



US 20030182439A1

(19) **United States**

(12) **Patent Application Publication**

Geshwind

(10) **Pub. No.: US 2003/0182439 A1**

(43) **Pub. Date: Sep. 25, 2003**

(54) **CLASS OF METHODS FOR IMPROVING PERCEIVED EFFICIENCY OF END-USER INTERACTIVE ACCESS OF A LARGE DATABASE SUCH AS THE WORLD-WIDE WEB VIA A COMMUNICATION NETWORK SUCH AS "THE INTERNET"**

(76) Inventor: **David Michael Geshwind, New York, NY (US)**

Correspondence Address:  
**DAVID M. GESHWIND  
111 FOURTH AVENUE  
NEW YORK, NY 10003 (US)**

(21) Appl. No.: **10/341,758**

(22) Filed: **Jan. 14, 2003**

**Related U.S. Application Data**

(60) Division of application No. 08/483,205, filed on Jun. 7, 1995, now Pat. No. 6,507,872. Continuation-in-part of application No. 08/110,230, filed on Aug. 23, 1993, now abandoned, which is a continuation-in-part of application No. 07/951,267, filed on Sep. 25, 1992, now Pat. No. 6,590,573.

**Publication Classification**

(51) **Int. Cl. 7 .....** G06F 15/16  
(52) **U.S. Cl. .....** 709/236; 709/230

(57) **ABSTRACT**

Procedures and systems for speeding up end-user access of large interlinked databases over communication networks such as the internet. Predicting user requests is accomplished by artificially intelligent assistant. These requests are prioritized and augmented protocols are used to download in a weighted interleaved fashion multiple information resources.

(1200) **SESSION IS OVER SESSION HISTORY AND DOCUMENTS RECALLED**

(1201) **SELECTION OF OPTIMIZATION CRITERIA IS SELECTED**

(1202) **SENSITIVITY ANALYSIS RUN ON VARIOUS PARAMETERS AND POLICIES AND REPLAY OF SIMULATED SESSION OF RUN FOR EACH SET TO BE CONSIDERED**

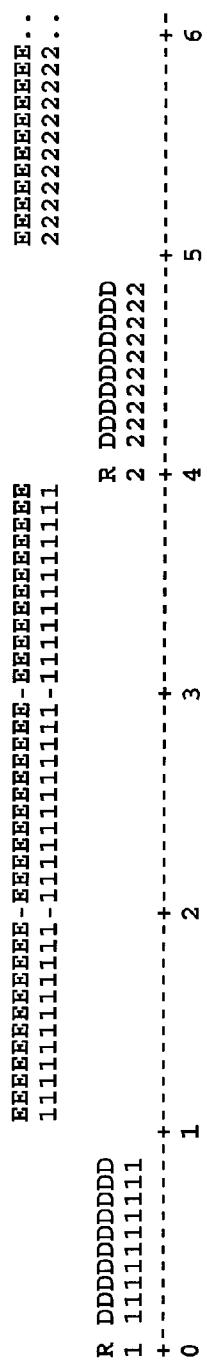
(1203) **RESULTS OF SIMULATION ARE EVALUATED IN LIGHT OF OPTIMALITY CRITERIA SPECIFIED IN (1201)**

(1204) **PARAMETERS ARE NOTED AND STORED FOR RESULTS OF THE BEST (AND WORST FOR AVOIDANCE KNOWLEDGE) FEW SIMULATIONS THESE ARE STORED IN ASSOCIATION WITH A CHARACTERIZATION OF THE SESSION IN SOME APPROPRIATE FEATURE SPACE**

(1205) **AT SUBSEQUENT SESSION, THE SESSION IS CHARACTERIZED AND THE CLOSEST MATCH (OR COMBINATION OF MATCHES) IS FOUND AND THE SET OF "OPTIMAL" PARAMETERS, POLICIES, ETC. STORED IN ASSOCIATION WITH THAT MATCH IS USED NOW**

(1206) **DURING CURRENT SESSION THE CHARACTERIZATION OF THE SESSION IS REEVALUATED AND THE FEATURE SPACE SEARCHED FOR A NEW MATCH, UPDATE PARAMETERS, POLICIES, ETC.**

POSTMORTEM FINE-TUNING PROCESS



**FIGURE 1:** END USER COMMUNICATION CHANNEL USAGE UNDER A FIRST CONDITION

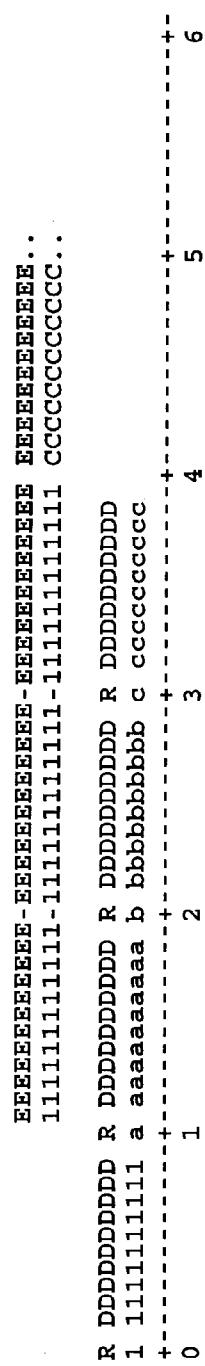
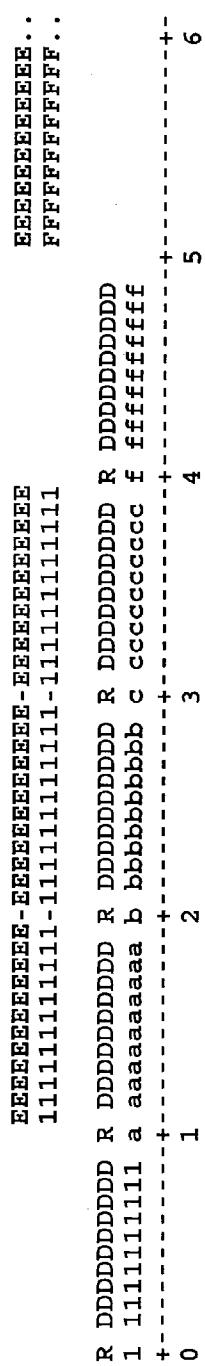


FIGURE 2: END USER COMMUNICATION CHANNEL USAGE UNDER A SECOND CONDITION



**FIGURE 3: END USER COMMUNICATION CHANNEL USAGE UNDER A THIRD CONDITION**

```

EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE.. .
1111111111-1111111111-1111111111-1111111111 cccccccc.. .

R DDDDDDDDD R DDDDDDDDDDDDDDDDDDDDr DDDDDDDDDDDDDDDDDDD
1 1111111111 a-f abcdefabcdefabcdefabcdefabcdefabcdef
+ +-----+ +-----+ +-----+ +-----+ +-----+
0 1 2 3 4 5 6

```

FIGURE 4: END USER COMMUNICATION CHANNEL USAGE UNDER A FOURTH CONDITION

```

EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE.. .
1111111111-1111111111-1111111111-1111111111 cccccccc.. .

R DDDDDDDDD R DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
1 1111111111 a-f abcdefabcdefabcdefabcdefabcdefabcdefabcdef
+ +-----+ +-----+ +-----+ +-----+ +-----+
0 1 2 3 4 5 6

```

FIGURE 5: END USER COMMUNICATION CHANNEL USAGE UNDER A FIFTH CONDITION

```

EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE-EEEEEEEEEE.. .
1111111111-1111111111-1111111111-1111111111 ffffffffff.. .

R DDDDDDDDD R DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD
1 1111111111 a-f abcdefabcdefabcdefabcdefabcdefabcdefabcde
+ +-----+ +-----+ +-----+ +-----+ +-----+
0 1 2 3 4 5 6

