

ADAPTIVE SERVICE TIMEOUTS

[0001] The following application claims priority to, and the benefit of, U.S. Patent Application entitled "Adaptive Service Timeouts" filed on September 18, 2012 having application number 13/622,094, which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] Timeouts can prevent excessive use of system resources in processing high-latency service calls. Specifically, when processes time out, they no longer consume system resources.

SUMMARY OF INVENTION

[0002a] According to an aspect of the present invention, there is provided a system, comprising:

- at least one computing device; and
- a timeout management application executable in the at least one computing device, the timeout management application comprising:
 - logic that monitors a plurality of latencies, ~~each latency individual~~ ones of the plurality of latencies corresponding to one of a plurality of services; and
 - logic that sets a timeout associated with a pair of the plurality of services based at least in part on the plurality of latencies, ~~the timeout defining a~~ maximum time to complete a call to a first service of the pair of the plurality of services when called by a second service of the pair of the plurality of services.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003] Many aspects of the present disclosure can be better understood with reference to the following drawings. The components in the drawings are not necessarily to scale, with emphasis instead being placed upon clearly illustrating the principles of the disclosure. Moreover, in the drawings, like reference numerals designate corresponding parts throughout the several views.

[0004] FIG. 1 is a drawing of a networked environment according to various embodiments of the present disclosure.

[0005] FIG. 2 is a flowchart illustrating one example of functionality implemented as portions of the timeout management application executed in a computing environment in the networked environment of FIG. 1 according to various embodiments of the present disclosure.

[0006] FIG. 3 is a flowchart illustrating one example of service dependency functionality implemented as portions of the timeout management application executed in a computing environment in the networked environment of FIG. 1 according to various embodiments of the present disclosure.

[0007] FIG. 4 is a schematic block diagram that provides one example illustration of a computing environment employed in the networked environment of FIG. 1 according to various embodiments of the present disclosure.

DETAILED DESCRIPTION

[0008] Service-oriented architectures enforce timeouts to prevent excessive use of system resources when making calls to a high-latency service. This prevents a high-latency service from

CLAIMS

Therefore, the following is claimed:

1. A system, comprising:

at least one computing device; and

a timeout management application executable in the at least one computing device, the timeout management application comprising:

logic that monitors a plurality of latencies, ~~each latency individual ones~~
of the plurality of latencies corresponding to one of a plurality of services; and

logic that sets a timeout associated with a pair of the plurality of services
based at least in part on the plurality of latencies, the timeout defining a maximum time
to complete a call to a first service of the pair of the plurality of services when called by a
second service of the pair of the plurality of services.

2. The system of claim 1, wherein the timeout management further comprises logic that calculates an aggregate latency relative to a time interval based at least in part on the plurality of latencies, and the timeout is further based at least in part on the aggregate latency.

3. The system of claim 1, wherein the timeout is further based at least in part on an amount of used service capacity.

4. The system of claim 3, wherein the logic that sets the timeout increases the timeout in response to the amount of used service capacity relative to a total service capacity falling below a threshold.

5. The system of claim 3, wherein the logic that sets the timeout decreases the timeout in response to the amount of used service capacity relative to a total service capacity exceeding a threshold.

6. The system of claim 3, wherein the timeout management further comprises logic that selects ~~one~~ the first service of the pair of the plurality of services based at least in part on a subset of the amount of used system resources used by the selected service.

7. The system of claim 1, wherein the timeout management further comprises logic that selects ~~one~~ the first service of the pair of the plurality of services based at least in part on a service priority list, and the timeout is associated with the selected one of the services.

8. The system of claim 1, wherein the timeout management further comprises logic that selects ~~one~~ the first service of the pair of the plurality of services based at least in part on a reliability threshold.

9. The system of claim 1, wherein the logic that monitors the plurality of latencies further comprises logic that obtains log data from the services.