Multicast Extensions to TP4

Kevin Thompson
MITRE Corporation
Reston, Va.

1.0 Introduction

This paper defines a set of extensions to TP4 for a reliable multicast service. The scope of this definition applies to a connection-oriented one-to-n Transport Layer multicast service operating over a connectionless network service (CLNS). Data flow is allowed in both directions of the multicast connection. Three group management functions are specified: member join, member leave, and membership poll request/response.

The domain of application of this service is communication between an initiating entity and multiple participants. Participants receive identical data and send data to the initiator only. Participants may join and leave the multicast connection depending on qualifications. MITRE has undertaken a simulation study and prototyping effort to examine the feasibility of these extensions in certain environments and to provide a proof-of-concept.

This description includes state transition diagrams and state tables of connection establishment/release and data transfer phases. A state transition diagram created for the basic TP4 service operating over CLNS is provided in Figure 1 as a point of reference. New service primitives, protocol data units, events, and actions are defined. Description of some procedures accompanies these definitions. Protocol mechanisms for possible implementation are also discussed.

The state transition diagrams appearing in this document use abbreviations and numbered codes to describe events, actions, and notes related to TP4 and the multicast extensions. These conventions follow those used in ISO 8073 and ISO 8073 Addendum 2. Tables identifying events, actions, and notes used in the extensions are also included in this section. The reader should refer to ISO 8073 and Addendum 2 for identifications of codes used in normal TP4 operations, all of which appear in these state diagrams.

2.0 Extensions

2.1 Connection Establishment and Release

Figure 2 depicts the state transition diagram for the connection establishment phase of the 1-to-n reliable transport multicast. The 1-to-n reliable transport multicast connection is initiated at a single End System, called the Initiator. All functioning responders, called Participants, receive a Multicast Connect Request TPDU. Those wishing to participate in the connection should respond with a Connect Confirm TPDU for successful multicast connection establishment. When the Transport Service user (TS-user) at the Initiator End System issues the Multicast Connect Request Service request to the Transport service provider, the service user may indicate to the service provider information regarding how long the service provider should wait in order to receive confirmation of connection acceptance as well as information, known as "Active Group Integrity" (AGI) information, regarding the minimum active group membership requirements that must be met in order for the establishment of the connection to be acceptable to the Initiator.

The AGI criteria are determined by application policy. All End Systems that respond affirmatively to the connection initiation request from the Initiator become part of what is known as the active group for the connection. Providing that the active group meets the AGI criteria, the Transport service provider will indicate successful connection establishment up to the Transport service user and, as part of this service primitive indication, the service provider may indicate to the service user the identities of the active group participants. Once the connection has been established, only the Initiator is able to send multicast data transfer on the connection. The n Participants to which the connection is made receive and positively acknowledge this multicast data transfer, providing full reliability. The multicast connection is released when the service user at the Initiator requests a disconnect. Individual Participants may leave the connection, but their departure will not trigger connection release unless their departure causes the active group membership characteristics to fall below AGI criteria.

2.1.1 New Protocol Data Units

Three additional multicast PDU types are required for the connection establishment phase of multicast service in the recommended extension to TP4: Multicast Connect Request/Confirm (MCR/MCC), Multicast Connect Set (MCSET), and Multicast Disconnect Request/Confirm (MDR/MDC).

Multicast Connect Request/Confirm (MCR/MCC)

The Multicast Connect Request PDU is sent by the Initiator of the multicast connection to either a group TSAP or a list of the individual TSAPs with which the connection is to be established. Once the Multicast Connect Request PDU has been initiated by the Transport Service provider (TS-provider), the Initiator has an outgoing multicast connection pending until he has received a Multicast Connect Confirm PDU from, at a minimum, some subgroup of recipients that meets the AGI criteria. When the AGI has been met and the connection pending timer (if supplied) expires, the TS-provider multicasts out an MCSET to the active group and confirms connection establishment to the TS-user. Any Multicast Connect Confirm PDUs that are received by the TS-provider after the connection has been established are treated as Join Connect Requests.

Multicast Connection Set (MCSET)

When the MCR is sent, it includes QoS parameter values proposed for use on the multicast connection by the Initiator. Specific QoS parameters include: connection establishment delay; connection establishment failure probability; throughput; transit delay; residual error rate; transfer failure probability; connection release delay; connection release failure probability; connection protection; connection priority; and resilience of the connection. If the service provider delivering the MCR wants to lower the proposed QoS offered on the connection, it may indicate this by lowering the QoS values in the Multicast Connect Indication service primitive that it delivers to each of the Participants. Each of the Participants in turn also has the prerogative to propose changes to the proposed quality of service, as long as the changes lower the proposed quality of service. The values proposed by the Participants, whether the same as the values received in the Multicast Connect Indication or whether changed by the Participant, are transmitted to the Initiator in the MCC PDU. Because the individual Participants may have proposed different QoS parameter values for the connection, the Initiator must determine the least common denominator of the proposed values it receives and send these values out to all Recipients in the Multicast Connect Set PDU as the final QoS parameter values for the connection. Hence, the purpose of the Multicast Connect Set Request is to ensure that all Participants are informed of the multicast connection QoS values that have been negotiated as acceptable to all connection participants.

Multicast Disconnect

The Multicast Disconnect Request PDU, when sent from the Initiator, indicates to the Participants that the multicast connection is being disconnected. If sent from one of the Participants, the MDR PDU indicates that either the Participant is rejecting a Multicast Connect Request that it has just received, or that it wishes to leave a connection of which it is already a member.

2.1.2 New Service Primitives

Seven additional multicast service primitives are required for the connection establishment/release phase of multicast service in this recommended extension to TP4: Multicast T-CONNECT request, Multicast T-CONNECT indication, Multicast T-CONNECT response, Multicast T-CONNECT confirm, Multicast T-CONNECT inform, Multicast T-DISCONNECT request, and Multicast T-DISCONNECT indication.

Multicast T-CONNECT Request (MTCONreq)

The MTCONreq service primitive is issued by the TS-user to request a multicast connection be established. By definition this primitive can take place at the Initiator only. A group identifier or a list of individual identifiers must be included in the primitive for the TS-provider to use in establishing a connection to the correct multicast group. This group identifier is equivalent to a called address in normal ISO transport terminology. Other parameters include the Initiator's calling address, expedited data option, QoS parameters, and TS user-data. In addition, the TS-user may include a connection pending timer value and/or an AGI criteria value that indicates to the TS-provider the minimum membership criteria that must be met by the group of end systems responding to the connection request in order for the connection request to be acceptable to the TS-user. If a connection pending timer value is specified by the TS-user, then upon expiration of this timer, (assuming the AGI criteria have been met), the TS-provider confirms the connection establishment to the

TS-user. If a connection pending timer value is not specified by the TS- user, then upon receipt of the last outstanding Multicast Connect Confirm required to meet the AGI the TS-provider confirms the connection establishment to the TS-user.

Multicast T-CONNECT Indication (MTCONind)

MTCONind is the primitive issued by the TS-provider of a Participant to the TS-user signalling the request for participation in the multicast connection. The same parameters as carried in MTCONreq still apply. If the TS-provider has reduced the QoS, new parameter values appear in MTCONind.

