

Performance and availability measurement of the IBM Information Network

by R. M. Bailey
R. C. Soucy

A key requirement of a network service is the management of specified service levels as perceived by the end user. A capability for measuring and reporting end user response time and availability is essential. This paper describes measurement techniques to track these key service-level attributes in the IBM Information Network (IBM/IN). These techniques apply to most complex SNA networks.

The IBM Information Network (IBM/IN) has been developed to allow users to attach terminals throughout the United States to interact with a central computer complex located in Tampa, Florida. When a terminal is attached, one can use applications running on these computers. For other users, the network provides access to their own computer systems at remote locations that are also attached to the network. One way of using the IBM Information Network might be for the duration of a project, when a user's own systems are temporarily overloaded and additional capacity is required. For long-term application, a user may develop programs, access specialized data banks, or perform such application-oriented functions as financial management. Some hospitals, for example, use a centralized Patient Care System application to perform such functions as patient admission, tracking, and billing. Other users attach to the network to gain access to applications on their own centralized processors. That is, a user's locations may be dispersed geographically and still require attachment to his own central computer complex. Rather than building and managing a new network, the geo-

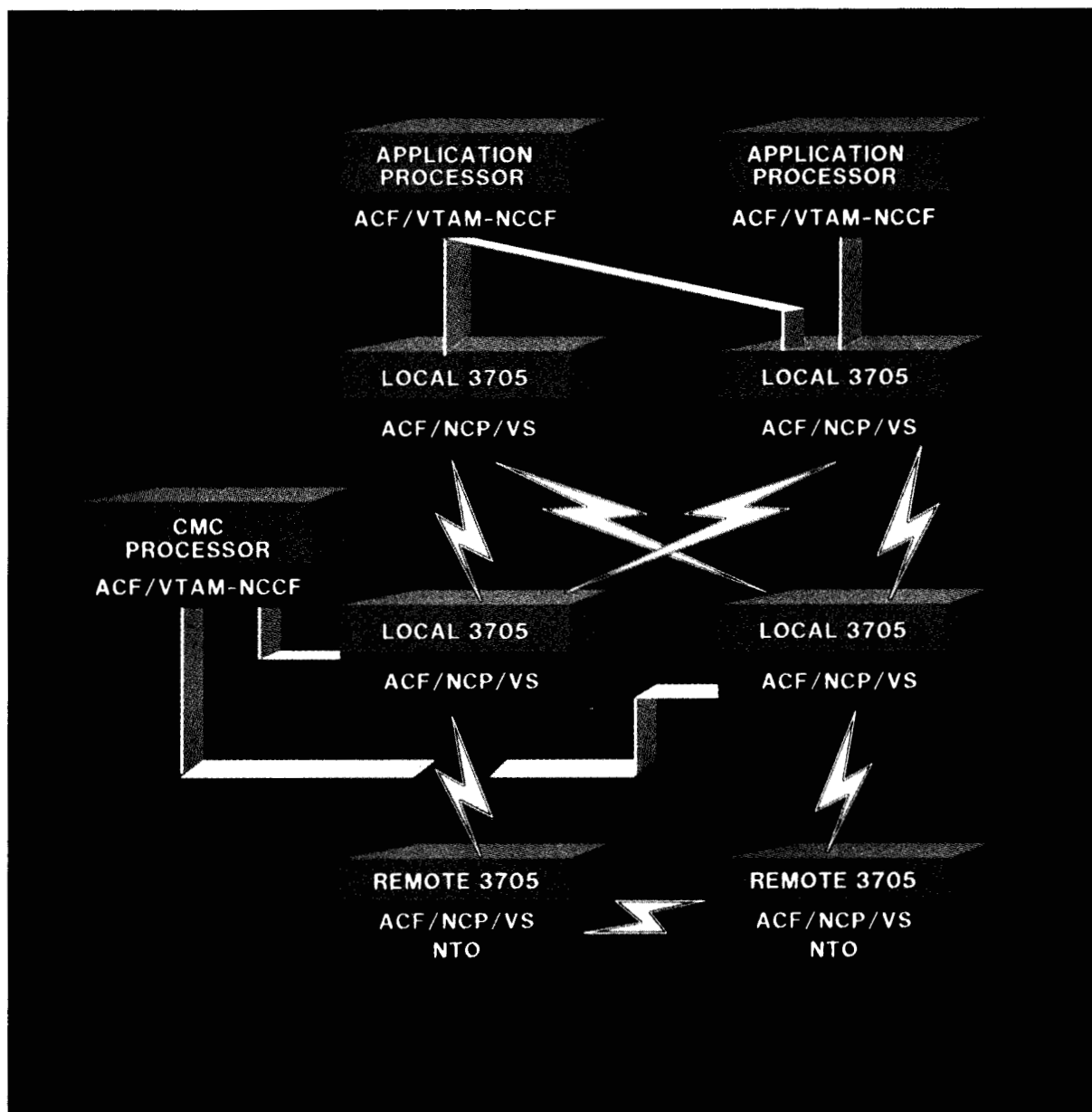
graphically dispersed terminals may access the central computers through the IBM Information Network.

As with any service, user satisfaction is a major concern of the management of the service. One key to user satisfaction can be determined by measuring the degree to which the system is meeting its operational objectives of availability and response time. *Availability* is the percentage of time that a user is able to access the system during the time it is scheduled for use. *Response time* is the elapsed time between the entry of a transaction at a terminal and the receipt of a usable response. This paper addresses techniques used for measuring these key service levels in the IBM Information Network.