```

FIGURE 6: END USER COMMUNICATION CHANNEL USAGE UNDER A SIXTH CONDITION

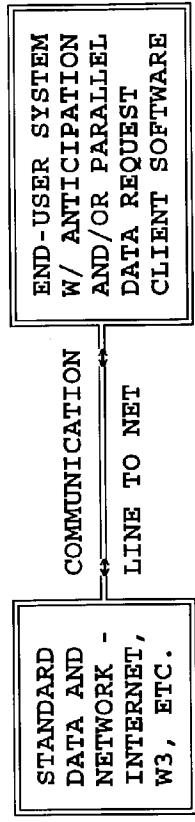


FIGURE 7: ANTICIPATORY AND/OR PARALLEL REQUEST CLIENT EMBODIMENT

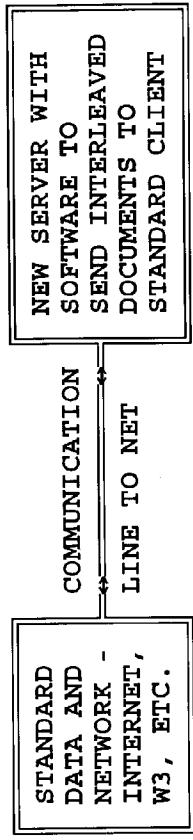


FIGURE 8: INTERLEAVED DOCUMENT SERVER EMBODIMENT

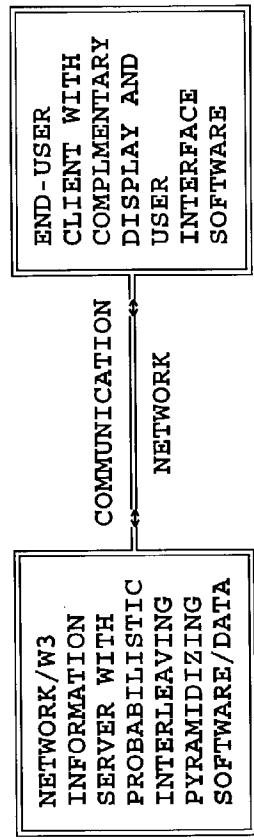


FIGURE 9: ENHANCED CLIENT/SERVER EMBODIMENT

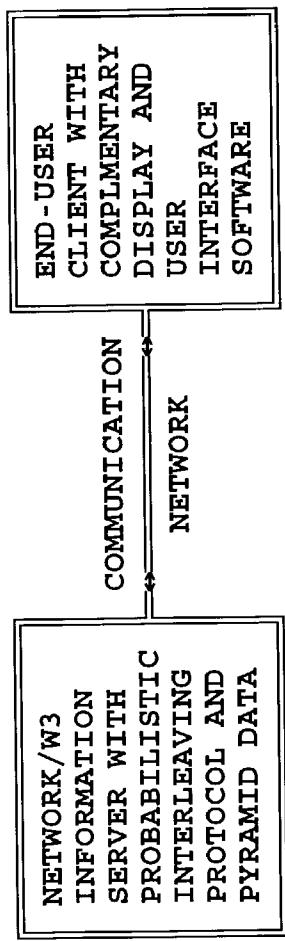


FIGURE 10: ENHANCED PROTOCOL EMBODIMENT

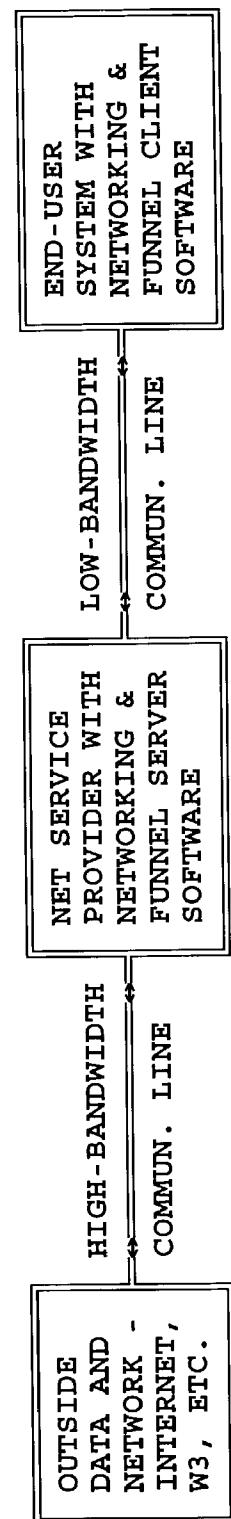


FIGURE 11: "FUNNEL" SERVICE EMBODIMENT

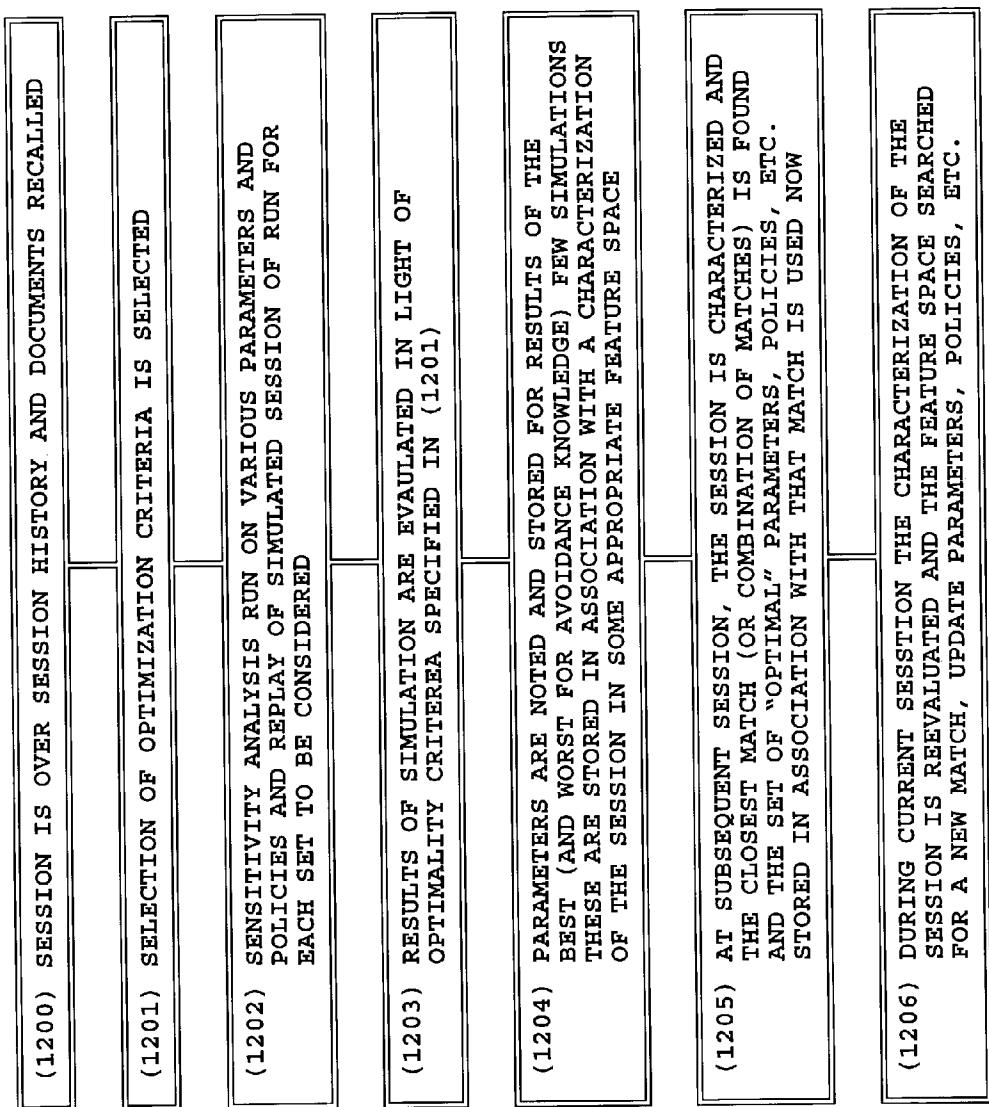


FIGURE 12: POSTMORTEM FINE-TUNING PROCESS

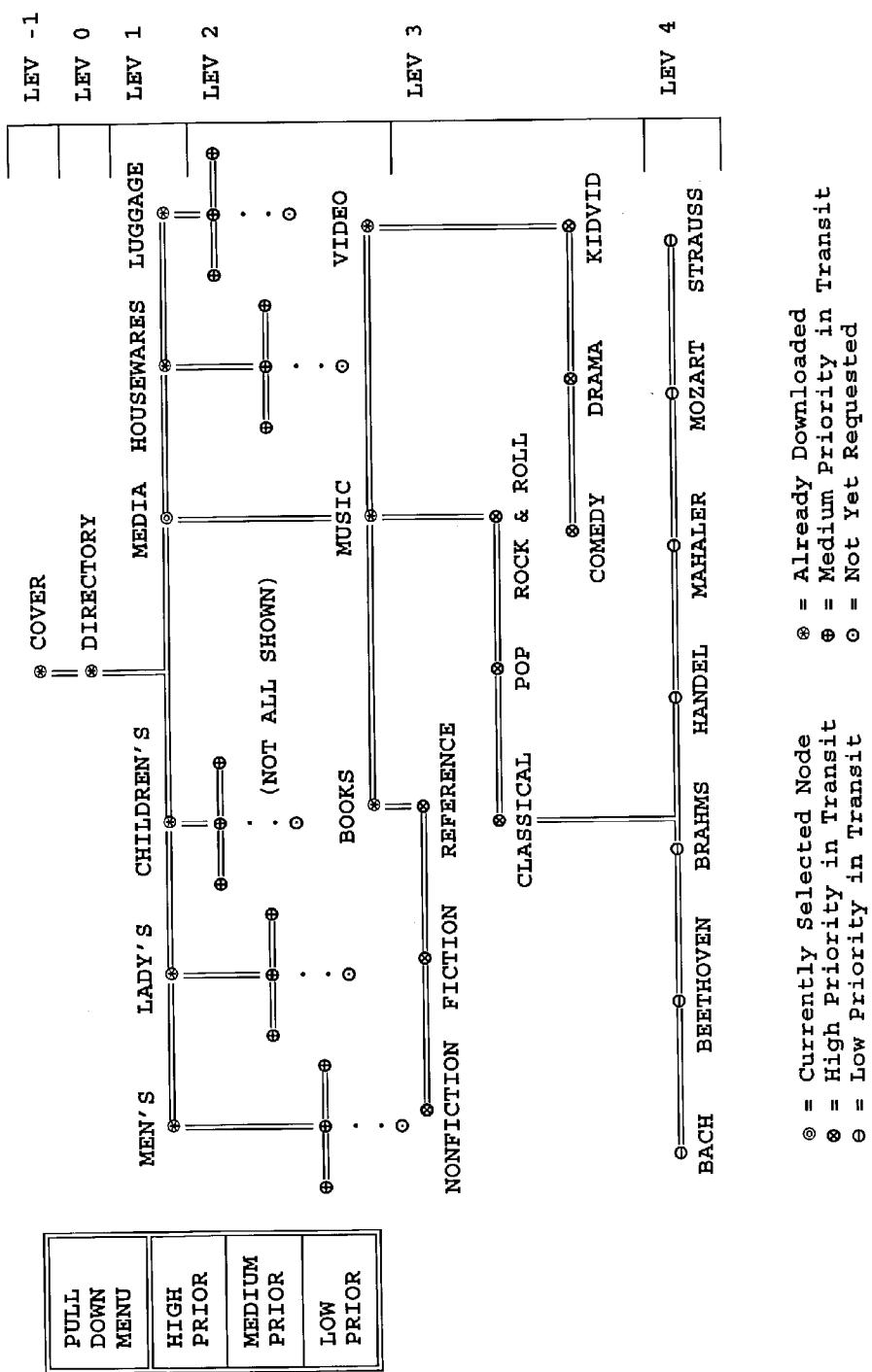


FIGURE 13: MULTI-CLICKABLE TREE NAVIGATION DISPLAY — CATALOG EXAMPLE

ONE METHOD: Sub-Packet Headers

I	T	M	DATA	M	DATA	M	DATA MICRO-PACKET (SUB-PACKET) 3
P	C	P	PACKET	P	PACKET	P	DELIVER TO: FIBR (FIBER)
P	P	1	DELIVER	2	DELIVER	3	ORDER BY: SEQ (SEQUENCE)
H	H	H	TO:	H	TO:	H	SIZE OF PACKET:
E	E	E	FIBR	E	FIBR	E	
A	A	A	ORDER:	A	ORDER:	A	OPTION A: ALL ARE UNIFORM
D	D	D	SEQ	D	SEQ	D	
E	E	E		R	SIZE IS	R	OPTION B: LEN LONG
R	R	R			UNIFORM		
M	S	F		F	UNIFORM	F	OPTION C: NO OF UNIFORM "CHUNKS"
A	O	I		I	OR:	I	
C	C	B		B	LEN	B	FOR HIGH-PRIORITY FIBERS:
H	K	R		R	LONG	R	
I	\	&		R	OR:	&	OPTION A: MORE FREQUENTLY
N	P	S		S	NO OF	S	OPTION B: LONGER SUB-PACKET
E	O	E		E	UNIFORM	E	OPTION C: MORE "CHUNKS" IN A
R	R	Q		Q	CHUNKS	Q	SUB-PACKET
A	T	-		L	MICRO-	L	
D	&	E		E	PACKET	E	
R	S	N	1	N		N	
E	S	/		/		/	
S	S	Q	O	O		O	

(IN THIS EXAMPLE MP 3 IS A LONG HIGH-PRIORITY SUB-PACKET)

ALTERNATE METHOD: Map or Directory in Header

I	T	"MAP"	1 UNIT OF	1 UNIT OF	3 UNITS OF DATA SPACE IS DEVOTED
P	C	OR	DATA	DATA	TO FIBER C, A HIGHER-PRIORITY
P	P	DIREC-	SPACE	SPACE	RESOURCE
H	H	TORY	DEVOTED	DEVOTED	
E	E	OF	TO	TO	
A	A	DATA	FIBER A	FIBER B	
D	D	USAGE	A LOW-	A LOW-	
E	E	EXAMPL	PRIORITY	PRIORITY	
R	R		RESOURCE	RESOURCE	
M	S	FIB A			
A	A	1 UNIT			
C	C	LOW			
H	H	PRIOR			
I	I	FIB B			
N	N	1 UNIT			
E	E	LOW			
R	R	PRIOR			
A	T	FIB C			
D	D	3 UNIT			
R	D	HIGH			
E	S	PRIOR			
S	S	HEADER			
S	S		FIBER A	FIBER B	FIBER C DATA SUB-PACKET
S	S		DATA	DATA	
S	S		SUB-	SUB-	
S	S		PACKET	PACKET	

FIGURE 14: MICRO PROTOCOL

I	M	T	S	
P	A	C	O	
C		P	C	
H		H	K	
H		H	E	
E	A	E	T	
A	D	A		
D	D	D	A	
E	R	E		
R	E	R	S	
S		S	E	
S		S	Q	

MACRO-PACKET CORRESPONDS TO TCP PACKET  
SOCKET ADDRESS A CORRESPONDS TO RESOURCE X  
SEQUENCE NUMBER IS SAME FOR BOTH A & X

I	M	T	S	
P	A	C	O	
C		P	C	
H		H	K	
H		H	E	
E	A	E	T	
A	D	A		
D	D	D	B	
E	R	E		
R	E	R	S	
S		S	E	
S		S	Q	

MACRO-PACKET CORRESPONDS TO TCP PACKET  
SOCKET ADDRESS B CORRESPONDS TO RESOURCE Y  
SEQUENCE NUMBER IS SAME FOR BOTH B & Y

I	M	T	S	
P	A	C	O	
C		P	C	
H		H	K	
H		H	E	
E	A	E	T	
A	D	A		
D	D	D	C	
E	R	E		
R	E	R	S	
S		S	E	
S		S	Q	

MACRO-PACKET CORRESPONDS TO TCP PACKET  
SOCKET ADDRESS C CORRESPONDS TO RESOURCE Z  
SEQUENCE NUMBER IS SAME FOR BOTH C & Z

FIGURE 15: MACRO PROTOCOL

I	M	T	S	A	F	AUGMENTED SUB-PACKET DATA		
P	A	C	O	U	I	FIBER A (NOT SOCKET N) CORRESPONDS TO RESOURCE A		
C		P	C	G	B	SEQUENCE NUMBER IS ASEQ NOT SEQ		
H				K	E			
H		H	E	H	R			
E	A	E	T	E				
A	D	A		A	A			
D	D	D	N	D				
E	R	E		E	A			
R	E	R	S	R	S			
S	S	S	E	E				
S	S	S	Q	Q				

I	M	T	S	A	F	AUGMENTED SUB-PACKET DATA		
P	A	C	O	U	I	FIBER B (NOT SOCKET N) CORRESPONDS TO RESOURCE B		
C		P	C	G	B	SEQUENCE NUMBER IS BSEQ NOT SEQ		
H		H	E	H	R			
H		E	T	E				
E	A	E	T	A	B			
A	D	A		A	B			
D	D	D	N	D				
E	R	E		E	B			
R	E	R	S	R	S			
S	S	S	E	E				
S	S	S	Q	Q				

I	M	T	S	A	F	AUGMENTED SUB-PACKET DATA		
P	A	C	O	U	I	FIBER C (NOT SOCKET N) CORRESPONDS TO RESOURCE C		
C		P	C	G	B	SEQUENCE NUMBER IS CSEQ NOT SEQ		
H		H	E	H	R			
H		E	T	E				
E	A	E	T	A	C			
A	D	A		A	C			
D	D	D	N	D				
E	R	E		E	C			
R	E	R	S	R	S			
S	S	S	E	E				
S	S	S	Q	Q				

FIGURE 16: AUGMENTED PROTOCOL

A1 A2 A3 A4 A5	LAYER ONE
2 1 B1 1 2 B2 B3	LAYER TWO
A1 A2 B1 A3 B2 B3 A4 A5	LAYER ONE + TWO
1 1 C1 4 1 C2	LAYER THREE
A1 C1 A2 B1 A3 B2 C2 B3 A4 A5	LAYER ONE + TWO + THREE

FIGURE 17: A SCHEME FOR DELTA-MODULATION-LIKE UPDATING OF DOCUMENTS

**CLASS OF METHODS FOR IMPROVING  
PERCEIVED EFFICIENCY OF END-USER  
INTERACTIVE ACCESS OF A LARGE DATABASE  
SUCH AS THE WORLD-WIDE WEB VIA A  
COMMUNICATION NETWORK SUCH AS "THE  
INTERNET"**

[0001] This application is a division of application Ser. No. 08/483,205 filed Jun. 7, 1995, issuing as U.S. Pat. No. 6,507,872, and is being filed pursuant to 37 CFR 1.129(b).

