by installing the appropriate plugin. TechCrunch; PCWorld.

164. Wong describes a video streaming system that includes "streaming the one or more video portals to a web-browser accessing the web server via a proxy of web-browser plug-in media." Wong, [0013]. For instance, "[t]he web browser can automatically download a plug-in that can implement the multimedia communication to **show the user the blog or the video cast.**" Wong, [0041]. Wong's browser-based video streaming technology can be used to join "a conferencing call or chat." Wong, [0100].

165. It would have been obvious to a POSITA for the relay/receivers 104 to use Wong's web browser to participate in the video conferencing system taught by the combination of Setton and Baker. Using the browser software to connect to and control the videoconference display would eliminate the need for participants to download and install specialized programs at each endpoint. Wong, [0103] ("Users who have only web browser also can join the casting through flash proxy servers. **Most PC web browsers have the adobe flash plug-in installed.**"). Internet browser applications are compatible with a variety of platforms, operating

systems, and hardware configurations. Internet browsers rely on standardized web programming protocols to display content and provide user interaction. Section X.B.

166. Further, a POSITA would have found it obvious for the browser to be installed and running on the relay/receiver 104 (*co-located with the non-mesh connected endpoint*). Setton explains that the relay/receivers 104 may be audience members, and may be much more numerous than the active speakers connected in the mesh-based portion of the peer-to-peer network. Baker specifically contemplates the storage of “software applications including ... browsers.” Baker, 3:29-34. It was well-known that Internet browsers installed at the PC (*co-located*) were one of the principal mechanisms for computer users to communicate over the internet.

167. A POSITA would have been motivated to implement the video conferencing system taught by the combination of Setton and Baker using a web browser because it would eliminate the need for participants to download and install specialized programs at each endpoint. Rather, participants could view and control the video conferencing session using an existing browser application increasing the flexibility and adoption rate of the system.

168. In addition, configuring the video conferencing application to be hosted on a browser advantageously promotes cross-platform compatibility,

supporting multiple operating systems. It also reduces storage requirements for individual participants by not requiring the installation of specialized software. This also eliminates the need for software updates to continue using the platform. Users are typically more familiar with web browsers as well, facilitating ease of use for participants.

169. *Second*, it would have been obvious to a POSITA that the browser would be installed and running on the relay/receiver 104 (*co-located with the non-mesh connected endpoint*). Baker specifically contemplates the storage of browser software applications in a computer-readable medium:

For example, the computer-readable medium 214 can be non-volatile memory, such as flash memory, an optical or a magnetic disk; and includes volatile memory. The computer-readable medium 214 can store **software applications**, including word processors, **browsers**, email. Instant Messaging, media players, and telephone software.

Baker, 3:29-34.

170. Setton explains that the peers can be internet-connected devices. Setton, [0014], [0034]. It was well-known that Internet browsers installed at the PC (*co-located*) were, and still are, one of the principal mechanisms for PC users to access content on the internet, including web applications. A POSITA would have found it obvious for the relay/receiver 104a, which may support an audience

member viewing the video conference, to include a browser that receives and displays the video feed (*a browser co-located with the non-mesh connected endpoint*)

171. Thus, the combination of Setton, Baker, and Wong render this claim obvious.

2. **Claim 5**

- a. **[5.0] *A teleconferencing system as claimed in claim 4, wherein the browser is displayed on a computer co-located with the non-mesh connected endpoint.***

172. The combination of Setton, Baker, and Wong render this claim obvious.

173. As explained in [3.1.1], Setton's relay/receivers 104 include computers.

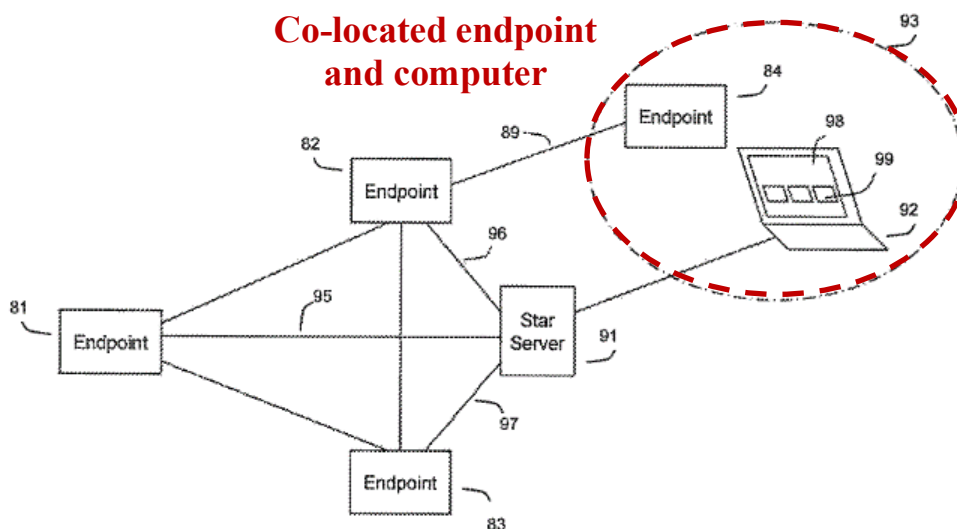
174. According to the '099 patent, and with reference to Fig. 9, an endpoint may be implemented and run on a computer, and also be "co-located" with that computer:

The computer 92 and star endpoint 84 are co-located 93 i.e. a user or users at 93 can view and access both devices.

...

Computer 92 could be any computer for example a laptop of Smart mobile phone (iPhone or IPad). Furthermore **End point 84 function may run on that same computer 92.**

'099 patent, 16:9-21.



'099 patent, Fig. 9 (annotated)

175. As explained in [4.0], the combination of Setton, Baker, and Wong renders obvious Setton's relay/receiver 104a (*non-mesh connected endpoint*) executing a *browser* to display Baker's strip of video media (*co-located*). A POSITA would have understood, in light of the '099 patent's description of co-located endpoints and computer, that the relay/receiver 104a comprises an endpoint that is also co-located with a computer. *See* '099 patent, 16:9-21.

176. It was well known that browsers provided a graphical user interface (*browser is displayed...*) for a user to view and interact with the content, including teleconferencing communications.

177. A POSITA would have therefore understood that Setton's relay/receiver 104a includes a computer co-located with the endpoint, and that the

computer running the browser would include the browser being *displayed on a computer co-located with* the relay/receiver 104a.

178. Thus, the combination of Setton, Baker, and Wong render this claim obvious.

3. Claim 6

- a. [6.0] *A teleconferencing system as claimed in claim 5, wherein the proxy and server are merged into a virtualized environment.*

179. The combination of Setton, Baker, and Wong render this claim obvious.

180. As explained above at [3.2] and [3.3], it would have been obvious to implement the proxy and server in active source 102e. A POSITA would have been motivated to implement the proxy and server at the active source 102e as a virtual machine or virtual server (*merged into a virtualized environment*). Virtual servers were well-known and commonly used at the time of the '099 patent. Virtual servers can be instantiated in an existing computer or server by software, and use existing computing resources to perform tasks. For instance, U.S. Patent No. 9,124,757 to Weber ("Weber") shows that a POSITA would have understood that server functions could be implemented using virtual servers and virtualized instances of hosts:

It will be apparent that the components portrayed in this figure can be arbitrarily combined or divided into separate

software, firmware and/or hardware components. Furthermore, it will also be apparent that such components, regardless of how they are combined or divided, can execute on the same host or multiple hosts or **several virtualized instances on one or more hosts**, and wherein the multiple hosts can be connected by one or more networks geographically distributed anywhere in the world.

Weber, 4:7-15.

...[A] computing device can be...a server machine that is a physical or **virtual server** and hosted in an Internet public or private data center by a service provider or a third party for the service provider....

Weber, 4:40-46.

181. A POSITA would have been motivated to use a virtual server to implement the *proxy* and *server*. Setton describes combining different functionalities in one peer, for example, “...[b]y incorporating the functionality of registration server 106 into members of the peer-to-peer network, embodiments in accordance with the present invention further reduce the infrastructure associated with the peer-to-peer network configuration.” Setton, [0027]. This implementation would have been particularly beneficial in Setton’s network, as it would allow for dynamic changes and would not require the same expensive, application-specific infrastructure. Setton, [0027]. Virtual servers can be deployed more cost-

effectively than dedicated servers. A POSITA would have known that employing virtual servers to configure multiple services at the same node would help reduce infrastructure costs. Using software to instantiate a virtual server or machine at the active source 102e would have been within the skill set of a POSITA and the results would have been predictable.

182. Alternatively, it would have been obvious to a POSITA for Setton's "content aggregator 304 for aggregating the content generated by the peer with content generated by other peers" (*proxy*) and content distribution software (*server*) to be supported by or otherwise in communication with a cloud (*merged into a virtualized environment*).

183. The '099 patent discloses that a "virtualized computer environment" may include a "cloud":

Taken to the extreme, which would be understood by persons skilled in the art any or all functions constituting the Star Proxy 106 and or Star Server 91 **could be implemented in a virtual computer or computers within a virtualized computer environment known as a cloud.**

'099 patent, 17:29-33.

184. A POSITA would have found it obvious for some of the *proxy* and *server* functions to be performed on a cloud server, in coordination with the *proxy* and *server* software running on the active source 102e.

185. It was well known that computational hardware, such as computers and/or servers, may either be physical or virtual (e.g., cloud-based). U.S. Patent No. 8,601,097 to Sartori (“Sartori”) explains that, in place of a physical server, a “cloud server system 220 runs virtualization software to provide a virtual machine.” Sartori, 5:50-52.