Network structure. The IBM Information Network (IBM/IN) provides for the attachment of various processor and terminal types to remote communication controllers located in a number of cities. The terminals are used for accessing either applications at the IBM/IN processor complex or processors attached to the remote communication controllers. These attachments are made through such standard

©Copyright 1983 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to republish any other portion of this paper must be obtained from the Editor.

Figure 1 Sample of the IBM/IN structure



IBM products as IBM 3705 communication controllers, Network Control Program (ACF/NCP/VS Release 3), Network Terminal Option (NTO), Network Communication and Control Facility (NCCF), and ACF/VTAM Release 3. The network consists of centralized processors and multitiered local and remote 3705 communication controllers, as illustrated in Figure 1. Thus the design provides the

capability to support the main network service functions of Remote Computing Services (RCS) and Network Services (NS) in major U.S. cities.

The IBM/IN centralized computer complex currently consists of three System/370 processors located in Tampa, Florida, that have two specific functions:

1. Two System/370 processors are dedicated as application processors for remote computing services. One processor provides services for the MVS environment; the other provides services for the System/370 VM environment. Services pro-

**Performance and availability
requirements have been defined
from the point of view of
the end user.**

vided by these processors are accessed by remote terminals attached to communication controllers located in major U.S. cities. Reference 1 discusses these remote computing services.

2. The third System/370, called the Communication Management Configuration (CMC) processor, is dedicated to network management and control. The CMC establishes sessions with attached terminals and passes sessions to requested applications in application processors after identification, password, and security checks. (Details on the CMC can be found in Reference 2.)

In addition to Remote Computing Services (RCS), the network allows the attachment of processors at the remote 3705s, to which remote terminals attached to the same or other remote 3705s may have access. The CMC provides session establishment and network management facilities similar to the RCS, which is discussed in Reference 3. Because this paper is concerned specifically with performance and availability, design philosophy discussion will relate to these service levels only.

We first describe the determination of requirements and the establishment of objectives for response time and availability. Then we discuss the obtaining of outage data used for reporting availability. Finally, we show how the IBM Personal Computer is used for measuring response time.

Response time and availability

Requirements. Performance and availability requirements have been defined from the point of view of the end user. Regardless of the city in which the end user enters the network, he must experience the following service levels:

- The ninetieth percentile end user response time does not exceed a specified number of seconds for each of the following three transaction classes per application: trivial, intermediate, and complex. These classes are defined later in this paper.
- System availability is not less than a specified 9X percent during business hours, as defined by announcements. Availability includes reliability effects of components between the end user and the target application.

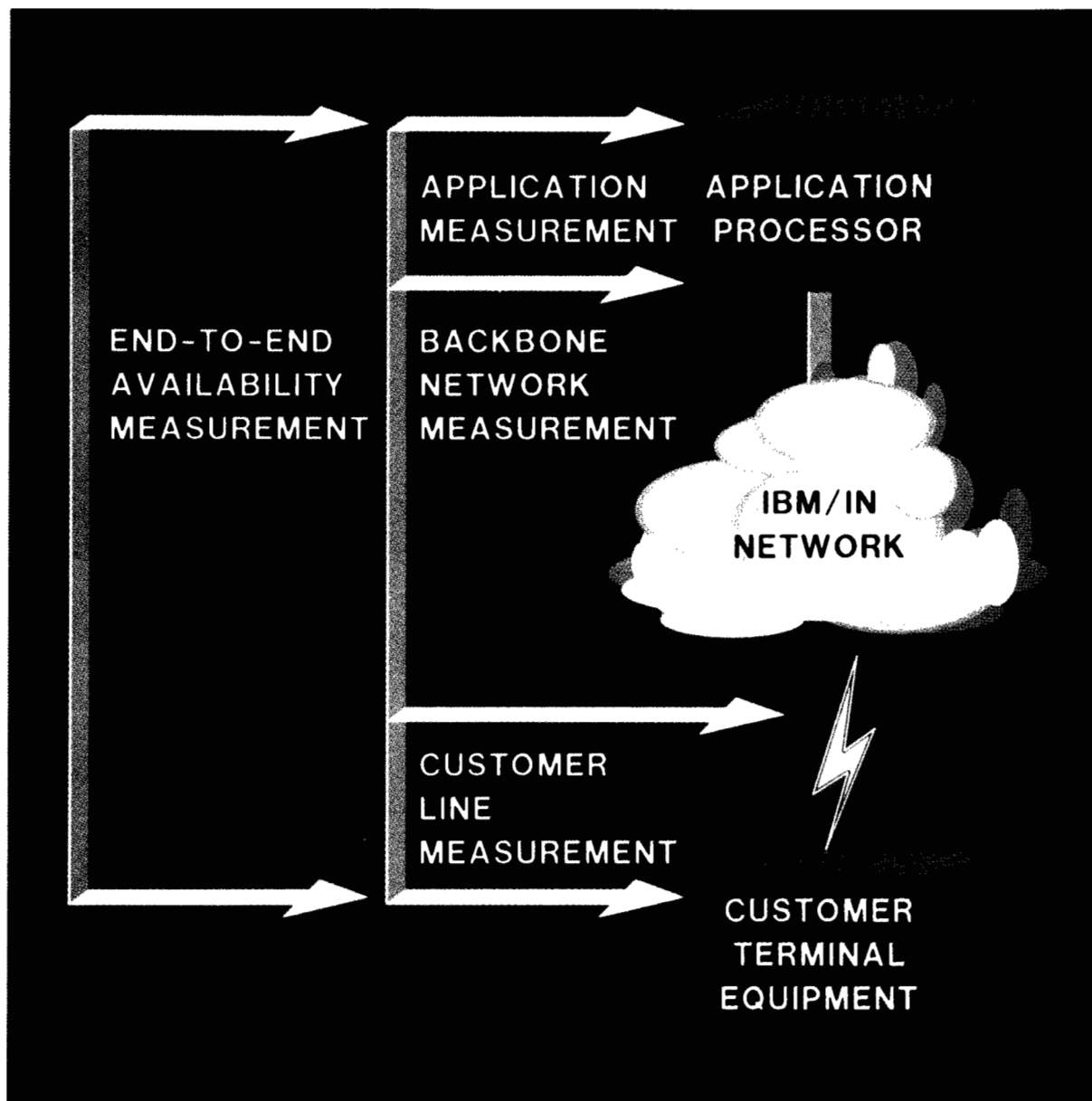
Performance objectives. Given these performance and availability requirements, the system has been divided into the following three parts, as illustrated in Figure 2:

