

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

MICROSOFT CORPORATION,
Petitioner

v.

SANDPIPER CDN, LLC,
Patent Owner

Case IPR2026-00174
U.S. Patent No. 9,762,692

DECLARATION OF PRASHANT SHENOY, PH.D.

Mail Stop "PATENT BOARD"

Patent Trial and Appeal Board
U.S. Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450

Sandpipper EX2007
Microsoft v. Sandpipper
IPR2026-00174

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	MY BACKGROUND AND QUALIFICATIONS	2
III.	LIST OF DOCUMENTS CONSIDERED	6
IV.	LEGAL UNDERSTANDING	9
	A. My Understanding of Claim Construction	9
	B. My Understanding of Obviousness	9
	C. My Understanding of a Person Having Ordinary Skill in the Art	11
V.	PERSON OF ORDINARY SKILL IN THE ART	12
VI.	OVERVIEW OF THE STATE OF THE ART	12
VII.	OVERVIEW OF THE '692 PATENT	15
VIII.	OVERVIEW OF THE APPLIED REFERENCES	17
	A. Seed	17
	B. Swildens	19
IX.	DETAILED DISCUSSION OF OPINIONS	20
	A. Independent Claims 1, 10, 15: Seed and Swildens, in any order and for both Grounds, do not teach the “first portion of the content provider’s library is distinct from the second portion,” as recited in claim 1 and similarly recited in claims 10 and 15.	22
	B. Independent Claim 1	27
	1. The Petitioner fails to show Seed-Swildens or Swildens- Seed renders obvious the claimed “if the resource is determined not to be popular, directing the client to a second server in a second tier of servers distinct from the first tier of servers.”	27
	2. The Petitioner fails to show either combination teaches the claimed “ <i>if the resource is determined to be popular, then the first server obtaining the resource and the first server serving the resource to the client.</i> ”	31
	a. Ground 1	31
	b. Ground 2	36

C.	Independent Claim 10	39
1.	Petitioner fails to show Seed-Swildens or Swildens-Seed renders obvious the claimed “ <i>provide the edge servers with a second set of information about which tier of servers should handle those requests for content.</i> ”	39
2.	Petitioner fails to show Seed-Swildens or Swildens-Seed renders obvious the claimed “ <i>determin[ing a tier of servers to handle a request] based, at least in part, on a measure of popularity of requested content .</i> ”	42
D.	Independent Claim 15: Seed and Swildens, in any order and for both Grounds, do not teach the claimed “ <i>redirecting being based, at least in part, on a current popularity value for the resource.</i> ”	44
X.	CONCLUSION.....	46

I, Prashant Shenoy, Ph.D., hereby declare as follows.

I. INTRODUCTION

1. I am over the age of twenty-one and otherwise competent to make this Declaration. This Declaration is based on my personal knowledge as an expert in the relevant field. I could testify competently if asked to do so.

2. I have been retained as an expert witness on behalf of Patent Owner Sandpiper CDN, LLC (“Patent Owner”) for the above-captioned *inter partes* review (“IPR”). I am being compensated for my time in connection with this IPR at my normal hourly rate. My compensation is not contingent on either my expert opinion or the outcome of this litigation. I have no other interest in this proceeding.

3. I understand that this Declaration accompanies Patent Owner’s Preliminary Response filed in this IPR involving U.S. Patent No. 9,762,692 (“the ’692 patent”) (EX1001), which resulted from U.S. Application No. 12/880,324 (“the ’324 application”), filed September 13, 2010. The ’324 application claims priority as a continuation-in-part to U.S. Application No. 12/408,681, filed March 21, 2009, which claims priority to U.S. Provisional Application No. 61/042,412, filed April 4, 2008. My opinions regarding the knowledge in the art at the time of the ’692 patent are based on the April 4, 2008 priority date.

4. In preparing this Declaration, I have reviewed the ’692 patent and each of the documents cited herein, in light of general knowledge in the art before

the '692 patent's priority date. In formulating my opinions, I have relied upon my experience, education, and knowledge in the relevant art. In formulating my opinions, I have also considered the viewpoint of a person of ordinary skill in the art ("POSITA") (as discussed further below in Section V) at the time of the '692 patent's priority date

II. MY BACKGROUND AND QUALIFICATIONS

5. I earned a Bachelor of Technology degree from the Department of Computer Science and Engineering at the Indian Institute of Technology in Bombay, India. I later received my Master of Science (M.S.) and Doctor of Philosophy (Ph.D.) degrees in 1994 and 1998, respectively, in Computer Science from the University of Texas at Austin.

6. Over the years, I have led research and industry collaboration projects in areas such as storage and file systems, Web and Internet systems, content distribution, streaming media, database systems, cloud computing, virtualization, RFID and sensor networks, and energy systems.

7. I served as an Associate Dean in the College of Information and Computer Sciences at the University of Massachusetts Amherst ("UMass") from 2016 to 2025. I am a Distinguished Professor of Computer Science at the University of Massachusetts and additionally hold a Visiting Professor

appointment at UCLA on account of a sabbatical. I have been a faculty member at UMass since 1998.

8. I received my Bachelor of Technology (B.Tech.) degree in 1993 from the Department of Computer Science and Engineering at The Indian Institute of Technology (“IIT”), Mumbai, India. IIT awarded me the Institute Medal for being the top ranking undergraduate student. In 1994, I received my Master of Science (M.S.) degree in Computer Sciences from The University of Texas at Austin. In 1998, I received my Doctor of Philosophy (Ph.D.), also from the Department of Computer Sciences at The University of Texas at Austin. My dissertation, which was in the area of file systems, was recognized by the department as the Best Doctoral Dissertation of 1998-1999.

9. I have been involved in the research and development of distributed computing systems for more than 30 years. At UMass, I perform research and teach in the areas of distributed systems, operating systems, computer networking, data management, and multimedia systems, among others.

10. A major focus of my research has been the design of large-scale distributed computing systems. Over the years, I have led research and industry collaboration projects in areas such as computer networking, distributed systems, storage and file systems, Web and Internet systems, content distribution, streaming media, database systems, cloud computing, virtualization, Radio Frequency

Identification (“RFID”) and sensor networks, and energy systems. During that time, I have consulted with both large companies and startups in a technical advisory role related to product architecture, intellectual property (“IP”) issues, and applied research.

11. For example, starting in the early 1990s, my research focused on the development of storage and file systems. During that time, I developed multimedia file system and data management technologies that were designed for efficient storage and retrieval of large amounts of video and audio data. I collaborated with AT&T Bell Laboratories to design a new media server for video streaming. These efforts resulted in U.S. Patent No. 6,079,028 on fault-tolerant architectures for continuous media servers.

12. In the late 1990s, I led the design of a new operating system called QLinux that was based on the open-source Linux operating system. As part of this research, I developed resource management and scheduling techniques that provided isolation and service guarantees to applications. This technology provided the underpinnings for fair resource management and isolation techniques used in operating system (“OS”) virtualization. It was adopted by a startup company to provide resource isolation in its web hosting products.

13. In the early 2000s, I led several projects on web systems and cloud computing. For example, I led research projects on web technologies such as web

caching and scalable web applications. I also led research projects on cloud computing platforms (which were known as web hosting platforms at the time) involving virtualization, dynamic resource management, replication, fault tolerance, and storage. My group at UMass collaborated with technology companies such as IBM, AT&T, Sprint, Intel, and many others as part of this research.

14. My research on virtualization has involved designing technologies for virtual machines and container resource management and deploying them in cloud computing systems and data centers. I am co-inventor on U.S. Patent No. 8,473,557, jointly assigned to AT&T and UMass, on virtual machine migration.

15. Over the years, I have authored or co-authored more than 300 articles in peer-reviewed journals and conference proceedings that have been cited over 27,000 times. Seventeen papers authored by myself and members of my research group on this project won awards for their technical contributions. Two papers have won Test-of-Time awards, an award given to a publication that has had the most impact over a ten-year period since its publication.

16. I have served as an Editor for six leading journals and have served as the Technical Program Chair for twenty conferences and workshops in my field. I founded the ACM Special Interest Group on Energy Systems and also served on a steering committee of conferences and journals.

