PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :		(11) International Publication Number:
H04M 15/00	**A1**	**WO 98/54887**
		(43) International Publication Date: 3 December 1998 (03.12.98)

(21) International Application Number: PCT/US98/11516

(22) International Filing Date: 1 June 1998 (01.06.98)

(30) Priority Data:
60/048,437 2 June 1997 (02.06.97) US

(71) Applicant: HARRIS CORPORATION [US/US]; 1025 West NASA Boulevard, Melbourne, FL 32919 (US).

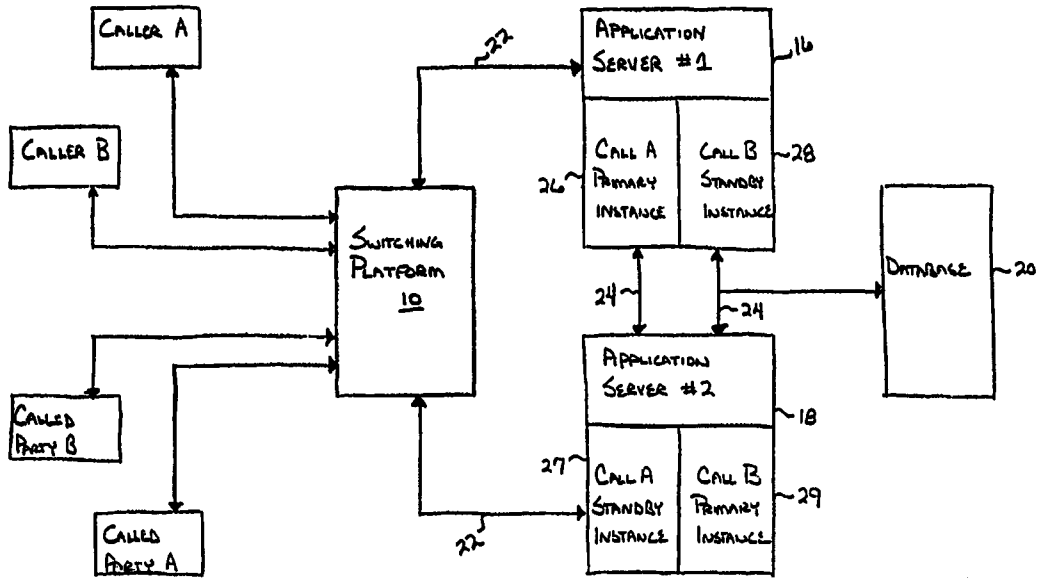
(72) Inventors: BAILIS, Jason; 40 Bridle Path Lane, Novato, CA 94945 (US). SAMSONOV, Max; – (**). HUDSON, Dan; 3445 Pebble Hill Drive, Marietta, GA 30062 (US). KUZNETSOV, Sergey; – (**).

(74) Agents: ROGERS, L., Lawton, III et al.; Rogers & Killeen, Suite 400, 510 King Street, Alexandria, VA 22314 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SERVER LOAD SHARING AND REDUNDANCY AND METHOD



(57) Abstract

A method for an enhanced services system is provided by a server load sharing and redundancy system. The system comprises a switching platform (10), a plurality of caller and called parties' telephones (12, 14) connected to the switching platform; a primary application server (16) containing caller A primary instance (26) and call B standby instance (28); a standby application server (18) containing call A standby instance (27) and call B primary instance (29) wherein both servers process and/or monitor calls placed within the system; and a database (20) storing and maintaining call specific information, and being connected to both primary and standby application servers. Call Tracking Links (22) are established between the primary and standby application servers and the switching platform to support the application server redundance scheme. Links (24) are established to transmit messages between the primary and standby application servers.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

WHAT IS CLAIMED IS:

1. A method of achieving load sharing and of providing redundancy in an enhanced services telephone system comprising the steps of:

(a) providing at least two servers independently linked to the switching platform of the enhanced services telephone system, each of the servers having the capability of monitoring when in an active state (i) the state of a call placed by a subscriber and (ii) the data necessary to bill the call to the subscriber;

(b) activating a first one of the servers to monitor a first call from the switching platform;

(c) establishing under the control of the first server a sync between the first server and a second server from the time that (i) the first call is connected through the switching platform to the receiving number to (ii) the time that the number is disconnected and billed, so that in the event of a failover or a controlled switchover of the first server the second server has all of the data necessary to monitor and bill the first call;

(d) effecting the billing of the first call by the first server in the absence of a failure of the first server ;

(e) activating a second server to monitor a second call from the switching platform thereby sharing the load of calls between servers;

(f) establishing under the control of the second server a sync between the two servers for the second call from the time that (i) the second dialing subscriber is connected through the

switching platform to the receiving number to (ii) the time that the number is disconnected and billed, so that in the event of a failure of the second server the first server has all of the data necessary to monitor and bill the second call; and

(g) effecting the billing of the second call by the second server in the absence of a failure of the second server.

2. In a system with N servers adapted to receive a large number of events to be processed by one of the N servers, the method of achieving both load sharing and redundancy among the N servers comprising the steps of:

(a) applying successive events sequentially to the N servers thereby sharing the load of events among the N servers,

(b) identifying the particular one of the N servers to whom an event is applied as the primary server for each such event;

(c) sharing data relating to each such event between the primary server and a second one of the N servers;

(d) identifying the second one of the N servers for each such event as the secondary server for such event; and

(e) for each event, causing the secondary server to take over the role of primary server in the event of a failure of the primary server thereby providing redundancy for all events.

3. The method of Claim 2 wherein N is 2.

4. The method of Claim 2 wherein N is greater than 2 and wherein each of the N servers is the primary server for 1/N events and the secondary server for 1/N events.

5. A method of achieving load sharing and redundancy in

an N server system adapted to collectively receive a large number of events to be processed comprising the steps of:

(a) applying 1/N of the events to each one of the N servers as the primary server for such event;

(b) sharing data relating to each such event between the primary server and a second one of the N servers as the secondary server for such event; and

(c) for each event, causing the secondary server to take over the role of primary server in the event of a failure of the primary server to thereby provide redundancy for all events.

6. A enhanced services telephone system with load sharing and call processing redundancy comprising:

first and second application servers for monitoring the operation of a telephone switching platform;

a database operatively connected to said first and second application servers, each of said servers having the capability when in an active state of monitoring the state of a call and billing the call upon the completion thereof, means for dividing the receipt of the calls between said servers, each of said servers being in an active state for each of the calls received and sharing the data associated with such call with the other one of said servers so that the other server may take over as the active server in the event of the failover or controlled switchover of the active server.