- *Application processors* include associated input/output equipment and channel connections to locally attached IBM/IN 3705 communication controllers.
- *Network* contains the locally attached IBM/IN 3705s, the various link connections to the remotely located IBM/IN 3705s, and the remotely located IBM/IN 3705s. This network of hardware coordinated through a definition of primary and alternate routes from the application processor channels to the remotely located IBM/IN 3705 is termed the *backbone network*.
- *Local end user attachment* to a remote 3705 includes the modems and dial or leased line but excludes the user control unit and terminal.

Of these system components, the backbone network transit delay is the most critical part to specify. This delay is the most complex system parameter and has to be analyzed in sufficient detail to ensure correctness, because of the necessary lead times to upgrade the system links if required. Furthermore, the backbone network delay is an integral subset for specifying service levels for both the Remote Computing Services and Network Services.

Given forecasted message rates entering the remote 3705s, message lengths, and application processor load values, a network specification has been derived for a backbone network transit delay objec-

Figure 2 Three components of end-to-end measurement



tive. If not exceeded, this specification ensures meeting the end user requirements. To do this, an IBM program called Systems Network Analysis Program/Simulated Host Overview Technique (SNAP/SHOT) has been developed. This is a digital simulator used extensively to model the backbone network, application processors, and local end user lines. SNAP/SHOT, discussed in Reference 4, is a

family of models developed to evaluate primary routes from end user to application hosts, effects of using alternate routes, and the delay of each of the three components under various load conditions. SNAP/SHOT has proved to be a most valuable tool for projecting individual component delay and capacities so as to not exceed the end user requirements.

Availability objectives. Using reliability information relative to processors, operating systems, 3705 communication controllers, links, lines etc., a system availability specification has been written to establish an *end user availability* that satisfies the user requirement. The end user availability has been derived as the product of the following quantities (adjusted for multiple component failures): processor availability (including the application and system programs), network availability, and local line connection. Of particular interest is the derivation for the network, whose availability depends not only on the individual components but also on the various primary and alternate routes that SNA provides between the end user and the desired application. The application is considered unavailable if all routes to a destination are unavailable. In addition to the routes, another key network component is the Communication Management Configuration (CMC) processor. By definition, the CMC owns all resources in the network. It is a processor dedicated to the management of the network and receives and stores error status relative to these resources. The CMC also receives all end user logon requests to an application, and it controls activation and deactivation of all resources. Because of the importance of the CMC to end user availability, redundant processors, communications controllers, and operational backup procedures have been implemented to meet the end user objective.

Availability reporting. Measurement and reporting of availability in a complex system can be a rather difficult and expensive undertaking because there is a general lack of tools for this purpose. Thus a significant effort is required to obtain comprehensive and accurate measurements. The following is one facet of the measurement and reporting process within the IBM Information Network.

A variety of reports ranging from the availability of individual components to the effect of outages on the user population have been developed. The capability of an end user to access an application is one of the primary service levels that must be monitored. A user's satisfaction is highly dependent on his ability to do productive work when required. To track these service levels, it is necessary to determine the availability of all system components from the user's terminal equipment to applications. This availability is measured in three separate parts, as shown in Figure 2. Outages for each part are determined and used to compute an overall end-to-end availability.

Outage data

The development of outage data for the determination of user line and application availability is straightforward. The determination of the backbone network availability is more complex. In a system configured of multiple processors that can be accessed from remote locations through several routes, an outage of a backbone network path occurs only when all redundant routes in the path fail at the same time. For the purpose of this paper, the term *path* means all valid routes from a processor to a remote node. When it is determined that one route has failed, the status of the redundant routes

The system to supply and collect the availability data is called the Network Programmed Operator.

must also be known to determine whether a path has become unavailable. This is a task that does not lend itself readily to manual procedures. Manual procedures are not usually accurate, and they require the time of an operations staff. Therefore, it is desirable to fully automate the functions of gathering and reducing the raw availability data and reporting on availability.