17. I am also a Fellow of the ACM, IEEE, AAAS, and AAIA, four professional societies in computing and science.

18. I am a named inventor on three U.S. Patents and one published application (U.S. Publication No. 2014/0040206 and U.S. Patent Nos. 8,510,596, 8,473,557, and 6,079,028). These documents are generally related to distributed and operating systems, virtualization, and cloud data centers.

19. As a faculty member at UMass, I have also taught undergraduate and graduate courses in operating systems, distributed systems, networking, and related topics for more than twenty years.

20. These courses teach concepts related to operating systems, virtual machines, containers, cloud computing, file systems, and distributed systems. They also involve hands-on course projects on these topics.

21. My *curriculum vitae* is attached as EX2008, which contains further details on my education, experience, publications, and other qualifications to render an expert opinion.

III. LIST OF DOCUMENTS CONSIDERED

22. I have considered the following documents listed below in addition to any other documents cited in this Declaration. To the best of my knowledge, these documents are true and accurate copies of what they purport to be. An expert in the field would reasonably rely on them to formulate opinions such as those set forth

in this Declaration.

Exhibit No.	Description
2001	First Amended Docket Control Order, <i>Sandpiper CDN, LLC v. Microsoft Corporation</i> , Case No. 2:25-cv-00664-JRG-RSP (E.D. Tex.), October 17, 2025
2002	Exhibit E101 to Microsoft’s Invalidity Contentions (Subject Matter Ineligibility Contentions for U.S. Patent No. 9,762,692), <i>Sandpiper CDN, LLC v. Microsoft Corporation</i> , Case No. 2:25-cv-00664-JRG-RSP (E.D. Tex.), December 5, 2025
2003	“Enforcement and Non-Waiver of 37 C.F.R. § 42.104(b)(4) and Permissible Uses of General Knowledge in <i>Inter Partes</i> Reviews,” United States Patent and Trademark Memorandum, July 31, 2025
2004	Success Rate of Motions to Stay Pending <i>Inter Partes</i> Review before Judge Rodney Gilstrap, Docket Navigator Report generated February 4, 2026
2005	Complaint for Patent Infringement, <i>Sandpiper CDN, LLC v. Microsoft Corporation</i> , Case No. 2:25-cv-00664-JRG-RSP (E.D. Tex.), June 26, 2025
2006	Microsoft’s Invalidity Contentions, <i>Sandpiper CDN, LLC v. Microsoft Corporation</i> , Case No. 2:25-cv-00664-JRG-RSP (E.D. Tex.), December 5, 2025
2009	Markman Order, <i>Sandpiper CDN, LLC v. Comcast Cable Comm’ns, LLC</i> , Case No. 2:24-cv-00886-JRG-RSP (E.D. Tex.), February 18, 2026
2010	Infringement Contentions, <i>Sandpiper CDN, LLC v. Microsoft Corp.</i> , Case No. 2:25-cv-00664 (E.D. Tex. Sept. 16, 2025)
2011	“Connect – Most Innovative New Products Award Winners Announced,” Connect NewsLetter, October 26, 1999, available at https://library.ucsd.edu/dc/object/bb60436752/_1.pdf (last accessed March 1, 2026)

Exhibit No.	Description
2012	“Sandpiper Named ‘Best of the Best’ at Networld + Interop 99 Las Vegas,” PRNewswire, May 12, 1999, available at https://ubmchannelnews.mediaroom.com/index.php?s=29573&item=85841 (last accessed March 1, 2026)
2013	“Entrepreneur of the Year,” Los Angeles Business Journal, June 19, 2000
2014	“Sandpiper Networks’ Footprint Takes Aim at Internet Traffic,” Wall Street Journal Interactive Edition, June 17, 1999, available at https://url.us.m.mimecastprotect.com/s/wTVWCxklA1iDQQWjSYi2Syt1OK?domain=wsj.com (last accessed March 3, 2026)
2015	“Web-Page Distribution System Could Unclog Internet Traffic Jams,” The New York Times, April 26, 1999, available at https://archive.nytimes.com/www.nytimes.com/library/tech/99/04/biztech/articles/26sand.html (last accessed March 3, 2026)
2016	“Sandpiper Puts Its Footprint in Marketplace,” Los Angeles Times, September 28, 1998, available at https://www.latimes.com/archives/la-xpm-1998-sep-28-fi-27201-story.html
2017	“Level 3 Lands Netflix Streaming Business, Will Double Its Storage Capability,” TechCrunch, available at https://techcrunch.com/2010/11/11/level-3-lands-netflix-streaming-business-will-double-its-storage-capability/ (last accessed March 1, 2026)
2018	“CenturyLink named a Leader in IDC MarketScape: Worldwide Commercial CDN Vendors Report,” PR Newswire, Oct. 18, 2019, available at https://www.prnewswire.com/news-releases/centurylink-named-a-leader-in-idc-marketscape-worldwide-commercial-cdn-vendors-report-300933112.html (last accessed March 1, 2026)
2019	U.S. Publication No. 2009/0252071 A1 (“Nurminen”)
2020	U.S. Patent No. 7,937,477 B1 (“Day”)

IV. LEGAL UNDERSTANDING

A. My Understanding of Claim Construction

23. I understand that during an *inter partes* review proceeding, claims are to be construed in light of the specification as would be read by a person of ordinary skill in the relevant art at the time of the invention. I understand that claim terms are given their plain and ordinary meaning as would be understood by a person of ordinary skill in the relevant art (a “POSITA”) in the context of the patent’s specification and prosecution history. A claim term, however, will not receive its ordinary meaning if the patentee acted as his or her own lexicographer and clearly set forth a definition of the claim term in the specification. In this case, the claim term will receive the definition set forth in the patent.

B. My Understanding of Obviousness

24. I understand that a patent claim is invalid if the claimed invention would have been obvious to a person of ordinary skill in the art at the time the application was filed. I understand that this means that even if all of the requirements of the claim cannot be found in a single prior-art reference that would anticipate the claim, the claim can still be invalid.

25. To obtain a patent, a claimed invention must have been, as of its priority date, nonobvious in view of the prior art in the field. I understand that a patent claim is obvious when the differences between the subject matter sought to be patented 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 person having ordinary skill in the art.

26. I understand that to prove that prior art or a combination of prior art renders a patent claim obvious, it is necessary to: (1) identify the particular references that, singly or in combination, render the patent claim obvious; (2) specifically identify which elements of the patent claim appear in each of the asserted references; and (3) explain how the prior-art references could have been combined in order to create the inventions claimed in the patent claim.

27. I also understand that when considering the obviousness of a patent claim, one should consider whether a teaching, suggestion, or motivation to combine the references exists so as to avoid impermissibly applying hindsight when considering the prior art. I understand this test should not be rigidly applied, but that the test can be important to avoiding such hindsight.

28. I understand that certain objective indicia can be important evidence as to whether a patent is obvious or nonobvious. Such indicia include: (1) commercial success of products covered by the patent claims; (2) a long-felt need for the invention; (3) failed attempts by others to make the invention; (4) copying of the invention by others in the field; (5) unexpected results achieved by the invention as compared to the closest prior art; (6) praise of the invention by the infringer or others in the field; (7) the taking of licenses under the patent by

others; (8) expressions of surprise by experts and those skilled in the art at the making of the invention; and (9) the patentee proceeded contrary to the accepted wisdom of the prior art.

29. I also understand that “obviousness” is a legal conclusion based on the underlying factual issues of the scope and content of the prior art, the differences between the claimed invention and the prior art, the level of ordinary skill in the pertinent art, and any objective indicia of non-obviousness. For that reason, I am not rendering a legal opinion on the ultimate legal question of obviousness. Rather, my testimony addresses the underlying facts and factual analysis that would support a legal conclusion of obviousness or non-obviousness, and when I use the term obvious, I am referring to the perspective of one of ordinary skill at the time of invention of the '692 patent.

C. My Understanding of a Person Having Ordinary Skill in the Art

30. I understand that a POSITA is presumed to be aware of all pertinent art, thinks along conventional wisdom in the art, and is a person of ordinary creativity.

31. I understand that the following should be considered when deciding the level of ordinary skill in the art that someone would have had at the time the claimed invention was made:

- the levels of education and experience of persons working in the field;

- the types of problems encountered in the field; and
- the sophistication of the technology.