7. A system for load sharing and redundancy among N servers adapted to collectively receive a large number of events to be processed comprising:

means for applying 1/N of the events to each one of the N servers as the primary server for such event;

means for sharing data relating to each such event between the primary server and a second one of the N servers as the secondary server for such event; and

means for each such event for causing the secondary server to take over the role of primary server in the event of a failure of the primary server to thereby provide redundancy for all events.

8. A system comprising:

a telephone call switching platform;

a first application server;

a second application server;

means for operatively connecting said switching platform to said first application server through a CTI link for the application of 50% of the telephone calls from said switching platform to said first server as the primary server for such 50% of the calls;

means for operatively connecting said switching platform to said second application server through a CTI link for the application of the other 50% of the telephone calls from said switching platform to said second server as the primary server for such other 50% of the calls; and

means for replication of the data related to each call for which one of said first and second servers serves as primary server so that such data is available to the secondary server in the event of a failure of said primary server.

9. A system comprising:

a telephone call switching platform operably connected to the telephones of calling party A, calling party B, called party A and called party B;

a first application server;

a second application server;

means for operatively connecting said switching platform to said first application server through a CTI link for the application of the call from calling party A to called party A ("Call A") for the monitoring thereof by said first application server as the primary server for Call A; and

means for operatively connecting said switching platform to said second application server by a CTI link for the application of the call from calling party B to called party B ("Call B") for the monitoring thereof by said second application server as the primary server for Call B;

a database shared by said first and second application servers,

said first server being operably connected to said database to monitor Call A in a stand-by status and said second server being operably connected to said database to monitor Call B in a stand-by status,

whereby said first and second servers service the same number of calls as primary and in a standby status and provide 100% redundancy for all calls.

10. The system of Claim 9 including means associated with said first and second servers for establishing a sync between said two application servers during only the active period of any call.

11. The system of Claim 9 including means associated with said first and second servers for establishing a sync between said two application servers during the period from call setup through call teardown for any call.

12. The system of Claim 9 wherein each of said first and second servers includes means to effect the billing of only the calls for which said server is the primary server.

SERVER LOAD SHARING AND REDUNDANCY AND METHOD

FIELD OF THE INVENTION

This application claims the benefit of U.S. Provisional Application No. 60/048,437 filed June 2, 1997, the disclosure of which is incorporated herein by reference.

The present invention relates to server redundancy and loadsharing, and more particularly, to a server redundancy and loadsharing system and method having application in an enhanced services environment.

BACKGROUND OF THE INVENTION

Redundancy in prior art server systems is accomplished through the use of a second or standby server which functioned solely as a backup for an on-line primary server. As shown in Figure 1, the primary server 30 performs all the necessary call processing, storing the associated data in shared memory 34. Only upon the failure of the primary server 30 does the standby server 32 become active, retrieving the necessary data from shared memory 34 to continue the call processing. As the standby server only functions in the event of a failed primary server, valuable resources are wasted in maintaining the standby server. Further, since only the information available in the shared memory can be used by the standby server, calls in the setup phase in the primary server may be lost/disconnected. Additionally, wear on the server system is not distributed equally among the individual servers necessitating the more frequent servicing of the primary server.

Also known in the prior art are server systems which perform load sharing. Figure 2 shows a typical loadsharing

system 44 in which multiple servers 40 share the call processing load based on a designated loadsharing scheme implemented by an automatic call distribution (ACD) mechanism. Failure of a server in such a system results in both decreased call processing capacity of the system and the interruption of service to the calls under progress in the failed server.

Accordingly, it is an object of the present invention to provide a novel method and system which obviates many of the problems of the prior art while providing both loadsharing and redundancy capabilities.

It is another object of the present invention to provide a novel method and system which preserves both stable and unstable communications during the controlled switchover from a primary application server to a standby application server.

It is yet another object of the present invention to provide a novel method and system which permits the preservation of communications during the uncontrolled failover of a primary application server.

It is still another object of the present invention to provide a novel method and system to insure the accurate billing of established calls occurring during a controlled switchover or failover of a primary application server.

It is a further object of the present invention to provide a novel loadsharing and redundancy method and system which reduces the amount of servicing required by the individual servers due to the equal distribution of wear on the servers.

These and many other objects and advantages of the present invention will be readily apparent to one skilled in the art to

which the invention pertains from a perusal of the claims, the appended drawings, and the following detailed description of the preferred embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a functional block diagram illustrating the redundancy scheme of a prior art server system.

Figure 2 is a functional block diagram illustrating the loadsharing among servers in a prior art server system.

Figure 3 is a functional block diagram of one embodiment of the system of the present invention.

Figure 4 illustrates the four stages of a typical debit call requiring synchronization between the primary and standby servers.

Figure 5 is a functional block diagram of the interface links supporting the embodiment of Figure 3.

Figure 6 illustrates the interaction between servers of Figure 3 under normal operating conditions.

Figure 7 illustrates the interaction of the servers of Figure 3 under switchover/failover conditions.

Figure 8 is a functional block diagram of an embodiment of the present invention illustrating a loadsharing scheme in which each application server functions in both a primary and a standby capacity.

Figure 9 is a functional block diagram showing a daisy chain implementation of a loadsharing redundancy scheme for one embodiment of the present invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

With reference to Figure 3 where an enhanced services platform is illustrated in the embodiment of a prepaid telephone system, a conventional telephone switching platform 10 such as the Harris Corporation 20/20 switch is connected in a conventional telephone system. Connected to the switching platform 10 are large numbers of subscriber telephones, such as the caller telephone 12 and the called party telephone 14, along with the primary application server 16 and the standby application server 18 which process/monitor the calls placed within the system. The application servers 16, 18 are further connected to the database server 20 which maintain the call specific information (e.g., account balance information).

Each call placed in the system is monitored/processed by a call processing instance of a primary application server 16. In one embodiment of the present invention, each primary call processing instance has a standby peer in a different primary application server 18 which takes over the call processing from the primary instance under certain conditions. For example, in one embodiment, when a failover occurs or a controlled switchover is performed transferring control of the call processing from the primary instance to the standby instance, all stable and unstable calls at the time of the failover/controlled switchover are preserved by the standby instance.

