(19) **United States**

(12) **Patent Application Publication** (10) Pub. No.: **US 2012/0016820 A1**

Walker (43) Pub. Date: **Jan. 19, 2012**

(54) **STOCHASTIC SEARCH STRATEGIES FOR MULTIMEDIA RESOURCE DISCOVERY AND RETRIEVAL SYSTEM**

(75) Inventor: **Reginald L. Walker**, Los Angeles, CA (US)

(73) Assignee: **Tapicu, Inc.**, Los Angeles, CA (US)

(52) **U.S. Cl.** **706/12**
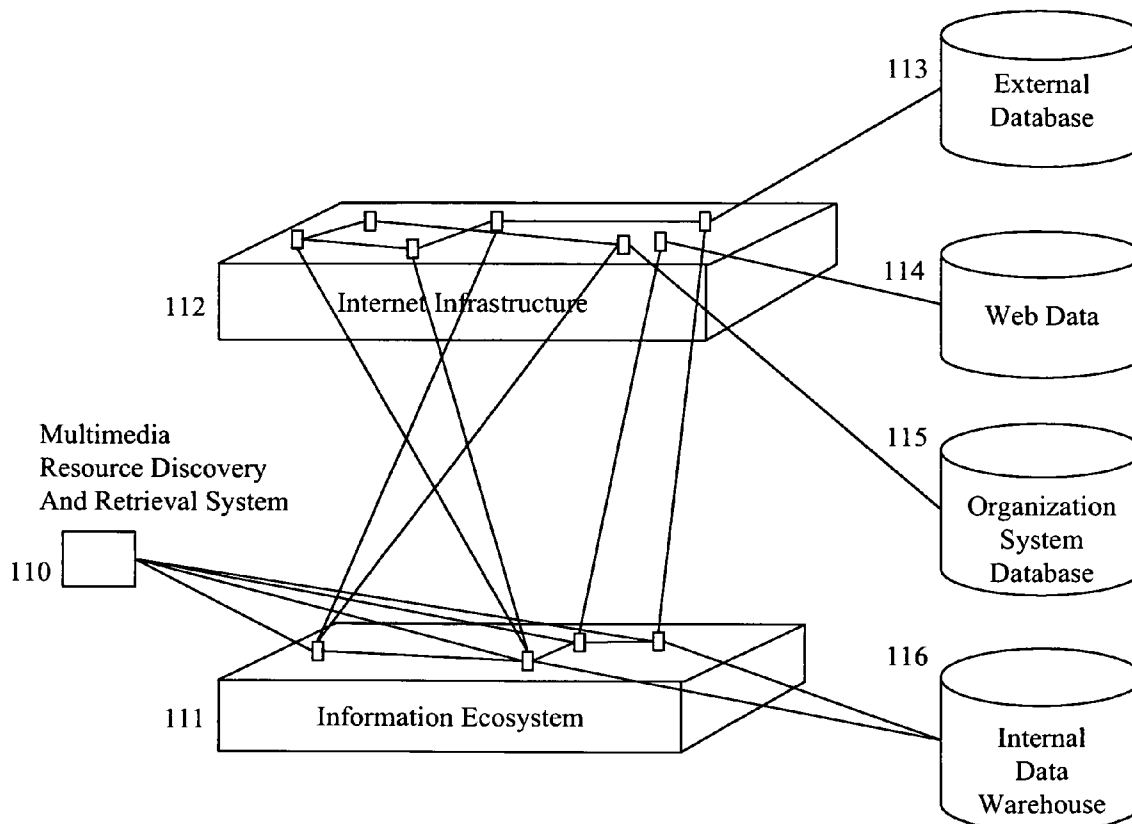
(57) **ABSTRACT**

A method is described for applying distributed stochastic optimization techniques of evolutionary computation using a plurality of servers and a plurality of clients machines being connected via a computer network such as the Internet. The stochastic optimization techniques of evolutionary computation seek to optimize a populations of individuals against one or more predetermined fitness criteria when applied to solving solve the network routing problem coupled with one or more information retrieval problems. The field of evolutionary computation encompasses stochastic optimization techniques, such as randomized search strategies, in the form of evolutionary strategies (ES), evolutionary programming (EP), genetic algorithms (GA), classifier systems, evolvable hardware (EHW), and genetic programming (GP). The stochastic optimization component objectives of the multimedia resource discovery and retrieval systems includes maximization of resource utilization and of overall LAN throughput.