32. My opinion below explains how a POSITA would have understood the technology described in the documents I have identified herein around the time of the '692 patent.

33. Regardless of whether I use “I” or “POSITA” during my technical analysis below, all of my statements and opinions are always to be understood to be based on how a POSITA would have understood or read a document at the time of the '692 patent.

V. PERSON OF ORDINARY SKILL IN THE ART

34. For purposes of this Declaration, I have applied Mr. Lanning’s articulation of a POSITA. EX1003, ¶¶85-87. It is my opinion that the definition of a POSITA would not materially affect the limited analysis required for deciding institution and presented in this Declaration. If called to give further opinions in this proceeding should trial be instituted, I reserve the right to reconsider the definition of a POSITA and arrive at my own conclusion.

VI. OVERVIEW OF THE STATE OF THE ART

35. Content delivery networks (“CDNs”) provide critical services that enable content providers to quickly deliver online content to millions of consumers simultaneously over the Internet. In the early 1990s, the World Wide Web saw

increasing adoption and became a household staple. This mass adoption led to data congestion issues from the ever-growing number of users seeking to simultaneously access Internet content.

36. A typical computer server in the 1990s could only handle a limited number of simultaneous connections before becoming overloaded. Moreover, data takes time to traverse routers and move through content-congested internet cables. Consumers living far from the physical server(s) hosting content experienced sluggish load times and high latency due to problems such as overloaded servers, congested network segments, and geographic separation.

37. The availability of CDN servers around the world allowed service providers to replicate appropriate content from customers' origin servers to appropriate CDN servers and transparently route end users requesting that content to the "best" CDN server to deliver that content, all while providing their customers with control over their content and the user experience. This service and its architecture were quickly adopted by many others in the industry.

38. End users would connect to an optimal edge server—a server that was closer to the end user or that had available capacity. Distributing content across a network of servers alleviated data congestion issues and allowed consumers to connect to edge servers located near them to reduce latency. Systems and methods for propagating data from origin servers to edge servers and storing them at end

servers for serving user requests are known as “caching.”

39. I understand that by 1996, a team at Sandpiper Networks Inc. had conceived techniques for delivering static and dynamic content, including streaming resources, such as audio and video, using Sandpiper’s CDN. EX2005, ¶23.

40. I also understand that by at least May 1998, Sandpiper Networks was caching content and delivering cached content to end users of content providers using its CDN. EX2005, ¶¶25-26 n. 12. Sandpiper Networks’ first paying customer was the L.A. Times, which paid Sandpiper Networks to host the report of Independent Counsel Ken Starr on his investigation of President Bill Clinton (“the Starr Report”) beginning on September 11, 1998. EX2005, ¶¶25-26. Sandpiper Networks’ CDN was capable of caching and was used to cache and deliver Internet resources including pictures, text files, dynamic resources, and streaming multimedia resources. EX2005, ¶¶25-26.

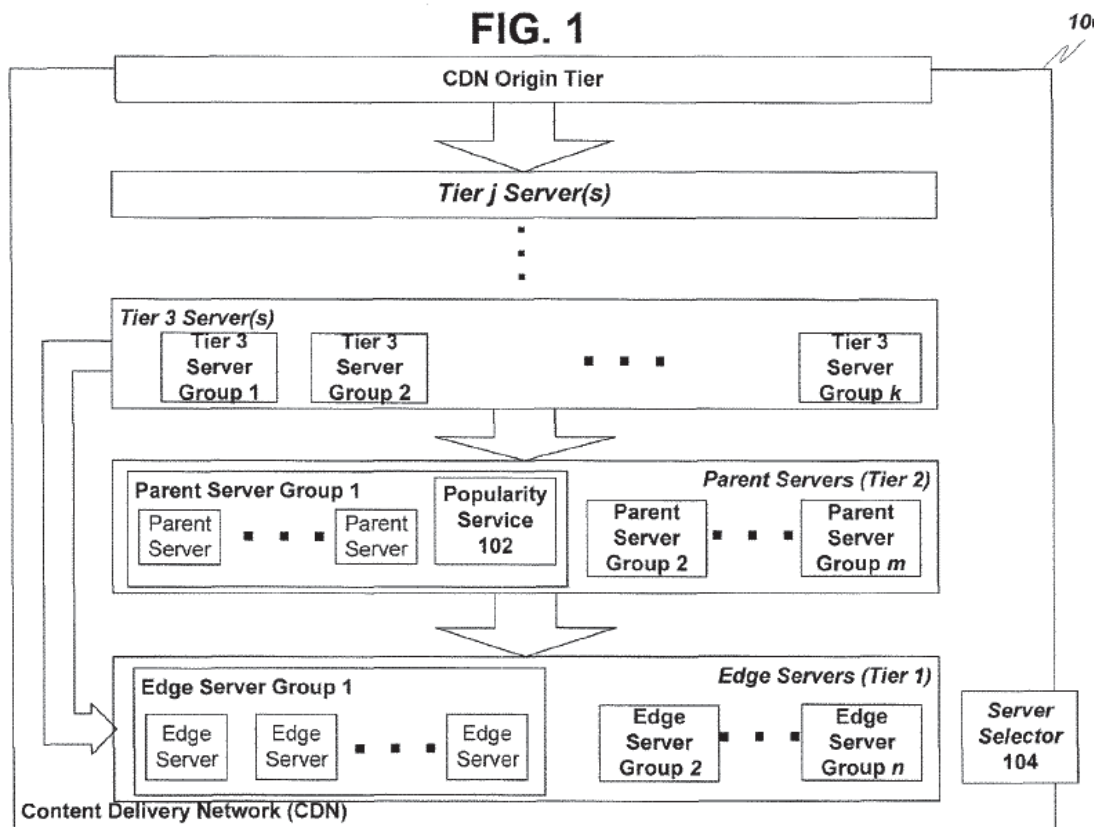
41. In the early-to-mid 2000s, demand for CDNs exploded. As the market expanded, other businesses entered the space—often by misappropriating Sandpiper’s innovations and directly competing with Sandpiper for market share.

42. One such company is Microsoft. Microsoft built a CDN, which uses technology described and claimed by Sandpiper’s patents. For example, Microsoft’s Azure CDN supports Xbox gaming, Office 365, Windows updates,

and web resources like LinkedIn and Bing. *See* EX2010, 2-3, 366-446.

VII. OVERVIEW OF THE '692 PATENT

43. I understand that the '692 patent presents a novel and innovative technique for serving content based on popularity levels. *See* e.g., EX1001, 6:41-7:10. As shown below, a CDN may include a plurality of tiers, such as a tier of edge servers, a tier of origin servers, and a tier of parent servers. EX1001, 2:52-66, 4:5-8, FIG. 1.



EX1001, FIG. 1.

44. When a client requests an object, that request is first directed to an edge server. EX1001, 5:45-48. The edge server checks if the requested object is

present and, if so, serves the object. EX1001, 5:66-6:3. If the requested object is not available, the CDN returns one of three possible responses to the edge server. EX1001, 6:41-50; *see also* EX1001, FIG. 5. First, if the object has not reached a first/minimal level of popularity, the popularity server instructs the edge server to redirect the request to an origin server. EX1001, 6:51-55. Second, if the object's popularity has exceeded the first/minimal level but has not yet exceeded a second, mid-tier threshold, the popularity server instructs the edge server to redirect the request to a parent server. EX1001, 6:56-61. Third, if the object's popularity has exceeded the mid-tier threshold, the popularity server instructs the edge server to serve the content itself. EX1001, 7:4-7.

45. Furthermore, when redirecting requests to parent servers, a determination of which particular parent server has the requested object is made. EX1001, 6:61-64. "Such a determination would be based on how the distinct subsets of a content provider's library are partitioned across parent servers in the same intermediate tier." EX1001, 6:64-67. For example, the '692 patent explains that when "a content provider's library comprises multiple content-types (e.g., movies, video games, music, software patches, etc.)," "it would be advantageous to store a proportionate amount of each content type on each participating intermediate tier server (e.g., 60% movies, 30% video games, 10% software downloads on each participating server in the tier)." EX1001, 12:29-36. "In such a

scenario, if the demand for a particular content-type increases during a given time period (and/or demand for a different content-type declines), then such fluctuations in demand will generally affect each participating server proportionately.”

