



US 20050246705A1

(19) **United States**

(12) **Patent Application Publication**

Etelson et al.

(10) **Pub. No.: US 2005/0246705 A1**

(43) **Pub. Date: Nov. 3, 2005**

(54) **METHOD FOR DYNAMICALLY ALLOCATING AND MANAGING RESOURCES IN A COMPUTERIZED SYSTEM HAVING MULTIPLE CONSUMERS**

(75) Inventors: **Garik Etelson, Kiryat Ono (IL); Gregory Bondar, Rishon Lezion (IL); Michael Stoler, Yahud (IL)**

Correspondence Address:

**FLEIT KAIN GIBBONS GUTMAN BONGINI & BIANCO
21355 EAST DIXIE HIGHWAY
SUITE 115
MIAMI, FL 33180 (US)**

(73) Assignee: **SPHERA CORPORATION, NEWTON, MA**

(21) Appl. No.: **11/042,478**

(22) Filed: **Jan. 25, 2005**

Related U.S. Application Data

(63) Continuation of application No. PCT/IL03/00619, filed on Jul. 25, 2003.

(30) **Foreign Application Priority Data**

Jul. 25, 2002 (IL) 150911

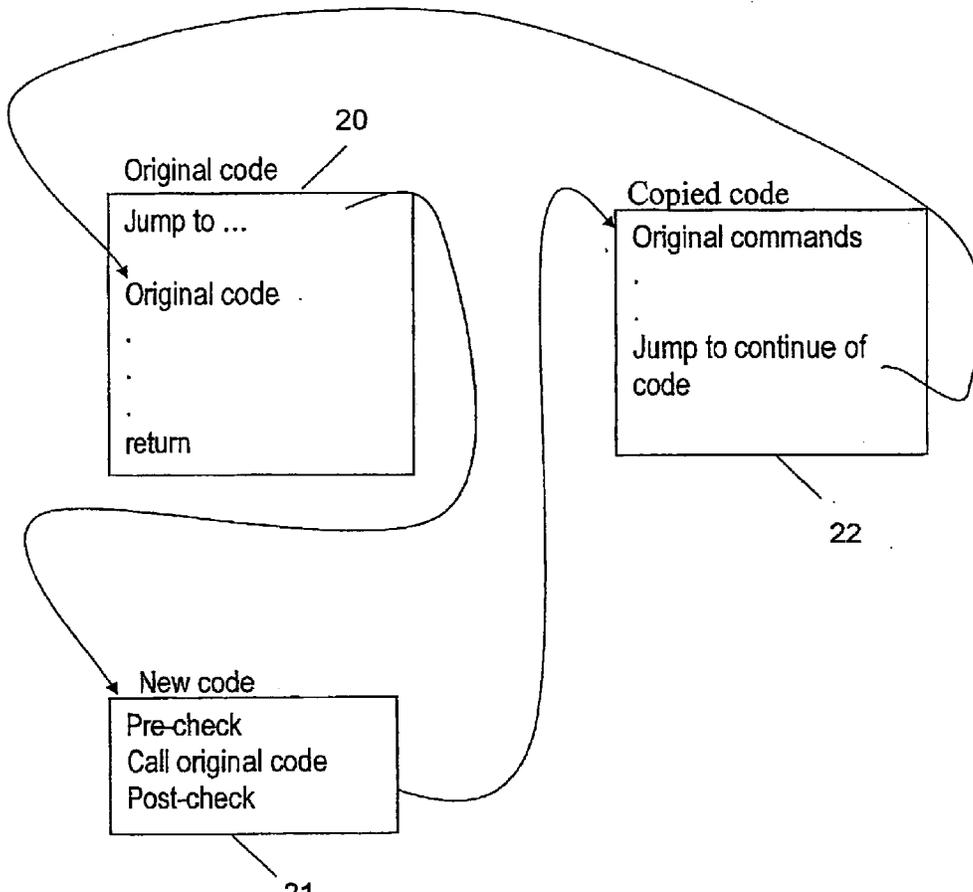
Publication Classification

(51) **Int. Cl.⁷ G06F 9/46**

(52) **U.S. Cl. 718/100**

(57) **ABSTRACT**

Method for dynamically allocating and managing resources in a computerized system managed by an operating system (OS) and having multiple accounts of consumers. Portions of the virtual memory address space are allocated, whenever desired, in a swap file, for each account associated with a consumer. The memory address space is limited for each account. The CPU usage is divided between the tasks requested from each account, and segments in the original code of the OS are changed by locating one or more specific procedures in the original code, and modifying the specific procedures to operate according to the allocation and/or the limitation of the memory address space and/or the limitation of the number of processes and/or the divided CPU usage.



