Trial@uspto.gov
571-272-7822

Paper 18
Entered: May 17, 2013

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

EMC CORPORATION
Petitioner,

v.

PERSONALWEB TECHNOLOGIES, LLC
Patent Owner.

Case IPR2013-00085 (JYC)
Patent 7,945,539

Before KEVIN F. TURNER, JONI Y. CHANG, and MICHAEL R. ZECHER, *Administrative Patent Judges*.

CHANG, *Administrative Patent Judge*

DECISION
Institution of *Inter Partes* Review
37 C.F.R. § 42.108

I. INTRODUCTION

EMC Corporation ("EMC") filed a petition, requesting an *inter partes* review of U.S. Patent 7,945,539 ("the '539 patent"). (Paper 5, "Pet.") In response, PersonalWeb Technologies, LLC ("PersonalWeb") filed a patent owner preliminary response. (Paper 11, "Prel. Resp.") We have jurisdiction under 35 U.S.C. § 314.

The standard for instituting an *inter partes* review is set forth in 35 U.S.C. § 314(a) which provides:

> THRESHOLD -- The Director may not authorize an inter partes review to be instituted unless the Director determines that the information presented in the petition filed under section 311 and any response filed under section 313 shows that there is a reasonable likelihood that the petitioner would prevail with respect to at least 1 of the claims challenged in the petition.

Upon consideration of the petition and patent owner preliminary response, we determine that the information presented in the petition demonstrates that there is a reasonable likelihood that EMC would prevail with respect to the challenged claims. Accordingly, we authorize an *inter partes* review to be instituted for claims 10, 21, and 34 of the '539 patent.

A. Related Proceedings

EMC indicates that the '539 patent is the subject of litigation styled *PersonalWeb Technologies LLC v. EMC Corporation and VMware, Inc.*, No 6:11-cv-00660-LED (E.D. Tex.). (Pet. 1.)

EMC also filed five other petitions seeking *inter partes* review of the following patents: Patent 5,978,791 (IPR2013-00082), Patent 6,415,280

(IPR2013-00083), Patent 7,945,544 (IPR2013-00084), Patent 7,949,662 (IPR2013-00086), and Patent 8,001,096 (IPR2013-00087). (Pet. 1.) According to EMC, those patents and the '539 patent share a common disclosure. (*Id*. citing to Ex. 1008.)

The '539 patent claims the benefit of various applications and patents under 35 U.S.C. § 120, but the earliest priority date claimed by the '539 patent is April 11, 1995, the filing date of U.S. patent application No. 08/425,160 (now abandoned). (Ex 1001, front page; *see also* Ex. 1008.)

B. The '539 patent

The '539 patent relates to a method for identifying and obtaining access to a data item (*e.g.*, a data file or record) in a data processing system, by using an identifier which depends on all of the data in the data item and only on the data in the data item. (Ex. 1001, 1:45-48; 3:52-56; 4:11-13.) Thus, the identity of a data item is said to be independent of its name, origin, location, and address. (Ex. 1001, 3:55-58.) According to the '539 patent, the system provides transparent access to any data item by reference only to its identity and independent of its present location (*e.g.*, local, remote, or offline). (Ex. 1001, 4:11-13.)

For a simple data item (a data item whose size is less than a particular given size) (S216 and S218), a data identifier (True Name) is computed using a function (*e.g.*, a message digest ("MD") function, such as MD4 or MD5, or a secure hash algorithm ("SHA") function). (Ex. 1001, 14:24-50, 15:37-48, Figs. 10(a) & 10(b).) As a result, a data item that has an arbitrary

length is reduced to a relatively small, fixed size identifier (True Name) that represents the data item. (*Id.*)

Figure 10(b) of the '539 patent, reproduced below, is a flow chart for determining an identifier (True Name) of a simple or compound data item.

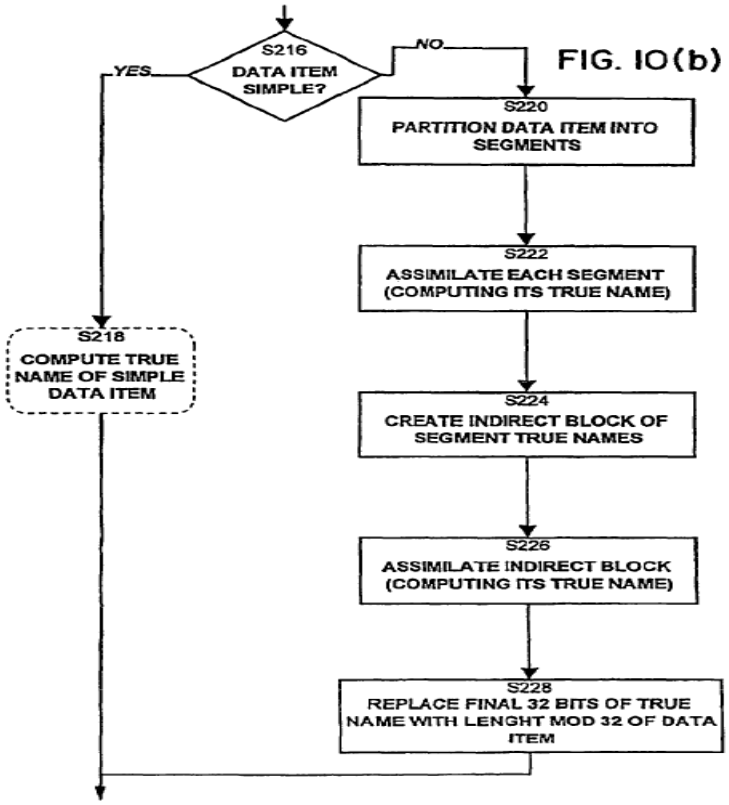

FIG. IO(b)

As shown in Figure 10(b), if the data item is a compound data item (a data item whose size is greater than the particular given size), the system will partition the data item into segments (S220); assimilate each segment (S222); compute the True Name of the segment; create an indirect block consisting of the computed segment True Names (S224); assimilate the indirect block (S226); and replace the final 32 bits of the resulting True Name by the length modulo 32 of the compound data item (S228).

(Ex. 1001, 15:49-67, Fig. 10(b).) The result is the True Name of the compound data item. (*Id*.)

