A method and apparatus has been shown and described which allows Quality of Service to be controlled at a temporal granularity. Time-value curves, generated for each task, ensure that mission resources are utilized in a manner which optimizes mission performance. It should be noted, however, that although the present invention has shown and described the use of time-value curves as applied to mission workflow tasks, the present invention is not limited to this application; rather, it can be readily appreciated by one of skill in the art that time-value curves may be used to optimize the delivery of any resource to any consumer by taking into account the dynamic environment of the consumer and resource.
U.S. PATENT DOCUMENTS

OTHER PUBLICATIONS

Lavian, et al., A Platform for Large-Scale Grid Data Service on Dynamic High-Performance Networks, Nortel Networks Labs, et al., pp. 1-10.


Pan, et al., Time in OWL-S, University of Southern California/Information Sciences Institute, pp. 1-8.


* cited by examiner
Figure 2
Start

Select Generic Task Profile

Calculate time-value curve using task specific variables and constraints with selected generic task profile

Store time-value curve with task

Finish

Figure 4
Figure 5
Figure 7A

Figure 7B
Infinite 1/0 and BW

Fixed QoS

Using Time Value Curves for QoS

Figure 9
Figure 10
Start

New Task

Yes

Generate TV-curve for Task

No

Will Task suppress Other task?

No

Drop Suppressed Tasks

Yes

Figure 12
Figure 13
Figure 14

- Fixed QoS
- TV QoS
According to a further aspect of the invention, a device comprising a computer readable medium accessible by a resource allocator is described. The computer readable medium stores a data structure comprising, for each one of a plurality of enqueued tasks associated with the resource, a time value curve defining a temporal priority to provide to the associated task, the data structure being accessed by the resource controller to control access to the resource. These and other aspects of the invention will be described with regard to the attached figures.

**BRIEF DESCRIPTION OF THE FIGURES**

FIG. 1 is a block diagram of a mission architecture in which the present invention may be implemented;
FIG. 2 is a block diagram illustrating several components that may be included in a policy engine to generate time-value curves of the present invention;
FIGS. 3A-3H are diagrams illustrating different types of time-value curves that may used as a basis for time-curve generation by a policy engine of the present invention;
FIG. 4 is a flow diagram illustrating exemplary steps that may be performed by the policy engine of FIG. 2 to generate time-value curves for one or more tasks in a workflow;
FIG. 5 is a diagram of an exemplary network or service resource manager which may use the time value curves of the present invention to control allocation of resources to workflow tasks;
FIG. 6 is a timing diagram illustrating several exemplary time-value curves that may be assigned to tasks in a workflow;
FIGS. 7A and 7B are timing diagrams provided to illustrate the addition of time value curves as tasks are added to a workflow;
FIGS. 8A and 8B are used to describe an embodiment of the present invention supporting a surveillance workflow;
FIG. 9 is a processing timing diagram which pictorially illustrates the bandwidth smoothing capabilities of time-value QoS servicing of a workflow;
FIG. 10 is a pictorial representation of task selection using time-value curves in the example of FIG. 9;
FIGS. 11A and 11B are timing diagrams used to illustrate a stalled task drop capability of the present invention;
FIG. 12 is a flow diagram illustrating exemplary steps that may be performed by a policy engine performing task dropping such as that illustrated in FIGS. 11A and 11B;
FIG. 13 is a graph illustrating the performance of TV QoS versus Fixed QoS as the offered load increases; and
FIG. 14 is a graph illustrating how TV QoS provides for graceful network failure.

**DETAILED DESCRIPTION**

FIG. 1 illustrates a mission architecture 1 in which the present invention may advantageously be used to optimize mission resource utilization. In one embodiment the mission architecture 1 is an adaptive network that uses Workflow Locked Loops (WLL) for mission resource selection and access management. The embodiment of FIG. 1 includes a Mission Resource Manager (MRM) 10 and a Software Controlled Network (SCN) 12. The MRM 10 essentially orchestrates the use of mission resources by a mission by identifying mission tasks and allocating mission resources to the mission tasks. Although the MRM may be described below as including certain components, it should be understood that an MRM is not a required to implement the present invention. For example, a set of Mission Goals statements may be input into...
the Policy Engine to help form the Time-Value curves via an administrator at a GUI or automated via a machine interface or even a Web Services Interface.