Multicast T-CONNECT Response (MTCONresp)

MTCONresp is issued by the TS-user of a Participant to its TS-provider in response to the received indication. This primitive causes the TS-user to issue a MCC TPDU back to the Initiator. Again, if the quality of service is reduced, new parameter values may appear. Parameters are quality of service, responding address, expedited data option, and TS-user data.

Multicast T-CONNECT Confirm (MTCONconf)

This confirmation primitive is given to the TS-user of the Initiator by the TS-provider. MTCONconf carries parameter values representing the minimum QoS values of those received by all participants through collection of CC TPDUs.

Multicast T-CONNECT Inform (MTCONinf)

The MTCONinf primitive is delivered to the TS-user of a Participant by the TS-provider to signal receipt of a MCSET TPDU. Its sole purpose is to inform the user of changes to the QoS parameters made during connection establishment. This change is made by the Initiator's calculation of a minimum set of QoS values from all Participant MCC responses.

Multicast T-DISCONNECT Request (MTDISreq)

This primitive is a legal response by a Participant's TS-user to a received MTCONind primitive; it is used by the Participant to refuse a multicast connection and also, in the data transfer phase, to leave a multicast connection in progress. If issued by the Initiator, the MTDISreq is used to close the multicast connection. Its only parameter is TS user data.

Multicast T-DISCONNECT Indication (MTDISind)

If received by the Initiator, a MTDISind indicates to the Initiator that the sending Participant is either rejecting a MCR that it has just received, or that it is leaving the multicast connection. If received by the Participant, a MTDISind indicates that the Initiator is closing the multicast connection. The single parameter carried by MTDISind is a Participant id.

2.2 Data Transfer

The data transfer phase of the 1-to-n reliable multicast comes after connection establishment, when the Initiator and all Participants have reached the Open state. Figure 3 shows the state transition diagram for data transfer. The new PDUs and service primitives described below relate mainly to the multicast data transfer in the 1-to-n direction, not unicast data in the n-to-1 direction. Figure 3 accounts for this part of the service, however, and it is specified according to normal data transfer procedures for unicast on connection-oriented transport found in ISO 8073 and Addendum 2.

2.2.1 New Protocol Data Units

Nine additional TPDUs are defined for use in the data transfer phase of this multicast connection: Multicast Data, Multicast Expedited Data, Join Connection Request, Join Connection Confirm, Multicast Connection Reset, Join Connection Reject, Multicast Connect Confirm, Unicast Membership Poll Request, and Unicast Membership Poll Response.

Multicast Data TPDU (MDT)

The Multicast Data TPDU carries user data from the Initiator to all participants. As with PDUs for connection establishment and release, a copy of an MDT must be retained by the Initiator to allow for later retransmission if necessary. An MDT TPDU is considered delivered only after acknowledgment by all participants.

Multicast Expedited Data TPDU (MED)

The Multicast ED TPDU is the multicast TPDU for expedited data transfer. See section 6.11 of ISO 8073 for the description of expedited data transfer for unicast data, to which the MED TPDU applies. Each MED TPDU received is acknowledged by an Expedited Acknowledge (EA) TPDU from each participant. Only one outstanding MED is allowed at any time.

Join Connection Request (JCR)

The JCR TPDU is generated by a potential Participant not yet part of the existing multicast connection. Like a CR TPDU and having the same parameter list, JCR is a direct request to the Initiator to join a multicast connection. A special parameter is needed to indicate which multicast connection to join. It is possible for the request to be deemed unacceptable by the Initiator in the same ways an ISO 8073 CR TPDU may be unacceptable, but additionally it may be rejected for specifying inadequate parameter values. The basis for such decisions, or policy, is left as a matter of local choice as per standard protocol definition practices.

Join Connection Confirm (JCC)

The JCC primitive supports the group management function of joining an active group. The JCC TPDU is sent by the Initiator to an ES who has requested to join an existing multicast connection. The JCC results in adding the requesting ES to the group of participants on the connection. The JCC TPDU parameter list is the same as that of the CC TPDU with two exceptions: 1) a parameter is introduced to specify by sequence number the current position in the sequence space of the multicast data; 2) there is a difference between CC and JCC TPDUs in how QoS parameter values are treated. The parameter values in the JCC TPDU represent an assignment to the new Participant and are not changeable. A JCC TPDU is sent either after an acceptable JCR TPDU has been received with parameter values at or above those set already for the existing connection, or, in special cases, if the Initiator's policy is to reset the QoS on the connection to lower values in order to accommodate the JCR. Thus, the Initiator simply unicasts back the set parameter values used by the existing multicast connection in the JCC TPDU. The last part of the three-way handshake must still occur from the new Participant before the Participant reaches the Open state.

Multicast Connection Reset (MCRST)

The MCRST TPDU is multicast by the Initiator to all Participants, signifying a resetting of QoS parameters for the connection. All participants are required to explicitly acknowledge receipt of this TPDU before the Initiator or any Participant returns to the states associated with data transfer. MCRST is used by the Initiator to lower QoS parameters of the existing multicast connection according to values proposed by a new Participant joining the group.

Join Connection Reject (JRJ)

The JRJ TPDU is the negative reply by an Initiator to a JCR TPDU, rejecting a request to join a multicast connection. The reason for the rejection (which may be unsupported QoS requirements by the group or disqualification based on member identification among other reasons) is part of the parameter list. The JRJ TPDU is unicast to the sender of the JCR TPDU.

Multicast Connect Confirm (MCC)

The MCC TPDU, which has already been defined in section 2.1.1 for multicast connection establishment, is also available in the data transfer phase. Depending on AGI criteria, it is possible for the Initiator to receive one or more MCC TPDUs after moving to the open state of data transfer. The Initiator's TS-provider treats an MCC TPDU as a request to join the existing connection (a JCR TPDU).

Membership Poll Request (PREQ)

The PREQ TPDU is sent by a Participant during the data transfer phase to the Initiator, asking for a list of TSAPs representing the current membership of the connection.

Membership Poll Response (PRESP)

The PRESP TPDU is the response to a PREQ TPDU. The TPDU contains a list of current members of the multicast connection identified by their TSAPs. This TPDU is unicast by the Initiator to the polling Participant, and it does not require acknowledgment.

2.2.2 New Service Primitives

Multicast T-DATA Request (MTDTreq)

The MTDTreq primitive comes from the Initiator's TS-user requesting multicast data transfer by the TS provider. The single parameter is TS user data.

Multicast T-DATA Indication (MTDTind)

The MTDTind primitive is given to the TS user of a participant by its provider carrying delivered TS user data.

Note that, as with ISO 8073 unicasts, no confirmation of data delivery by the receiving entity is represented in the service primitives. Reliability is considered an issue for the protocol mechanisms.

Multicast T-DISCONNECT Indication (MTDISind)

If received by the Initiator, a MTDISind indicates to the Initiator that the sending Participant is either rejecting the MCR that it has just received, or that it is leaving the connection. If departure of this participant causes the active group membership characteristics to fall below AGI criteria, then the Initiator will start the release of the connection. If received by the Participant, an MTDISind means that the Initiator is releasing the connection. The single parameter carried by MTDISind is a Participant id.

3.0 Procedures and Mechanisms

3.1 Flow Control

3.1.1 ACK Collection

For the Initiator's transport entity to ensure delivery to all n active Participants, multicast data must be ACKed by all n Participants. This means an ACK collection procedure at the Initiator examines all acknowledgments received for multicast data, and records, according to the source and acknowledgment number, which of the n Participants have acknowledged any given unit of multicast data.