C. Illustrative Claim

All of the challenged claims are independent claims. For the purposes of this decision, claim 21 is illustrative and reproduced as follows (emphasis added):

> 21. A computer-implemented method of obtaining access to a data item at a first computer in a network of computers, said data item comprising a plurality of segments, each of said plurality of segments being stored on at least one of a plurality of computers in said network, said plurality of computers being distinct from said first computer, the method comprising the steps of:
>
> (A) by hardware in combination with software, using a ***first data identifier*** to obtain a ***plurality of segment identifiers***, each of said segment identifiers corresponding to one of said plurality of segments, the segment identifier for each particular segment being based at least in part on a first given function of the data comprising said particular segment and only the data in said particular segment, where any two identical segments will have identical segment identifier as determined using said first given function, and
>
> ***wherein said first data identifier is based, at least in part, on a second given function of data comprising the plurality of segment identifiers***;
>
> (B) using the plurality of segment identifiers obtained in step (A) to obtain at least one of said plurality of segments by, for at least one particular segment identifier of said plurality of segment identifiers:
>
> (b0) using said particular segment identifier to ascertain one or more locations in said network of computers that should have the corresponding particular segment;

(bl) using said particular segment identifier to request said corresponding particular segment from at least one of said one or more locations ascertained in step (b0); and

(b2) obtaining said corresponding particular segment from at least one location in said network.

D. Prior Art Relied Upon

EMC relies upon the following prior art references:

Fischer	U.S. Patent 5,475,826[1]	Dec. 12, 1995	(Ex. 1036)
Woodhill	U.S. Patent 5,649,196[2]	July 15, 1997	(Ex. 1005)

Albert Langer, "*Re: dl/describe (File descriptions)*," post to the "alt.sources" newsgroup on Aug. 7, 1991 ("Langer," Ex. 1003)

Frederick W. Kantor, "*FWKCS (TM) Contents-Signature System Version 1.22*," FWKCS122.REF (Aug. 10, 1993) ("Kantor," Ex. 1004)

S. Browne et al., "*Location-Independent Naming for Virtual Distributed Software Repositories*," University of Tennessee Technical Report CS-95-278 (Feb. 1995) ("Browne," Ex. 1002)

E. The Asserted Grounds

EMC challenges the patentability of claims 10, 21, and 34 of the '539 patent based on the following grounds:

1. Claims 10, 21, and 34 are unpatentable under 35 U.S.C. § 102(a) as

[1] Fischer claims the benefit of U.S. patent application No. 08/154,520, filed on Nov. 19, 1993.

[2] Woodhill claims the benefit of U.S. patent application No. 08/085,596, filed on July 1, 1993.

anticipated by Browne;

2. Claim 34 is unpatentable under 35 U.S.C. § 103(a) over Browne and Woodhill;
3. Claims 10, 21, and 34 are unpatentable under 35 U.S.C. § 102(b) as anticipated by Langer;
4. Claim 34 is unpatentable under 35 U.S.C. § 103(a) over Langer and Woodhill;
5. Claims 10, 21, and 34 are unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor;
6. Claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor and Browne;
7. Claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor and Langer;
8. Claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Woodhill and Fischer; and
9. Claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Woodhill and Browne.

II. ANALYSIS

A. Claim Construction

As a first step in our analysis for determining whether to institute a trial, we determine the meaning of the claims. In an *inter partes* review, claim terms in an unexpired patent are given their broadest reasonable construction in light of the specification of the patent in which they appear.

37 C.F.R. § 42.100(b). Under the broadest reasonable construction standard, claims are to be given their broadest reasonable interpretation consistent with the specification, and the claim language should be read in light of the specification as it would be understood by one of ordinary skill in the art. *In re Am. Acad. of Sci. Tech. Ctr.*, 367 F.3d 1359, 1364 (Fed. Cir. 2004). In that regard, an inventor is entitled to be his or her own lexicographer of a patent claim term by providing a definition of the term in the specification with reasonable clarity, deliberateness, and precision. *In re Paulsen*, 30 F.3d 1475, 1480 (Fed. Cir. 1994). However, we must be careful not to read a particular embodiment appearing in the written description into the claim if the claim language is broader than the embodiment. *In re Van Geuns*, 988 F.2d 1181, 1184 (Fed. Cir. 1993).

The parties submit a claim construction for each of the following claim terms: (1) "data" and "data item," (2) "data identifier," (3) "True Name," and (4) "location." (Pet. 5-6; Prel. Resp. 3-5.) We address each of these claim terms in turn.

1. "Data" and "Data item" (Claims 10, 21, and 34)

The parties appear to agree that the term "data item" should be construed as "sequence of bits." (Pet. 5-6; Prel. Resp. 3.) EMC also directs our attention to the following portions of the specification of the '539 patent (Pet. 5-6):

> Thus a *data item* may be the contents of a *file*, a portion of a file, a *page* in memory, an *object* in an object-oriented program, a digital *message*, a digital scanned *image*, a part of a *video* or

> audio *signal*, or any *other entity* which can be represented by a sequence of bits.

(Ex. 1001, 2:16-21, emphasis added.)

> In all of the prior data processing systems the names or identifiers provided to identify *data items (the data items being files, directories, records in the database, objects in object-oriented programming, locations in memory or on a physical device, or the like*) are always defined relative to a specific context.

(Ex. 1001, 2:26-38, emphasis added.)

Upon consideration of the parties' constructions and the specification of the '539 patent, we construe the claim term "data item" as "sequence of bits," which includes one of the following: (1) the contents of a file; (2) a portion of a file; (3) a page in memory; (4) an object in an object-oriented program; (5) a digital message; (6) a digital scanned image; (7) a part of a video or audio signal; (8) a directory; (9) a record in a database; (10) a location in memory, on a physical device, or the like; and (11) any other entity which can be represented by a sequence of bits.

Further, we note that claims 10 and 21 recite the claim language "said data item comprising a plurality of segments" and "the *data* comprising said particular segment and only the *data* in said particular segment" (emphasis added). As such, these claims treat the claim terms "data" and "data item" as separate and distinct elements. "In the absence of any evidence to the contrary, we must presume that the use [of] different terms in the claims connotes different meanings." *CAE Screenplates Inc. v. Heinrich Fiedler GmbH & Co. KG*, 224 F.3d 1308, 1317 (Fed. Cir. 2000). Therefore, we

construe the claim term "data" as a subset of a "data item."