[0002] This application is a continuation-in-part of application Ser. No. 08/110,230, filed Aug. 23, 1993 which is itself a continuation-in-part of Ser. No. 07/951,267, filed Sep. 25, 1992.

[0003] This application is also a continuation-in-part of an application entitled A METHOD FOR UTILIZING POINT-TO-POINT COMMUNICATION DEVICES INCLUDING FAX MACHINES AS BI-DIRECTIONAL INTERNET TERMINALS Ser. No. \_\_\_\_ filed Jun. 7, 1995, which is itself a continuation-in-part of Ser. No. 07/800,325 filed Dec. 2, 1991 and Ser. No. 07/933,377 filed Jul. 10, 1992, which were a continuation-in-part and continuation respectively of Ser. No. 07/378,799, filed Jul. 14, 1989, now abandoned.

[0004] Inventor also references his European Patent Number EP 0144352B1. Inventor also references his U.S. Pat. Disclosure Document No. 2,875,39.

[0005] All of these documents (except for those abandoned) are hereby incorporated by reference.

**TECHNICAL FIELD**

[0006] The instant invention comprises a related class of methods, techniques and system designs for improving at least the perceived, if not the actual, efficiency of an end-user's interactive access of large (perhaps cross-referenced, cross-linked or "hyper-linked") databases, such as the World Wide Web, via a communications network, such as "The Internet"; and which can also be applied to the computerized access of local databases.

[0007] This document, to some extent, builds on inventive work disclosed by Inventor in earlier patent applications, issued patents, and disclosure documents, as well as during proprietary development. Where that earlier material is relevant it will, generally, be pointed out in the disclosure.

**SCOPE OF INVENTION**

[0008] Generally, the domain of the instant invention is an extant (generally packet-switching) communications network such as "The Internet". [Alternately, the invention can be adapted, in a straight-forward manner, and applied to a locally accessible database.] Such a network comprises a multiplicity of computers connected via a network over which information is sent: from a first program running on a first computer to a second program running on a second computer; by (if sufficiently large) breaking that information into a multiplicity of packets; and, "wrapping" each packet in addressing information that specifies at least its destination. The specification of how information is parsed, addressed, transmitted, received and reassembled are generally called communication protocols and include, on the internet, TCP/IP, UDP, FTP and HTTP; as well as protocols

that permit an end user to dial-up a service provider's computer on the internet including SLIP, CSLIP and PPP.

[0009] The instant invention relates to a group of related methods for selecting and arranging information prior to its being submitted for distribution via the network and/or re-assembling the information after its being received at the destination. The network and its communications protocols operate on one or more levels below the operation of the instant invention and, thus, the instant invention may be considered a "meta-protocol" or higher level protocol.

[0010] The intended practitioner of the present invention is someone who is skilled in designing such networks and implementing the software that affects and/or utilizes these communication protocols within the environment of a particular operating system (e.g., UNIX). That is, one skilled in the art of this invention is capable of writing program code and/or utilizing and coordinating existing programs within a multi-computer environment. Examples of such related tasks would be the programming of software to implement: a communications protocol such as TCP/IP, UDP, FTP, HTTP, SLIP, CSLIP or PPP; or, an application that utilizes such a protocol, such as an information providing server (e.g., an EMail distribution program, or an FTP or World Wide Web site) or information requesting and receiving client program (e.g., a user EMail program or Web Browser). Many of these capabilities or programs are provided as part of operating systems (e.g., UNIX) and such practitioners are knowledgeable about and skilled in implementing the internals of such operating systems. Often they are referred to as system programmers or software engineers.

[0011] Those skilled in the art know how to accomplish such tasks as to: coordinate and pass information between programs operating within a single computer or on several interconnected computers; initiate requests for local or remote information or database searches; and, generally, to accomplish these tasks by utilizing services provided by operating systems, communications packages and protocols, and other application programs. The details of accomplishing such standard tasks are well known and within the ken of those skilled in the art; are not (in and of themselves) within the scope of the instant invention; and, will be referred to but not described in detail in the instant disclosure.<sup>1</sup>

[0012] Rather, what will be disclosed are novel higher-level methods, system designs and information arrangements, which rely and build upon those known and understood processes. Disclosed will be what one skilled in the art will need to know, beyond that with which he is already familiar, in order to implement the instant invention.

[0013] The above discussion relates to the instant invention in general. However, certain aspects of the instant invention comprise elements including database searching, statistical simulation or artificial intelligence (e.g., learning software and neural networks). Similarly, those skilled in these specialized areas of computer science are well versed in the details of those fields; and, generally, those details are not the substance of this invention. Rather, the disclosure will point out the specific elements of the instant invention to which the standard practices of those fields are to be applied; or, specifically how the standard practices are to be modified to practice the instant invention.<sup>2</sup>

**[0014]** In summary, the disclosure of the instant invention will focus on what is new and novel and will point out, but not repeat the details of, what is relevant of that which known in the art.

#### BACKGROUND AND PRIOR ART

**[0015]** This invention is disclosed in an technological environment that already includes, without limitation, large packet-switching communication networks, such as "The Internet", and large interconnected databases, such as the World Wide Web (WWW). Much of the disclosure will be cast as relating specifically to accessing the WWW on the internet; however, references made to "The Internet" or "internet" may be considered to apply to any similar data network (as well as to the process of accessing a local database) and references made to the WWW may be considered to apply, generally, to any type of local or remote, distributed or coherent, database. The WWW has been chosen as an example because of its complexity as a hypermedia database (which comprises text, images, sound, etc.) which is interconnected and distributed in nature; and, also, because of the current interest in, and increasing usage of, this particular database.

**[0016]** In addition there are many extant programs and mechanisms for accessing and initiating computer mediated searches of such databases. For example, gopher, WAIS, archie, veronica, web browsers (including Mosaic and Netscape).

**[0017]** Communication between various computers on the internet (and computer programs running on those computers) is affected by various communication protocols, including TCP/IP, FTP, UDP, HTTP, SLIP, CSLIP, PPP, etc.

**[0018]** There are already extant many programs and techniques for: running computer simulations and analyses of systems such as communication networks; compiling and analyzing statistics, including sensitivity analysis; and artificial intelligence techniques such as complex searching algorithms, programs that learn or train, can recognize patterns, neural networks, etc.

**[0019]** Lastly, there are a large number of service providers who will make available to users, who have personal computers or workstations, a connection via modem over a dial-up telephone line (or ISDN or dedicated line) through their computer to the internet. However, accessing, searching and browsing the WWW through such a provider can be a painfully slow process.

#### THE PROBLEM TO BE SOLVED—IN BRIEF

**[0020]** As explained, above, the process of "surfing the Web", let alone trying to accomplish serious work, via the internet can often be a slow and frustrating process; especially if accessing via a modem through a service provider but, often, even if directly connected to the internet. Periods of being flooded with information may alternate with long stretches of waiting for information to arrive.

**[0021]** As working, playing or just browsing over communications networks, via digital computers, become ubiquitous, the number of users, number of "sites", and the shear bulk of data being accessed, is growing exponentially.

**[0022]** The introduction of graphical or hypertext document systems (such as the HTML language, the HTTP

protocol and browsers such as Mosaic) to the Internet, and the transformation of that network into the World Wide Web (WWW), have tremendously increased the requirements for bandwidth: both because of the increased requirements of graphic (or even video and audio) components of documents, "pages" and files; and, because this richer, multimedia content is attracting users beyond a threshold where the commercial potential becomes obvious. This, in turn, prompts more "information providers" to take the plunge, increasing the size of the browsable pool of content, and a positive-feedback spiral results.

**[0023]** The increasing demands for raw bandwidth and "network infrastructure" are creating a technical and logistical problem which is, on the other hand, perhaps an economic and business boon. The class of inventions described herein will likely only exacerbate this situation, as will become clear later.

**[0024]** Rather than being directed toward reducing requirements for bandwidth—i.e. traditional bandwidth reduction—these techniques are designed to increase the performance, efficiency and value of the work that can be done using whatever level of such communications technology is available. That is, the goal here is not to reduce the size of the channel needed, but to—apparently—squeeze more information down the available user channel.

**[0025]** This is similar to an approach used in Inventor's invention which is subject of U.S. patent application Ser. No. 07/077,916 where Inventor states:

**[0026]** The goal of most bandwidth compression schemes is to "squeeze" a given information signal into a smaller channel. A main goal of the instant invention is to "squeeze" a larger picture into a given channel. Although this may at first sound as a semantic difference only, there is a significant practical difference.

**[0027]** Broadcast television in the US is a "given channel" that is unlikely to change for the foreseeable future. Bandwidth compression schemes that would allow the use of a smaller transmission channel and require complex decoding at destination could not be used with the many millions of existing home receivers.

**[0028]** Similarly, most users are tethered to communications networks via telephone dial-up lines and modems which—although, in general, modems get better and higher-bandwidth telephone service is becoming available—for any given user, at any given time, are limited and usually not sufficient for the user's desires. That is, working over a communications network is usually perceived as slower than working locally (or over a high bandwidth network, such as a Local Area Network or ISDN or T1 digital telephone line).

**[0029]** However, whatever the given network connection, these techniques can be used to advantage to increase actual or perceived performance (of work, if not strictly of bandwidth) for many types of work done over such networks.

**[0030]** In particular, searching or browsing through large amounts of information (cohesive databases, distributed masses of information such as the WWW, or otherwise) is a task that often goes in fits and starts. The user makes some decision about what is desired and makes a request for more remote information: as a formal database search request; by clicking on a "link" on a WWW page; or otherwise. The user

then waits for the information to be delivered to his local screen. During this period the user often sits idle. The user then reviews the new information. During this period the local computer and communication line often sits idle. And so on.

[0031] The goal here, then, is to: speed up the overall process and make it more efficient; make the user feel that the process is more immediate and valuable; shorten work time; or, alternately, get more done in the same time.

#### CONCEPTUAL ORIGINS

[0032] In an early stage of the development of the techniques described herein, Inventor recognized that: remote computer databases that are searched for professional purposes are becoming more widely used, particularly in the legal, engineering/scientific and medical fields. Further, the amount of data to be down-loaded is increasingly large.

[0033] Communications and on-line costs are now often a significant part of the cost of doing business. Efficient use of communication and database search facilities is therefore desired.

[0034] We have developed a group of related techniques (technical details are described below) that will make more efficient use of the open communication line, saving money. These will also speed the search procedure, saving the working time of the lawyer or other professional using the system, which professional time is much more valuable than communications time.

[0035] These techniques will also be useful for local searching of databases contained on magnetic disk or CD-ROMs.

