(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 August 2003 (14.08.2003)

PCT

(10) International Publication Number
WO 03/067731 A2

(51) International Patent Classification[7]: H02J

(21) International Application Number: PCT/US03/03297

(22) International Filing Date: 4 February 2003 (04.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/354,588 4 February 2002 (04.02.2002) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/354,588 (CIP)
Filed on 4 February 2002 (04.02.2002)

(71) Applicant *(for all designated States except US)*: **ROUTESCIENCE TECHNOLOGIES, INC.** [US/US]; 167 2nd Avenue, San Mateo, CA 94401 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LLOYD, Michael, A.** [US/US]; 160 Arundel Road, San Carlos, CA 94070 (US). **KARAM, Mansour, J.** [LB/US]; 707 Continental Circle, #421, Mountain View, CA 94040 (US). **VILLAVERDE, Jose-Miguel, Pulido** [ES/US]; 1020 Bryant Street, Palo Alto, CA 94301 (US). **FINN, Sean, P.** [US/US]; 1533 Escondido Way, Belmont, CA 94002 (US). **BALDONADO, Omar, C.** [US/US]; 700 Alester Avenue, Palo Alto, CA 94303 (US). **MCGUIRE, James, G.** [US/US]; 2312 Gough Street, San Francisco, CA 94019 (US). **MADAN, Herbert, S.** [US/US]; 347 Blackfield Drive, Tiburon, CA 94920 (US).

(74) Agent: **SUZUE, Kenta**; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/067731 A2

(54) Title: LOAD OPTIMIZATION

(57) Abstract: Methods, computer code, and means are described that can control load in a network. In some applications, the monetary cost of operating the network can be reduced. Utilization of links in the network can be monitored. A degree of suboptimality with respect to some criteria can be assessed. In some instances, the criteria could be based at least partly one or more monetary billing structures of some subset of two or more links. A subset of the forwarding decisions of one or more forwarding nodes in the network can be adjusted automatically, based at least partly on the assessing. The adjustment can attempt to reduce the degree of suboptimality.

LOAD OPTIMIZATION

BACKGROUND OF THE INVENTION

By changing the forwarding decision of a network, a network user can decrease the cost of using the network, or otherwise enhance the load distribution of the network. One approach to decreasing the cost of using the network is for a person to periodically intervene and adjust the forwarding decisions of the network.

Unfortunately, manually adjusting the forwarding decisions of particular network nodes is an imperfect solution. First, manual adjustments are labor intensive. Second, manual adjustments are slow. Because of the dynamic nature of network traffic, manual adjustments that may have had the result of decreasing cost at one point in time may not have the effect of decreasing cost at a later time - or worse, even increase the cost.

Another difficulty with adjusting forwarding decisions is that monetary billing structures can be complicated, such as when the monetary billing structure is not flat. Particularly when multiple monetary billing structures (e.g., of multiple providers such as internet service providers) of multiple links are considered with the dynamic nature of network traffic, correctly adjusting forwarding decisions while attempting to decrease the cost of using the network can present a significant challenge.

What is needed is an effective solution for adjusting the load distribution in a network, for example to decrease the cost of using the network.

BRIEF SUMMARY OF THE INVENTION

Some embodiments of the invention control load in a network.

Some embodiments of this invention reduce the monetary cost of operating the network. Some embodiments include at least part of one or more of:

- Monitoring at least a first utilization of a first subset of two or more links in the network
- Assessing the degree of suboptimality with respect to some criteria. In some instances, the criteria could be based at least partly one or more monetary billing structures of a second subset of two or more links, wherein:
  - at least one of the one or more monetary billing structures receives as input at least a second utilization of the second subset of two or more links,
  - at least one of the one or more monetary billing structures includes variable cost, and
  - the first utilization of the first subset of two or more links is at least partly indicative of the second utilization of the second subset of two or more links
- Adjusting automatically a subset of the forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing, wherein the adjusting attempts to reduce the degree of suboptimality.

In some embodiments of this invention, the steps of monitoring, assessing, and adjusting are independent – in such embodiments, no causal relationship exists between the steps of monitoring, assessing, and adjusting.

In some embodiments of this invention, adjustments can be made as to control load without excessively compromising performance. In some embodiments of this invention, the assessment of suboptimality is based at least partly on the monitoring, hence providing a closed loop system. (e.g., in such embodiments of the invention, the adjusting could affect load; the reading of the monitoring could then be reflected by the consequent changes in load, resulting in a modification in the results of the assessment, which in turn provokes new adjustments.) In other embodiments of this invention, the assessment of suboptimality is not necessarily based on the monitoring. In some embodiments of this invention, the steps of monitoring, assessing, and adjusting are continually repeated so that the latest information provided by the monitoring can be used in adjusting the forwarding decisions.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a computer programmed from program media.

Figure 2 illustrates a computer programmed from a network.

Figure 3 illustrates a network with nodes and links that are adjusted, links that are assessed, and links that are monitored.

Figure 4 illustrates a network with links that are both assessed and monitored.

Figure 5 illustrates a network with links that are both assessed and adjusted.

Figure 6 illustrates a network with links that are both assessed and monitored, links that are assessed but not monitored, and links that are monitored but not assessed.

Figure 7 illustrates an example of a first degree of unacceptability function.

Figure 8 illustrates an example of monetary billing structures.

DETAILED DESCRIPTION OF THE INVENTION

Various embodiments of the invention include methods, software, hardware, and/or a combination.

The software can be on any of various program media, such as an optical medium (e.g., a DVD, CD), a magnetic medium (e.g., a floppy or hard disk), an electrical medium (e.g., flash), a nanoscale medium, or some combination. The software can also be in a transitory medium, such as an optical signal, magnetic signal, electrical signal, or some combination, such as an electromagnetic wave. The software can also be stored on a computer, such as on long term storage or short term storage, such as in volatile or nonvolatile memory.

The hardware can be any of various mechanisms, such as a computer, personal digital assistant, cell phone, or embedded device. The hardware may be implemented on program media such as an integrated circuit or chip that can be added to a computer.

Some embodiments are a combination of hardware and software, such as hardware with some of the instructions implemented in the hardware, combined with software for some of the instructions executing on the hardware.

Computer code in various embodiments can be implemented in hardware, software, or a combination of hardware and software.

Figure 1 illustrates a computer 110, which is programmed by code stored on program media 120. The program media 120 is used to place code on the computer 110.

Figure 2 illustrates a computer 210, which is programmed by code from a network 230. The network 230 is used to place code on the computer 210

In this document, we describe mechanisms that can be used to control load in a network. In some embodiments of this invention, these mechanisms will be used to reduce the monetary cost of operating the network.

Some embodiments include at least part of one or more of:

- Monitoring at least a first utilization of a first subset of two or more links in the network
- Assessing the degree of suboptimality with respect to some criteria. In some instances, the criteria could be based at least partly one or more monetary billing structures of a second subset of two or more links, wherein:
    - at least one of the one or more monetary billing structures receives as input at least a second utilization of the second subset of two or more links,

- at least one of the one or more monetary billing structures includes variable cost, and
- the first utilization of the first subset of two or more links is at least partly indicative of the second utilization of the second subset of two or more links

➢ Adjusting automatically a subset of the forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing, wherein the adjusting attempts to reduce the degree of suboptimality.

In the following sections, we describe how, in some embodiments of the invention, the steps of monitoring, assessing, and adjusting would be performed.
In some embodiments of this invention, the steps of monitoring, assessing, and adjusting are independent – in such embodiments, no causal relationship exists between the steps of monitoring, assessing, and adjusting.

In some embodiments of this invention, adjustments can be made as to control load without excessively compromising performance. In some embodiments of this invention, the assessment of suboptimality is based at least partly on the monitoring, hence providing a closed loop system. (E.g., in such embodiments of the invention, the adjusting could affect load; the reading of the monitoring could then be reflected by the consequent changes in load, resulting in a modification in the results of the assessment, which in turn provokes new adjustments.) In other embodiments of this invention, the assessment of suboptimality is not necessarily based on the monitoring. In some embodiments of this invention, the steps of monitoring, assessing, and adjusting are continually repeated so that the latest information provided by the monitoring can be used in adjusting the forwarding decisions.

In some embodiments, load and utilization. can be inter-related. Load can include a measure of traffic, for example, in bits per second, flowing across a resource. Utilization can include a measure of the load portion of resource capacity. For example, the load of a link could be 200 bits per second. If the link capacity is 500 bits per second, then the link utilization can be 200 / 500 = 0.4 = 40%. So in this case, for some embodiments, a load of 200 bits per second and a utilization of 40% are equivalent statements about the rate of traffic flowing through the link. In some embodiments, utilization can include an absolute portion without reference to the resource capacity, such as a load, rather than a

relative portion with reference to the resource capacity. In some embodiments, utilization can include a relative portion of another value besides the resource capacity.

In some embodiments of this invention, monitoring is used to provide load information upon which, in some systems, the assessing will partly be based. In some embodiments of this invention, the monitoring uses the Simple Network Monitoring Protocol (SNMP); in other embodiments, the monitoring is based partly on flow information export. One such flow information export is NetFlow. In other embodiments of this invention, monitoring is based at least partly on a source external to the subset of forwarding decisions used in the adjusting. In some embodiments of this invention, the monitoring is based at least partly on span port.

In some embodiments, systems are included to deal with the case where monitoring is done for a subset of set of two or more links, but not for another subset of the two or more links. In some embodiments, in instances where SNMP is used for monitoring, systems are included to deal with timeouts in SNMP polling.

In some embodiments, monitoring can be done using bye counts over a time interval of specified length. In other embodiments, monitoring can be done using rates.

In some embodiments of the invention, a minimum limit is imposed on the number of utilization samples obtained from the monitoring before assessing can proceed.

In some embodiments of this invention, the method takes into account the load corresponding to subsets of the objects. In some such embodiments, the subsets of objects correspond to one or more prefixes. This information can be obtained through monitoring systems that will be recognized by the skilled in the field. Such mechanisms include NetFlow, RMONI/II, span port, and other external monitoring sources. Such monitoring systems can also include systems based at least partly on web server logs; for example, rate of requests per destination can be counted for different applications. If the subsets of objects include one or more prefixes, one can also use the size of the prefix as an estimate of the contribution of that prefix to the total utilization. For example, a /8 would

be estimated to have twice the traffic than a /9, itself having twice the traffic of a /10.

In some embodiments of this invention, the monitoring combines the utilization samples in some fashion. In some embodiments of this invention, the monitoring estimates a percentile of load samples. In some embodiments, an estimation of the n^{th} percentile includes, given a sampling rate r and a billing period b, storing the largest (1-n)*b*r samples during a billing period.

The assessing is done on a set of two or more links that, in some embodiments of this invention, are the same as the set of two or more links being monitoredIn some embodiments, the two sets are equal; In some embodiments, the two sets may overlap; yet in other embodiments, they can be different. In some embodiments, the load utilization of the set of links used for the assessing can be deduced from the load utilization of the set of links that are used for the monitoring. For example, in some embodiments of this invention, the utilization on the links that are monitored can be equal to the utilization on the links that are assessed.

In some embodiments of this invention, forwarding decisions are adjusted as to control load. In some embodiments of this invention, forwarding decisions are adjusted as to strike an adequate balance between load control and performance.