2. "Data identifier" (Claim 21)

The specification of the '539 patent provides that "[i]n the following, the terms 'True Name', 'data identity' and '*data identifier*' refer to the *substantially unique data identifier* for a particular data item." (Ex. 1001, 6:20-22, emphasis added.) We observe that the claim term "data identifier" itself is part of the definition "substantially unique *data identifier*," and two different claim terms, namely "data identifier" and "True Name," have the same meaning.

We find it necessary to ascertain the scope of the term "substantially unique data identifier." Ordinarily, the word "identifier" means "any text string used as a label."[3] The discussion in the Background section of the '539 patent also provides a context as to how a person of ordinary skill in the art would have understood the term "identifier" (Ex. 1001, 1:50-2:5, emphasis added):

> Data processing (DP) systems, computers, networks of computers, or the like, typically offer users and programs *various ways to identify the data* in the systems.
>
> Users typically identify data in the data processing system by giving the data some form of name. For example, a typical operating system (OS) on a computer *provides a file system in which data items are named by alphanumeric identifiers*. Programs typically identify data in the data

[3] *See e.g.*, *Microsoft Computer Dictionary* at 264 (5th ed. 2002).

> processing system using a location or address. For example, a program may identify a record in a file or database by using a record number which serves to locate that record.

Taking into account the specification, we construe the term "substantially unique data identifier" broadly, but reasonably as "a substantially unique alphanumeric label for a data item." Consequently, we determine the claim term "data identifier" to mean "a substantially unique alphanumeric label for a particular data item."

Nothing in the language of the term "substantially unique data identifier" or the claim term "data identifier" suggests that a particular generating or calculating method is required. Nevertheless, PersonalWeb asserts that these terms should be interpreted as "an identity for a data item generated by *processing all of the data in the data item, and only the data in the data item, through an algorithm*." (Prel. Resp. 4 and 15.)

To support its assertion, PersonalWeb contends that the term "substantially unique data identifier" already has been construed in this manner regarding U.S. Patent 5,978,791 ("the '791 patent") by the U.S. District Court for the District Massachusetts. (Prel. Resp. 4.) The *'791 patent* is said to be a family member of the '539 patent, the involved patent, of the instant proceeding. (*Id*.) PersonalWeb also alleges that EMC also agrees with this construction given EMC's identical construction of this term in its petition in *IPR2013-00082*. (*Id*.)

We decline to adopt PersonalWeb's construction, as it would import limitations from the claims of the *'791 patent* into the claims of the '539

patent. In *IPR2013-00082*, every challenged independent claim of the *'791 patent* expressly recites the following claim language:

> a substantially unique identifier, the identifier being determined using and depending on all the data in the data item and only the data in the data item

(Claims 1, 30, and 33 of the '791 patent.)

In contrast, none of the challenged claims in the instant proceeding recite this language or limitation. Most importantly, PersonalWeb's proposed construction would be inconsistent with the language of claim 21 of the '539 patent, which provides "wherein said first *data identifier* is based, at least *in part*, on a second given function of data comprising the plurality of *segment identifiers*," rather than requiring "*only* the data in the *data item*." Further, we have reviewed the cited portions of the specification of the '539 patent, and we do not find a special definition for the term "substantially unique identifier" in the '539 patent.

Therefore, PersonalWeb's proposed construction for the claim term "data identifier" as recited in the challenged claims of '539 patent is overly narrow.

3. "True Name" (Claim 34)

PersonalWeb alleges that the claim term "True Name" should be construed as "substantially unique identifier for a particular data item, *calculated in accordance with the description at col. 14:24-45*." (Prel. Resp. 4-5, emphasis added.) PersonalWeb fails to provide a sufficient explanation or credible evidence to establish such a claim construction. (Prel. Resp. 4-5,

citing to Ex. 1001, 6:20-22; 14:24-45.) Upon reviewing the cited portion of the specification, we do not find an explicit or special definition for the claim term "True Name" that would support PersonalWeb's assertion.

In light of the specification of the '539 patent, we construe the claim term "True Name" broadly, but reasonably as a "substantially unique alphanumeric label for a particular data item."

4. "Location" (Claims 21)

Both parties construe the claim term "location" with respect to a data processing system as "any of a particular processor in the system, a memory of a particular processor, a storage device, a removable storage medium (such as a floppy disk or compact disk), or any other physical location in the system." (Pet. 6; Prel. Resp. 3-4, citing to Ex. 1001, 6:12-17). Because that claim construction is consistent with the specification of the '539 patent, we will adopt it as our own.

B. Whether Kantor and Langer are "Printed Publications"

As an initial matter, we address the issue of whether Kantor and Langer are available as prior art for the purposes of this decision. EMC contends that Kantor and Langer are prior art under 35 U.S.C. § 102(b). (Pet. 35-36 and 42.) According to EMC, "Kantor is a published manual that describes a software program called the Frederick W. Kantor Contents-Signature System Version 1.22 ('FWKCS')" and has been publicly available since August 1993. (Pet. 4 and 42, citing to Ex. 1004, Title Page.) EMC

also submits that Langer was made available on the "alt.sources.d" and "comp.archives.admin" newsgroup distribution lists on August 7, 1991. (Pet. 3, n2.)

In its preliminary response, PersonalWeb urges the Board to deny the asserted grounds of unpatentability on the basis that Kantor and Langer are not prior art "printed publications." (Prel. Resp. 5-10.) In particular, PersonalWeb argues that EMC has presented no testimony, declaration, or other evidence that either reference "was catalogued or indexed in a meaningful way prior to the critical date, or that [it] would have turned up in a customary search prior to the critical date, or that persons interested and ordinarily skilled in the art exercising reasonable diligence would have located [it] prior to the critical date." (Prel. Resp. 7.)

We are not persuaded by PersonalWeb's arguments. Rather, on this record, we determine that EMC has made a threshold showing to establish that Kantor and Langer are "printed publications" within the meaning of 35 U.S.C. § 102(b). As a consequence, Kantor and Langer are available as prior art for the purposes of this decision to demonstrate that the challenged claims are unpatentable under 35 U.S.C. §§ 102(b) and 103(a).