Availability data collection system. One problem we had to solve devolved from the fact that no standard outage data as such are provided by the system at either the component or path level. Standard ACF/VTAM error messages tell when components have failed and when they are active again. But where it is necessary to understand the status of multiple routes, the reduction of the raw data is cumbersome. Another problem is the accumulation of the required data. The data used in determining backbone network availability are stored in the NCCF log. These data are generated at each processor in the complex. Generally, each processor has its own NCCF log, which would require these logs to be gathered during the operating day and at end-of-day processing. During the day, when a log fills, a switch is made to an alternate log. The full log

must be stored before another switch occurs. If NCCF is to be restarted, the old log must be saved or NCCF will destroy it during startup. In addition, procedures to handle the archiving of these pieces of daily activity would have to be expanded to handle more logs, as additional processors are added to the system. For these reasons, an *availability data collection system* has been developed that not only generates the required data but also accumulates them automatically at a central location at the time they are created.

Inasmuch as the Communication Management Configuration (CMC) processor owns all network resources, it is thus the central control point for the network. This fact makes the CMC the logical place for a common logging function. The creation of a common logging function requires communications between the processors and the CMC to allow data of interest to be gathered at the common point. The fact that the CMC function might have to be moved to another processor in the event of failure or planned outage had to be considered.

Network Programmed Operator. The system we developed to supply and collect the availability data is called the Network Programmed Operator (NPO). The NPO runs as an application under the Network Communications Control Facility (NCCF). NCCF facilities used for NPO include the User Exit, User Subtask, and CLIST interfaces, described in Reference 5.

NCCF and NPO reside in all of the processors. However, the CMC-NPO function is in the CMC processor, which is the control point for the collection system. The other processors contain subordinate NPO functions, which are all basically the same. Thus, as additional processors are added to the complex, NCCF and NPO can be added to the processor without change to the collection system. The CMC-NPO establishes sessions through NCCF with the other NPO tasks. These sessions are used to forward commands from the CMC for execution at the other processors and to return data to the CMC for logging. Some of the functions executed at each processor are the following:

- The processing of commands sent by the CMC for execution at that processor.
- The examination of acquired unsolicited error messages and solicited responses.
- The time stamping and transmission of selected messages to the CMC-NPO.

- The establishment and maintenance of NCCF timers for the execution of commands at timer expiration.

The collection of network status information requires that the NPO tasks in the application processors are always in session with the CMC-NPO

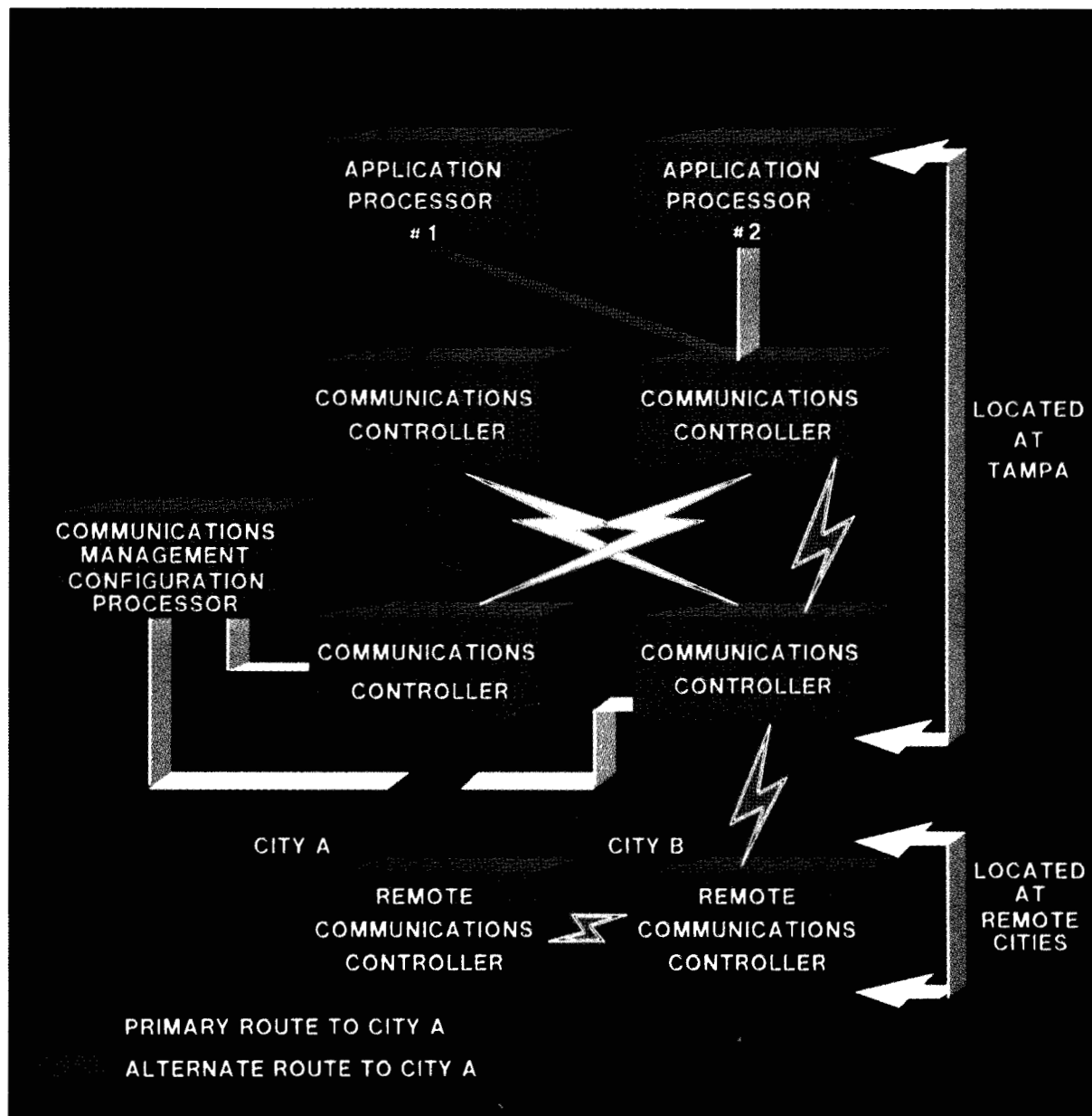
The network design allows for sessions to be re-established on one of the alternate routes, upon failure of a primary route.