[0036] A narrative example is as follows: A lawyer has requested from a remote database, and is reviewing on screen, a particular legal case. In the most simple embodiment, all cases referred to by said case would be retrieved by the remote computer and initial sections (pages or paragraphs for example) of each case would be transmitted to the lawyer's workstation. In this way when the lawyer was ready for a next case, it is highly likely that some of that material would already be available at the lawyer's station for review. As soon as a selection is made by the lawyer, transmission time is wholly devoted to that selection rather than multiplexed between multiple potential selections. The relative amounts transmitted of each potential selection, and the list of potential selections, can both be modified by factors such as: the length of each document; statistical "knowledge" of the lawyer's prior access to that particular document; statistical "knowledge" of the lawyer's prior access to that type of document; document dates; search of abstracts and key words; prior statistical knowledge of lawyer's work style; and other factors.

#### SUMMARY OF INVENTION

[0037] As explained, above, communication while accessing remote databases over the internet often runs in fits and starts; the communication channel is often left fallow. [At least the communication line between a service provider and an end user; the communication line between the service provider and the internet, in general, may be in more uniform use, serving many users simultaneously.] The intent of the instant invention is to make use of that fallow time

and, thereby, improve the perceived (if not actual) responsiveness and efficiency of the network to the user's demands.

[0038] A specific deficiency that arises when accessing WWW "pages" is that each page consists of a text script and, optionally, a number of associated image or other multimedia resource files. A separate communication connection (a logical data construct, not a physical wired connection) is established, used, and then broken down for the delivery of the script and each resource. This use of a multiplicity of connections results in a significant amount of overhead and delay. One element of the present invention will be to establish and leave extant a communication connection that will survive the delivery of a single file, and which can be used not only for delivery of all the resource files comprising a document or "page", but for the delivery of multiple documents from the same source to the same destination. This technique, in and of itself, may not constitute inventive matter; however, the composite method, which this and related techniques comprise, and new uses to which they are put, are.

[0039] Similarly, it is standard that an end user's client program (that program which permits the user to "browse" and which initiates system requests for information resources) permits a single request to be made, waits for the requested information to be delivered, and then waits further while the user reviews the delivered item and decides what to do next. Only then is the user permitted to make a next request. In the instant invention multiple, parallel requests will not only be permitted, but in some embodiments the system will make those multiple requests in anticipation of user requirements. Again, this technique, in and of itself, may not constitute inventive matter; however, the composite method, which this and related techniques comprise, and new uses to which they are put, are.

[0040] The more substantive elements of the present invention comprise the following:

[0041] 1. The interleaving or time-division multiplexing of multiple data resources (generally, files) into a single (or at least composite) communication stream of information, by fragmenting the resources into portions and interleaving those portions.

[0042] 2. The weighting of the presence of the individual data resources within the composite information stream by modulating the size and/or frequency of representative portions to be included.

[0043] 3. Attempting to anticipate the user's likely future requirements on a "predictive probabilistic" basis. Further, the flexible and/or artificially intelligent adjustment or customization of the algorithms that predict and determine the interleaving weights for a particular user.

[0044] 4. Determining and updating the interleaving weights (2) based on the anticipation (3) and actual events.

[0045] 5. The "pyramidization" of data resources in such a way that a partially delivered data resource is, at least, of some use to the end user; perhaps, if only, to evaluate whether the balance of the data resource is required.

**[0046]** These five elements will be disclosed in more detail below, as will their integration into a class of methods and protocols that will permit more efficient usage of communication networks by end users. In particular, new communication protocols and methods will be disclosed that function: as a layer above existing protocols; as a layer below existing protocols; as addons or extensions to existing protocols; or, as entirely new protocols or methods.

**[0047]** Further, it will be disclosed how to apply these new methods to a number of different system architectures. Thus, a group of related embodiments will be disclosed which can be appropriately applied to different situations. Not all of these variations will embody each of the above elements to the same degree, if at all.

#### END USER COMMUNICATION CHANNEL USAGE UNDER VARIOUS CONDITIONS

**[0048]** Referring, now, to **FIGS. 1 through 6** (reading from top), a six minute interval during an end user's interaction with the WWW is depicted, in a schematic fashion, under various conditions. These figures are simplifications and exemplary in nature. They are being presented to better explain how several of the inventive elements, introduced above, can be applied to increase the efficiency of end user interaction with the network.

**[0049]** **FIG. 1** (top) shows the standard situation where, in response to an end user action (e.g., specifying the designation for a WWW site home page) the system requests (R, at time zero) a first page (1). About ten seconds elapse before the network begins to deliver page 1, which delivery (D/1) continues for another 50 seconds being complete at 1 minute. The user has waited for a full minute of "thumb-twiddling" and, at that point, takes three minutes to examine (E) the information on that first page (1) and decides to link (by clicking on a highlighted word or icon, denoting hyperlinks to additional pages elsewhere on the network) to a second WWW page (2). In this case this choice of a second page was made from among six possible (a, b, c, d, e or f) secondary pages. The system then requests (R, at time 4 min.) the second selected page (2). About ten seconds elapse before the network begins to deliver page 2, which delivery continues for another 50 seconds being complete at 5 minutes. End user then examines (E) the second page (2).

**[0050]** **FIG. 2** (second from top) shows that, in response to an end user action (e.g., specifying the designation for a WWW site home page) the system requests (R, at time zero) a first page (1). About ten seconds elapse before the network begins to deliver page 1, which delivery (D/1) continues for another 50 seconds being complete at 1 minute. The user has waited for a full minute of "thumb-twiddling" and, at that point, takes three minutes to examine (E) the information on that first page (1). However, during the three minutes that end user is examining page one, the system notes that page one has six potential hyperlinks to pages A, B, C, D, E and F and, in a serial fashion, begins to request those pages from the network. Each request takes one full minute (10 seconds to request (R) and 50 seconds to deliver(D)) and by the time end user has made a decision, three minutes later, the first three of those pages (A, B & C) have been requested and delivered. Thus, when end user "clicks" on the hyper link to page C, the network's response is apparently instantaneous—a vast (100%) improvement over the situation shown in **FIG. 1**.

**[0051]** **FIG. 3** is identical to **FIG. 2** except that, when end user makes a decision, it is to click on the link to page F instead of page C. Since page F has not yet been downloaded, a request goes out at the four minute mark and, by the five minute mark, the page is delivered with performance no worse than was exhibited in **FIG. 1**.

**[0052]** In both **FIGS. 2 and 3**, the system is employing a simple embodiment of the inventive element of predictively anticipating the end user's future requirements. In **FIG. 2**, the system succeeds in its prediction and greatly improves response time; in **FIG. 3** the system has failed and responds no worse than in the standard situation. However, we would like to improve performance in all cases.

**[0053]** In **FIG. 4**, another inventive element is added and the requests (R) for all six pages (a-f) go out to the network simultaneously (or, at least in very quick succession) in parallel, and packets from all six pages begin to arrive soon after and continue to do so. The specifics of internet communication protocols (e.g., TCP/IP) permit the packets from each separate page to be identified and sequenced. At the four minute mark, end user makes a decision (again, page C but, in this case the results will be the same for any of the six pages) and the information for page C is already half there, indicated by (c). At that point (4 min.) the end user's system requests (r) the service provider's computer (or in other embodiments, the network or information provider) to now focus all available communication bandwidth to forwarding remaining packets of page C (c). [Affecting such a request and response is not part of the standard capabilities of internet operation and will be explained in greater detail below.] Since each page's delivery over the network takes approximately one minute (in this simplified schematic example) by the four minute mark all of the packets for all of the pages will likely already have arrived at the internet service provider's computer, and be awaiting delivery, via modem, to the end user's computer. It takes about an additional 20 seconds to download the remaining packets for page C from the internet service provider to the end user.

**[0054]** Thus, at the 4:20 mark page C is completely there. At worst, performance has improved over the standard situation by more than 66%. The advantage here (over the case in **FIGS. 2 & 3**, where performance is improved either 100% or 0%) is that the 66% improvement is achieved no matter which of the six pages are selected. The worst case wait is 20 seconds where, previously, the worst case wait was a full minute (even though the best case was no wait at all). Generally, the perception of system performance is tied to worst case. That is, always waiting about a half minute will be perceived as less annoying than waiting a full minute half the time and not at all half the time; even though the average wait is the same (in fact, the average is :30, more than the :20 here).

**[0055]** Further, since more than half of page C is already there (at time 4:00) with zero wait, an additional inventive element permits the first part of the file to be displayed while the balance is still in transit. Thus, effective performance can still be perceived as zero wait, or 100% improvement because, by the time the end user progresses through the document beyond the half-way mark, the complete file will, in all likelihood, be available. Document "pyramidization", an additional inventive element described in detail below, can increase the perception of performance further, by

reordering the information in the document in order to make a partial delivery more useful, especially for making decisions such as “whether to wait for the completion of the document” or “which page to link to next”.

[0056] **FIG. 5** depicts an additional inventive element. Previously, the “prediction” was uninformed; all potential choices were treated as equally likely. In **FIG. 5**, “probabilistic prediction” and weighting of the interleaving are added. Based on some inference (e.g., the displayed words associated with the hyperlink to page C match some entry in a list of key words that the system has on file associated with subjects of interest to this particular user) that page C has a higher probability of being what the end user will select, the number of packets devoted to page C is doubled, compared to any of the other five pages. Thus, when C is selected, it is already there, complete.

[0057] **FIG. 6** depicts the same situation as **FIG. 5** but, in this case, the prediction was off and page F was selected. Performance in this case is approximately the same as for any page selected in the case depicted in **FIG. 4**; the maximum wait till completion is .25, an approximate performance increase of about 58%; but a perceived performance increase of up to 100% with partial display or “pyramidization”.

#### EMBODIMENTS FOR VARIOUS SYSTEM CONFIGURATIONS

[0058] Depending upon where, within the network, the inventive techniques may be applied, there will be different embodiments. These are depicted in **FIGS. 7 through 11**. When the computer(s) collaborating on the invention are few, the inventive embodiments are similarly limited, when the improvements are system wide, the most capable and flexible embodiments result.

[0059] In particular, **FIG. 7** shows a minimal embodiment where the client software, running on the end user system, contains the only inventive elements, and these are, thus, limited. The client attempts to anticipate potential user requirements and requests delivery of those resources, during fallow communication time, while the user is examining prior information. These may be requested serially, as shown in **FIG. 2** or in parallel, as shown, for example, in the early part of **FIG. 4**. Since, in this embodiment, the invention resides solely in the client software, the interleaved reception of packets from parallel requests will be haphazard, at the variegations of network traffic. Further, the shift in transmission, depicted at time 4:00 in **FIG. 4** cannot be counted upon as being possible.

[0060] **FIG. 8** shows another minimal embodiment where the information provider server software, running on the computer “at the opposite end” of the network, contains the only inventive elements, and these are, thus, also limited. The server attempts to anticipate potential user requirements in an uninformed way, as best it can, and starts delivery of those resources, in addition to that which was requested, in response to a simple request for the “tip of the iceberg”. This approach will be disclosed in greater detail later herein.

[0061] **FIG. 9** depicts a situation where both the client and server software share complementary inventive elements, with much improved performance. However, in general, other computers on the network are unaware of the invention

and, as will be disclosed in greater detail later, this enhanced “meta-protocol” operates within the confines of existing communication protocols. Thus, while specific client/server pairs may operate in an enhanced mode, such operation cannot be counted upon to be generally available.

[0062] **FIG. 10** depicts a situation quite similar to **FIG. 9**. The difference here is that the entire network is made “aware” of the inventive techniques by the adoption of a novel or enhanced communication protocol. In that case, enhanced operation may, generally, be expected; and, further other computers which pass the communication up and down the network can be counted upon to help.

[0063] **FIG. 11** depicts an embodiment (called THE FUNNEL™ by Inventor) where an enhanced client program running on an end user’s system, and an enhanced communications utility running on a service provider’s computer, collaborate to provide performance approaching that available from either of the two prior embodiments, although the various information provider’s server computers run standard software and use standard communication protocols. As will be explained more fully below, with this embodiment overall network bandwidth used by the information provider and service provider servers, on behalf of the end user, may increase significantly, in order to make modest additional use of the end user’s communication line; but, this will provide marked improvement as far as the end user is concerned.

#### INTELLIGENT ASSISTANT

[0064] The enhanced communication protocol technology disclosed herein will, to a great extent, be made useful by an “Intelligent Assistant” program that will anticipate a user’s potential needs for information. That assistant will search ahead looking for potential information resources to download. It will then make probability assessments based on the document in question, what is currently being looked at, the stated and “inferred” goals of the current work session, the stated and “inferred” general goals and work style of the user, etc. Next the assistant will initiate parallel requests for multiple information resources and will weight the amount of available bandwidth that should be applied to each request based on the probabilistic assessment made previously.

