Congestion Handling Paper 4-30-1998

This paper addresses congestion handling in NPAC SMS communications. The intent of the description doesn't assume any specific implementation of the hardware platform, OSI stack, and CMIP Tool kit. This paper has been written by the NANC T&O change management administrator in an attempt to document discussions to date and provide information that will lead to closure on all open congestion issues. This paper is a work in progress and is not meant to be included in the IIS or any other formal/official documentation. It is for the sole use of the NANC T&O group working on congestion.

The contents of this paper include:

- A description of the problem
- A description of key communication components where congestion can occur
- Typically causes of congestion in the communication components
- Recommendation for congestion control
- Open Issues

Problem Description

Congestion occurs when one system is attempting to send to another system and a resource limitation has been reached preventing the delivery of the request.

The following issues/questions have been raised during NANC T&O analysis of congestion:

1. How is congestion handled from the sender's standpoint?
2. How is congestion handled from the receiver's standpoint?
3. How is congestion recognized when the problem is network related?
4. What is the effect of congestion on the 3X2 timer?
5. What is the effect of congestion on recovery?
6. What is the invoke id of the next message after congestion has cleared?
7. How is thrashing prevented? Thrashing occurs when the congestion condition has been entered, exited, and re-entered within a short period of time.
8. How are linked-replies handled during a congestion condition?
9. Does everyone need to implement congestion?
10. What if an abort occurs during a congestion condition?

Communication Components

Many different components exist to enable communication between the NPAC and the SOA/LSMS. The section will focus on these components at a relatively high level. Communications between the NPAC and SOA/LSMS require the following components:

1. LNP Application (NPAC, SOA or LSMS functionality).
2. CMIP Tool kit
3. OSI Stack
4. Network Interactions and Physical Connectivity

The following descriptions have been intentionally kept at a high level.

LNP Application

This communications component is the entity that performs LNP processing. It executes the business rules associated with the respective LNP request, read/writes to the database, and communicates with the CMIP Tool Kit.

CMIP Tool Kit

This communication component implements the CMIP protocol. It provides an interface that is used by LNP applications to send/receive data. Upon receiving a message for delivery, the CMIP tool kit will perform whatever packaging it must perform and send the request to the OSI Stack. When the CMIP Tool Kit receives a message from the OSI Stack, the CMIP Tool kit either sends a notification to the LNP applications indicating data to available for reception or the LNP applications polls the CMIP Tool Kit.

OSI Stack

For the sake of this discussion, the OSI Communication Reference Model as defined by ISO is being broken into two communication components. The first communication component, ‘OSI Stack’, will be discussed here.

The OSI Stack accounts for the application layer through the transport layer as defined by OSI Communication Reference Model. The CMIP Toolkit sends and receives data to the OSI Stack. Upon receiving a request from the CMIP Toolkit the request to filtered down the protocol stack with each respective layer of the protocol stack adding it information to the message. Since the IIS implements RFC 1006, the OSI Stack sends/receives data to/from the network using TCP. Once the message gets down to the transport layer it is delivered using TCP.

When the OSI stack receives a request from the TCP driver, it is passed up the stack and forwarded on to the CMIP Tool Kit.

Network Connectivity

For the sake of this discussion, the second part of the OSI Communication Reference Model is being called “Network Connectivity”. This section accounts for the Network layer through the physical layers of the OSI Communication Model. Actual communications to the destination computers is accounted for by this communication component. This includes:

- The size the communication pipe being used (T1, 56KB, ISDN)
- The configuration of the network (number of gateways that may exist between the source and destination.
- The routing protocols used with the respective gateways.
- The different types of LANs that may exist between the source and destination.

Causes of Congestion

A variety of reasons can cause congestion. The following table takes into account each of the communication components described above. In addition the table addresses the causes of congestion that can occur with in each communication component. It also states how the respective communication components notify its ‘user’ of the congestion condition. Note, the term ‘user’ implies the systems that is requesting service from another systems (i.e. request the OSI stack to send a message.