task. These sessions are automatically established by the CMC-NPO when the network is activated. However, these sessions are subject to loss due to failing network components or processors. To ensure against the loss of status data, the CMC-NPO task checks the status of these sessions and periodically attempts to re-establish any that have been lost.

The availability of a path from an application processor to a remote node of the network is the most difficult to determine. The difficulty lies in path definitions of up to six alternate routes from each application processor to each remote node. The network design allows for sessions to be re-established on one of the alternate routes, upon failure of a primary route. Because several routes can be activated to a remote node, the path is considered available if any one of the defined routes and the node itself are operative. Figure 3 shows that cities A and B can access both the CMC and the application processor through several routes.

It is possible to determine network availability from standard message data produced by ACF/VTAM. However, this requires that all status and error data be mapped against a data base that defines the resources and the connectivity of the resources of the network. Aside from the complexity of reducing the data in this manner, there is also the problem of maintaining the currency of the data base as the network configuration changes.

Figure 3 Example of primary and alternate routes between Application Processor 1 and remote City A



Operability snapshot. A more desirable method is that of obtaining a snapshot of the operability of all the routes to a remote node when an event occurs that changes the status of any component in the routes to the node. To obtain these status data, we used the TEST option of the ACF/VTAM DISPLAY NET, which is referred to as *route test*. When the TEST option is used in conjunction with the *class of*

service name (COSNAME) option, the command tests the operability of all routes from a processor to a designated remote node. The COSNAME option allows the designation of a reference to an entry in a Class Of Service Table. These entries include the transmission priority for this class of service and all of the defined virtual routes from a processor to an IBM 3705 or 3725 communications controller. Since a

route test adds network traffic, it is desirable to limit its use only to those times when a network status change occurs. The determination of when to issue the commands is based on the use of a subset of the standard network messages known as *triggers*.

Trigger messages fall into two general categories, *solicited messages* and *unsolicited messages*. Solicited messages are usually the result of an operator command to activate or deactivate a resource. Examples of solicited message responses are the following:

```
IST105I nodename NOW INACTIVE
```

```
IST093I nodename NOW ACTIVE
```

Here, the term nodename is a user-defined name given to a network resource. Unsolicited messages used as triggers are those that indicate either the failure of a resource on a network route or the recovery of a resource. Examples of these messages are the following:

```
IST259I INOP RECEIVED FOR nodename CODE  
= 02
```

```
IST129I UNRECOVERABLE OR FORCED ERROR  
ON NODE nodename
```

```
IST619I ID = nodename FAILED - RECOVERY  
IN PROGRESS
```

Because the CMC owns all network resources, trigger messages for all resources in the network arrive normally at the CMC.

The CMC-NPO task intercepts each message that is returned and checks whether it is in the list of trigger messages. When a trigger is found, the CMC-NPO converts the nodename in the message to the subarea number of the remote node. The nodename may be other than a communication controller, but our naming conventions allow the names of other resources to be related to a specific communication controller. The subarea number, which is part of the SNA resource address, is used in the destination subarea (DESTSUB) parameter of the DISPLAY command. If the trigger message contains a subarea number, it is used in the route test directly. The following form of the DISPLAY command is used for the route test:

```
DISPLAY NET,ROUTE,DESTSUB = nn,TEST  
= YES,COSNAME = name
```

The COSNAME option of the DISPLAY command directs the command to obtain the routes to be

tested from the named entry in the Class Of Service Table. The Class Of Service Table lists all valid routes from a processor to a remote subarea in priority order. The use of COSNAME results in a test of all of these routes, allowing the creation of the route test by NPO to be independent of changes in the route structure of the network. Once created, the route test is issued to test all routes from the CMC to the remote node. Because there are usually common resources in the paths from the CMC to a remote node and the paths from other processors to that node, the paths from all processors must also be tested. To accomplish this, the CMC-NPO then transmits the same command to all other IBM/IN processors in the network on its NCCF-to-NCCF sessions to test their routes to the same remote node.

Responses to route tests originating at the CMC arrive there directly and are recorded through NPO in the NCCF log. Responses to the route tests sent to the other processors for execution are returned to those processors. These responses are then retransmitted by NPO to the CMC-NPO to be logged. Responses to the test show the route tested and the results of the test, as in the following example:

```
IST533I ER 3 FAILED IN ROUTE TEST 125 FROM  
SA = 3 TO SA = 45
```

```
IST533I ER 1 SUCCEEDED IN ROUTE TEST 3  
FROM SA = 1 TO SA = 36
```

These messages show the route number tested, the number of the route test, the origin and destination subareas, and the status of the route. The route test number appears on each response of one command and ties together the responses from that command. It is these responses that provide a complete picture of all the paths to a remote location at the instant of network status change event. These data are collected for a twenty-four-hour period and are reduced to give an accurate picture of the availability of the backbone network. The status of network paths is maintained across collection periods so that data from a previous period coupled with new data show a continuous picture of network availability.

