UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSETTS

CABLE & WIRELESS INTERNET SERVICES, INC. (formerly known as DIGITAL ISLAND, INC.),	§	Civil Action No. 02-11430 RWZ
	§	
	§	Judge Rya W. Zobel
and	§	JURY TRIAL REQUESTED
KINETECH, INC.,	§	
Plaintiffs,	§	
vs.	§	
AKAMAI TECHNOLOGIES, INC.,	§	
Defendant.	§	

CWIS' OPENING MARKMAN BRIEF CONSTRUING THE TERMS AT ISSUE IN U.S. PATENT NO. 6,415,280

Pursuant to the Order of this Court, plaintiff Cable & Wireless Internet Services, Inc. ("CWIS") submits this claim construction brief, construing the seven terms at issue in U.S. Patent No. 6,415,280 ("the '280 patent"). The terms are *data files, data identifier, given function, comprises, cached versions of data files, hash,* and *value.*

THE TECHNOLOGY OF THE CLAIMS AT ISSUE

The claims of the '280 patent at issue in this case are directed at various methods for using a *data identifier* to ensure that a fresh, rather than stale, data file is served by the cache server of a Content Delivery Network ("CDN").[1]

[1] The claims at issue in this case are claims 9, 18, 19, 20, 21, 23, 24, 25, 35, 36, 37, 38 and 39 of the '280 patent. The claims generally refer to the method described therein as a "content delivery method" that involves "cached" data files. *See Declaration of Tim Walker* ("Walker Decl."), Tab A: '280 patent.

EXHIBIT
tabbies
2004

The '280 patent is a continuation of U.S. Patent No. 5,978,791 ("'791 patent"), which was the subject of a previous trial before Your Honor. The claims at issue in the '791 trial were directed broadly to identifying a data item throughout an entire data processing system using an "identifier depending on and being determined using all of the data in the data item and only the data in the data item." *Declaration of Timothy Walker* ("Walker Decl."), Tab B; '791 patent, claims 30, 31, 33 and 41.[2] The claims at issue in this case are different. They are directed to serving fresh data files from a CDN cache server, not identifying data items throughout an entire data processing system.[3] The claims in this case describe using "*data identifiers*" to ensure that the cached data files are fresh, where the "*data identifiers*" are determined using at least the contents of the data file.

In everyday terms, suppose that the *Washington Post* website is maintained on an origin server in Boston, you are at a computer terminal in Cambridge, there is a late-breaking news story in Washington, D.C., and you want to see the front-page, lead picture from the *Post's* website. You send a request from your computer in Cambridge to get the *Post's* front-page, lead picture. If the data file that includes that lead picture is on a cache server in Boston, your computer's browser can retrieve the data more quickly from the nearby cache server, rather than having to send the request all the way to the *Post's* origin server in Northern Virginia. This quicker retrieval, however, is dependent upon the cache server in Boston having a data file that contains the current front-page lead picture, and not the lead picture from 4 hours ago. The '280

[2] The claims at issue in the '791 trial were claims 30, 31, 33 and 41, and each claim required an "identifier depending on and being determined using all of the data in the data item and only the data in the data item."

[3] The patentee, in response to a request from the Examiner to explain the claimed invention, described the '280 patent as involving a "content delivery method" that uses "cached versions" of data files. *Walker Decl.*, Tab C: Amendment, 8/22/01, at p. 46.

patent is directed to using *data identifiers* to ensure that the data file served by the cache server is fresh, and contains the current picture that exists on the *Post's* origin server.