186. Setton proposes to “reduce the infrastructure” required for media streaming. Setton, [0027]. As explained above, the *proxy* and *server* would be implemented as software, which can be run locally on the active source 102e. It would have been obvious to a POSITA for a cloud server to support the *proxy* and *server* functions by handling administrative tasks and user data, such as authorization, registration, and meeting configurations. A POSITA would have understood that merging the *proxy* and *server* into a *virtualized environment* through the cloud in this way would achieve Setton’s objective of reducing infrastructure.

187. Further, having a cloud server (*virtualized environment*) perform computational and administrative services in a conference would also provide the added benefit of freeing up computational resources of the active sources to meet its higher video streaming quality demands. For example, an objective of Setton is to “accommodate the different quality of service requirements of the peers” including providing low latency to the active sources 102. Setton, [0028], [0014].

Offloading the tasks of the proxy and server from the active source 102 to a virtualized environment would advantageously free up resources of the active source 102. Further, it would have been obvious for the cloud server to provide common access to non-streamed content and resources, such as files or presentation graphics (*see* Baker, Fig. 10A) that can be accessed by the conference participants individually as desired.

188. Thus, the combination of Setton, Baker, and Wong render this claim obvious.

XI. THE COMBINATION OF APOSTOLOPOULOS AND BAKER RENDERS OBVIOUS CLAIMS 3 AND 10

A. Summary of Apostolopoulos

189. Apostolopoulos relates to a “communication network” that allows “communication between devices and nodes” in, for example, an “N-way collaborative video conference.”

FIG. 1 is a diagram of a [*sic*] exemplary **communication network 100 that is capable of supporting a communicative environment** (e.g., an N-way collaborative virtual environment), in accordance with one embodiment of the present invention. As shown in FIG. 1, the **network 140 facilitates communication between devices and nodes that are communicatively coupled with the network 140.** Within the network 140,

communication traffic is transmitted through various devices, such as, routers and/or switches.... Embodiments of the present invention are well suited to application within a class of communication systems and networks that allow multiple numbers of users or participants to interact in a communicative environment, e.g., the **N-way collaborative virtual environment**.

Apostolopoulos, [0021].

While the present embodiment is described within the context of an **N-way collaborative environment** (e.g., an **N-way collaborative video conference**), other embodiments are well suited to other environments that provide for interaction between multiple collaborators within the virtual environment.

Apostolopoulos, [0032].

190. Apostolopoulos and the '099 patent are from the same field of endeavor or video teleconferencing systems. *See* '099 patent, Abstract, 15:42-45.; Apostolopoulos, [0021], [0033], [0025]. Apostolopoulos also addresses the particular problem of the '099 patent, which includes video teleconferencing in mixed topology networks. *See* '099 patent, 15:46-49; Apostolopoulos, [0021], [0033], [0025], [0052], [0069], [0070], Fig. 3C.

191. In Apostolopoulos, the nodes are each associated with a collaborator and may produce audio and/or video data streams for the other collaborators in the conference.

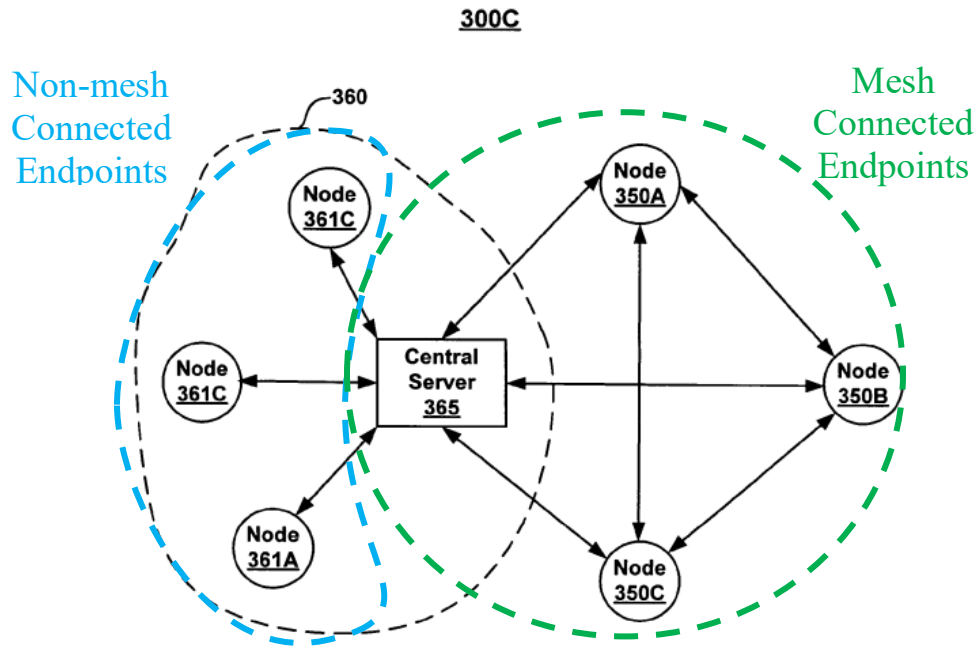
Within the N-way collaborative environment, a collaborator is associated with each of the nodes 120A-N in the communication network 100. **Each of the collaborators at each node interacts with the remaining collaborators in order to participate within the N-way collaborative virtual environment.**

Apostolopoulos, [0024].

In the communicative environment, **each participant can possibly transmit data streams (e.g., audio and video data streams)** to some or all of the other participants.

Apostolopoulos, [0033]; *see also* [0025].

192. Apostolopoulos describes multiple topologies for the communication among the nodes. Apostolopoulos, [0052] (“There are a number of topologies possible for supporting N to N communication.”). For example, the nodes may communicate in a “hybrid topology,” which combines features of a “full-mesh topology” and a “star topology,” as shown in Fig. 3C below.



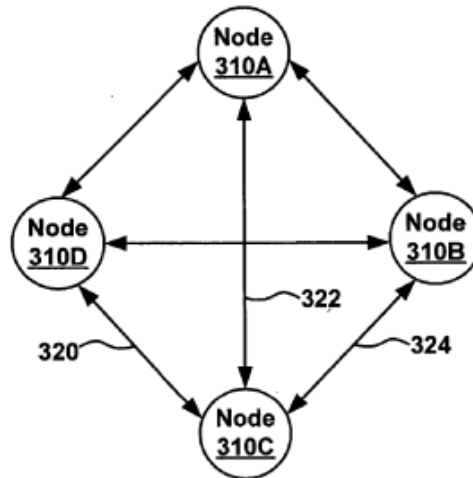
Apostolopoulos, Fig.3A (annotated)

FIG. 3C is a diagram illustrating the virtual representation of communication paths in a hybrid communication topology 300C, in accordance with one embodiment of the present invention. **The hybrid topology 300C combines features of the full-mesh topology of FIG. 3A and the star topology of FIG. 3B.** That is, the hybrid topology 300 comprises the simultaneous use of a central node/server in a star topology 360 connected to other nodes that are running a full-mesh topology. The full-mesh topology comprises node 350A, node 350B, node 350C, as well as central server 365 of the star topology 360.

Apostolopoulos, [0069].

193. In a “full-mesh topology,” each node is directly connected to each other node, for example, as shown in Fig. 3A below.

300A



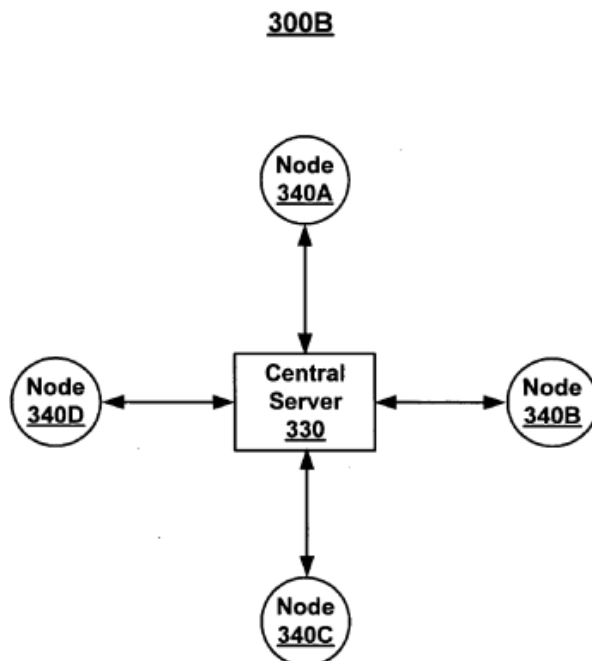
Apostolopoulos, Fig. 3A

In a full-mesh topology, **a plurality of communication paths couple each of the plurality of collaborators together** for transmitting and receiving the view dependent image streams for each of the plurality of collaborators.

Apostolopoulos, [0054]

194. On the other hand, in a “star topology,” each node is not directly connected to each other node. Apostolopoulos, [0056]. Instead, each “node” is

connected to a “central node/server” that routes data streams to each other node,
for example, as shown in Fig. 3B.



Apostolopoulos, Fig. 3B

The communication topology 300B comprises a central node/server 330 that is coupled to each of the N collaborators participating in an N-way collaborative environment (e.g., a communicative environment). The star topology 300B comprises the central server 330 for receiving and sending communication traffic between the plurality of nodes, and N connections coupling the plurality of nodes to the central server 330.