To determine whether to deny a ground on the basis that a reference is not a "printed publication," we decide each case on the basis of its own facts. More specifically, the determination of whether a given reference qualifies as a prior art "printed publication" involves a case-by-case inquiry into the facts and circumstances surrounding the reference's disclosure to

members of the public. *In re Klopfenstein*, 380 F.3d 1345, 1350 (Fed. Cir. 2004).

Here, EMC asserts that Kantor has been publicly available since August 1993, which is one year before April 11, 1995, the earliest priority date claimed by the '539 patent. (Pet. 3-4, citing to Ex. 1004.) EMC also provides the following explanation (*id*.):

> Kantor's FWKCS user manual has been publicly and freely available continuously since August 1993. Kantor distributed the user manual with the FWKCS program as shareware and posted it online to electronic Bulletin Board Systems including "The Invention Factory" and "Channel 1" for an extended period of time, where it could be downloaded by anyone. As such, the document was accessible to others in the relevant community of BBS users and system operators. (*See* Kantor at 3; *see also* 158-59; Ex. 1004.)

Further, the title page of Kantor clearly shows the posted date of August 10, 1993. (Ex. 1004, Title Page "FWKCS (TM) Contents_Signature System[,] Version 1.22[,] **1993 August 10** [,] (C) Copyright Frederick W. Kantor 1988-1993." Emphasis added.) Kantor also provides the following:

> The FWKCS(TM) Contents_Signature System has become a robust platform for supporting contents_signature functions. FWKCS provides many functions and options for application in a public, commercial, school, institutional, or governmental environment. Extensive technical support is of special value in helping such users to benefit more fully from these many features.
>
> Registered FWKCS hobby BBS users are able to receive a modest amount of assistance, and are invited to participate in

> the FWKCS conference on The Invention Factory BBS, echoed via Execnet.
>
> Commercial, school, institutional, and governmental users, with their special support needs, are invited to discuss terms for obtaining such assistance.
>
> To get a new version of FWKCS, download FWKCSnnn.ZIP from The Invention Factory BBS, where nnn is the new version number without a decimal point. These special downloads are available at no fee, from a 43_line hunt_up group of USR Dual Standard modems, at 2400-16800 bits/sec (including V32.bis)[.]

(Ex. 1004, 158-159.)

Given that disclosure, Kantor appears to convey that the reference was posted on a publicly accessible site well known to those interested in the art – the electronic Bulletin Board Systems – and could be downloaded and retrieved from that site. *In re Wyer*, 655 F.2d 221, 226 (CCPA 1981) (An electronic publication, including an on-line database or Internet publication, is considered to be a "printed publication" "upon a satisfactory showing that such document has been disseminated or otherwise made available to the extent that persons interested and ordinarily skilled in the subject matter or art, exercising reasonable diligence, can locate it and recognize and comprehend therefrom the essentials of the claimed invention without need of further research or experimentation.").

As to Langer, the header on the first page of Langer, reproduced below, indicates that Langer was distributed on August 7, 1991 to the

newsgroups "alt.sources.d" and "comp.archives.admin" (Ex. 1003, 1):

From: cmf851@anu.oz.au (Albert Langer)
Newsgroups: alt.sources.d,comp.archives.admin
Subject: Re: dl/describe (File descriptions) posted to alt.sources
Message-ID: <1991Aug7.225159.786@newshost.anu.edu.au>
Date: 7 Aug 91 22:51:59 GMT
References: <1991Aug7.124457.6814@csv.viccol.edu.au>
<1991Aug7.131048.6817@csv.viccol.edu.au>
Sender: news@newshost.anu.edu.au
Followup-To: comp.archives.admin
Organization: Computer Services Centre, Australian National University,
Canberra, Australia.
Lines: 291

EMC asserts that those newsgroups were disseminated widely and readily accessible to the relevant technical community, and the newsgroups were devoted to technical discussions relating to source code repository and computer archive administration. (Pet. 3, n2.) Therefore, it also appears that Langer was posted on publicly accessible sites well known to those interested in the art – the newsgroups "alt.sources.d" and "comp.archives.admin" – and could be downloaded and retrieved from those sites.

PersonalWeb cites *Synopsys, Inc. v. Mentor Graphics Corp.*, IPR2012-00042 (Paper No. 16), at *35-36 (PTAB Feb. 22, 2013) for the proposition that any asserted grounds of unpatentability based on an electronic reference should be denied, unless the reference is presented in the petition with a declaration from one of the authors or other evidence that someone accessed or received the reference prior to the critical date. (Prel. Resp. 5-8.) PersonalWeb's reliance on *Synopsys* is misplaced because *Synopsys* did not involve a reference, such as Kantor or Langer, which has a

posted date. In fact, *Synopsys* involves a brochure that did not include any indication of when it was created or whether it was disseminated publicly, and the only evidence submitted by the petitioner was that it was cited in an Information Disclosure Statement filed in an unpublished patent application. *Synopsys*, IPR2012-00042 (Paper No. 16), at *35.

As to PersonalWeb's argument that there is no evidence that either reference was catalogued or indexed in a meaningful way prior to the critical date, we are not convinced. "[W]hile often relevant to public accessibility, evidence of indexing is not an absolute prerequisite to establishing online references [] as printed publications within the prior art." *Voter Verified, Inc. v. Premier Election Solutions, Inc.*, 698 F.3d 1374, 1380 (Fed. Cir. 2012).

PersonalWeb further argues that Kantor and Langer are inadmissible evidence, as the copy of the references submitted by EMC has not been authenticated or certified. (Prel. Resp. 9.) In that regard, PersonalWeb has not followed the proper procedures for objecting to and/or excluding evidence. *See* 37 C.F.R. § 42.64(b); *LKQ Corp. v. Clearlamp, LLC*, IPR2013-00020 (Paper No. 17), at *3-4 (PTAB Mar. 5, 2013).

Under the procedure set forth in 37 C.F.R. § 42.64(b), when a party objects to evidence that was submitted during a preliminary proceeding, such an objection must be served within ten business days of the institution of trial. The objection to the evidence must identify the grounds for the objection with sufficient particularity to allow correction in the form of supplemental evidence. This process allows the party relying on the

evidence to which an objection is served timely, the opportunity to correct, by serving supplemental evidence within so many days of the service of the objection. If, upon receiving the supplemental evidence, the opposing party is still of the opinion that the evidence is inadmissible, the opposing party may file a motion to exclude such evidence. The time for filing a motion to exclude is typically several months into a trial. *See, e.g., Office Patent Trial Practice Guide*, 77 *Fed. Reg*. 48756, 48768-69, Scheduling Order – Due Date 4. Therefore, PersonalWeb will have full opportunity to object, serve, and reconsider any supplemental evidence and, finally, file a motion to exclude evidence.