In a CDN system, data is typically stored in data files on the origin server, and a copy of the data files may be stored on a cache server. *Declaration of Robert Dewar* ("Dewar Decl."), ¶ 20. Ordinarily, a data file is identified using a Uniform Resource Locator ("URL") that typically includes a pathname. For example, the *Washington Post* front-page lead picture may be identified as "http://media.washingtonpost.com/wp-srv/photo/homepage/hp7-25-03b.jpg." The '280 patent explains that when data is simply identified by a URL, any change to the data on the origin server may not be reflected on the CDN cache server. To ensure that the cache data is fresh, the CDN system typically uses a costly "synchronization" process. The '280 patent describes this "synchronization" process:

> Before using a cached item, a cache client must either reload the cached item, be informed of changes to the cached item, or confirm that the master item corresponding to the cached item has not changed. In other words, a cache client must synchronize its data items with those on the cache server. This synchronization may involve reloading data items onto the cache client. The need to keep the cache synchronized or reload it adds significant overhead to existing caching mechanisms.

Walker Decl., Tab A: 2:63-3:4.[4] The claims at issue in the '280 patent use a *data identifier* to avoid this costly "synchronization" process.

The '280 patent explains that when a *data identifier* "is being used to cache data items, the problems of maintaining cache consistency are avoided." 37:24-26; 3:58-60.[5] The

---

[4] Throughout this brief, the '280 patent is referred to by column and line numbers. The '280 patent is attached as Exhibit A to the Walker Declaration.

[5] For example, the '280 patent explains the *data identifier*, in the context of a key, as follows:

> To access a cache and to fill it from its server, a key is required to identify the data item desired. Ordinarily, the key is a name or address [in this case, it would be the pathname of a file]. If the data associated with such a key is changed, the client's cache

consistency problems are avoided because the *data identifier* is determined using at least the contents of the data file. As a result, when the underlying data changes (*i.e.*, the *Post's* lead picture changes) a new data file is created on the origin server and a new *data identifier* is determined using at least the contents of that new data file. According to the '280 patent, the *data identifier* is created using a "given function of the data," and the "data used by the given function comprises the contents of the particular data file." 41:11-14, 41:49-52, 42:10-13.

The '280 patent further explains that the *data identifier* is a "substantially unique" identifier for a data file. 6:8-9. The '280 patent does not require absolute uniqueness of the *data identifier*, but only "sufficient uniqueness" for the application. 13:49-50. The '280 patent recognizes that, in the context of a CDN cache, the actual number of cached data files "form a very sparse subset of all possible inputs." 13:7-10. Accordingly, the likelihood of a *data identifier* collision (*i.e.*, the same *data identifier* for different data files) in the context of a CDN is extremely remote. 13:19-29. Given this small subset of cached data files, the '280 patent recognizes that "lower probabilities of uniqueness may be acceptable." 13:31-33.

The '280 patent also contemplates combining the *data identifier* with other information to provide an additional level of uniqueness, if necessary. According to the '280 patent, the *data identifier* may "use tagged, typed, categorized or classified data items and use a combination of both the [data identifier] and the tag, type, category or class of the data item as an identifier."

---

becomes inconsistent; when the cache client refers to that name, it will retrieve the wrong data...

By using an embodiment of the present invention, the cache key uniquely identifies the data it represents. When the data associated with a name changes, the key itself changes. Thus, when a cache client wishes to access the modified data associated with a given file name, it will use a new key [the True Name of the new file] rather than the key to the old file contents in its cache.

37:27-44.

13:41-44. The patent explains that these "tags provide an additional level of uniqueness." 13:50-51.

THE TERMS AT ISSUE

The seven terms at issue in this case are: (1) *data files*; (2) *data identifier*; (3) *given function*; (4) *comprises*; (5) *cached versions*; (6) *hash*; and (7) *value*. The terms appear repeatedly in the 13 separate claims that CWIS asserts, and Claims 9 and 35 are illustrative. The text of Claim 9 provides as follows (with the terms at issue in italicized red):

> 9. In a system in which a set of *data files* are distributed across a network of servers, some of the *data files* being cached from a source server distinct from the servers in the network, a content delivery method comprising:
>
> determining a *data identifier* for a particular *data file* on the source server, the *data identifier* being determined using a *given function* of the data, wherein said data used by the *given function* to determine the *data identifier comprises* the contents of the particular *data file*; and
>
> responsive to a request for the particular *data file*, the request including at least the *data identifier* of the particular *data file*, causing a copy of the particular *data file* to be provided from a given one of the servers of the network of servers.