In such embodiments, assessing includes at least partly an assessment of load and/or an assessment of performance. In some embodiments, load and performance information can be combined in a metric that can be used to rate one or more of the two or more links in the network. In some embodiment, metrics can be computed for one or more links for objects controlled by forwarding decisions based at least partly on performance information for these objects on the one or more links; the metric for each of these links can then be penalized by an amount that is based, at least partly on the desired utilization of the one or more links. In some embodiments, the penalty associated for at least one of the one or more links can be at least partly fixed; in other embodiments, at least one of the one or more penalty values corresponding to the one or more links can be at least partly variable.

In some embodiments of the invention, the objects controlled by the forwarding are prefixes. In some embodiments of the invention, the objects controlled by the forwarding are flows. In some embodiments of the invention, the objects controlled by the forwarding are network applications.

In some embodiments of this invention, computing the object penalties of the one or more links is based at least partly on the amount the corresponding metric needs to be degraded by so that the metric on this link is deemed unacceptable. In some embodiments, the standard of unacceptability is based at least partly on the concept of a winner set, the width of this set including metric values that are deemed acceptable.

First degree of unacceptability functions

In some embodiments of this invention, the assessing includes generating one or more sets of functions, wherein at least one function in the one or more sets of functions gives a first degree of unacceptability of at least one link from the first subset of two or more links, wherein the first degree of unacceptability is based at least partly on utilization of the at least one link in the network.

In some embodiments of this invention, at least one function in the one or more sets of functions outputs at least a varying value. In some embodiments, at least one function in the one or more sets of functions is continuous or piecewise continuous with respect to utilization. In some embodiments, the at least one function in the one or more sets of functions is non-decreasing with respect to load.

In some embodiments of the invention, at least one degree of unacceptability function in the at least one set of degree of unacceptability functions receives at least one input, the at least one input at least partly depending on load, wherein the at least one degree of unacceptability function outputs at least:

1) a first constant value for values of the at least one input ranging from a second constant value to an third constant value

2) a linear function of at least one input for values of the at least one input ranging from the third constant value to a fourth constant value

3) a fifth constant value when the values of the at least one input exceeds the fourth constant value

In some embodiments, the first degree of unacceptability function can be computed as follows: (We denote the first degree of unacceptability p.)

p=0 if load <= startAvoidance

p=maxProbability * (load - startAvoidance) / (maxAvoidance - startAvoidance) if startAvoidance < load <= maxAvoidance

p=maxProbability if load > maxAvoidance

(See Figure 7, wherein threshold = maxAvoidance.)

Figure 7 illustrates an example of a first degree of unacceptability function.

In some embodiments of the invention, at least one degree of unacceptability function in the at least one set of functions receives at least one input, the at least one input at least partly depending on load , wherein the at least one degree of unacceptability function outputs at least:

1) a first constant value for values of the at least one input up to a threshold value

2) a second constant value for values of the at least one input above the threshold value

In some embodiments, the first degree of unacceptability function can be computed as follows: (We denote the first degree of unacceptability p.)

p=0 if load <= avoidance

p =maxProbability if load > avoidance

The load value is based at least partly on the monitoring. In some instances of the invention, the load value is based at least partly on inbound utilization. In some instances of the invention, the load value is based at least partly on outbound utilization. In some embodiments of the invention, load value is based at least partly on max(inbound,outbound); in some instances of the invention, load value is based at least partly on avg(inbound,outbound); in some instances of the invention, the load value is based at least partly on inbound + outbound. In some instances of the invention, the load value can be based on the instantaneous load values that result from the monitoring. In some instances of the invention, the load values are based at least partly on a percentile of a subset of load values that result from the monitoring. In some instances of the invention, the load values are based at least partly on the average of a subset of load values that result from the monitoring.

In some embodiments of the invention, different first degrees of unacceptability curves are applied to different forwarding decisions. More than one degree of unacceptability can exist. Selection of a set of functions can be done per forwarding decision. In some embodiments of the invention, no degree of unacceptability is applied

to at least one link for at least one forwarding decision. For example, not all functions that are being assessed must have one or more sets of functions assigned to them.

In some instances, the assessing also includes the computation of a second degree of unacceptability for a link that can be dependent at least partly on the first degree of unacceptability. In some embodiments, determining of the second degree of unacceptability includes treating the first degree of unacceptability as a probability value, and assigning, using the probability value, one of a plurality of states to the second degree of unacceptability. In some such embodiments, the second degree of unacceptability can be assigned two states, that we denote here "hot" and "cold" based at least partly on the result of a random selection based at least partly on the first degree of unacceptability.

In some embodiments of the invention, the winner sets are constructed in an ordered list of one or more winner sets, where the elements of a winner set are links from the set of two or more links. In such embodiments, the elements of a winner set are comparable in quality for an object influenced by the forwarding decisions. In such embodiments, links that have a second degree of unacceptability that is large enough are not included in at least one winner set. In the instances of the invention in which the second degree of unacceptability includes one of the two states, "hot" and "cold", hot links are removed from at least one winner set in a list of one or more winner sets.

In some instances of the invention, the ordered list of one or more winner sets includes two winner sets, denoted the basic winner set and the extended winner set. If such instances also include a second degree of unacceptability that includes two states, "hot" and "cold", and if, for an object, the basic winner set is empty and the extended winner set is non-empty, then the forwarding decision that influences this object is adjusted to point to at least one of the one or more links in the extended winner set. In some embodiments of the invention, all winner sets are empty in the ordered list of winner sets, no adjustment is done for this object, and an attempted adjustment may be done to the following object. In other embodiments, an adjustment is performed that is based solely on performance. In other embodiments, a new ordered list of winner sets is constructed, based on a new set of first degree of unacceptability functions for each link. (See the section on more than one set of functions.). In other embodiments, one or more links in the set of two or more links can be chosen using a probabilistic approach. In one such embodiment, one link in the set of two or more links can be chosen randomly among the various links in the set of two or more links. In such embodiments, the probability density function used for the random selection can be biased towards some links and

away from other links, based at least partly on the monetary cost of the one or more links. When all winner sets are empty in the ordered list of winner sets, other possible choices of action will be visible to those skilled in the art.

In some embodiments, assessing is based at least partly on monitoring a degree of suboptimality with respect to one or more monetary billing structures of a subset of two or more links in the network, wherein:

- at least one of the one or more monetary billing structures receives as input at least a utilization of the subset of two or more links, and
- at least one of the one or more monetary billing structures includes at least variable cost.

The monetary billing structures are applied to a set of two or more links that, in some embodiments of this invention, are related to the set of two or more links being assessed.

Monetary billing structures can include one or more rules which determine a monetary bill resulting from the use of network links.

In some embodiments, the two sets are at least partly equal and/or unequal; in some embodiments, the load utilization of the set of links on which the monetary billing structures are based can be deduced from the load utilization of the set of links that are used for the assessing. For example, in some embodiments of this invention, the utilization on the links that are monitored can be equal to the utilization on the links on which the monetary billing structures are based. In some embodiments, the utilization of the links that are monitored overlap the utilization on the links on which the monetary billing structures are based. In yet other embodiments, the utilization of the links that are monitored are different from the utilization of the links on which the monetary billing structures are based.

Suboptimality can mean the existence of a state, and/or can mean the degree of a state,. respect to one or more of the monetary billing structures, such that the cost of operating the network, as given by the monetary billing structures, is not minimized. Reducing the suboptimality with respect to one or more of the monetary billing structures therefore includes minimizing the discrepancy between the current load distribution and the optimal load distribution for which the cost of operating the network is minimized.

In some embodiments, at least one of the one or more monetary billing structures receives as input at least a utilization of at least one link from the second subset of two or more links, wherein the utilization may be determined over time. In some embodiments, the utilization is computed at least partly from at least one of: 1a) a maximum and 1b) an average, of at least one of: 2a) one or more percentiles and 2b) one or more averages, of one or more sets of utilization samples of the at least one link from the second subset of two or more links. In some embodiments, the billing structure is based on some amount such as a percentage, e.g. 95%, of the link utilization, measured over a billing period. In some embodiments, the billing period is equal to a regular period, such as a month, week, day, hour, or fraction or multiple thereof. In some embodiments, load is controlled by taking into account, at least partly, the same formula used in utilization for billing. For example, in the instance where the billing structure is based on the 95% of a link utilization, some embodiments of the invention can choose to only react when some estimation of the 95% of the link utilization is about to jump beyond a value that could cause in an increase in the bill. In some such embodiments, this can be achieved by having the first degree of unacceptability only increase once such thresholds are reached. Once such a threshold is exceeded, a second set of first degree of unacceptability functions are used, where the threshold now becomes the next point in the billing structure for the link where the bill increases again.

In some embodiments of this invention, the billing structures are based at least partly on the 95^{th} percentile of a function of both the inbound and outbound load of the at least one link. In some embodiments, the function of both the inbound and outbound load is a combining function, such as the averaging function.

In some embodiments, the billing structures are based at least partly on a function of both the 95^{th} percentile of the inbound load and the 95^{th} percentile of the outbound load. In some embodiments, the function of both the 95^{th} percentile of the inbound load and the 95^{th} percentile of the outbound load is the averaging function; in some embodiments, the function of both the 95^{th} percentile of the inbound load and the 95^{th} percentile of the outbound load is the max function.

The 95^{th} percentile value is illustrative. Other values in the range of 0-100%, or an absolute, non-percentage-based value, can be used.

In some embodiments of this invention, the assessing is done using more than one set of functions. In some embodiments, the system would select, for a given object, a first set of functions from the one or more sets of functions; if the first degree of

unacceptability fails a threshold of acceptable unacceptability for all functions in the set of functions, then a second set is chosen. In some embodiments, one example of a degree of unacceptability can be a degree of unacceptability. In some embodiments, one example of a threshold of acceptable unacceptability can be a threshold of unacceptability. In some embodiments, examples of failing a threshold of acceptable unacceptability can include any of: passing a threshold of unacceptable unacceptability, failing a threshold of unacceptable acceptability, and/or passing a threshold of acceptable acceptability.

Alternatively, in some embodiments where performance considerations also taken into account, so that the assessing is further based at least partly on quality characterizations of the one or more objects, then the assessing further includes selecting at least one object from the one or more objects, selecting at least one set of functions from the one or more sets of functions, and constructing one or more winner sets for the at least one object and the at least one set of functions, wherein each winner set from the one or more winner sets includes a corresponding quality characterization threshold, wherein constructing includes:

1. including in at least one of the one or more winner sets one or more links from the subset of two or more links,

2. excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets

3. excluding, from the at least one or more winner sets, unwanted links, wherein the unwanted links have a degree of unacceptability failing a threshold of acceptable unacceptability, wherein the degree of unacceptability is based at least partly on the first degree of unacceptability given by the at least one set of functions

In various embodiments, an example a quality characterization can indicate quality and/or lack of quality. In some embodiments, an example of failing a quality characterization threshold can be passing a quality characterization.

Finally, in such embodiments, the links that are selected are from the a non-empty winner set from the one or more winner sets, wherein the non-empty winner set has a low corresponding quality characterization threshold (such as a lowest corresponding quality characterization threshold) from all corresponding quality characterization thresholds included by all winner sets from the one or more winner sets.