Apostolopoulos, [0056].

195. Thus, in the hybrid topology, the central node/server of the star topology is mesh-connected to the other nodes in the full-mesh topology and is connected in a star topology to the nodes in the star topology.

In the hybrid topology 300C of FIG. 3C, a nodes 361A-C are communicatively coupled to the central node 365. The central node is capable of streaming communication traffic to a group of other collaborators that communicate amongst themselves. That is, the collaborators 350A-C along with the central node/server 365 communicate in a full-mesh topology. For example, the star topology 360 may comprise a group of low-power cellphone clients at nodes 361A-C. In this case, the central server/node 365 provides the processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.

Apostolopoulos, [0070].

196. The functionality of central node/server in the hybrid topology may be implemented on a stand-alone server or may be integrated into a node in the full-mesh configuration.

The nodes within the communication network 100 can produce data streams for some or all of the other nodes within the communication network 100.

Apostolopoulos, [0025].

The star topology 300B comprises the **central server 330 for receiving and sending communication traffic between the plurality of nodes**, and N connections coupling the plurality of nodes to the central server 330. Specifically, the **central node/server 330** is communicatively coupled to collaborators 340A, 340B, 340C, and 340D.

Apostolopoulos, [0056].

For instance, the **central server/node 330** is capable of sharing some or most of the computational tasks for each of the collaborators at the nodes 340A-D.

Apostolopoulos, [0058].

In this case, the **central server/node 365** provides the processing power for **generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

For instance, collaborators 371A-N are communicatively coupled to the central **server/node 375** of the star topology 370; collaborators 381A-N are communicatively coupled to the **central server/node 385** of the star topology 380; and collaborators 391A-N are communicatively coupled to the **central server/node 395** of the star topology 390.

In addition, the **central nodes 375, 385, and 395 are interconnected via a full-mesh topology.**

Apostolopoulos, [0072].

In another embodiment, in a communication topology that includes a **central server/node** for processing communication data streams, confidentiality can be introduced.

Apostolopoulos, [0088].

B. Reasons to Combine Apostolopoulos with Baker

197. A POSITA would have been motivated to combine Baker's multi-participant display and display control techniques with Apostolopoulos' communication network. The incorporation of Baker's features into Apostolopoulos' network would be simply combining prior art elements (Baker's "strip of miniaturized remote participant viewing areas" and "touch or tap" control interface with Apostolopoulos' central node 365) according to known methods to yield predictable results. The methods for combining Apostolopoulos and Baker would involve well-known, routine software application programming to specify how the video streams are presented on screen, and to provide user interface objects (e.g., buttons, clickable areas) that would modify the presentation of the video streams to enlarge at least one of the video streams. This type of software programming, including defining user interface visual objects and their associated actions, would have been within the skill set of a POSITA. The results would have

been predictable, since the additional software features simply modify how the content is displayed (e.g., the size of each video feed on the screen), and when (e.g., in response to a touch or tap).

198. The combination of Baker and Apostolopoulos would be applying a known technique (Baker's video strip and display control features) to a known device (Apostolopoulos' central node 365) ready for improvement to yield predictable results. Apostolopoulos describes a dynamic, hybrid arrangement of full-mesh nodes and star-connected nodes that allows for an exchange of communication data streams between the full-mesh portion and star portion of the network, such as in a video conference. But Apostolopoulos does not go into great detail about the format or arrangement of the video conferencing content displayed by the nodes.

199. Baker, however, discloses one possible implementation of Apostolopoulos' network. In Baker, multiple video conferencing stream are aggregated and displayed in a single strip of multiple video streams, with some of the video streams having a reduced size. In this way, Baker's features would further define the functionalities of Apostolopoulos' nodes in a way that furthers Apostolopoulos' purposes, such as reducing network load, accommodating network and power constraints of different participants, and reducing delay. Apostolopoulos, [0039], [0040], [0048]. As explained above, incorporating these

functionalities of Baker into Apostolopoulos' system would involve routine software programming that would have been well within the skill set of a POSITA. Thus, the results would have been predictable.

200. A POSITA would be motivated to provide the film strip of reduced-size video feeds to allow for some video feeds deemed more “important” (e.g., active speaker) to be prioritized while operating within network and power use limitations. *See, e.g.*, Apostolopoulos [0082], [0083]. A POSITA would also be motivated to provide the film strip of reduced-size video feeds because the film-strip advantageously allows participants to simultaneously focus on the speaking participant, while also monitoring all participants at once for various purposes, such as documenting attendance or monitoring engagement, making the video conferencing system disclosed more suitable for a wide variety of applications, including education, corporate environments, or social settings. Thus, the combination of Apostolopoulos and Baker render this limitation obvious.

201. A POSITA would have had a reasonable expectation of success by combining Apostolopoulos and Baker. Incorporating Baker's user interface elements (strip of miniaturized images and touch or tap functionality) into Apostolopoulos' communication network and its nodes would involve straightforward computer programming techniques that a POSITA would be capable of implementing. Apostolopoulos provides limited detail on how to display and

control the aggregated video streams. However, a POSITA would have had this requisite background knowledge, and Baker simply provides implementation details that Apostolopoulos expects. The implementation details of Baker include well-known video layouts and UI objects (e.g., buttons, clickable areas) that control how the video streams are served and presented on screen. Baker, Figs. 10A, 10B, 8:40-62. These UI objects were common and familiar to both application developers and users. Thus, a POSITA would have a reasonable expectation of success in implementing the features of Baker into Apostolopoulos' communication network.

C. Detailed Analysis of Claims

1. Claim 3

a. [3.0] *A teleconferencing system comprising:*

202. To the extent the preamble is limiting, Apostolopoulos discloses it.

203. Apostolopoulos discloses a “communication network” that allows “communication between devices and nodes that are communicatively coupled with the network”:

FIG. 1 is a diagram of a [sic] exemplary **communication network 100 that is capable of supporting a communicative environment** (e.g., an N-way collaborative virtual environment), in accordance with one embodiment of the present invention. As shown in FIG. 1, the **network 140 facilitates communication between**

devices and nodes that are communicatively coupled with the network 140. ... Embodiments of the present invention are well suited to application within a class of communication systems and networks that allow multiple numbers of users or participants to interact in a communicative environment, e.g., the **N-way collaborative virtual environment.**

Apostolopoulos, [0021].

204. The network may be used for “N-way collaborative video conferenc[ing]”:

While the present embodiment is described within the context of an **N-way collaborative environment** (e.g., an **N-way collaborative video conference**), other embodiments are well suited to other environments that provide for interaction between multiple collaborators within the virtual environment.

Apostolopoulos, [0032].

205. The participants, via the nodes, produce and transmit video data streams to the other participants/nodes:

In the communicative environment, **each participant can possibly transmit data streams (e.g., audio and video data streams)** to some or all of the other participants.

Apostolopoulos, [0033], [0025].

206. Apostolopoulos' "communication network," including "nodes," teaches the claimed "*teleconferencing system*."

207. Thus, Apostolopoulos discloses this limitation.

- b. [3.1.1] *a plurality of endpoints, at least some of which are connected in a mesh configuration and at least one of which is connected in a non-mesh configuration to a selected one of the mesh-connected endpoints, and*

208. *First*, Apostolopoulos discloses multiple "nodes" 350A, 350B, 350C, 365 (*endpoints*) and 361C and 361A (*endpoints*) that use its network. Each "node" within Apostolopoulos' network can produce a "data stream" including "audio and video data streams."

The nodes within the communication network 100 **can produce data streams for some or all of the other nodes** within the communication network 100.

Apostolopoulos, [0025].

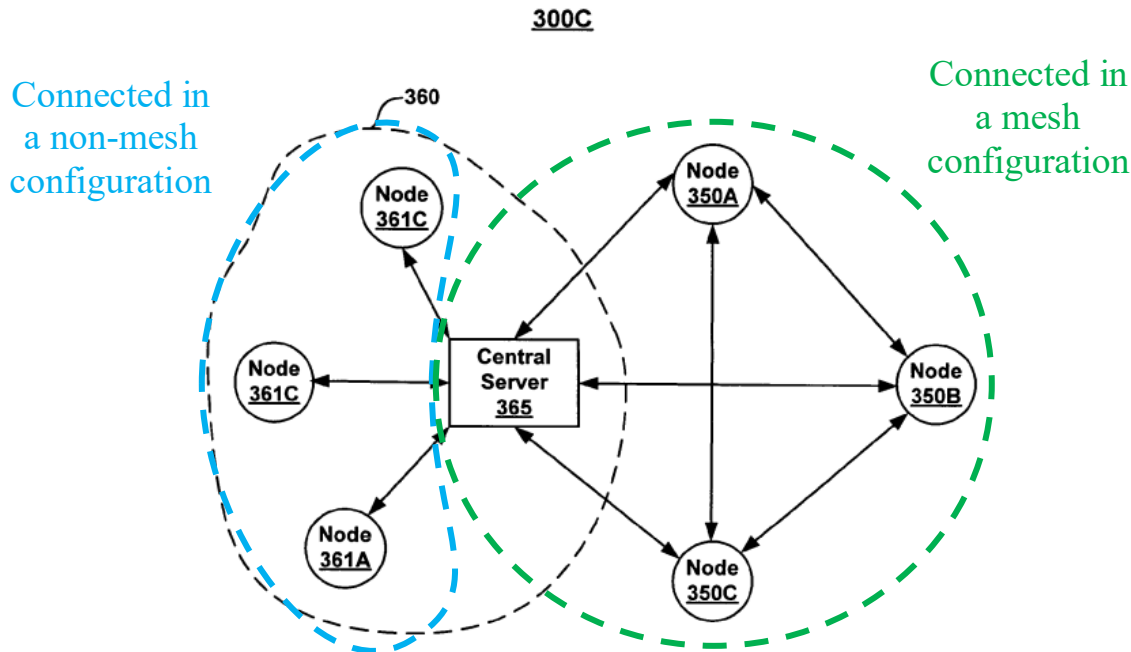
In the communicative environment, **each participant can possibly transmit data streams (e.g., audio and video data streams)** to some or all of the other participants.