3.1.2 Sliding Window

Multicast data is not considered delivered until all needed Participants acknowledge receipt. A send window procedure is used to control the flow of data on the multicast connection, and like other window procedures, this window slides across the sequence space of data according to three basic parameters: the recipient's advertised window, data which has been sent, and acknowledgment of data from the recipient. In the multicast case, however, the advertised window may differ across the n Participants; the procedure used for sliding the send window must always act on the minimum window advertised by any recipient. This method prevents an overflow of data to any Participant and improves overall efficiency by reducing the possibility of retransmissions. But it lowers throughput for those individual participants that could tolerate a larger window.

It is possible for the window to reduce to 0, meaning no more data may be sent until the window is reopened, either by acknowledgment from all Participants of previously sent data, or by new advertised window information. Figure 3 and the accompanying tables account for this state, labeled "AKWAIT".

3.1.3 Multicast Retransmissions

Depending on topology characteristics of the multicast connection and the environment, failure to deliver data to one Participant may or may not imply failure of delivery to other Participants. Here it is assumed that the Initiator knows, at the time of retransmission, exactly which Participants have not received the data units in question. In the target environment for this multicast service, it is possible to unicast or multicast data from the Initiator if all TSAPs of the Participants are available to the Initiator's transport entity. When a retransmission timer expires, the choice of unicasting a DT TPDU to each unacknowledged Participant, or sending an MDT TPDU to the entire group, is made on the basis of the number of unacknowledged Participants.The default procedure for retransmissions is to multicast to all Participants. Participants receiving multicast TPDUs that they have already acknowledged will discard duplicate data, but must respond with an explicit acknowledgment in case previous acknowledgments were lost.

3.1.4 Implosion Prevention

A procedure for implosion prevention should be incorporated into the protocol to stagger receipt of acknowledgments at the Initiator, preventing local flooding of resources by the simultaneous arrival of up to *n* ACKs. This issue has been studied in several papers, although a pragmatic solution has not been implemented and tested in an internet environment. Implosion prevention is another procedure whose operation is highly dependent on topology and delay characteristics of the connection.

3.2 Group Management

3.2.1 Join Normal

A normal join operation in this service is simply a sequence of events resulting in a new Participant joining the existing multicast connection, but having no effect on the Participants or any other characteristics of the connection. In response to an acceptable JCR TPDU with QoS requested at or above the level serviceable by the existing connection, the Initiator unicasts back an JCC TPDU confirming its addition to the list of Participants. The JCC TPDU also includes any parameters describing the current state of the connection, such as current sequence number of multicast data. The new Participant is not required, or able, to receive data that was sent previous to its joining the group.

3.2.2 Join with Accepted Lowering of QoS

A special condition join is allowed by this service: a join resetting the multicast connection's QoS to some lower level. An ES wishing to join a multicast connection sends an JCR TPDU, but requests a level of QoS lower than the current state of the connection. The decision to accept or reject the join request based on this condition is a local policy choice. If the request is rejected, the Initiator sends a JRJ to the ES. If the request is accepted, the Initiator must perform a reset of the connection, informing all participants by multicasting an MCRST which describes the new QoS parameters. The Initiator then remains in the AKWAITN state (see Figure 3) until all Participants have acknowledged the reset. This reset affects only the QoS levels of the connection.

3.2.3 Leave

A Participant may leave the connection by sending a MDR TPDU to the Initiator. The Initiator responds with a MDC TPDU confirming the release, and removes the Participant from the connection entirely.

3.2.4 Membership Polling

Participants in the connection are allowed to poll the Initiator for a membership status. The poll request is in the form of a PREQ TPDU. The response, a PRESP TPDU, is unicast back to the Participant, and includes a list of current Participants in the connection. The PRESP TPDU does not require acknowledgment.

4.0 Areas for further study

The following list describes some of the open issues in the development of this multicast service:

- Acknowledgment strategy minimizing overhead without compromising full reliability
- Acknowledgment time slotting to prevent network implosion
- Flow control mechanisms
- Decision algorithms for retransmissions by unicast or multicast
- AGI functionality
- Group addressing and implications of member addresses being unknown to the Initiator's transport service
- Group management functionality to enhance the services provided here
- Addition of unreliable transport within this service

Table 1: New Incoming Events (to TS-Provider)

Abbreviated Name	Category	Name
MTCONreq	TS-user	Multicast T-CONNECT Request primitive
MTDISreq	TS-user	Multicast T-DISCONNECT Request primitive
MTCONresp	TS-user	Multicast T-CONNECT Response primitive
MTDTreq	TS-user	Multicast T-DATA Request primitive
MCR	TPDU	Multicast Connection Request TPDU
MCC	TPDU	Multicast Connection Confirm TPDU
MDR	TPDU	Multicast Disconnect Request TPDU
MDC	TPDU	Multicast Disconnect Confirm TPDU
JCR	TPDU	Join Connection Request TPDU
JCC	TPDU	Join Connection Confirm TPDU
MDT	TPDU	Multicast Data TPDU
MED	TPDU	Multicast Expedited Data TPDU
MCRST	TPDU	Multicast Connection Reset TPDU
MCSET	TPDU	Multicast Connection Set TPDU
JRJ	TPDU	Join Reject TPDU
PREQ	TPDU	Unicast Membership Poll Request TPDU
PRESP	TPDU	Unicast Membership Poll Response TPDU

Table 2: New Outgoing events (from TS-Provider)

Abbreviated name	Category	Name
MTCONconf	TS-provider	Multicast T-CONNECT Confirm primitive
MTCONind	TS-provider	Multicast T-CONNECT Indication primitive
MTDISind	TS-provider	Multicast T-DISCONNECT Indication primitive
MTDTind	TS-provider	Multicast T-DATA Indication primitive
MCR	TPDU	Multicast Connection Request TPDU
MCC	TPDU	Multicast Connection Confirm TPDU
MDR	TPDU	Multicast Disconnect Request TPDU
MDC	TPDU	Multicast Disconnect Confirm TPDU
JCR	TPDU	Join Connection Request TPDU
JCC	TPDU	Join Connection Confirm TPDU
MDT	TPDU	Multicast Data TPDU
MED	TPDU	Multicast Expedited Data TPDU
MCRST	TPDU	Multicast Connection Reset TPDU
MCSET	TPDU	Multicast Connection Set TPDU
JRJ	TPDU	Join Reject TPDU
PREQ	TPDU	Unicast Membership Poll Request TPDU
PRESP	TPDU	Unicast Membership Poll Response TPDU

Table 3: New Specific Actions

Name	Description
[10M]	Set initial credit for controlling reception according to the sent MCR/MCC TPDU
[23]	ncount = ncount + 1
[24]	ncount = 0
[25]	set initial credit for sending according to minimum advertised value from received MCR/MCC TPDUs
[26]	send a multicast TPDU to complete 3-way handshake
[27]	exercise expedited data transfer procedures as described in text
[D1]	readjust send window
[D2]	set a retransmission timer
[D3]	add DT to send queue
[D4]	if needed remove acknowledged TPDUs from retransmission queue
[D5]	stop timer for acknowledged TPDU if all recipients have now acknowledged
[D6]	remove MDT TPDU copies from send queue, send, and readjust send window until send queue is empty or send window is 0
[D7]	add MDT to retransmission queue
[D8]	stop retransmission timer
[D9]	ncount = ncount - 1

