SWITCHING SYSTEMS MANAGEMENT
NO. 2 ELECTRONIC SWITCHING SYSTEM
ABNORMAL TRAFFIC LOAD CONDITIONS

CONTENTS PAGE

Figures

1. GENERAL

1.01 This section describes the response of the No. 2 Electronic Switching System (ESS) to abnormal traffic load conditions. It provides the network administrator with descriptive information for use in recognizing and dealing with an abnormal traffic load.

1.02 When this section is reissued this paragraph will contain the reason for reissue.

1.03 The title of each figure includes a number(s) in parentheses which identifies the paragraph(s) in which the figure is referenced.

2. DESCRIPTION

2.01 An overload condition can exist in a central office when the usage of the equipment exceeds the engineered capacity or when the call load is so heavy that service deteriorates in the form of excessive dial tone delays or matching loss.

2.02 Many types of overloads can occur within the system. They may be broadly classed as:

(a) Real-time overloads

(b) Software overloads

(c) Hardware overloads.

REAL-TIME OVERLOADS

2.03 The call processing capacity of the No. 2 ESS is dependent upon the small but finite amount of real time that is required by the central processor to perform the tasks required in processing a call. The originating plus incoming call rate per main station and the call mix (ie, percentage intraoffice, percentage automatic message accounting, etc) have a significant influence on the call processing capacity.

2.04 Unless manual intervention or the detection of trouble occurs, both central processors in a No. 2 ESS operate simultaneously, executing the same instructions in synchronism. The outputs of the off-line processor to the peripheral equipment and maintenance administration center are inhibited.

2.05 Real-time overloads are the result of the failure of the on-line central processor to cycle through all classes of base level work (those functions executed during the base level scan; that is, call processing, maintenance, and routine work operations) within the required time interval. The central processor becomes overloaded when the number of requests for service results in abnormally long base level scans. The base level scan is normally initiated every 100 milliseconds (or more precisely 100.08 milliseconds). As mentioned above, call processing, maintenance, and routine work operations are performed during the base level scan. Base level work is classified as either periodic deferrable or nondeferrable. During the period allocated for deferrable work, referred to as ***periodic deferrable time*** (PDT), audits, detection tests, and call store updating are performed. The nondeferrable operations consist of those functions that the program is obligated to perform during the scan, regardless of the call volume. They may be categorized as fixed overhead tasks (scanning for supervisory inputs, accessing call records, and making usage measurements), variable overhead tasks (system monitors), and the tasks required to process calls offered to the system (detection of originations, setting up path connections, detection of call terminations, etc). The base level scan will exceed 100 milliseconds whenever there is more work in the nondeferrable area than can be completed in this time.

2.06 A principal task performed at the base level scan is the distribution of supervision to appropriate transient call records (TCRs). A TCR is associated with each active nontalking call in the system and consists of eight call store words. During the base level scan each TCR block, consisting of 70 TCRs, is scanned for any new information associated with each call. A heavy traffic load will increase the time required to survey the TCR blocks and, consequently, lengthens the base level scan time. At the start of each TCR block scan, a hardware timer is set in the central processor. If a TCR block scan is not completed within 320 milliseconds, a central processor switch will occur. Excessively long base level scans, then, result in an equipment time-out which causes the on-line central processor to switch to the off-line central processor. Consult Dial Facilities Management Practices, Division H, Section 10d(4), Service Overloads and Interrupts, for a detailed explanation of central processor overload.

SOFTWARE OVERLOADS

2.07 Software overloads result chiefly from the inability of the program to adequately process information accumulated in various hoppers, registers, peripheral order buffers, etc.

2.08 An overload control program, entered once each base level scan, defers work operations that require considerable (greater than 10 milliseconds) processing time. ***A deferral may be defined as the delay of certain call processing tasks during a time interval. It should not be confused with the PDT work operation.*** Those call processing tasks which also may be deferred include:

(a) Connection of dial tone

(b) Connection of ringing and audible ringing tone

(c) Connection of a transmitter to an outgoing trunk.

2.09 The primary call processing control in No. 2 ESS rests with the set of programs associated with the TCRs. The TCRs control the progress of a call for the interval starting with dialing, through ringing, and until an answer signal has been received and the connection is in the talking state. A software timer, acting in union with the overload control program, times each TCR block scan. If the time exceeds 175 milliseconds all TCRs remaining in the monitored block which indicate that they contain deferrable work (discussed in 2.08) are by-passed during the base level scan. Separate program actions ensure that individual calls will not experience permanent deferral. The action of the overload control program avoids a central processor switch such as might be initiated by a timed-out hardware timer.

2.10 Another deferral mechanism is initiated when the processor approaches its capacity load. To account for those periods when the processor is offered more calls than it can complete, the No. 2 ESS has been designed to accept a maximum of three new calls per base level scan.