C. Claims 10, 21, 34 – Grounds Based on Langer

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 102(b) as anticipated by Langer, and that claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Langer and Woodhill. (Pet. 35-41.) In support of its assertions, EMC provides detailed explanations as to how each claim limitation is met by the cited references, including a declaration of Professor Douglas W. Clark[4] ("Dr. Clark") and claim charts. (Pet. 35-41, citing to Ex. 1009, ¶¶ 26-31; Ex. 1046.)

PersonalWeb opposes and argues that Langer does not describe all of the elements recited in each challenged claim. (Prel. Resp. 11-20.)

[4] Dr. Clark has a Ph.D. in computer science and extensive experience in computer systems architecture and design. (Ex. 1009, ¶¶1-6.) We conclude that Dr. Clark is qualified to testify as to the understanding of a person of ordinary skill in the art.

However, PersonalWeb does not challenge the obviousness ground of unpatentability based on Langer and Woodhill. (Prel. Resp. 20.)

PersonalWeb's arguments are not persuasive. Rather, we determine that EMC has made a threshold showing to establish that claims 10 and 21 are anticipated by Langer, and that claim 34 is obvious over Langer in view of Woodhill.

"A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros. v. Union Oil Co. of California*, 814 F.2d 628, 631 (Fed. Cir. 1987). An anticipatory reference is not required to duplicate word for word what is in the claims. *Standard Havens Prods, Inc. v. Gencor Indus, Inc.*, 953 F.2d 1360, 1369 (Fed. Cir. 1991). Further, a reference anticipates a claim if it discloses the claimed invention "such that a skilled artisan could take its teachings *in combination with his own knowledge of the particular art and be in possession of the invention*." *In re Graves*, 69 F.3d 1147, 1152 (Fed. Cir. 1995) (emphasis in the original)(citation omitted). In other words, the prior art references must be considered from the perspective of a person of ordinary skill in the art.

Langer discloses a method of accessing files in a network of computers. (Ex. 1003, 3.) For instance, a file request may be embedded in a news article and includes a unique identifier for the file. (Ex. 1003, 3-4.) As a result, users are automatically informed of the nearest location of the file. (*Id*.)

Langer further discloses that a unique identifier for a file is calculated using a hash function (*e.g.*, MD5, a cryptographic hash function) on the entire contents of the file, rather than the file's location. (Ex. 1003, 2-3.) For a package (*e.g.*, an archive) that is divided into its component files, a unique identifier for each component file is calculated by using an MD5 hash function on the contents of the component file. (Ex. 1003, 5.) The unique identifier for the entire package is calculated by applying *an MD5 hash again to the concatenation of the MD5 hashes* of the component files, in numeric order ("a hash of hashes"). (*Id*.)

Whether Langer Describes "Obtaining a Plurality of Segment Identifiers"

PersonalWeb argues that Langer fails to disclose the claim limitation "in response to a request, said request comprising a first identifier, obtaining a plurality of segment identifiers" as recited in claim 10, and similarly recited in claims 21 and 34. (Prel. Resp. 13, 17-19.) In particular, PersonalWeb maintains that Langer fails to describe how a package would be requested, what would be included in such a request to access a package, and what would be provided in response to such a request. (*Id*.)

We are not persuaded by PersonalWeb's arguments. As stated previously, we conclude that Dr. Clark is qualified to testify as to the understanding of a person of ordinary skill in the art. Notably, citing to certain portions of Langer, Dr. Clark testifies:

> Langer discloses that *a user computer can access a particular file data item by sending a query to a central database server*, such as Archie or WAIS. (Langer at 3–4.) The *query includes*

> *an MD5 hash for the requested file*. (Langer at 3–4.) For example, Langer discloses that a file server can simply assign filenames ("hard links") equal to the files' MD5 hashes, in order to allow clients to call files up by their MD5 hash. "A simple ftp implementation would just hardlink every file available for ftp to a filename encoding of it's [sic] MD5 token." (Langer at 4.) *In response, the central database server uses the MD5 hash to return the locations, such as FTP sites, that store a copy of the file corresponding to the identifier*. (Langer at 4; Ex. 1003) ("An archie or similar lookup could first determine which nearby systems have the file (though come to think of it, that database lookup may as well also provide the local directory and filename for it).") *In order to retrieve a particular file in an archive, a person of ordinary skill in the art would understand that the MD5 hash of the archive could be used to obtain the MD5 hashes that were computed for each of the files within the archive. The user computer can then access the particular file using the FTP protocol.* (Langer at 4.) ("Users would then ftp the directory path and filename of the MD5 token and obtain the file").

(Ex. 1009, ¶ 29, emphases added.)

On the record before us, we credit Dr. Clark's testimony. In particular, we agree with Dr. Clark that one of ordinary skill in the art would have understood that the MD5 hash of the archive could be used to obtain the MD5 hashes that were computed for each of the files within the archive. Given the disclosure of Langer and Dr. Clark's testimony, we determine that EMC has made a threshold showing that Langer describes the disputed limitation.

Whether Langer Describes a "Data Identifier"

In the context of claim 21, PersonalWeb also argues that Langer fails to disclose a "data identifier." (Prel. Resp. 14-15.) PersonalWeb maintains that the claim term "data identifier" should be construed to mean "an identity for a data item generated by processing *all* of the data in the data item, and *only the data in the data item*, through an algorithm." (Prel. Resp. 15.) Applying that claim construction, PersonalWeb alleges that Langer's identifier for a package is *not* based on *all of the data in the package* as required by the claim. (*Id*.) According to PersonalWeb, Langer does not hash *all of the data in a package* because Langer does not apply a hash to the headers, directory, or directory tree of a package. (Prel. Resp. 16-17.)

We are not persuaded by those arguments as they hinge on PersonalWeb's overly narrow construction of the claim term "data identifier," which we decline to adopt (*see supra*). As stated previously, PersonalWeb's claim construction improperly imports additional limitations from the claims of '791 patent from another proceeding into the challenged claims in the instant proceeding.