In such embodiments, the excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets can include:

identifying at least one best link from the one or more links from the third subset of two or more links, wherein the at least one best link has a high quality characterization from at least one of the one or more links from the third subset of two or more links, and-

determining the corresponding quality characterization threshold based at least partly on the high quality characterization.

In such embodiments, the selection of a second set can also occur when the constructing of the first one or more winner sets corresponding to the first set of functions yields all empty winner sets. In this case, a second set of functions from the one or more sets of functions is chosen, and a second one or more winner sets is constructed for the second set of functions from the one or more sets of functions

In some embodiments, the one or more sets of functions are ordered into an ordered list of, for example, functions that are nontrivial to the embodiment. In this case, the first and second set of functions referred to above are adjacent in the ordered list of the one or more sets of functions. Adjacent functions can have in between one or more functions that are trivial to the embodiment. In some embodiments,

In some embodiments, the ordering includes the following steps:

- computing the first degree of unacceptability function using the following function of load: (We denote the first degree of unacceptability p.)

p=0 if load <= startAvoidance

p=maxProbability * (load - startAvoidance) / (maxAvoidance - startAvoidance) if startAvoidance < load <= maxAvoidance

p=maxProbability if load > maxAvoidance

- computing, for each set of functions in the one or more sets of functions, a level, wherein a level is based at least partly on a sum of maxAvoidance values across the one or more functions in each set of functions

- performing the ordering based at least partly on the level computed for each set of functions

In some embodiments, the approach above is combined in a table that we denote the threshold table. In some embodiments, the table consists of multiple rows, wherein each row in the table includes information regarding one set of functions, i.e., corresponding to one level. For each set of functions, the parameters corresponding to each function are described. If the functions include a minAvoidance and maxAvoidance as described above, then the minAvoidance and maxAvoidance parameters are included in the row for each function. In addition, if assessing is based at least on a second degree of acceptability, then in some embodiments, the value of the second degree of acceptability can also be stored along with each function. Each set of functions includes functions for a number of links in the network.

In some embodiments, one level is selected at any one time. In some embodiments, the selection includes the following steps:

- compute a total load across links of interest.
- Select the minimum level that is larger than the total load.

In some embodiments, the example below applies: if the total load is 90, the probability of rejection for link L1 will be computed using start-avoidance 40, max-avoidance 44. The (x, y) pairs represent the minAvoidance and maxAvoidance for each function for each set of functions corresponding to each level.

	link L1	link L2	link L3
level 85	(30,35)	(20,25)	(20,25)
level 94	(40,44)	(20,25)	(20,25)
level 132	(40,44)	(40,44)	(20,44)

Load threshold table

In some embodiments of this invention, a function for at least a link receive for input at least one of the values of outbound loads for the at least one link.

In some embodiments of this invention, a function for at least a link receive for input at least one of the values of inbound loads for the at least one link.

In some embodiments of this invention, a function for at least a link receive for input at least one of the values of a combination of inbound loads and outbound loads for the at least one link.

In some embodiments, the system, upon receipt of a new load sample on a link, can do the following:

- Update the load info on the link
- Select the active level on each load-threshold-table based on the updated sampled total load
- Update the first degree of unacceptability for the link, for the active level

Some embodiments of this invention have different sets of functions for different objects.

In some embodiments, when the monitoring results in a new load sample that triggers a change in the active level, the assessing also includes re-computing the first degree of unacceptability based at least partly on the new level.

In some embodiments of this invention that include a second degree of unacceptability that includes two states "hot" and "cold", the assessing includes at least one of the following steps:

- evaluating the value of the second degree of unacceptability based at least partly on treating the first degree of unacceptability as a probability value, and assigning, using at least the probability value, one of "cold" and "hot" to the second degree of unacceptability.
- Excluding from the winner set the links that are "hot"
- If the winner set is empty after excluding the hot links, an extended winner set having a larger winner set width is used.
- Excluding from the extended winner set the links that are "hot"

If the extended winner set is empty after the excluding of the hot links, various embodiments can do different things:

- In some embodiments, the system selects another object in the list.
- In some embodiments, a selection of a link based solely or primarily on the quality characterization of the links is done.

- In some embodiments, if none of the probabilities derived from the first degree of unacceptability functions are larger than one for all the links in the performance-only winner set (prior to the excluding steps above), at least one of the following steps is included:
    - For those links in the performance-only winner set for which the probability is less then one, reevaluate the probabilities until at least one links' second degree of unacceptability is assigned the "cold" state.
    - Select at least one link from the one or more links that are assigned the "cold" state.

- In some embodiments, move to the set of functions corresponding to the next level, and re-evaluate the second degree of unacceptability for this next set of functions.
- In some embodiments,
    - For those links in the performance-only winner set for which the probability is less then one, reevaluate the probabilities until at least one of links' second degree of unacceptability is assigned the "cold" state.
    - Select at least one link from the one or more links that are assigned the "cold" state.
- In some embodiments, select from any subset of the links at random
- In some embodiments, compute a second probability based on a first degree of unacceptability assigned to each link, wherein the second probability is based at least partly on the distance between one and the value of the first degree of unacceptability. In some embodiments, the following example applies: if the first degrees of unacceptability for two links are 0.9 and 0.8, respectively, then assign to the two links a second probability value proportional to 1-0.9 = 0.1 and 1-0.8 = 0.2, respectively, leading to a second probability value of 0.5 and 1 for the two links, respectively. In some embodiments, the second probability corresponds to the probability for the link to be "cold".
-

In some embodiments of this invention, the set of functions from which one derives the first degree of unacceptability based at least partly on the monetary billing structures.

In such embodiments, assessing includes generating, from at least one of the one or more monetary billing structures, one or more sets of functions, wherein at least one

function in the one or more sets of functions gives a first degree of unacceptability of at least one link from a subset of two or more links, wherein the first degree of unacceptability is based at least partly on a utilization of the at least one link from the subset of two or more links.

In some embodiments, the generating of the sets of functions includes

- compiling a list of sums of loads (i.e., total load), wherein at least one sum of the list adds up the different combinations of load on the links,

- determining, for different values of total load, an optimal utilization distribution based at least partly on the at least one of the one or more monetary billing structures, and

- constructing the one or more sets of functions based at least partly on the utilization distribution

In some embodiments, determining the optimal utilization involves solving for the minimum monetary cost of operating the network, with respect to the at least one of the one or more monetary billing structures

In some embodiments, determining the optimal utilization involves a steepest descent strategy with respect to the at least one of the one or more monetary billing structures. (See example on steepest descent approach.)

In some embodiments of this invention, the determining of the adequate set of functions includes at least one of the following steps:

1. Determining an estimate of the sum of the individual amounts, e.g., 95th percentiles, from prior billing intervals
2. Round the estimate up by approximately one billing interval (e.g., 3 Mbps)
3. Using a calculation program (e.g., Excel, Mathematica) to figure out the best allocation of the estimated load, and assigning the level and the maxAvoidance values based at least partly on the estimated load
4. For at least one other level, assigning the max avoidance of one of the functions in the level to be the link capacity.

In some embodiments, Step 4 can be repeated for all links of interest.

In some embodiments, if the number of links that include first degree of unacceptability functions is N, then we have N+1 levels.

In some embodiments, if the number of links that include first degree of unacceptability functions is N, then we have at most N levels.

Those skilled in the art will recognize other ways of constructing the sets of first degree of unacceptability functions based on the billing structures.

In some embodiments of this invention, startAvoidance and maxAvoidance are related as follows:

StartAvoidance = maxAvoidance*(1-percentageBelowMax)

In some embodiments of the invention, the problem of finding an optimal load distribution can be posed as a linear programming problem. That is, given:

- N the total number of links
- $C(x_1)$, $C(x_2)$, ..., $C(x_N)$ the cost function of each link as a function of the load on each of these links x_1, x_2, ... and x the total load,

Find x_1, x_2, ..., x_N (the load on each of the links) such that:

1. $x_1 + x_2 + ... + x_N = x$
2. $x_1, ..., x_N >= 0$
3. $C(x_1) + C(x_2) + ... + C(x_N)$ is minimized

In some embodiments of this invention, linear programming techniques can be applied to solve this problem.

One can take advantage of the cost functions on the links, and the fundamental theorem of linear programming, to transform the search of target loads in a table lookup. The fundamental theorem of linear programming states that optimal points in an

optimization problem are extreme points of the feasible regions, that is the regions where a valid solution can be found. A valid solution is a combination of load values such that the cost is optimal, for a given total load. Linear programming algorithms such as the simplex algorithm speed up the calculation of solutions by restricting the search for optimal values on the set of extreme points only.

In some embodiments, the problem can be converted into a table lookup using a heuristic approach. In some such embodiments, for each load sample, a table of optimal solutions is stored, wherein the table of optimal solutions includes the combinations of load values that lead to optimal cost. In some embodiments, the appropriate row is retrieved each time a new load sample comes in. In some embodiments, the choice of the optimal solution is based on a proximity factor, wherein the proximity factor selects the optimal solution that minimizes the load changes among links, for the current combination of individual loads that lead to the total load that's being looked up. In some embodiments, the proximity factor can be based on at least one of the following functions:

$PF(OPj) = \text{sum } i \; (\text{current\_load\_i} - \text{target\_load\_j\_i})^2$

square error

$OP = \min j \; PF(OPj)$

least square error

In some embodiments, computing this table is a one-time effort. In some embodiments, the computation of this table is done off-line. In some embodiments, the computation of this table is done periodically. In some embodiments of this invention, the computation of this table is triggered by an external event.

In some embodiments, determining the optimal utilization involves a steepest descent strategy with respect to the at least one of the one or more monetary billing structures.

In some embodiments of the invention, the one or more sets of function that give a first degree of unacceptability use at least one of the following:

- 1) Defining the first load tier to be the minimum commit level of all providers

- 2) Defining the next bandwidth level by selecting the provider that represents the *smallest incremental cost increase.* In some embodiments of the invention, utilize that provider for the full duration of that cost tier.
- 3) In some embodiments, in instances where the incremental cost increase is identical, select the provider that maintains that billing level for the longest duration (greatest capacity.)

In some embodiments of this invention, Steps 2 and 3 are repeated. In some embodiments of this invention, Steps 2 and 3 are repeated until the maximum cost tier is reached for all providers. In some embodiments, the maximum cost tier constitutes the physical link capacity

In some embodiments, a set of function in the one or more sets of functions that give the first degree of unacceptability is set at the actual level of transition, wherein the actual level of transition is based at least partly on the provider's billing model. In some embodiments, it is not necessary to cautiously set thresholds lower than the actual provider bandwidth tiers. In some embodiments, the maxAvoidance is set to the actual transition levels for all links. In some embodiments, startAvoidance is set to an amount, such as 10% lower than the true threshold. In some embodiments, a value for startAvoidance is selected automatically.

For this example, we will assume that the enterprise has active links to three service providers, who bill according to the following utilization tiers:

	usage level	cost
Service Provider 1		
minimum commitment:	up to 10 mbps	$100
billing tier 1	11 – 20 mbps	$250
billing tier 2	21 – 45 mbps	$400
Service Provider 2		
minimum commitment:	up to 10 mbps	$150
billing tier 1	11 – 15 mbps	$200
billing tier 2	16 – 45 mbps	$350

Service Provider 3

minimum commitment:	up to 5 mbps	$200
billing tier 1	6 – 40 mbps	$300
billing tier 2	41 – 45 mbps	$450

Figure 8 illustrates an example of monetary billing structures.