QUEUE OVERLOADS

2.11 When call supervision is required, as indicated by a supervisory signal (on-hook, off-hook, or flash), the program attempts to locate an idle TCR. When there are none available, an effort is made to store the signal in a queue (an 8-entry list) located in call store. Entries are then taken individually from the queue for assignment to a TCR. If a TCR is available, the queue entry is cleared; if there is not one available, the entry is restored.

2.12 Under heavy loads the queue is filled. When the queue overflows, a timed-out entry is made in either the off-hook hopper or the on-hook timing list (also in call store). The location in which this entry is stored is determined by the type of supervisory change. Off-hooks and flashes are placed in the off-hook hopper which contained the original off-hook transition. On-hooks are placed in the on-hook timing list which contained the original on-hook transition. The program then examines the off-hook hopper and on-hook timing list for timed-out entries. If any such entries are found, an attempt is made to assign supervision to a TCR or, failing this, to the queue. If the queue is filled, a second identical timed-out entry is rewritten over the original.

HARDWARE OVERLOADS

2.13 Hardware overloads can occur when all network paths are blocked (ie, when the total demand for available trunk and service circuits is exceeded). Any piece of equipment that is under-provided for a given offered load tends to show up as ineffective calls. When there are not enough customer digit receivers (CDRs) dial tone delays occur which cause a higher incidence of partial dials. When there are not enough outgoing trunks the effect is similar to that for inadequate CDRs. Once digits have been received, failure to seize a trunk results in an incompleted call. A network that is experiencing matching loss as a result of poor line balance or improper assignment practices will also produce an overload situation that may be considered to be hardware related.

OVERLOAD INDICATORS

A. Data

2.14 The No. 2 ESS is programmed to generate traffic data useful in determining the performance of an office. The traffic data recorded include peg counts, overflows, and usage. These are defined as follows.

- ***Peg count*** is a cumulative count of the number of times a given event occurs during a fixed time interval.
- ***Overflow*** is a cumulative count of the number of times events failed because of network blocking or lack of facilities.
- ***Usage*** is a cumulative count of the number of circuits or registers that is in a busy state during each periodic scan of a particular group of circuits or registers.

2.15 Traffic measurements which indicate overload conditions are printed on several schedules. The Q-schedule, which automatically prints out at the clock quarter-hour during a system overload (deferrals occuring), shows dial tone delays and incoming matching loss. The H- and C-schedules show dial tone delays, incoming matching loss, and overflow counts. (Consult Bell System Practices, Section 232-120-301, Traffic and Plant Measurements; Translation Guide TG-2H, Division 10, Traffic Measurements; and Dial Facilities Management Practices, Division H, Section 10i, No. 2 ESS Traffic Measurements, for items recorded on H- and C-schedules. Q-schedule items are shown in Figure 1.

2.16 Analysis based on the quarter-hour is perferred since it more closely reflects the effects of traffic peaks over the hour.

B. Lamps

2.17 The overload control program causes the OVERLOAD lamp on the maintenance center frame control and display panel to light when the number of call processing deferrals in a base level scan is nonzero. The lamp is extinguished when the count returns to zero.

LOAD SERVICE CURVES

2.18 When an office begins to gather historical data, it is advantageous for the network administrator to develop load service curves. Load service curves are graphs of an item versus the service which that item directly affects. For example, CDRs ***versus*** dial tone speed test failures or junctor CCS ***versus*** matching loss might be considered. To derive a valid load service curve, sufficient service data are required. This pictorial representation of how an office operates under heavy traffic pressure is an invaluable aid. By using graphed data, the actual performance of the call processing functions can be observed.

2.19 Of all the Q-schedule measurements, Q14 is perhaps the most significant register from the network administration standpoint. Q14 represents the time spent doing the periodic ***deferrable*** tasks; this time was referred to in 2.05 as PDT. Specifically, it represents the real time remaining in the base level scan once fixed overhead tasks (scanning for supervisory inputs, accessing call records, making usage measurements), variable overhead tasks (system monitors), and the tasks required to process calls offered to the system (detection of originations, setting up path connections, detection of call terminations, etc) have been completed. It should be evident that PDT decreases as the result of an increase in call volume. The reading is valid only if the system remained in synchronism during the measurement period and the audits (programs that locate and attempt to correct inconsistencies between different memory records in call store) did not have to run in a priority mode. If the reading exceeds 9999, the measurement should be ignored because one or both of the above conditions have not been satisfied. For readings less than 9999, division by 100 yields the percentage of time spent in the ***periodic deferrable portion*** of the base level scan.*

Example: If Q14 = 6667, the percentage of time spent in the ***periodic deferrable portion*** (PDT) of the total base level scans during the last 15 minutes equals 6667/100 or 66.67 percent (or 10 minutes out of the last 15 minutes).

*For offices with the LO-1 generic program, Q14 is valid only for Issue 4.6 and higher.