Communication Management Configuration (CMC) processor availability. Another critical part of the determination of backbone network availability is the CMC processor itself. Users are initially connected to an application in the CMC called the *Service Manager*. The user selects the application to be used, and the Service Manager establishes a session for the user with that application. If the CMC processor fails, new users are unable to gain access

Figure 4 Example printout of a daily report of backbone network path availability for the previous day

REPORT: 422-11		NETWORK AVAILABILITY REPORTING NETWORK AVAILABILITY TO BOUNDARY NODE (DAILY)		10/27/82	
DATE: 10/26/82 ORIGIN: CMC PROCESSOR TAMPA		JULIAN: 299 SUBAREA: 001		24 HOURS	
DESTINATION CITY	SUB AREA	NCP NAME	PERCENT AVAILABLE	NUMBER OF OUTAGES	AVERAGE LENGTH OF OUTAGE (HRS)
ATLANTA	045	IBMNGRT	100.00	0	.00
CHICAGO-1	043	IBMTTAT	98.68	1	.46
DALLAS	041	IBMDSAT	100.00	0	.00
HOUSTON	037	IBME7AT	100.00	0	.00
LOS ANGELES	039	IBMWGRT	100.00	0	.00
NEW YORK-1	048	IBMF3AT	99.43	1	.40
PHILADELPHIA	047	IBMMBAT	100.00	0	.00
SAN FRANCISCO	040	IBMNORT	99.99	2	.00
TAMPA-1	049	IBMWHRT	100.00	0	.00
TAMPA-2	038	IBMFLAT	100.00	0	.00
WASH D.C. -1	046	IBMLCRT	100.00	0	.00

Figure 5 Example report printout of the average backbone network availability for the previous thirty days

REPORT: 422-35		NETWORK AVAILABILITY TO BOUNDARY NODE 30 DAYS AVERAGE FOR THE PERIOD ENDING 10/27/82			
ORIGIN: CMC TAMPA		SUBAREA: 001		24 HOURS	
				10/27/82	
DESTINATION CITY	PERCENT AVAILABLE	OF OUTAGES	LENGTH OF OUTAGE (HRS)	TOTAL LENGTH OF OUTAGES	AVERAGE TIME BETWEEN FAILURES
ATLANTA	99.59	4	.56	2.24	129.79
CHICAGO-1	99.81	4	.30	1.20	130.20
DALLAS	99.63	3	.62	1.86	153.13
HOUSTON	98.91	5	1.11	5.55	89.67
LOS ANGELES	99.12	3	.68	2.04	141.44
NEW YORK-1	99.96	1	.25	.25	336.00
PHILADELPHIA	99.32	2	.71	1.42	224.97
SAN FRANCISCO	99.84	1	.44	.44	336.00
TAMPA-1	99.76	1	.52	.52	334.45
TAMPA-2	99.93	1	.46	.46	336.00
WASH D.C. -1	99.87	1	.41	.41	336.00
AVERAGE ALL CITIES	99.59	23	.55	1.19	231.69

to the system. For this reason, an outage of the CMC processor is applied as an outage to all remote nodes even though some existing sessions are preserved. A CMC outage is detected by the reduction programs by the existence of a network startup NCCF *command list* (CLIST) in the log. The duration of an outage is determined by the time stamp on the last entry in the log prior to the CLIST and the time

stamp on the message that indicates the completion of network startup procedure.

Formal reports are prepared each day detailing backbone network availability for the previous day. Another report shows the average backbone network availability for the previous thirty days, and a thirty-day rolling average report provides availabil-

ity trends. These reports, as illustrated in Figures 4 and 5, provide a picture of how the backbone network is being managed to meet service-level objectives.

Application and user-line availability. The backbone network availability is one of the pieces of the end-to-end availability. To complete the picture, end-to-end availability must be combined with the availability of both the application and the user line from a remote node to a user's terminal equipment.

Application and application processor outages are determined through messages that indicate the loss or activation of a Cross Domain Resource Manager (CDRM) session between ACF/VTAM in the CMC and ACF/VTAM in the application processor. In addition, periodic samples of the status of an application are taken using the ACF/VTAM DISPLAY NET, . . . ID = (application name) command. These responses are also stored in the central NCCF log for later reduction.

Outages for user lines are derived from messages pertaining to the line, indicating that the line has failed or has been taken out of service, and from

messages indicating that the line has been returned to service. The message data in the log are accumulated for each line to develop an overall outage picture for the line.

The line, application, and backbone network outage data provide the basis for a report showing the user's availability through the network to all applications. This report, as illustrated in Figure 6, shows the availability of each part of the path—the line, the network, the application, and the combination of all three. The combination of the availability percentages is computed, based on the unoverlapped outages of the components. The same data used for the previously mentioned reports are also used to develop additional exception reports on the availability of such individual resources as lines, communication controllers, and virtual routes.