In addition, PersonalWeb's claim construction is inconsistent with claim 21 of the '539 patent, which recites "wherein said first *data identifier* is based, at least *in part*, on a second given function of data comprising the plurality of *segment identifiers*." That claim language does not require a "data identifier" to be generated by processing "*all* of the data in the data item, and *only the data in the data item*." Rather, claim 21 requires that the data identifier is based, at least *in part* on "*segment identifiers*."

As we discussed above, Langer's unique identifier of the entire package is calculated by applying an MD5 hash (a second function) to the concatenation of the MD5 hashes (segment identifiers) of the component files in numeric order. (Ex. 1003, 5.) Given the disclosure of Langer and our claim construction of the claim term "data identifier," we determine that, on this record, EMC has made a threshold showing that Langer describes the disputed claim limitation "data identifier" as recited in claim 21.

Claim 34 – Obvious Over Langer and Woodhill

As to claim 34, EMC relies on Woodhill to satisfy the claim limitation "dividing a particular data item into a plurality of segment" for this asserted ground of unpatentability. (Pet. 40-41.) In that regard, EMC relies on Woodhill's procedure that divides large files into smaller segments (*e.g.*, "binary objects," and "binary objects" into "granules") to reduce the amount of data that must be transmitted (*i.e.*, smaller segments instead of entire files are transmitted). (Pet. 41.) Indeed, Woodhill describes that if a data stream is larger than the maximum size (*e.g.*, 1 megabyte), the Distributed Storage Manager ("DSM") program divides the data stream into multiple binary objects (Ex. 1005, 4:19-30), and for large files (*e.g.*, database files), the DSM program divides the large file into "granules" and creates a "shadow file" which contains the content identifiers for the "granules" (Ex. 1005, 14:62-65; 15:9-24). Dr. Clark testifies that a person of ordinary skill in the art would have modified Langer's system to use

Woodhill's techniques for more efficient handling of large data items and improve file access. (Ex. 1009, ¶ 31.)

Upon consideration of the disclosures of Langer and Woodhill, and Dr. Clark's testimony, we determine that EMC has made a threshold showing that: (1) Woodhill describes "dividing a particular data item into a plurality of segment;" and (2) the combination of Langer and Woodhill collectively renders the claimed subject matter in claim 34 obvious.

D. *Claims 10, 21, and 34 – Obvious Over Kantor and Langer*

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor, and that claim 34 is obvious over Kantor in view of Langer. (Pet. 42-49.) In its preliminary response, PersonalWeb does not challenge these asserted grounds of unpatentability.

Kantor discloses a method for processing electronic files on a network of computers. (Ex. 1004, 4-5.) In particular, Kantor discloses the operations of the FWKCS program in a client and host configuration. (Ex. 1004, 1.) Kantor's system includes a set of programs for creating and manipulating contents signatures (unique data identifiers), and performs operations based on the contents signatures. (Ex. 1004, 2-3.)

The contents signature for each file is obtained by applying a hash function (*e.g.*, a cyclic residue check or cyclic redundancy check (CRC)) to each file. (Ex. 1004, 6-8, 48-49.) For each zipfile, Kantor creates "zipfile contents signatures" by *hashing the contents signatures* for the files contained within the zipfile ("a hash of hashes"). (Ex. 1004, 2, 9, and 55.)

As Kantor points out, this is done by "adding together all the 32_bit CRC's for the files in the zipfile, modulo 2^32, separately adding together their uncompressed file_ lengths module 2^32, and then arranging the two resulting hexadecimal number as a single structure." (*Id*. at 9.) Dr. Clark testifies that "additional modulo 2^32" is another well-known simple hashing function that uses addition to calculate a value for a file based on the file's contents. (Ex. 1009, ¶ 35.)

Kantor describes several operations that use file contents signatures and zipfile contents signatures. (Ex. 1004, 9.) For instance, when uploading a zipfile, the system determines whether that zipfile and individual component files of that zipfile already exist in the system by using the zipfile contents signature and the contents signatures for the individual files. (*Id*.) As noted by EMC, those signatures also can be used to find zipfiles or files on the network system, to delete duplicate zipfiles or files uploaded under different names, and to determine if files are contained in a larger zipfile or spread among different zipfiles. (*Id*.)

Although Kantor does not teach a read or download request that identifies a zipfile by its zipfile contents signature expressly, EMC asserts that "a person of ordinary skill in the art would have found it obvious to modify the BBS commands, including the download and/or read commands, to permit identifying files based on contents-signatures or zipfile contents-signatures." (Pet. 46.) Dr. Clark also testifies that in response to such commands, the system would use the file contents signatures for the component files of the zipfile to obtain the files. (Ex. 1009, ¶ 41.) EMC

submits that "this would facilitate integrity checking by more precisely specifying the file of interest by its content, and thus improve accuracy." (Pet. 46.)

As to claim 34, EMC also relies on Langer to teach applying an *MD5 hash* function on the contents of a file to obtain a unique identifier for the file. (Pet. 49-50, Ex. 1003, 4-5.) EMC further asserts that in light of Langer, a person of ordinary skill in the art would have modified Kantor to utilize an *MD5 hash* "to improve the statistical error rate of KWKCS further using a hash function widely accepted in the community." (Pet. 49.)

We have reviewed EMC's analysis and supporting evidence, and we determine that EMC's contentions are persuasive. Therefore, on this record, we conclude that EMC has demonstrated that there is a reasonable likelihood that it would prevail with respect to claims 10 and 21 based on the ground that these claims are unpatentable over Kantor, and with respect to claim 34 based on the ground that claim 34 is unpatentable over Kantor and Langer.

E. Claims 10 and 21 – Obvious Over Woodhill and Fischer

EMC asserts that claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Woodhill and Fischer. (Pet. 50-57.) EMC cites Woodhill to meet all of the claim limitations of claims 10 and 21, except EMC acknowledges that the embodiment of Woodhill for restoring a file does not use a hash of the granule identifiers to identify the database file that contains the granules. (Pet. 50-55.) Nevertheless, EMC indicates that using a "*hash of hashes*" technique for identifying large database files or

compound data files is a well-known technique as evidenced by Fischer (Ex. 1036, 7:49-8:38), and other cited references such as Langer (Ex. 1003, 5) and Kantor (Ex. 1004, 9). (Pet. 56.)

