［Original Paper］

Configuration of a Power-saving High-availability Server System Incorporating a Hybrid Operation Method

Mitsuyoshi Kitamura

Abstract

Internet services continue to expand and become increasingly indispensable in our daily lives. For a server system to realize power savings and high availability, which involves reducing the frequency of a condition in which offers of service cannot be made, is important. The present study proposes a hybrid operation method and examines a method for constructing a power-saving high-availability server system (PSHASS) incorporating the hybrid operation method. In addition, the present study exhibits the characteristics of both a power-saving server system and a high-availability server system. These systems can operate independently. An experimental server system using the proposed method was constructed. Experiments to reproduce several types of trouble in the server system in moderate or power-saving operation state were conducted. The proposed PSHASS was found to be able to automatically recover to a normal condition from a condition in which trouble occurs. Since both the power savings and high availability of the server system can be effectively realized by incorporating the proposed method, expansion of Internet services can be expected, providing greater convenience in our daily lives.

1. Introduction

The use of communications equipment, such as smartphones or tablet terminals, is rapidly developing with the development of a highly information-oriented society. An advanced future network system is needed because a number of services that use the Internet will expand. In order to provide users with reliable services, the optimal design and construction of a server-client system, which is the fundamental unit of the network system, is considered to be important [1]. Furthermore, in order to deal with the increasing number of services, the number of server units will generally increase. As a result, the possibility of server freezes or failures will increase. Thus, constructing safer server systems is necessary and the number of server units that include backup devices must increase.

Research to investigate server trouble from various viewpoints has been conducted. For example, modeling of multi-server systems has led to the development of the concept of server recovery [2], and transparent fault tolerance for device drivers for virtual machines and a recovery mechanism have been proposed [3]. Moreover, optimization of software for a server virtualization system constituting a cloud system has been reported [4], and a high-reliability design method that provides users with continuous service even when several servers in a data center go down has been proposed [5].

In constructing a server system, it is necessary to consider a power savings as well as high speed and security [6],[7]. In a report by the Ministry of Economy, Trade, and Industry of Japan, it was estimated that the power consumption by IT devices will increase ninefold by 2025, as compared with that in 2006 [8]. A number of important studies on power-saving network systems have been conducted. The

Tokyo Polytechnic University
［Original Paper］ Received 22 January 2014
Accepted 5 February 2015

realization of power savings through the combination of a dispersion cluster and a directional antenna in a radio sensor network [9], the proposal of optimal power savings and performance for a network interface [10], a passive light network architecture to decrease energy consumption [11], the development of a power-saving decoder in a radio sensor network [12], and dynamic control via traffic prediction in a router [13] have all been considered.

In addition, several studies on server systems have been conducted. Realization of low cost and power savings by introducing a system that can back up the functions of several real servers by means of one real server [14], the examination of processing performance and power consumption in a server having an ARM processor [15], and control of a server power supply in a data center using server virtualization [16] have been achieved.

However, although approximately half of all IT devices are composed of network devices and servers, very little research has been done in this field. There has been little research on power savings in server systems. Furthermore, the consideration of power savings and high availability of an Internet system, which is a significant social infrastructure, is important, and power savings and high availability of the server-client system, which is the fundamental unit of the Internet system, is very important. Research on power-saving high-availability server systems appears to be scarce. If the generation of server trouble is considered, it is generally believed that a control method to realize power savings and high availability by one control program would be extremely complicated.

Therefore, the present study proposes a hybrid operation method, and a power-saving high-availability server system (PSHASS) that incorporates the method is configured. The power-saving system measures the load of a virtual server that provides a service to clients and judges the optimal server condition depending on the environment of a server system on the basis of the load. Even if various server troubles occur, the high-availability system recovers the function of the inoperative server. In order to shorten the server function recovery time, a background process is performed in an attempt to return a server system that includes an inoperative server to a normal state. In the PSHASS, the power-saving system adopts a judgment method for the optimal server condition, and the high-availability system adopts the recovery method of the server function and the server system. The power-saving system and the high-availability system are proposed herein. In addition, the power-saving and high-availability systems adopt a mode segmentation method [14], which is introduced in order to simplify system control design and improve the level of applicability to various other server systems.

The PSHASS consists of a load balancing system, virtual servers offering service to clients, real servers, which start virtual servers, and a control server. A power-saving control program and a high-availability control program are installed on the control server. Each control program can operate independently. The proposed method enables a server system to realize power savings and high availability by alternately operating each control program. The PSHASS contains configuration files and shared files. In order to start the PSHASS, configuration files (e.g., an IP address or host name) are used. Shared files (e.g., information of a suspended server by the power-saving system and the host name of a virtual server that has experienced trouble) are used by power-saving and high-availability systems and are edited automatically according to the condition of the PSHASS. These systems can grasp each server condition through the use of shared files. Thus, the high-availability system does not judge the server to be experiencing trouble, even if the power-saving system suspends the server in order to realize power savings.

We assume that the virtual server generates more trouble, e.g., hang-ups, than the real server because the virtual server is started as software, whereas the virtual server has numerous advantages. In order to realize power savings and high availability, a virtual server is adopted for use in the PSHASS as a server to provide services to clients. The power-saving system can gather virtual servers on a particular real server. Therefore, the power-saving system suspends real servers, which do not start virtual servers, because virtual servers are started from the real server as software. As a result, the PSHASS can realize power savings. Furthermore, the high-availability system recovers a function of the virtual server that is generating trouble by means of a virtual server that is prepared as a backup. As a result, the PSHASS can realize high availability.

Section 2 describes the mode segmentation method. Section 3 describes the construction of the proposed system, the structure of the hybrid operation method, and the configuration files. Section 4 describes the features of the power-saving system, the measurement of server loads, and the judgment method for the optimal server condition. Section 5 discusses the details of the recovery method of the server function and the server system in the high-availability system and the method used to deal with trouble in virtual or real servers. Section 6 describes the construction of an experimental system that adopts the proposal method and the mechanism of the experimental system and demonstrates that the proposed PSHASS can deal with various types of trouble, including hardware problems in the target server. In addition, the recovery times from various types of trouble are measured.

2. Mode segmentation method

The mode segmentation method was proposed in a previous study [14]. The power-saving and high-availability systems in the proposed PSHASS adopt the mode segmentation method. A mode is a software system used to execute a specific task, and each mode can be executed independently. This method constitutes one system through the use of multiple modes and is introduced in order to simplify system control design and improve the level of applicability to various other server systems. In the previous study, the mode segmentation method was adopted in the development of a low-cost power-saving multiple server backup system.

Figure 1 shows an outline of the mode segmentation method. The system is configured for four different modes, Modes A through D. The modes are software systems that are used to execute specific jobs, and each mode is independent. These modes execute operations according to the situation by transmitting necessary information between modes. In Mode A, a network and services on a target server to be monitored are regularly monitored. If trouble with the target server network is detected, Mode A reports the IP address of the server to Mode B, and the control state shifts to Mode B. If trouble with the target server offering services to clients is detected, Mode A reports the IP address of the server and the name of service that is abnormal to Mode C, and the control state shifts to Mode C. Mode B examines the target server network and the control state returns Mode A if the network is normal. Mode C recovers the service function on the target server and the control state returns Mode A if the service function is normal. If the target server is judged to be abnormal in Mode B or Mode C, these modes report the IP address of the server to Mode D, and the control state shifts to Mode D. Since Mode C performs recovery of the service functions, it is necessary to perform a task depending on the OS of the target server to be monitored. Therefore, the system can control the target server of various OSs by editing a control program in Mode C.