In order to take over the call processing from the primary application, the standby application must have access to the call data of the calls and must independently monitor the call

states on the switching platform 10. As shown in Figure 3, this is accomplished via redundant host interface links ("HIL") Call Tracking Links 22 into the Switching Platform 10 and proprietary links 24 between application servers 16 and 18.

Just before the active application tells the switching platform 10 to put the call through, the active application synchs up with its standby peer and instructs the standby to start monitoring the necessary switching platform ports. The active application waits until the standby application is synched before placing the call. If the active application should fail, the standby application is then able to complete the call transactions.

In one embodiment of the present invention, there are four stages during a typical debit call where the standby instance of the call must be synched up with the primary instance in order for the standby instance to take over administration of a call and barge/disconnect the call to prevent a negative caller account balance from resulting, or to bill the call accurately upon completion. These four stages are the stages B-E shown in Figure 4.

Just before the primary application server 16 tells the switching platform 10 to put the call through (stage B), the primary application server 16 synchs up with its standby application server 18, instructing the standby server to start monitoring the necessary switching platforms ports. Since both servers are using HIL Call Tracking ("HCT"), they are both informed of an answer (stage C) and a disconnect (stage D). At stage E, depending upon the implementation, either one or both

of the call processing instances may then bill the call. In the latter implementation, a carefully constructed primary key is used in the database server to permit only one of the billing attempts and thus the call is billed only once. Either method guarantees that if there had been a failure of the primary after the disconnect (stage D), but before the call had been billed (stage E), the call would certainly have been billed by the standby instance.

With further reference to Figure 4, the points during the processing of a call at which the primary application server 16 may synch with the standby application server 18 are not bounded by the stages B and E and may begin, for instance, during the call setup phase (stage A) and extend through the call tear down phase (stage F).

As shown in the embodiment of Figure 5, two HIL Call Tracking Links 22 may be established to support the application server redundancy scheme. Links T-A and T-B are used for call tracking where it is possible to track the same circuit from two links at the same time.

The application servers will establish the HIL Call Tracking Links 22 by using the Telephony Application Programming Interface (TAPI) available with the NT Server operating system to interface to the HIL2TAPI service providers 24 and 26 and instruct the service provider to open the necessary HIL links 22. The primary application server 16 opens TAPI Line T-A while the standby application server 18 only opens TAPI Line T-B.

While calls are in the setup phase, they are maintained over the links. The primary application server 16, responsible for shuttling all primary call data to the standby application server 18, replicates all internal call data to the standby server. Simultaneously in the role as a standby server, application server 16 receives redundant call information from application server 18. When the calls are established, both application servers 16 and 18 turn on HCT via their lines Link T-A and Link T-B respectively.

The decision to perform switchover is made by the application servers with the HIL2TAPI service providers simply reporting line conditions via LINE CONNECTED/LINE DISCONNECTED TAPI messages. Since the standby application server 18 has already established HCT prior to switchover, the calls in conversation are preserved and can be accommodated by the standby system.

Figures 6-7 illustrate the interaction between the application servers and the remaining system under normal and switchover conditions respectively.

As shown in Figure 6, upon receipt of an incoming call by the active/primary application server, the active server replicates the call data to the standby server, and upon confirmation from the standby server, both servers begin monitoring the call. When the call has been completed, the active server controls the billing of the call, and upon confirmation by the database that the call has been billed, the active server instructs the standby server to terminate further processing of the call.

The call processing under switchover conditions shown in Figure 7 differs only from that of the normal operating conditions shown in Figure 6 in that it is the standby server which controls the billing of the call.

As shown in Figure 8, each application server may function in both a primary and standby capacity. Application server 16 may maintain a primary call instance 26 for CALL A with a standby call instance 27 in application server 18 while maintaining a standby calling instance 28 for CALL B for a primary call instance 29 in server 18. In this loadsharing embodiment, all servers perform primary call processing while continuing to provide redundancy within the system.

Further, the loadsharing redundancy scheme of Figure 8 is not limited to the grouping of servers into pairs, but may be applied in a daisy chain fashion as shown in Figure 9.

While preferred embodiments of the present invention have been described, it is to be understood that the embodiments described are illustrative only and the scope of the invention is to be defined solely by the appended claims when accorded a full range of equivalence, many variations and modifications naturally occurring to those of skill in the art from a perusal hereof.

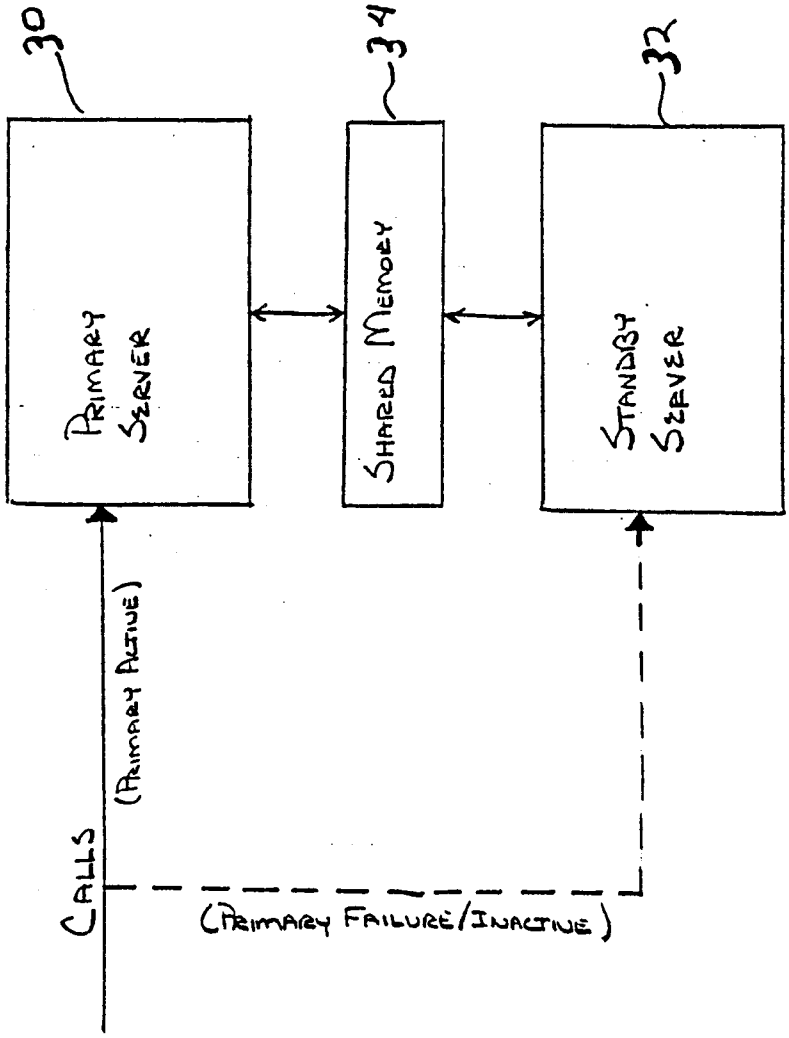


FIG. 1
(PRIOR ART)