The text of Claim 35 provides as follows (with the terms at issue in italicized red):

> 35. A content delivery method, comprising:
>
> distributing a set of *data files* across a network of servers, at least some of the *data files* being *cached versions* of *data files* from another server, said other server being distinct from the network of servers;
>
> determining a *data identifier* for a particular *data file*, the *data identifier* including a *hash* of the contents of the particular *data file*; and
>
> in response to a request for the particular *data file*, the request including at least the *data identifier* of the particular *data file*, providing the particular *data file* from a given one of the servers of the network of servers.

CWIS, consistent with controlling Federal Circuit law, explains below that each of the terms at issue in this case is defined according to its plain meaning, consistent with the intrinsic record of the patent specification, claims and file history, the dictionary definition and the understanding of one skilled in the art.

ARGUMENT

I. CLAIM TERMS ARE CONSTRUED CONSISTENT WITH THEIR PLAIN MEANING, THE INTRINSIC EVIDENCE AND UNDERSTANDING OF ONE SKILLED IN THE ART.

Claim construction begins with the words of the claim. *Vitronics Corp. v. Conceptronic, Inc.*, 90 F.3d 1576, 1582 (Fed. Cir. 1996).[6] "The analytical focus must begin and remain centered on the language of the claims themselves, for it is that language that the patentee chose to use to particularly point out and distinctly claim the subject matter which the patentee regards as his invention." *Texas Digital Systems Inc. v. Telegenix, Inc.*, 308 F.3d 1193, 1201-02 (Fed. Cir. 2002).

When construing the claims, courts begin with an examination of the intrinsic evidence, *i.e.*, the claims, the other portions of the specification, and the prosecution history. *Altiris, Inc. v. Symantec Corp.*, 318 F.3d 1363, 1369 (Fed. Cir. 2003). Additionally, dictionary definitions may be consulted in establishing a claim term's ordinary meaning. *Id.* A patentee may choose to be his own lexicographer and use terms in a manner other than their ordinary meaning, as long as the special definition of the term is clearly stated in the patent specification. *Inverness Medical Switzerland GmbH v. Princeton Biomeditech Corp.*, 309 F.3d 1365, 1371-72 (Fed. Cir. 2002), *citing Vitronics Corp.*, 90 F.3d at 1582.

[6] The Federal Circuit, in *Intellectual Property Development, Inc. v. UA-Columbia Cablevision of Westchester, Inc.*, 2003 WL 21688043 at 4 (Fed. Cir. July 21, 2003), recently confirmed the claim construction analysis articulated in *Vitronics*.

The terms used in the claims bear a "heavy presumption" that they mean what they say and have the ordinary meaning that would be attributed to those words by persons skilled in the relevant art. *Texas Digital*, 308 F.3d at 1202. Courts may also review extrinsic evidence, always to assist them in comprehending the technology in accordance with the understanding of skilled artisans and as necessary for actual claim construction. *Id.* Extrinsic evidence may never be relied upon, however, to vary or contradict the clear meaning of terms in the claims. *Id.*

In accord with these legal principles, CWIS provides its construction of the seven claim terms at issue in this case.

II. CONSTRUCTION OF THE DISPUTED TERMS.

A. Data File.

The term *data file* means "a named sequence of bits."[7]

The '280 patent refers to a *data file* as being a "named data item" (5:45-50) and refers to a "data item" as a "sequence of bits" (1:53-54). Accordingly, a *data file*, as that term is used in the claims at issue, means a named sequence of bits.

The '280 patent provides further context. A data item is "the contents of a file." 1:54. Specific examples include "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." 1:54-58. Supplying the data item with a name creates a data file. According to the '280 patent: "[A] file is a named data item which is either a data file (which may be simple or compound) or a directory file." 5:47-50. The term "name," as used in the '280 patent, simply means an alphanumeric identifier for data that typically uses a location or address to identify the data. 1:22-27.