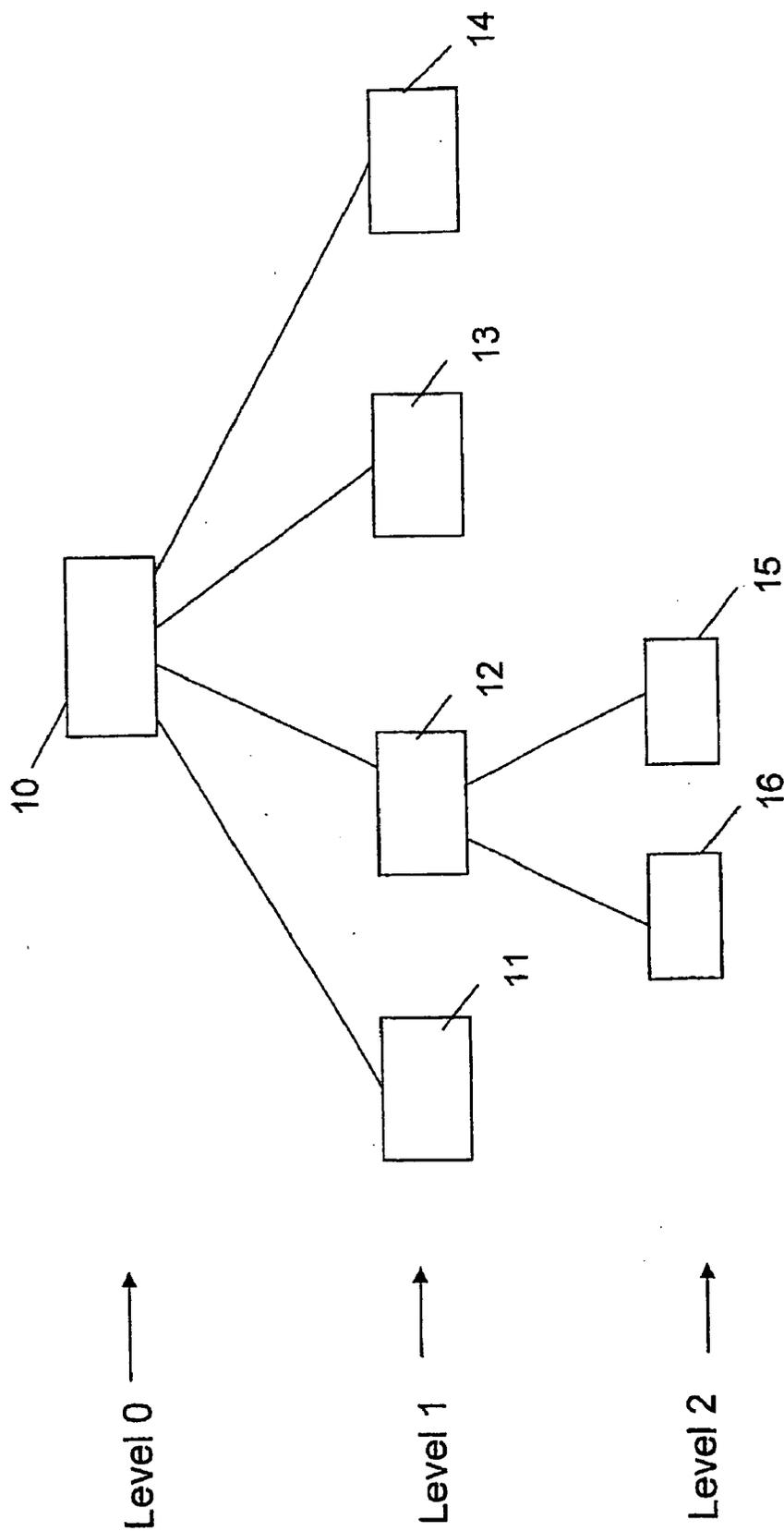


Fig. 1

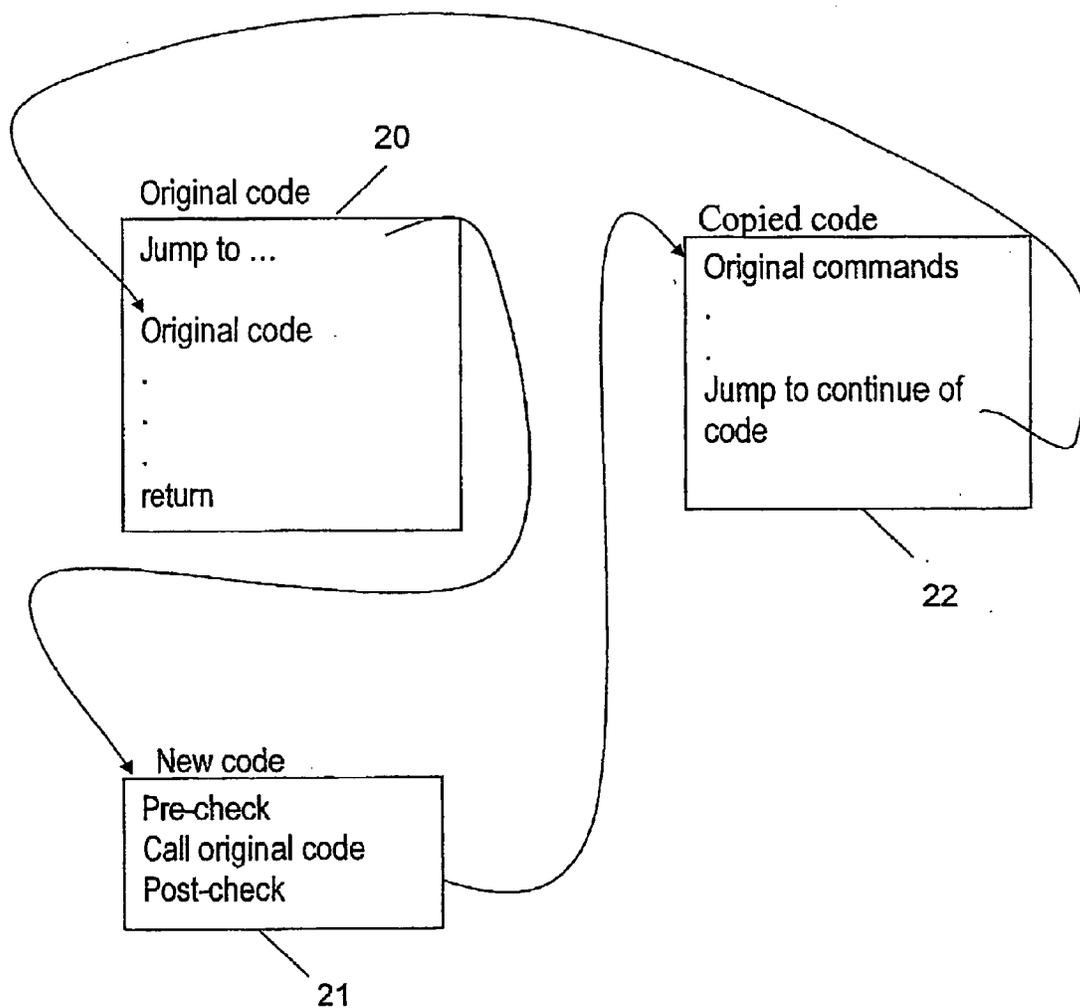


Fig. 2

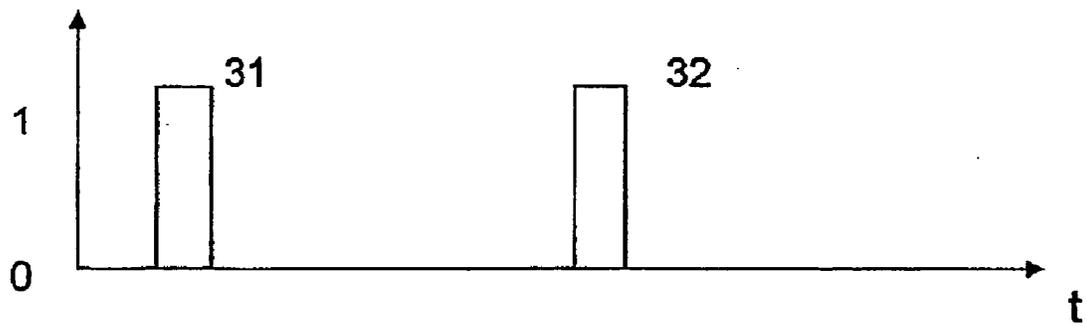


Fig. 3

**METHOD FOR DYNAMICALLY ALLOCATING
AND MANAGING RESOURCES IN A
COMPUTERIZED SYSTEM HAVING MULTIPLE
CONSUMERS**

RELATED APPLICATION

[0001] This application is a continuation of International Patent Application Serial number PCT/IL2003/000619 filed Jul. 25, 2003, the contents of which are here incorporated by reference in their entirety. The benefit of 35 USC Section 120 is here claimed.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to the field of managing a computerized system. More particularly, the invention relates to a method for limiting the resources that are used by consumers, systems and web services of a given computerized system.