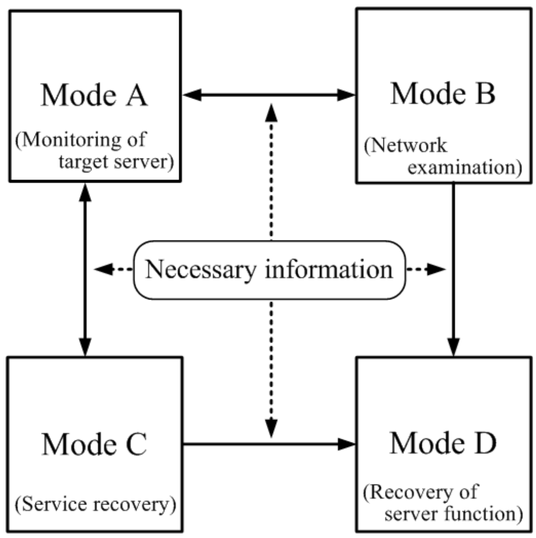


Fig. 1 Mode segmentation method.

In the previous study, a server management system program, which is constructed by the mode segmentation method, manages one target real server and can start one virtual server to recover a function of the real server if the target real server is abnormal. Therefore, in the method of the previous study, the preparation of virtual servers having a function for target real servers and multiple starts of the server management system program are needed in order to manage numerous target real servers. Since these tasks are performed with one management server, a low-cost power-saving multiple-server backup system is realized. In the present study, one server management system program, which is constructed by the mode segmentation method, can manage several target servers, including virtual servers and real servers that start virtual servers. Power savings and high availability are realized depending on the operation condition of the target servers.

The mode segmentation method is based on a mode that monitors the state of the target server, and the control state shifts to another mode based on necessary information depending on the type of trouble encountered. The operation of a power-saving regular mode, described in Section 4, is different from that of Mode A. The operation regularly measures the load state of target servers, and the power savings is realized by means of shifting the control state to the appropriate mode depending on the operation state. In addition, Mode A monitors the network and service of one target server, whereas the high-availability regular mode, described in Section 5, monitors the networks and services of several target servers. If trouble occurs, high availability is realized by means of shifting the control state to the appropriate mode according to the trouble state.

3. Outline of the PSHASS

The present study proposes a hybrid operation method, and the PSHASS that incorporates the method is configured. The method alternately operates two systems and realizes the function obtained by the systems. The power-saving system and the high-availability system used in the present study are independent systems, and the proposed method alternately operates these systems in order to realize a power-saving high-availability server system. In addition, the proposed method uses shared files to grasp each server condition because these systems operate alternately. Thus, the high-availability system does not judge the server to be experiencing trouble even if the power-saving system suspends a server to realize power saving. The PSHASS is based on a load balancing system, which is generally used to construct high-availability load-distribution server systems.

3.1 Configuration of PSHASS

Figure 2 shows the configuration of the PSHASS. A client accesses a target server via the load balancing system. Therefore, the system is made redundant by installing multiple load balancers. Virtual servers are used to offer services to clients, for example, mail, web, and FTP services. Virtual servers are distributed redundantly to each service group, and clients access the virtual servers via the round robin method. These virtual servers are included in the virtual server group (VSG). A VSG real server (VSGRS) is installed in the VSG start server group (VSGSSG) to start VSG. A control server monitors the states of the VSG and the VSGSSG and controls these groups and the load balancer according to the situation. The power-saving control program and the high-availability control program are installed on the control server. Each control program can operate independently. The

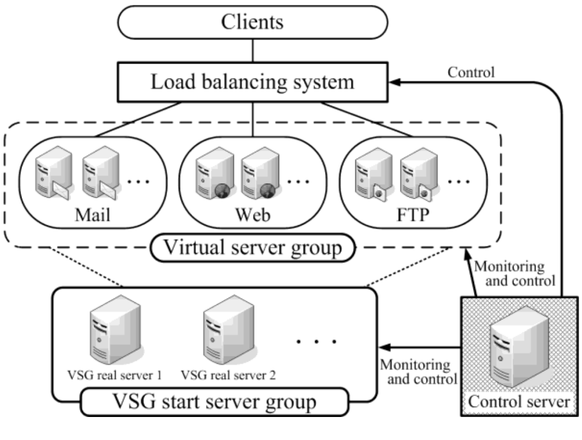


Fig. 2 Configuration of the PSHASS.

hybrid operation method alternately operates each control program. The power-saving control program suspends real servers that do not start virtual servers by means of gathering virtual servers on a particular real server, and the high-availability control program recovers the function of the virtual server that is generating trouble by means of a virtual server that is prepared as a backup. As a result, the PSHASS realizes both power savings and high availability.

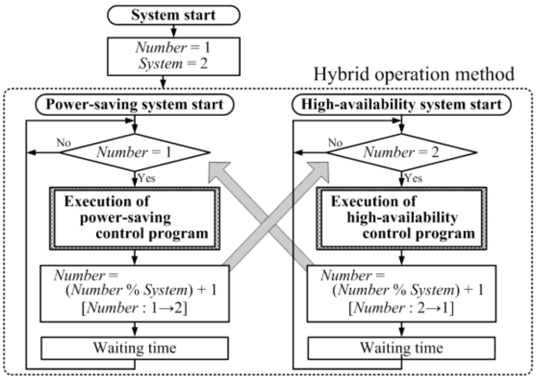
3.2 Hybrid operation method

The basic concept of the hybrid operation method is shown in Fig. 3. The hybrid operation method alternately operates two systems and realizes the function obtained by the systems. Figure 3(a) shows the control timing for the power-saving system and the high-availability system, and Fig. 3(b) shows the configuration files that are used to start the PSHASS and the shared files that are used by these systems. In Fig. 3(a), the percentage sign (%) represents a modulus operator. The variable "*System*" indicates the number of start systems, and the variable "*Number*" indicates the control program to be executed by either system. At the time at which the system starts, the value of *Number* is 1 and the value of *System* is 2, and the two systems start at the same time. Since *Number* = 1, the power-saving control program is started. When the program ends, the *Number* becomes 2, according to the following equation:

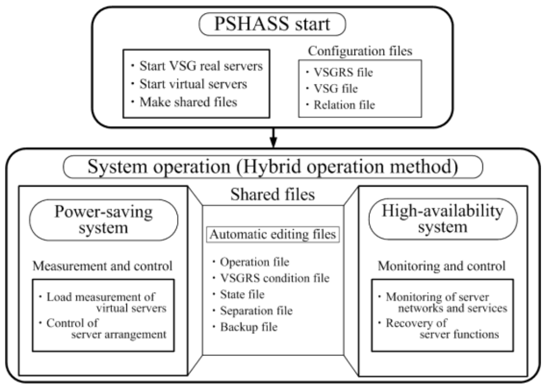
$$Number = (Number \,\%\, System) + 1. \quad (1)$$

The processing program of the power-saving system enters the waiting time state. Since *Number* becomes 2, the high-availability control program is started. When the program ends, *Number* becomes 1 according to Equation (1). The processing program of the high-availability system enters the waiting time state. The two systems are executed alternately by this method.

Figure 3(b) shows the required constitution files to start and the operation of the PSHASS. The VSGRS file is recorded in the detailed configuration of the real servers, and the VSG file is recorded in the detailed configuration of the virtual servers. The relation data of virtual servers and real servers that start virtual servers are recorded in the Relation file. The shared files that the power-saving system and the high-availability system use are automatically generated at the time the PSHASS starts, and the files are edited by the two systems. The power-saving system and the high-availability system are independent systems. Therefore, the PSHASS uses shared files that can grasp the state of each system in order to prevent false operation of either system. The power-saving system measures the load of a virtual server to offer services to clients and realizes power savings by means of the optimum server arrangement depending on the state of the load. The high-availability system examines the state of the networks and services for the virtual servers and the real servers, and if trouble occurs, the system recovers its function automatically.