Following the implementation steps above, as used by some embodiments of the invention, the chart above would yield the following load tiers for some embodiments of the invention:

level (aggregate bandwidth)		provider 1	provider 2	provider 3
Tier 1	25	10	10	5
Tier 2	30	10	15	5
Tier 3	65	10	15	40
Tier 4	95	10	45	40
Tier 5	105	20	45	40
Tier 6	130	45	45	40
Tier 7	135	45	45	45

Tier 1: In some embodiments of this invention, the first tier is configured to make optimal use of the minimum commit level of each provider. In some embodiments, the level value is simply the sum of all provider thresholds.

Tier 2: In some embodiments of this invention, the second tier is configured to use provider 2 for any traffic that exceeds the minimum commit levels of tier (1). In some embodiments, Provider 2 was selected by comparing the incremental cost increase of all three providers at the next utilization level, and selecting the cheapest:

provider 1: $100 → $250 = $150 increase
provider 2: $150 → $200 = $50 increase
provider 3: $200 → $300 = $100 increase

In some embodiments, once provider 2 is identified, it is utilized to its full capacity at the next cost tier. In this example, provider 2 is used until that link approaches 15 mbps.

Tier 3: In some embodiments, if bandwidth utilization exceeds the 30 mbps aggregate of tier (2), the same heuristic is used to determine the next provider to bear an increase on tier (3):

provider 1: $100 → $250 = $150 increase
provider 2: $200 → $350 = $150 increase
provider 3: $200 → $300 = $100 increase

In this example, provider 3 will be the next link utilized. Provider 3 is utilized to its full capacity at this cost level, which is 40 mbps.

Tier 4: In this example, at tier (4), there is a tie among the cost increments:

provider 1: $100 → $250 = $150 increase
provider 2: $200 → $350 = $150 increase
provider 3: $300 → $450 = $150 increase

In such a case, in some embodiments, the provider that provides the most capacity at the next billing level is selected.

In this example, Provider 2's cost remains at this cost level from 15 mbps – 45 mbps, which is the longest duration of the three.

Tier 5: In this example, at tier (5), Provider 1 is selected using the same logic as tier (4).

Tier 6: In this example, note that although provider 1 is again selected at tier (6), this tier is not combined with tier (5).

Tier 7: In this example, the last tier represents the full link capacity of each provider.

Adjusting can be done automatically to a subset of the forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing, wherein:

- at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a subset of two or more links in the network,
- the adjusting attempts to reduce the degree of suboptimality

"Automatic" adjustment may mean that human intervention may not be required prior to completing a change of forwarding decision.

In some embodiments of the invention, systems are included to prevent flapping that could incur from repeated adjustments of forwarding decisions. In some embodiments, a minimum limit can be imposed on the interval separating consecutive reevaluations of one or more of their first and second degrees of unacceptability for an object. In embodiments of the invention in which the second degree of unacceptability for an object includes the states "hot" and "cold", a minimum limit can be imposed on the interval separating consecutive hot/cold reevaluations. (In the context of this document, we denote the minimum time to reevaluate degrees of unacceptability the "reevaluation interval" for the object.) In some embodiments of this invention, the reevaluation interval can be chosen randomly with respect to some probability distribution function. In some embodiments of the invention, the reevaluation interval is chosen as to be larger than the minimum interval between two consecutive monitoring actions. In some such embodiments in which the second degree of unacceptability includes the states "hot" and "cold", the probability distribution functions in respect to which the reevaluation interval are computed can be chosen differently for hot to cold transitions, and cold to hot transitions, respectively. In some such embodiments, the probability distribution function for cold to hot transitions has a lower median than the probability distribution function for hot to cold transitions.

In some embodiments of the invention, the probability distribution function with respect to which the reevaluation interval is computed can include an exponential distribution function. In some embodiments, a minimum limit can be imposed on the range of values that is allowed by the distribution. In some embodiments, a maximum limit can be imposed on the range of values allowed by the distribution.

In some embodiments of this invention, the subset of two or more forwarding decisions in the network that are to be adjusted automatically does not consist of all

forwarding decisions. Load often varies randomly in unpredictable ways. Computing a target that provides an optimal solution to the problem, and adjusting the forwarding decisions to meet this target seldom leads to the optimal solution, because the conditions at the time when the target was computed, and at the time the forwarding decisions were adjusted are not the same.

Therefore, in some embodiments of this invention, the incremental approach is used, wherein a subset of the forwarding decisions are selected for adjustment at any one time. In some embodiments, continuously monitoring and assessing, and continuously adjusting in an incremental fashion a subset of the forwarding decisions allows for stable load movements towards the optimal load distribution.

In some embodiments of this invention, the subset of the forwarding decisions of one or more forwarding nodes is done automatically.

In some embodiments of this invention, the selecting of the subset of the forwarding decisions is random

In some embodiments, the selecting of the subset of the forwarding decisions is independent from the assessing.

In some embodiments, the selecting of the subset of the forwarding decisions uses a flow monitoring device

In some embodiments of this invention, at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application; in some such embodiments, the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the third subset of two or more links. In some such embodiments, the selecting of the subset of the forwarding decisions is based at least partly on a measuring of the quality characterizations of the one or more objects.

In some embodiments, the selecting of the subset of the forwarding decisions is based at least partly on a source external to the third subset of two or more links.

In some embodiments of this invention, the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 3 Protocol

In some embodiments of this invention, at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Internet Protocol (IP).

In some embodiments of this invention, the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 2 Protocol

In some embodiments of this invention, the adjusting is described at least partly by at least one Border Gateway Protocol (BGP)

In some embodiments of this invention, the adjusting is described at least partly by Border Gateway Protocol (BGP) Version 1

In some embodiments of this invention, the adjusting is described at least partly by Border Gateway Protocol (BGP) Version 2

In some embodiments of this invention, the adjusting is described at least partly by Border Gateway Protocol (BGP) Version 3

In some embodiments of this invention, the adjusting is described at least partly by Border Gateway Protocol (BGP) Version 4

What is claimed is:

1. Computer code reducing the monetary cost of operating a network, comprising:
code that performs monitoring of at least a first utilization of a first subset of two or more links in the network;
code that performs assessing, based at least partly on the monitoring, of a degree of suboptimality with respect to one or more monetary billing structures of a second subset of two or more links in the network;
wherein at least one of the one or more monetary billing structures receives as input at least a second utilization of the second subset of two or more links; and
at least one of the one or more monetary billing structures includes at least variable cost; and
code that performs adjusting, automatically, of a subset of forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing;
wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a third subset of two or more links in the network; and
the adjusting attempts to reduce the degree of suboptimality.

2. The computer code of claim 1, wherein the first utilization of the first subset of two or more links is at least partly indicative of the second utilization of the second subset of two or more links.

3. The computer code of claim 1, wherein at least one link from the first subset of two or more links is included in at least one of: 1) the second subset of two or more links and 2) the third subset of two or more links.

4. The computer code of claim 1, wherein at least one link from the second subset of two or more links is included in the third subset of two or more links.

5. The computer code of claim 1, wherein at least one link from the first subset of two or more links is not included in at least one of: 1) the second subset of two or more links and 2) the third subset of two or more links.

6. The computer code of claim 1, wherein at least one link from the second subset of two or more links is not included in the third subset of two or more links.

7. The computer code of claim the computer code of claim 1, wherein at least one of the monitoring, the assessing, and the adjusting repeats.

8. The computer code of claim 1, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 3 Protocol.

9. The computer code of claim 8, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Internet Protocol (IP).

10. The computer code of claim 1, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 2 Protocol.

11. The method of claim 1, wherein the adjusting is described at least partly by at least one Border Gateway Protocol (BGP).

12. The computer code of claim 1, wherein at least one of the one or more monetary billing structures is for at least one Internet Service Provider (ISP).

13. The computer code of claim 1, wherein each link of at least one link from the second subset of two or more links has a third utilization, and at least one of the one or more monetary billing structures receives as input at least the third utilization.

14. The computer code of claim 13, wherein the second utilization and the third utilization are equal.

15. The computer code of claim 13, wherein the second utilization and the third utilization are unequal.

16. The computer code of claim 13, wherein the third utilization is being determined over time.

17. The computer code of claim 16, wherein the third utilization is computed at least partly from:

at least one of: 1a) a maximum and 1b) an average;

of at least one of: 2a) one or more percentiles and 2b) one or more averages; and

of one or more sets of utilization samples of the at least one link from the second subset of two or more links.

18. The computer code of claim 16, wherein the at least one of the one or more monetary billing structures is continuous or piecewise continuous with respect to the third utilization.

19. The computer code of claim 1, wherein the monitoring uses one or more of Simple Network Monitoring Protocol (SNMP), flow information export, NetFlow, span port, and a source external to the first subset of two or more links.

20. The computer code of claim 13, wherein the assessing includes generating, from the at least one of the one or more monetary billing structures, one or more sets of functions, wherein at least one function in the one or more sets of functions gives a first degree of unacceptability of at least one link from the first subset of two or more links, wherein the first degree of unacceptability is based at least partly on a fourth utilization of the at least one link from the first subset of two or more links.

21. The computer code of claim 20, wherein the first utilization and the fourth utilization are equal.

22. The computer code of claim 20, wherein the first utilization and the fourth utilization are unequal.

23. The computer code of claim 20, wherein the generating includes:

compiling a list of at least two sums, wherein at least one sum of the list adds at least two of the third utilizations;

determining, for a subset of the list, a utilization distribution based at least partly on the at least one of the one or more monetary billing structures; and

constructing the one or more sets of functions based at least partly on the utilization distribution.

24. The computer code of claim 23, wherein the utilization distribution minimizes a monetary cost of operating the network, with respect to the at least one of the one or more monetary billing structures.

25. The computer code of claim 23, wherein the utilization distribution uses at least a steepest descent strategy with respect to the at least one of the one or more monetary billing structures.

26. The computer code of claim 20, wherein the at least one function in the one or more sets of functions outputs at least a varying value.

27. The computer code of claim 26, wherein the at least one function in the one or more sets of functions is continuous or piecewise continuous with respect to the fourth utilization.

28. The computer code of claim 26, wherein the at least one function in the one or more sets of functions is non-decreasing with respect to the fourth utilization.

29. The computer code of claim 26, wherein the at least one function in the one or more sets of functions receives at least one input, the at least one input at least partly depending on the fourth utilization, wherein the at least one function outputs at least

1) a first constant value for values of the at least one input up to a threshold value; and

2) a second constant value for values of the at least one input above the threshold value.

30. The computer code of claim 26, wherein the at least one function in the one or more sets of functions receives at least one input, the at least one input at least partly depending on the fourth utilization, wherein the at least one function outputs at least:

1) a first constant value for values of the at least one input ranging from a second constant value to a third constant value;

2) a linear function of at least one input for values of the at least one input ranging from the third constant value to a fourth constant value; and

3) a fifth constant value for values of the at least one input exceeding the fourth constant value.

31. The computer code of claim 20, wherein the assessing includes:

selecting a first set of functions from the one or more sets of functions, wherein at least one function in the first set of functions gives the first degree of unacceptability; and

selecting a second set of functions from the one or more sets of functions if:

1) the one or more sets of functions includes at least two sets of functions; and

2) for each function in the first set of functions that gives the first degree of unacceptability, the first degree of unacceptability fails a first threshold of acceptable unacceptability.