2.20 Q14, then, is related to the "call capacity remaining" in the system. It is recommended that the network administrator graph Q14 versus calls. (Additional information is contained in Dial Facilities Management Practices, Division H, Section 10h, No. 2 ESS Machine Capacity Management—Call Processor Capacity.) By definition, a call is any line origination that dials at least one digit or any trunk origination. Using data from the quarter-hour traffic schedule, calls would be the sum of the Q6 plus Q7 registers. As long as the number of base level scans remains constant, Q14 will decrease linearly with calls (assuming the mix of traffic does not change). When the number of base level scans starts to decrease, Q14 decreases at a slower rate as the number of calls increases. An example of a load service curve developed using these parameters is shown in Figure 2. Point E on Figure 2 indicates the point at which the number of base level scans starts to decrease.

2.21 Q11 records the total number of base level scans occurring in a 15-minute time period. Since the minimum length of a base level scan is 100.08 milliseconds, the maximum reading would be 8993. A load service curve using Q11 as one variable might be useful. As the load (call processing) increases, the time required for base level functions increases. This results in a decrease in the number of base level scans.

STATUS INDICATORS

2.22 It is advantageous for the network administrator to graph ***important*** items from the Q-schedule. The graphs can be used as valuable tools in capacity determinations, pinpointing call processing problem areas, determining busy periods, and providing machine performance records. The chosen items, which are termed ***status indicators,*** could include: originating plus incoming peg count (Q6 + Q8), dial tone speed test failures (Q3 + Q4), number of base level scans (Q12), number of 25-millisecond interrupts in all base level scans (Q11), PDT (Q14), etc.

2.23 It is recommended that the status indicators be graphed in every office under various load conditions. The graphing may be done directly from printout data. The various load conditions under which graphs should be prepared are:

(a) Busy periods—average load day

(b) Busy periods—high load day

(c) Overload condition

(d) Abnormal condition.

A. Busy Period—Average Load Day

2.24 To graph the busy period of an average day, preselect one day a month which is considered average in terms of the load it generally carries. In selecting the day, avoid, if possible, the selection of the same day each month (such as the second Tuesday). Once the day is selected, collect Q-schedule data during the busy hours and graph significant items. An example of an average load busy hour graph appears in Figure 3. The items chosen for the graph include Q6 + Q8 and Q11 registrations. It should be noted that as the demand for service increases, the magnitude of Q11 drops below the maximum reading of 8993.

B. Busy Period—High Load Day

2.25 To graph the busy period of a high day, preselect one day per month, considered a high day, and graph pertinent Q-schedule data. This day may well be the first Monday of the month or the first day after a 3-day weekend. An example of a graph from a busy period of a high day appears in Figure 4. Observation of the graph reveals a sharp decline in Q11 registrations, matched by a substantial increase in load (Q6 + Q8). During the peak load period dial tone speed test failures (Q3 + Q4) begin to manifest themselves.

C. Overload Period

2.26 The graphing of an overload condition can provide a valuable tool in determining No. 2 ESS equipment capacity deficiencies and can aid in the detection of call processing problem areas. An overload graph is perhaps the most meaningful of the four status indicators. It portrays a true profile of how the office performs under severe traffic pressures. An example is shown in Figure 5. First, note the drastic increase in traffic registrations (Q6 + Q8), as well as the appearance of a great many dial tone speed test failures (Q3 + Q4). Secondly, the diminution in the number of base level scans (Q11) and an increase in Q12 can be observed. Thirdly, and predictably, the amount of time spent on deferrable work operations (Q14) shows a decrease while Q13 registrations (maximum length of nondeferrable portion of one base level scan, measured in 25-millesecond increments) have increased. Also to be observed is the activation of dynamic service protection (DSP) and, later, the partial clearing of call store memory (MI SY CLR PCL). For additional information and usage of the overload period graphs in capacity determination refer to Dial Facilities Management Practices, Division H, Section 10h, No. 2 ESS Machine Capacity Management.

> ***Note:*** In preplanning for overload periods, the network administrator should determine the state of DSP (allow or deny mode). Following any stable clear initialization with LO-1, Issue 4.6, and EF-1, Issue 3.4, the system will automatically deny DSP and set dial tone speed test failures to zero. In offices with earlier generic programs, DSP is automatically allowed. Therefore, the deny message must be inputted on either the network adminstration or maintenance teletypewriter to prevent the activation of DSP.

D. Abnormal Period

2.27 It is also appropriate to graph data taken under abnormal conditions. It is understandable that the occurrence of abnormal conditions cannot be predetermined. However, when an abnormal condition arises and data are obtained, a graph of the data may provide useful information about the system operation. An example of an abnormal condition would be when the number of base level scans (Q11) and incoming and originating registrations (Q6 + Q8) are not inversely proportional (as one increases, the other should decrease). This is relatively easy to observe if normal Q-schedule reviews are conducted, assuming that the network administrator has developed a tolerance band within which these parameters fall. An example of an abnormal period graph is provided in Figure 6. It is observed that the normal relationship between Q11 and Q6 + Q8 does not exist, indicating possible machine irregularities.