Table 4: New Predicates

Name	Description
PD0	Multicast T-CONNECT Request is acceptable
PD1	send window = 0
PD2	send queue is empty
PD3	AK is for previously unacknowledged TPDU
PD4	readjusted send window
PD6	AK is out of sequence
PD7	readjusted send window is of greater size than send queue
PD8	JCR is acceptable
PD9	parameter values presented by new Participant fall below minimum set at multicast connection establishment
PD12	ncount = maximum
PD13	ncount = 1

Table 5: New States

Abbreviated Name	Name
WFMCC	Wait for n CC TPDUs
WFMTRESP	Wait for multicast T-CONNECT response
MCIWAIT	Wait for MCSET indication from multicast Initiator
SROPEN	Open for sending and receiving multicast and unicast data
AKWAITN	Wait for n AK TPDUs of new QoS for multicast connection

Table 6: New Specific Notes

Name	Description
5M	not a duplicated MCR TPDU

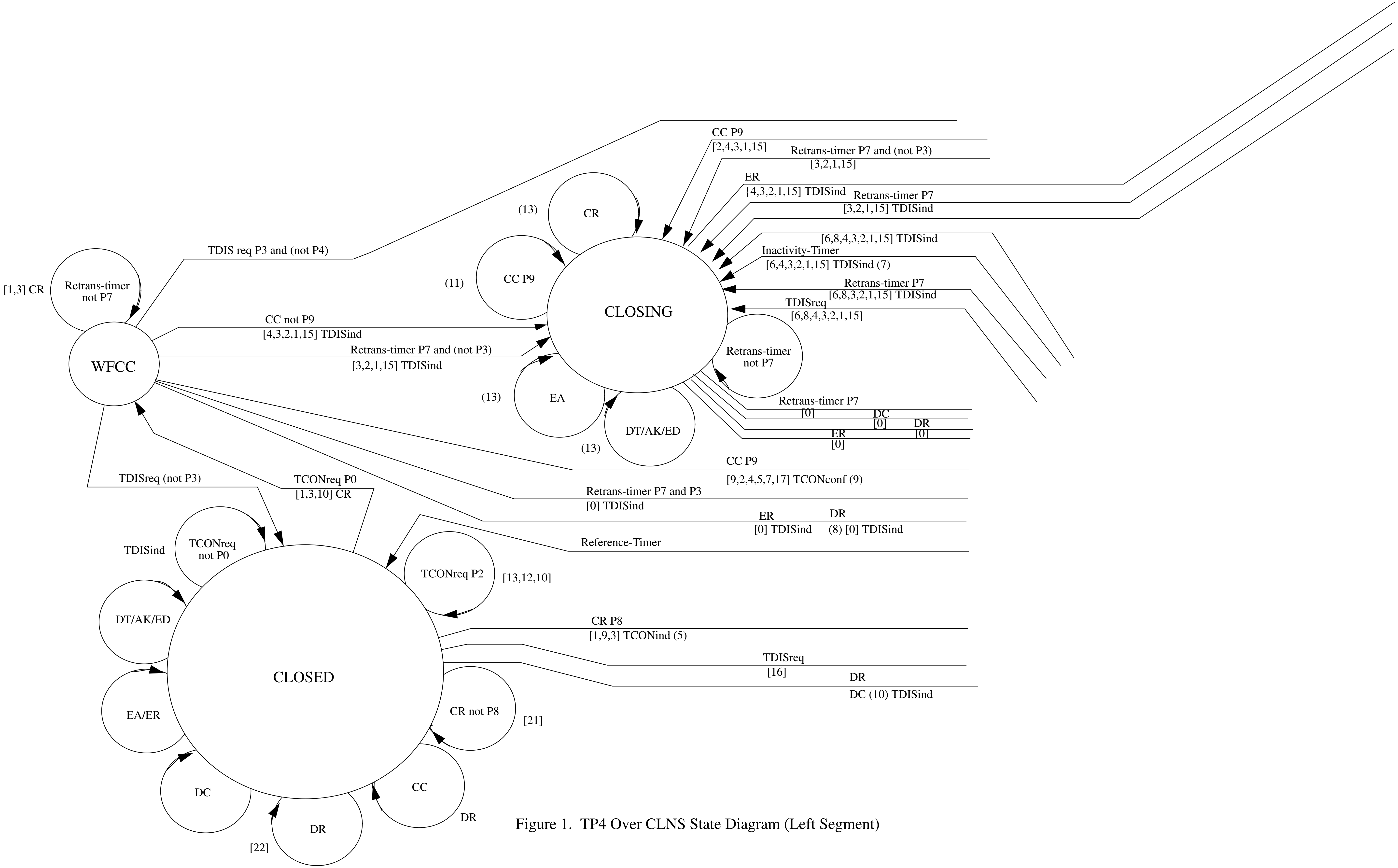

Figure 1. TP4 Over CLNS State Diagram (Left Segment)