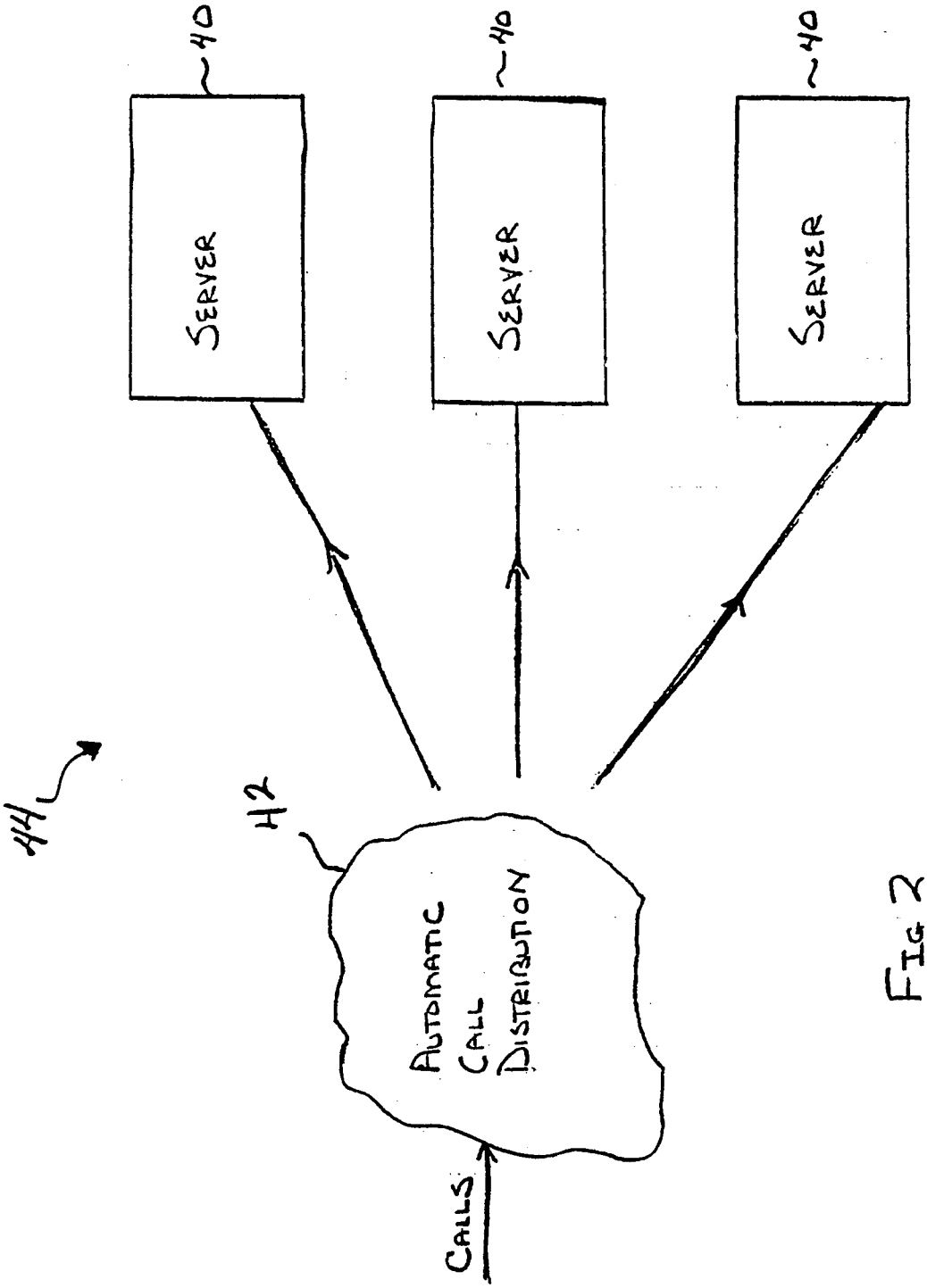


FIG 2
(PRIOR ART)