2.28 The following points may be made regarding the relationship of the status indicators to the traffic load.

(a) As incoming and originating peg counts (Q6 + Q8) increase, the number of base level scans (Q11) should decrease. Q12 and Q13 probably will increase, while Q14 will decrease.

This results from the fact that as calling volume increases, the amount of work at each base level scan is greater.

(b) When the number of base level scans drops significantly, dial tone delays should increase.

(c) An extremely heavy load, requiring continuous, nondeferrable work during each base level scan, would cause the Q12 reading to increase drastically (maximum = 35,972). Q14, which measures periodic deferrable time, would decrease.

(d) If the Q13 reading rises above 16, this will usually be an indication that service degradation is occurring.

(e) The mimimum reading for Q11 is defined as (Q5 + Q7) peak divided by three. The number of base level scans and the traffic load will continue to decrease and increase, respectively, until the number of originations divided by three equals Q11 minimum.

(f) For certain trouble situations the Q14 register (PDT) will print out an invalid number as a "flag" to the network administrator. The normal range for Q14 is between 0 and 9999. The Q14 register is jammed to a very large number (in excess of 32,000) when an action is taken that could adversely affect the deferred time reading. This happens whenever the processors go out of synchronization, a teletypewriter request is placed for an audit, or the processor self-initiates a high-priority audit.

(g) As long as Q11 remains constant, Q14 will decrease linearly with an increase in calls (assuming the mix of traffic does not change). Typical idle system values of Q14 and Q11 are 6550 (65.5 percent) and 8993, respectively.

3. EMERGENCY PLAN

3.01 With the awareness of how the No. 2 ESS machine functions during periods of overload and abnormal conditions, it is absolutely vital that there is communication between the maintenance personnel and network administrator during these periods. It is also necessary that prior to an actual overload situation a prearranged plan be developed. Figure 7 contains considerations of what might comprise an emergency plan.

REGISTER	DESCRIPTION	COMMENT
Q01	Dial pulse dial tone speed tests performed.	The register pegged is determined by the type of receiver required by the line translation for the last call traced by the dial tone speed test program, not the type of receiver actually used; ie, in an office with 100% TOUCH-TONE receiver, a dial pulse line will peg a dial pulse count, even though a TOUCH-TONE receiver was used.
Q02	TOUCH-TONE® dial tone speed tests performed.	
Q03	Dial pulse dial tone speed test failures.	A failure is a dial tone speed test that was greater than 3 seconds.
Q04	TOUCH-TONE dial tone speed test failures.	
Q05	Total line originations.	Line originations are line off-hooks, counted before a CDR is connected — does not include dial tone speed tests (includes coin originations).
Q06	Total originating calls.	CDR connections plus at least one dialed digit (includes coin connections).
Q07	Total trunk originations.	Incoming seizures on bylink, one-way incoming, and 2-way trunks.
Q08	Incoming call attempts.	An incoming call is a call that originates from a trunk and is going to terminate to a line.
Q09	Incoming path overflow.	Overflow is scored if the path hunt fails between incoming trunk and terminating line.
Q10	Incoming calls made stable.	Incoming calls placed into talking state by the system.
Q11	Total number of base level scans.	A peg count of the number of base level scans.
Q12	Total length of the nondeferrable portion of all base level scans measured in 25-ms counts.	The accumulative value of all base level scan lengths measured by 25-ms intervals during this quarter hour. Does not include periodic deferrable time.
Q13	Maximum length of the nondeferrable portion of one base level scan.	Measured by 25-ms intervals during the nondeferrable portion of the base level scans.
Q14	An indication of the amount of time spent in the deferrable portion of the base level scan.	Q14 represents the periodic deferrable time only if over the past 15 minutes: (1) The system remained in sync (2) The audits did not have to run in a priority mode. If Q14 is greater than 9999, one or both of the above conditions could not be met. Consequently Q14 does not represent the time in the deferrable portion of the scan and should be ignored.