[0065] In this way the access to the fallow communication link capacity which is made available through the novel (parallel and weighted) communication protocols can be put to optimal use by the intelligent assistant program while the user is otherwise engaged.

[0066] Further, by analyzing its performance, and tracking the user’s habits, the intelligent assistant will continually increase its performance and keep from becoming outdated.

[0067] Extensive discussion of how these techniques can be applied will follow below; however, first, the techniques will be described.

[0068] Probability Assessment:

[0069] As will be described in much detail below, probabilistic assessments may be made for many documents, by matching keywords and other elements of documents, and many other techniques. These processes are controlled by a complex and sophisticated set of options, parameters and

policies. These, in turn determine how the probabilities will be calculated and what policies determine how weights will be set for various document types, specific documents, or pyramidal levels. For example, a document with a high probability may still be given a low priority if it is of a particular type (say an image).

[0070] Whichever these mechanisms are chosen to be implemented (and, details of choices are presented below) the important aspect to consider here is that the mechanisms are very flexible due to the many parameters, choices and policies that may be combined with the various “search and match” mechanisms that add up to a probability assessment.

[0071] Post Mortem:

[0072] After each session the intelligent assistant will perform a “post mortem” or “Monday morning quarterback” type analysis of the session.

[0073] This will consist of running a simulation of the session, over and over, with the various parameters being adjusted for each simulation. The goal, by whatever searching and pruning techniques are applied, is to find (as best as possible or, at least, reasonable) what would have been the best parameters and policies to have implemented in the prior session in order to optimize performance.

[0074] [What constitutes optimize is another matter entirely. That is do we wish to minimize the average wait time, minimize the maximum weight time, shorten the work session, etc. These are a matter human choice and it will be assumed here that some preferred measure of performance has been chosen by the user.]

[0075] A “simulation” will consist of re-playing the session. A history of the session and documents being kept by the system for this purpose and then, if the documents have not been tagged as to be saved, by the user, they will be discarded. For each simulation the parameters and policies will be adjusted and the system will calculate the various times to download documents (based on the data kept from the actual session history) with different weights and policies in effect. It will then match what documents “arrived” and when, and match these against which documents the user actually decided to look at (as well as how long he looked at them, if he kept the document, if he used it to make another “hyperlink” jump to another document, etc.). In this way the machine may construct a simulated work session and determine how much and when the user would have been without whatever he needed.

[0076] Whatever optimal criteria will then be applied to come up with a figure of how convenient or inconvenient the session would have been with the particular policies and parameters in effect.

[0077] A “sensitivity analysis” will then be run on the various parameters to see which effect the session positively or negatively; alone or in combination.

[0078] Then, the system will keep in its long term storage the results of this sensitivity analysis in association with a characterization of the work session in some feature space (e.g., what type of work was being done; how did the user “feel” (or at least reported he felt); what types of documents were accessed; how active was the user, how continuously was he working; how long was the session; what types and amounts of information was kept vs. discarded; etc. etc.).

[0079] Then, the next session the machine will determine a characterization for the current session (which can be continually updated during the session) and try to match that characterization to the characterization of other sessions in feature space. For such a task artificial intelligence techniques including pattern recognition and neural networks are well suited.

[0080] The system will then chose the parameters and policies that are associated with the closest match in feature space (or, may combine parameters and policies from several partial matches if there is not a close match).

[0081] However, the system will, of course, be programmed to give override priority to explicit user requests. For example, if the user selects from a menu that he wants pictures with his data today, the system will not deny him just because it will speed up the session. However, the system will, in fact, be programmed to “suggest” such a course of action. [However, it will also be made possible to suppress such suggestions, in order to prevent user agitation.] The above method may be combined with the method of the intelligent edge creation (a system that “watches over the shoulder of the user”) disclosed in Inventor’s earlier patent application. The combination will provide a trainable intelligent assistant shell that can be applied to many applications where computers and humans interact on a repeated basis.

[0082] What follows are descriptions and a number of narrative examples of how the parameters, policies, probabilities and weights may be determined and utilized. See FIG. 12 for a flow diagram of this “post-mortem simulation” technique.

#### PROBABILISTIC PYRAMIDAL INTERLEAVING

[0083] Emulsification of Blobs:

[0084] The section CONCEPTUAL ORIGINS, above, outlines a process that Inventor calls “Emulsification of BLOBS”. A BLOB is (someone else’s) acronym for a Binary Large Object—that is, a large binary file that can be a large text document or, as is common now, a file or document that contains graphics, sound, video or other multimedia or high-volume components.

[0085] As was described, in the previous section, a number of files—or BLOBs—are interleaved for transmission (multiplexed). The present Inventor has been working for some time with novel ideas for interleaving and/or multiplexing signals, dating at least back to the mid-1970s and an audio gaming project.

[0086] A process in physics where blobs of oil are dispersed in water (or with two other immiscible liquids) is referred to as emulsification. The blobs are broken up and a uniformly mixed fluid, containing bits of material from all the blobs, suspended in the other medium, can be smoothly poured.

[0087] By analogy, bits (only informally, the file “chunks” likely being larger than a single bit) of many different BLOBs will be interspersed in the digital transmission medium so that a “smoothly mixed information fluid can be poured” down a transmission channel.

[0088] In a “naive” version, the amount of information interleaved from each file would be the same: e.g., the first

8 bytes from file 1, followed by the first 8 bytes of file 2 . . . followed by the first 8 bytes from file N, followed by the second 8 bytes from file 1, etc. In a, perhaps slightly less naive, alternative the chunks might be weighted by the length of the file: e.g., the first 1 % of file 1, followed by the first 1% of file 2 . . . followed by the first 1% of the file N, followed by the second 1% of file 1, etc. In this way, information from all the files being transmitted run out approximately at the same time.

[0089] Neither of these schemes take into account what Inventor calls Probabilistic Predictive Weighting.

[0090] Probabilistic Predictive Weighting:

[0091] With the above two examples, the size of portions, taken from a multiplicity of files to be interleaved, were: a) fixed in absolute uniform size; or, b) fixed in a size proportional to the size of each file.

[0092] A second major component of the present group of inventive techniques comprises techniques for weighting the size of portions to be interleaved.

[0093] First, each file can be given a base portion size, or weight, of 1 unit. And this can be of the absolute or proportional type.

[0094] Next, the weight (i.e., portion size, or frequency of inclusion of portion) can be adjusted, in a straightforward or "unintelligent" fashion, depending upon known factors. For example, graphic images can be given precedent because of their, generally, larger size compared to text files. Or, the exact opposite conclusion can be reached because graphic files generally take much more bandwidth than text but yield less equivalent information. By cliche, a picture may be worth a thousand words, but it is often possible to hold tens or hundreds of thousands, or even millions, of words in the same file space. Other straightforward policies can be made on similar, or other unrelated, bases.

[0095] However, more important (or at least more inventive) other factors can be used to adjust the interleaved weighting factors for files; and, these can be updated and adjusted on an on-going basis during a session.

[0096] The basic idea is that while a particular item of information is being reviewed by the user, the system will be making a best guess at which several items the user may be interested in accessing next. These multiple items will then be downloaded in interleaved fashion and the weights continually updated so that, stated simply: at any given time, there is the greatest likelihood of the most information required next, being correctly anticipated and delivered, or at least in the process of being delivered.

[0097] Weights can be adjusted depending on: a) file size (as described above); b) file type (also, touched on above); or, c) an estimated probabilistic prediction of the likelihood of the particular item being later required by the user (described following).

[0098] The general principle for Probabilistic Predictive Weighting is to weight more heavily (i.e., increase the size or frequency of included portions) from files that are more likely to be (i.e., have a higher probability of being) the (or, at least, one of the) next item that the user will want to access. In this way, the expectation can be increased of having more of what will be needed next already down-

loaded before it is even requested. Thus, apparent instantaneous response can be achieved.

[0099] Once an item is actually selected for viewing by the user, its weight is increased (perhaps, but not necessarily, getting all of the available bandwidth) so that the completed document will be available ASAP or, at least, sooner.

[0100] The probability estimates can be made based on the interaction of several factors. Each file can then be given a composite probability index, perhaps with the sum of the indexes being normalized for a total of 1 unit or 100%. The relevant factors can interact in a non-trivial way and elements of Artificial Intelligence (AI) can be applied. Related AI-enhanced user interfaces have been described by others; and also by Inventor, for example, in his U.S. patent application Ser. No. 08/435,487 and Ser. No. 07/227,403 since issued as U.S. Pat. No. 4,925,294.

[0101] The basic factors include a straightforward or more "intelligent" assessment of:

[0102] a. The to be weighted document's size, type and particularly its content or other associations;

[0103] b. The currently viewed document's content and relationship or linkages to the weighted document(s);

[0104] c. The user's stated goals, work style, and session "profile" as stated by the user;

[0105] d. The system's own historical/statistical assessment of the user's "profile" or habits and activities, and the content, linkages and relationships of documents accessed by the user, during the current session; including the content, linkages and relationships of prior documents in the current session;

[0106] e. The user's general work habits, preferred work style, and "profile" as stated by the user;

[0107] f. The system's own historical/statistical assessment of the user's general "profile" or habits and activities, and the content, linkages and relationships of documents accessed by the user, in earlier sessions; in particular, an assessment must be made of the type of earlier sessions (documents) and the relevance of those earlier sessions (documents) to the current session (document).

[0108] Note that the term document as used above may be taken rather loosely and can refer to an entire file or coherent document (text or multimedia) in the usual sense; part of such a file/document, divisions being made on the basis of sequence, content type, format type, media type, etc.; a monomedia graphic, video, sound or text file; a "page", "card", or other subdivision of a linked hyper-media or multimedia document; a W3 page; a referenced or linked document of any of the types above, etc.

[0109] The above 6 factors were stated in rather abstract fashion. In the interests of clarity, some concrete, specific and/or narrative examples of each type, and of some of the complex interactions between types, will follow. All of the "policy" type issues mentioned can be explicitly specified by the user, or can be "statistically intuited" by the system based on its "watching over the user's shoulder". A number of different collections of policies (profiles) can coexist on

the system, and can be selected from (or combined) by the user; or, the system can choose what it determines is the best match of profile(s) to the current user work session.

[0110] Under item (a) which depends solely (primarily) on factors of the document to be weighted:

[0111] 1. Large documents can be weighted more (or less) heavily; etc.

[0112] 2. Graphics can be favored over text, sound over graphics, etc.; abstracts can be favored over full text; reviews can be favored over abstracts; large-circulation documents can be favored over small-circulation; on-line versions of print publications can be favored over electronically-originated documents; newer (or older, or in a particular date range of) documents can be favored; etc.

[0113] 3. Content (or association) decisions can be made on the basis of: key-words, abstracts, headings, full-text, etc. searches; similar assessment of documents linked to or from, or referenced by or in, or otherwise related to, this document; authorship; assessments of the identity or profiles of those others who have accessed this document (if known); etc. Each "hit" can count equally, or different types of hits can themselves be sub-weighted, or a complex or non-linear relationship between number and type of hits can be made.

[0114] Under item (b) where factors are related between the current and weighted documents:

[0115] 1. Relating content factors (key words, etc. as described above) between the two documents—for example, matching more key words would be weighted higher; the number of cross-references—for example, a document cited several times can be weighted more heavily than one referenced once, or one referenced forward and backward would be weighted additionally; both documents in question being referenced by (or, more easily determined, referencing) the same third document would increase the weighting; etc.

[0116] Under item (c) where factors are related between user session specifications and weighted documents:

[0117] 1. Stated goals such as: key words (or any of the policies or items described above [or below]); organization types (commercial, academic, research); author names or types (researchers, MDs, etc.); publication types; etc.

[0118] 2. Instructions such as: deliver only abstracts (or other sub-set of text) first, and full text later, or only if the document is actually viewed, or only when specifically requested (if at all); deliver text first and pictures or graphics later, or only if the document is actually viewed, or only when specifically requested (if at all); etc.

[0119] 3. Instructions as to how to weight, or which to use of, the policies or factors described above or below: such as "only pre-send if at least two factors of a particular type and any other mix of three factors are "hits" (e.g., two key-word matches and any three types of references)"; or, "send abstracts of all ref-

erenced documents, but only send full text if at least one third of key words match"; or, "send full text of any referenced document that matches any three factors (from a detailed list not show here) with any of the documents I have already looked at this session"; etc.