32. The computer code of claim 20, wherein the adjusting includes attempting to reduce the degree of suboptimality based at least partly on the first degree of unacceptability.

33. The computer code of claim 31, wherein the assessing further includes determining a second degree of unacceptability based at least partly on the first degree of unacceptability.

34. The computer code of claim 33, wherein the determining of the second degree of unacceptability includes treating the first degree of unacceptability as a probability value, and assigning, using the probability value, one of a plurality of states to the second degree of unacceptability.

35. The computer code of claim 31, further comprising:

ordering the one or more sets of functions into an ordered list of the one or more sets of functions; and

wherein the first set of functions and the second set of functions are adjacent in the ordered list of the one or more sets of functions.

36. The computer code of claim 35, wherein:

at least one function in the one or more sets of functions receives at least one input, the at least one input at least partly depending on the fourth utilization, wherein the at least one function outputs at least:

1) a first constant value for values of the at least one input ranging from a second constant value to a third constant value;

2) a linear function of at least one input for values of the at least one input ranging from the third constant value to a fourth constant value; and

3) a fifth constant value for values of the at least one input exceeding the fourth constant value, further comprising:

computing, for each set of functions in the one or more sets of functions, a level, wherein the level is based at least partly on a sum of at least the fourth constant values across the one or more functions in each set of functions; and

wherein performing the ordering is based at least partly on the level computed for each set of functions.

37. The computer code of claim 35, wherein the sum of at least the fourth constant values across the one or more functions in each set of functions, sums at least one function of the one or more functions in each set of functions.

38. The computer code of claim 35, wherein the sum of the fourth constant values across the one or more functions in each set of functions, sums all functions of the one or more functions in each set of functions.

39. The computer code of claim 32, wherein the adjusting includes attempting to reduce the degree of suboptimality by changing at least one forwarding decision from the subset of the forwarding decisions:

wherein prior to the changing, the at least one forwarding decision from the subset of the forwarding decisions points to at least a first link from the third subset of two or more links in the network;

wherein after the changing, the at least one forwarding decision from the subset of the forwarding decisions points to at least a second link from the third subset of two or more links in the network; and

wherein the first degree of unacceptability of the at least the first link from the third subset is more unacceptable than the first degree of unacceptability of the at least the second link from the third subset.

40. The computer code of claim 1, wherein at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application.

41. The computer code of claim 40, wherein the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the third subset of two or more links.

42. The computer code of claim 20,

wherein at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application;

the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the third subset of two or more links; and

the assessing includes:

selecting at least one object from the one or more objects;

selecting at least one set of functions from the one or more sets of functions; and

constructing one or more winner sets for the at least one object and the least one set of functions, wherein each winner set from the one or more

winner sets includes a corresponding quality characterization threshold, wherein the constructing includes:

1) including in at least one of the one or more winner sets one or more links from the third subset of two or more links;

2) excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets; and

3) excluding, from the at least one or more winner sets, unwanted links, wherein the unwanted links have a third degree of unacceptability failing a second threshold of acceptable unacceptability, wherein the third degree of unacceptability is based at least partly on the first degree of unacceptability given by the at least one set of functions;

selecting one or more links from a non-empty winner set from the one or more winner sets, wherein the non-empty winner set has a low corresponding quality characterization threshold from all corresponding quality characterization thresholds included by all winner sets from the one or more winner sets.

43. The computer code of claim 42, wherein the first threshold of acceptable unacceptability and the second threshold of acceptable unacceptability are equal.

44. The computer code of claim 42, wherein the first threshold of acceptable unacceptability and the second threshold of acceptable unacceptability are unequal.

45. The computer code of claim 42, wherein the low corresponding quality characterization threshold is the lowest corresponding quality characterization threshold from all corresponding quality characterization thresholds included by all winner sets from the one or more winner sets.

46. The computer code of claim 42:

wherein the constructing of a first one or more winner sets is done for a third set of functions from the one or more sets of functions; and

the constructing of a second one or more winner sets is done for a fourth set of functions from the one or more sets of functions if;

1) the one or more sets of functions includes at least two sets of functions, and

2) all of the first one or more winner sets are empty.

47. The computer code of claim 42:

wherein the constructing of a first one or more winner sets is done for a first object from the one or more objects; and

the constructing of a second one or more winner sets is done for a second object from the one or more objects if:

1) the one or more objects includes at least two objects, and

2) all of the first one or more winner sets are empty.

48. The computer code of claim 42, wherein the excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets, is further comprised of:

identifying at least one best link from the one or more links from the third subset of two or more links, wherein the at least one best link has a high quality characterization from at least one of the one or more links from the third subset of two or more links; and

determining the corresponding quality characterization threshold based at least partly on the highest quality characterization.

49. The computer code of claim 48, wherein the high quality characterization is the highest quality characterization from the at least one of the one or more links from the third subset of two or more links.

50. The computer code of claim 1, further including selecting the subset of the forwarding decisions of one or more forwarding nodes automatically.

51. The computer code of claim 50, wherein the selecting of the forwarding decisions is at least partly random

52. The computer code of claim 50, wherein selecting the subset of the forwarding decisions is independent from the assessing.

53. The computer code of claim 50, wherein the selecting of the subset of the forwarding decisions uses a flow monitoring device.

54. The computer code of claim 50:
wherein at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application;
the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the third subset of two or more links; and
the selecting of the subset of the forwarding decisions is based at least partly on measuring the quality characterizations of the one or more objects.

55. The computer code of claim 50, wherein the selecting of the subset of the forwarding decisions is based at least partly on a source external to the third subset of two or more links.

56. The computer code of claim 1, wherein the computer code is at least partly software.

57. The computer code of claim 1, wherein the computer code is all software.

58. The computer code of claim 1, wherein the computer code is at least partly hardware.

59. The computer code of claim 1, wherein the computer code is all hardware.

60. Computer code that attempts to ensure a desired load distribution in a network, comprising:

code that performs monitoring of at least a first utilization of a first subset of two or more links in the network;
code that performs assessing, based at least partly on the monitoring, of a degree of suboptimality with respect to the desired load distribution, the assessing including:

generating at least two sets of functions; and

selecting a first set of functions from the at least two sets of functions:

wherein at least one function from the first set of functions gives a first degree of unacceptability of at least one link from the first subset of two or more links,

wherein the first degree of unacceptability is based at least partly on a second utilization of the at least one link from the first subset of two or more links; and

the at least one function in the first set of functions outputs at least a varying value;

code that performs selecting of a second set of functions from the at least two sets of functions if, for each function in the first set of functions that gives the first degree of unacceptability, the first degree of unacceptability fails a first threshold of acceptable unacceptability; and

code that performs adjusting, automatically, of a subset of forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing:

wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a second subset of two or more links in the network; and

the adjusting includes code that attempts to reduce the degree of suboptimality.

61. The computer code of claim 60, wherein the first utilization and the second utilization are equal.

62. The computer code of claim 60, wherein the first utilization and the second utilization are unequal.

63. The computer code of claim 60, wherein at least one link from the first subset of two or more links is included in the second subset of two or more links.

64. The computer code of claim 60, wherein at least one link from the first subset of two or more links is not included in the second subset of two or more links.

65. The computer code of claim 60, wherein at least one of the monitoring, the assessing, and the adjusting repeats.

66. The computer code of claim 60, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 3 Protocol.

67. The computer code of claim 66, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Internet Protocol (IP).

68. The computer code of claim 60, wherein at least one of the forwarding decisions of the one or more forwarding nodes are described at least partly by at least one Layer 2 Protocol.

69. The computer code of claim 60, wherein the adjusting is described at least partly by at least one Border Gateway Protocol (BGP).

70. The computer code of claim 60, wherein the at least two sets of functions are generated from one or more monetary billing structures of a third subset of two of more links in the network.

71. The computer code of claim 70, wherein at least one link from the third subset of two or more links is included in at least one of: 1) the first subset of two or more links and 2) the second subset of two or more links.

72. The computer code of claim 1, wherein at least one link from the third subset of two or more links is not included in at least one of: 1) the first subset of two or more links and 2) the second subset of two or more links.

73. The computer code of claim 70, wherein at least one of the one or more monetary billing structures is for at least one Internet Service Provider (ISP).

74. The computer code of claim 70, wherein each link of at least one link from the third subset of two or more links has a third utilization, and at least one of the one or more monetary billing structures receives as input at least the third utilization.

75. The computer code of claim 74, wherein the first utilization of the first subset of two or more links is at least partly indicative of the third utilization of the third subset of two or more links.

76. The computer code of claim 74, wherein the third utilization is being determined over time.

77. The computer code of claim 352, wherein the third utilization is computed at least partly from
at least one of: 1a) a maximum and 1b) an average;
of at least one of: 2a) one or more percentiles and 2b) one or more averages; and
of one or more sets of utilization samples of the at least one link from the third subset of two or more links.

78. The computer code of claim 352, wherein the at least one of the one or more monetary billing structures is continuous or piecewise continuous with respect to the third utilization

79. The computer code of claim 60, wherein the monitoring uses one or more of Simple Network Monitoring Protocol (SNMP), flow information export, NetFlow, span port, and a source external to the first subset of two or more links.

80. The computer code of claim 74, wherein the generating includes:
compiling a list of at least two sums, wherein at least one sum of the list adds at least two of the third utilizations;
determining, for a subset of the list, a utilization distribution based at least partly on the at least one of the one or more monetary billing structures; and
constructing the at least two set of functions based at least partly on the utilization distribution.

81. The computer code of claim 80, wherein the utilization distribution minimizes a monetary cost of operating the network, with respect to the at least one of the one or more monetary billing structures.

82. The computer code of claim 80, wherein the utilization distribution uses at least a steepest descent strategy with respect to the at least one of the one or more monetary billing structures.

83. The computer code of claim 60, wherein at least one function in the at least two sets of functions is continuous or piecewise continuous with respect to the second utilization.

84. The computer code of claim 60, wherein at least one function in the at least two sets of functions is non-decreasing with respect to the second utilization.

85. The computer code of claim 60, wherein at least one function in the at least two sets of functions receives at least one input, the at least one input at least partly depending on the second utilization, wherein the at least one function outputs at least:
1) a first constant value for values of the at least one input up to a threshold value; and
2) a second constant value for values of the at least one input above the threshold value.

86. The computer code of claim 60, wherein at least one function in the at least two set of functions receives at least one input, the at least one input at least partly depending on the second utilization, wherein the at least one function outputs at least:
1) a first constant value for values of the at least one input ranging from a second constant value to a third constant value,
2) a linear function of at least one input for values of the at least one input ranging from the third constant value to a fourth constant value, and
3) a fifth constant value for values of the at least one input exceeding the fourth constant value.

87. The computer code of claim 60, wherein the adjusting includes attempting to reduce the degree of suboptimality based at least partly on the first degree of unacceptability.

88. The computer code of claim 60, wherein the assessing includes determining a second degree of unacceptability based at least partly on the first degree of unacceptability.

89. The computer code of claim 87, wherein the determining of the second degree of unacceptability includes treating the first degree of unacceptability as a probability value, and assigning, using the probability value, one of a plurality of states to the second degree of unacceptability.