An exemplary MRM 10 includes a Goal to Service translation Unit 12. The Goal to Service translation Unit may be a semantic tool which maps a high level goal definition, such as a natural language definition, into real-language mission goals into policy statements. One exemplary method for performing Goal to Service translation is described in Ser. No. 11/469,416, entitled Mission Goal Statement to Policy Statement Translation, by Travostino et al, filed Aug. 31, 2006, incorporated herein by reference.

Policy statements are translated into workflows using the Goal to Service Translation Unit (GSTU) 12, the Constraint Analyzer (CA) 15, the Discovery Engine and Scheduler (DES) 13 and the Directory 16. Together these components identify, quantify and qualify services that are reserved for a mission as described in patent application Ser. No. 11/469,422 for Workflow Locked Loops to Enable Adaptive Networks, by Travostino et al, filed Aug. 31, 2006.

For example in one embodiment the directory 16 identifies network resources and services that may be allocated to missions. The DES 13 identifies resources and services that are available for use as an infrastructure of the mission. The CA 15 evaluates the infrastructure identified by the DES 13 in view of mission constraints. The CA 15 and DES 13 work in concert with the goal translator to identify the mission infrastructure. An Engine uses the infrastructure and mission information to generate mission workflows, which are forwarded to the SCN 14.

The SCN 14 generally controls the execution of the mission and the use of the mission resources during execution of the mission. The SCN 14 includes an Execution Manager (EM) 22, a Policy Engine (PE) 24, a Network Services Manager (NSM) 26, a Network Resource Manager (NRM) 28 and a Network Resource and Topology Discovery resource 27.

For the purpose of this application a mission workflow is a collection of tasks that are performed by a collection of actors, wherein an actor may be any entity that uses a mission dedicated resource. A mission dedicated resource is any resource that is at least partially dedicated to a mission.

Mission tasks are dynamic, and therefore during the course of the mission different tasks are initiated and terminated. At any given time during the mission, mission tasks may seek access to a shared mission resource. According to one aspect of the invention, each task has an associated time-value curve indicating a temporal priority of the task over the task lifetime. The time-value curves are used during resource allocation to intelligently allocate resources to tasks to optimize the performance of the mission. Each point in a time-value curve identifies a priority of QoS to be associated with the task for the point in time. This allows QoS to be adjusted in real time based on conditions in the field, thereby adding the degree of temporal granularity to the QoS structure. Controlling temporal granularity of QoS helps resolve the conflicts of congestion and while supplying all the resources to the mission within the mission time-frame, without having to expand the size of the network.

In one embodiment the PE 24 generates a time-value curve for each mission task. The time value curves are forwarded to the EM 22. At resource allocation intervals the EM references the relative values of the time-value curve of each task that is active at that time instance to retrieve a time-value for each task, wherein an active task is any task that has been initiated and not terminated. Using any one of a variety of selection methods, such as highest value, highest weighted value, etc., the time values are compared to select one of the active tasks.

FIG. 2 illustrates several exemplary components that may be included in a PE to generate time-value curves for tasks. The PE 24 includes a workflow parser 31, a generic task profile store 34, and time-value (TV) curve generation logic 32. The workflow parser receives workflows from the Execution Machine (EM). The task may be initiated in response to a mission trigger; for example, in a surveillance mission, a mission trigger may be the capturing of an image at a sensor point in a surveillance area. The task may be triggered by partial or thorough completion of another task. However the task is triggered, it is forwarded to the workflow parser 31 which extracts task information that may be used to build the time-value curve. Any obtainable information which may affect a priority or quality of service to provide to the task may be provided to the time-value curve generation logic, including, for example, information regarding an actor of the task, a location of the task, a security level of the task, constraints of the tasks, dependencies of the task, power requirements of the task, execution time for the task, etc. Information regarding services and network elements that are associated with the task may also be used when calculating the time-value curve. The present invention is in no way limited to the use of any particular task information in generating a time-value curve for priority scheduling of the task.