Communication Component	**Cause of Congestion**	**Congestion Notification**
LNP Application	Processing High Volumes Processing a Large Request (Mass Update, Large Split, etc)	Depends on Vendor products. Typically a congestion notification would be generated either by the vendor software or by common

	System Not tuned for volumes (high CPU, memory thrashing, poor disk utilization, etc) Received a notification from the CMIP Took Kit that a resource threshold has been exceeded.	operating system utilities. The results would be sent to an NMS for proper assessment.
CMIP toolkit	Maximum buffer space has been exceeded. Maximum number of buffers has been used. Received a notification from the OSI Stack that a resource threshold has been exceeded.	Depends on how the Tool Kit vendor implemented this feature. DSET does have a mechanism to support this concept. They are providing a white paper describing how an LNP Application vendor can take advantage of this functionality.
OSI Stack	Maximum buffer space has been exceeded. Maximum number of buffers has been used. Peer flow control encountered based on the sliding window protocol being exceeded.	Based on the service definition defined for each layer of the OSI stack (application layer through transport), a mechanism exists to provide Inter-layer flow control.
Network Connectivity	A slow transmission pipe. Network configuration problem. Packet routing problems. I.e. Each packet being routed through an excessive number of gateways.	The remote side doesn’t acknowledge the receipt of the message.

Recommendations for Congestion Control

This section will first describe the recommendation and then the question raised in the problem statement will be addressed.

Congestion Control

The recommendation for Congestion Control follows the “Flow Control” mechanism described in OSI Communication Reference Model (ISO/IEC 7498). Two types of flow control are defined:

1. Peer Flow Control
2. Inter-Layer Flow Control

Peer Flow Control can be used when two peer layers of the OSI Stack talk to each either. The most common form of Peer Flow Control is the sliding window protocol. This protocol is implemented by TCP. The sliding window protocol prevents the sender from over-running the destination by placing a limit on the number of unacknowledged messages that can be outstanding at one time.

Inter-Layer Flow Control operates on the messages that are transmitted between each layer of the protocol stack. The lower layers of the OSI model will return an error to the caller when it has exceeded a resource limitation.

From the NPAC standpoint, these two mechanisms work as follows. When TCP encounters a condition where its sliding window is preventing the delivery of any new messages, TCP will STOP accepting any new messages. If a sender (in this case is the session layer) tries to send a message, TCP will return a resource limitation error. Eventually, the sender (again in this case the session layer) will reach a resource limitation because its queue of messages is too large and will have to return a resource limitation to its sender. This process proceeds up the stack until the CMIP Tool Kit is considered the sender.

When the CMIP Tool Kit receives a resource limitation error, it must either queue the message up or return the error to the LNP Application. If the message is queued-up, then a threshold must exist where the LNP Application will eventually be notified of the condition.

Once the LNP Application receives this error message it will temporarily suspend delivery of the message and try again at a later time. How the try again later algorithm is implemented is left up to the each vendor. The key point is that the message in NOT dropped and delivery is suspended. What this means is the LNP Application will be ultimate holding reservoir.

Addressing the Issues

1. How is congestion handled from the sender's standpoint?

   Section 'Congestion Control' describes how congestion is handled from a sender's point of view.

2. How is congestion handled from the receiver's standpoint?

   From the receiver's standpoint, no special processing is necessary to handle congestion. If the receiving LNP Application is unable to process the request as fast as they can come in then the CMIP Toolkit on the receiver will back and eventually stop accepting new messages from the OSI stack. Likewise, the OSI stack will backup. Eventually, TCP will be unable to acknowledge new request. This will cause the senders TCP to become flow controlled. If the conditional continues, the LNP application will ultimately become suspended.

3. How is congestion recognized when the problem is network related?

   Since TCP uses the sliding window protocol, if the network is congested then the sender won't be receiving any TCP-level acks from the destination.

4. What is the effect of congestion on the 3X2 timer?

   This depends on the implementation of the LNP Application. The NPAC SMS currently enforces the 3X2 timer during congestion situations.

5. What is the effect of congestion on recovery?

   If the link becomes congested during recovery, the NPAC will adhere to the paradigm it has in place for congestion.

6. What is the invoke id of the next message after congestion has cleared?

   The next sequential invoke id will be used.

7. How is thrashing prevented? Thrashing occurs when the congestion condition has been entered, exited, and re-entered within a short period of time.

   This is implementation dependent. The NPAC will take advantage of DSET high/low watermark mechanism. A Congestion condition is entered when the high water is achieved. The congestion condition is exited when the low water mark is achieved.

8. Handling of Linked-replies during a congestion condition

   Linked Replies will follow the same paradigm. A linked reply won't be sent until the congestion has been cleared. Once the condition has been cleared, the sender will pick up where they left off.

9. Does everyone need to implement congestion?

   Congestion control is optional by the local systems. However it is required by the NPAC.

10. What if an abort occurs during a congestion condition?

When a new association is created the LNP Application must initiate recovery as is done today.