(a) Control timing of the power-saving and high-availability systems



(b) Shared files to be used by the power-saving and high-availability systems

Fig. 3 Basic concept of the hybrid operation method.

3.3 Configuration file

The formats of the fundamental server files for the PSHASS are listed in Table 1. The PSHASS starts virtual servers based on the contents of these files and constitutes the VSG. The specifications of each VSGRS are recorded in the VSGRS file, which also contains the memory sizes, the IP address, and the MAC address. The information of each virtual server for constructing the VSG is recorded in the VSG file, which contains the service group name, the directory path of the virtual server system file, the host name, the IP address, the memory sizes, the number of CPUs, and the name of the service offered by the virtual server. The memory sizes and the number of CPUs recorded in the VSG file are set automatically at the start time of the virtual server. The Relation file indicates the relations between the VSGRS, the virtual server, and the backup server, which is a virtual server. The IP address of the VSGRS, the host name of the virtual server, and the host name of the backup server are recorded in the Relation file. If trouble occurs on the virtual server that is started from the VSGRS, the backup server is started based on the contents of the Relation file and recovers the function of the server that is generating the trouble. The PSHASS uses the backup server because the backup server can recover the function of the virtual server that experienced trouble, even if a system disk of the virtual server is corrupted.

Table 1 Formats of the VSGRS, VSG, and Relation files.

VSGRS file

Memory [MB]	Real: IP address	MAC address
8,192	192.168.1.1	00:24:81:87:F6:AF
⋮		

VSG file

Service group	Path	Host name	Virtual: IP address	Memory [MB]	CPU	Service name
Mail	/dir/mail1/	mail1	192.168.1.51	2,048	1	smtp
FTP	/dir/ftp1/	ftp1	192.168.1.52	1,024	1	ftp
⋮						
Web	/dir/link/web/	webb	192.168.1.60	2,048	2	http

Relation file

Real: IP address	Virtual: host name	Backup: host name
192.168.1.1	mail1	mailb
192.168.1.2	ftp1	ftpb
⋮		

3.4 Automatic editing file

The PSHASS starts based on the VSGRS, VSG, and Relation files listed in Table 1, and generates the automatic editing file listed in Table 2. The detailed information of the starting virtual server is recorded in the Operation file, which contains the IP address of the VSGRS starting the virtual server, the directory path of the virtual server system file, the host name of the virtual server, the IP address of the virtual server, the service group name, and the name of the service offered by the virtual server. The VSGRS condition file indicates the start condition of the VSGRS, and the IP address of the VSGRS, the memory sizes, which can be assigned to the virtual server, and the number of starting virtual servers are recorded in the VSGRS condition file. The State file indicates the mode condition of the power-saving system, and the service group name, the mode condition of the group (Heavy, Moderate, or Light), and the IP address of the VSGRS, which starts the virtual server in the power-saving state, are recorded in the file. The Separation and Backup files may be necessary, depending on the situation of the system operation. The data of the virtual server, which is not accessed in the power-saving state by the control of the load balancer, are deleted from the Operation file, and the data are recorded in the Separation file. In addition, if there is trouble with the network or the service of the virtual server, then the PSHASS automatically recovers the function by means of

Table 2 Formats of automatic editing files for the PSHASS.

Operation file

Real: IP address	Path	Host name	Virtual: IP address	Service group	Service name
192.168.1.1	/dir/mail1/	mail1	192.168.1.51	Mail	smtp
192.168.1.2	/dir/ftp1/	ftp1	192.168.1.52	FTP	ftp
⋮					

VSGRS condition file

Real: IP address	Available memory size [MB]	Number of virtual servers
192.168.1.1	3,072	2
192.168.1.2	2,048	2
⋮		

State file

Service group	Condition	Base server Real: IP address
Mail	Moderate	192.168.1.1
FTP	Moderate	192.168.1.1
⋮		

Separation file

Real: IP address	Path	Host name	Virtual: IP address	Service group	Service name

Backup file

Virtual server host name (Trouble server)	Virtual server host name (Backup server)

the backup server. The host name of the virtual server that experienced trouble and the host name of the backup server are recorded in the Backup file.

4. Construction of the power-saving system

4.1 Mode operation of the power-saving system

The power-saving system adopts a mode segmentation method and is configured for four different modes: Power-saving regular, Moderate, Light, and Heavy. In the Power-saving regular mode, the load of the service group, which is composed of virtual servers providing services to clients, is measured and is classified as belonging to the Light load class, the Moderate load class, or the Heavy load class, and the control state shifts to a mode depending on the load class. If the load belongs to the Light load class, the system deals with the state using the Light mode. If the load belongs to the Moderate load class, the system deals with the state using the Moderate mode. Finally, if the load belongs to the Heavy load class, the system deals with the state using the Heavy mode.

Figure 4 shows an example of construction for each mode on the power-saving server system. In the figure, RS1, RS2, and RS3 indicate real servers, and A, B, and C indicate the group name of the service that is offered by the virtual server. When the system starts, each service group is constructed with redundant virtual servers. The regular mode judges the load of each service group, and the Moderate mode is maintained if the normal load state continues. In addition, Fig. 4 shows a shift to the Light mode when the load of all service groups is judged to belong to the Light load class. In order to realize effective power-savings, the virtual server for service group C is started on RS1, and RS2 and RS3 are thereafter suspended. In this way, the power consumption becomes a third of that in the Moderate mode. When service group A is judged to belong to the Heavy load class in the regular mode, the control state shifts to the Heavy mode,

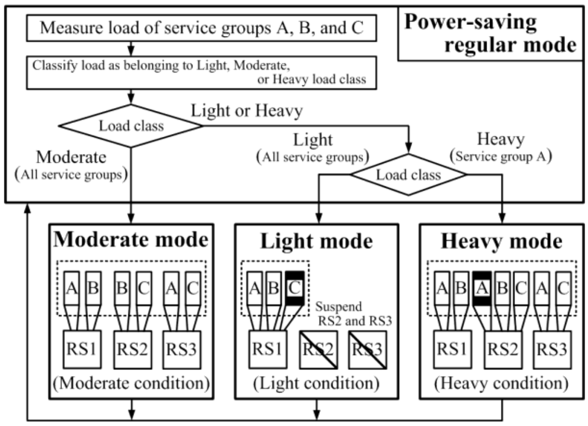


Fig. 4 Configuration of the power-saving system.

and the virtual server for service group A is started on RS2 in order to distribute the load. As a result, the load per virtual server is reduced. The state of the virtual server in each mode is recorded in the Operation file, the VSGRS condition file, the Separation file, and the State file, all of which are listed in Table 2. The server condition configured by the Moderate mode is represented by the Moderate condition, and the server condition configured by the Heavy mode is represented by the Heavy condition, and the server condition configured by the Light mode is represented by the Light condition.

4.2 Load measurement and calculation of load

This power-saving system measures the load of every service group, each of which has a virtual server that provides services to clients, and judges when to shift to each mode based on the measured loads. Generally, each server provides several types of services. Therefore, in this system, the CPU utilization rate and the bytes transferred by the network are considered to be the load.