EX1001, 12:36-40.

VIII. OVERVIEW OF THE APPLIED REFERENCES

46. I understand that Petitioner’s expert, Mr. Lanning, relies on numerous references in his proposed obviousness grounds for the challenged claims of the ’692 patent. Two of those references are particularly relevant to my opinions provided in this Declaration: Seed (EX1005, hereinafter referred to as “Seed”) and Swildens (EX1006, hereinafter referred to as “Swildens”). I provide brief overviews of these references below.

A. Seed

47. Seed describes a system for “managed object replication and delivery over a network.” EX1005, ¶1. Seed explains that “[a] typical content delivery network (CDN) operator deploys one or more parent” and “one or more edge servers at the edge of the network to facilitate more cost-effective and efficient delivery of such objects to an end-user (client).” EX1005, ¶10. As shown below, “the network also optionally has access to one or more origin server sites 110” that “are typically owned and/or maintained by the network provider’s customers for storing and serving one or more objects.” EX1005, ¶20, FIG. 2.

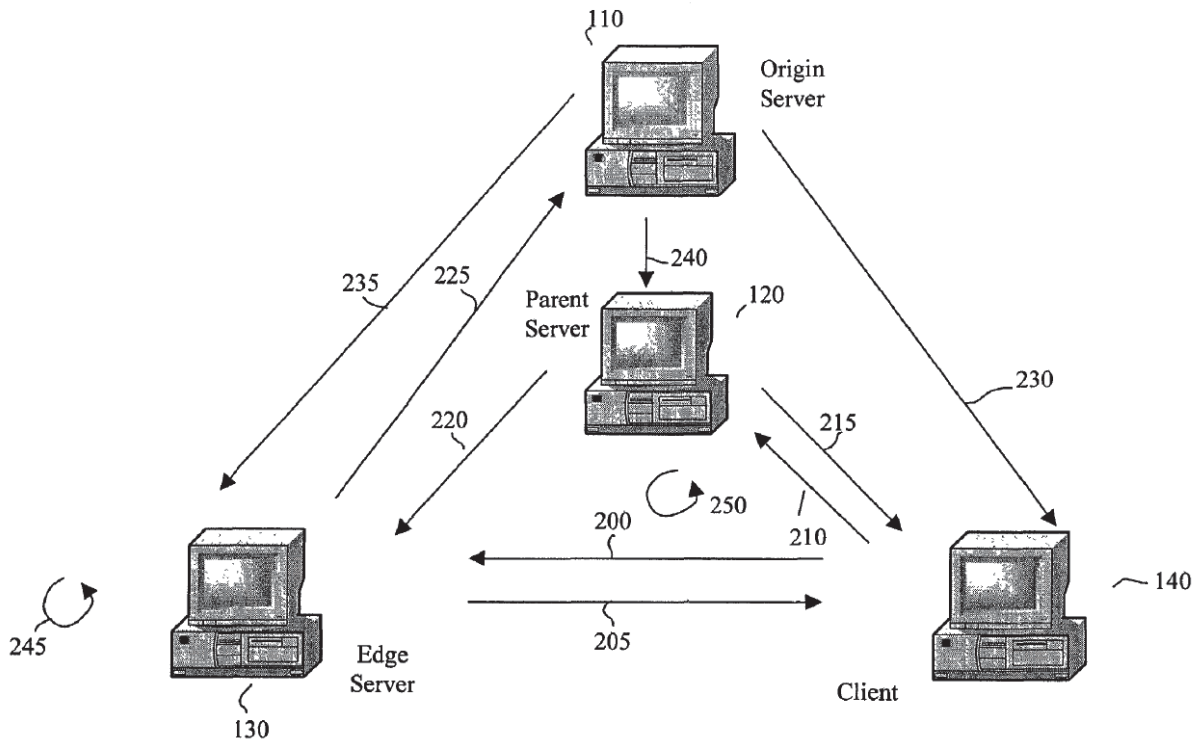


Fig. 2

EX1005, FIG. 2

48. Seed describes that it “is typically not feasible to store all objects on the edge servers” because it can be cost-prohibitive and may not even be possible. EX1005, ¶11. One solution is to pre-populate edge servers with objects that are likely significant and in high demand. EX1005, ¶12. “However, it is difficult to predict popularity and difficult to manage pre-populating.” EX1005, ¶12. “Other solutions fetch objects on demand”—“when [the] requested object is not available on a handling edge server,” the handling edge server fetches the requested object from a parent server. EX1005, ¶13. Seed describes that doing so can use too much bandwidth and create congestion. EX1005, ¶14.

49. In view of the issues, Seed's edge server redirects the client's request to a parent server when the requested object is not available. EX1005, ¶¶17-18, 22, 27. If the requested object is popular, it is replicated from a parent server to the edge server to be served in the future. EX1005, ¶¶17-18, 26-27.

B. Swildens

50. Swildens describes a system for “the automatic migration of data via a distributed computer network.” EX1006, ¶18. Swildens' system includes edge servers, an origin server, and a staging site. EX1006, ¶¶42, 46, 53-54, FIG. 1. Swildens' system is depicted below in Figure 1.

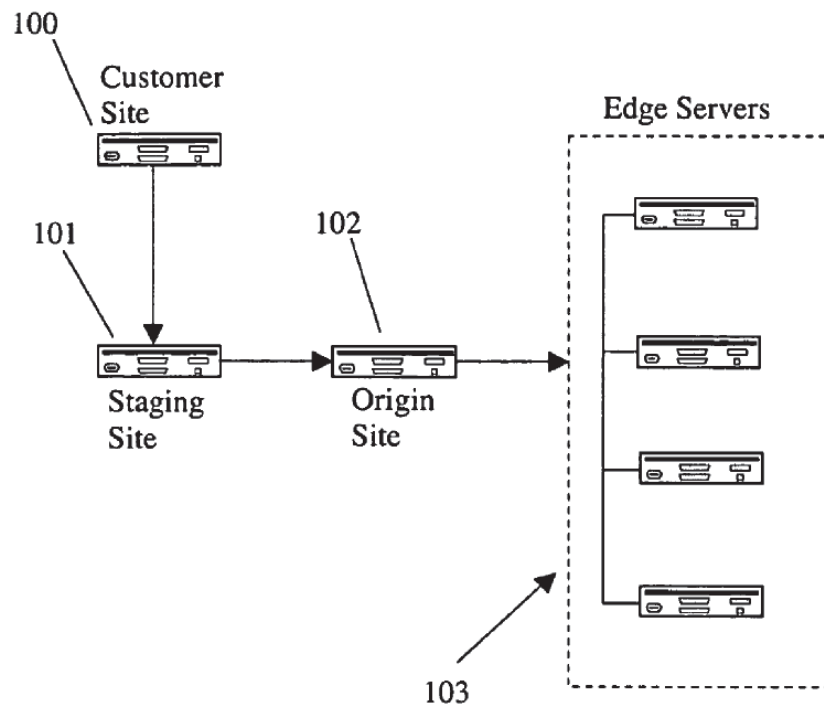


Fig. 1

EX1006, FIG. 1.

51. Swildens’ “edge server receives a user’s request for content.”

EX1006, ¶46. If the edge server does not have the content, it redirects the user to an origin server. EX1006, ¶46. If the origin server does not have the content, it redirects the user to a staging server. EX1006, ¶47. Swildens describes that servers “should typically have at least the combined storage and delivery capacity to store one copy of all of the content files.” EX1006, ¶58.

52. Swildens’ system also includes an auto migration server that periodically (e.g., every hour) collects information from the edge servers to determine popularity levels of content. EX1006, ¶¶126-127. The auto migration server then selects the few most popular content items for auto migration to the edge servers. EX1006, ¶¶126-127. In addition, “[s]ingle edge serve[r]s can perform their own periodic evaluation of popular files.” EX1006, ¶128.

IX. DETAILED DISCUSSION OF OPINIONS

53. For purposes of this Declaration, I have been asked to opine on certain positions taken by Petitioner and Mr. Lanning. That I omit certain claims, claim elements, or presented Grounds from my analysis should not be viewed as agreement that any claims or claim elements are obvious.