Response time measurement by Personal Computer

Response time is another service level that must be tracked continuously. Response time directly affects the productivity of a user and is one of the main criteria by which an interactive communica-

Figure 6 Example printout of users' availability of each part of the path through the network to the application

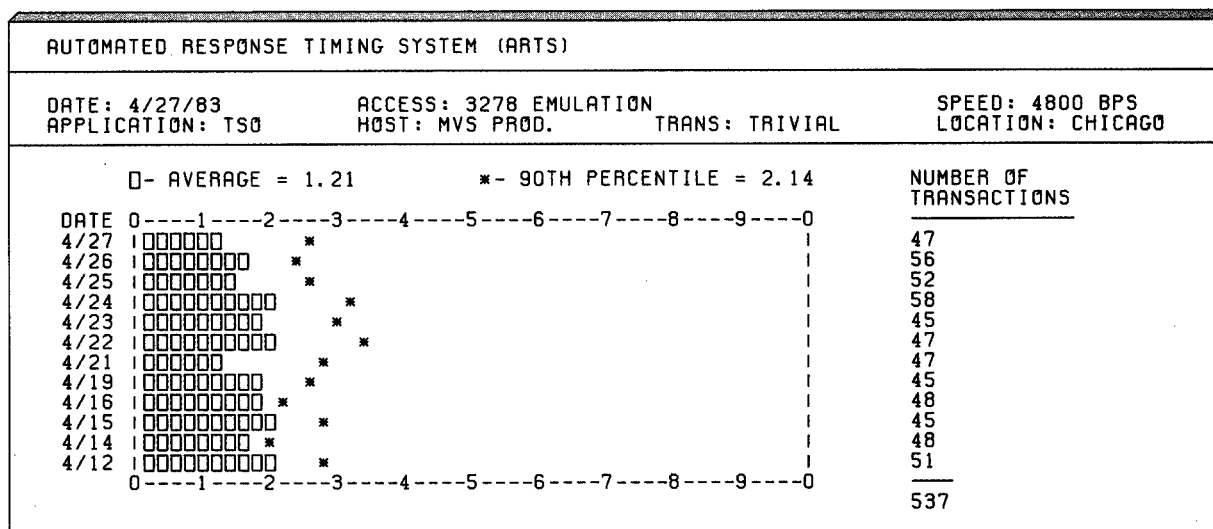
REPORT: 450-21		END TO END AVAILABILITY PRODUCTION SYSTEMS (DAILY)						9/21/82	
DATE: 9/20/82		JULIAN 263						24 HOURS	
CITY	PROD. HOST NAME	APPL. NAME	LINE NAME	LINE AVAIL. %	NETWRK. AVAIL. %	APPL. AVAIL. %	END TO END %	NBR. OF OUT.	CUSTOMER
ATLANTA	CMC	NETM2	N6861	99.46	100.00	99.83	99.29	2	#3641
	MVS-A	APPL-A	N6861	99.46	100.00	100.00	99.46	1	"
	MVS-A	APPL-B	N6861	99.46	100.00	100.00	99.46	1	"
	VM-A	APPL-C	N6861	99.46	100.00	98.86	98.32	2	"
CHICAGO-1	CMC	NETM2	TT0A1	99.91	100.00	99.83	99.84	2	#2921
	CMC	NETM2	TT121	98.06	100.00	99.83	97.89	2	#2866
	MVS-A	APPL-A	TT0A1	99.91	100.00	100.00	99.91	1	#2921
	MVS-A	APPL-A	TT121	98.06	100.00	100.00	98.06	1	#2866
	MVS-A	APPL-B	TT0A1	99.91	100.00	100.00	99.91	1	#2921
	MVS-A	APPL-B	TT051	98.06	100.00	100.00	96.46	1	#2866
	VM-A	APPL-C	TT051	99.91	100.00	100.00	99.91	1	#2921
	VM-A	APPL-C	TT051	98.06	100.00	100.00	98.06	1	#2866
	CMC	NETM2	DS1A1	98.88	100.00	99.83	98.71	2	#6123
	MVS-A	APPL-A	DS1A1	98.88	100.00	100.00	98.88	1	"
DALLAS	MVS-B	APPL-E	DS1A1	98.88	100.00	100.00	98.88	1	"
	VM-B	APPL-D	DS1A1	98.88	100.00	99.26	98.14	2	"

AUTOMATED RESPONSE TIMING SYSTEM (ARTS)					
DATE: 4/15/83		ACCESS: LINE MODE		SPEED: 1200 BPS DIAL	
APPLICATION: WBO		HOST: VM PRODUCTION		TRANS: COMPLEX LOC: DALLAS	
□- AVERAGE = 4.33 * - 90TH PERCENTILE = 7.95				NUMBER OF TRANSACTIONS	
HOUR FROM	0-----1-----2-----3-----4-----5-----6-----7-----8-----9-----0				
8:00	██████████*				12
9:00	██████████*				23
10:00	██████████*				52
11:00	██████████*				60
NOON	██████████*				45
1:00	██████████*				47
2:00	██████████*				47
3:00	██████████*				45
4:00	██████████*				48
5:00	██████████*				45
6:00	██████████*				30
7:00	██████████*				35
	0-----1-----2-----3-----4-----5-----6-----7-----8-----9-----0				
					489

We have found that delivered response time (to an end user) is difficult to measure automatically without having logical capability in the terminal device, not only to record transaction times but also to record the types of transactions from which they result. Devices and systems are available that attach to user lines or terminals to provide some level of

A representative set of transactions has been defined for each of the three transaction classes for each application, as specified in the objectives. These transactions are executed periodically from various remote locations and the response times for each transaction are recorded by the PC. At the end of each session, the PC forwards these measurements to a central location for storage and for report creation. Since our objective for measurement has

Figure 8 Example on-line report showing response time trends



been to collect data automatically, the PC response-time measuring system must operate unattended.