The generic task profile store 34 may be used to store pre-determined task profiles. The predetermined task profiles are tailored to the resource requirements of various pre-defined tasks. For example, a surveillance task that performs high resolution image capture would have a time-value curve that temporally aligns the resource priority with the resource usage characteristics of the task. The resource requirements of a task may be selected from a set of pre-defined profiles, or obtained by monitoring and profiling resource usage in advance to generate a time-value profile for the task. FIGS. 3A-3H illustrates several representative time value curve profiles. A time-value curve profile such as that of FIGS. 3A and 3B may be applied to a specific task identified as having a relatively flat QoS. Profiles 3C and 3D may be used to apply a linear QoS to a task. Impulse profiles 3E and 3F may be used to provide high priority to a task for a short time. Time value curves of 7G and 7H illustrate curves that are constructed for a particular QoS profile.

FIG. 4 illustrates an exemplary process that may be performed by the PE 24 to construct a time-value curve for a task. At step 42 a generic profile associated with the task is retrieved from the profile store 34. At step 44, the time-value curve is piece-wise constructed using the base profile in conjunction with task specific variables, constraints and resource information using techniques known to those in the art. At step 46 the time-value curve is stored with the task. The TV augmented task is then forwarded to the EM 22.

FIG. 5 illustrates exemplary components that may be included in an Execution Manager 22 that uses the time-value curves of the invention. The EM 22 would generally include task queues 52 for storing active mission tasks. The task queues may be organized by mission resource. A resource controller 56 may include or otherwise have access to the Time-Value curves for the tasks, for example in a time-value queue associated with the task queue. At each mission resource allocation instance the resource controller examines the time-values of each active task associated with a resource at that instance, and selects one of the tasks based on the time-value curve.
FIG. 6 is a timing diagram illustrating illustrative time value curves for 3 tasks A, B and C. The generic task profile for each of the curves is shown by below Equation 1:

\[ P(T) = \text{Min}(mT + b) / \text{Max}(P) \]

Where \( P \) is the priority, or quality of service, associated with the task at time \( T \), \( m \) is derived based on the value of the task, the lifetime of the task and the execution time of the task and \( b \) is determined according to a value of the task. The flat portion on each of the time-value curves in this embodiment represents the time needed to complete the task. \( \text{Max}(P) \) represents the highest priority to allocate to the task.

Thus in FIG. 6, time value curve \( C \) is a lower value task, which takes a relatively short period of time to complete. Curve \( B \) has the highest initial priority task, with a shorter lifetime than curve \( C \). Curve \( A \) has the second highest initial priority, but the shortest lifetime. Using strict priority allocation technique, task \( B \) will obtain access to the resource until time \( T_1 \). At time \( T_1 \), access begins to be granted to task \( A \), with task \( A \) reaching its maximum priority at time \( T_2 \) to allow it to be at maximum priority for the time period associated with executing the task. The time-value curves may thus be used provide a temporal priority of the task which ensures that the task can complete in its lifetime.

FIGS. 7A and 7B are pictorial representations of a mission workflow, provided to illustrate the dynamic nature of workflow tasks; tasks are continually instantiated and terminated during the workflow. For example, at time \( T_4 \), a low priority task \( 68 \) is triggered. At time \( T_7 \) a higher priority task is triggered. At any resource allocation period, the time value curves may be used to real-time, dynamic resource management for a dynamic mission workflow.