Apostolopoulos, [0033].

209. The nodes (350A, 350B, 350C, 365, 361C, and 361A) are the *plurality of endpoints*. This is consistent with the '099 patent, which similarly describes endpoints as a single device or "collection of devices." '099 patent, 5:4-7.

210. *Second*, Apostolopoulos' full-mesh topology nodes 350A, 350B, 350C, 365 (endpoints) are connected in a mesh configuration and, the star topology nodes 361C and 361A (endpoints) are connected in a non-mesh configuration.

211. In Fig. 3C, Apostolopoulos shows a hybrid topology 300C that “combines features of the full-mesh topology of Fig. 3A and the star topology of Fig. 3B.” Apostolopoulos, [0069].



Apostolopoulos, Fig. 3C (annotated)

212. In Apostolopoulos' full-mesh topology, each “node” is directly connected to each other node:

In a full-mesh topology, **a plurality of communication paths couple each of the plurality of collaborators together** for transmitting and receiving the view

dependent image streams for each of the plurality of collaborators.

Apostolopoulos, [0054].

213. Apostolopoulos' description of how its nodes in the full-mesh topology communicate is consistent with the '099 patent's description of a "mesh configuration," which is one where "each endpoint is connected directly to each other endpoint in the conference." '099 patent, 2:34-37.

214. In a star topology, Apostolopoulos explains, individual nodes are not directly connected to each other:

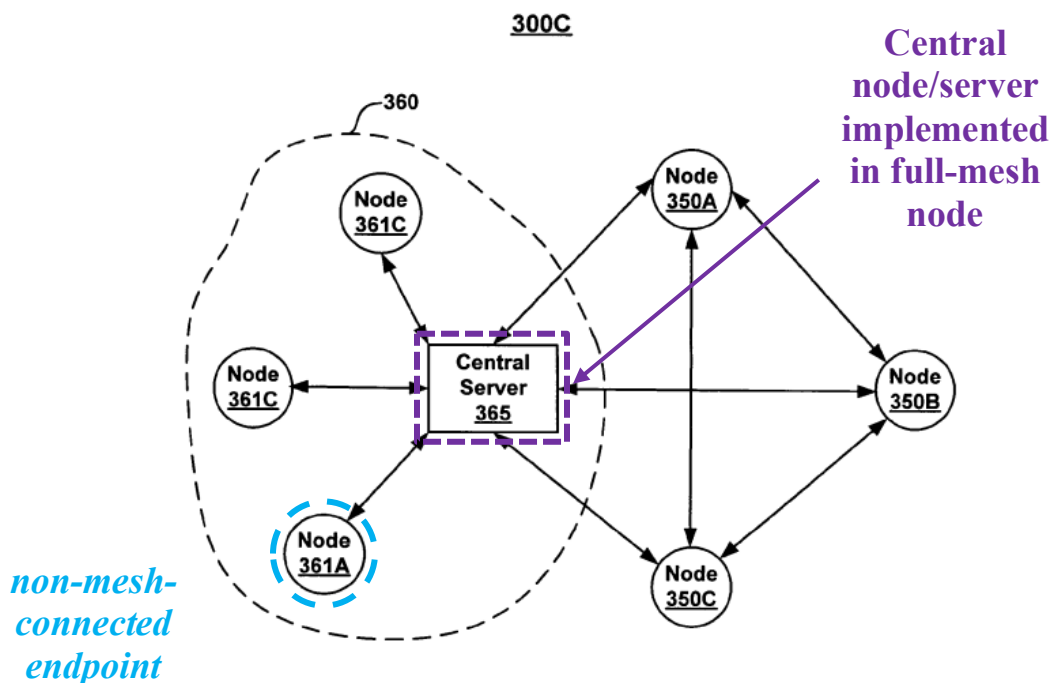
The communication topology 300B comprises a central node/server 330 that is coupled to each of the N collaborators participating in an N-way collaborative environment (e.g., a communicative environment). The star topology 300B comprises the central server 330 for receiving and sending communication traffic between the plurality of nodes, and N connections coupling the plurality of nodes to the central server 330.

Apostolopoulos, [0056].

215. Apostolopoulos' star topology corresponds to the '099 patent's non-mesh configuration, where each node is connected to a "central server" that routes data streams to each node, rather than being directly connected to each of the other nodes.

216. As depicted in Fig. 3C, nodes 350A, 350B, 350C, 365 (*endpoints*) are connected in a mesh configuration and nodes 361C and 361A (*endpoints*) are connected in a non-mesh configuration.

217. **Third**, Apostolopoulos renders obvious that at least one of the star topology nodes 316 is connected in a non-mesh configuration to a selected one of the mesh-connected endpoints. Fig. 3C of Apostolopoulos shows that node 361A, for example, is connected to central node/server 365, which is *mesh-connected* to the rest of the full-mesh nodes 350.



218. Apostolopoulos refers to the central server 365 as a “central **node** 365,” and states that the central node/server is part of the “full-mesh topology”:

The full-mesh topology comprises node 350A, node 350B, node 350C, **as well as central server 365** of the star topology 360.

In the hybrid topology 300C of FIG. 3C, a nodes 361A-C are communicatively coupled to **the central node 365**.

Apostolopoulos, [0069], [0070].

219. To the extent that Apostolopoulos does not disclose, explicitly or implicitly, that the central node/server 365 is a *mesh-connected endpoint*, a POSITA would have found it obvious to implement the central node/server as a full-mesh connected node (e.g., similar or identical to nodes 350A-350C). The central server 365 could be implemented by server software to perform server functionalities. Apostolopoulos explains that the central node/server 365 shares the same computational tasks for each of the collaborator nodes. Apostolopoulos, [0058]. Indeed, all nodes (and participants) in its topologies generate audio and video data streams:

The nodes within the communication network 100 **can produce data streams for some or all of the other nodes** within the communication network 100.

Apostolopoulos, [0025].

In the communicative environment, **each participant can possibly transmit data streams (e.g., audio and video data streams)** to some or all of the other participants.

Apostolopoulos, [0033].

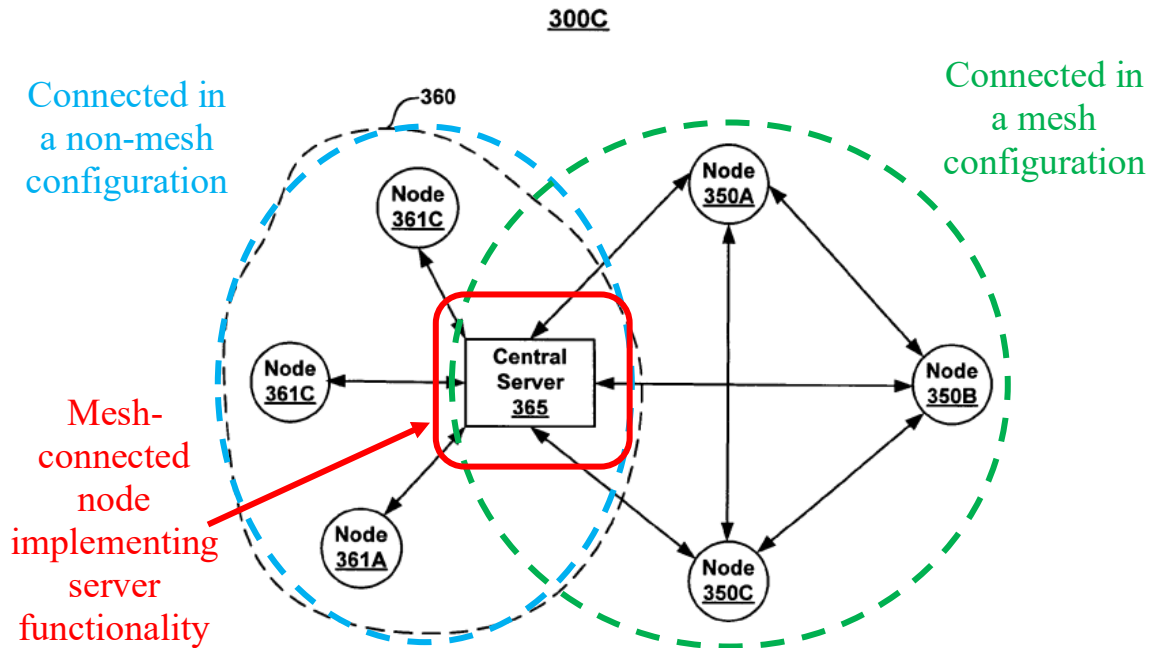
220. Consistent with the other nodes, central server 365 is also described as generating data streams. Apostolopoulos explains that “central server/node 365” “generat[es] **communication data streams** for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.” Apostolopoulos, [0070]. A stated goal of Apostolopoulos is to facilitate dynamic topology changes based on the number of users:

In another embodiment, as the communication topology is adapted and configured according to performance metrics, the locations where computation is performed within the communication network shifts around. For example, when **the topology shifts from a decentralized topology (e.g., mesh topology) to a more centrally located topology (e.g., star topology)**, much of the communication shifts from the end-user nodes to the central node. **As a result, the computation is moved dynamically, as necessary, to meet various constraints, such as, computation, bandwidth, or low-power constraints.**

Apostolopoulos, [0048].