Each PC at a remote location operates independently, as though it were an end user of services competing for system resources. The PC establishes access to applications in three modes:

- Dial access as a line mode device via the Network Terminal Option (NTO).
- Dial access into a Series I that has a special program for simulating an IBM 3101 as a full-screen device.
- Simulation of an IBM 3278 via an IBM 3274.

For dial access, the PC is attached to an auto-dial modem.

The basic operation cycle starts with a timer interrupt at the PC. The network is then accessed through the auto-dial modem or through the 3274. The PC requests access to each application from the CMC in each application host and executes the selected transactions for the application. The transaction is time-stamped when sent to the host and the response is time-stamped when returned to the PC. Finally, the response times are stored at the PC until all requested active applications have been accessed. At the completion of the cycle, the PC accesses a control application on the VM development processor at Tampa and sends it the accumulated response times. The PC then waits for the

next timer interrupt to start the cycle again. Accumulated data can be viewed on line at any time. Exception reports are produced for areas that fail specifications. Because of the unattended mode of operation, maintenance is also a major consideration in the design of the PC response time measurement program. Any nonprogram data in the PC that are subject to change are maintained in tables. When changes are required, the host program passes the new tables to the PC. Figures 7 and 8 are examples of two of the on-line response time reports produced by the system.

The PC response-time measurement system requires no support from any of the host operating systems or network programs. A performance data base program receives the response-time data from the PCs. Report programs create response-time displays from current data periodically throughout the day.

This response-time measurement system provides a level of data precision that was previously unavailable, thereby allowing direct comparison of actual response times with preset response-time objectives. The use of standard transaction scenarios over time makes possible accurate and useful trend reporting.

Concluding remarks

When telecommunication capability was first added to traditional batch operations, there was

little planning for system measurement. As interactive functions grew to become a major part of the processing, measurements were gradually added, usually as the parameters of response time and availability of the central computer. As networks have grown in size, complexity, and geographic distribution, the interaction of the network and system components has become more difficult to understand and manage. Operations personnel require comprehensive information for tuning, capacity planning, and performance management. Service providers must respond to an increasing demand for committed service levels. It is no longer sufficient to view service in terms of the response time of the central computer. Information must be available concerning how the entire system is performing, as viewed by the end user. This requires that management be provided with daily, period, and trend information on the attainment of important service levels. To provide an adequate reporting capability, measurement requirements should be considered as an integral part of the system design. The measurements discussed in this paper are but a part of the information provided to IBM Information Network management and departments concerned with performance and capacity planning. This variety of measurements, together with user service calls and periodic user surveys, provide a broad picture of the adequacy of service delivery.

Cited references

1. *IBM/IN Network Services*, G534-2228, IBM Corporation; available through IBM branch offices.
2. G. Ackridge, "CMC usage at IBM Information Network," *Proceedings of Guide International, Guide/56, Chicago, IL, May 15-20, 1983*, Form No. DB-25, 257-278 (1983).
3. *IBM/IN Network Services*, G534-2208, IBM Corporation; available through IBM branch offices.
4. *SNAP/SHOT—An Executive Perspective of Capacity Planning*, G520-0083, and *Performance Prediction and Capacity Planning Using SNAP/SHOT*, G520-0084, IBM Corporation; available through IBM branch offices.
5. *NCCF Installation*, SC27-0430, and *NCCF Customization*, SC27-0433, IBM Corporation; available through IBM branch offices. See also H. M. Stewart, "Performance analysis of complex communications systems," *IBM Systems Journal* **18**, No. 3, 356-373 (1979).

Reprint Order No. G321-5202.

Richard M. Bailey *IBM Information Network, P.O. Box 30021, Tampa, Florida 33630.* Mr. Bailey began working in the computing industry in 1957, on the development of compilers for various programming languages. He joined IBM in 1963 and began in a program development project, working specifically on compiler techniques. He spent several years in various manage-

ment and technical positions related to OS and TSS. Mr. Bailey was later associated with the field of organization for Public Sector accounts, working with public safety customers in the development of Regional Criminal Justice Information Systems. He later joined the Public Sector Industry staff in Bethesda, Maryland. Mr. Bailey joined the IBM Information Network in 1980 in the network development department, where one of his responsibilities was that of planning for and design of measurements for the network. Mr. Bailey received a B.S. degree in mathematics from the University of Maine, Orono, in 1957.

Richard C. Soucy *IBM Information Network, P.O. Box 30021, Tampa, Florida 33630.* Mr. Soucy joined IBM in 1962 at Poughkeepsie, New York. During the following ten years, he helped develop and manage the development of such programming aids as GPSS, OSPT1, CNDP, and AMAP to support systems engineering. In 1972 he transferred to the Programming Center at Kingston, New York, where he held several management positions involved in the development, performance, and planning of the SNA products VTAM and NCCF. He is currently the manager of the network performance and availability department for the IBM Information Network in Tampa, Florida. Mr. Soucy received his B.S. degree in electrical engineering from the Worcester Polytechnic Institute, Worcester, Massachusetts, in 1959 and his M.S. degree in electrical engineering from the University of Massachusetts, Amherst, in 1962.