The measurements of the CPU load and the network load are shown in detail in Fig. 5. The following assumes that the OS of the virtual server is CentOS. The directory "/proc" of CentOS stores basic data files for system information. The utilization state of the CPU is recorded as accumulative data from starting a server in the "/proc/stat" file. The system uses the accumulative CPU idle time of all CPUs to calculate the CPU load. The accumulative idle time is obtained by

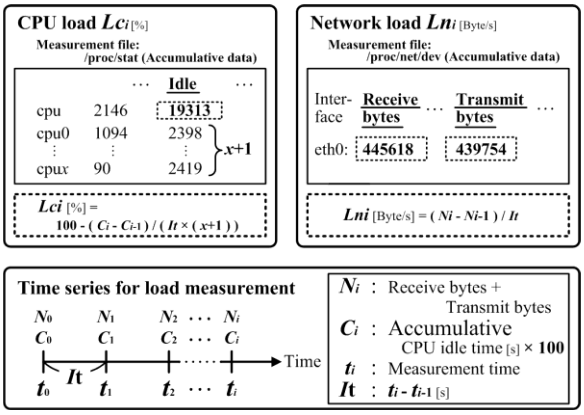


Fig. 5 Loads of the CPU and the network.

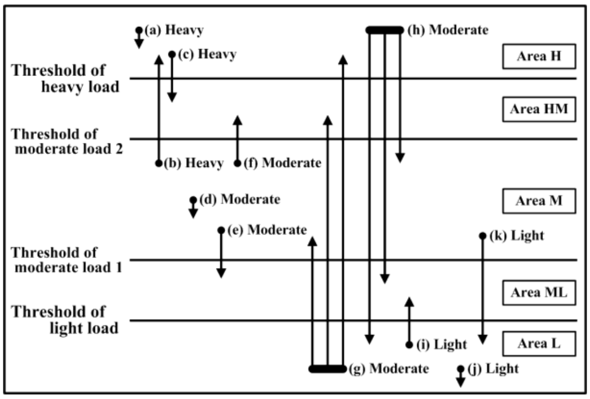


Fig. 6 Threshold of load judgment.

multiplying the real CPU idle time by 100, and 19313 in the figure indicates an accumulative CPU idle time of 193.13 s. The load measurement time is denoted as t_i, and the accumulative CPU idle time at t_i is denoted as C_i. The measurement interval time is denoted as It, and the number of CPUs of the virtual server is denoted as x+1. The CPU load Lc_i is given as follows:

$$Lc_i\,[\%] = 100 - (C_i - C_{i-1}) / (It \times (x + 1)). \quad (2)$$

The utilization state of the network is recorded as accumulative data in the "/proc/net/dev" file. The system uses the total value of receive bytes and transmit bytes as the calculation data of the network load. The load measurement time is denoted as t_i, and the total receive bytes and transmit bytes at t_i is denoted as N_i. The measurement interval time is denoted as It. The network load Ln_i is defined as follows:

$$Ln_i\,[\text{Byte/s}] = (N_i - N_{i-1}) / It. \quad (3)$$

4.3 Classification of load class

Figure 6 shows the threshold by which to judge the load. The power-saving system classifies the load as belonging to the Light, Moderate, or Heavy load class, and sets four thresholds, i.e., heavy load, moderate load 1, moderate load 2, and light load. As shown in the figure, loads that exceed the heavy load threshold are represented by area H. Loads that fall between the heavy load and moderate load 2 thresholds are represented by area HM. Loads that fall between the moderate load 2 and moderate load 1 thresholds are represented by area M. Loads that fall between the moderate load 1 and light load thresholds are represented by area ML. Finally, loads that are lower than the light load threshold are represented by area L. The vectors indicating loads (a) through (k) show the changes in the loads. Black dots indicate the initial state, and the arrowhead indicates the final state. If the load state shifts to area HM or area ML, as in the case of loads (c), (e), (f), and (i), the load is classified as belonging to the initial state. When the load state shifts from area L to area H, as in the case of load (g), or the load state shifts from area H to area L, as in the case of load (h), or the load state shifts to area M, the load is judged to belong to the Moderate load class. When the load state shifts from area M to area H, as in the case of load (b), the load is judged to belong to the Heavy load class. When the load state shifts from area M to area L, as in the case of load (k), the load is judged to belong to the Light load class.

4.4 Judgment of mode change

The load judgment of the service group and the judgment of a mode change are shown in Fig. 7. As shown in Fig. 5, the load for each service group consists of both a CPU load and a network load. Thus, the load class of the service group is defined as belonging to one of heavier classes in the Load class (CPU) or in the Load class (network). For example, when the Load class (CPU) is light and the Load class (network) is heavy, the load of the service group is judged to be heavy. Regarding the mode change, we propose

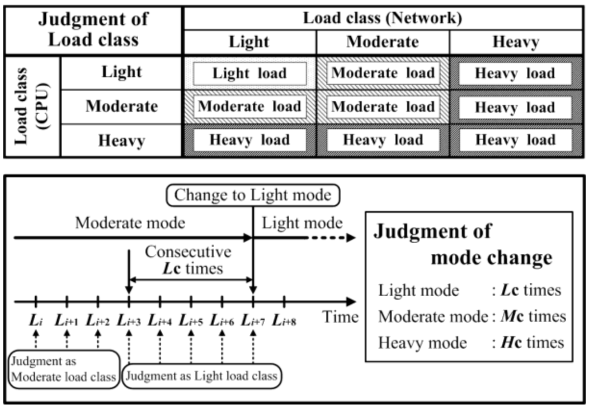


Fig. 7 Load judgment and mode change.

basing it on the number of consecutive times that the load class is judged to be the same, as shown in the figure, because the environment of the server systems varies. The control state changes to the Light mode from the Moderate mode when the load is judged to be in a light load state Lc times consecutively in the Moderate mode. The control state changes to the Moderate mode if the load is judged to be moderate Mc times consecutively, and the control state changes to the Heavy mode if the load is judged to be heavy Hc times consecutively.

5. Characteristics of the high-availability server system

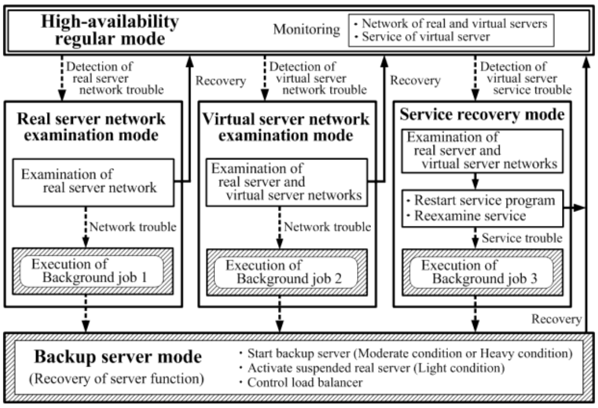
The PSHASS has Heavy, Moderate, and Light modes, which are controlled by the power-saving system, and high availability is realized in each state. As shown in Fig. 4, the server condition configured by the Moderate mode on the power-saving system is represented by the Moderate condition, and the server condition configured by the Heavy mode is represented by the Heavy condition. Finally, the server condition configured by the Light mode is represented by the Light condition.