113 External Database

112 Internet Infrastructure

114 Web Data

Multimedia
Resource Discovery
And Retrieval System

115 Organization System Database

110

111 Information Ecosystem

116 Internal Data Warehouse

FIG. 1

221 Stochastic Optimization Agent

222 Stochastic Optimization Monitor Agent

231 Messaging Passing Mechanism

223 Stochastic Optimization File Server Regulatory Agent

224 Stochastic Optimization Probe Regulatory Agent

225 Stochastic Optimization Scout Regulatory Agent

226 Stochastic Optimization Forager Regulatory Agent

231 Messaging Passing Mechanism

227 Stochastic Optimization File Server Further Agent

228 Stochastic Optimization Probe Further Agent

229 Stochastic Optimization Scout Further Agent

230 Stochastic Optimization Forager Further Agent
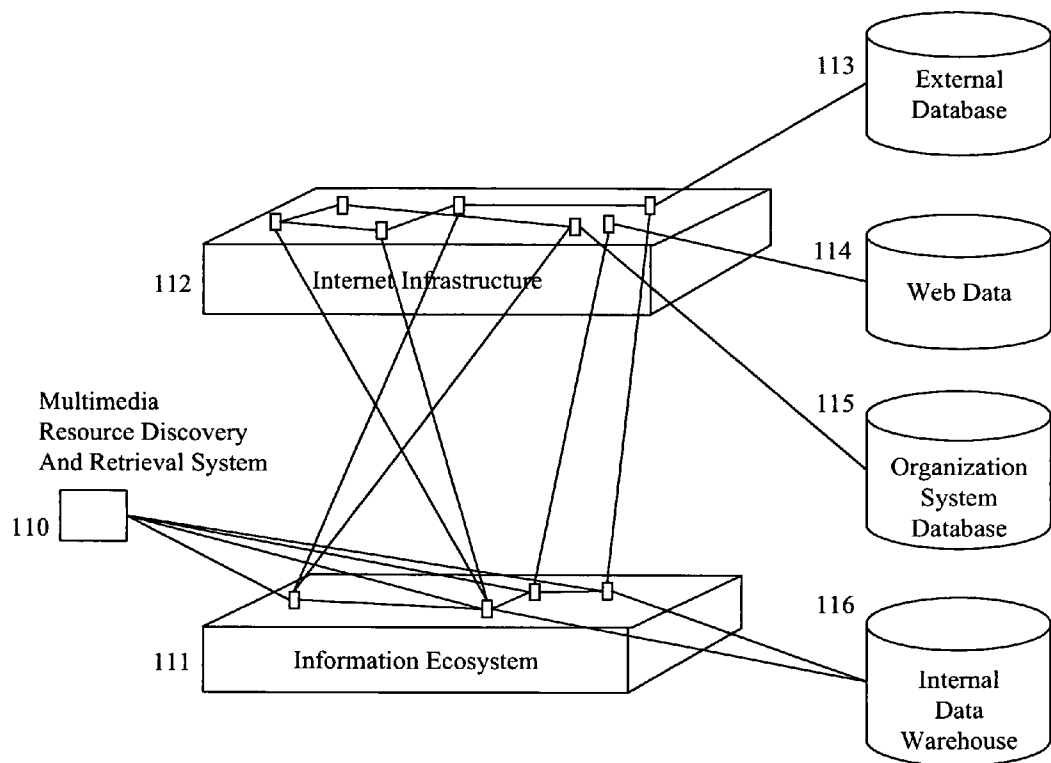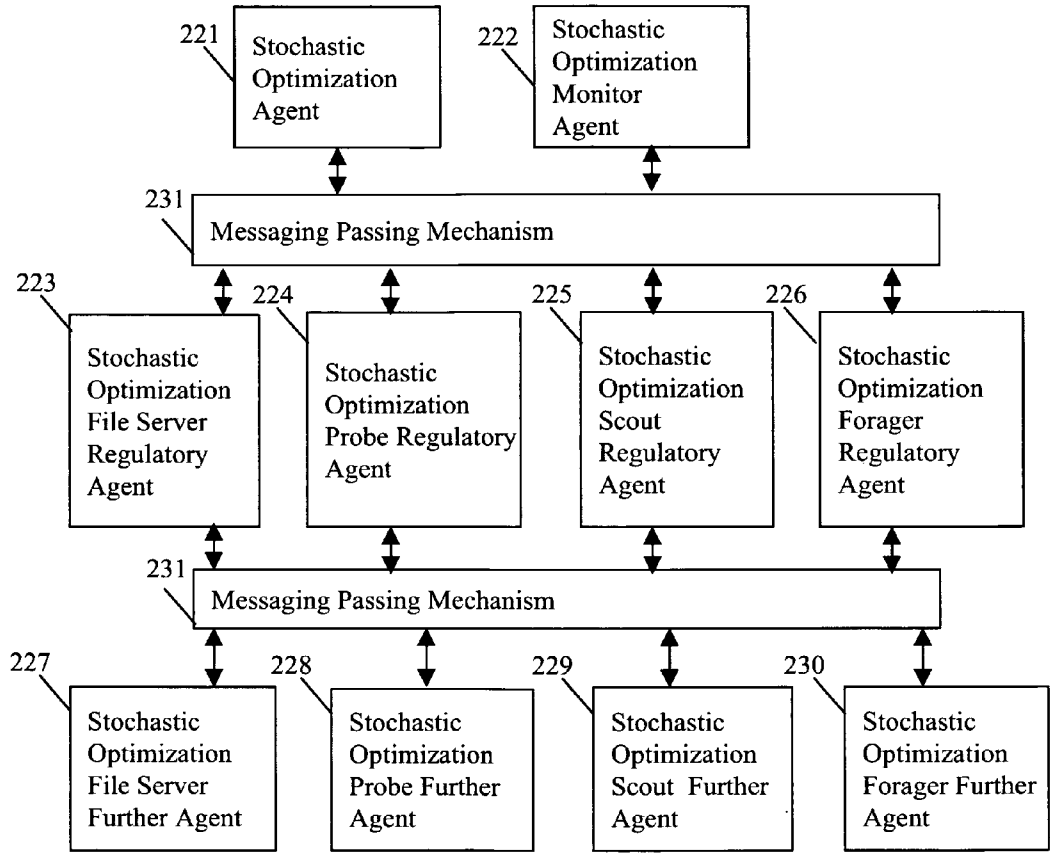
FIG. 2

STOCHASTIC SEARCH STRATEGIES FOR MULTIMEDIA RESOURCE DISCOVERY AND RETRIEVAL SYSTEM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001]   This application claims benefit of provisional application Ser. No. 61/399,937, filed on Jul. 19, 2010 by the present inventor.

FEDERALLY SPONSORED RESEARCH

[0002]   Not Applicable

SEQUENCE LISTING OR PROGRAM

[0003]   Not Applicable

BACKGROUND OF THE INVENTION

[0004]   The present invention relates to the stochastic optimization of evolutionary computation (EC) search strategy parameters in a distributed fashion across a plurality of servers and a plurality of clients which are connected together by a network, such as the Internet. In particular, it relates to stochastic optimization search strategies for a multimedia resource discovery and retrieval system for computer information sharing systems such as search engines, data warehouses, and service oriented architectures (SOAs). The multimedia resource discovery and retrieval system uses a plurality of stochastic optimization agents to adaptively solve the network routing problem coupled with one or more information retrieval (IR) problems. The field of evolutionary computation encompasses stochastic optimization techniques, such as randomized search strategies, in the form of evolutionary strategies (ES), evolutionary programming (EP), genetic algorithms (GA), classifier systems, evolvable hardware (EHW), and genetic programming (GP).