FIGS. 8A and 8B will be used to describe an exemplary embodiment of the present invention for use in a particular mission. The mission of FIG. 8 is a surveillance mission which seeks to identify individuals within a defined radius. The mission comprises dependent tasks, with a first set of tasks being associated with the detection and collection of information for each individual that comes within a radius, a second set of tasks being associated with collecting high resolution information regarding individuals within the radius, a third set of tasks being associated with collecting increasingly images at very high resolution. Each of the tasks of in the mission use network bandwidth for the exchange of information associated with the target. The individuals are grouped into low value targets (initial priority 0); mid value targets (initial priority value 1) and high value targets (initial priority value 2). The targets are detected using mission resource sensors distributed within the radius. A detection of the target results in an attempt to collect information about the target to determine if the target is a target of interest. The collection of information may include the transmission of identification information, such as a low resolution or high resolution image. The type of collection to be performed for each target is a task in the above described architecture. The task is assigned a time-value curve which provides a temporal priority to the collection of information for the particular target.

The time value curves of FIG. 8B are associated with tasks to be performed on three targets. A first task \( 72 \) is associated with a mid priority target which has been sensed in area DS3. There are only 3 minutes to collect information regarding the target before the target leaves the area. Thus the time-value curve is adapted for this target to ensure that data collection for this target can occur within the target tracking task lifetime. A second task \( 74 \) is associated with a target in area DS2.

The target has been identified as a high value target, perhaps as a result of a low-resolution image mapping task. The target is exiting the area in 4 minutes. The slope and intercept of the time-value curve reflect the perceived target value and detected target location, and ensures that adequate processing of the target may be completed before the target is lost. Time-value curve \( 76 \) is associated with a task generated in response to target detection in area DS2b. There are 5 minutes until the target quits the area, and thus time-value curve \( 76 \) reflects the lower quality of service level to be provided to the collection of information regarding this target.

FIG. 9 is a data flow diagram intuitively illustrating the benefits of using the time-value curves of the present invention to provide real-time temporally prioritized resource allocation in to mission tasks. As will be seen from the below description, the time-value curves effectively smooth and optimize resource allocation among the tasks. Optimizing the resource allocation improves the completion rate of mission tasks and increases the chances of mission success.

Data series \( 80 \) illustrates a stream of resource requests that are received in a data stream. The performance of the mission depends upon being able to process each task in the data stream during its lifetime; if tasks are dropped or delayed, mission information is lost. In FIG. 9, each data slot represents a one minute interval, and assume that the Burst traffic received in interval 1 includes high, medium and low priority tasks having respective lifetimes of three, four and five minutes.

Data flow \( 90 \) illustrates an exemplary result of resource allocation for a mission using Fixed QoS resource allocation methods. In a typical Fixed QoS resource allocation method, a fixed bandwidth percentage is allocated to each task. During interval 1 of Data flow \( 90 \), Burst Traffic for three tasks is received. The prior art resource allocator, receiving no information from the field regarding a temporal priority of the tasks, uses a Fixed QoS to allocate resources to the tasks. Each of the three tasks is provided one third of the bandwidth. At the end of interval 1 (or one minute into the task), each of the three pending tasks in group A is \( 1/3 \) complete. During interval 2, an additional task group B is received which includes two additional tasks. Assume group B includes a medium and high priority task with respective lifetimes of four and five minutes. During interval 2 the resource allocator allocates the resource equally among the tasks in group A and group B, providing one fifth of the bandwidth to each task. The Fixed QoS structure has no way to determine which task is temporally more important due to conditions in the field. After two minutes, the tasks in group A are now just over half completed. During interval 3, an additional task is received, and each task in interval 3 is assigned one sixth of the bandwidth. At the end of three minutes the high priority task of group A is dropped. Using the example of FIG. 8A, this would equate with tasks initiated for individuals in DS3 failing to complete before the individual had left the radius.

In contrast, when time value curves are applied to the QoS paradigm, resource allocation would proceed as is shown by data flow \( 95 \). FIG. 10 is a pictorial representation of the time-value curves per task for the data flow of FIG. 9. During interval 1, when burst task set A is received a resource allocator of the present invention compares the time-value curves of each task to identify the task with temporal priority. As shown in FIG. 10, during interval 1 the high level task is given control of the resource. During interval 2, new tasks are added. However, during time interval 2 the time value curve of the mid-range task of group A has temporal priority. Thus the mid-level task of group A is completed at two minutes. It the high-level task of group A is completed at two minutes. At interval 3, at least a portion of the mid-level task of...
group B is allocated access to the resource. Thus it can be seen 8 that integrating time value curves into the policy and quality of service enforcement decisions increases the completion rate of high value tasks.