221. In view of this disclosure, it would have been obvious to a POSITA to implement Apostolopoulos’ “central server 365” as a full-mesh-connected node, similar to nodes 350A-350C, that performs the server functionality. In this configuration, Apostolopoulos’ nodes 361 would be *connected in a non-mesh*

configuration to a full-mesh node implementing the central server 365 (a selected one of the mesh-connected).



Apostolopoulos, Fig. 3C (annotated)

222. Implementing Apostolopoulos’ “central server 365” as a node, similar to the other end nodes 350 in the hybrid topology 300C would further Apostolopoulos’ stated goal of facilitating dynamic topology changes, as the function and role of the central node/server 365 would more flexibly support different network topologies and needs. Further, implementing server functionality into an endpoint was also well-known in the art. For example, Baker discloses a “centralized computing device 302” that receives and transmits video and audio streams. The “centralized computing device 302” can be “operated by one of the remote participants” (i.e. by an *endpoint* or “node”):

Each remote participant has video conferencing equipment that includes a camera, a microphone, a display, and a **computing device that runs a client-side video conferencing application that sends video and audio data streams generated by the participant to a centralized computing device 302 via the network 100.** The **computing device 302 can be operated by one of the remote participants** or can be located at a separate location. The device 302 receives the video and audio streams from each of the participants and includes a server-side video conferencing application that processes each of the video and audio streams supplied by the participants. The server-side application sends to each participant the video and audio streams of the other participants and sends any shared content to the participants.

Baker, 3:58-4:4.

223. Similarly, Kenoyer (U.S. Patent No. 7,692,683) discloses “video conferencing endpoints” that couple to a “Multipoint Control Unit (MCU),” which transmits and receives the data streams from the “video conferencing endpoints.” The MCU “may be embedded in one of the video conferencing endpoints”:

FIG. 2 illustrates an embodiment of multiple endpoints supporting various resolutions coupled to a **Multipoint Control Unit (MCU) 201**. In some embodiments, video conferencing endpoints 203 a-203 c may transmit high

resolution images captured through their high definition cameras 205 a-205 c. In some embodiments, **one or more lower resolution endpoints (e.g., endpoint 203 d)** with, for example, a standard definition camera 207 **may also be coupled to the MCU 201**. In some embodiments, **the MCU 201 may be a separate device or may be embedded in one of the video conferencing endpoints**.

Kenoyer, 3:55–65.

224. Further, it would have been obvious for the mesh node implementing the central server 365 to be *selected* as the node that provides the connection point between the full-mesh nodes 350 and the star nodes 361. Apostolopoulos' network topology is "dynamically selected":

A network topology is dynamically **selected** that most efficiently allows communication between the collaborators in the N-way collaborative environment.

Apostolopoulos, [0052].

225. A POSITA would have found it obvious for a network administrator, for example, to select the full-mesh node implementing the central server to be part of the dynamic network topology selection process.

226. For these reasons, Apostolopoulos renders this limitation obvious.

c. **[3.1.2] *which are configured such that a film strip with frames containing reduced-size video feeds from the***

conference participants is displayed at the mesh endpoints;

227. The combination of Apostolopoulos and Baker renders this limitation obvious.

228. Apostolopoulos describes generating and transmitting “video streams” of participants for “an N-way collaborative video conference.” Apostolopoulos, [0032], [0026]-[0028]. Apostolopoulos does not explain the details of how the compressed video signal is presented at the receiving nodes.

229. As explained in Ground 1, [3.1.2], Baker’s “strip of miniaturized remote participant viewing areas” displayed on PED’s display is the claimed *film strip with frames containing reduced-size video feeds from the conference participants*.

230. It would have been obvious to a POSITA to display Baker’s film strip on the display of one of Apostolopoulos’ nodes (*a film strip with frames containing reduced-size video feeds from the conference participants is displayed*). As explained, above displaying some remote participants with reduced-size video feeds while others are displayed at a larger size, or full sized, was generally known in the art. *See, e.g.*, Hwang, [0060], Figs. 9, 17A-18B; Pepperell, [0047], Figs. 4 and 9; King, Fig. 10. The ’099 patent acknowledges that videoconferencing displays that have a combination of large viewing areas and reduced-size viewing areas are “prior art.” *See* ’099 patent, Figs. 3b-3c, 4:42.

231. Apostolopoulos' goal is to offer "low complexity" video collaboration where a central server/node compresses a video signal from each collaborator and then sends it to other nodes:

In addition, the synthetic N-way collaborative environment for each collaborator can be generated by the central node 330, and compressed as a video signal. This compressed video signal is then sent to the corresponding collaborative node. In this case, the collaborative node only needs a simple video decoder, leading once again to a **very low complexity** collaborative node.

Apostolopoulos, [0062]

232. Apostolopoulos has the stated goals to accommodate communication bandwidth constraints, and to generate a collaborative visual environment for all of the participants. *See* Apostolopoulos, [0037], [0052], Further Apostolopoulos contemplates prioritizing "important" people in the conference (e.g., person speaking), which may include displaying some objects/participants but not others, and providing higher resolution images for objects/people with higher importance:

In still another embodiment, the communication topology that is selected is capable of prioritizing data traffic based on an importance of objects associated with the data traffic. As such, data traffic in the communication topology comprising objects with higher importance have priority over data traffic for objects with lower

importance. For example, given a number of objects, for example **N head-and-shoulder objects when N people are collaborating, it is often the case that a subset of the objects are more important at a given time as compared to the others.** In another example, **the person who is speaking may be identified as having higher importance than someone in the back row who is simply listening.**

In another embodiment, resolution of the objects associated with the data traffic is varied as a function of the importance. That is, **data traffic for objects with higher importance have higher resolution than data traffic for objects with lower importance.**

Apostolopoulos, [0082], [0083].

233. It would have been obvious to a POSITA to present Apostolopoulos' compressed video stream in the form of Baker's strip of miniaturized remote participant viewing areas at the receiving nodes to address bandwidth and processing constraints while prioritizing important objects/speakers in the conference. *See* Section XI.B.

234. Thus, the combination of Apostolopoulos and Baker renders this limitation obvious.

- d. ***[3.2] a proxy at the selected mesh-connected endpoint for the non-mesh connected endpoint and connected to each of the mesh-connected endpoints to receive video***

feeds from the mesh-connected endpoints for the non-mesh connected endpoint;

235. Apostolopoulos renders obvious this limitation.

236. *First*, as discussed in [3.1.1], it would have been obvious Apostolopoulos' central node/server 365 to be implemented in a full-mesh node (*the selected mesh-connected endpoint*), similar to the nodes 350A-350C, where the central node performs sever functionality in the hybrid topology 300C.

237. *Second*, it would have been obvious to a POSITA for the server functionality of the central node/server 365 (*at the selected mesh-connected endpoint*) to be implemented by software that functions a *proxy* for at least one of the star topology nodes 361 (*for the non-mesh connected endpoint*).

Apostolopoulos explains that part of the function of a central server is to “**receive[]** and **send[]** communication traffic between the plurality of nodes associated with the plurality of collaborators that are participating in the N-way collaborative environment.” Apostolopoulos, [0057]. The central node/server 365 generates data streams “for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.” And, as discussed in [3.1.1], “central server 365” is connected to and communicates with each of the nodes 350A-C (*mesh-connected endpoints*) and communicates with the connected nodes 361A-C (*non-mesh connected endpoints*).

In this case, the **central server/node 365** provides the **processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

238. Apostolopoulos, FIG. 3D illustrates multiple nodes 371, 381, and 391 connected in non-mesh configurations respectively to servers 375, 385, and 395. These non-mesh configurations operate in a similar way to the nodes 361 and server 365 described:

For instance, collaborators 371A-N are communicatively coupled to the central **server/node 375** of the star topology 370; collaborators 381A-N are communicatively coupled to the **central server/node 385** of the star topology 380; and collaborators 391A-N are communicatively coupled to the **central server/node 395** of the star topology 390. In addition, the **central nodes 375, 385, and 395 are interconnected via a full-mesh topology.**

Apostolopoulos, [0072].

239. Apostolopoulos also explains that a “central node” can assume processing and computation functions on behalf of other end-nodes after a topology change:

As such, computations may be performed at a different location within the same topology. For example, within a star topology, **some computation may be changed from an end-node to a central node**. This occurs without changing the star topology, and corresponds to adapting the operation of the system.

Apostolopoulos, [0051].