90. The computer code of claim 60, further comprising:
ordering the one or more sets of functions into an ordered list of the one or more sets of functions; and
wherein the first set of functions and the second set of functions are adjacent in the ordered list of the one or more sets of functions.

91. The computer code of claim 90, wherein:
at least one function in the one or more sets of functions receives at least one input, the at least one input at least partly depending on the second utilization, wherein the at least one function outputs at least:

1) a first constant value for values of the at least one input ranging from a second constant value to a third constant value;

2) a linear function of at least one input for values of the at least one input ranging from the third constant value to a fourth constant value; and

3) a fifth constant value for values of the at least one input exceeding the fourth constant value, further comprising:

computing, for each set of functions in the one or more sets of functions, a level, wherein the level is based at least partly on a sum of at least the fourth constant values across the one or more functions in each set of functions; and

performing the ordering based at least partly on the level computed for each set of functions.

92. The computer code of claim 90, wherein the sum of at least the fourth constant values across the one or more functions in each set of functions, sums at least one function of the one or more functions in each set of functions.

93. The computer code of claim 90, wherein the sum of the fourth constant values across the one or more functions in each set of functions, sums all functions of the one or more functions in each set of functions.

94. The computer code of claim 87, wherein the adjusting further includes attempting to reduce the degree of suboptimality by changing at least one forwarding decision from the subset of the forwarding decisions, wherein:
prior to the changing, the at least one forwarding decision from the subset of the forwarding decisions points to at least a first link from the second subset of two or more links in the network;
after the changing, the at least one forwarding decision from the subset of the forwarding decisions points to at least a second link from the second subset of two or more links in the network; and
wherein the first degree of unacceptability of the at least the first link from the second subset is more unacceptable than the first degree of unacceptability of the at least the second link from the second subset.

95. The computer code of claim 60, wherein at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application.

96. The computer code of claim 95, wherein the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the second subset of two or more links.

97. The computer code of claim 95, wherein the assessing further includes:
selecting at least one object from the one or more objects;
selecting at least one set of functions from the one or more sets of functions; and
constructing one or more winner sets for the at least one object and the least one set of functions, wherein each winner set from the one or more winner sets includes a corresponding quality characterization threshold, wherein the constructing includes:

1) including in at least one of the one or more winner sets one or more links from the second subset of two or more links;

2) excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets; and

3) excluding, from the at least one or more winner sets, unwanted links, wherein the unwanted links have a third degree of unacceptability failing a second threshold of acceptable unacceptability, wherein the third degree of unacceptability is based at least partly on the first degree of unacceptability given by the at least one set of functions;

selecting one or more links from a non-empty winner set from the one or more winner sets, wherein the non-empty winner set has a low corresponding quality characterization threshold from all corresponding quality characterization thresholds included by all winner sets from the one or more winner sets.

98. The computer code of claim 97, wherein the first threshold of acceptable unacceptability and the second threshold of acceptable unacceptability are equal.

99. The computer code of claim 97, wherein the first threshold of acceptable unacceptability and the second threshold of acceptable unacceptability are unequal.

100. The computer code of claim 97, wherein the low corresponding quality characterization threshold is the lowest corresponding quality characterization threshold from all corresponding quality characterization thresholds included by all winner sets from the one or more winner sets.

101. The computer code of claim 97, wherein:

the constructing of a first one or more winner sets is done for a third set of functions from the one or more sets of functions; and

the constructing of a second one or more winner sets is done for a fourth set of functions from the one or more sets of functions if:

1) the one or more sets of functions includes at least two sets of functions; and

2) all of the first one or more winner sets are empty.

102. The computer code of claim 97:

wherein the constructing of a first one or more winner sets is done for a first object from the one or more objects; and

the constructing of a second one or more winner sets is done for a second object from the one or more objects if:

1) the one or more objects includes at least two objects, and

2) all of the first one or more winner sets are empty.

102. The computer code of claim 97, wherein the excluding, from the at least one or more winner sets, links for which the quality characterizations of the at least one object fails the corresponding quality characterization threshold included by each winner set from the one or more winner sets is further comprised of:

identifying at least one best link from the one or more links from the second subset of two or more links, wherein the at least one best link has a high quality characterization from at least one of the one or more links from the second subset of two or more links, and

determining the corresponding quality characterization threshold based at least partly on the highest quality characterization.

104. The computer code of claim 103, wherein the high quality characterization is the highest quality characterization from the at least one of the one or more links from the second subset of two or more links.

105. The computer code of claim 60, further including selecting the subset of the forwarding decisions of one or more forwarding nodes automatically.

106. The computer code of claim 105, wherein the selecting of the subset of the forwarding decisions is at least partly random.

107. The computer code of claim 105, wherein the selecting of the subset of the forwarding decisions is independent from the assessing by the code that assesses.

108. The computer code of claim 105, wherein the selecting of the subset of the forwarding decisions uses a flow monitoring device.

109. The computer code of claim 105:
wherein at least one forwarding decision from the subset of the forwarding decisions at least partly influences one or more objects, wherein the one or more objects includes at least one of a prefix, a flow, and a network application;
the assessing is further based at least partly on quality characterizations of the one or more objects, wherein the quality characterizations are with respect to at least one link from the second subset of two or more links; and
the selecting of the subset of the forwarding decisions is based at least partly on a measuring of the quality characterizations of the one or more objects.

110. The computer code of claim 105, wherein the selecting of the subset of the forwarding decisions is based at least partly on a source external to the second subset of two or more links.

111. The computer code of claim 60, wherein the computer code is at least partly software.

112. The computer code of claim 60, wherein the computer code is all software.

113. The computer code of claim 60, wherein the computer code is at least partly hardware.

114. The computer code of claim 60, wherein the computer code is all hardware.

115. Computer code for reducing the monetary cost of operating a network, comprising:

means for monitoring at least a first utilization of a first subset of two or more links in the network;
means for assessing, based at least partly on the monitoring, a degree of suboptimality with respect to one or more monetary billing structures of a second subset of two or more links in the network;

wherein at least one of the one or more monetary billing structures receives as input at least a second utilization of the second subset of two or more links; and

at least one of the one or more monetary billing structures includes variable cost means for adjusting automatically a subset of the forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing;

wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a third subset of two or more links in the network, the means for adjusting attempts to reduce the degree of suboptimality

116. Computer code for ensuring a desired load distribution in a network, the method comprising:

means for monitoring at least a first utilization of a first subset of two or more links in the network;

means for assessing, based at least partly on the means for monitoring, a degree of suboptimality with respect to the desired load distribution, the means for assessing including:

means for generating a list of at least two sets of functions;

means for selecting a first set of functions from the list of at least two sets of functions:

wherein at least one function from the first set of functions gives a first degree of unacceptability of at least one link from the first subset of two or more links, wherein the first degree of unacceptability is based at least partly on a second utilization of the at least one link from the first subset of two or more links, and

at least one function in the first set of functions outputs at least a varying value, and

means for selecting a second set of functions from the at least two sets of functions if:

1) at least one function in the first set of functions gives the first degree of unacceptability; and

2) for each function in the first set of functions that gives the first degree of unacceptability, the first degree of unacceptability fails a first threshold of acceptable unacceptability.

means for adjusting automatically a subset of the forwarding decisions of one or more forwarding nodes in the network based at least partly on the means for assessing:

wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a second subset of two or more links in the network, the means for adjusting attempts to reduce the degree of suboptimality.

117. A method of reducing the monetary cost of operating a network, comprising:

monitoring at least a first utilization of a first subset of two or more links in the network;

assessing, based at least partly on the monitoring, a degree of suboptimality with respect to one or more monetary billing structures of a second subset of two or more links in the network;

wherein at least one of the one or more monetary billing structures receives as input at least a second utilization of the second subset of two or more links; and

at least one of the one or more monetary billing structures includes at least variable cost; and

adjusting automatically a subset of forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing;

wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a third subset of two or more links in the network; and

the adjusting attempts to reduce the degree of suboptimality.

118. A method that attempts to ensure a desired load distribution in a network, comprising:

monitoring at least a first utilization of a first subset of two or more links in the network;

assessing, based at least partly on the monitoring, a degree of suboptimality with respect to the desired load distribution, the assessing including:

generating at least two sets of functions; and

selecting a first set of functions from the at least two sets of functions:

wherein at least one function from the first set of functions gives a first degree of unacceptability of at least one link from the first subset of two or more links, wherein the first degree of unacceptability is based at least partly on a second utilization of the at least one link from the first subset of two or more links; and

the at least one function in the first set of functions outputs at least a varying value;

selecting a second set of functions from the at least two sets of functions if, for each function in the first set of functions that gives the first degree of unacceptability, the first degree of unacceptability fails a first threshold of acceptable unacceptability; and

adjusting automatically a subset of forwarding decisions of one or more forwarding nodes in the network based at least partly on the assessing:

wherein at least one forwarding decision from the subset of the forwarding decisions points to at least one link from a second subset of two or more links in the network; and

the adjusting includes code that attempts to reduce the degree of suboptimality.