[0005]   There has always been a need to iteratively improve the retrieval of remotely located multimedia files (images, text, audio, and video). The stochastic optimization techniques of evolutionary computation (EC) contain mechanisms which enable the representation of certain unique aspects of individual (agent) behavior to adaptively improve multimedia resource discovery and retrieval mechanisms. Principles of the stochastic optimization techniques of EC can be found for example in Walker (2003) "*Tocorime Apicu: Design of an Experimental Search Engine Using an Information Sharing Model*", University of California Dissertation, UMI Dissertation Publishing, Ann Arbor, Mich. 48106-1346 (www.proquest.com) or rwalker@tapicu.com, which is incorporated by reference herein in its entirety.

[0006]   The chief differences among the various types of EC stem from: 1) the representation of solutions (known as individuals in EC), 2) the design of the variation operators (mutation and/or recombination—also known as crossover), and 3) selection mechanisms. A common strength of these optimization approaches lies in the use of hybrid algorithms derived by combining one or more of the evolutionary search methodologies. The underlying optimization methodologies of EC are used to implement unique stochastic aspects of search strategies.

[0007]   A hybrid of the stochastic optimization techniques of evolutionary computation (EC) provides a computational optimization strategy for problems which are difficult to solve using conventional mathematical techniques, such as the net-

work routing problem coupled with one or more information retrieval problems. They are particularly applied to highly complex situations, where a very large number of variable parameters prevent any formal attempt at a solution. A typical example of such a problem is where a Web crawlers must visit all of the possible number of IP addresses (there are 2^{32} possible Ipv4 addresses) supporting multimedia services and retrieval all text documents—based on the class structures associated with the Ipv4 and eventually, the IPv6 protocols (2^{128} or 3.4×10^{38} possible network addresses). Class A can support up to 2^7 (approximately 128) networks, class B can support up to 2^{14} (approximately 16384) networks, and class C can support up to 2^{21} (approximately 2 million) networks. When the number of sites to be visited is very large, it is not feasible to determine a solution by an exhaustive calculation of the distance for every possible route (this is sometimes referred to as the "traveling salesman" problem).

[0008]   Information processing associated with the stochastic optimization component of multimedia resource discovery and retrieval systems is hampered by the stochastic information fluctuations that occur each second within the Internet. The multimedia resource discovery and retrieval model encompasses policies and processes that require mechanisms capable of adaptively adjusting operational parameters. These parameters are required to filter, organize, and index any large-scale data set—information stored on a single computer, a local area network (LAN), and a wide area network (WAN) that encompasses the whole Internet—that may consists of constantly fluctuating information content over relatively short periods of time. Additionally, the model (or its individual components) can be applied to hosts of existing multimedia systems that require mechanisms to locate new and updated information in a timely manner, mechanisms to integrate new information into existing knowledge bases, and software to extract, analyze, and disseminate information from large-scale data sets.

[0009]   Instead, one way of addressing network routing problems is to use stochastic optimization agents that are derive from hybrids of the stochastic optimization techniques of evolutionary computation. According to this approach, sets of different possible routes are selected, each being typically represented in practical implementations by a fixed length string, and the distance for each selected route is calculated (this is known as the "fitness"). Then, a new set of routes is generated by forming essentially random combinations of routes from the preceding set, with those routes having the shortest distances (i.e. best fitness) being preferentially selected for the reproduction of routes for the next generation. The fitness of each member of the new generation is then calculated, and the process is repeated, for example until a predetermined number of generations are reached, or until an individual having a particular fitness level has been produced. Thus the hybrid stochastic optimization techniques of evolutionary computation mimics natural evolution, in that for each generation, the least successful members are destroyed, and only the most successful members produce progeny for the next generation. In time, it is hoped that the quality of the population steadily improves, eventually leading to an optimal solution. In simplest terms, each cycle of the hybrid stochastic optimization techniques of evolutionary computation involves an evaluation phase for the current population, and then a generation phase, for producing the next generation.