[0004] 2. Prior Art

[0005] Hosting a Website locally is relatively expensive, as it requires allocating sufficient bandwidth for Internet traffic to the site, as well as allocating resources for keeping the site available all the time (both in terms of software and hardware) and handling security aspects, such as a firewall.

[0006] Web Hosting Providers (WHP), which are the consumers of a computerized system, use a variety of service models to address different types of customers, depending on the required class of service. The Web sites of small and medium-sized businesses normally do not preempt the resources afforded by a dedicated server, and therefore might settle for a shared server model. However, as the requirements of the WHP change and their sites conduct more and more activity, they become more resource-consuming. When WHPs become more resource consuming, they usually, hire more resources, or keep the same resources with decreased performance. As the demand for the site's services is not constant over a time period, the customer might prefer to keep the same resources rather than hiring more resources, assuming that a relatively high demand for resources might occur for only a relatively short duration.

[0007] Typically, each dedicated server runs an instance of the OS (Operation System). However, running an instance of the OS for each dedicated server comparatively requires a large amount of resources, which is required for each instance of the OS.

[0008] Hereinafter, the term "computerized system" refers to a server that hosts a plurality of virtual dedicated servers that execute a plurality of services, wherein each virtual dedicated server utilizes a substantial portion of the computer resources.

[0009] A virtual dedicated server in such a computerized system is actually an emulation of a computer system's interface in which a remote client can access its system utilities and programs, and it will be called hereinafter a Virtual Dedicated Server (VDS). A plurality of VDS instances can be executed simultaneously on a single hosting computerized system.

account might share its allocated resource with other accounts, but together they can not utilize more than their allocated share. An "account" can be allocated to a user, a domain, a VDS, a service, a specific processes or process groups or to any other suitable user of the machine's resources.

[0011] One of the existing solutions for limiting the resources consumption of an account is to use a static division of the computer resources. The hosting computer resources are divided in a static manner between the virtual computers. The result is that if, for example, the real computer is split into 10 identical virtual computers, then 10% of the system resources are allocated to each virtual computer, even if only one virtual computer is being operated. A dynamic resource allocation would result in a better performance per virtual computer (if not all the VDSs are activated at the same time), with an appropriate allocation to each VDS (according to predefined parameters) in the case that a plurality of VDSs are activated at the same time. Therefore, the dynamic resource allocation results in a better performance from the user's point of view. The dynamic resource allocation can be used by any consumer of the computer resources, such as different services, different users, etc.

[0012] Resources of a computerized system are limited due to several factors such as budget, spatial restrictions, etc. Resources of a computerized system comprise, among others, the usage of a Central Processing Unit (CPU), the size of a memory address space, storage capabilities of data, etc. A computerized system used by multiple consumers, whether they are WHP or regular consumers, needs to provide to each of its consumers, at least, a predefined percentage of its resources according to predefined terms or agreements between each consumer and the corresponding resources owner in the computerized system. A WHP may offer more than the actual available resources, based on the low probability that all consumers will concurrently demand maximum resources. Therefore, in order to enable different consumers to have their predefined share of resources, there is a need to limit the resources available to a specific consumer according to those predefined for him. Additional reasons for limiting the resources consumption for each consumer in a multiple consumer computerized system may be as follows:

[0013] If the resources are not of a preempt kind (i.e., non-preemptable), then a suitable process in the computerized system should free those resources by itself, upon receipt of such a resource. For example, the memory or a suitable storage disk of a computerized system is usually non-preemptable. Granting a higher number of resources might prevent a process, before the previous resources were freed, from getting its share. Unfortunately, it is relatively complicated to remove the resources, once granted.

[0014] If the resources are of a preempt kind (i.e., preemptable), then in every time-slice they are divided between the requesting processes. For example, a CPU is usually a preemptable resource. When dealing with preemptable resources, there are two possibilities to deal with the unused resources, as the allocation is performed on every time-slice from scratch. The first, granting the process more than his

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.