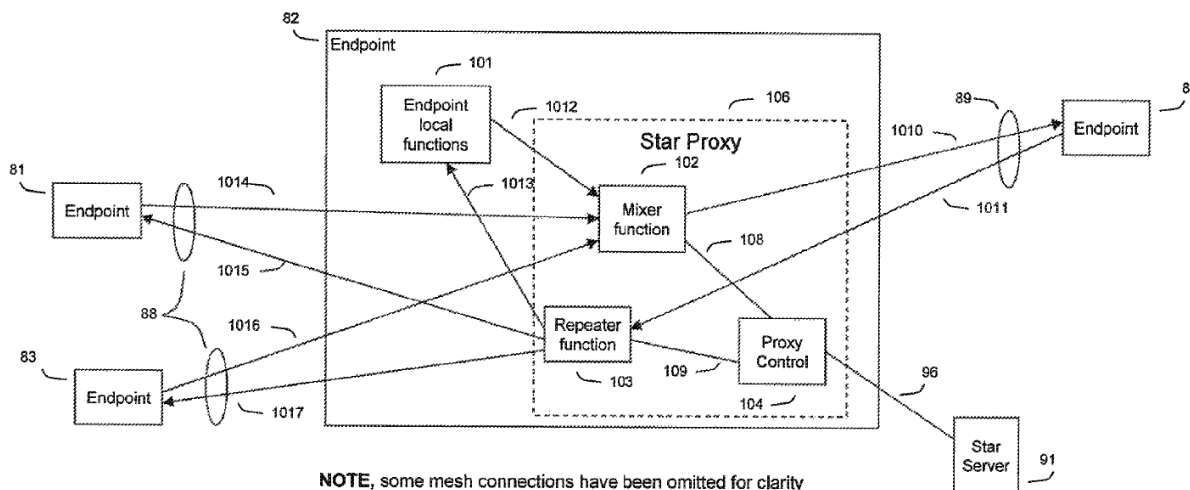
240. Apostolopoulos' central server functionalities, including the receiving and processing functionalities, are consistent with the '099 patent's description of a *proxy*. As explained further below, claim 3 of the '099 patent specifies that the *proxy* is provided to *receive video feeds from the mesh-connected endpoints for the non-mesh connected endpoint*. Similarly, the specification of the '099 patent explains, with respect to the embodiment of Fig. 10, that a proxy includes "well-known" software running on a mesh-connected endpoint to mix and repeat communications between endpoints:

Referring to FIG. 10, a mesh endpoint 82 is shown, it has been adapted to include **a Star Proxy function, similar to an embedded MCU**. Connections such as 96 might terminate in the Proxy Control function 104.

Once an endpoint, in this case endpoint 82, has been designated as a proxy for the star endpoint 84 it will instantiate local Mixer and Repeater functions which will act as a mesh proxy for the star endpoint. Control

connections 108 and 109, for example software ‘handles’, control the connection and operation of the Mixer and Repeater respectively using **well-known MCU technology**.

’099 patent, 16:36-46.



’099 patent, Fig. 10

241. The ’099 patent further explains that the MCU “consolidates” video feeds from the various end points in the conference:

Multi-party videoconferences are typically hosted by a Multiparty Control Unit (MCU) also known as a Multipoint Conference Unit. **The MCU consolidates all the video feeds from the end points** and presents appropriate video images to each of the end points of all the video camera and collaboration material video feeds in the conference.

’099 patent, 1:34-39.

242. It would have been obvious to a POSITA for the mesh node implementing the central server 365 (*selected mesh-connected endpoint*) to run software that performs the functionalities of a “central server,” including receiving and consolidating video feeds from the participants in the N-way conference. The receiving and consolidating functionalities running on the central node 365 would be a *proxy at the selected mesh-connected endpoint*.

243. Third, central server receiving functionality (*proxy*) of the central node 365 “stream[s] communication traffic to a group of other collaborators.” Apostolopoulos, [0070]. It would have been obvious to a POSITA, based on this disclosure, that central node 365’s functionalities would include receiving video streams from the full-mesh-connected collaborators (*proxy...receive[s] video feeds from the mesh-connected endpoints*). The central server 365 generates data streams, which include audio and video streams, for nodes 350A-C (*mesh-connected endpoints*) and nodes 361A-C (*non-mesh connected endpoints*) (*proxy...receive[s] video feeds from the mesh-connected endpoints for the non-mesh connected endpoints*):

In the hybrid topology 300C of FIG. 3C, a [sic] nodes 361A-C are communicatively coupled to the central node 365. **The central node is capable of streaming communication traffic to a group of other collaborators that communicate amongst themselves.**

That is, the collaborators 350A-C along with the central node/server 365 communicate in a full-mesh topology. For example, the star topology 360 may comprise a group of low-power cellphone clients at nodes 361A-C. In this case, **the central server/node 365 provides the processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

244. It would have been obvious to a POSITA that the central server 365 would first receive the video streams in order to transmit the video streams to the other nodes. Thus, Apostolopoulos renders obvious this limitation.

- e. **[3.3] *a server associated with the proxy for serving the film strip with frames containing reduced-size video feeds from the mesh-connected conference participants to the non-mesh connected endpoint; and***

245. The combination of Apostolopoulos and Baker render this limitation obvious.

246. As discussed in [3.1.2], Baker discloses “a strip of miniaturized remote participant viewing areas” (*film strip with frames containing reduced-size video feeds from conference participants*).

247. First, as discussed in [3.2], it would have been obvious for the mesh node implementing the central server 365 (*selected mesh-connected endpoint*) to

receive video streams from other collaborator nodes 350A-C (*mesh-connected endpoints*) and to transmit the video stream to nodes 361A-C (*serv[e] ... video feeds from the mesh-connected conference participants to the non-mesh connected endpoint*). A POSITA would have understood that nodes 350 (*mesh-connected endpoints*) host *conference participants*.

248. The '099 patent itself describes the reduced-size video feeds as “typical prior art screen layouts.” *See* '099 patent, Figs. 3b-3c, 4:42. A POSITA also would have found it obvious to combine Baker’s miniaturized images with Apostolopoulos’ system because showing miniaturized images of videoconferencing participants was a common UI technique.

249. It would have been obvious to mix the received videos such that at least some of the video feeds of the conference participants (e.g., audience members) have a “*reduced size*.” As explained above, displays that include a combination of full size viewing areas and reduced-size viewing areas were known in the art. *See, e.g.,* Hwang, [0060], Figs. 9, 17A-18B; King, Fig. 10.

250. *Second*, Apostolopoulos describes that the central node 365 takes on some of the processing tasks for generating and streaming communication traffic:

In the hybrid topology 300C of FIG. 3C, a nodes 361A-C are communicatively coupled to the central node 365. **The central node is capable of streaming communication traffic to a group of other collaborators that**

communicate amongst themselves. That is, the collaborators 350A-C along with the central node/server 365 communicate in a full-mesh topology. For example, the star topology 360 may comprise a group of low-power cellphone clients at nodes 361A-C. In this case, **the central server/node 365 provides the processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

251. A POSITA would have understood that the portion of the central server/node 365 (e.g., software module, communication port) that streams the video streams to the other nodes includes a *server*.

252. Additionally, it would have been obvious to a POSITA for the central node/server 365 to include a *server*. In addition to *proxy* software explained above, it would have been obvious for the central node 365 to also incorporate server software (*server*) to transmit the resulting video stream(s) to nodes 361A-C (*serv[e] ... video feeds from the mesh-connected conference participants to the non-mesh connected endpoint*). Apostolopoulos explains that the central server/node 365 provides processing power for video streams to the other nodes:

For example, the star topology 360 may comprise a group of low-power cellphone clients at nodes 361A-C. In this

case, **the central server/node 365 provides the processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

253. Thus, Apostolopoulos discloses that the central node 365 comprises a *server*. [0023 seems to say so clearly]

254. Additionally, it would have been well-known and obvious to a POSITA to run a software server (*server*) on the central node 365 to transmit streams to the nodes 361A-C. A POSITA would have understood that software could be installed on a networked computer system, such as the central node/server 365, to control how, when, and to whom the video streams are transmitted. Including server software at the central node 365 would have been a straightforward and obvious approach for distributing the video streams to the star nodes 361, which are coupled to the mesh nodes 350 via the central node 365. Further, the '099 patent itself acknowledges that the star server “could be a program running on a computer within one of the mesh endpoints e.g. 11 in Fig. 1.” '099 patent, 17:3-5.

255. Thus, it would have been obvious for the central node to include software for both the proxy and server functions as explained above (*server*

associated with the proxy). As explained above at [3.1.2], it would have also been obvious A POSITA would have understood that the server of Apostolopoulos' central node/server 365 would facilitate streaming of the video streams "for the collaborators at nodes 361A-C ... as well as with the collaborators at nodes 350A-C" (*for serving the film strip with frames containing reduced-size video feeds from the mesh-connected conference participants to the non-mesh connected endpoint*) Apostolopoulos, [0070].

256. Thus, the combination of Apostolopoulos and Baker render this limitation obvious.

- f. **[3.4]** *a control connection permitting a user located at the non-mesh connected endpoint to select a video feed on the film strip for promotion to full size and to initiate transfer of the full size video feed to the non-mesh connected endpoint.*

257. The combination of Apostolopoulos and Baker render this limitation obvious.

258. *First*, as discussed in [3.3], the data stream receiving and processing functionality (*proxy*) of Apostolopoulos' central node 365, which is connected to the other full-mesh nodes 350, transmits the video stream to star nodes 361 (*serv[es] the video media stream from the mesh-connected conference participants to the non-mesh connected endpoint*).

259. *Second*, as discussed in Ground 1, [3.4], Baker discloses that a local participant can “touch or tap” a conference participant’s reduced-sized video feed on the strip (*select a video feed on the film strip*) to bring up that conference participants viewing area as full sized on the display (*for promotion to full size*). Further, it was known in the art to translate users’ touch screen inputs selecting areas of a display into commands to control the display. *See* Kristiansen, [0037], [0039].