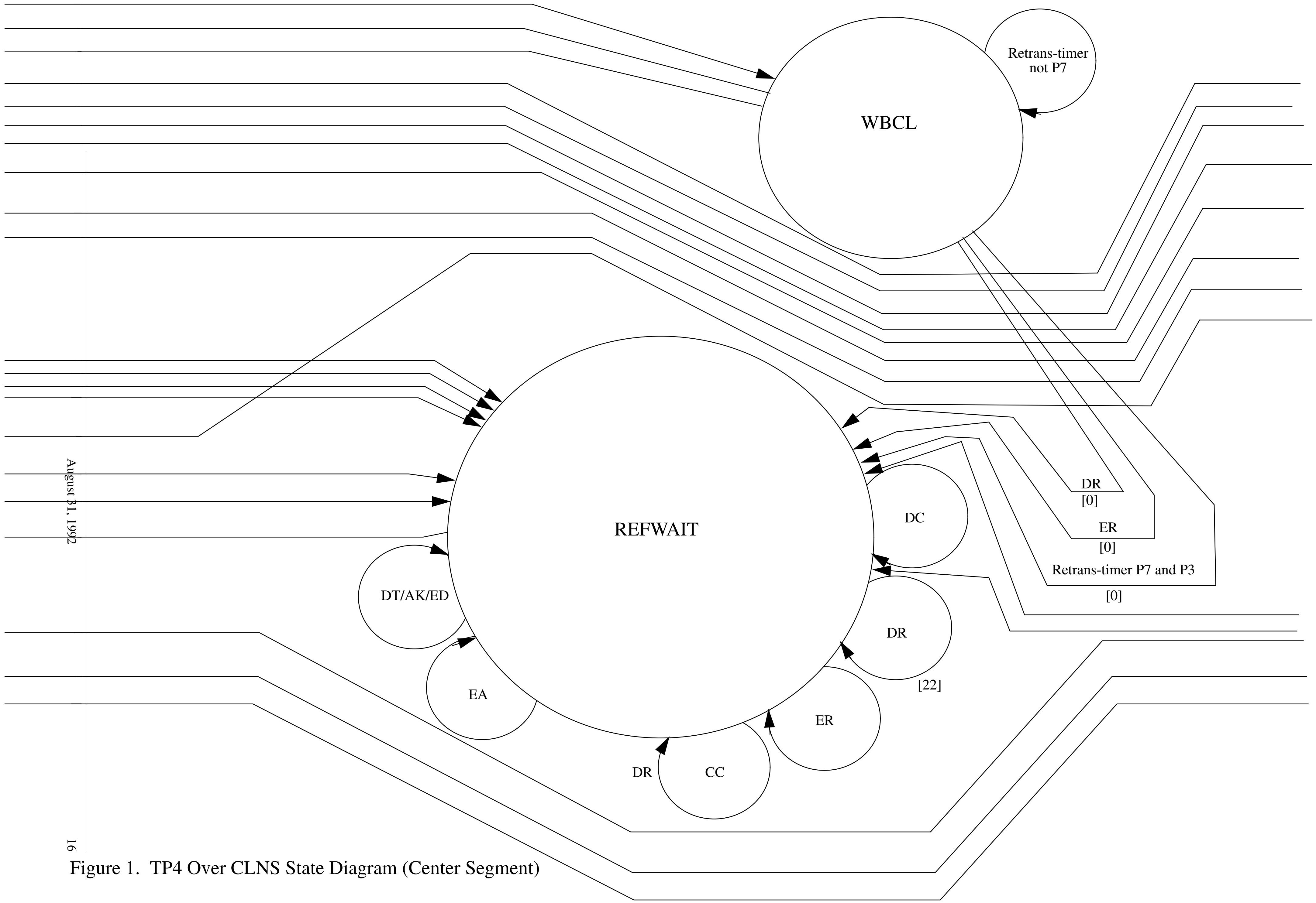

Figure 1. TP4 Over CLNS State Diagram (Center Segment)

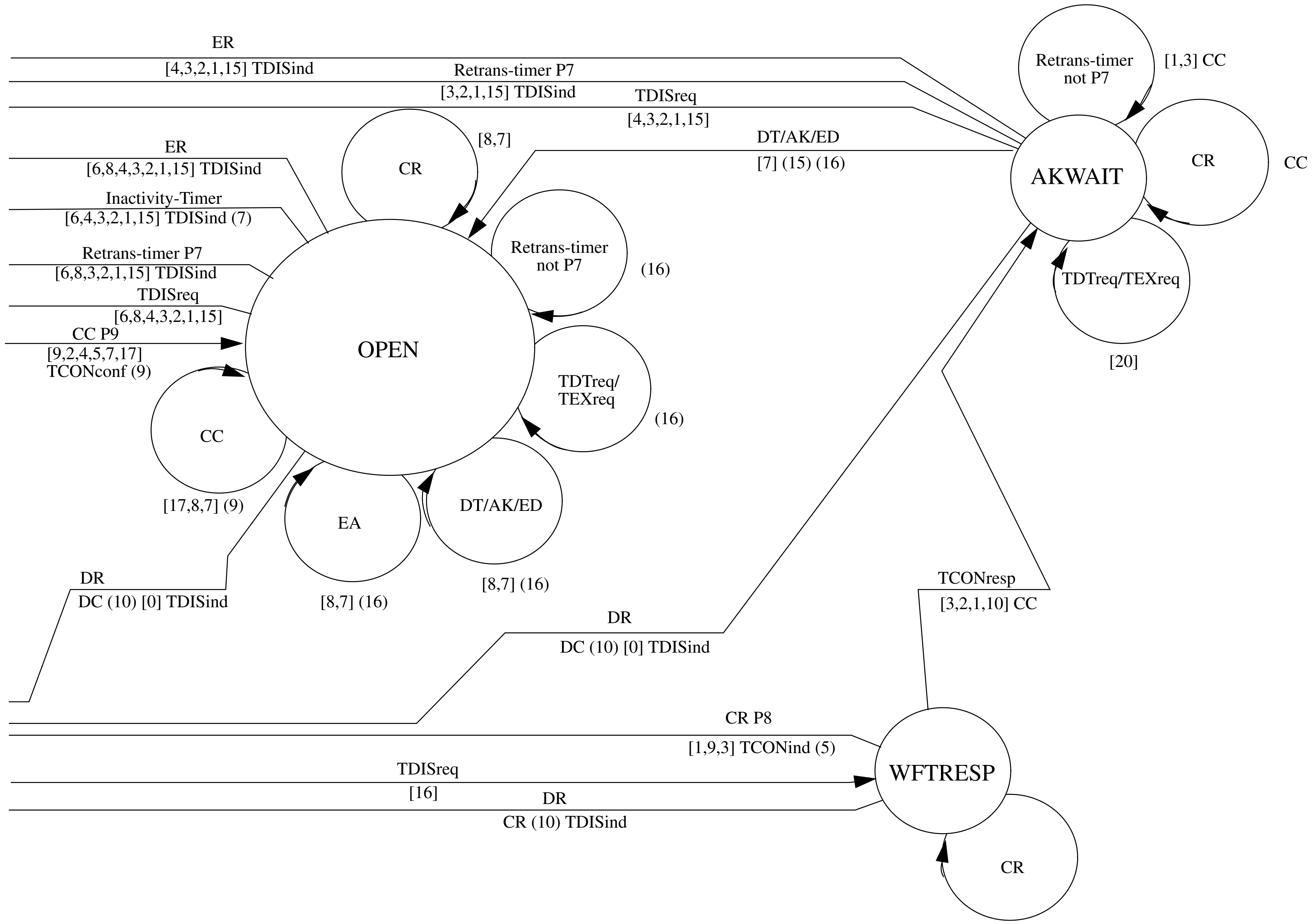

Figure 1. TP4 Over CLNS State Diagram (Right Segment)