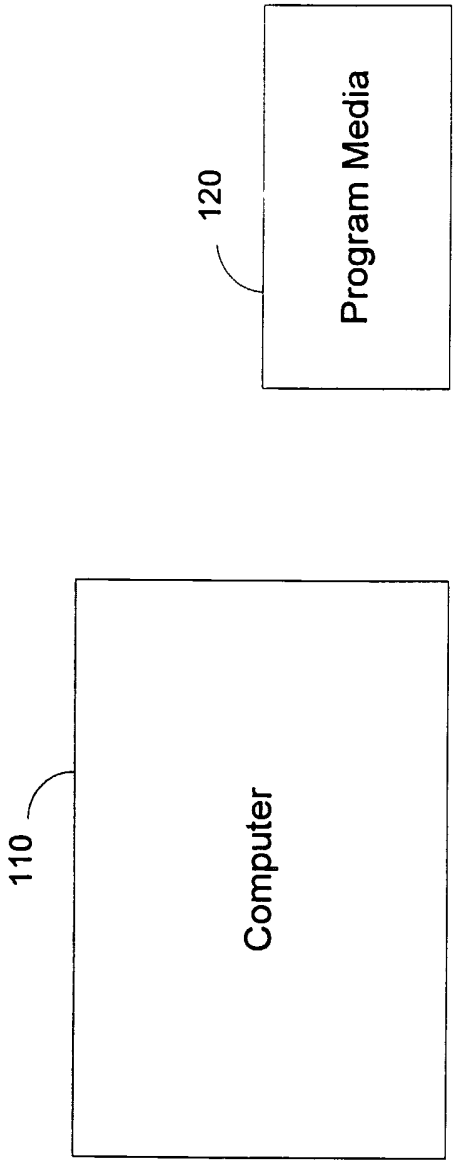


FIG. 1

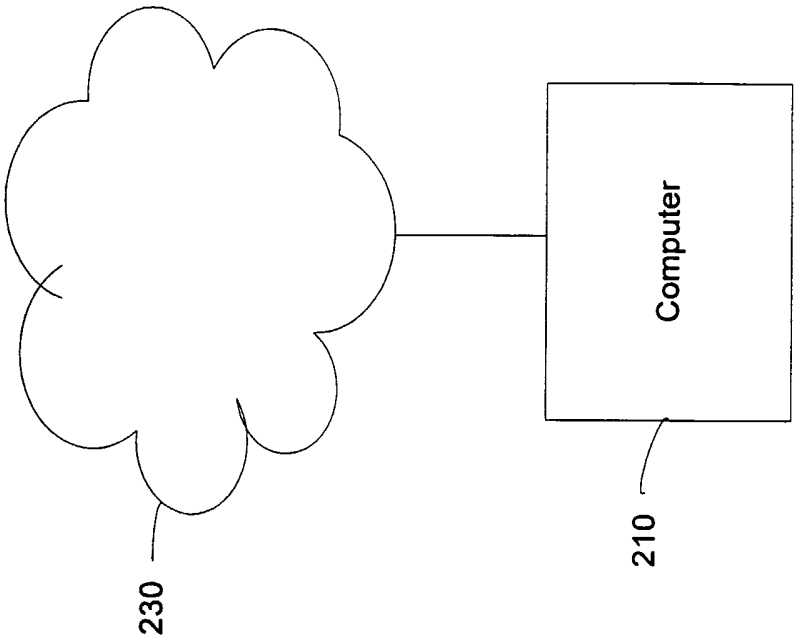


FIG. 2

FIGURE 3

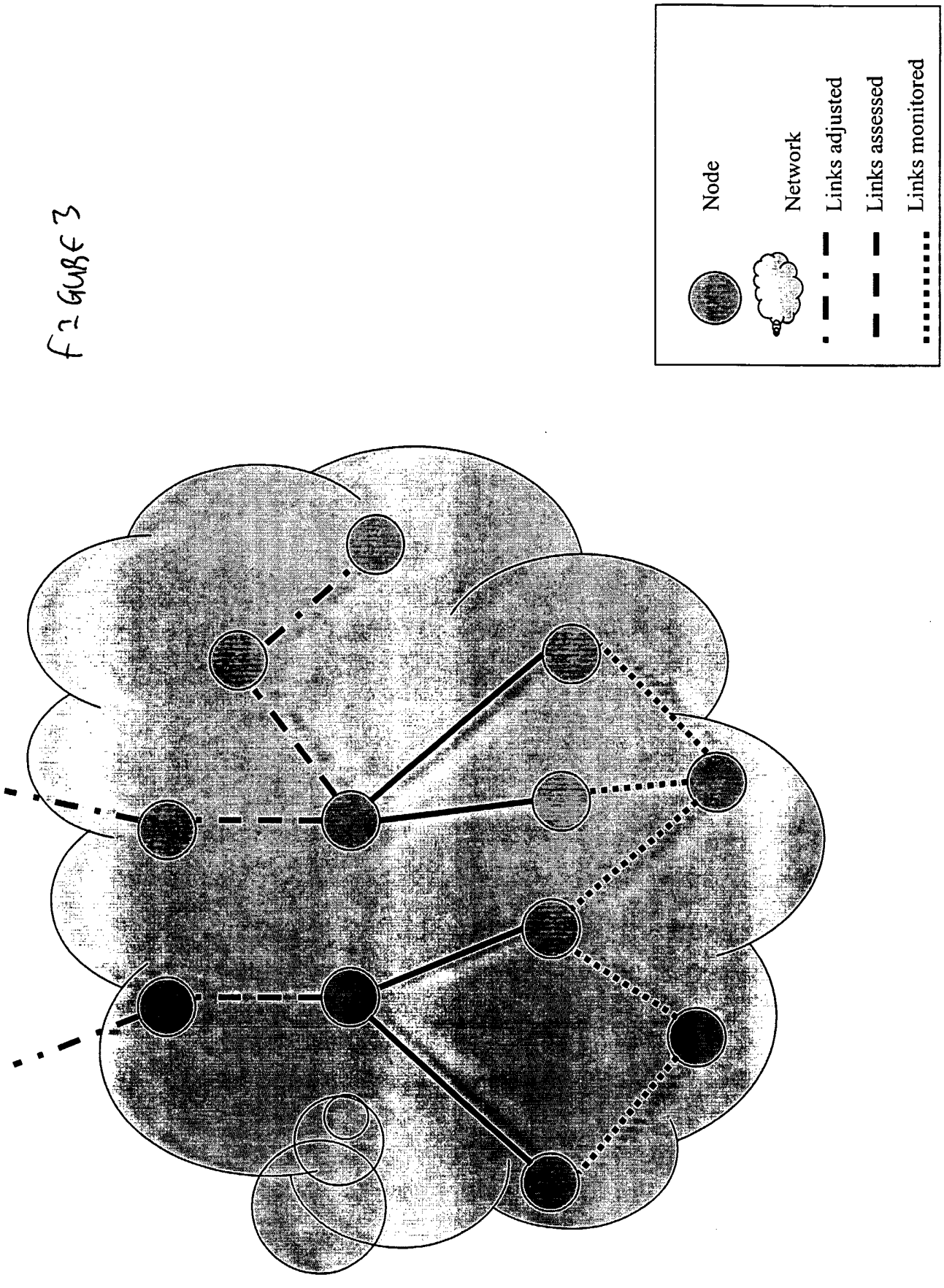

Node
Network
Links adjusted
Links assessed
Links monitored

FIGURE 4

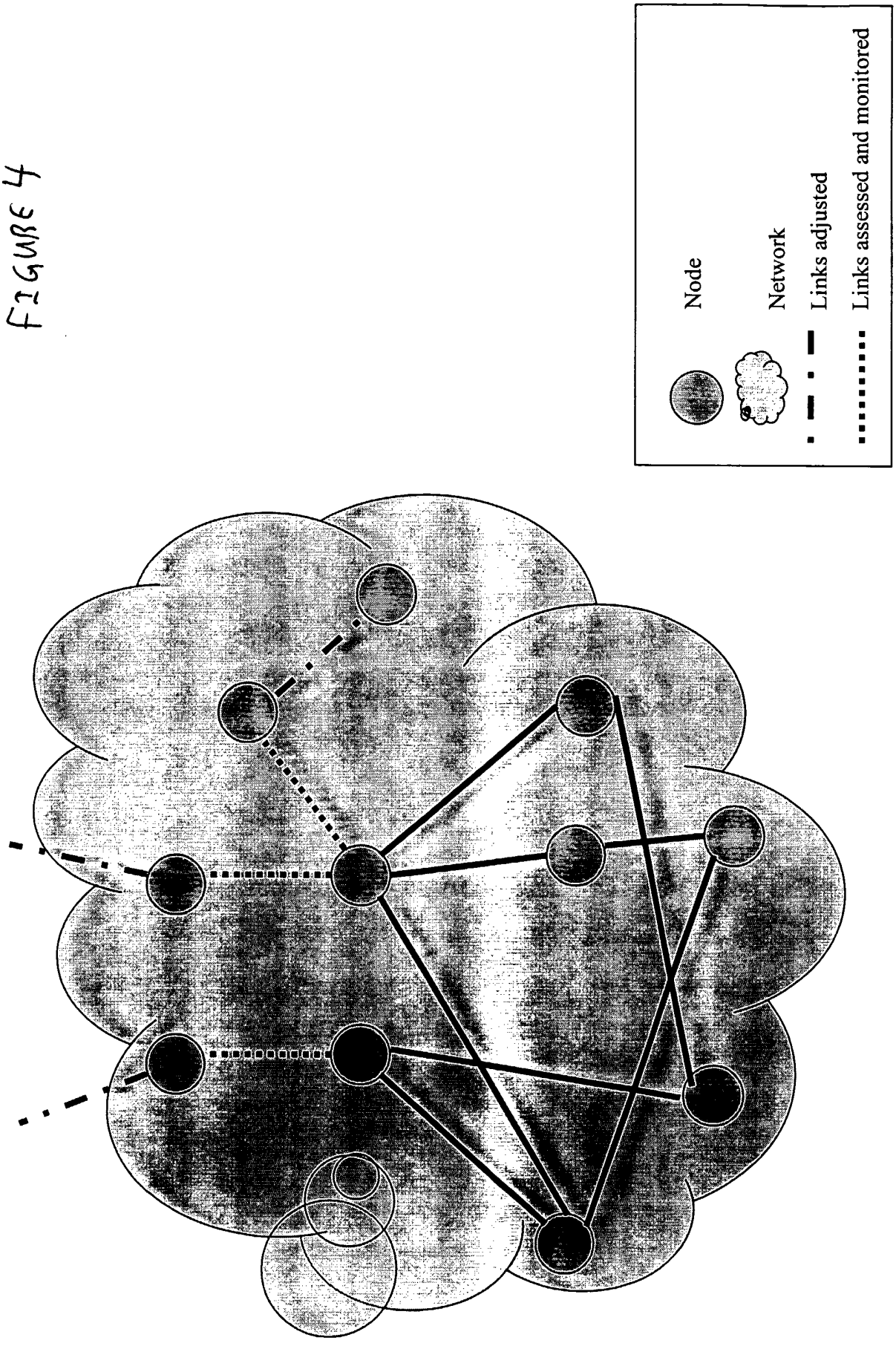


FIGURE 5

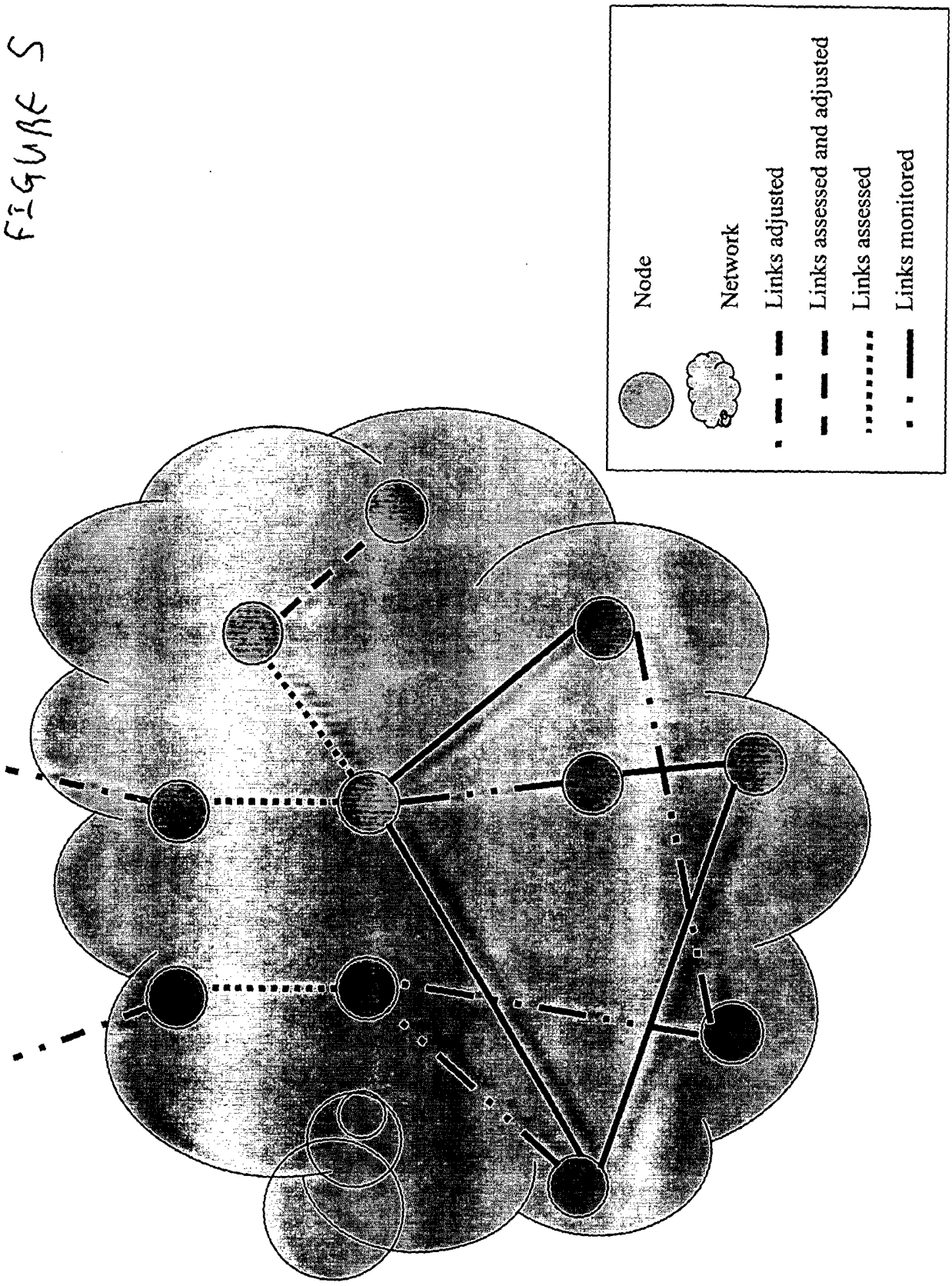
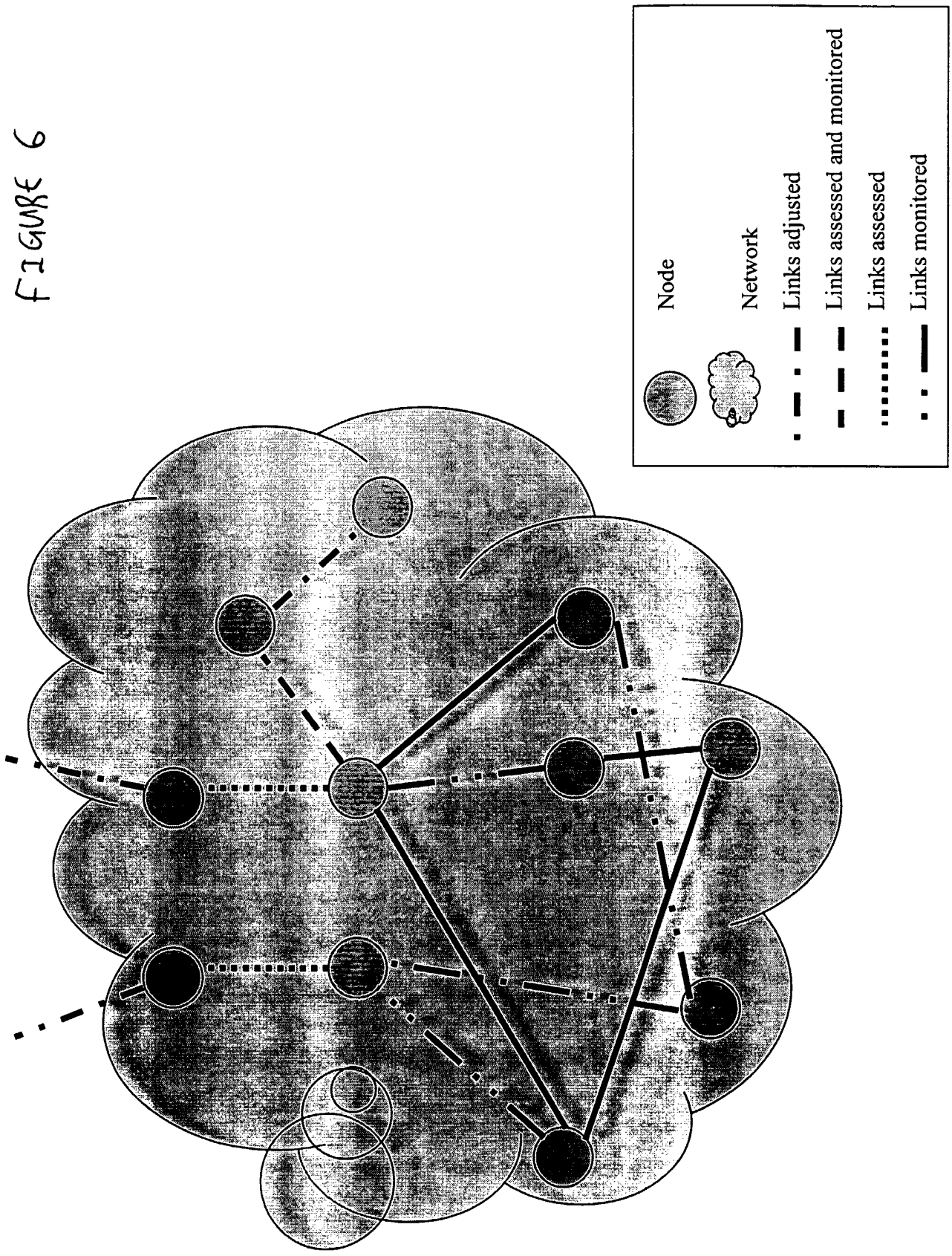

Node
Network
Links adjusted
Links assessed and adjusted
Links assessed