[0010]   The efficiency of Internet applications is often tested by adding new applications that compete for shared network resources. Studies associated with network traffic show the need for adaptive congestion control and avoidance at the

application level. The side-effects resulting from non-adaptive applications include burstiness (degree of self-similarity) in network transmissions. New applications, such as the transmission of multimedia data coupled with current network traffic, provide a hierarchical ordering based on the priority and diversity of data transmissions.

[0011] The file types associated with multimedia Web transmissions are reduced to images, text, audio, and video. It has been shown that Internet traffic associated with multimedia (text only) file transfers introduced the least network congestion. The addition of image, audio, and/or video transmissions associated with document/file transmissions tends to cause heavy-tailed traffic distributions. The heavy-tailed distributions reflect active/inactive periods and varying sizes of multimedia files. The coupling of information organization with retrieval systems contributes to self-similarity in Internet traffic.

[0012] It is the goal of this invention to provide a system and method for reducing the computational effort to achieve periodic optimal multimedia resource discovery and retrieval from a host of diverse multimedia sources.

BACKGROUND OF THE INVENTION

Objectives

[0013] Accordingly, the objectives and advantages of the invention are as follows:

[0014] It is an objective of the present invention to use stochastic optimization agents formed from hybrid algorithms derived by combining one or more of the evolutionary computation search methodologies.

[0015] It is another objective of the present invention to encompass policies and processes that require mechanisms capable of adaptively adjusting operational parameters required to filter, organize, and index any large-scale data set—information stored on a single computer, a local area network (LAN), and a wide area network (WAN)—that may consists of constantly fluctuating information content over relatively short periods of time.

[0016] It is another objective of the present invention to represent solutions as memes to reduce in the computational effort to achieve the periodic optimal document clusters. The fitness of a species can be improved by the non-genetic transmission of cultural information that uses a meme as the transmission mechanism rather than the genetically based gene. The difference between the two includes the fact that genetic transmissions (stochastic selection process) evolve over a period of generations, whereas cultural transmissions result from an educational process.

SUMMARY OF THE INVENTION

[0017] According to the invention there is provided a multimedia resource discovery and retrieval system comprising:

[0018] a. Decentralized retrieval of multimedia files located throughout a local area network (LAN), and a wide area network (WAN) combine adaptive solutions achieved by the various the stochastic optimization agents of multimedia resource discovery and retrieval system to the network routing problem coupled with one or more information retrieval problems. The objectives are to achieved: 1) maximization of resource utilization and of overall LAN throughput, and 2) minimization of rejected request packets and guarantee quality of service (QoS). The network routing procedure requires shortest path routing that minimizes "hops" between the source and randomly chosen Internet service providers (ISPs).

Factors that must be considered are connection requirements (end-to-end delay, delay variation, mean rate) and network conditions, and

[0019] b. Real-time system constraints that require timeliness and correctness of periodic tasks, sporadic tasks, and aperiodic tasks executed by the stochastic optimization agents. The periodic tasks are those with fixed time requirements, sporadic tasks are governed by some minimum inter-arrival time, and aperiodic tasks are governed by some stochastic assumptions. Generally, skip-over scheduling is applied to all these tasks since the Internet's workload fluctuates. The skip-over policy defers those jobs which jeopardize the new job's deadlines. These deferred jobs are placed at the end of the scheduling queue in order to reassess the stochastic optimization parameters of optimization agents in order to meet the QoS requirements. The stochastic optimization search and retrieval strategies require that QoS and quality of information sharing be maintained for each ISP hosting multimedia services in order to efficiently retrieve multimedia files.