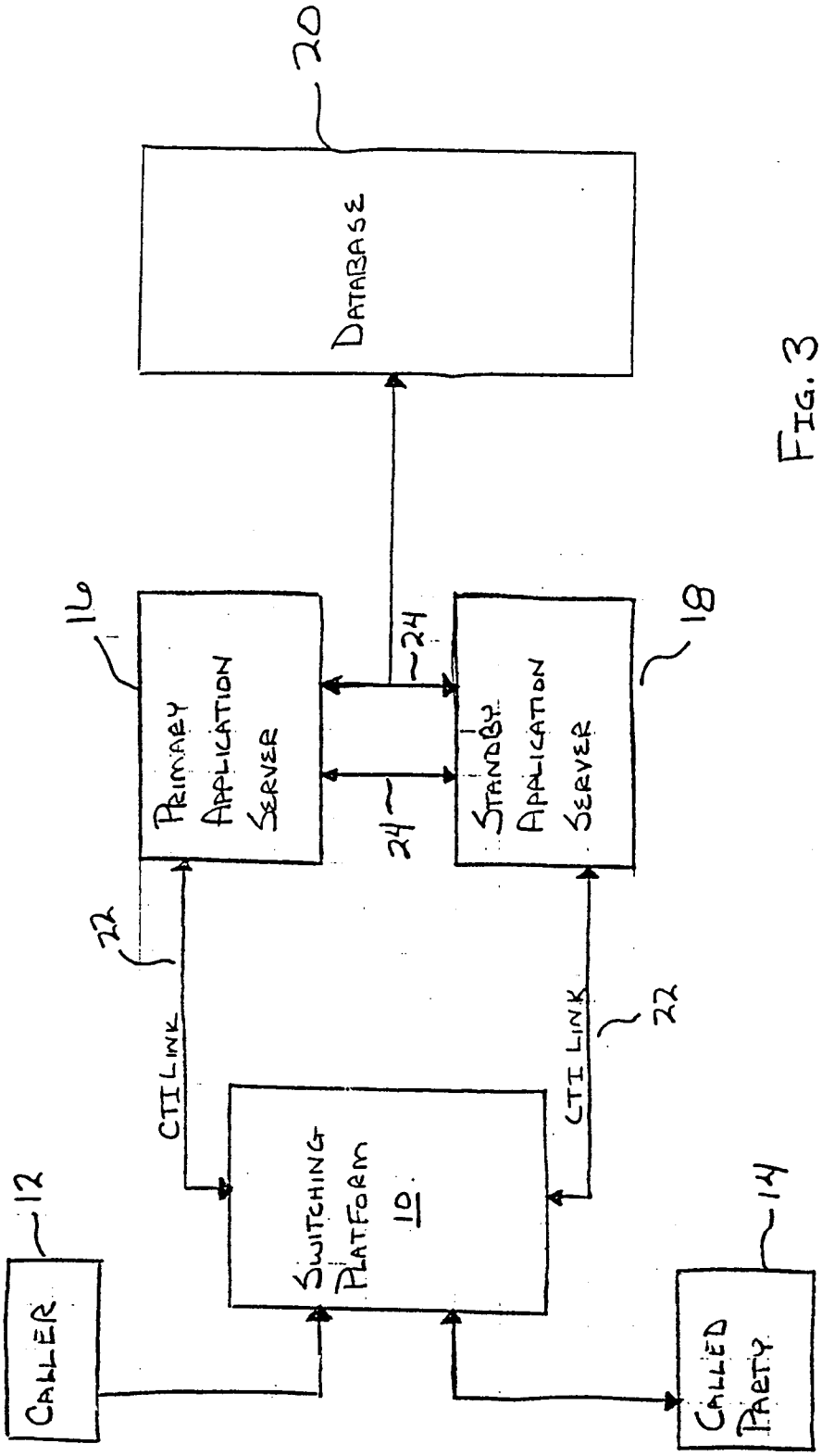

20
DATABASE
16
PRIMARY APPLICATION SERVER
24
24
STANDBY APPLICATION SERVER
18
22
CTI LINK
22
CTI LINK
SWITCHING PLATFORM 10
12
CALLER
14
CALLED PARTY