Time-value curves can also be used to identify tasks that may not successfully complete in their lifetime. These tasks are hereinafter referred to as stalled tasks. Identifying stalled tasks allows them to be removed from resource allocation consideration, thereby reducing performance and providing a graceful degradation of performance in the face of congestion.

FIG. 11A illustrates how dynamic time-value curve analysis may be used to identify stalled tasks. In FIG. 11A, task 126 is initiated at time. However, given the temporal priorities of the other tasks in the workflow, it is unlikely that the low priority task 129 will be able to complete in its lifetime. In the present invention, the Execution Manager identifies stalled flows and removes them from the task queue, thereby simplifying the decision making process of the EM and reducing congestion. FIG. 11B is a pictorial representation of the time-value curves that are under consideration following the abortion of the stalled task.

FIG. 12 is a flow diagram illustrating several exemplary steps that may be performed by an Execution Manager or other resource allocator, to drop task execution using time-value curve information of the present invention. At steps 132 and 134 a time value curve is generated for each new task as described with regards to FIGS. 2-4. At step 136 it is determined whether the added task is going to result in the stalling of an existing task. If so, at step 138 the stalled task is removed from the task queue.

FIG. 13 is a graph that is provided to illustrate the performance of a TV QoS system versus that of a fixed QoS as the offered load increases. As shown in FIG. 13, the number of missed targets increases linearly with the increase in load. In contrast, the TV QoS does not begin to drop targets until the average traffic approaches the offered load (as indicated by the dashed line X in FIG. 13).

FIG. 14 is a graph that is provided to illustrate the graceful performance degradation that can be realized in a network which implements time-value curves of the present invention to abort stalled queues. As the size of the packet payload is reduced the frequency of resource accesses is increased. Once the offered load is less than the average traffic, resource congestion quickly ensues, and data stream are dropped. The present invention uses time-value curves to identify the most temporally relevant tasks even during congestion, thereby resulting in graceful network performance degradation in the face of congestion as shown in FIG. 13.

Accordingly a method and apparatus has been shown and described which allows Quality of Service to be controlled at a temporal granularity. Time-value curves, generated for each task, ensure that mission resources are utilized in a manner which optimizes mission performance. It should be noted, however, that although the present invention has shown and described the use of time-value curves as applied to mission workflow tasks, the present invention is not limited to this application; rather, it can be readily appreciated by one of skill in the art that time-value curves may be used to optimize the delivery of any resource to any consumer by taking into account the dynamic environment of the consumer and resource. The allocation of any resource, including network bandwidth, storage, processing capability, etc., may benefit from the teachings of the present invention.

Having described various embodiments of the invention, it will be appreciated that many of the above figures are flowchart illustrations of methods, apparatus (systems) and computer programs according to an embodiment of the invention. It will be understood that each block of the flowchart illustrations, and combinations of blocks in the flowchart illustrations, can be implemented by computer program instructions. These computer program instructions may be loaded onto a computer or other programmable data processing apparatus to produce a machine, such that the instructions which execute on the computer or other programmable data processing apparatus create means for implementing the functions specified in the flowchart block or blocks. These computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide steps for implementing the functions specified in the flowchart block or blocks.

Those skilled in the art should readily appreciate that programs defining the functions of the present invention can be delivered to a computer in many forms; including, but not limited to: (a) information permanently stored on non-writable storage media (e.g read only memory devices within a computer such as ROM or CD-ROM disks readable by a computer I/O attachment); (b) information alterably stored on writable storage media (e.g floppy disks and hard drives); or (c) information conveyed to a computer through communication media for example using baseband signaling or broadband signaling techniques, including carrier wave signaling techniques, such as over computer or telephone network via a modem.