54. Upon considering the materials listed above and discussed herein, based on my extensive experience and education in the relevant art, in view of my understanding of the governing legal principles and the arguments presented in the

Petition, I have formed the following opinions.

55. It is my opinion that the Petition's obviousness challenges to independent claims 1, 10, and 15 are deficient because Petitioner and Mr. Lanning fail to establish that either proposed combination (Seed-Swildens or Swildens-Seed) teaches:

- **All claims:** *“wherein the second server comprises a first portion of the content provider’s library ... wherein at least one other server in the second tier comprises a second portion of the content provider’s library, wherein the first portion of the content provider’s library is distinct from the second portion”* (element 1[d][ii], and similarly recited in elements 10[d] and 15[c]).
- **Claim 1:** (a) *“if the resource is determined to be popular, then the first server obtaining the resource and the first server serving the resource to the client”* (element 1[c]) and (b) *“if the resource is determined not to be popular, directing the client to a second server in a second tier of servers distinct from the first tier of servers”* (element 1[d][i]).
- **Claim 10:** (a) *“provid[ing] the edge servers with a second set of information about which tier of servers should handle those requests for content”* (element 10[c]) and (b) *“the popularity service making its*

determinations [of which tier of servers should handle those requests for content] based, at least in part, on a measure of popularity of requested content” (element 10[c]).

- **Claim 15:** *“selectively redirecting the request from the client to a second tier server ... based, at least in part, on a current popularity value for the resource” (element 15[b]).*

A. Independent Claims 1, 10, 15: Seed and Swildens, in any order and for both Grounds, do not teach the “first portion of the content provider’s library is distinct from the second portion,” as recited in claim 1 and similarly recited in claims 10 and 15.

56. Independent Claim 1 recites that *“the second server comprises a first portion of the content provider’s library ... wherein at least one other server in the second tier comprises a second portion ... the first portion of the content provider’s library is distinct from the second portion.”* Independent claim 10 recites that *“distinct portions of the content provider’s library are logically partitioned across parent servers in the second tier.”* Independent claim 15 recites that *“the second tier server comprises a first portion of the content provider’s library ... wherein at least one other second tier server comprises a second portion of the content provider’s library distinct from the first portion.”* For all three claims, Petitioner relies on its arguments for claim element 1[d][ii]. See Pet., 25-28, 37, 44-45, 58-59, 69-70, 77. However, I believe Petitioner’s arguments regarding the claimed *distinct portions* are fundamentally flawed.

57. First, Petitioner argues that Seed discloses this feature. Pet., 25. This argument is incorrect for several reasons. To begin with, both the Examiner and the Board previously found that Seed does not disclose this feature. EX1002, 252, 2123, 2338-2342. Petitioner has not explained why the Examiner and Board were incorrect—nor can it because Seed simply does not disclose the claimed feature.

58. Petitioner fails to identify any portion of Seed that discloses partitioning of content amongst parent servers as claimed. Petitioner’s theory solely relies on paragraph 29 of Seed. Pet., 25. But the cited portion of Seed discusses how an individual parent server handles a request when it does not have the requested object—it says nothing about how content is distributed. EX1005, ¶29. Petitioner acknowledges that Seed “does not expressly disclose this limitation.” Pet., 25. Facing the absence of disclosure, Petitioner provides a conclusory sentence that Seed suggests its servers have distinct content portions. Petitioner states that “Seed recognizes that not every parent-level server has the same objects.” Pet., 25. I disagree because Seed does not discuss parent servers having different objects. Mr. Lanning also acknowledges that “Seed does not expressly specify the extent of any overlap between parent servers.” EX1003, ¶124.

59. Faced with this gap, Petitioner argues that this limitation “would have been obvious in view of Swildens.” Pet., 25-26, 58-59. This argument hinges on an

assumption that Swildens' hash function would uniquely identify a single server.

See Pet., 25-26, 58-59 (“[A] POSITA would have understood that a hash function would typically result in each server having distinct, non-overlapping portions of the library, because each content identifier would be assigned to one server based on its hash value.”). Mr. Lanning similarly argues that “running a content identifier (key) through a hash function would typically produce only one output, which a POSITA would have understood would be the server to which it is assigned (e.g., a one-to-one key-to-slot mapping).” EX1003, ¶134. I disagree because a hash function can identify multiple servers. Although a hash function would typically produce only one output, that output can correspond to multiple servers, not just a single server. Taking Dr. Lanning's example of a hash function being a modulo operation of 10 and assuming that Swildens' content identifiers are numerical values, the hash function would divide a content identifier by 10 and take the remainder as the output, which is a hash value. *See* EX1003, ¶132. For a content identifier that is a number 51, the hash value is 1 because 51 divided by 10 produces a remainder of 1. In this way, the hash value is uniquely produced. However, multiple servers can share the hash value. For example, servers may have identifiers that include a section to indicate corresponding hash values. Multiple servers can set that section to be 1. One way to do that is mapping the identifiers that end with 1 to the hash value of 1. In such a case, all servers having

the identifiers that end with 1 are identified by the hash value of 1. Thus, even though the output of the hash function is unique, such an output is still associated with multiple servers. In my opinion, this is a well-known technique to assign objects to servers. For example, U.S. Publication No. 2009/0252071 to Nurminen explains that

An object can be inserted into or retrieved from the DHT logical space by calculating its hash value (e.g., in accordance with SHA-1) to produce a key. This key is then used to locate one or more particular node IDs that receive the object, or from which the object may be retrieved. Kademlia stores the objects in replica in several nodes.

EX2019, ¶21. Here, a uniquely produced hash value can be used to locate multiple nodes that store objects in replica. Thus, the mere fact that the hash function is used does not mean only one server can be identified. On the contrary, it is my opinion that a uniquely produced hash value can identify multiple servers. In fact, this is what Swildens describes. EX1006, ¶60 (“[A] hash included in the URL ... [is used] to determine which other of the LAN-connected **origin servers** in its group should hold the content on its local Storage device.”). This understanding is also consistent with Swildens’ consideration of origin servers. EX1006, ¶58 (“More LAN-connected groups of origin servers can be added to increase the minimum amount of redundancy required and to improve the performance of unpopular or rapidly popular content.”).

60. Furthermore, Petitioner and Mr. Lanning’s argument misread Swildens. Swildens describes that “[s]ervers should typically have at least the combined storage and delivery capacity to store one copy of **all** of the content files.” EX1006, ¶58 (emphasis added). Swildens further describes that a group of LAN-connected servers can be configured to “access each other’s local storage devices.” EX1006, ¶¶88-89. In other words, the servers share common storage and all the servers have access to the same content. Swildens explains that using “a hash included in the URL (e.g., in the host name or first component of the path),” an origin server is able to “determine which **other of the LAN-connected origin servers in its group** should hold the content on its local storage device.” EX1006, ¶60 (emphasis added).

61. Therefore, in my opinion, Swildens does not disclose, teach or suggest the claimed “first portion of the content provider’s library is distinct from the second portion,” but rather describes the content may be stored on multiple local storage devices in the LAN (not a single server as Petitioner alleges). Swildens describes that the origin server to which the content request was redirected may then use the hash function to determine which local storage devices the origin server itself can access to serve the content (i.e., it does not redirect the request because all of the LAN servers have access to the same content). EX1006, ¶¶60, 88-89. In my opinion, all origin servers share the same set of local storage devices

connected via a LAN. The hash value included in the URL of the redirected request is used by the receiving origin server to determine which local storage devices to access for the requested content. Thus, consistent with Swilden's requirement, each origin server in the LAN has "the combined storage and delivery capacity" to serve "all of the content files and satisfy the peak delivery load." EX1006, ¶58.

62. Accordingly, it is my opinion that Petitioner fails to show that Seeds or Swildens, either separately or in combination, teach the claimed "*first portion of the content provider's library is distinct from the second portion,*" recited in claims 1, 10, and 15.

B. Independent Claim 1

63. In my opinion, there are two additional incorrect arguments for claim 1.

- 1. The Petitioner fails to show Seed-Swildens or Swildens-Seed renders obvious the claimed "if the resource is determined not to be popular, directing the client to a second server in a second tier of servers distinct from the first tier of servers."**

64. Independent claim 1 recites "*if the resource is determined not to be popular, directing the client to a second server in a second tier of servers distinct from the first tier of servers.*" Thus, when the resource is unpopular, the first server redirects the request to another server. It is my opinion that Petitioner fails to show that either Ground teaches this element.