260. It would have been obvious to a POSITA that the local participant touching or tapping Baker’s strip would be facilitated by a *control connection* between the mesh node implementing the central server 365 and the star node 361 displaying the video feeds. The act of touching or tapping results in commands being sent to the server (mesh node implementing the central server 365) to control what content is displayed on the local participant’s display, including triggering a full size video feed to be served to and displayed on the local participant’s device.

261. Like the ’099 patent, all conference participants are displayed in Baker’s strip. *See* ’099 patent, 5:42-47 (“Showing all sources from all distant end point(s) all of the time, even if they have also been promoted to a larger size, is a key differentiator from prior art endpoints (FIG. 3 described above) in which promotion means that a smaller video is removed on promotion.”). Indeed, as depicted in Baker’s Fig. 10B, the strip miniaturization of participant 5 remains in

the strip even when participant 5 is selected and moved to the center of display
106.

262. *Third*, as discussed in Ground 1, [3.4], when the conference participant's viewing area is brought up on the display, the full size video feed for that participant is displayed (*initiate transfer of the full size video feed*).

263. It would have been obvious, as described above, to provide Baker's strip display and selection user interface to Apostolopoulos' nodes 361A-C. Baker's discussion of displaying reduced quality miniaturization of non-speaking participants, and only displaying full size videos of a speaking participant, would be beneficial in Apostolopoulos' topology because Apostolopoulos' nodes 361A-C "may comprise a group of low-power cellphone clients" that lack the processing power of nodes 350A-C:

In the hybrid topology 300C of FIG. 3C, a [*sic*] nodes 361A-C are communicatively coupled to the central node 365. **The central node is capable of streaming communication traffic to a group of other collaborators that communicate amongst themselves.** That is, the collaborators 350A-C along with the central node/server 365 communicate in a full-mesh topology. For example, the star topology 360 may comprise a group of **low-power cellphone clients at nodes 361A-C**. In this case, **the central server/node 365 provides the processing power for generating communication data**

streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.

Apostolopoulos, [0070]

264. Further, Apostolopoulos contemplates prioritizing “important” people in the conference (e.g., person speaking), which may include displaying some objects/participants but not others, and providing higher resolution images for objects/people with higher importance:

In still another embodiment, the communication topology that is selected is capable of prioritizing data traffic based on an importance of objects associated with the data traffic. As such, data traffic in the communication topology comprising objects with higher importance have priority over data traffic for objects with lower importance. For example, given a number of objects, for example **N head-and-shoulder objects when N people are collaborating, it is often the case that a subset of the objects are more important at a given time as compared to the others.** In another example, **the person who is speaking may be identified as having higher importance than someone in the back row who is simply listening.**

In another embodiment, resolution of the objects associated with the data traffic is varied as a function of the importance. That is, **data traffic for objects with**

higher importance have higher resolution than data traffic for objects with lower importance.

Apostolopoulos, [0082], [0083].

265. It would have been obvious to a POSITA to implement Baker's strip user interface to *initiate transfer of the full size video feed*, as this would prioritize important objects/speakers in the conference. Allowing a participant to select which video feed to promote to full size allows participants the flexibility to focus on a speaker or shared content, such as slide presentations, as needed. In addition, allowing participants to select a video for full size presentation helps to reduce visual complexity for larger meetings with many participants and increases accessibility of the video conference platform by allowing participants to modify the display to meet their needs. For example, participants may view particular feeds in full size, such as the video feed of an interpreter. *See* Apostolopoulos, [0082] ("In another example, **the person who is speaking** may be identified as having **higher importance** than someone in the back row who is simply listening ... data traffic for objects with higher importance have **higher resolution** than data traffic for objects with lower importance.")

266. Thus, the combination of Apostolopoulos and Baker render this limitation obvious.

2. Claim 10

- a. ***[10.0] An endpoint for use as a mesh-connected endpoint in a teleconferencing system comprising a plurality of endpoints, at least some of which are connected in a mesh configuration and at least one of which is connected in a non-mesh configuration to a selected one of the mesh-connected endpoints, wherein a film strip with frames containing reduced-size video feeds from the conference participants is displayed at the mesh endpoints, the mesh-connected endpoint comprising:***

267. This limitation is substantively identical to limitations [3.0], [3.1.1], and [3.1.2] and is rendered obvious for the same reasons.

- b. ***[10.1] a proxy for connection to each of the mesh-connected endpoints to receive video feeds from the mesh-connected endpoints for the non-mesh connected endpoint;***

268. This limitation is substantively identical to limitation [3.2] and is rendered obvious for the same reasons.

- c. ***[10.2] a server associated with the proxy for serving the film strip with frames containing reduced-size video feeds from the mesh-connected conference participants to the non-mesh connected endpoint; and***

269. This limitation is substantively identical to limitation [3.3] and is rendered obvious for the same reasons.

- d. ***[10.3] a proxy controller permitting a user located at the non-mesh connected endpoint to select a video feed on the film strip for promotion to full size and to initiate***

***transfer of the full size video feed to the non-mesh
connected endpoint.***

270. The combination of Apostolopoulos and Baker render this limitation obvious.

271. As explained in [3.2], Apostolopoulos' central node 365 (*mesh-connected endpoint*) has server functionality that includes receiving and processing functionality (*proxy*) and video stream transmission functionality (*server*).

272. First, as explained in Ground 1, [3.4], it would have been obvious to a POSITA for the central node to also facilitate Baker's "touch or tap" functionality to allow participants at one of the star nodes 361 to control what content is displayed, including triggering a full size video feed to be transmitted to and displayed on the local participant's device. As similarly explained in Ground 1, [10.3], if the *proxy controller* is interpreted to include the *control connection* of [3.4], Apostolopoulos and Baker render obvious the *proxy controller* for the same reasons described above in [3.4]. If the *proxy controller* is interpreted as being different from the *control connection*, the combination of Apostolopoulos and Baker also renders this feature obvious as explained below.

273. The central node 365 integrates the server functionalities explained above for receiving, processing, and transmitting video feeds to the conference participants. Fig. 10B of Baker shows that the strip miniaturization of participant 5 remains in the strip even when participant 5 is selected and moved to the center of

display 106. It would have been obvious to a POSITA for these server functionalities to also include functionalities for receiving the “touch or tap” commands (*proxy controller*) from the conference participants to trigger the changes to the film strip display. Apostolopoulos’ nodes 361A-C “may comprise a group of **low-power** cellphone clients” that lack the processing capabilities of nodes 350A-C and 365. Apostolopoulos, [0070]. It would have been obvious for the central node 365 to assume the “touch or tap” command processing functions (*proxy controller*) to alleviate processing demands at the nodes 361A-C, which may have power or processing limitations.

274. *Second*, as discussed in Ground 3, [3.4], when the conference participant’s viewing area is brought up on the display, the full size video feed for that participant is displayed (*initiate transfer of the full size video feed*). It would have been obvious, as described above, to provide Baker’s strip display and selection user interface to Apostolopoulos’ nodes 361A-C. Baker’s discussion of displaying reduced quality miniaturization of non-speaking participants, and only displaying full size videos of a speaking participant, would be beneficial in Apostolopoulos’ topology because Apostolopoulos’ nodes 361A-C “may comprise a group of low-power cellphone clients” that lack the processing power of nodes 350A-C. Apostolopoulos, [0070]

275. Further, Apostolopoulos contemplates prioritizing “important” people in the conference (e.g., person speaking), which may include displaying some objects/participants but not others, and providing higher resolution images for objects/people with higher importance:

In still another embodiment, the communication topology that is selected is capable of prioritizing data traffic based on an importance of objects associated with the data traffic. As such, data traffic in the communication topology comprising objects with higher importance have priority over data traffic for objects with lower importance. For example, given a number of objects, for example **N head-and-shoulder objects when N people are collaborating, it is often the case that a subset of the objects are more important at a given time as compared to the others.** In another example, **the person who is speaking may be identified as having higher importance than someone in the back row who is simply listening.**

In another embodiment, resolution of the objects associated with the data traffic is varied as a function of the importance. That is, **data traffic for objects with higher importance have higher resolution than data traffic for objects with lower importance.**

Apostolopoulos, [0082], [0083].

276. It would have been obvious to a POSITA to implement Baker's film strip user interface to *initiate transfer of the full size video feed*, as this would prioritize important objects/speakers in the conference. Allowing a participant to select which video feed to promote to full size allows participants the flexibility to focus on a speaker or shared content, such as slide presentations, as needed. In addition, allowing participants to select a video for full size presentation helps to reduce visual complexity for larger meetings with many participants and increases accessibility of the video conference platform by allowing participants to modify the display to meet their needs. For example, participants may view particular feeds in full size, such as the video feed of an interpreter.

277. Thus, the combination of Apostolopoulos and Baker render this limitation obvious.

XII. THE COMBINATION OF APOSTOLOPOULOS, BAKER, AND WONG RENDERS OBVIOUS CLAIMS 4-6

A. Reasons to Combine Wong with Apostolopoulos and Baker

278. The combination of Wong with Apostolopoulos and Baker would have been combining prior art elements (Wong's web browser the communication network of Apostolopoulos and Baker) according to known methods (computer programming) to yield predictable results. *See* Section XI.B (describing reasons to combine Apostolopoulos and Baker). The incorporation of Wong's web browser software into the communication network taught by the combination of

Apostolopoulos and Baker could be achieved in a similar way, and for similar reasons, as explained with respect to the combination of Setton and Baker described above. *See* Section X.B (describing reasons to combine Wong with Setton and Baker).