Fig. 1—Quarter-Hourly (Q) Schedule (2.15)

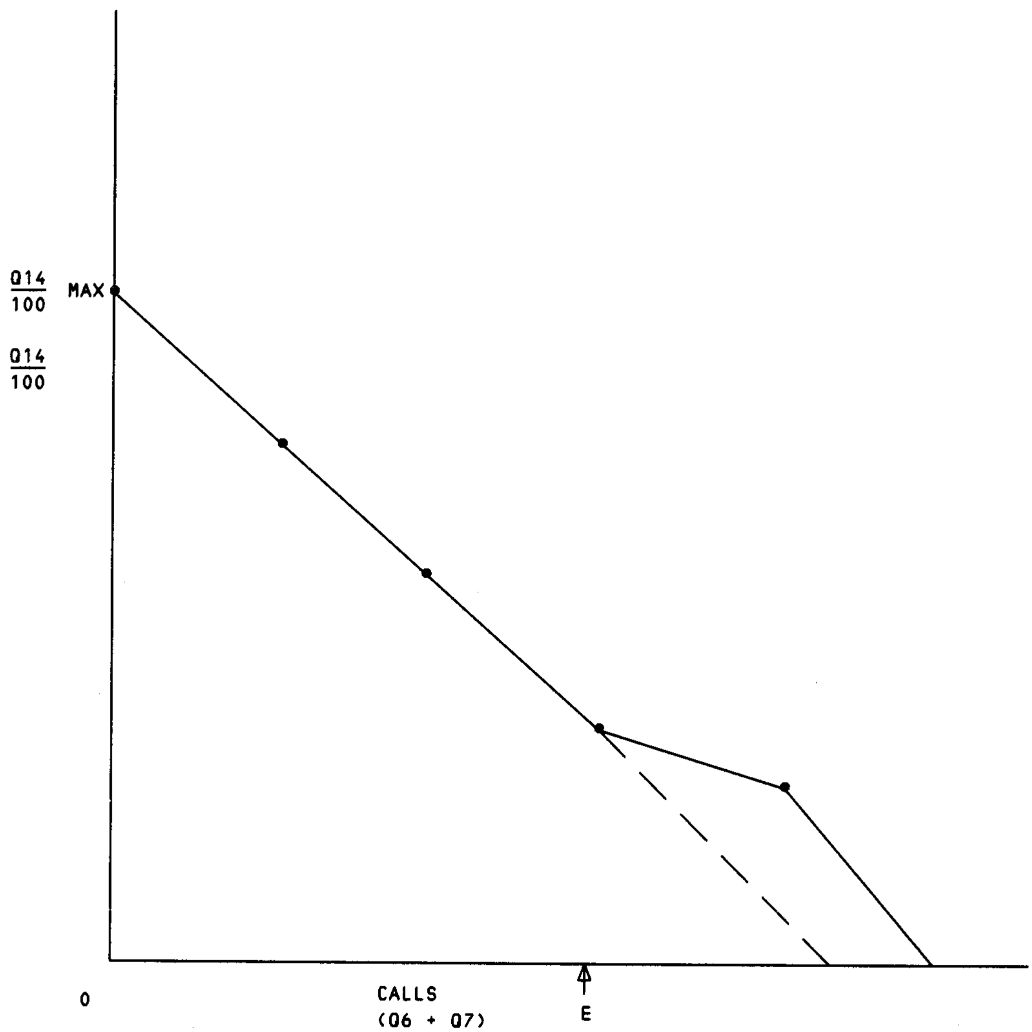

Fig. 2—Load Service Curve (2.20)