65. In **Ground 1**, Petitioner relies on Seed for this element. Pet., 23-24.

Seed explicitly discusses that the edge server redirects the request when the requested object is not available. EX1005, ¶¶17-18, 27. Thus, contrary to Petitioner's argument, Seed's redirecting is based on availability, not popularity. Seed states that it does not include pre-populating, even with content likely to be popular. EX1005, ¶12. Furthermore, Seed explicitly explains that an unavailable object can be either popular or unpopular. For example, Seed explains that "[i]f the selected edge server does not have the requested object, a check is initiated (at 315) for the edge server to determine whether the requested object is popular and if so, to replicate the popular requested object to the edge server." EX1005, ¶26. The fact that Seed checks to determine whether the unavailable object is popular indicates that both popular and unpopular objects may be unavailable based on the specific disclosure in Seed. Otherwise, if the unavailable object can only be unpopular, as the Petitioner suggests, the check is unnecessary and the check step 315 should not exist. This point is further illustrated in Figure 3(b), reproduced below, which illustrates the steps employed (at 315) to determine whether the requested object is popular. EX1005, ¶26, FIG. 3B. Specifically, Seed explains that "the method depicted in FIG. 3(b) and discussed in more detail below is employed to determine whether the requested object is popular and if so, to replicate the popular requested object to the edge server." EX1005, ¶26. Thus, Figure 3(b) describes detailed

actions of step 315 that check the popularity.

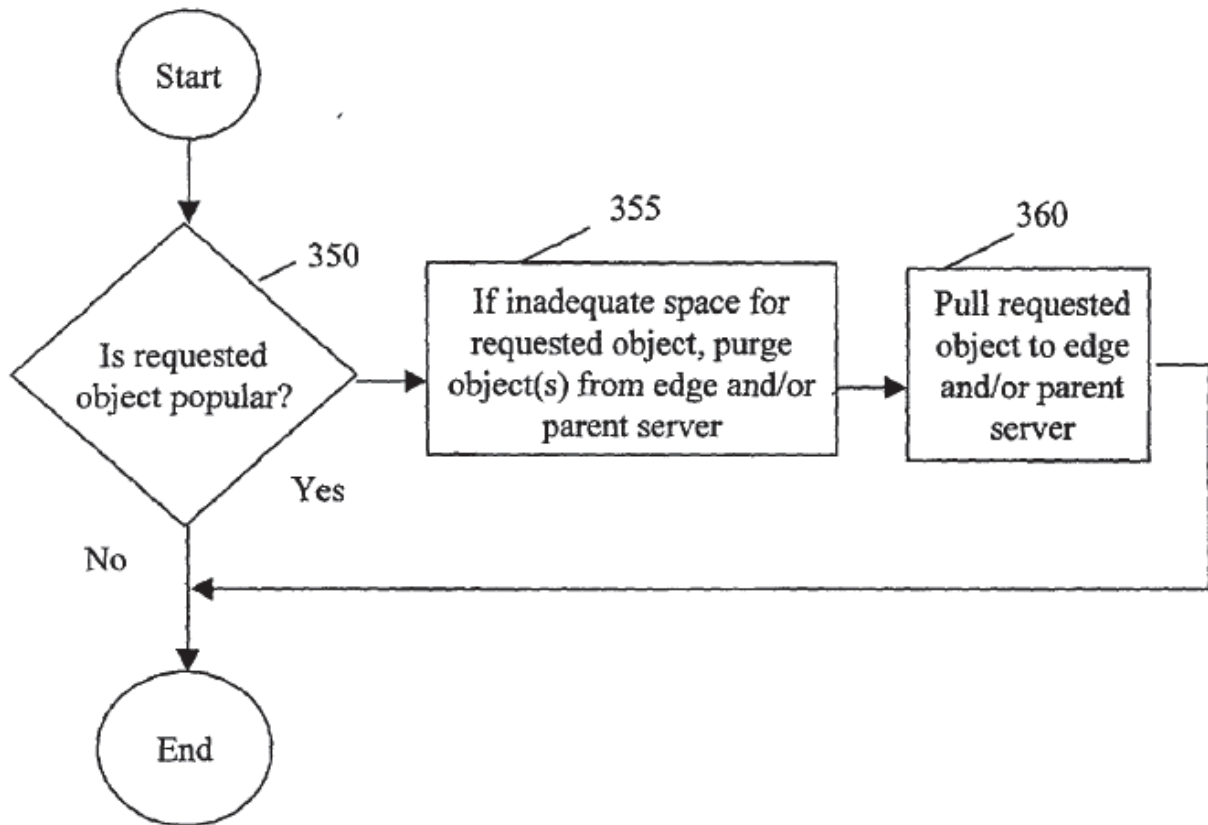


Fig. 3(b)

EX1005, FIG. 3B.

66. As shown in Figure 3B, check step 350 has a “Yes” outcome branch indicating that the requested object is unavailable and popular. Thus, contrary to Petitioner’s assumption, when the requested object is unavailable, it can be popular and does not need to be unpopular. This is a common issue where the requested object has become popular recently, so servers have not had a chance to migrate it. For example, the ’692 patent explains that “[c]ontent can become popular (by

various measures of popularity) or fade into relative obscurity dynamically, so a content library cannot easily be explicitly partitioned.” EX1001, 2:46-48. Seed recognizes this common issue and thus uses the check step 350 to catch objects that are unavailable but popular. However, in contrast to the ’692 patent and Seed, Petitioner and Mr. Lanning’s arguments ignore this common issue and incorrectly assume that all unavailable objects are not popular. Because the Petition’s entire theory is based on this flawed assumption, it is my opinion that the obviousness argument based on Seed-Swildens is incorrect.

67. In **Ground 2**, Petitioner asserts that Swildens discloses the limitation. Pet., 54-55. However, just like Seed, Swildens also explains that the edge server redirects the request when the requested content is unavailable and is not based on whether the content is popular. EX1006, ¶68. Furthermore, Swildens recognizes the common issue of dynamically changing popularity identified by the ’692 patent. For example, Swildens’ “edge server checks the frequency of requests over a certain amount of time (configurable by an administrator) to see if any new files have gained enough popularity to replace another file in the local storage.” EX1006, ¶68. When a new file becomes popular, it may not be available at the edge server. Petitioner and Mr. Lanning’s arguments do not address this issue and, therefore, in my opinion, do not establish that Swildens’ edge server redirects the request when the requested content is not popular. Thus, it is my opinion that

Petitioner fails to show that Swildens teaches directing the client to a second server based on a determination that the resource is not popular.

2. The Petitioner fails to show either combination teaches the claimed “*if the resource is determined to be popular, then the first server obtaining the resource and the first server serving the resource to the client.*”

68. Independent claim 1 recites “*if the resource is determined to be popular, then the first server obtaining the resource and the first server serving the resource to the client.*” Thus, when the resource is popular, the first server takes two actions: “*obtaining*” and “*serving.*” It is my opinion that Petitioner fails to show that either Seed-Swildens or Swildens-Seed renders obvious this feature.

a. Ground 1

69. Petitioner makes several arguments in Ground 1 based on Seed supporting when the resource is popular, the first server takes two actions: “*obtaining*” **and** “*serving.*” Pet., 18-23. It is my opinion that neither argument is correct.

70. **First**, in my opinion, Petitioner’s theory improperly separates the claim elements and relies on two separate actions: (1) in response to a first request for a popular but not-yet-replicated object, obtain the content and (2) in response to a subsequent—and different—request after replication, serving the content. Pet., 19-20. However, these two actions are mutually exclusive because a request is either a first request or a subsequent request, but it cannot be both. Thus, the edge

server either performs “*obtaining*” in response to the first request or performs “*serving*” in response to the subsequent request, not both. Petitioner’s theory thus fails to explain how Seed’s edge server discloses both “*obtaining*” and “*serving*” if the resource is determined to be popular. Thus, in both of Petitioner’s scenarios Seed’s edge server performs only “*serving*” or “*obtaining*.” Petitioner’s obviousness theory fails to explain how Seed discloses both “*obtaining the resource*” and “*serving the resource*” when “*the resource is determined to be popular*.”