FIG. 3

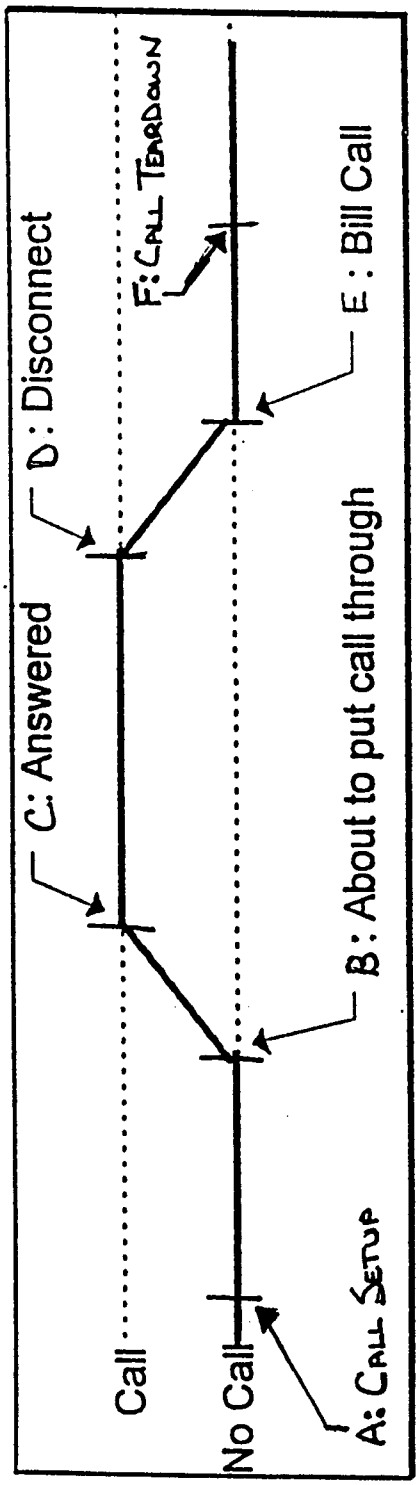


Figure 4

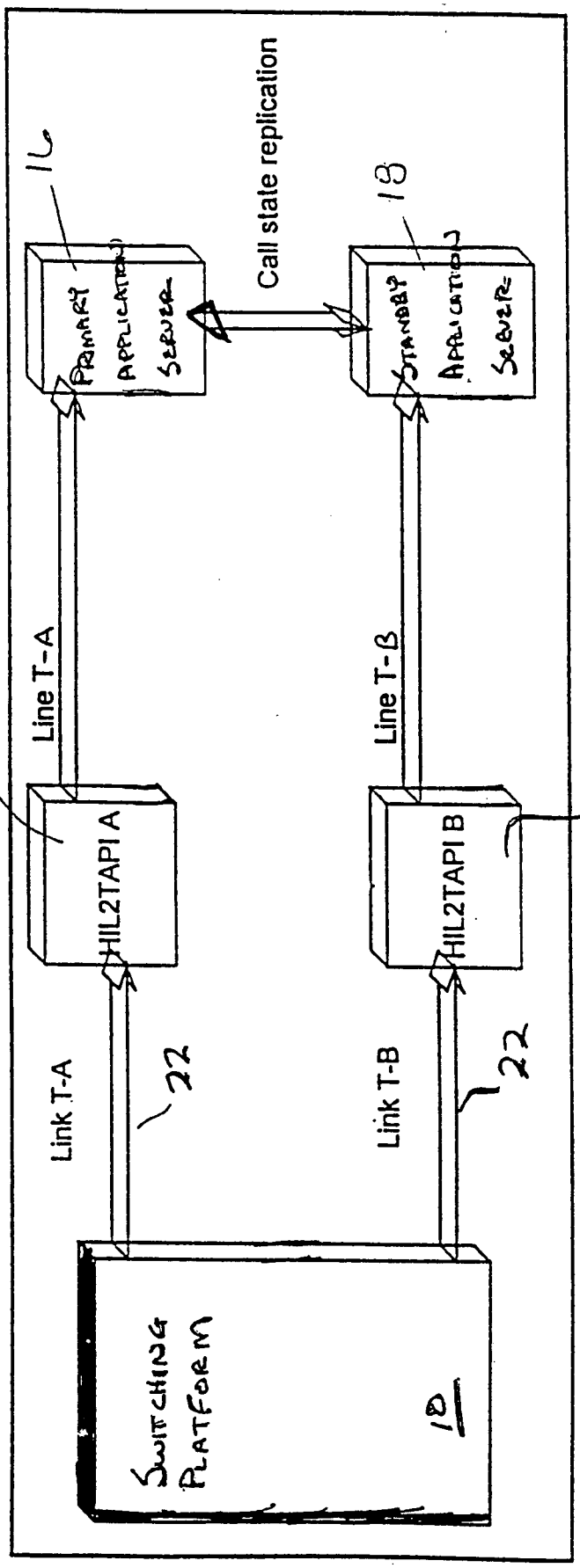


Figure 5

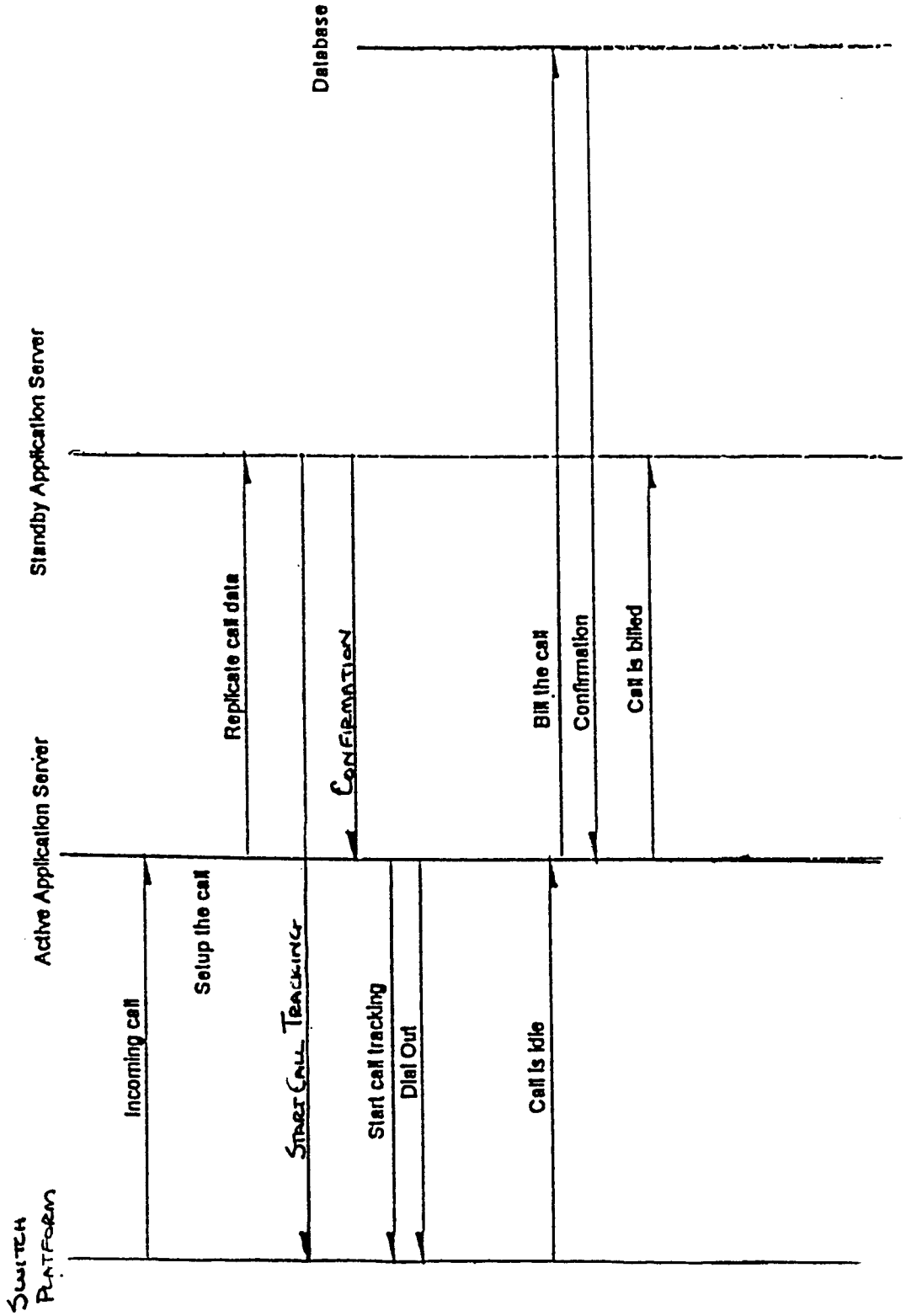

Switch Platform
Active Application Server
Standby Application Server
Database
Incoming call
Setup the call
Start Call Tracking
Replicate call data
Confirmation
Start call tracking
Dial Out
Call is idle
Bill the call
Confirmation
Call is billed