For this asserted ground, EMC relies upon Fischer to teach the limitation "wherein said *first data identifier* is based, at least in part, on a second given function of the plurality of *segment identifiers*" as recited in claim 21, and similarly recited in claim 10. (Pet. 56-57.) Directing our attention to Dr. Clark's testimony, EMC contends that a person of ordinary skill in the art would have modified Woodhill's data file restoring process by calculating the identifier for the large data file based on a function of the granule identifiers ("a hash of hashes"), as taught by Fischer, because this would improve the efficiency and performance of Woodhill's data processing for restoring a file. (Pet. 57, citing to Ex. 1009, ¶ 59.)

PersonalWeb counters that the combination of Woodhill and Fischer does not describe all of the claim limitations, and one of ordinary skill in the art would not have combined Woodhill and Fischer. (Prel. Resp. 43-45.) Upon consideration of the parties' contentions, we are not persuaded by PersonalWeb's arguments. Instead, we determine EMC's analysis and supporting evidence have merit.

Prior Art – Woodhill

Woodhill discloses a system for distributed storage management on a computer network system. (Ex. 1005, 1:11-17.) Figure 1 of Woodhill, reproduced below, depicts a computer network system that includes a

distributed storage management system.

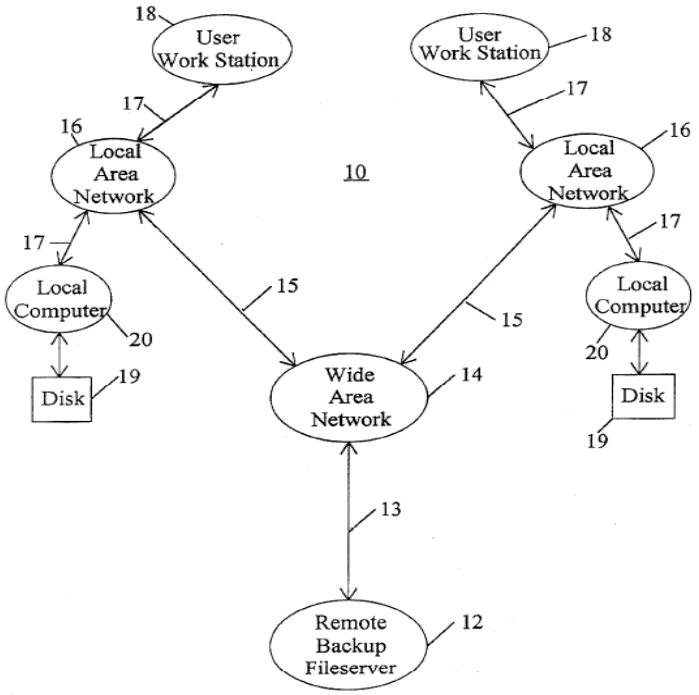

FIG. 1

As illustrated in Figure 1 of Woodhill, each local area network 16 includes multiple user workstations 18 and local computers 20. (Ex. 1005, 3:24-44.) Woodhill's system includes a DSM program for building and maintaining the File Database. (Ex. 1005, 3:45-49.)

For large database files on the network computer system, the DSM program utilizes a technique of subdividing the large database files into "granules" and then tracks changes from the previous backup copy of the "granule" level. (Ex. 1005, 14:53-65.) The "granularization" technique is used to reduce the amount of data that must be transmitted to the remote backup file server. (Ex. 1005, 15:4-8.) If this is the first time that a binary object is being backed up using the "granularization" technique, the DSM

program creates a "shadow file," which contains a contents identifier for each "granule" in the binary object. (Ex. 1005, 15:9-24.) Each contents identifier includes a 32-bit hash number which is calculated against the contents of the "granule." (Ex. 1005, 15:24-30.)

Using an Identifier to Locate and Request a Segment

PersonalWeb contends that Woodhill does not describe the following claim limitations recited in claim 21 (Prel. Resp. 41-43, emphasis added):

> (b0) using said particular segment identifier to *ascertain one or more locations* in said network of computers that should have the corresponding particular segment;
>
> (bl) using said particular segment identifier to *request said corresponding particular segment* from at least one of said one or more locations ascertained in step (b0);

We are not persuaded by PersonalWeb's contention. Woodhill describes using the technique of "granularizing" large files to *restore* a current version of a file *to a previous version* of that file. (Ex. 1005, 17:18-18:9.) This embodiment of Woodhill describes how the DSM program on a network of computer systems restore the previous version of the file by retrieving the "granules" (segments) from the remote server and the local computer. (*Id*.)

Woodhill's DSM program uses the *contents identifiers* of the "granules" within the *previous version* of the file to determine the location of the "granules." (Ex. 1005, 17:50-55; box 450 of Fig. 5I.) Notably, it compares the *contents identifiers* of the "granules" within the *previous version* of the file with the contents identifiers for the "granules" within the

current version of the file. (*Id*.) The comparison of the contents identifiers allows the DSM program to *ascertain whether the "granule" is located on the local computer or on the remote server*. (*Id.*)

The DSM program locates the "granule" on the local computer when the contents identifiers match. (Ex. 1005, 17:58-60.) If the DSM program determines that the contents identifiers do not match, the DSM program locates the "granule" on the remote server and *transmits the "granule" from the remote server to the local computer*. (Ex. 1005, 17:60-64; box 454 of Fig. 5I.) After all of the "granules" that are located on the remote server have been transferred to the local computer, the file on the local computer is restored to its previous version. (Ex. 1005, 18:6-9.)

Given the disclosure of Woodhill, we determine that EMC has made a threshold showing that Woodhill describes the disputed claim limitations.

Whether Woodhill Teaches Away the Proposed Combination

PersonalWeb argues that Woodhill teaches away from the technique employed by Fischer. (Prel. Resp. 43-44.) According to PersonalWeb, Woodhill teaches that Fischer's technique is undesirable because Woodhill emphasizes that "the key notion is that the Binary Object Identifier is calculated from the contents of the data instead of from an external and arbitrary source," whereas Fischer calculates the "fileHash" using external and arbitrary sources. (*Id*. citing to Ex. 1005, 8:40-42; Ex. 1036, 8:4-55.)