71. **Second**, Petitioner alleges that Seed’s edge server is configured to fetch and serve the requested object based on popularity in response to the first request. Pet., 19-20 (“[i]f, however, the claim requires the edge server serve the object to the client after only a first request ...”). I believe this is also incorrect. For the first request, Petitioner argues that “in response to a first request for a popular but not-yet replicated object, the client is redirected to another server that has the object, and the object is then replicated to the edge server for ‘future’ requests for the same object.” Pet., 19. But Seed explains that

[A] request for a streaming object will be served by a handling edge server if that handling edge server has a copy of that object. Otherwise, the request is redirected, directly or indirectly, to a parent server for service of the requested streaming object to the client. If the requested streaming object is popular, the object is replicated from a parent server

that has the requested streaming object to the handling edge server so that the handling edge server will serve the object from the edge of the network when the object is requested in the future.

EX1005, ¶18. In other words, in response to the first request, Seed's edge server *redirects the request* to a parent server to serve the object and then replicates the object for future requests. EX1005, ¶18. In my opinion, when the request is redirected to and served by a parent server, the edge server does not serve the same request for at least two reasons. First, the edge server does not have the requested object yet because the replication is for future requests. Thus, the edge server is incapable of serving the request. Second, when the redirected parent server serves the request, it is not efficient for the edge server to also serve the same request because doing so would increase the resources used for each request. For example, in that case, the requested object would be transmitted to the client twice by the edge server and the parent server respectively, which requires more bandwidth and computational power and is unnecessarily duplicative. The increase in resource use would burden the edge server, the parent server, and the entire network. Thus, I believe that Seed's edge server only performs "*obtaining*," not "*serving*," because the request is redirected to and served by a parent server.

72. Petitioner then relies on Seed's discussion of prior-art solutions that fetch objects on demand. Pet., 20. However, Seed does not disclose or suggest that fetching objects on demand is performed based on the popularity level. EX1005,

¶¶13-14. In fact, Seed does not mention popularity at all when describing these prior-art solutions. *See* EX1005, ¶¶13-14. This is not surprising because in the cited paragraphs, Seed discusses request-handling approaches that existed at a previous time when popularity was not considered. Naturally, the fetching in these approaches is not related to popularity. Therefore, it is my opinion that Petitioner fails to demonstrate that *the first server* performs “*obtaining*” **and** “*serving*” when the resource is popular, as claimed.

73. Attempting to cure this deficiency, Petitioner asserts that “[b]ased on these disclosures, a POSITA would have understood that Seed suggests having edge servers fetch and then serve popular objects.” Pet., 20. However, no explanation is provided to support this conclusory assertion; rather, Petitioner relies on declarant testimony that itself provides no additional supporting evidence for this assertion. *See* Pet., 20; EX1003, ¶113. In my opinion, a POSITA would not have considered popularity in the prior art fetching on demand when reading Seed. Seed explains that

Such fetching suffers however from having to go through the parent path (the network path between the handling edge server and the parent server with the object) **whenever** a client requests an object that is not already at the particular edge server.

EX1005, ¶13 (emphasis added). Thus, the issue here is constant fetching whenever the requested objects are unavailable at the edge server. Seed also explains that

such constant fetching can use too much bandwidth and create congestion.

EX1005, ¶13. As explained in Day, “[w]hen the origin server faces heavy traffic, the burden of managing persistence can adversely affect its ability to respond to client requests.” EX2020, 2:60-62. In response to this issue, Seed explains that

So, where the requested object is not available at the handling edge server, the client's request is redirected directly to the parent server that can provide the requested object to the client or indirectly via one or more parent servers to a parent server that can provide the requested object to the client.

EX1005, ¶17. Thus, Seed’s solution to the issue of constant fetching is to redirect the request. In contrast, the ’692 patent, as recited in claim 1, addresses the issue of constant fetching by performing “*obtaining*” and “*serving*” based on the popularity level so that only popular objects are obtained. In this way, the edge servers would not be overloaded, and resources for obtaining and storing the objects in the edge servers are efficiently used. Thus, I believe the on-demand fetching approach described in Seed lacks any consideration of content popularity, rendering it critically deficient and materially different from the approach required by claim 1.

74. **Finally**, Petitioner asserts that a POSITA would have been motivated to modify Seed to satisfy this element. Pet., 20-21. The Petition argues that there were only three typical ways to fulfill a request for a missing object: (1) obtain (fetch) the requested object from another server and then serve it to the client; (2)

redirect the client to another server that has the object; or (3) forward the request to another server that has the object. Pet., 20-21. However, none of the three typical ways identified by the Petition satisfies the limitation. Thus, modifying Seed based on these alleged typical ways still does not satisfy the limitation.

75. The first typical way involves fetching and serving, which ignores the claimed popular objects feature, as was shown was missing in Seed. Additionally, Seed explains that the fetching and serving approach is costly. EX1005, ¶¶13-15. This was one problem the '692 patent specifically addresses and overcomes by “*obtaining*” and “*serving*” popular objects. EX1001, 2:35-51, 6:48-7:10. As I previously explained, the claimed approach of “*obtaining*” and “*serving*” based on the popularity level obtains a small portion of objects that are popular to avoid overloading the edge servers and achieve efficient use of resources. Thus, the first typical way fails to meet the claim.

76. The second and third alleged typical ways are both related to redirecting the request and thus are not relevant to the claimed “*obtaining*” and “*serving*” the resource “*if the resource is determined to be popular.*”

77. Thus, it is my opinion that Ground 1 does not establish that claim 1 is obvious.

b. Ground 2

78. Petitioner makes several arguments for when the resource is popular,

the first server takes two actions: “*obtaining*” **and** “*serving*” in Ground 2 based on Swildens. Pet., 51-54. It is my opinion that these arguments do not show that claim 1 is obvious for similar reasons as the arguments based on Seed.

79. **First**, like Ground 1, Petitioner’s arguments are based on two separate actions: (1) in response to a first request for a popular but not-yet-replicated object, and (2) in response to subsequent requests after replication. Pet., 51-52. However, these two actions are mutually exclusive because a request is either a first request or a subsequent request, but it cannot be both. Thus, the edge server either performs “*obtaining*” in response to the first request or performs “*serving*” in response to the subsequent request, not both. Petitioner’s theory thus fails to explain how Seed’s edge server discloses both “*obtaining*” and “*serving*” if the resource is determined to be popular. In both of Petitioner’s scenarios Seed’s edge server performs only “*serving*” or “*obtaining*.” Petitioner’s obviousness theory fails to explain how Seed discloses both “*obtaining the resource*” and “*serving the resource*” when “*the resource is determined to be popular*.”

80. **Second**, Petitioner argues that a POSITA would have been motivated to modify Swildens to fetch popular content rather than redirect clients based on Seed’s disclosure of fetching. Pet., 52-53. Petitioner again relies on Seed’s disclosure of fetching on demand in paragraph 13 to meet the claims. Pet., 52-53. As I discuss for Ground 1, Seed describes prior-art solutions that fetch objects on

demand but does not disclose that the fetching is performed based on the popularity level. EX1005, ¶¶13-14. In fact, Seed does not mention popularity at all when describing these prior-art solutions. *See* EX1005, ¶¶13-14. As I explained in Ground 1, Seed’s prior art solutions lack the consideration of popularity level as offered by the ’692 patent, which makes it critically deficient and materially different from the approach required by claim 1. Therefore, it is my opinion that Petitioner fails to demonstrate that *the first server* performs “*obtaining*” **and** “*serving*” when *the resource is popular*, as claimed.

81. Finally, Petitioner asserts that a POSITA would have been motivated to modify Swildens to satisfy this element. Pet., 53-54. Petitioner argues that there were only three typical ways for a server to ensure that requests for an object missing from its cache are fulfilled: (1) fetch-and-serve; (2) redirect the client; or (3) forward the request. Pet., 53. However, none of the three typical ways identified by the Petition satisfies the limitation of the claims.