Figure 6

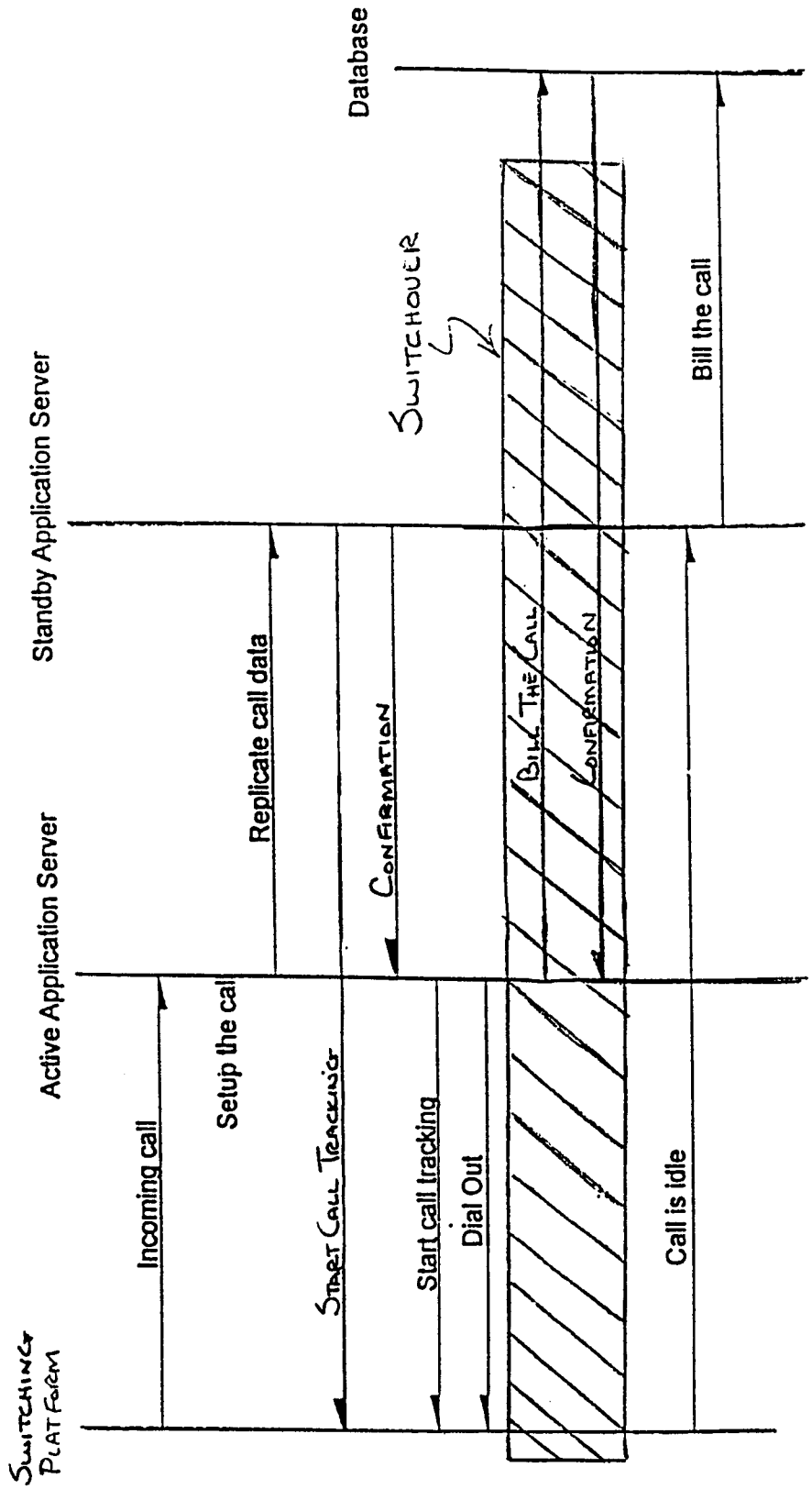

Switching Platform
Active Application Server
Standby Application Server
Database
Incoming call
Setup the call
Replicate call data
Start Call Tracking
Confirmation
Start call tracking
Dial Out
Bill The Call
Confirmation
Switchover
Bill the call
Call is idle

Figure 7

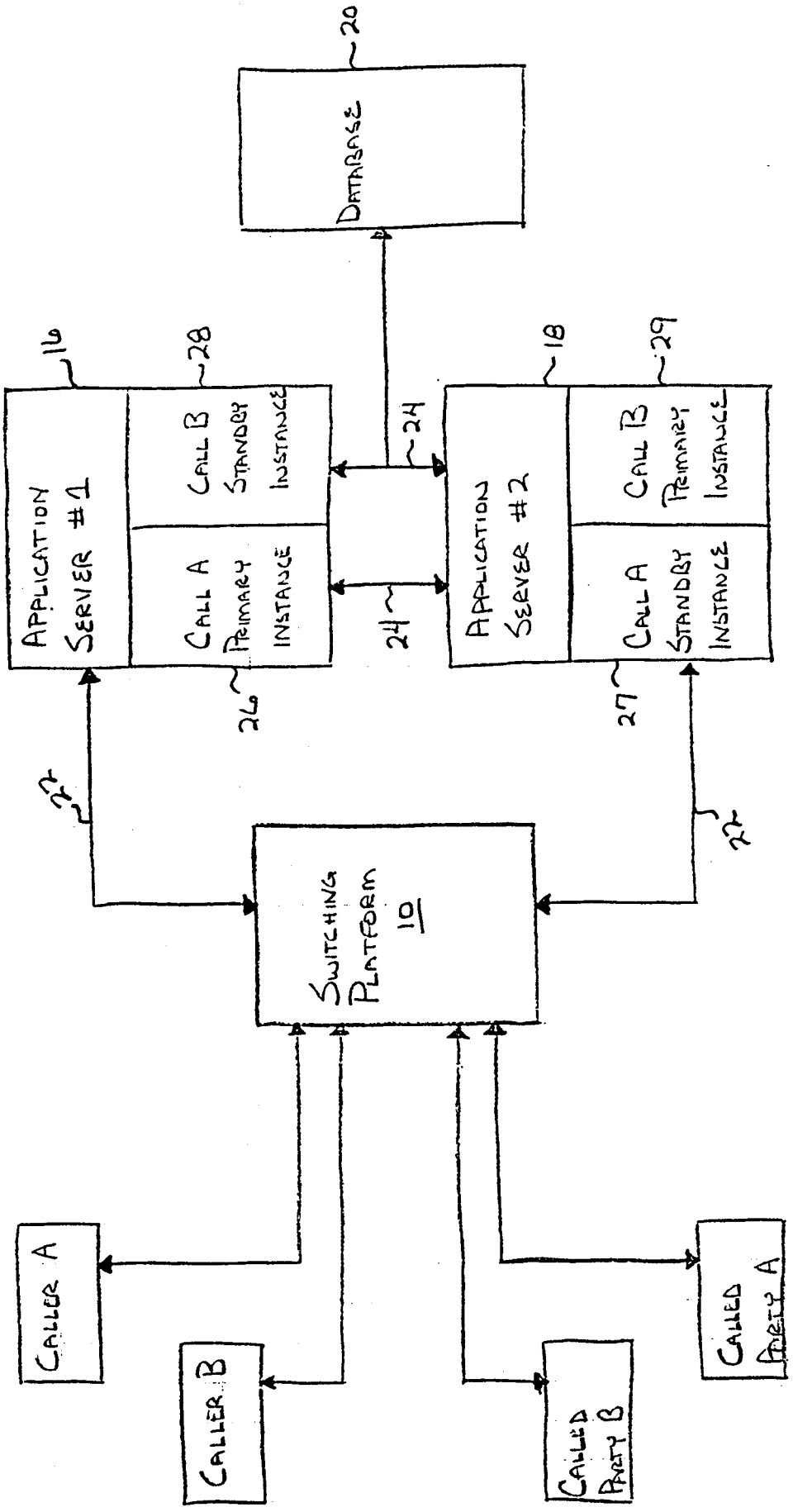

20
DATABASE
16
APPLICATION SERVER #1
28
CALL B STANDBY INSTANCE
CALL A PRIMARY INSTANCE
24
18
APPLICATION SERVER #2
29
CALL B PRIMARY INSTANCE
CALL A STANDBY INSTANCE
26
27
22
22
SWITCHING PLATFORM 10
CALLER A
CALLER B
CALLED PARTY A
CALLED PARTY B