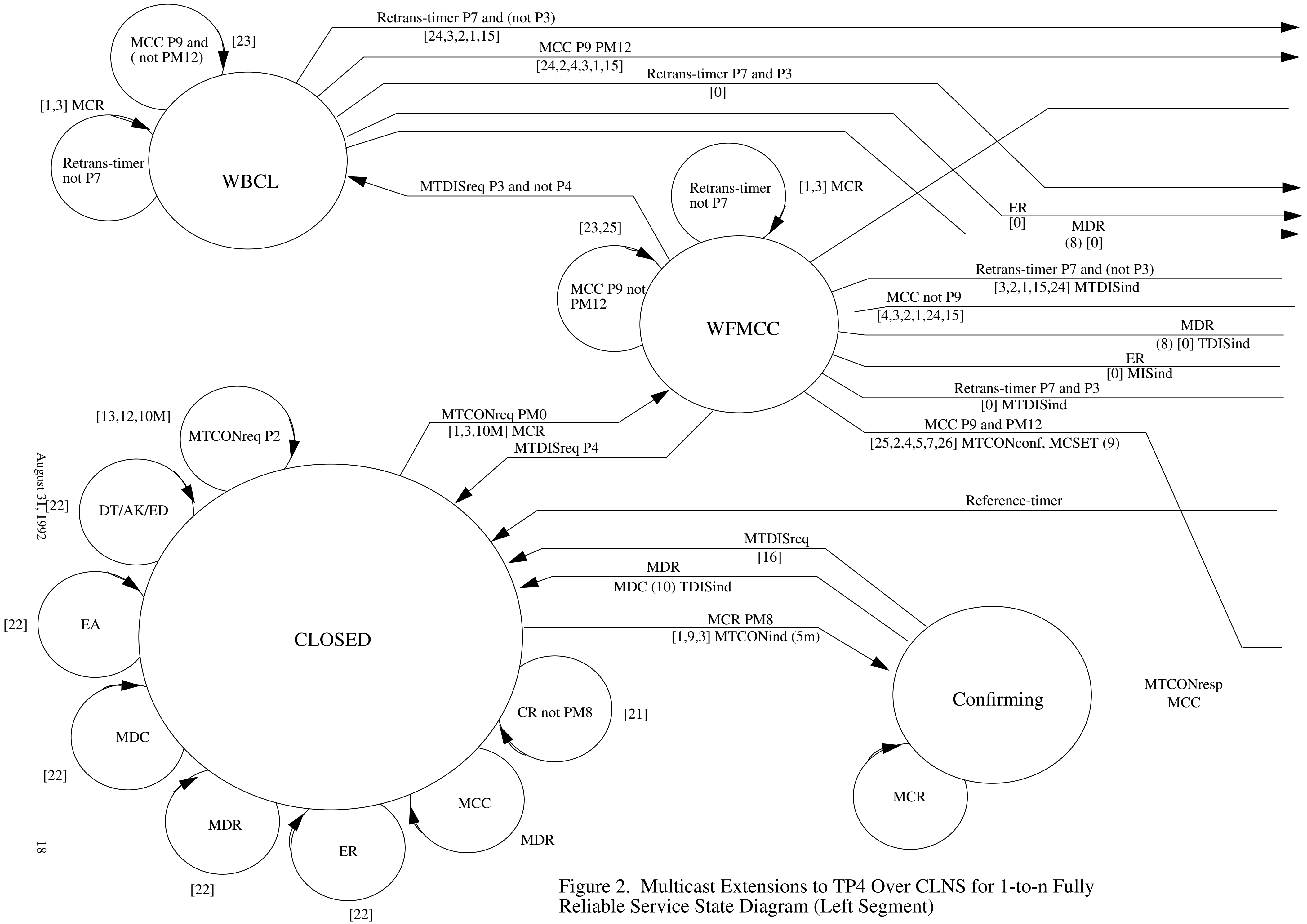

Figure 2. Multicast Extensions to TP4 Over CLNS for 1-to-n Fully Reliable Service State Diagram (Left Segment)

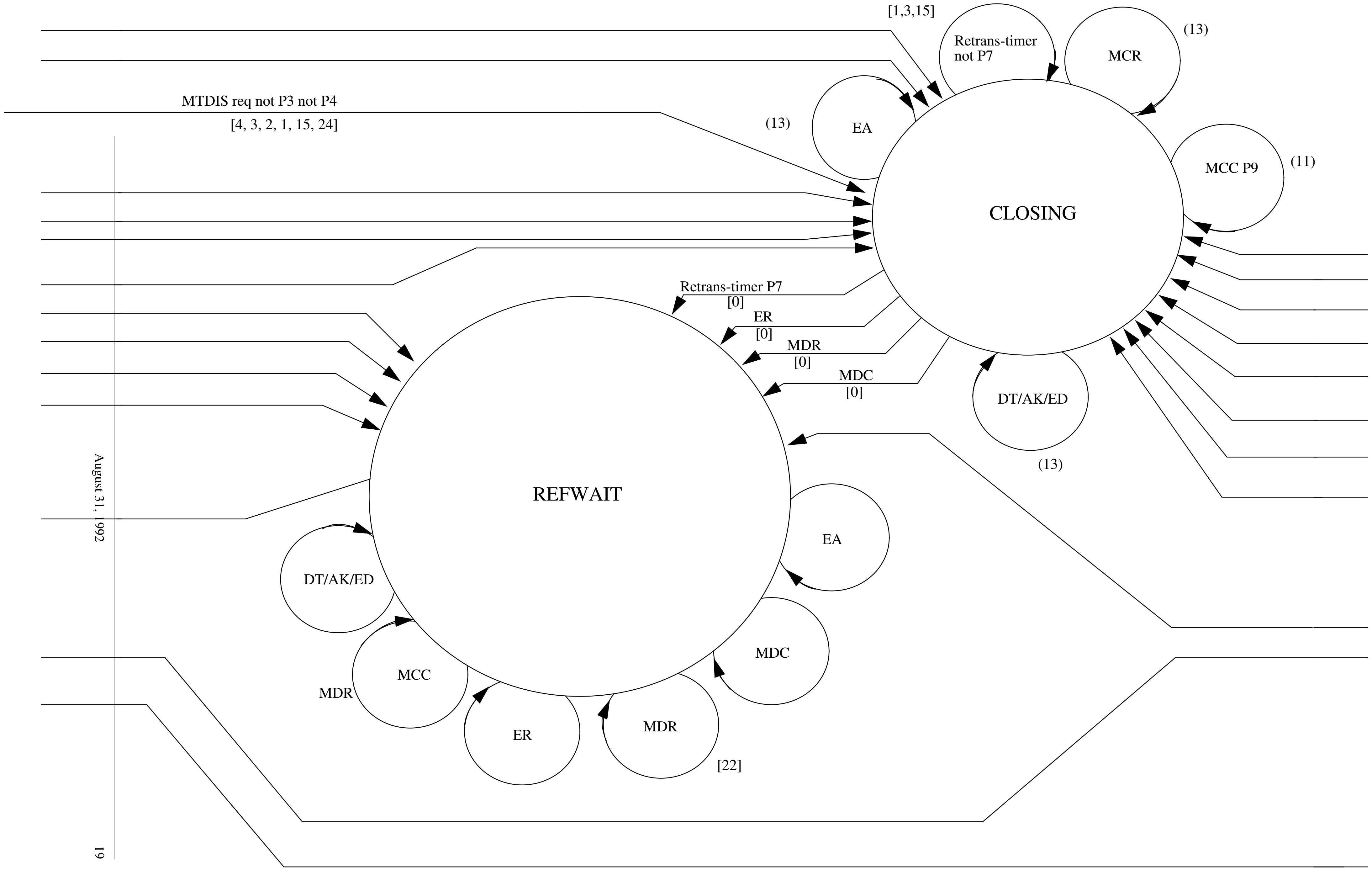

Figure 2. Multicast Extensions to TP4 Over CLNS for 1-to-n Fully
Reliable Service State Diagram (Center Segment)