82. The first typical way involves fetching and serving, which ignores the claimed popular objects feature, as was shown to be missing in Swildens. Thus, the first typical way fails to meet the claim. The second and third alleged typical ways are both related to redirecting the request, and thus are not relevant to the claimed “*obtaining*” and “*serving*” the resource “*if the resource is determined to be popular*.” Because none of the three typical ways satisfy the claim limitation,

modifying Swildens based on these alleged typical ways does not satisfy the limitation.

83. Accordingly, Petitioner fails to show the Swildens-Seed combination teaches “*obtaining*” and “*serving*” the resource “*if the resource is determined to be popular.*”

C. Independent Claim 10

84. In my opinion, Petitioner does not establish that claim 10 is obvious for two additional reasons.

1. Petitioner fails to show Seed-Swildens or Swildens-Seed renders obvious the claimed “*provide the edge servers with a second set of information about which tier of servers should handle those requests for content.*”

85. Independent claim 10 recites “*at least one popularity service constructed and adapted ... to provide the edge servers with a second set of information about which tier of servers should handle those requests for content.*”

Petitioner admits that Seed does not expressly disclose a separate “popularity service” that receives information from edge servers and provides information back to edge servers. Pet., 36. Instead, for both Grounds 1 and 2, Petitioner relies on Swildens for purportedly teaching this feature. Pet., 36-37, 69. Yet Petitioner then simply provides a single sentence that wrongly asserts Swildens’ auto migration is the same as the claimed “*provid[ing] the edge servers with a second set of information about which tier of servers should handle those requests for*

content.” See Pet., 36-37, 69.

86. Swildens auto migration and the claimed “*provid[ing] the edge servers with a second set of information about which tier of servers should handle those requests for content*” are two disparate operations. Swildens’ auto migration server selects the content and performs the auto migration without providing any information to the edge servers. For example, Swildens describes that at a specified time interval, content that falls above a threshold value is sorted and the “top few” are selected for auto migration. EX1006, ¶126. In my opinion, a POSITA would have understood that Swildens’ auto migration is a data transmission process that transfers content to edge servers from other servers. EX1006, Abstract. On the other hand, providing tier information required by claim 10 is a notification process that gives instructions to an edge server about which tier of servers the request is redirected to. EX1001, 6:48-7:10. Furthermore, Swildens’ auto migration server does not provide information to the edge server during the auto migration process. Swildens recites that “[f]or all the content that falls above this threshold value, the daemon [of auto migration server] sorts them based on the miss count and then selects the top few (say 20) for auto migration.” EX1006, ¶126. Thus, Swildens’ auto migration server selects the content and performs the auto migration by transferring the selected content to the edge servers. Here, the edge servers receive the selected content itself, not information about the selected

content. There is no need for the edge servers to know what content is being transferred because that decision has already been made by the auto migration server.

87. This is further evident by Swildens' characterization of the auto migration server. For example, Swildens explains that "[a] centralized server can periodically collect logs from all of the edge serve[r]s." EX1006, ¶127. To do that, "[a] single auto migration daemon can combine data from multiple edge servers located in a single edge location to determine the most popular content" and "the popularity is based on the aggregate data from all of the edge servers that are accessing that common storage." EX1006, ¶130. Thus, centralization is the fundamental concept of the auto migration process. In my opinion, providing "*the edge servers with a second set of information about which tier of servers should handle those requests for content*" would defeat the entire reason for Swildens' invention.

88. Furthermore, even if Swildens' auto migration server provides information to the edge server, it would not be the claimed "*second set of information about which tier of servers should handle those requests for content.*" Swildens explains that the auto migration server selects a list of content, such as the top 20 popular content. EX1006, ¶126. In my opinion, a POSITA would have understood that to provide a list of content, the auto migration server can provide

identifications corresponding to the content in the list. However, these identifications do not point to a tier of servers or any servers, and thus cannot be information about a tier of servers. Thus, in my opinion, any information provided by Swildens' auto migration server would be that list of content, not "*a second set of information about which tier of servers should handle those requests for content,*" as recited in claim 10.

89. In sum, it is my opinion that Petitioner fails to show "*at least one popularity service constructed and adapted ... to provide the edge servers with a second set of information about which tier of servers should handle those requests for content,*" and thus fails to establish Seed-Swildens or Swildens-Seed teaches this element of independent claim 10.

2. Petitioner fails to show Seed-Swildens or Swildens-Seed renders obvious the claimed "*determin[ing a tier of servers to handle a request] based, at least in part, on a measure of popularity of requested content.*"

90. Independent claim 10 recites "*the popularity service making its determinations,*" which includes "*which tier of servers should handle those requests for content,*" "*based, at least in part, on a measure of popularity of requested content.*" The Petition admits that Seed does not disclose this feature, but attempts to address the deficiency using Swildens in both Grounds 1 and 2. Pet., 36, 69. In my opinion, Swildens does not teach the claimed "*which tier of servers should handle those requests for content,*" "*based, at least in part, on a measure of*

popularity of requested content.”

91. Swildens uses popularity level to determine which content to auto-migrate, not the claimed “*which tier of servers should handle those requests for content.*” EX1006, ¶126. Rather than addressing this mismatch, Petitioner simply equates, without any explanation, Swildens’ tabulation and selection of content for auto-migration with the claimed determining a tier of server to handle the request. Pet., 36-37, 69. However, this conclusion is incorrect and contradicted by Swildens. In my opinion, determining a list of content to auto-migrate is different from the claimed determining “*which tier of servers should handle those requests for content*” because a list of content is digital data, but servers are physical devices. In other words, digital data is virtual and is different from servers that physically exist. Thus, I believe determining digital data, such as the list of content, is different from determining “*which tier of servers should handle those requests for content,*” as recited in claim 10.

92. Furthermore, Petitioner does not address Swildens’ explicit explanation of how it determines which server to handle a request. Swildens explains that when its edge server receives a request, it redirects the user to an origin server if the edge server does not have the requested content. EX1006, ¶46. If the origin server does not have the requested content, it redirects the user to a staging server. EX1006, ¶47. Thus, it is my opinion that Swildens determines

which server to handle a request based on availability (with respect to redirecting the user to a staging server), not the claimed “the *popularity service making its determinations.*”

93. In my view, Petitioner thus fails to establish either Ground 1 or 2 renders obvious “the *popularity service making its determinations based, at least in part, on a measure of popularity of requested content.*”

D. Independent Claim 15: Seed and Swildens, in any order and for both Grounds, do not teach the claimed “*redirecting being based, at least in part, on a current popularity value for the resource.*”

94. Independent claim 15 requires that “the *redirecting being based, at least in part, on a current popularity value for the resource.*” It is my opinion that the Petition fails to show that Seed-Swildens or Swildens-Seed teaches this feature.

95. To begin with, claim 15 does not have multiple meanings. Petitioner misconstrues the relationship between terms in claims 15 and 18. Pet., 40-44. Petitioner assumes that the “*second tier server*” is a server in the “*second tier.*” Pet., 41. In contrast, element 15[b] recites “the *second tier server being in a tier distinct from the tier containing the first server*” (i.e., it does not recite the second tier server is in the second tier, but rather that it is a server in a tier distinct from the tier of the first server). Claim 18 further recites that “the *first server is a parent server in the second tier and the second [tier] server is a server in the origin tier.*” Thus, claim 18 also places the “*first server*” and the “*second tier server*” in

different tiers, which does not contradict element 15[b].

96. Regardless, Petitioner’s obviousness theory for element 15[b] relies on Seed to teach “*redirecting would be based, at least in part, on a current popularity value for the resource.*” Pet., 44 (citing EX1005, ¶26). But, as I discuss in Section IX.B.1, this is not what Seed discloses. Rather, Seed discloses that redirecting is done because “the selected edge server does not have the requested object,” which applies equally to both popular and unpopular content according to Seed. EX1005, ¶26. Thus, Petitioner fails to show element 15[b] is obvious over Swildens-Seed or Seed-Swildens.

X. CONCLUSION

97. In signing this declaration, I recognize that the declaration will be filed as evidence in an *inter partes* review before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in the case and that cross-examination will take place within the United States. If cross-examination is required, I will appear for cross-examination within the United States during the time allotted.

98. I hereby 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 further that these statements were made with the 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.

Executed on this 19 day of March 2026, in Los Angeles, CA.



Prashant Shenoy, Ph.D.