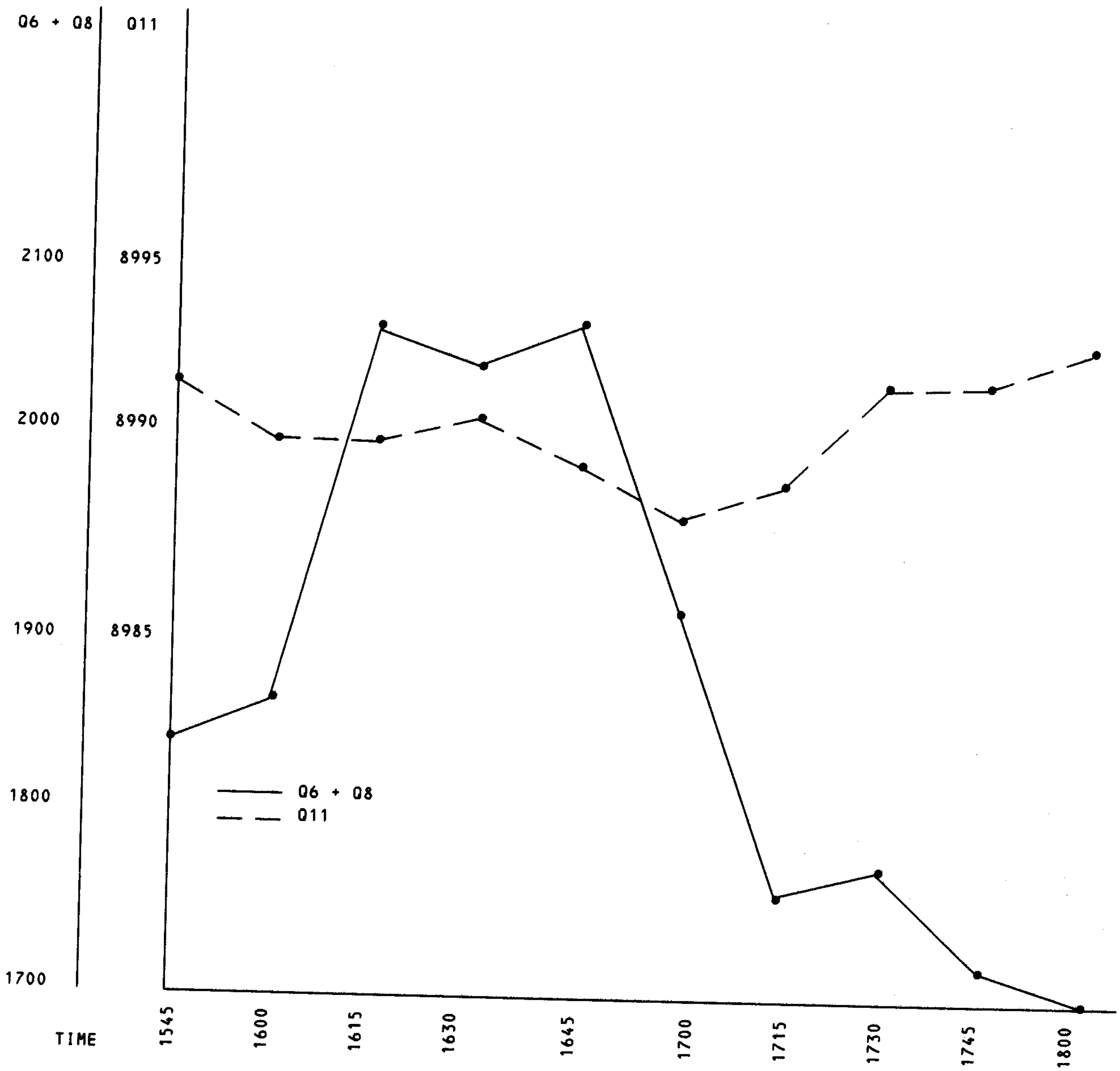

Fig. 3—Office XYZ Normal Day (Average-Load-Day Busy Period) (2.24)