The above description and figures have included various process steps and components that are illustrative of operations that are performed by the present invention. These components may be implemented in hardware, software or a combination thereof. However, although certain components and steps have been described, it is understood that the descriptions are representative only, other functional delineations or additional steps and components can be added by one of skill in the art, and thus the present invention should not be limited to the specific embodiments disclosed. In addition it is understood that the various representational elements may be implemented in hardware, software running on a computer, or a combination thereof.

While the invention is described through the above exemplary embodiments, it will be understood by those of ordinary skill in the art that modification to and variation of the illustrated embodiments may be made without departing from the inventive concepts herein disclosed. Accordingly, the invention should not be viewed as limited except by the scope and spirit of the appended claims.

The invention claimed is:

1. A method of allocating a resource to a plurality of resource consumers includes the steps of:
   generating, with a network device, a time-value curve defining a temporally dynamic priority of a resource consumer over a time period, wherein the priority is calculated as a function of time; and
   using the time-value curve to allocate access to the resource to the consumer over the time period.
2. The method of claim 1 wherein the temporally dynamic priority defines a temporally dynamic quality of service to provide to the resource consumer.

3. The method of claim 1 wherein each of the plurality of resource consumers has an associated time-value curve and wherein at least two of the time-value curves are different.

4. The method of claim 3 wherein the step of using the time-value curve to allocate access to the resource includes the steps of:
   - sampling time-value curves at periodic intervals to obtain sampled time-values for each of the plurality of resource consumers; and
   - using the sampled time-values to select one of the plurality of resource consumers as an access consumer for accessing the resource for a periodic interval.

5. The method of claim 4 wherein the time-value curves are differently weighted.

6. The method of claim 4 wherein the time-value curves have different lifetimes.

7. A method of allocating a resource to a plurality of resource tasks in a workflow includes the steps of:
   - generating, with a network device, a time-value curve for each resource task in the workflow, the time-value curve defining a temporally dynamic quality of service to be provided to the corresponding resource task, wherein the quality of service is calculated as a function of time, and using the time-value curve to select one of the plurality of resource tasks for resource allocation.

8. The method of claim 7 wherein the time-value curves are associated with dynamic tasks which have a lifetime.

9. The method of claim 8 further including the step of using the time-value curves to identify stalled tasks.

10. The method of claim 9 including the step of aborting stalled tasks to reduce congestion.

11. The method of claim 7 wherein the step of generating a time-value-curve for each task includes the steps of:
   - selecting a generic task profile; and
   - modifying the generic task profile with task specific information to provide a task specific time-value curve.

12. The method of claim 11 wherein the generic task profile is defined by the below equation 1:

\[
P(T) = \text{Min}(mT+b, \text{Max}P)
\]

wherein \(P\) is a priority of the task at time \(T\), \(m\) is derived through a combination of a value of the task, a lifetime of the task and an execution time of the task, \(b\) is determined according to the value of the task and \(\text{Max} \ (P)\) represents the highest priority to allocate to the task.

13. The method of claim 7 wherein the time-value curve is used to calculate a quality of service for the task.

14. A device comprising a non-transitory computer readable medium accessible by a resource allocator, the computer readable medium for storing a data structure comprising, for each one of a plurality of enquired tasks associated with the resource, a time value-curve defining a temporal priority to provide to the associated task as a function of time, the time value curve of the data structure being accessed by the resource controller to control access to the resource.

* * * * *
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title Page 1, Front Page, and Column 1 in the Title, replace “TRANSFER” with --TRANSFERS--.

Title Page 2, Column 1, under “Other Publications”, line 9 (approx.) replace “Gird” with --Grid--.

In Column 10, Line 25, Claim 14, replace “time value-curve” with --time-value curve--.

In Column 10, Line 26, Claim 14, replace “time value” with --time-value--.

Signed and Sealed this Fifth Day of June, 2012

David J. Kappos
Director of the United States Patent and Trademark Office