[0120] Under item (d) where factors are related between user's session history/statistics and weighted documents:

[0121] 1. As just one example: the system might note that (92% of) all documents actually looked at during this session had the words "virtual reality" in the title, key word list, or abstract and, thus, more heavily weight any potential document that has those words somewhere in the text (or even more heavily for the title, abstract or key word list). Further, the system might notify the user that such an "system inferred point of interest" were noticed.

[0122] Under item (e) where factors are related between user general (not session specific) profile specifications and weighted documents:

[0123] 1. For specifics see item (c) above. Such policies or preferences can be specified in a general profile; and can further be specified for separate profiles for different types of sessions, e.g., academic research, commercial research, legal research, browsing; different states of mind, e.g., feeling sharp vs. feeling fuzzy, feeling like being detailed and in control vs. feeling like letting system make more suggestions, etc.

[0124] Under item (f) where factors are related between user general (not session specific) history/statistics and weighted documents:

[0125] 1. For specific example see item (d) above. However, such statistics and inferences can be made over many sessions, perhaps only considering sessions that match the current session in type, goals, state of mind, etc.

[0126] 2. As another example: the system might note that during work-related research sessions, the user almost always reviews the first two documents in toto, irrespective of content, and then (as the user gets frustrated and more discriminating—although the system will be working with statistics rather than psychological analysis) the user views abstracts first and (usually) only downloads documents that match the stated session keywords.

[0127] Pyramidal Ordering of Information Elements:

[0128] The next component of the instant inventive system is what Inventor refers to as "Pyramidal Ordering of Information Elements".

[0129] In short, this term covers a class of techniques (some existent, some new to this invention) that permit small versions of various information elements to be delivered, which contain reduced detail but which are still useful and, in particular, useful for the task of evaluating whether the user wants to access a fuller version of the same information element or document.

[0130] This is not the same as information compression, in which a complete information element is encoded into a

smaller form, and can be losslessly (or, in some cases, notably for video and graphics, with some loss) reconstructed at a later time. Any sort of information compression, including those disclosed by Inventor in his U.S. Pat. application Ser. No. 07/077,916, can be used in conjunction with the “pyramidal” techniques disclosed herein.

[0131] The term pyramidal is being used here in an informal somewhat prosaic manner. There are a class of mathematical functions with a similar name and Inventor is not necessarily drawing an association with those specifics.

[0132] Rather, the intent here is to use the sense of the term pyramidal as is used to describe a desired structure for journalistic writing. Essential facts are disclosed first, followed by a somewhat more complete restatement, followed by a wealth of detail; thus, progressively more detailed and larger versions of the same story follow each other, creating an expanding or pyramidal structure. One can get the gist by a quick perusal of the opening of the article, and may then decide whether to pursue the fuller story by continuing with the more complete and detailed sections which follow.

[0133] It may already be common practice to organize some information elements, notably graphics, in a manner that may correspond to Inventor's intention for pyramidal organization, and Inventor is not necessarily claiming the invention of such techniques here. Nevertheless, the use of such techniques for the purposes described herein may, in fact, constitute a new use or inventive improvement over the current art and, to that extent at least, Inventor is claiming inventorship. In particular, some graphics that are delivered in what might be described as pyramidal fashion do not exploit that structure and: are neither interleaved with other information elements; nor is there the ability to make a stop/continue decision between layers or levels delivered.

[0134] Further, Inventor believes that several other pyramidal organization methods are new and unique to the instant invention.

[0135] Beyond what is described above, it is the further intent of pyramiding that each subsequent “layer” of information (as much as possible) not repeat the information in the prior level. Rather, as much as possible, the new level of information should provide additional detail, not already present, which can be interleaved, or otherwise combined, with the prior level(s) of information already delivered.

[0136] See FIG. 17 for a diagram of how several pyramidal layers may be integrated.

[0137] 1. Examples (which can be used separately or in combination) of pyramiding with a raster graphic image:

[0138] Some pixels in a line (and some lines of the image) can be sent first, constituting a low-resolution version of the image. This can be displayed small, or large with large blocky pixels, or large with smooth (low-pass filtered) large pixels. Then additional pixels from already sent lines, and additional lines can be sent and interleaved with the prior information for a higher-resolution version of the image. Third and more levels of detail can also be sent subsequently. The parsing process may not necessarily be as simple as taking every Nth pixel and every Nth line; the information can be low-pass filtered or more complexly

filtered prior to sampling; other data sampling/compression schemes can be utilized within the pyramiding process.

[0139] Some bits of a pixel can be sent first, followed by more bits later. For example, a 24-bit (8 bits each for Red, Green and Blue) image can be reduced to 8 bits total by sending the most significant 3 bits each of Red and Green and 2 bits for Blue. Alternately, a custom optimized 8-bit look-up table (LUT) can be calculated for the image. Later, additional significant bits (3G, 2R, 3B; followed by 2G, 3R, 3B) can be sent. This is just one example, and bit-parsing can be combined with pixel or line parsing.

[0140] In addition, for strings of raster images, i.e., video, the temporal element can be pyramided. For example, very simply, every 64th, 32nd, 16th, 8th, 4th, 2nd, 1st frames can be sent (if not sent in a previous level) as separate levels; or, frames containing major content changes (i.e., key frames) can be sent first; etc. Again, Inventor's U.S. patent application Ser. No. 07/077,916 should be consulted.

[0141] Also, a vector, contour or outline version (for example, with flat color or textured fills of enclosed areas) can be derived by analysis of a raster image. This reduced information version could be sent first (perhaps itself sent at several levels of detail) to be replaced by the full raster image at a later level(s).

[0142] In general, much more sophisticated approaches for separating information into levels can be applied to this process and, in particular, “wavelet” or other multi-resolution techniques, or related mathematical procedures, may prove useful for this purpose.

[0143] 2. Similarly, vector graphics or motion sequences can be “thinned” and parsed into pyramidal levels of detail. In general, large (low-spatial frequency) structures will be sent first with smaller details sent later.

[0144] 3. Audio signals can be pyramidally parsed by: sending sub-sets of samples; sending sub-sets of bits for given samples; sending samples from only 1 of several tracks or channels (again, see Inventor's prior inventions) which may involve stereo, separate languages, etc.; fractal or other compression can be used; reduced phoneme codec-oriented compression might be used. Further, only part of a message content may be sent at each level, for example every other word. Although that particular option would probably be less useful, the techniques described for text can be applied to audio content.

[0145] 4. For text sent as ASCII or other coded form, rather than as an image of text, the following techniques can be applied.

[0146] First, fonts, drop shadows, colors and other graphic or formatting details can be omitted for a first pass; although, the amount this will reduce bandwidth requirements may not be as significant as with other techniques described herein.

[0147] However, a more important pyramidization of text will consist of ordering document elements in a pyramidal fashion, and transmitting them with sufficient placement information to reconstruct the entire document in sequence once elements from several layers are available. For example, the TITLE, ABSTRACT, TABLE OF CONTENTS, INDEX and CHAPTER HEADINGS can be sent in

that order. This might be followed by the first paragraph of each chapter; then the first sentence of each paragraph; the large words of sentences first (followed by the context and connector words); etc.

[0148] What is described, are a number of simple and straightforward concepts which may be combined and expanded upon. In addition, AI and search style techniques can be used to search the text and look for (likely) relevant portions, chapters or paragraphs. These would then be the only ones sent, the ones sent first, or the ones weighted more heavily in an interleaving or emulsification scheme. And, they can also be flagged as "of potential interest" to the user.

[0149] 5. For W3 or other multimedia pages or documents, graphics can be omitted at a first level (and then sent in a pyramidal fashion, perhaps only when specifically requested (e.g., clicked on) or the document becomes active, i.e., is actually being viewed by the user) being replaced by a text element that says GRAPHIC, PICTURE, BUTTON, etc. In general, a complex scheme for ordering and weighting the various elements of a multimedia document will be possible. And, the user's preferences will be specified in one (or more, for different work purposes or moods) profile as described above.

[0150] Combining These Components:

[0151] The three inventive components so far described—Emulsification (Interleaving) of BLOBS; Probabilistic Predictive Weighting; and, Pyramidal Ordering of Information Elements—can be combined to achieve a technique for increasing the efficiency of end-user work and end-user communication line usage.

[0152] A list of potential "next needed" files, documents or other information resources is assembled—by tracing links, or other references in prior documents; or, by searching directories, abstracts or actual documents.

[0153] This list of potential files, documents, pages or other resources available on the net, is ordered by importance and the probability that each will be the next resource the user requests to see. These measures of the importance and likelihood of a document are combined to form a weight which modulates the size and/or frequency of portions of these documents which are then to be interleaved and transmitted over a communications line to the user.

[0154] Once a specific next document is selected as the current document, its weight is increased (perhaps monopolizing the communication line) until it (or at least the selected portion(s)) are delivered. Further, once a new document (document portion) is selected and made active, new potential documents are added to the list and the probabilities of all potential documents are revised.

[0155] Thus, the expectation that as much as possible of the resource most likely to be needed is maximized.

[0156] The pyramidal structure of the various documents guarantees (or at least increases the likelihood) that, although the next needed document has only been partially delivered, the content of the document will still be (at least to some extent) usable and understandable by the user.

[0157] All of these techniques would be fine-tuned depending upon the bandwidth of the available communica-

cations channel. For clarity, two narrative examples follow which exhibit, at least, simple versions of the techniques described abstractly above.

[0158] Patent Document Search (Narrative Example):

[0159] Assume, for the sake of discussion, a patent search is being conducted for an invention in the area of using computer simulation to design kitchens. A first patent number is specified and the document is downloaded. While it is being read by the searcher, abstracts of the patents cited in that first patent are downloaded in an interleaved manner; and, after that, abstracts of secondary references and full-text of primary references are downloaded in an interleaved manner.

[0160] After digesting the first document, the searcher chooses only two of the primary references for full text search. The secondary referenced documents, descending from those two primary references chosen, are then weighted more heavily in terms of interleaving. And, abstracts of tertiary references descending the two secondary documents are downloaded with relatively heavier weights compared to abstracts of other tertiary references. This followed by full-text of the high-priority tertiary references. And, so on.

[0161] As soon as each element is downloaded (or, perhaps, as the text comes in from the communications line) a search of its content is performed concurrently with additional downloading.

[0162] While looking for significant words (e.g., ignoring small and connecting words—the, and, but, however, etc.; ignoring terms of art—embodiment, system, invention, etc.) in those documents actually reviewed by the searcher, the system notes that the two-word sequence "virtual reality" was mentioned repeatedly in 60% of those references. In looking at the text of abstracts of quaternary references as they are downloaded, the system notes that three of them mention that term, "virtual reality", and those that do are weighted more heavily for having their full-text and referenced documents downloaded.

[0163] Further, it is noted that three of the patent documents not yet reviewed are cited in many of the documents already looked at or in documents anticipated as having a high probability of being looked at. The weights of those documents are also increased.

[0164] When a document is actually looked at, the amount of review time is monitored by the system. Also, the searcher can tag the document as: highly useful, useful, not very useful, not at all useful; or some other similar scale. The weight of that document, and all weight decisions relating to that document are appropriately adjusted up or down as a result.

[0165] Further, a set of tools will be made available to the searcher to highlight words, terms, phrases, whole paragraphs or arbitrary sections of text, and references, as being of particular interest; or of non-interest/anti-interest in order to downplay connections that might be made on a statistical/linguistic basis but which are, in fact, not relevant.

[0166] Research Document Review (Narrative Example):

[0167] During the review of a pyramided downloaded research paper, the TITLE, ABSTRACT and SECTION

HEADINGS are downloaded first. The headings are Introduction, Methods, Experiment 1, Experiment 2, Experiment 3, Experiment 4, Discussion of Results, and Conclusion; not terribly informative. While the abstract is being perused by reviewer, the first paragraph of each section heading is downloaded.

[0168] The first paragraph of Experiment 1 reads, in part, that "... was administered to twenty-seven laboratory mice ...". Now, our hypothetical researcher is interested in reviewing results of administering a particular class of drugs to primates. He would highlight (or from a menu choose) Section 1 and indicate it as "not of interest"; similarly he would highlight the word mice and "dis-emphasize" its interest level.