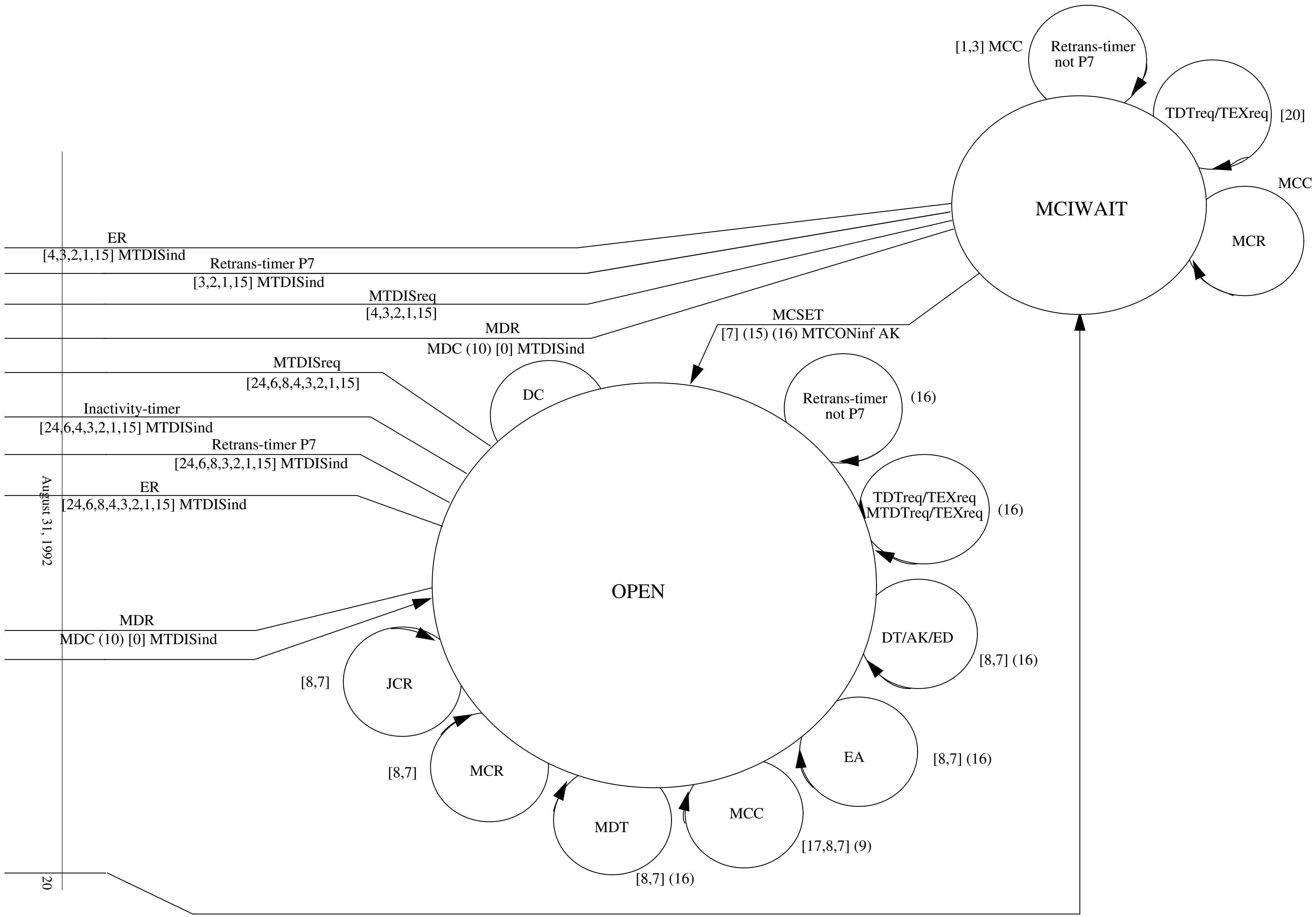

Figure 2. Multicast Extensions to TP4 Over CLNS for 1-to-n Fully Reliable Service State Diagram (Right Segment)

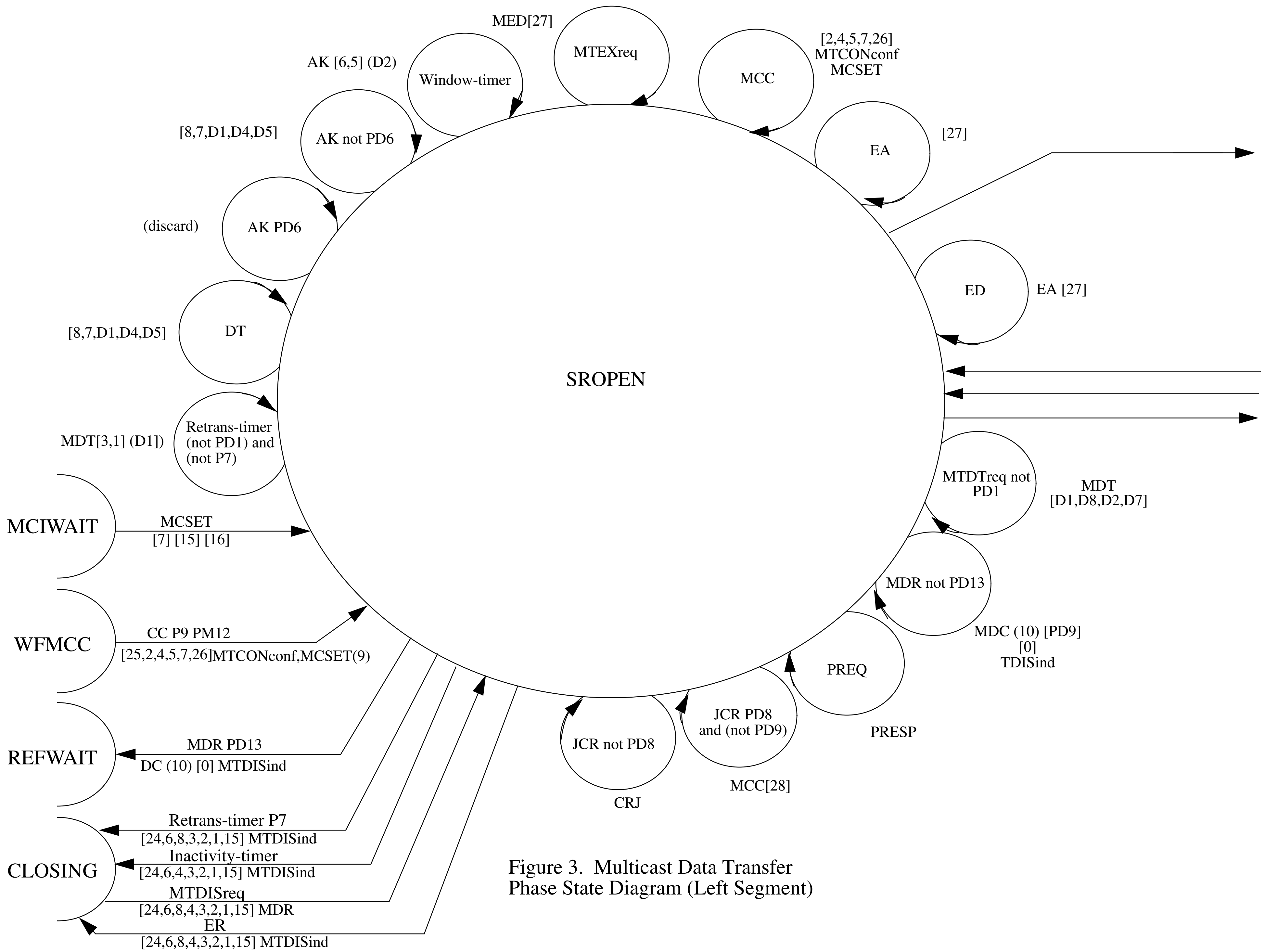

Figure 3. Multicast Data Transfer
Phase State Diagram (Left Segment)