[7] The term *data file* or *data files* appears in claims 9, 18, 19, 20, 21, 23, 24, 25, and 35 of the '280 patent.

The dictionary definition of the term "file," as used in the context of computers and the Internet, is consistent with this construction. The *Random House Webster's Computer & Internet Dictionary* (3d ed. 1999) defines "file" as "a collection of data or information that has a name." *Walker Decl.*, Tab D: p. 211. Similarly, the *Microsoft Computer Dictionary* (4th ed. 1999) defines "file" as "a complete, named set of information, such as a program, a set of data used by a program, or a user-created document." *Id.* at Tab E: p. 183.

This construction of the term *data file* is also consistent with the understanding of one skilled in the art. *Dewar Decl.*, ¶¶ 9-10.

B. Data Identifier.

The term *data identifier* means "a substantially unique identifier for a data file that is determined using at least the contents of the data file."[8]

The term *data identifier*, as used in the '280 patent, refers to "the substantially unique data identifier for a particular data item." 6:7-9. The '280 patent contains a lengthy description about what it means to be "substantially unique." The patent explains that it is impossible to have a truly "unique" identifier because, in some circumstances, the number of possible data files may be larger than the number of possible identifiers, creating the possibility of a "colliding set" of data files.

Indeed, the '280 patent recognizes that absolute uniqueness cannot be achieved so long as there are more theoretical data files than *data identifiers*. If the contents of a data file can be up to 1,000 bits in length, but the *data identifier* can only be 100 bits in length, a mathematical possibility exists that the same *data identifier* may identify different data files. The '280 patent recognizes that it is theoretically "impossible to define a function having a unique output for

[8] The term *data identifier* appears in claims 9, 18, 19, 23, 24, 25, and 35 of the '280 patent.

each possible input when the number of elements in the range of the function is smaller than the number of elements in its domain." 13:4-7.

For this reason, the '280 patent requires only that the *data identifier* be "substantially unique." The '280 patent explains that substantial uniqueness is sufficient because "the actual data items that will be encountered in the operation of any system embodying this invention form a very sparse subset of all the possible inputs." 13:7-10. The '280 patent further explains "lower probabilities of uniqueness may be acceptable, depending on the types of applications and mechanisms used" (13:31-33), and that in some embodiments it is acceptable to have "different degrees of uniqueness" (13:33-35). The patent simply requires "sufficient uniqueness" for the application. 13:49-50.

In addition, the '280 patent requires that the *data identifier* be determined using at least the contents of the data file: "the data identifier determined using a given function of the data, where said data used by the given function comprises *the contents of the particular data file.*" 41:11-14 (emphasis supplied).[9] The '280 invention is directed to using a *data identifier* that, in a particular application, changes when the contents of a data file change. A *data identifier* that is determined using at least the contents of the data file has this quality. Whenever the contents of the data file change, a new *data identifier* is created using the contents of the new, changed data file. As described in the '280 patent, this ensures that "problems of maintaining cache consistency are avoided." 3:60-61, 37:25-26.

It is important to note that the *data identifier* at issue here is not the same as the "substantially unique identifier" at issue in the '791 trial. The claims at issue in the '791 trial all required "determining a substantially unique identifier for the data item, the identifier depending

[9] Similarly, the Abstract of the '280 patent explains that the *data identifier* is "determined using a given function of the data comprising the particular data file."

on and being determined using all of the data in the data item and only the data in the data item." *Walker Decl.*, Tab B: '791 patent, claims 30, 31, 33 and 41. Thus, the claims at issue in the '791 case required that the "substantially unique identifier" depend on and be determined using *only* the data in the data item. The claims at issue in this case have no such requirement.

Instead, the only requirement of the claims at issue in this case is that the data used to determine the *data identifier* include the contents of the data file: "the *data identifier* determined using a given function of the data, wherein said data used by the given function comprises the contents of the particular data file." *Walker Decl.*, Tab A: '280 Patent, at 41:11-14; 41:49-52; 42:10-13. The *data identifier* in this case is not limited to *only* the contents of the data file.