279. As explained above, Web browsers, like Wong's web browser, were well-known and commonly used for internet-based communication applications. *See* TechCrunch (Ex.1015); PCWorld (describing Google Hangouts and Facebook Skype services that use web browsers for video calls) (Ex.1016). Using the web browser disclosed by Wong would have been a natural choice as a platform for a videoconferencing application.

280. The combination would be achieved by routine computer programming techniques, including web application programming (e.g., HTML) to create the interface and functionality involved with Apostolopoulos' and Baker's communication network and nodes (e.g., central node 365).

281. The results of the combination would have been predictable, since a POSITA would have been very familiar with web browsers and web application development. A POSITA would have web application development skills and capabilities to create UI objects, specify how video streams are displayed, and how the display may have been controlled.

282. Further, combining Wong with Apostolopoulos and Baker would be applying a known technique (Wong's web browser) to a known device ready for improvement (the combined communication network system of Apostolopoulos and Baker) to yield predictable results. While Baker describes a miniaturized film strip and tap or touch functionality controlling it (Baker, Figs. 10A, 10B, 8:40-62), Apostolopoulos and Baker provide limited details on what specific type of application or platform is used. Wong specifically describes using a web browser to view live video casts and conferences, and even contemplates similar display techniques as Baker, such as displaying "thumbnails of live video casts", and that a "user can click on a thumbnail to view that particular blog or cast." Wong, [0041].

283. The results of the combination would have been predictable, and a POSITA would have been motivated to combine Wong's web browser with the communication network of Setton and Baker for the same reasons explained above. *See* Section X.B.

284. Finally, a POSITA would have had a reasonable expectation of success in combining Wong's web browser-based video share casting technique with Apostolopoulos' communication network and Baker's film strip. A POSITA would have had the skills and general knowledge (e.g., computer programming and user interface design) to implement a web browser-based videoconferencing application. Apostolopoulos provides limited detail on what type of software (e.g.,

programming language) or platforms could be used to view and control the conference. Baker describes storing “software applications including...browsers.” Baker, 3:29-34, but otherwise does not go into specifics on the software implementation. However, a POSITA would have had this requisite background knowledge, and Wong simply provides implementation details that Apostolopoulos and Baker expect. Modifying Apostolopoulos and Baker to serve Baker’s film strip to a web browser, as explained in Wong, installed at Apostolopoulos’ nodes would have been within a POSITA’s skill set. Wong, [0013], [0041], [0100]. Further, a POSITA would have understood that using well-known and commonly-used web browsers to access the videoconferencing service would have make the service available to more people.

B. Detailed Analysis of Claims

1. Claim 4

- a. **[4.0] *A teleconferencing system as claimed in claim 3, wherein the server is configured to serve the filmstrip to a browser co-located with the non-mesh connected endpoint.***

285. The combination of Apostolopoulos, Baker, and Wong render this limitation obvious.

286. *First*, as discussed in [3.3], it would have been obvious to a POSITA that central node 365 would include and run server software (*server*) to accomplish Apostolopoulos’ transmitting purpose.

287. *Second*, it would have been obvious for the server software running on Apostolopoulos' central node 365 to serve Baker's strip of video media to a *browser* on Apostolopoulos' nodes 361, including node 361A (*non-mesh connected endpoint*).

288. Apostolopoulos explains that "network 140" facilitates communication between the components of the "N-way collaborative virtual environment." The "network 140" may include the "Internet":

As shown in FIG. 1, the network 140 facilitates communication between devices and nodes that are communicatively coupled with the network 140. Within the network 140, communication traffic is transmitted through various devices, such as, routers and/or switches. A communication protocol (e.g., TCP/IP protocol) implements the communication through the network 140. **For example, the network 140 may comprise the Internet,** or a local area network (LAN), a wide area network (WAN), etc.

Apostolopoulos, [0021].

289. Moreover, nodes may be low-powered devices such as cell-phones, in which case, Apostolopoulos teaches that such device would only need a "simple video decoder" to view the collaborative video streams:

In the hybrid topology 300C of FIG. 3C, a [*sic*] nodes 361A-C are communicatively coupled to the central node

365. The central node is capable of streaming communication traffic to a group of other collaborators that communicate amongst themselves. That is, the collaborators 350A-C along with the central node/server 365 communicate in a full-mesh topology. For example, the star topology 360 may comprise a group of **low-power cellphone clients at nodes 361A-C**. In this case, **the central server/node 365 provides the processing power for generating communication data streams for the collaborators at nodes 361A-C to communicate amongst themselves, as well as with the collaborators at nodes 350A-C.**

Apostolopoulos, [0070].

In addition, a node such as a laptop or PDA may only have 1, 2, or 0 cameras for capturing real-time video streams of the local collaborator. However, the **low complexity node can still collaborate because the node only requires a simple video decoder to view the synthetic N-way collaborative environment.**

Apostolopoulos, [0063].

290. As explained in Ground 2, [4.0], it would have been obvious to a POSITA that the video conference application could operate within a browser. Browsers were one of the principal mechanisms for users to access the Internet and internet-based applications, including video teleconferencing applications. *See*

TechCrunch (Ex.1015); PCWorld (describing Google Hangouts and Facebook Skype services that use web browsers for video calls) (Ex.1016). It would have been obvious for each of the nodes 361A-C to have installed a browser application that is stored on local memory (*co-located with the non-mesh connected endpoint*).

291. Wong provides additional implementation details related to the software and application platform for conference participants in Apostolopoulos' communication network to view and engage in the conference. It would have been obvious for each of the nodes 361, including node 361A (*non-mesh connected endpoint*) to have installed Wong's web browser (Wong, [0013], [0041], [0106]) on local memory (*co-located with the non-mesh connected endpoint*) to connect to and participate in a video conference. Accordingly, it would have been obvious for the central node's content distribution software (*server*) to serve the filmstrip to a *browser* installed at the star node 361A. *See* section XII.A.

292. Thus, the combination of Apostolopoulos, Baker, and Wong renders this limitation obvious.

2. Claim 5

- a. [5.0] *A teleconferencing system as claimed in claim 4, wherein the browser is displayed on a computer co-located with the non-mesh connected endpoint.***

293. The combination of Apostolopoulos, Baker, and Wong renders this limitation obvious.

294. The '099 patent explains that an endpoint 84 can be both “co-located” with a computer and “run on that same computer.” '099 patent, 16:10-12, 20-21, Fig. 9; section X.C.2.

295. As explained in [3.1.1], Apostolopoulos' nodes are computational devices. *See* Apostolopoulos, [0060] (laptops), [0063] (laptop, or PDAs), [0070] (cellphones).

296. As explained in [4.0], the combination of Apostolopoulos, Baker, and Wong renders obvious Apostolopoulos' node 361A (*non-mesh connected endpoint*) using a *browser* to display Baker's strip of video media. A POSITA would have understood, in light of the '099 patent's description of co-located endpoints and computers, that the star node 361A comprises an endpoint that is also co-located with a computer. *See* '099 patent, 16:9-21. Thus, it would have been obvious to a POSITA for the star node 316A to include a *computer co-located with the non-mesh connected endpoint*, and to display Wong's *browser* on the computer.

297. Thus, the combination of Apostolopoulos, Baker, and Wong renders this limitation obvious.

3. Claim 6

- a. [6.0] *A teleconferencing system as claimed in claim 5, wherein the proxy and server are merged into a virtualized environment.***

298. The combination of Apostolopoulos and Baker render this limitation obvious.

299. As explained above, it would have been obvious to a POSITA for Apostolopoulos' "central node 365" to perform the functionalities of a *proxy* and a *server* by software running on the central node 365. Apostolopoulos contemplates that its "N-way collaborative virtual environment" may be implemented in software:

Embodiments of the present invention can be implemented on software running on a computer system. The computer system can be a personal computer, notebook computer, server computer, mainframe, networked computer, handheld computer, personal digital assistant, workstation, and the like. This software program is operable for providing topology configuration and adaptation to support communication in a communicative environment.

Apostolopoulos, [0017].

300. As similarly explained in Ground 2, it would have been obvious to instantiate the server software as a virtual server or virtual machine. In this way,

the receiving and processing functionalities (*proxy*), and the video transmission functionalities (*server*) would be *merged into a virtualized environment* at the central node 365.

301. Alternatively, and as also explained in Ground 2, [6.0], it was also well known that computational hardware, such as computers and/or servers, may be implemented in the cloud. *See, e.g., Sartori*, 5:50-52. Similar to Ground 2, it would have been obvious for the *proxy* and *server* software running on the central node 365 to be at least partially administered or supported by a cloud (*merged into a virtualized environment*).

302. For example, a POSITA would have been motivated to incorporate a cloud server to support the *proxy* and *server* functions by handling administrative tasks and user data, such as authorization, registration, and meeting configurations. Further, offloading the tasks of the proxy and server from the active source 102 to a virtualized environment would advantageously free up resources of the active source 102. In one example, the cloud server would provide common access to non-streamed content and resources, such as files or presentation graphics (*see Baker*, Fig. 10A) that can be accessed by the conference participants individually as desired. A POSITA would have understood that using a cloud server to support the *proxy* and *server* would advance Apostolopoulos' objective of reducing complexity in collaboration. Apostolopoulos, [0006].

303. The combination of Apostolopoulos and Baker render this limitation obvious.

XIII. CONCLUSION

304. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and that these statements were made with knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code.

Dated: November 17, 2023

A.L. Narasimha Reddy
Narasimha Reddy, Ph.D.