DETAILED DESCRIPTION OF THE DRAWINGS

Figures

[0020] FIG. 1 is a diagram showing application layer portion of the software architecture of the present invention.

[0021] FIG. 2 is a block diagram of the multimedia resource discovery and retrieval system embodying the invention.

DETAILED DESCRIPTION

Preferred Embodiments

[0022] A preferred embodiment of the present invention is now described with reference to the figures where like reference numbers indicate identical or functionally similar elements.

[0023] Some portions of the detailed descriptions that follow are presented in terms of stochastic optimization agents which can be implemented by those skilled in data processing art to most effectively convey the substance of their work to others skilled in the art. It should be noted that the stochastic optimization agents of the present invention could be embodied in software, could be downloaded to reside on and be operated from different platforms used by a variety of operating systems.

[0024] The present invention also relates to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, or it may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in a computer. Furthermore, the computers referred to in the specifications may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0025] The stochastic optimization agents presented herein are not inherently related o any particular computer of other apparatus. Various general-purpose systems may also be used with programs in accordance with the teaching herein, or it may prove convenient to construct more specialized apparatus. The required structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the

present invention as describe herein, and any references below to specific languages are provided for disclosure of enablement and best mode of the present invention.

[0026] Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the claims.

1. Notational Conventions

[0027] FIG. 1 provides an overview of the processing completed by the stochastic optimization agents of the multimedia resource discovery and retrieval system. Processing starts in this system 110 when the data extraction portion of the stochastic optimization component extract multimedia files from a mapping of an information ecosystem 111 to selected areas of the network structure of the Internet infrastructure 112 which provides access to external databases 113, Web data 114, organization system databases 115, or internal data warehouses 116.

[0028] FIG. 2 presents the adaptive stochastic optimization agents 221-230 for the stochastic optimization component of multimedia resource discovery and retrieval system based on the information discovery model. This model has built-in mechanisms that allow each stochastic optimization agent to adapt its operational parameters to a changing environment. The stochastic optimization agents 221-230 are applied to the ever-changing environment of local area network (LAN) and/ or a wide area network (WAN) traffic which varies considerably, depending on: 1) time of day, 2) time zones, 3) various holiday and/or vacation patterns that exist throughout the world, and 4) naturally occurring disasters. The stochastic optimization further agents 228-230 communicate via message passing mechanisms 231,232, and emulate stochastic optimization information search strategies for: locating forage sources, and detecting and avoiding foraging congestion.

[0029] Stochastic optimization probe further agents 228 are deployed throughout the information ecosystem 111 in search of ISPs hosting multimedia services in order to initiate the development of customized routes for the retrieval of multimedia files by stochastic optimization forager further agents 230. Stochastic optimization scout further agents 229 use information obtained by the stochastic optimization probe further agents 228 to detect network congestion. The various objectives just mentioned are monitored by the stochastic optimization agents 221,222 and stochastic optimization regulatory agents 223-226 using rescaled adjusted range (RS) statistics.

[0030] The stochastic optimization agents 221,222 focus on performance monitoring of the interaction between stochastic optimization regulatory agents 223-226. The stochastic optimization probe agents 224,228 and stochastic optimization forager agents 225,229 form crucial components of the stochastic optimization component of the multimedia resource discovery and retrieval systems. Each located ISP is identified as useful by stochastic optimization probe regulatory agents 224 if it provides the desired multimedia services. These marked sites are provided to the stochastic optimization scout regulatory agent 225. The stochastic optimization probe further agents 228 are not concerned with network congestion or any other aspect of file retrieval. The stochastic optimization scout regulatory agents 225 proceed with the site metrics provided by only its group stochastic optimization probe regulatory agents 224. Stochastic optimization scout further agents 229 are released to each selected site periodically to gather and update information ecosystem 111 congestion traffic metrics. The stochastic optimization scout regulatory agent 225 use these metrics on a per-site basis to determine the feasibility of retrieving multimedia files from a