Indeed, the '280 patent explains that the *data identifier* may also be "a combination of both the True Name and the tag, type, category or class of the data item." *Id.* at 13:42-44.[10] The patent explains that this tag provides additional uniqueness:[11]

> (O)ther preferred embodiments use tagged, typed, categorized or classified data items and use a combination of both the True Name and the tag, type, category or class of the data item as an identifier. ...In such a system, a lower degree of True Name uniqueness is acceptable over the entire universe of data items, as long as sufficient uniqueness is provided per category of data items.

Id. at 13:39-50. The description of an embodiment using a *data identifier* with a tag to provide additional "uniqueness" confirms that for purposes of the '280 patent, there is no requirement that the *data identifier* be determined using only the contents of the data file.

This construction of the term *data identifier*, as a substantially unique identifier that is determined using at least the contents of the data file, is consistent with the understanding of one skilled in the art: *Dewar Decl.*, ¶ 11-13.

---

[10] The '280 patent uses the terms True Name and *data identifier* interchangeably. *Id.* at 6:7-9.

[11] Akamai's expert, Eugene Spafford, recognizes that that language of the '791 patent specification (which is identical to the '280 patent specification) loosely defines "identifier" to include other information that could be added to the True Name. *Walker Decl.*, Tab C: *Spafford Declaration*, 8/6/01, ¶48.

C. Given Function.

The term *given function* means "a specified algorithm whose input comprises the contents of a data file."[12]

The '280 patent equates a *function* with an algorithm: "These functions (or algorithms)...." 12:64. By *given*, the '280 patent simply means that the function specified must be used consistently: "the same function must be employed on a system-wide basis." 13:1-3. Accordingly, a *given function* is a specified algorithm.

The dictionary definition is in accord. The *Microsoft Computer Dictionary* (4th ed. 1999) defines "function" as "the action carried out by a program or routine." *Walker Decl.*, Tab E: p. 199.[13] The *Webster's II New College Dictionary* (1995) defines the term "given" as "specified." *Id.* at Tab H: p. 473.

The claims at issue in this case add that the input for this specified algorithm is comprised of the contents of a data file. According to the claims, "data used by the given function comprises the contents of the particular data file." 41:13-14, 41:51-52, 41:66-67, 42:12-13. This is consistent with the patentee's own explanation of the invention, that the input to the function be comprised of the data item, which is the contents of the data file: "the data identifier being determined using a given function of the data comprising the particular data item." *Walker Decl.*, Tab C: Amendment, 8/22/01, at p. 46.

This proposed construction, that the *given function* is a specified algorithm whose input comprises the contents of the data file, is also consistent with the understanding of one skilled in the art. *Dewar Decl.*, ¶¶ 14-16.

---

[12] The term *given function* appears in claims 9, 18, 19, 20, 21, 23, 24, 25, 38 and 39.

D. Comprises.

The term *comprises* means "includes, but not limited to."[14]

The term *comprises* is a term of art used in claim language that means that the named elements are essential, but that other elements may be added and still form a construct within the scope of the claim. *Amgen, Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1344-45 (Fed. Cir. 2003). The Federal Circuit recently addressed the term "comprise" in *Amgen*, and confirmed that it means including, but not limited to. The court explained that "a claim reciting 'a widget comprising A and B,' for example, would be infringed by any widget containing A and B, no matter that C, D, or E might be present." *Id.* at 1345. Similarly, the Manual of Patent Examination Procedure § 2111.03 (8th ed. 2001) explains that the term "comprising" is synonymous with "including," and that use of the term "comprising" in a patent claim does not exclude additional elements or method steps.

As used in the claims at issue in this case, the given function *comprises* the contents of the data file, meaning the function includes, but is not necessarily limited to, the contents of the data file.