Links monitored

FIGURE 6


Node
Network
Links adjusted
Links assessed and monitored
Links assessed
Links monitored

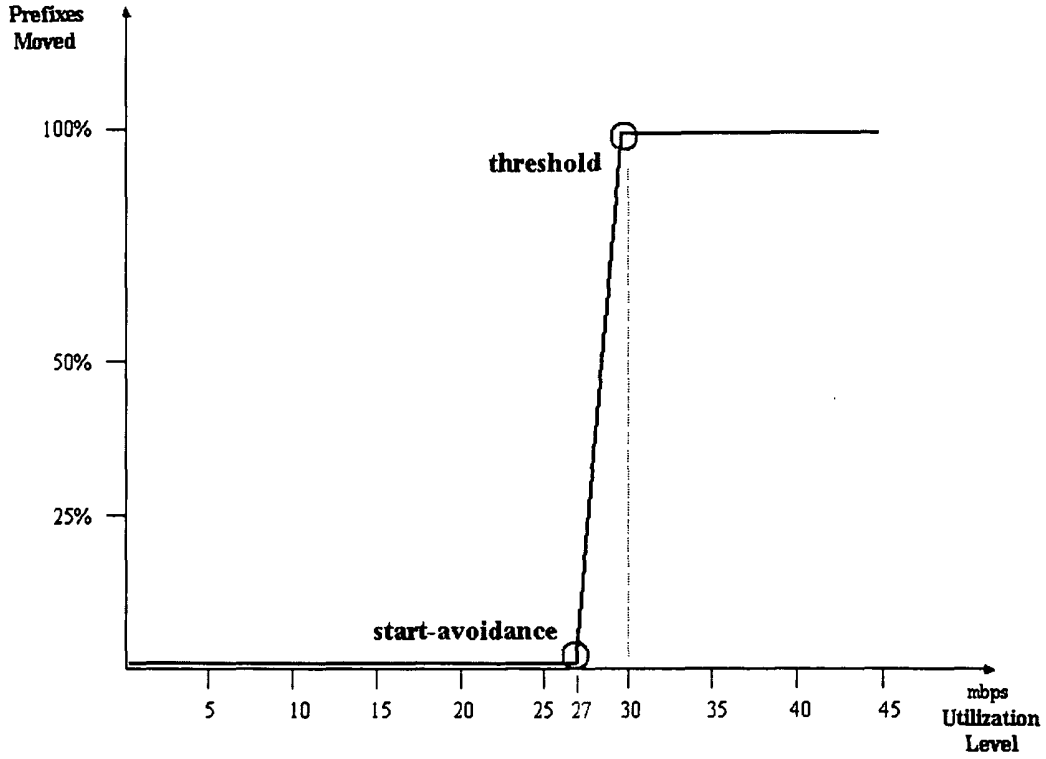

Prefixes Moved
100%
50%
25%
threshold
start-avoidance
5
10
15
20
25
27
30
35
40
45
mbps Utilization Level

FIGURE 7

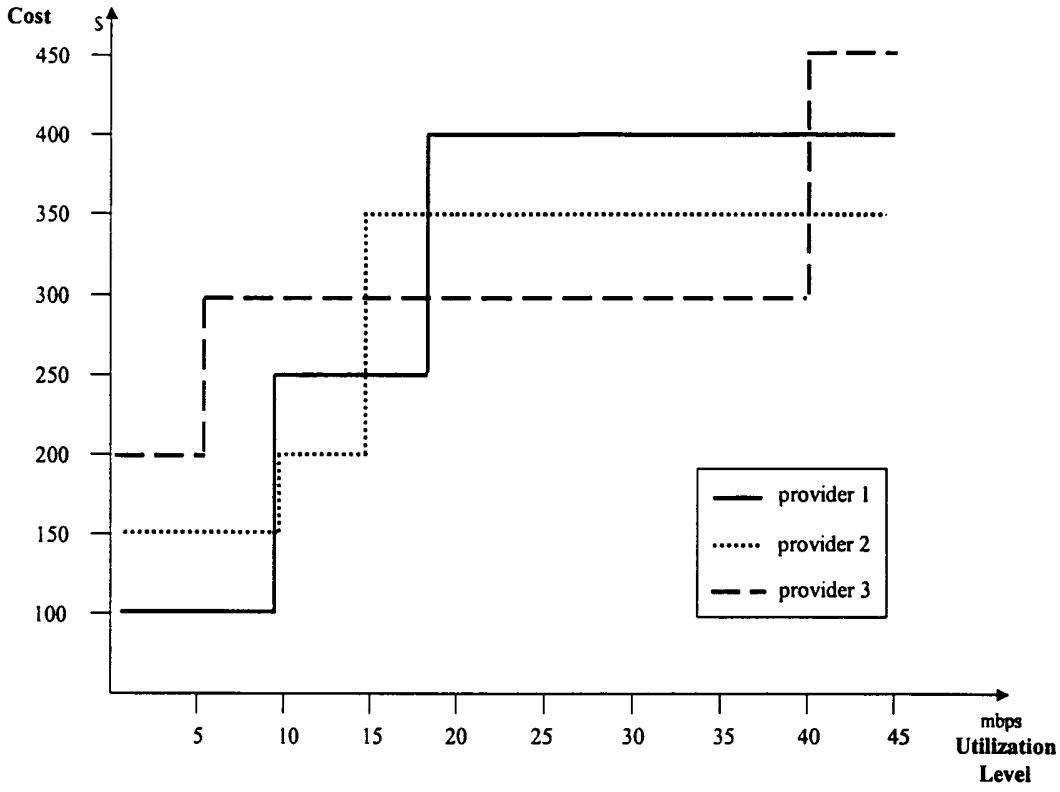

Cost
$
450
400
350
300
250
200
150
100
5
10
15
20
25
30
35
40
45
mbps
Utilization
Level
provider 1
provider 2
provider 3

FIGURE 8