selected site. Site rankings are based on the results of the feasibility tests which use RS statistics to perform time series analysis on each site's congestion metrics. The feasibility results for each site vary based on the time of day, time zone location with respect to the location of the stochastic optimization agents of multimedia resource discovery and retrieval system, localized holiday and vacation patterns, and natural disasters. The periodic feasibility update of each site occurs within a random time period and is based on the workload of each stochastic optimization scout regulatory agent 225 and stochastic optimization forager regulatory agent 226 coupled with the update rate of the newly located sites provided by its corresponding stochastic optimization probe regulatory agents 224.

[0031] The stochastic optimization regulatory agents 223-226 contain those features essential for releasing and coordinating the stochastic optimization further agents 227-230. Each stochastic optimization regulatory agents 223-226 has a finite scope, limiting its activity to those ISPs inscribed within an area whose radius is given by a value V (its visibility).

[0032] The stochastic optimization probe regulatory agent 224 provide stochastic optimization scout regulatory agent 225 with results in the form of IP addresses reflecting initial visits to random ISPs. The stochastic optimization scout regulatory agent 225 use the IP address of the appropriate ISP in order to start the process of determining/charting the optimal (customized) route using RS statistics. The stochastic optimization forager regulatory agent 226 uses the selected IP address if the RS statistics indicate that the corresponding information server meets the required QoS. This methodology has the ability to discover new ISPs as well as new sub-hosts, thus providing services to both new and existing information clients—this in turn resulting in faster discovery of new and updated documents. Each stochastic optimization further agent 227-230 is:

[0033] 1. Reactive—can interact with the information ecosystem 111 within appropriate time limits

[0034] 2. Independent—can act on its own

[0035] 3. Robust—can cope with the ever-changing network environment within the information ecosystem

The efficiency of the stochastic optimization forager further agents 230 is due to customized ISP router tables which are discovered by the stochastic optimization scout further agents 229—the result of periodic searches for optimized routes that exist for short periods of time. The initial step in this methodology is the releasing of stochastic optimization probe further agents 228 for all ISPs in a manner similar to reliable flooding.

[0036] The rapid release of a series of stochastic optimization further agents (probe/scout/forager sets 228-230) can have an adverse affect on the receiving host (ISP server) as well as on the releasing stochastic optimization regulatory agent 224-226. Each stochastic optimization regulatory agent 223-226 creates a series of stochastic optimization further agents 227-230 which can exhaust the resources of the system resources allocated to the stochastic optimization regulatory agent 223-226. The stochastic optimization further agents 228-230 are most effective in the event/case of reliable flooding, where monitoring stochastic optimization agents 221, 222 and stochastic optimization regulatory agents 224-226 are used to adequately control and coordinate valuable information returned by each individual process. The ISP hosting multimedia services may interpret the simultaneous requests as a form of flooding, resulting in requests being queued at the router level and/or server level. In the worst-case scenario, the life-span of a stochastic optimization further agent agents 228-230 will exceed the amount of time needed to establish

communication with the selected ISP and retrieve the requested information. Attempts to avoid worst-case scenarios are made through the use of RS statistics provided by the stochastic optimization scout further agents **229**.

[0037] The stochastic optimization regulatory agents **223-226** needed for retrieving multimedia documents require some form of adaptive methodology since each stochastic optimization further agent **227-230** searches for efficient paths (routes) to an uncongested source of information (documents) in order to build the stochastic optimization component of multimedia resource discovery and retrieval system ISP router tables.