E. Cached Versions of Data Files.

The phrase *cached versions of data files* simply means "data files on a cache server."[15]

As explained above, in a CDN system, data files reside on the origin server and copies of those date files may reside on the cache server. The '280 patent describes a cache server as a server where copies of data files are stored. 2:62. CWIS' expert, Robert Dewar, further explains

---

[13] The *Microsoft Computer Dictionary* (4th ed. 1999) defines "algorithm" as "a finite sequence of steps for solving a logical or mathematical problem or performing a task." *Walker Decl.*, Tab E; p. 19.

[14] The term *comprises* appears in claims 9, 18, 19, 23, 24 and 25.

[15] The term *cached versions* appears in claims 19, 23, 24 and 35.

the way in which cache servers work, including how the cache servers store, retrieve and serve data files. *Dewar Decl.*, ¶ 20.

The phrase, *cached versions of data files*, is a descriptive phrase that is given its plain, ordinary meaning -- the data files on a cache server.

This construction is consistent with the understanding of one skilled in the art. *Dewar Decl.* at ¶¶ 19-20.

F. Hash.

The term *hash* means "a number generated from input data that is substantially smaller than the data itself."

A hash is the name given to the value computed by a "hash function." According to the *Random House Computer and Internet Dictionary* (3d ed.1999) "a hash value (or simply hash) is a number generated from a string of text. The hash is substantially smaller than the text itself and is generated in such a way that it is extremely unlikely that some other text will produce the same hash value." *Walker Decl.*, Tab D: p. 250. A formal definition of a hash function may be found in the Telecom Glossary 2000 (American National Standard T1.523-2001): "A mathematical function that maps values from a large (or very large) domain into a smaller range, and that reduces a potentially long message into a 'message digest.'" *Id.* at Tab I. The term "domain" refers to the possible inputs, which in this case would be the contents of the data file in question. *Dewar Decl.*, ¶ 22. The "range" is the result of the function, which, in the context of the claims, is a small string that is combined with other data to form the data identifier. *Id.*

The term *hash* appears only in three of the claims at issue (35, 36 and 37) and in all three claims the term is used consistent with its plain meaning in computer science. The '280 patent adds, in general, that a data block of arbitrary length is reduced to "a relatively small, fixed size

identifier" that is "virtually guaranteed to represent the data block." *Walker Decl.*, Tab A: 12:39-42.

This plain, ordinary computer science meaning of the term *hash* is consistent with the understanding of one skilled in the art. *Dewar Decl.*, ¶¶ 21-22.

G. Value.

The term *value*, as used in claims 38 and 39, simply means "the output of a given function."

Both claims 38 and 39 refer to "the *value* determined by the given function." The term *value* in this context has a plain, ordinary meaning as the result produced by the given function, which is properly referred to as the output of the given function.

The dictionary definition of the term *value*, when used in this mathematical sense, is consistent. *Webster's II New College Dictionary* (1995) defines the term *value*, in the context of mathematics, as a "calculated numerical quantity." *Walker Decl.*, Tab H: p. 1219.

This plain, ordinary meaning is of the term *value* is also consistent with the understanding of one skilled in the art. *Dewar Decl.*, ¶¶ 23-24.

CONCLUSION

For the reasons stated above, Cable & Wireless Internet Services, Inc. respectfully requests that the Court construe the terms at issue in this case consistent with the definitions set forth above.

Respectfully submitted,

CABLE & WIRELESS INTERNET SERVICES, INC.

By its Attorneys,

Daniel P. Tighe (BBO 556583)
Scott McConchie (BBO 634127)
Griesinger, Tighe & Maffei, LLP
176 Federal St.
Boston, Massachusetts 02110
(617) 542-9900

Michael J. Bettinger (*pro hac vice*)
Timothy P. Walker (*pro hac vice*)
Preston Gates Ellis LLP
55 Second St., Suite 1700
San Francisco, California 94105
(415) 882-8002

Dated: July 25, 2003

I HEREBY CERTIFY THAT A TRUE COPY OF THE ABOVE DOCUMENT WAS SERVED UPON THE ATTORNEY OF RECORD FOR EACH OTHER PARTY BY MAIL/HAND ON:
7/25/03 SMcC