5.1 Mode operation of the high-availability system

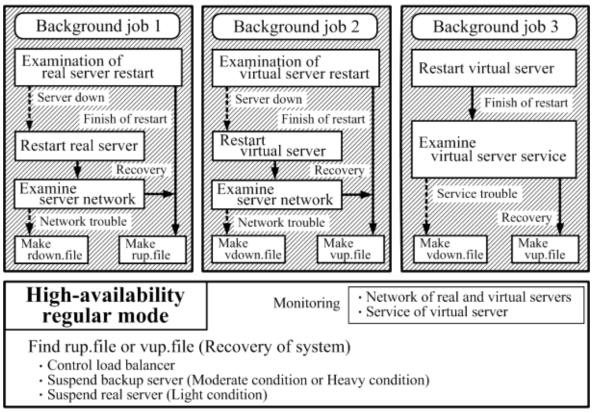
5.1.1 Monitoring and recovery of server function

Figure 8(a) shows the operation flow of the high-availability system. The system introduces a mode segmentation method and is configured for five different modes: High-availability regular, Real server network examination, Virtual server network examination, Service recovery, and Backup server. In the High-availability regular mode, networks and services on virtual servers in the VSG and networks on each VSGRS are regularly monitored based on the Operation file listed in Table 2. The control state shifts to the Real server network examination mode if trouble with the VSGRS network is detected in the regular mode. The Real server network examination mode examines the VSGRS network. The control state returns to the regular mode if the network is normal. However, if the network is judged to be abnormal, the control state shifts to the Backup server mode after executing Background job 1 as a background process. The control state shifts to the Virtual server network examination mode if trouble with the virtual server network is detected in the regular mode. This mode examines the real server that starts the virtual server and the virtual server networks. The control state returns to the regular mode if the network is normal. However, if the network is judged to be abnormal, the control state shifts to the Backup server mode after executing Background job 2 as a background process.

The control state shifts to the Service recovery mode if trouble with the virtual server offering services to clients is detected in the regular mode. This mode recovers the service functions by restarting the service program and reexamines the offering state of services to clients. If the state is normal, the control state returns to the regular mode, but if the state is judged to be abnormal, the control state shifts to the Backup server mode after executing Background job 3 as a background process. The Backup server mode starts a backup server in the Moderate or Heavy condition and activates a suspended VSGRS in the Light condition. Thereafter, this mode controls the load balancer and recovers the function of the inoperative server. If the server function is recovered by the backup server, the information is recorded in the Backup file listed in Table 2.



(a) Operation flow for dealing with various types of trouble by the mode segmentation method



(b) Roles of Background jobs 1, 2, and 3

Fig. 8 Operation flow of the high-availability system.

5.1.2 System recovery

Figure 8(b) shows the operation of Background jobs 1, 2, and 3, as well as the operation according to the results of each Background job in the regular mode. If the operation of a target server to be monitored is abnormal, intentionally restarting the server is considered. Examining and restarting a server that has been shut down because of trouble is time consuming. Hence, if the server function is recovered after the examination, the server function recovery time increases. Therefore, in the present study, the examination is performed as a background process.

In Background job 1, the VSGRS that detected the network trouble is examined in order to determine whether this VSGRS is in the restart state, and when the VSGRS network recovers after restart, a rup.file that records the IP address of the VSGRS is created. The VSGRS is restarted if the VSGRS is judged not to be in the restart state and the VSGRS network is reexamined. If the VSGRS network is recovered, a rup.file that records the IP address of the VSGRS is created. However, if the VSGRS network is not recovered, the rdown.file that records the IP address of the VSGRS is created.

In Background job 2, the virtual server that detected the network trouble is examined in order to determine whether it is in the restart state, and when the virtual server network recovers after restart, the vup.file that records the IP address of the virtual server is created. The virtual server is restarted if it is judged not to be in the restart state and the virtual server network is reexamined. If the virtual server network is recovered, the vup.file that records the IP address of the virtual server is created. However, if the virtual server network is not recovered, the vdown.file that records the IP address of the virtual server is created.

In Background job 3, the virtual server that detected the service trouble is restarted, and the service for the virtual server is examined. If the virtual server service is recovered, the vup.file that records the IP address of the virtual server is created. However, if the virtual server service is not recovered, the vdown.file that records the IP address of the virtual server is created. In the regular mode, if the rup.file or the vup.file is found, this mode controls the load balancer to return the server system to the normal condition, and if the server system is in the Moderate or Heavy condition, this mode suspends starting the backup server, whereas if the server system is in the Light condition, this mode suspends activating the VSGRS.

5.2 Service examination

Figure 9 shows the method used to examine the service in the high-availability server system. The Expect programming language is used for the service examination. This enables the PSHASS to correctly monitor the offering state of services to clients because this language has a time-out

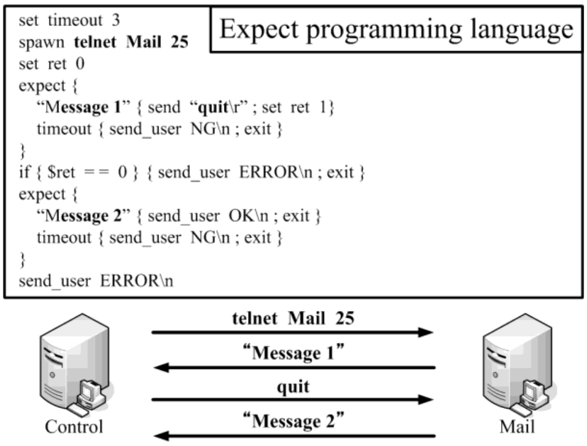


Fig. 9 Service examination using the Expect programming language.

function and performs processing based on the response from a server, which is accessed by a client. In the figure, Mail is the target server to be monitored, and the Control server monitors the simple mail transfer protocol (SMTP) on the target server. First, the service examination routine on the Control server accesses port number 25 on the Mail server by a telnet command, and Message 1 is sent in reply by the server program dealing with SMTP. If the message is normal, the service examination routine sends the command “quit” to Mail. When Message 2 is sent in reply by Mail, the routine returns the response “OK” to the PSHASS. If Message 1 or Message 2 is not normal, the routine returns the response “ERROR” to the PSHASS, and when Message 1 or Message 2 is not sent before time-out, the routine returns the response “NG” to the PSHASS. The PSHASS judges that the service operates normally when the routine responds “OK”, and judges that the service operates abnormally when the routine responds “ERROR” or “NG”.

5.3 Service recovery method under Moderate and Light conditions

The method of dealing with trouble that occurs on the VSGRS or the virtual server with the Moderate and Light conditions is shown in Fig. 10. The Moderate and Light conditions in this figure assume the case in which all service groups are in the same condition. In addition, we omit the method for dealing with trouble in the Heavy condition because the method in the Heavy condition is the same as that for dealing with trouble in the Moderate condition. We herein assume that VSGRS 1 is RS1, that VSGRS 2 is RS2, and that VSGRS 3 is RS3. We also assume that V_{M1} and V_{M2}, which are virtual servers, are included in the Mail service group, that V_{W1} and V_{W2} are included in the Web service group, and that V_{F1} and V_{F2} are included in the FTP service group. Moreover, we assume that V_{Mb} is the backup server of the Mail service group, V_{Wb} is the backup server of the Web service group, and V_{Fb} is the backup server of the FTP service group. In the Moderate condition, two types of virtual servers are started on three real servers, and the virtual servers of each service group are made redundant.