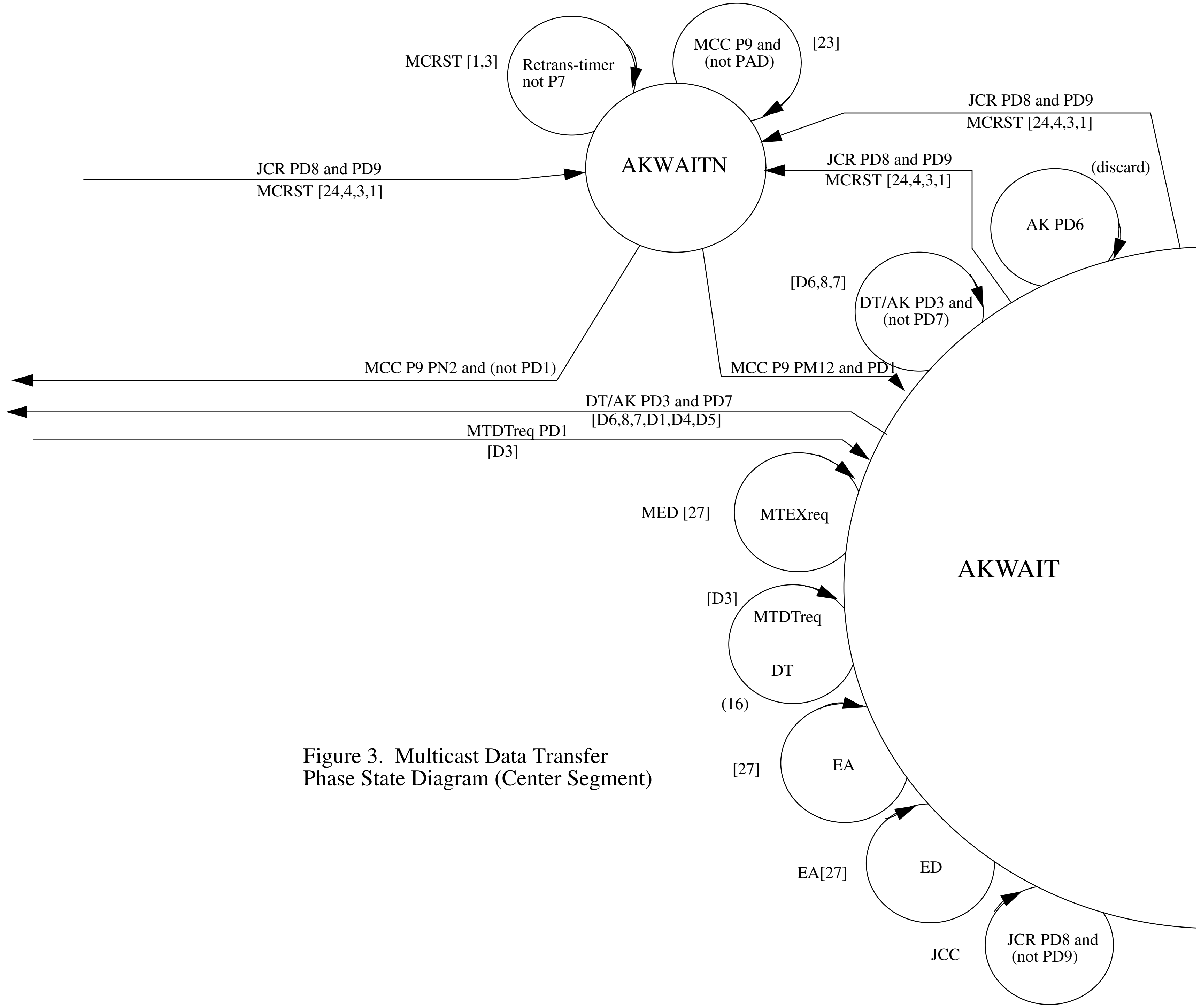

Figure 3. Multicast Data Transfer
Phase State Diagram (Center Segment)

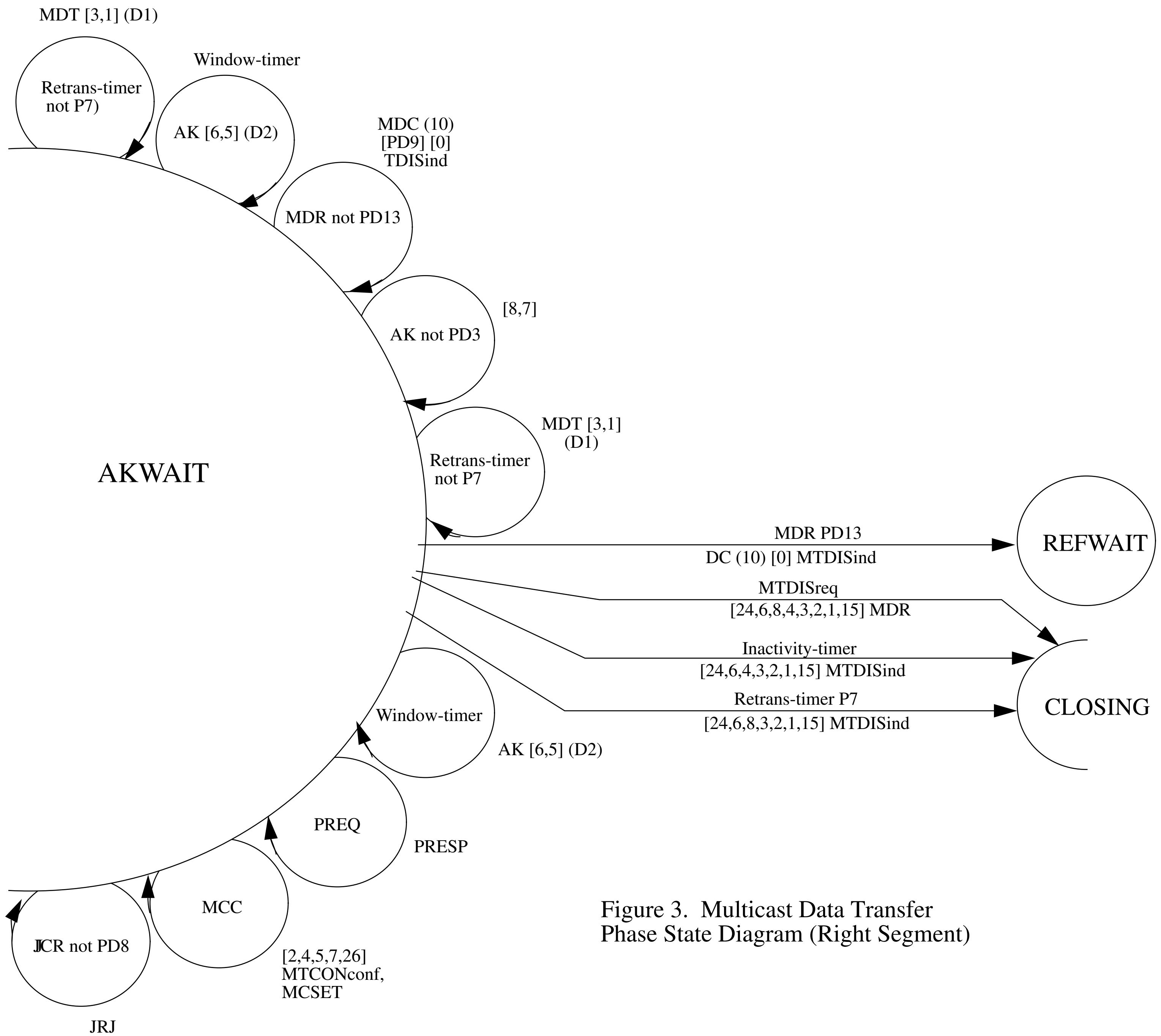

Figure 3. Multicast Data Transfer
Phase State Diagram (Right Segment)