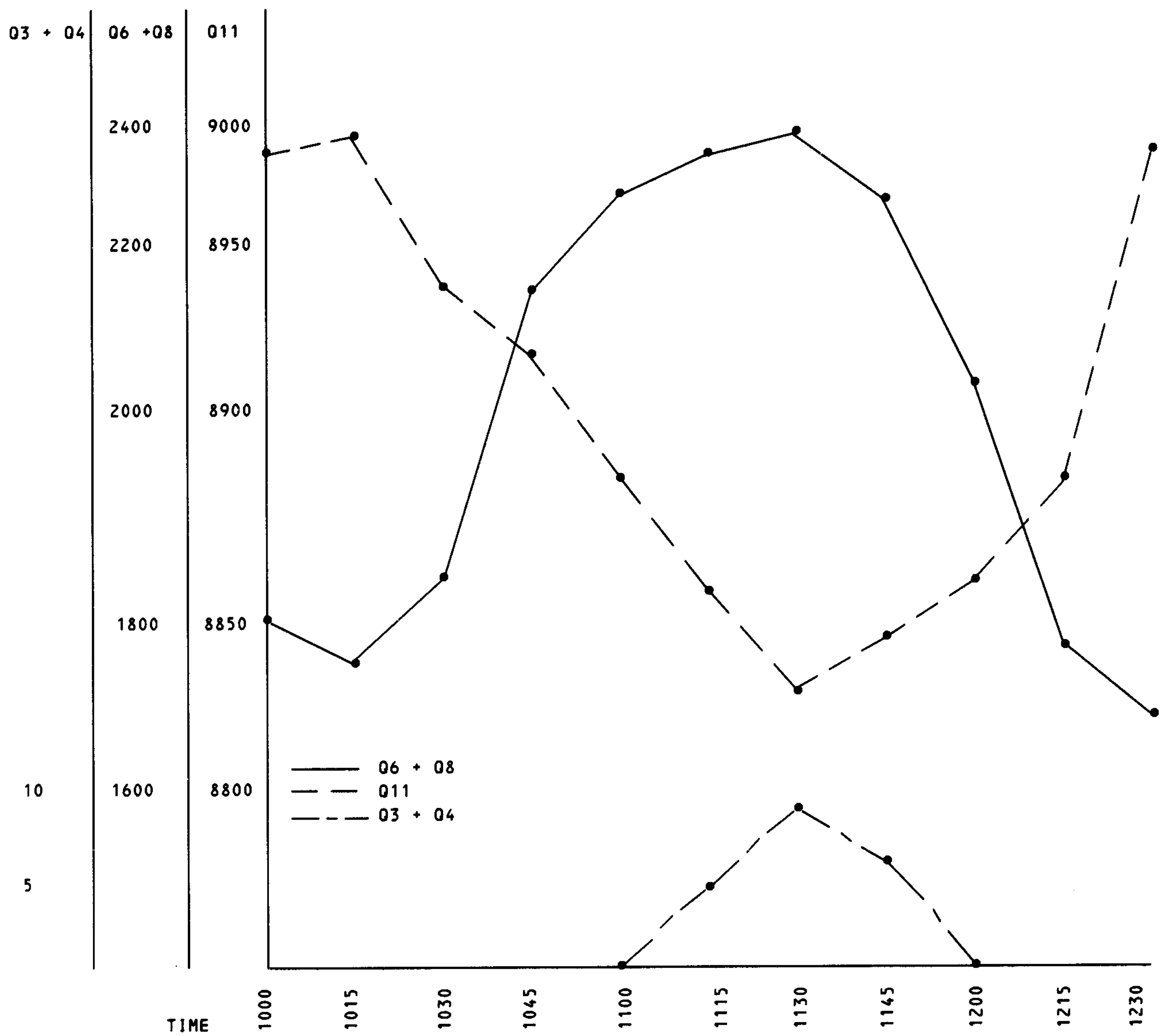

Fig. 4—Office XYZ High-Load-Day Busy Period (2.25)

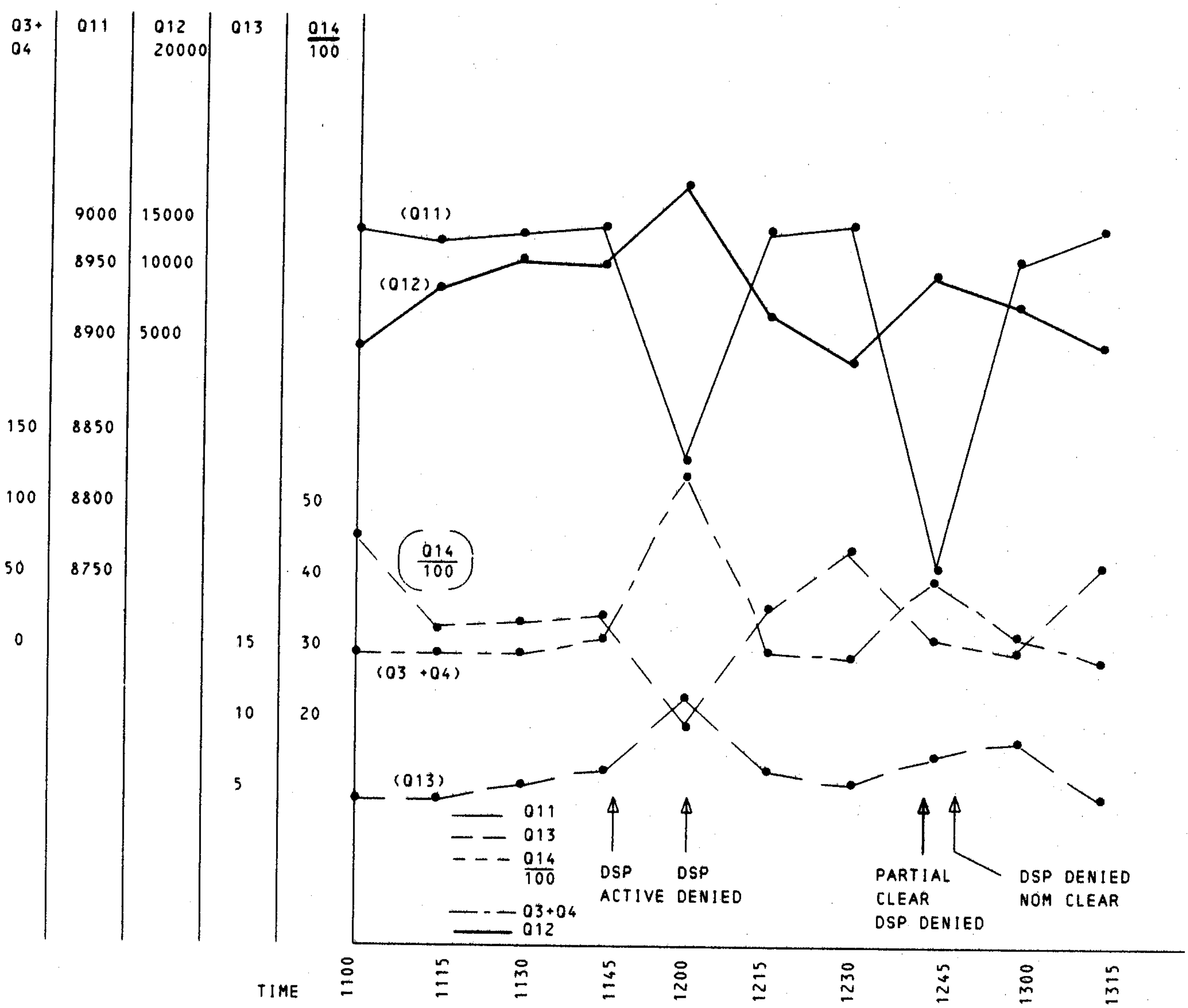

Fig. 5—Office XYZ Overload (2.26)