Figure 10(a) shows the method used to deal with trouble in one virtual server in the Moderate condition, and V_{Mb} is started on RS3 to recover the function of V_{M1}. Figure 10(b) shows the method used to deal with trouble in one real server in the Moderate condition, and V_{Mb} and V_{Wb} are started on RS2 and RS3 to recover the functions of V_{M1} and V_{W1}. In the Light condition, the virtual server of each service group is not made redundant in order to realize high-efficiency power savings. V_{M1}, V_{W1}, and V_{Fb} are started on RS1, and RS2 and RS3 enter the suspended state. V_{F1} and V_{M2}, which are started on RS2, and V_{W2} and V_{F2}, which are started on RS3, are shown in dashed rectangles, indicating the state whereby there is no access from the client to control the load balancer. Figure 10(c) shows the method used to deal with trouble in one virtual server in the Light condition. In order to recover the function of V_{M1}, RS2 is activated from the suspended state, and access from clients to V_{M2} on RS2 is enabled by controlling the load balancer. Figure 10(d) shows the method used to deal with trouble in one real server in the Light condition. In order to recover the functions of V_{M1}, V_{W1}, and V_{Fb} which are started on RS1, both RS2 and RS3 are activated from the suspended state, and access from clients to V_{F1} and V_{M2} on RS2 and V_{W2} on RS3 is enabled by controlling the load balancer. If there is trouble with the real server or the virtual server, the client is not affected in the Moderate condition,

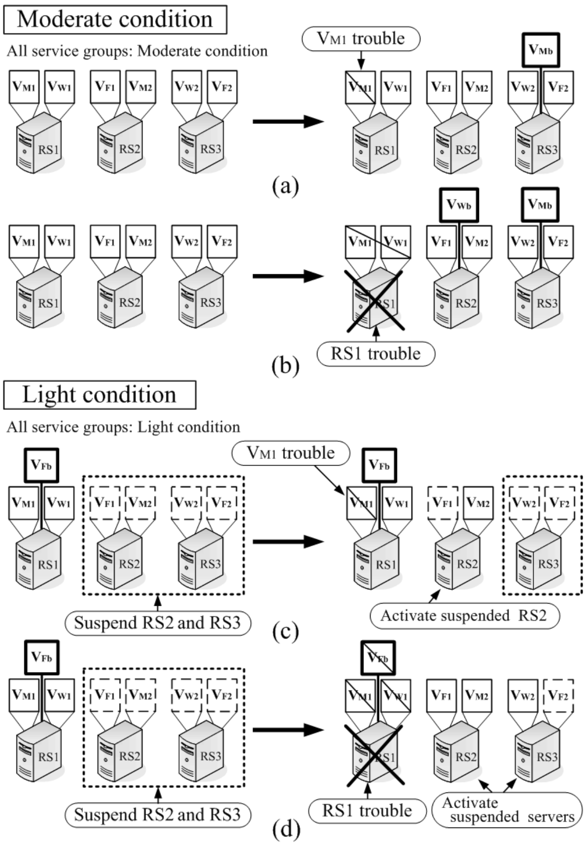


Fig. 10 Service recovery function in the Moderate and Light conditions.

whereas the time until recovery of the server function affects the client in the Light condition, because the virtual server of each service group in the Moderate condition is redundant, whereas this is not the case in the Light condition.

6. Power-saving High-availability Server System

6.1 Experimental system

An experimental PSHASS system is shown in Fig. 11. The experimental system consists of a Control server, which has the function of the network file system (NFS), three VSGRSs, a load balancer, a client, two 1000BASE-T switching HUBs, and a 100BASE-TX switching HUB. The VSG has virtual servers that offer services to clients, and the VSGSSG has VSGRSs that start virtual servers in the VSG. The VSG includes three service groups: Mail, Web, and FTP. Two virtual mail servers are included in the Mail service group, two virtual web servers are included in the Web service group, and two virtual FTP servers are included in the FTP service group. Each service group is made redundant via virtual servers. In the experimental system, a postfix server program that deals with SMTP is installed on the virtual mail server, an httpd server program that deals with HTTP is installed on the virtual web server, and a vsftpd server program that deals with FTP is installed on the virtual FTP server. These server programs offer server connections to clients. Two types of system data for starting virtual servers are saved in each VSGRS, and three types of system data for starting backup servers are saved in the Control server. In order to offer the same service to clients, the redundant virtual servers connect to the Control server with NFS connections, and each VSGRS connects to the Control server with an NFS connection to start the backup server that uses the system disk in the Control server. Each VSGRS has three network interface cards (NICs), and the virtual server in the VSG is connected to Switching HUB 1 or Switching HUB 2 through the VSGRS starting the virtual server. Switching HUB 1 is used to offer services to clients, and Switching HUB 2 is used as the NFS connection to the virtual server and the Control server. Furthermore, Switching HUB 2 is also used as the NFS connection to the VSGRS and the Control server.

The specifications of the server hardware and the virtual server on the experimental PSHASS system are listed in Table 3. CentOS (64-bit), which is often adopted as an OS for servers, is installed on both the load balancer and

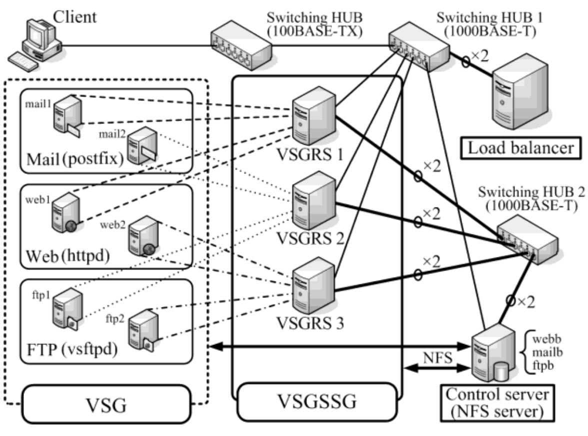


Fig. 11 Experimental PSHASS system.

the real server. The specifications of each VSGRS are similar, and the CPU is a Core i7-960. The memory size is 8,192 MB, and VMware Player and VIX API are installed in each VSGRS. The Control server has a Core i7-930 CPU and 6,144 MB of memory. Each VSGRS and Control server has three NICs, and the load balancer has two NICs. UltraMonkey-L7 is installed on the load balancer for software processing. Each virtual server is assigned one CPU, 1,024 MB of memory, and two NICs, and 64-bit CentOS is installed on each virtual server. As a service daemon, the Mail server uses postfix, the Web server uses httpd, and the FTP server uses vsftpd. Each VSGRS uses a character user interface, and each virtual server starts in the state in which the window is not displayed.

The configuration of the network in the PSHASS is shown in detail in Fig. 12. Each VSGRS and Control server has three NICs representing Ethernet 0, Ethernet 1, and Ethernet 2. Ethernet 0 for the VSGRS connecting Switching HUB 1 is used to offer services to clients, and Ethernet 1 and Ethernet 2 connecting Switching HUB 2 are used for the NFS connection and the control with the PSHASS. Ethernet 0 for the virtual server establishes a bridge connection to Ethernet 0 for the VSGRS, and Ethernet 1 for the virtual server establishes a bridge connection to Ethernet 2 for the VSGRS. Furthermore, M, W, and F in this figure represent the virtual server system data. The regular virtual server, which

Table 3 Specifications of the experimental PSHASS system.

Specifications	VSGRS 1 / VSGRS 2 / VSGRS 3 (VSGSSG)	Control server (NFS server)
CPU	Core i7-960 (3.20 GHz / TB: 3.46 GHz)	Core i7-930 (2.86 GHz / TB: 3.06 GHz)
Memory [MB]	8,192	6,144
Hard disk	SATA 2 (7,200 rpm)	
NIC	3	
System software	VMware Player 5.0.2 VIX-API 1.12.2	
OS	CentOS 6.4 (64-bit)	

Specifications	Virtual server (VSG)	Specifications	Load balancer
CPU	1	CPU	Core i7-3770 (3.40 GHz / TB: 3.90 GHz)
Memory [MB]	1,024	Memory [MB]	8,192
NIC	2	NIC	2
Service daemon	postfix, httpd, vsftpd	System software	UltraMonkey-L7
OS	CentOS 6.4 (64-bit)	OS	CentOS 6.4 (64-bit)

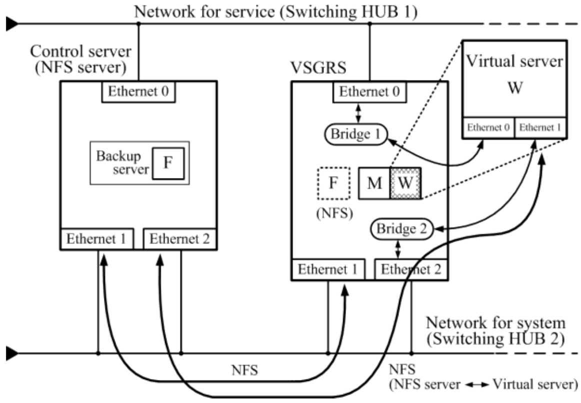


Fig. 12 Configuration of the server network.

comprises the VSG, uses system data M and W, and the backup server starts using system data F. The VSGRS uses system data F through the NFS connection because system data F is saved in the Control server, which means that the backup server can be started using system data F in each VSGRS. In addition, Ethernet 1 of the virtual server has access to the Control server through the NFS connection. Therefore, the redundant virtual server in the service group offers the same services to the clients.