PersonalWeb's argument is misplaced. Even if Fischer's technique includes a calculation using an external source, obviousness does not require

that all of the features of the secondary reference be bodily incorporated into the primary reference. *In re Etter*, 756 F.2d 852, 859 (Fed. Cir. 1985) (en banc); *In re Keller* 642 F.2d 413, 425 (CCPA 1981).

In any event, PersonalWeb's argument takes Woodhill's "key notion" statement out of context. That statement of Woodhill is directed to the *contents identifiers*, which are calculated against *the contents of the "granules,"* where a large database file has been divided into a plurality of "granules" (Ex. 1005, 15:24-30; Fig. 5A, step 138), or the contents identifiers of files in a non-granularization situation (Ex. 1005, 8:22-32). EMC's proposed modification does not change Woodhill's process for calculating the *contents identifiers of the "granules"* and files in non-granularization situations. (Pet. 55-57.)

By applying "a hash of hashes" technique in light of Fischer to Woodhill's restoring process as suggested by EMC, Woodhill's *identifier for the database file* itself would be calculated using a hash function (second function) against the *contents identifiers* of the "granules." (*Id*.) Therefore, Woodhill's "key notion" statement does not teach away from that proposed modification.

Whether the Proposed Modification Renders Woodhill Inoperative

PersonalWeb asserts that the alleged combination of replacing Woodhill's Binary Object Identifier with Fischer's "fileHash" technique would have resulted in an inoperative system. (Prel. Resp. 44-45.) Specifically, PersonalWeb argues that many parts of Woodhill's system rely

on Binary Object Identifiers to detect changes in binary objects that are not granularized, and thus one of ordinary skill in the art would not have modified Woodhill's Binary Object Identifiers to base them on granule identifiers because this would have resulted in much of Woodhill's system being inoperative. (*Id*.)

That argument is inapposite, as it implies that EMC's proposed combination would apply a "hash of hashes" technique to non-granularization situations, but that is not what EMC proposes. Instead, EMC's proposed modification is limited to Woodhill's embodiment for restoring a file that has been divided into a plurality of "*granules*." (Pet. 54-56, citing to Ex. 2005, 17:7-50, 17:60-18:4.) More precisely, the identifier for the file would be calculated based on a function of *the content identifiers of the "granules."* (Pet. 56.) Further, PersonalWeb's general allegation of inoperability lacks sufficient specificity. PersonalWeb fails to provide an adequate explanation or credible evidence to show how a file identifier calculated using a hash of content identifiers would prevent Woodhill's system from detecting changes in the file.

As to PersonalWeb's argument that Woodhill is concerned with uniquely identifying a binary object, whereas Fischer is concerned with security, we do not find that persuasive. The fact that the two references have different objectives does not mean that a person of ordinary skill in the art would not combine their teachings. *In re Heck*, 699 F.2d 1331, 1333 (Fed. Cir. 1983) ("The use of patents as references is not limited to what the patentees describe as their own inventions or to the problems with which

they are concerned.") (quoting *In re Lemelson*, 397 F.2d 1006, 1009 (CCPA 1968)). *See also KSR Int'l Co. v. Teleflex Inc*., 550 U.S. 398, 417 (2007) (("[W]hen a patent 'simply arranges old elements with each performing the same function it had been known to perform' and yields no more than one would expect from such an arrangement, the combination is obvious.") (Citing *Sakraida v. AG Pro, Inc.*, 425 U.S. 273, 282 (1976)).

Accordingly, we are persuaded by EMC's analysis and supporting evidence. On this record, we conclude EMC has demonstrated that there is a reasonable likelihood that it would prevail with respect to claims 10 and 21 on the ground that these claims are obvious over Woodhill and Fischer.

F. Other Asserted Grounds

EMC also asserted that claims 10, 21, and 34 are unpatentable under 35 U.S.C. § 102(a) as anticipated by Browne; claim 34 is unpatentable under 35 U.S.C. § 102(b) as anticipated by Langer; and claim 34 is unpatentable under 35 U.S.C. § 103(a) over Kantor, Kantor and Browne, Browne in view of Woodhill, or Woodhill in view of Browne. However, these asserted grounds of unpatentability are unnecessary in light of the determination that there is a reasonable likelihood that the challenged claims are unpatentable based on the ground on which we institute an *inter partes* review. We therefore exercise our discretion to deny these grounds as redundant. *See* 37 C.F.R. § 42.208.

III. CONCLUSION

For the forgoing reasons, we determine that the information presented in EMC's petition shows that there is a reasonable likelihood that EMC would prevail with respect to its assertion that claims 10, 21, and 34 of the '539 patent are unpatentable. Accordingly, the petition is granted.

IV. ORDER

It is

ORDERED that pursuant to 35 U.S.C. § 314, an *inter partes* review is hereby instituted as to claims 10, 21, and 34 of the '539 patent for the following grounds:

1. Claims 10 and 21 are unpatentable under 35 U.S.C. § 102(b) as anticipated by Langer;
2. Claim 34 is unpatentable under 35 U.S.C. § 103(a) over Langer and Woodhill;
3. Claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor;
4. Claim 34 is unpatentable under 35 U.S.C. § 103(a) as obvious over Kantor and Langer; and
5. Claims 10 and 21 are unpatentable under 35 U.S.C. § 103(a) as obvious over Woodhill and Fischer;

FURTHER ORDERED that pursuant to 35 U.S.C. § 314(d) and 37 C.F.R. § 42.4, notice is hereby given of the institution of a trial; the trial is commencing on the entry date of this decision; and

FURTHER ORDERED that an initial conference call with the Board is scheduled for 2:00 PM Eastern Time on June 3, 2013; the parties are directed to the Office Trial Practice Guide, 77 *Fed. Reg.* at 48765-66, for guidance in preparing for the initial conference call, and should come prepared to discuss any proposed changes to the Scheduling Order entered herewith and any motions the parties anticipate filing during the trial.

PETITIONER:

Peter M. Dichiara, Esq.
David L. Cavanaugh, Esq.
WILMER CUTLER PICKERING HALE & DORR LLP
peter.dichiara@wilmerhale.com
david.cavanaugh@wilmerhale.com

PATENT OWNER:

Joseph A. Rhoa, Esq.
Updeep. S. Gill, Esq.
NIXON & VANDERHYE P.C.
jar@nixonvan.com
usg@nixonvan.com