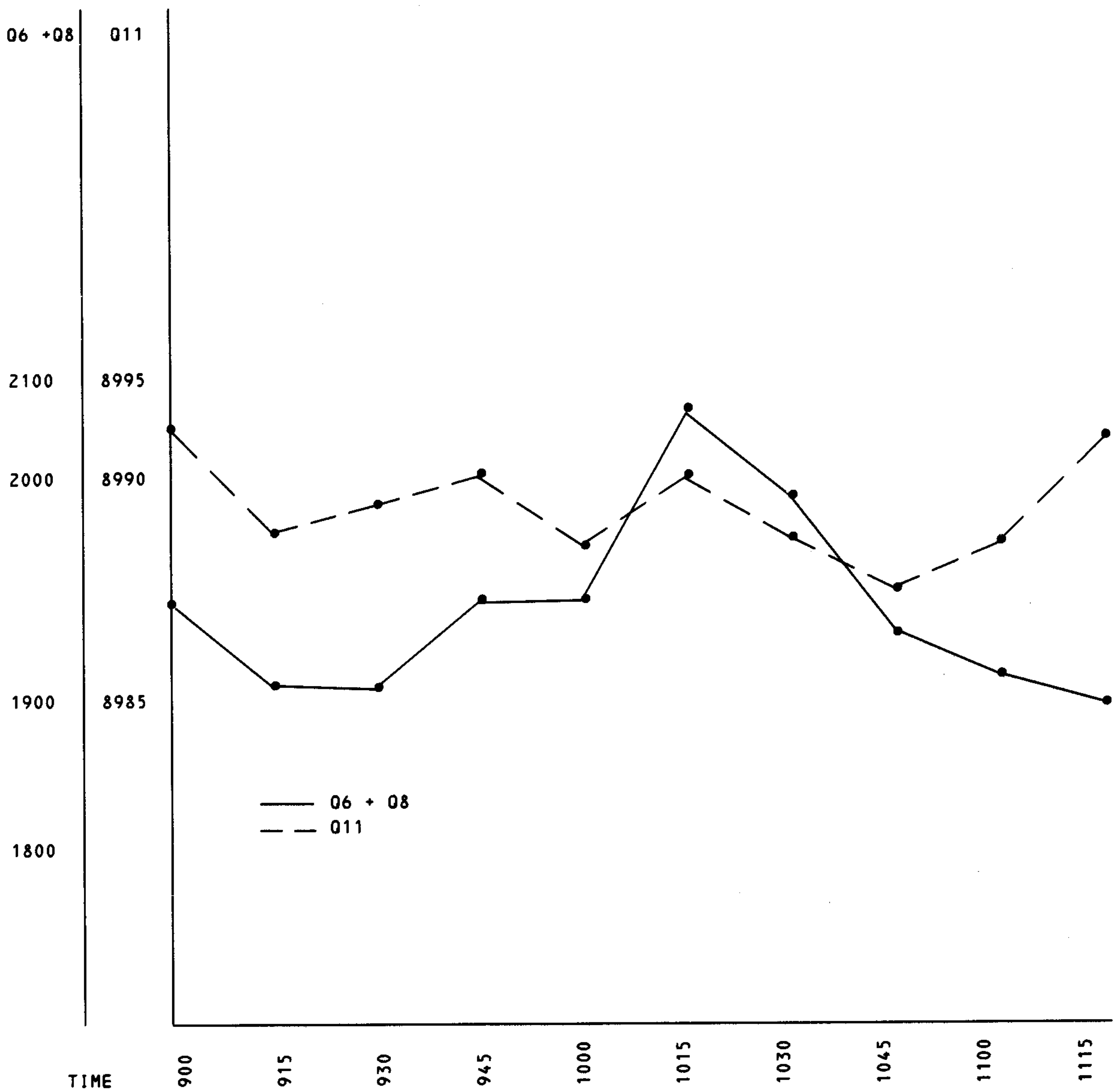

Fig. 6—Office XYZ Abnormal Condition (2.27)

1. Emergency call-out list — who should be notified
2. Abnormal conditions report
3. Dynamic service protection (DSP) report
4. Meanings of TTY messages
5. Operation of maintenance administration center (MAC) controls
6. Use of Q-schedule
   (a) When to implement DSP
   (b) Availability of graph paper
7. When traffic work table (TWT) is affected
8. Instructions as to how to make readings
9. If a remote office, information as to how to enter the building and a diagram of the office

Fig. 7—Emergency Plan (3.01)