6.2 Characteristics of the experimental system

The characteristics of the experimental system are listed in Table 4. In this experiment, VSGRS 1 is a real server that encounters several types of trouble, and mail1 is a virtual server that encounters several types of trouble. The times in Table 4 are average times given by the command “time” in UNIX. The start time of VSGRS 1 is 54.7 s, and the time to activate the suspended VSGRS 1 is 8.9 s. “Virtual server (Local)” represents the characteristics of mail1 starting from VSGRS 1, and the first start time is 29.1 s, the second start time is 20.4 s. This time lag is assumed to be affected by disk cache processing of the computer system because the virtual server is software. “Backup server (NFS)” represents the characteristics of the backup server starting from VSGRS 1 using the NFS connection with the Control server that has the backup server system data. The first start time is 41.0 s, and the second start time is 25.0 s.

Table 4 Characteristics of the experimental PSHASS system.

Target server	Real server: VSGRS 1 / Virtual server: mail1				
Performances	VSGRS 1	Virtual server (Local)		Backup server (NFS)	
		First	Second	First	Second
Start time [s]	54.7	29.1	20.4	41.0	25.0
Restart time [s]	61.2	29.2		31.5	
Shutdown time [s]	10.5	9.4		9.4	
Time to suspend active server [s]	6.5			4.0	
Time to activate suspended server [s]	8.9			10.3	

Power-saving system		
Mode (State)	Active real server (Number of active virtual server)	Electric power [w] (Total of 3 VSGRSs)
Light (All service groups)	VSGRS 1 (3)	105
Moderate (All service groups)	VSGRS 1, 2, 3 (6)	287

Condition	Operation for condition change	Necessary time [s]
Moderate → Light	Start virtual server → control load balancer → suspend two active real servers	21.1
Light → Moderate	Activate two suspended real servers → control load balancer → suspend virtual server	18.4

The time to activate the suspended backup server is 10.3 s. Backup servers are started at the start time of the PSHASS and are suspended. For this reason, the start time for the backup server in the PSHASS is shortened by this method. In the power-saving system, the electric power for the VSGSSG is 287 W because three VSGRSs start for the case in which all of the service groups are in the Moderate condition. The electric power for the VSGSSG is 105 W because only VSGRS 1 starts and VSGRS 2 and VSGRS 3 are suspended for the case in which all of the service groups are in the Light condition. The time required to shift all service groups from the Moderate condition to the Light condition is 21.1 s, and the time required to shift all service groups from the Light condition to the Moderate condition is 18.4 s.

6.3 Experimental results

The recovery times in the PSHASS and access states on the client are measured in an experiment in which trouble is intentionally generated on the network or the service program on mail1 and the network on VSGRS 1. The recovery times are calculated from log data recorded by the PSHASS and the access state is measured based on the examination results for client access using the Expect programming language. Each measured time is the average of multiple experimental results. The recovery times and access states for the case in which trouble occurs on VSGRS 1 or mail1 in the Moderate or Light conditions are listed in Table 5. The time required to recover the server function and the time required to return the server system to the normal state and the access disconnection time form the client are indicates as the "Measured time". "Recovery of server function" indicates the time until the server function is recovered by the backup server after trouble has occurred, and "Recovery of system" indicates the time until returning to the normal state from the trouble state by Background jobs 1, 2, and 3, as shown in Fig. 8(b). The inter-monitoring time is 10 s in the experimental system, and trouble is generated when the inter-monitoring time exceeds five seconds.

We examine two methods by which to generate trouble in the service program, i.e., recovery by service program restart and non-recovery by service program restart. The mode shift for the case in which the server function can be recovered by service program restart is High-availability regular mode → Service recovery mode (Service program restart) → High-availability regular mode, and the necessary recovery time is 6.2 s in the Moderate condition and is 5.6 s in the Light condition. The access disconnection time with the client is 15.7 s in the Light condition. The mode shift for the case in which the server function cannot be recovered by service program restart is High-availability regular mode → Service recovery mode (Non-recovery by service program restart) → Backup server mode → High-availability regular mode, and the necessary recovery time is 19.8 s in the Moderate condition and is 15.9 s in the Light condition. The time until the server system returns to the normal state is 38.7 s in the Moderate condition and is 36.7 s in the Light condition. The access disconnection time with the client is 25.8 s in the Light condition, which is approximately ten seconds longer than the server function recovery time, because of the inter-examination time (five seconds), the examination time of the VSGRS and virtual server networks and the services on virtual servers, and the recovery time.

We examine two methods by which to generate trouble on the virtual server network, i.e., an intentional restart and a shutdown of mail1 are examined. The mode shift for the case of an intentional restart is High-availability regular mode → Virtual server network examination mode → Backup server mode → High-availability regular mode, and the required recovery time is 18.7 s in the Moderate condition and is 14.0 s in the Light condition. The time until the server system returns to the normal state is 19.7 s in the Moderate condition and is 15.5 s in the Light condition. The access disconnection time with the client is 23.3 s in the Light condition. The mode shift for in case of a shutdown is High-availability regular mode → Virtual server network examination mode (Restart virtual server) → Backup server mode → High-availability regular mode, and the required recovery time is 15.9 s in the Moderate condition and is 14.5 s in the Light condition. The time until the server system returns to the normal state is 66.3 s in the Moderate condition and is 65.3 s in the Light condition. The access disconnection time with the client is 22.9 s in the Light condition. The access disconnection time is approximately nine seconds longer than the server function recovery time, because of the inter-examination time (five seconds), the examination time of the VSGRS and virtual server networks, and the recovery time.

We examine two methods by which to generate trouble on the real server network, i.e., an intentional restart and a shutdown of VSGRS 1. The mode shift in the case of an intentional restart is High-availability regular mode → Real server network examination mode → Backup server mode → High-availability regular mode, and the required recovery time is 27.2 s in the Moderate condition and is 28.3 s in the Light condition. The time until the server system returns to the normal state is 110.1 s in the Moderate condition and is 120.5 s in the Light condition. The access disconnection time with the client is 34.6 s in the Light condition. The mode shift in the case of a shutdown is High-availability regular mode → Real server network examination mode (Restart real server) → Backup server mode → High-availability regular mode, and the necessary recovery time is 27.3 s in the Moderate condition and is 27.9 s in the Light condition. The time until the server system returns to the normal state is 182.7 s in the Moderate condition and is 178.7 s in the Light condition. The access disconnection time with the client is 34.0 s in the Light condition. The access disconnection time is approximately six seconds longer than the server function recovery time because of the inter-examination time (five seconds), the examination time of the VSGRS network, and the recovery time.