[0169] The first paragraph of Experiment 2 reads, in part, that "... was administered to seven chimpanzees ...". Our user would first highlight (or from a menu choose) Section 2 and indicate it as "of primary interest"; similarly he would highlight the word chimpanzee and "emphasize" its interest level.

[0170] Immediately the system would download Section 2 at full-speed (if it were not yet there) for reviewer to look at; this section is both the current active section of the current active document, but has also been explicitly "emphasized" as of "interest".

[0171] Further, the system would search the additional sections and would: download the Experiment 3 section which mentioned chimpanzee 22 times and mouse twice; fail to download the Experiment 4 section which, although it mentioned chimpanzee once, mentioned the words mouse (a form of mice—determined from a look-up dictionary of related forms, synonyms, antonyms, etc.) and mice 32 times. However the sentence that contained the word chimpanzee (as well as the previous and following sentence) would be downloaded and the word chimpanzee would be highlighted; and download the last two sections (but at lower priority) which mentioned chimpanzee and mouse each a few times. The occurrences of chimpanzee would be highlighted (unless reviewer's preference profile indicated not to, which would be an option to those who found such "system-supplied" hints annoying).

[0172] Information Impedance Matching:

[0173] As described above, the three inventive components disclosed herein are combined to increase the efficiency of work done over a communications network. The various operating parameters of the composite technique would be fine-tuned depending upon the bandwidth of the available communications channel.

[0174] The technique can be applied within a single system, to two communications lines of different bandwidths, to implement a method whereby an end-user, working over a low-bandwidth communication line will have the perception of working over a much higher-bandwidth communication line. Apparent performance and response time is improved. See FIG. 1.

[0175] The interaction between a large time-sharing computer (serving multiple customers) and the global outside network will be via a (or many) high-bandwidth, high-speed communications line(s). The information resources accessed by that "server" will be from diverse sources and will be of

diverse type, content, composition, form and format. Thus, the techniques described herein will not necessarily be practical to apply in all instances.

[0176] In particular, the Probabilistic Prediction of documents—the importance and likelihood of need for them—can be done. Linkages and references from current documents can be traced down several levels. Those documents (once accessed) can be further searched for clues as to their potential relevance to the user during the current work session, work in general, general interest, or for clues as to further documents to access.

[0177] Thus, the server computer will be acting as an "intelligent agent" for the user, anticipating his needs and searching out potential relevant information.

[0178] However, the information may have to be downloaded—as far as the server—in fairly standard fashion. That is because the various information resources being accessed will not necessarily be organized in a pyramidal fashion; and, coming from diverse sources, may not be interleavable, let alone interleaved with weighting factors.

[0179] However, once these diverse information resources are assembled in the server, all of the techniques described herein can be used, when downloading from the server to the end user's client computer over the low-bandwidth communication line.

[0180] One of the major tasks of the server will (can) be to take the various resources that are not in pyramidal format and parse them for pyramidal transmission. This will be done by the various techniques described above, and the process will be fine-tuned based on the user profile, and the particulars of the current work session, as well as an analysis of the document in question and the specifications of the available communications line.

[0181] Overall, this process will likely not diminish (and may even greatly increase) the amount of network traffic done on behalf of a particular user during a particular work session. Much of the information downloaded to the sever in anticipation, may never need to be downloaded as far as the user's client system; much of it may never be seen by the user. However, much more efficient (continuous) use will be made of the communication line between the server and the end user's client system. See FIG. 11.

[0182] The result will be an apparent increase in the performance and responsiveness of the overall system to the user's needs.

[0183] The task of the large server computer will be to match the high-bandwidth traffic available to it for downloading from the world-wide network with the low-bandwidth channel available to it for uploading to the end user. It must do this in a way that will maintain a maximum of intelligibility of the information being provided to the end user.

[0184] In a conceptually similar situation, an electronic signal traveling down a high-impedance (resistance) line must be processed in order to interface it to a low-impedance line so that the signal does not become distorted. The device used to match the characteristics of the two lines is called an impedance matching transformer.

[0185] Impedance (resistance) of a wire through which electricity flows, is the electronic equivalent of the radius or

bore of a pipe through which water flows, is the equivalent of the bandwidth of a channel through which information flows. Hence, Inventor refers to the process performed by the server computer, in matching the performance of the high-bandwidth input line with the low-bandwidth output line, while maintaining understandability (i.e., without too much distortion of the meaning) of the information, as "information impedance matching".

[0186] A commercial name for such a system might be the INFORMATION FUNNEL™.

[0187] The Funnel™—an Extension of Emulsification (Narrative Description):

[0188] You sit at a desktop system, connected via telephone line and modem to an account on the Funnel. The Funnel is also the site for the TOURING PAGE™ or YELLOW PAGE™ (see next section for details). From your account at the Funnel an ultra-fast/wide "spigot" to the web/net is available to an "intelligent software agent" who downloads its "personality" from your local profile file. This agent anticipates, probabilistically, your demands from the web/net and downloads to its own Funnel site, in a probabilistic emulsified (when possible), fashion. The agent, in turn, downloads to you at your local desktop system in a probabilistic emulsified fashion (definitely), stripping BLOBS, and pyramidizing them.

[0189] For example, if you were looking at a page from which you may go to any of five other pages, pyramidized tips of those five pages will be sent to your local machine while you are reading the original page and deciding what to do next. The size (detail level) of each pyramid tip can be adjusted, based upon estimated probabilities; i.e., likelihoods of your desiring to access that page, based on complex and sophisticated functions matching each page's content with goals and statistics describing your current work session and your general work profile. The process can be carried out several levels, downloading the pages/documents linked to or referenced by the five first-level referenced pages, etc.

[0190] Stripping a document or page (only one level is described) follows: pictures can be replaced by a box that says "pix" or a small, low-res version of the graphic; whole text can be replaced by some combination of headlines, first sentence of paragraph, first few words of sentence, first line, key words, etc.; and, fonts and other formatting "prettification" can be stripped out. Similarly, buttons on other graphics can be simplified. In short: translation of the page description to a restricted sub-set description language; data compression; and, knowledge that a smart, sophisticated, "fill in the blanks" display driver is at your local machine—these will permit a faster delivery of a low-information-density version of the document quickly, with later detail to follow as/if needed. This, in turn, will permit you to have (apparently) much faster access and/or browsing capabilities. When a higher level of interest in a particular document (or document section) is probabilistically determined, progressively more detail will be downloaded to your system (or the weight for that item will be increased). Actually looking at a document (or section of document) is the highest level of interest and, if all detail is not yet available, a displayed page will take on detail and completeness as you view it.

[0191] When possible, the section of document being viewed would have highest priority within the selected

document. That is to say, if a large (compared to the size of the screen) document is being viewed, the position of the currently viewed portion of the document would be fed back at least as far as the Funnel server. If the W3 site feeding the Funnel were capable of the behavior described in this paragraph, the current window position would be fed back to the W3 site as well. In either or both cases, the document pyramidization would be (re-)structured so that information describing the details of the currently viewed section would be sent before (or, at least with higher weighting or priority) information describing the other parts of the document.

[0192] The Yellow Page™ and Network Caching™:

[0193] Briefly, the INFORMATION FUNNEL, as described herein, will be more useful and effective if combined with two other system components which will be described briefly below. A very large computer service is envisioned; a large customer base will make the entire operation more effective and efficient.

[0194] The first of the two other system components is a directory or yellow pages for the Internet, W3, etc. A number of printed and electronic directories are available which are intended as directories of the network. What is envisioned here is a comprehensive list of sites on the network cross-indexed to a large number of categories and key-words which will facilitate user-directed searches as well as facilitate the AI or intelligent agent functions described above. This database can be compiled by automated searching and analysis of the information resources available at the various network sites.

[0195] In addition, various commercial or other sites might be invited to provide "listings". And, further, local (to this system) copies of other's catalogs and other W3 resources might be kept for downloading to customer's client systems. These local copies may be useful because (at least for often accessed resources) communication requirements would be reduced and response time reduced if copies were already available at the server system.

[0196] The system would, when a request for such an item were made, see if a local copy were available and, only if not, get it from the network. Further, even if a local copy were available, a request to the remote site might be made to ask if the resource had been modified since last retrieved. If it had been updated, a new copy (of at least the modified part) would be downloaded. For catalogs and such, which were intentionally deposited by others at the sever site, it may be the responsibility of the depositor to update the local copy.

[0197] Such an Internet/W3 directory might be called THE YELLOW PAGE™ a hybrid of HOME PAGE—the root entry for a W3 site—and the YELLOW PAGES telephone directory. Alternately, it might be called THE TOURING PAGE™. An additional list of potential trademarks is listed at the end of this document.

[0198] The second system component Inventor calls NETWORK CACHING™. As explained above, a large service and system is envisioned with a very large customer base. Many customers will be accessing a portion of the same information as other customers. Thus, in order to reduce the need for communication time and bandwidth, copies of network information can be kept at the server system in anticipation of later customers requiring the same informa-

tion. Again: the system would, when a request for such and item were made, see if a local copy were available and, only if not, get it from the network. Further, even if a local copy were available a request to the remote site might be made to ask if the resource had been modified since last retrieved. If it had been updated a new copy (of at least the modified part) would be downloaded.

[0199] Which items to keep locally, and which to discard immediately, would be made by the same kind of probabilistic analyses as described above; in particular, by keeping cross-statistics between the interests and behavior of the many customers on the system. Home pages for many types of sites (which, generally, every access passes through) would be given additional weight. Further, if an information resource was particularly hard to get again (e.g.: it is large; it comes over a low-bandwidth, slow, hard to reach (often busy) or unreliable communications line; etc.) it would also be given additional weight as far as deciding whether to keep it around. Items often requested (in general) would be weighted more heavily. Items often requested (in the short term) would be weighted more heavily.

[0200] One user can even be notified when a second user has downloaded information that might be interesting to the first. Such a determination can be made based on the content of the document, and/or on statistics about the interests (and shared interests) of the two customers.

[0201] An item might be kept locally for a set period and then discarded. In that way, often accessed information would stay around pretty much continuously. A FIFO policy can otherwise be applied as the information cache becomes filled.

#### [0202] An Alternate Embodiment:

[0203] The group of embodiments, described above, employ—in addition to the more standard communications network itself—a computer system running non-standard information processing software (i.e., the Funnel) that is interposed between the information source systems available on the W3 (or elsewhere) and the end user's computer system.

[0204] An alternate embodiment would comprise document source (server) and document browsing (client) systems that each operated with communications software that had been augmented to include the techniques described above. See FIG. 2.

[0205] In this case the same basic techniques can be applied, that is: interleaving of multiple documents or document parts (e.g., web pages); probabilistic weighting of the size and/or frequency of interleaved document portions; “pyramidization” of documents so that documents that have only been partially delivered are still useful to review.

[0206] However, there will be differences in that the software running on the server site will, in general, be more familiar with its own source information, and less familiar with the user's statistical history profile than as described above for the Funnel embodiment. In particular, this will mostly effect the determination of the probabilities used to calculate the emulsification (interleaving) weights. A narrative example will illustrate.

[0207] On-Line Catalog (Narrative Example):

[0208] The following narrative example will describe an on-line catalog with content typical of a department store.

[0209] A naive predictive or probabilistic approach will be described first, and this may be the mode of the earliest interaction with a catalog system, even if that system uses more sophisticated prediction in later stages of interaction. Reference to FIG. 4 will be made during this discussion.

[0210] One or more preliminary introductory pages may come first, analogous to: a catalog cover; a logon process that can include user identification, method of payment, (intent description, questionnaire or profile downloading as described later), etc.; a list of last minute changes or “special” sales; etc. After that true interactivity will begin.

[0211] What comes next will be what might be referred to as a root node, home page, top interactive level, or table of contents (TOC) page; call this Level 0. Its contents (besides any prettifying graphics or icons) might be as simple as a list of “departments” While the user is browsing this storewide directory, unbidden the system will download a series of Level 1 department (e.g., men's clothing, books, housewares, lingerie) directories. Since the machine knows nothing about the user, they will all be downloaded with equal probability. However, the text headlines (pyramid tip) will be sent first from all pages; then more text from all pages interleaved; finally icon pictures. The headlines will be available to look at even before the full text or icons get there. Additionally, the hyperlinks associated with the icons will be hot, just without pictures.