FIG. 8

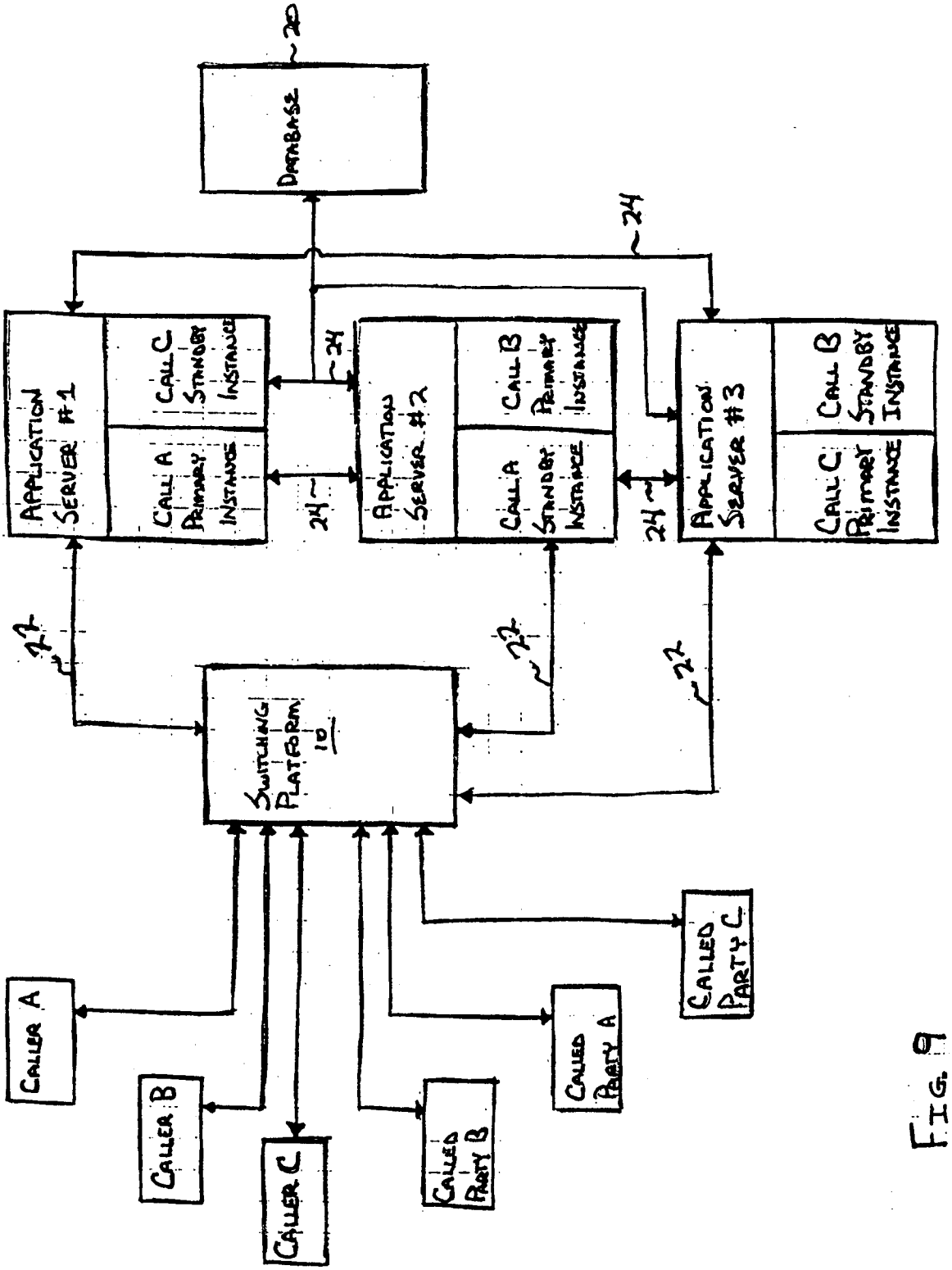

DATABASE
20
24
APPLICATION SERVER #1
CALL C STANDBY INSTANCE
CALL A PRIMARY INSTANCE
APPLICATION SERVER #2
CALL B PRIMARY INSTANCE
CALL A STANDBY INSTANCE
APPLICATION SERVER #3
CALL B STANDBY INSTANCE
CALL C PRIMARY INSTANCE
22
SWITCHING PLATFORM 10
CALLER A
CALLER B
CALLER C
CALLED PARTY B
CALLED PARTY A
CALLED PARTY C

FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/11516

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :H04M 15/00

US CL :379/112, 113, 115, 134

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4,912,698 A (BITZINGER et al.) 27 MARCH 1990 (27.03.90), see entire reference.	1-12
Y, P	US 5,664,090 A (SEKI et al.) 02 SEPTEMBER 1997 (02.09.97), col.3, line 27 - col.4, line 21.	1-12
Y	US 4,949,373 A (BAKER et al.) 14 AUGUST 1990 (14.08.90), see Figure 4, col.9, lines 10-67.	1-12
Y, P	US 5,661,719 A (TOWNSEND et al.) 26 AUGUST 1997 (26.08.97), see entire reference.	1-12
Y	US 5,369,680 A (BORBAS et al.) 29 NOVEMBER 1994 (29.11.94), col.3, lines 50-60.	1, 6, 12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
21 AUGUST 1998	13 OCT 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer CURTIS KUNTZ Diane Smith for Telephone No. (703) 305-4708

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/11516

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

379/112, 113, 115, 134, 1-2, 9-18, 32, 34, 207, 210-212, 217, 220-222, 229-230; 370/216, 217-220, 395/616, 618, 181, 182.02, 182.04, 182.08-182.09, 182.11, 182.13

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS
search terms: processor#, server#, platform#, bill?, charg?, load?, backup, back up, redundan?, fail?, storage, database, shar?, sync?