In all cases, the server function is recovered in the server function recovery time, which is indicted as "Recovery of server function" in the table. For trouble on the virtual server, the PSHASS can recover the server function in less than 20 s and can recover the server function in less than 30 s for trouble on the real server. The time until the virtual server and the real server that experienced trouble recover to the normal state is longer than the server function recovery time because examining the server that has shut down for trouble and then restarting the server, as shown in Fig. 8(b), is time consuming. Moreover, for the case in which the server network and the service offered to clients are judged to be in the normal state, the PSHASS controls the load balancer and returns the system to the original state. Since the virtual server providing services to the client is redundant in the Moderate condition, the service stop time for the client is 0 s in all cases. In order to increase the power savings, the virtual server for the service group is not redundant in the Light condition. If trouble occurs in the virtual server or the real server, the service stop time is less than 35 s. However, this is not a significant problem because the server load is light in the Light condition.

The proposed method realizes power-savings and high-availability by alternately operating a power-saving system and a high-availability system. In addition, each system adopts the mode segmentation method. Thus, in comparison with a control method to realize power savings and high availability by means of one control program, the control program for the PSHASS is not complicated. In the experiment of the present study, the Moderate condition and the Light

Table 5 Recovery time of the PSHASS for different types of trouble.

Inter-monitoring time: 10 [s]					
	Trouble list	Condition (Power saving)	Measured time [s]		
			Recovery of server function	Recovery of system (Background job)	Access disconnection (Client)
Service program	Service program stop (Recovery by service program restart)	Moderate	6.2		0.0
		Light	5.6		15.7
	Service program trouble (Non-recovery by service program restart)	Moderate	19.8	38.7	0.0
		Light	15.9	36.7	25.8
Virtual server network	Intentional restart of mail1	Moderate	18.7	19.7	0.0
		Light	14.0	15.5	23.3
	Shutdown of mail1	Moderate	15.9	66.3	0.0
		Light	14.5	65.3	22.9
Real server network	Intentional restart of VSGRS 1	Moderate	27.2	110.1	0.0
		Light	28.3	120.5	34.6
	Shutdown of VSGRS 1	Moderate	27.3	182.7	0.0
		Light	27.9	178.7	34.0

condition are constructed by the control of the power-saving system, and we generated network and service troubles in the server that offers services to clients. Since the PSHASS automatically recovers the server function, even if trouble occurs under both the Moderate and Light conditions, the hybrid operation method can be applied to a practical server system for both power savings and high availability. Furthermore, since the recovery of each server function is performed by starting the backup server or the VSGRS in the Backup server mode, a good understanding of these start times listed in Table 4 is important when constructing the PSHASS.

7. Conclusions

We herein proposed a hybrid operation method that realizes power savings and high availability and demonstrated that power-saving and high-availability systems that are independent can be alternately operated using a shared file. A PSHASS that adopts the proposed method was constructed.

In the power-saving system, the functions of the Power-saving regular mode, the Moderate mode, the Light mode, and the Heavy mode were demonstrated. The control state shifts to various modes depending on the load state of the virtual servers offering services to clients. The Heavy mode is entered when loads are heavy and the Light mode realizes power savings by suspending real servers. The electric power required by the real server that starts the virtual servers was 287 W in the Moderate condition and 105 W in the Light condition.

In the high-availability system, the functions of the High-availability regular mode, the Virtual server network examination mode, the Real server network examination mode, the Service recovery mode, and the Backup server mode were demonstrated. The control state shifts to each mode depending on the type of trouble encountered. The recovery method differs depending on whether the system is in a normal condition or a power-saving condition.

Furthermore, an experimental PSHASS system was configured, and experiments were performed. It was demonstrated experimentally that the PSHASS has sufficient reliability to deal with various types of trouble, including failure of the virtual server and failure of the real server that starts the virtual servers. The times for the service recovery time were measured. Even if a real server shuts down, the server function was recovered in less than 30 seconds.

The results of the present study indicate that both power savings and high availability of the server system, which is the fundamental unit of the Internet system, can be realized effectively. Therefore, expansion of Internet services can be expected, affording greater convenience in our daily lives.

References

[1] M. Kitamura and C. Fujihashi, "Analysis of Response Time Characteristics of a Real Computer by Heterogeneous-Task Bursty-Arrival Models", IEICE Trans. Vol. J81-B-I, No. 4, pp. 243-253, 1998-04.

[2] C. Lalit and K. Hemangee K., "Formal Approach for DVS-Based Power Management for Multiple Server System in Presence of Server Failure and Repair", IEEE Trans. Ind. Inform., Vol. 9, No. 1, pp. 502-513, 2013-02

[3] J. Heeseung, K. Hwanju, J. Jae-Wan, L. Joonwon, and M. Seungryoul, "Transparent Fault Tolerance of Device Drivers for Virtual Machines", IEEE Trans. Comput. Vol. 59, No. 11, pp. 1466-1479, 2010-11

[4] K. Ogura and H. Miwa, "Method of Finding Protected Data Centers to Keep Loads of Servers and Links Low during Server Failures", IEICE Technical Report NS2012-271, pp. 617-622, 2013-03

[5] B. Dario, D. Salvatore, L. Francesco, P. Antonio, and S. Marco, "Workload-Based Software Rejuvenation in Cloud Systems", IEEE Trans. Comput. Vol. 62, No. 6, pp. 1072-1085, 2013-06

[6] W. Shengquan, L. Jun, C. Jian-Jia, and L. Xue, "PowerSleep: A Smart Power-Saving Scheme With Sleep for Servers under Response Time Constraint", IEEE Emerg. Sel. Top. Circuit. Syst. Vol. 1, No. 3, pp. 289-298, 2011-09

[7] Y. Souda, T. Mikoshi, and T. Takenaka, "High-Speed Topology Constructive Method for Network Power Saving", IEICE Trans. Vol. J96-B, No. 2, pp. 92-101, 2013-02

[8] Ministry of Economy, Trade and Industry, http://www.meti.go.jp/committee/materials/downloadfiles/g80520c03j.pdf, 2008-05

[9] C. Ying-Chih and W. Chih-Yu, "Distributed Clustering With Directional Antennas for Wireless Sensor Networks", IEEE Sens. Jour. Vol. 13, No. 6, pp. 2166-2180, 2013-06

[10] L. Saulo O. D., P.Angelo, C. Bruna M. J., N. Breno H. M., and A. Gabriela M. da S., "Optimization of Timeout-based Power Management Policies for Network Interfaces", IEEE Trans. Consum. Electron. Vol. 59, No. 1, pp. 101-106, 2013-02

[11] V. Luca, V. Dung P., R. Pier G., C. Piero, C. Divanilson R., W. Shing-Wa, Y. She-Hwa, K. Leonid G., and Y. Shinji, "Energy Efficiency in Passive Optical Networks: Where, When, and How ?", IEEE Netw. Vol. 26, No. 6, pp. 61-68, 2012-11

[12] M. Xu, X. Ji, J. Wu, and M. Zhang, "A Low-Power LDPC Decoder for Multimedia Wireless Sensor Networks", IEICE Trans. Commun. Vol. E96-B, No. 4, pp. 939-947, 2013-04

[13] H. Ito, H. Hasegawa, and K. Sato, "Router Power Reduction though Dynamic Performance Control Based on Traffic Predictions", IEICE Trans. Commun. Vol. E95-B, No.10, pp. 3130-3138, 2012-10

[14] M. Kitamura, "Configuring a Low-cost, Power-saving Multiple Server Backup System: Experimental Results", IEICE Trans. Commun. Vol. E95-B, No. 1, pp. 189-197, 2012-01

[15] S. Kurumatani and M. Toyama, "A Report of Performance per Power Consumption of ARM Based Servers", IEICE Technical Report IN2012-154, pp. 1-6, 2013-03

[16] W. Arahata and M. Bandai, "Fundamental Study on Green Data Center with Server Virtualizations", IEICE Technical Report NS2012-17, pp. 7-10, 2012-05