[0212] After the level 1 pages get there level two pages (groupings within a department, e.g., coats, pants, small appliances, dishes, etc.) will be sent in a similarly interleaved fashion; all of them. However, as soon as the user clicks on a level one department link and gets that departmental level one page displayed (if and when this information can be fed back to the catalog server) the level 2 pages associated with the level 1 page being displayed will get top priority as the expense of other level 2 pages. Similarly, the level three pages (single items) associated with the current level one page will be downloaded in a similar equal probability fashion until a specific level two page is selected. Etc. As the user traverses the nodes of the catalog tree, the items on the next level, attached to the current node will always be sent unbidden. Of course, a specific request from the user will override this barrage of guessed data.

[0213] A more intelligent guessing process may be facilitated by a questionnaire in a early page which could determine for example: the user is a 55 year old male; shopping for his teenaged daughter's birthday; wants to spend about 50\$ to 75\$; characterizes her taste as “unknown”. In that case, before the user even selects a level one page, level 3 pages containing sweaters, jeans, CDs and rock concert tickets will already be downloaded.

[0214] Similarly, (perhaps via a published standard, or by registering with a store) a user's preference profile may be downloaded (or that of a person that is being shopped for) and size, color and “likes” may be codified for the system to match with items of likely interest.

[0215] Additionally, that information can be compiled from previous purchases at that store or generally (perhaps by a credit card company) although the legality and advisability of collecting such data is not certain.

[0216] Similarly, “suggestions” may be made by the systems of recently “popular items” and this may be done on an “since you liked (i.e., bought) this item you may like this other item which was bought by many people who bought the first item”.

[0217] One particularly nice feature for such a catalog (or more generally the web) will be in the web browser display a tree-like structure of the entire catalog indicating which node is the current node and the state (already downloaded, in transit, not requested, etc.) of other nodes. See FIG. 13.

#### NEW PROTOCOL

[0218] The next subject will be how to provide communications facilities that will facilitate such probabilistic, pyramidal, interleaving.

[0219] Preliminaries:

[0220] The first thing to note is that there are at least three distinct ways to implement the basic invention utilizing internet-like communication protocols. These are: micro-protocol, macro-protocol and augmented protocol and will be described below.

[0221] Additionally for most optimal results, all computer on the network should be aware of the new standards. In that case TCP/IP-like facilities will be provided.

[0222] However, in the absence of that, specific pairs of servers and (populations of) users may with foreknowledge use a not-widely-adapted private standard. Or, as in the catalog system described previously, a provider may (in a minimal fashion) provide such probabilistic interleaving services. However, one cannot ask for probabilistic interleaving if the server has no such facilities. In that case a meta-protocol will piggyback on (or below) TCP/IP like standards.

[0223] Similarly, if a service provider and a population of end-users are that pair, the Funnel service may be implemented. In this case the service provider will contact the network with standard TCP/IP-like facilities. However, in this case a new version of [C]SLIP/PPP dial-up protocol would be used.

[0224] In all cases the intention will be to establish multi-thread communication.

[0225] Micro-Protocol:

[0226] In this case standard TCP/IP-like packets will be received over a single socket or port. But within that packet will be a number of sub-packets, each belonging to a sub-connection. These may then be assigned in different sizes or at different rates in order to affect priorities for different documents, document levels, document types, document part (MIME) types, etc.

[0227] In this case the inner packet information would be split up into individual sub-packets. Each would have a header containing, at least, what sub-connection it belonged to. Alternately, the whole packet might start with a “map” of the packet’s contents.

[0228] See FIG. 14 for a diagram of such a data structure.

[0229] Macro-Protocol:

[0230] In this case standard TCP/IP-like packets will again be used. However, in this case a multiplicity of sockets, ports or connections will be established. In this way, different document streams may be distinguished by referring to the logical line that they are being received on. See FIG. 15 for a diagram of such a data structure.

[0231] Augmented Protocol:

[0232] In this case an additional layer of protocol will be established as an inside wrapper underneath the IP address and the TCP packet structure.

[0233] Here there can be a single socket per client/server pair (unlike the macro-protocol case) and there will be a one-to-one correspondence between logical and actual packets (unlike the micro-protocol case). See FIG. 16 for a diagram of such a data structure.

[0234] The inside wrapper will, at a minimum, have a designation of which sub-connection the packet is coming in on, so that the document or document part can be identified.

[0235] Inventor has adopted the terminology of CABLE and FIBER to indicate that a multi-thread connection is established with sub-connection signal capability.

[0236] The following table indicates a list of some of the inner layer packet header designations that can be implemented with this approach.

DESIGNATION NAME	EQUIVALENT OR USE
CABLE NUMBER	SOCKET or PORT NUMBER
FIBER NUMBER	INDIVIDUAL SUB-CONNECTION
DOCUMENT NAME	USEFUL TO DOUBLE UP IF
DOCUMENT PART NAME	RUN OUT OF FIBERS
DOCUMENT TYPE	MIME-LIKE OR MAY INDICATE ITEMS LIKE CATALOG OR REVIEW
DOCUMENT PART TYPE	AS IN HEADINGS, ABSTRACT, ETC.
DOCUMENT LEVEL REQUESTED	AS PER PYRAMIDIZATION
PRIORITY	AS PER USER REQUEST
SENDING PRIORITY	SAY, FROM CATALOG SYSTEM

[0237] With priority information embedded in the packet, if system-wide usage is adopted then every node on the network may act as a funnel. That is not to say that my high-priority request will be served before your low-priority request; that would defeat the purpose, everybody would ask for everything as high priority. However, for a particular user or site, deliveries may be prioritized.

[0238] In that case, however, low-priority packets may not be allowed to pile up. So not only is labeled priority a factor, but how long has a packet waited, and how big a backlog for that user or site is there.

[0239] However, in general, the overall network may be sped up if generally, people want to get their web-page text fast, make a decision, and wait for pictures only if this looks interesting. If most transactions were so prioritized that overall traffic of text would improve; and, then adopting such a policy would be effective for individuals.

**[0240]** In order to make this particularly effective, once you kill a page (decide it is not of interest) that information would be sent back to the server in order to terminate the transmission. With a single connection that is easy, just kill the connection. However, with a cable, individual fibers will have to be activated and deactivated by request.

**[0241]** Additionally, when requesting for an information resource or, more particularly a multiplicity of them to be interleaved, each requested resource will have a requesting priority associated with it. a mechanism will be provided in the protocol to change the priority of any in progress transmission including, specifically, the ability to: kill the resource; suspend transmission of the resource but do not kill it; make this resource, temporarily until told otherwise, the selected resource and devote all possible bandwidth to it at the expense off all other requests on this cable. Application programs will be able to make great use of such facilities by upping the priority on selected items in a hyper-media session.

**[0242]** Other requests will be built into this protocol. One is validation. Particularly for web-caching, but for other purposes, one may instead of requesting a resource, request to validate a resource for a particular date/time or version number. What will be sent will then be an OK (indicating that the document is current) or the recent document itself. In addition, for applications that are mutually aware of this, what may be sent is not the entire document but, rather, instructions to update a small portion of the document. Such delta-modulated versions can be kept for each revision and, by comparing the requested date with the current date the appropriate set of update directions sent. Of course, the receiving client program will have to be able to accept and affect such change directions for this to work. See FIG. 17 for a diagram of how several updates may be made on this bases, or for adding pyramidal data levels.

**[0243]** In addition, when a web browser is displaying a not-yet validated cached resource, it will indicate by flashing a "not yet valid indicator" until such time as the date is confirmed or the resource updated.

**[0244]** Similarly, a "request summary" flavor of request may be adopted where in response to such a request an index, summary or card-catalog-like entry precede the transmission of the actual resource. Options here will permit the resource to follow: unless otherwise killed; not until interest is confirmed; automatically sent (or killed) after a delay to permit user decision.

#### AVAILABILITY OF ELEMENTS OF TECHNOLOGY

**[0245]** It should be noted that the vast majority of individual elements of the foregoing invention are available "off-the-shelf" as hardware and software components, and are easily integrateable by one of ordinary skill in computer systems programming, integration and administration.

**[0246]** Specifically, computer add-in hardware boards are available to provide FAX/Modem capabilities, and "voice mail" capabilities (i.e., voice recording and playback, sensing of DTMF keypad entry, and reading of embedded DID and Caller ID signals). These components are available from many vendors and many developer-level software routines and whole application programs are available to operate

such boards. These software routines can be coordinated and made to communicate and exchange information under operating systems such as UNIX, DOS, Windows, or System 7.

**[0247]** Similarly, the communication, EMailing and other software functions that make up "The Internet" are widely known and available and their use and adaption are well within the ken of one of ordinary skill in computer network administration.

**[0248]** Thus, the various technological elements that comprise the invention disclosed herein are (except as described otherwise) standard, well-known, widely available elements. Their construction, operation, use and integration are not, in general, the subject of the instant invention. For the details of the construction, operation, use and integration of these components the practitioner is directed to the available published literature, as well as software and equipment manuals, in the areas of hardware, software and information formats relating to, without limitation: FAX machines, FAX/Modems, FAX servers, voice mail devices, barcode printing, barcode scanning, optical character recognition, computer communication and networking (including, specifically, EMail, The Internet, and related communication and information exchange protocols such as HTTP, UDP, TCP/IP, etc.) and computers and operating systems.

**[0249]** Many of the elements of the present invention are available "off-the-shelf" and, in and of themselves, not inventive. However, the design, construction and operation of the composite systems and techniques disclosed herein, as well as the use to which they are put, are the subject of the instant invention and are presented as novel and useful. These novel, composite, high-level system designs are described herein and depicted in the accompanying system and flow diagrams.

**[0250]** Finally, it is noted that the capabilities and system components described herein may be provided by equivalently functioning substitutes. For example: signals and hardware may be of analog or digital construction; certain functions may be exist as special purpose hardware, or as programming residing on general purpose hardware; software may be provided on PROMs, or be stored in memory or on magnetic or optical disk; various capabilities may, in different situations, reside in a FAX machine, FAX/Modem in a computer, in a local computer, or in a remote computer; many options exist for FAX-to-FAX, FAX-to-computer or computer-to-computer interconnection; etc. The use of any appropriate system component, now in use or later developed, to affect the methods and systems described herein, is considered to be within the scope of the instant invention.

**[0251]** The flows depicted in the software flow diagrams herein are exemplary, some items may be ordered differently, combined in a single step, skipped entirely, or accomplished in a different manner. However, the depicted flows will work. In particular, some of these functions may be carried out by hardware components, or by software routines residing on, or supplied with, such a component.

**[0252]** Similarly the systems depicted in the system diagrams herein are exemplary, some items may be organized differently, combined in a single element, split into multiple elements, omitted entirely, or organized in a different manner. However, the depicted systems will work. In particular,

some of these functions may be carried out by hardware components, or by software routines residing on, or supplied with, such a component.

[0253] It will thus be seen that the objects set forth above, among those made apparent from the preceding description, are efficiently attained and certain changes may be made in carrying out the above method and in the construction set forth. Accordingly, it is intended that all matter contained in the above description or shown in the accompanying figures shall be interpreted as illustrative and not in a limiting sense.

Now that the invention has been described, what is claimed as new and desired to be secured by Letters Patent is:

1. A method for intelligent assistance from a computer program comprising:

- a. retention of session history and resources;
- b. selection of optimization criteria;
- c. sensitivity analysis of parameters;
- d. multiple simulation with different sensitivity settings;
- e. retention of optimal parameters;
- f. matching later session with features of stored prior sessions;
- g. using optimal simulated parameters from matched prior session with current session.

2. A system for computer communication comprising a client, service provider and network at large whereby:

- a. client makes multiple prioritized requests of service provider;

b. service provider makes multiple requests to network at large on behalf of client without prioritizing;

c. network responds with information in an arbitrary manner;

d. service provider prioritizes returned data prior to passing it on to client.

3. A method of data communication whereby multiple information resources are transmitted over a single channel in a prioritized manner such that larger or more frequent packets are sent to higher priority resources.

4. A method of data communication whereby within a packet switching network packet and additional multi-thread packet header is insinuated between the network header and the user data.

5. An improved method for data communication, via a packet-switching digital communications network, utilizing a communications protocol wherein the normal mode of operation is to transmit each data file as a data stream of packets separate from the data streams of other data files, wherein the improvement comprises the transmission of a multiplicity of data files via a single data stream connection whereby, within a packet switching network packet, an additional multi-thread packet header is insinuated between the network header and the user data.

6. A method as in claim 5, wherein said data files comprise HTTP document elements.

7. A method as in claim 5, wherein said data packets are divided into sub-packets and said multi-thread packet header specifies an association between said data files and said sub-packets.

\* \* \* \* \*