[0038] The stochastic optimization forager regulatory agent **226** receives input from the stochastic optimization scout regulatory agent **225** which makes retrieval decisions based on the conversion of congestion detection information into high-level congestion avoidance mechanisms before releasing stochastic optimization forager further agents **230**. The release of stochastic optimization forager further agents **230** can only occur if the stochastic optimization scout regulatory agent **225** indicates that the feasibility results pass the QoS requirements imposed by the stochastic optimization component of multimedia resource discovery and retrieval systems. This layer of congestion avoidance incorporates network metrics from mechanisms used to customize routes between the location of the stochastic optimization component of multimedia resource discovery and retrieval systems and each selected ISP. Snapshots of source/destination traffic flow can change drastically over relatively short periods of time—depending on the release and return of each stochastic optimization scout further agents **229**. The second layer of congestion avoidance is handled implicitly by information ecosystem and Internet routers and switches **111,112** between the source **110** and destination **113,114,115,116**.

I claim:

1. A method of applying distributed stochastic optimization techniques of evolutionary computation using a plurality of servers and a plurality of clients machines being connected via a computer network, said stochastic optimization techniques of evolutionary computation aiming to optimize a populations of individuals against one or more predetermined fitness criteria, wherein the computer code instructions is invoked by stochastic optimization agents:

  a. said method for applying distributed stochastic optimization techniques of evolutionary computation including the steps of:

    i. requesting multimedia services provided by a plurality of servers to generate a population individuals by the client machines;

    ii. requesting QoS statistics to the multimedia services provided by a plurality of servers to generating a population of individual routing tables to a plurality of servers by said client machines;

    iii. requesting multimedia documents provided by a plurality of servers to generate a population of multimedia documents downloaded from a plurality of servers by said client machines;

  b. responsive to said requesting step, distributing individuals from said population of individuals from the plurality of servers across plurality of client machines by downloading individuals from the requesting client machine agents to the corresponding non-requesting client machine agents;

  c. evaluating the fitness of each distributed individual at a client machine by a stochastic optimization agent calling said method to invoke said computer code instructions; and

  d. identifying the individual(s) best satisfying said one or more predetermined fitness criteria, based on results of said evaluating step.

2. The method of claim **1**, where each of said plurality of client machines runs a text-only World Wide Web browser.

3. The method of claim **2**, where the said computer network is the Internet.

4. The method of claim **1** or **2**, further comprising the step of using the individuals evaluated as having the best fitness to produce a new generation of individuals.

5. The method of claim **4**, further comparing the step of sharing the new generation of individuals with non-requesting client machine agents.

6. The method of claim **4**, further comprising an initial step of creating the first generation of individuals at said plurality of client machines.

7. The method of claim **1**, wherein each of said plurality of client machines can commence or terminate their request from the plurality of servers at any time.

8. The method of claim **1**, wherein each of said plurality of server machines can commence or terminate their response with the plurality of client machines at any time.

9. The method of claim **1** or **2**, further comprising an initial step of creating the first generation of individuals at said plurality of client machines.

10. A system for applying distributed stochastic optimization techniques of evolutionary computation using a plurality servers and a plurality of clients machines, the server and client machines being connected via a computer network, said stochastic optimization techniques of evolutionary computation aiming to optimize a population of individuals, against one or more predetermined fitness criteria, each individual being generated by a stochastic optimization agent using computer code instructions to evaluate the fitness of individual, the system including:

  a. means, in the client machines for requesting multimedia services provided by a plurality of servers to generate a population individuals by the client machines;

  b. means, in the client machines for requesting QoS statistics provided by a plurality of servers to generate a population individuals by the client machines;

  c. means in the requesting client machine responsiveness to requests from the non-requesting client machines for distributing individuals from said population of individuals generated by the plurality of requesting client machines across said plurality of non-requesting client machines by downloading individuals from the plurality of client machines to the corresponding requesting client machines;

  d. means for evaluating the fitness of each distributed individual at a client machine by calling said method in said stochastic optimization agent(s) to invoke said computer code instructions; and

  e. means for identifying the individual(s) best satisfying said one or more predetermined fitness criteria, based on the results of their fitness evaluations.

* * * * *