Analysis of CCSDS File Delivery Protocol: Immediate NAK Mode*

Wonseok Baek and Daniel C. Lee

* This work was supported in part by the Jet Propulsion Laboratory, Pasadena, CA, under Contract 1214036.

Authors' addresses: W. Baek, University of Southern California, Department of Electrical Engineering, 3740 McClintock Avenue, Los Angeles, CA 90089-2565, USA, E-mail:wbaek@usc.edu; D. C. Lee, Wright State University, Department of Computer Science and Engineering, 3640 Colonel Glenn Hwy, Dayton, OH 45435-0001, USA, E-mail:dclee@cs.wright.edu.

Abstract

We analyze the Immediate NAK mode of the CCSDS File Delivery Protocol for the single-hop file transfer operation. We propose a timer setting rule that minimizes the expected time taken to transfer a file under the constraint that the throughput efficiency is maximized. Then, we derive the expected file delivery time and compare it with that of the Deferred NAK mode. The main contribution of this paper is the closed-form expression for evaluating the performance metrics.

Key words: communications, protocol, internetworking, throughput, delay

1. Introduction

In recent years, the Consultative Committee for Space Data Systems (CCSDS) has expended considerable effort toward providing flexible and efficient transfer of various data in a wide variety of mission configurations, from relatively low-earth-orbit spacecraft to complex arrangements of deep-space orbiters and landers supported by multiple transmission links. In many mission scenarios, space networking faces extremely long propagation delays, intermittent link connectivity, limited bandwidth, and limited power budgets. [1]-[3] In response to these factors and the need to automate the communication among spacecraft, the CCSDS File Delivery Protocol (CFDP) has been developed [4]–[9]. The aforementioned mission environments make the conventional Automatic Repeat reQuest (ARQ) schemes impractical [10]. (In an environment of long propagation delays, the *Stop-and-Wait* ARQ scheme has very low throughput [11]-[12]. Also, intermittent link connectivity and link occultation make it difficult to fix a window size of the sliding window ARQ scheme [13]-[17].) The most salient feature of ARQ schemes used in the CFDP, in comparison with conventional ARQ schemes, is that acknowledgement (ACK) is not issued for most protocol data units (PDUs). For these PDUs, only negative acknowledgements (NAKs) are issued in cases where the receiver perceives a PDU delivery anomaly. ACKs are only used for ancillary data PDUs such as End-of-File (EOF) and Finished (FIN) PDUs, which are used for closing the file transfer operation.

In CFDP, the file transfer is termed a “transaction,” and the sender assigns a transaction ID for each file transfer operation. The transaction ID, along with the source ID, etc., is contained in the header of each PDU. The sender informs the receiver of the start of the file transfer by transmitting the meta data PDU, which contains information such as the source ID, destination ID, file name, file size, etc. Like most PDUs in CFDP, there is no ACK for the meta data PDU, and the sender is allowed to transmit file data PDUs (PDUs carrying the actual content of the file) after transmitting the meta data PDU without confirming the meta data PDU's successful delivery (i.e., there is no handshaking for initiating a "transaction.") The receiver detects the failure in delivering a file data PDU or the meta data PDU by noticing missing PDUs in the sequence of PDUs correctly received (each file data PDU has a field that specifies the starting byte number and ending byte number of the file data carried by the PDU, so the receiver can detect missing PDUs by observing the ending byte numbers and the starting byte numbers of the correctly received PDUs.) The receiver reacts to missing PDUs by sending NAK messages, which contain the list of PDUs requested by the receiver for retransmission. Upon receiving a NAK, the sender re-transmits the PDUs requested. When the sender runs out of file data PDUs to send, the sender sends an EOF PDU, thus initiating the closure of the file transfer. After receiving the EOF PDU, the receiver acknowledges it with an ACK(EOF) and waits until all the file data PDUs and the expected meta data PDU of the transaction are received. (All data are eventually received due to the NAK mechanisms, and the receiver knows such completion from the file size information contained in the meta data PDU and the EOF PDU.) Then, the receiver sends a FIN PDU. After receiving the FIN PDU, the sender acknowledges it with an ACK(FIN) and closes the transaction. When the ACK(FIN) is successfully delivered back to the receiver, the receiver closes the transaction. As mentioned, there are ACK mechanisms for EOF and FIN PDU, so their exchange is reliable. According to CFDP, the receiver/sender must transmit an ACK message in response to each EOF/FIN PDU even after closing the transaction in order to prevent possible anomalies in closing the transaction (e.g., one described in [18].)

Depending upon mission requirements and transmission capability, four selectable ARQ schemes are offered by CFDP. The four ARQ schemes (Immediate NAK mode, Deferred NAK mode, Asynchronous NAK mode, and Prompt NAK mode) share a common mechanism for initiating and closing the file transfer operation but differ in their time of issuing NAK messages and their listing of PDUs requested for retransmission. In the Deferred NAK mode, the receiver defers issuance of NAKs until it successfully receives the EOF PDU from the sender. (Deferred NAK operation does not allow retransmission of a PDU prior to the delivery of EOF PDU. For more details of Deferred NAK mode, refer to [19].). In the Immediate NAK mode, the receiver generates NAKs for missing PDUs before receiving the EOF PDU unlike the Deferred NAK operation. In this paper, we focus on the analysis of the Immediate NAK mode.

The operation of the Immediate NAK mode is chronologically divided into two procedures: an "incremental lost segment detection procedure" followed by a "deferred lost segment detection procedure" [6]. (From now on, we will call the incremental lost segment detection procedure the "incremental procedure" and the deferred lost segment detection procedure the "deferred procedure.") The incremental procedure starts at the sender's transmission of the meta data PDU initiating the transaction and ends after EOF reception at the receiver's end. The incremental procedure ends at the end of ACK(EOF) transmission or at the end of NAK transmission immediately following ACK(EOF) transmission. If the last file data PDU is successfully received prior to EOF reception, the incremental procedure ends upon the transmission of ACK(EOF). If the reception of EOF finds the last file data PDU missing (possibly with other PDUs), then the incremental procedure ends upon the transmission of the NAK. The deferred procedure starts at the end of the incremental procedure and ends when the meta data PDU and all file data PDUs are successfully received by the receiver. The major difference between the operations of the incremental procedure and the deferred procedure is in their NAK mechanism.

During the incremental procedure of Immediate NAK mode, the receiver generates NAK whenever there is a missing PDU followed by a successful PDU. This NAK only requests the newly found missing PDU(s), and the NAK timer is not required during the incremental procedure. Upon receiving a NAK, the sender immediately retransmits all PDUs that the NAK requests. Note that certain PDUs might remain still missing even after the incremental procedure due to loss/error of NAK or loss/error of retransmitted PDUs. The receiver requests those missing PDUs during the deferred procedure. Immediately after EOF reception, the receiver acknowledges it with an ACK(EOF) (and transmits a NAK if necessary) and starts a NAK timer if there still exist any missing PDUs. Then, the deferred procedure starts.

Upon expiration of the NAK timer, the receiver examines the record of missing PDUs. If missing PDUs exist, the receiver requests them all and sets a NAK timer upon transmission of NAK. Again, upon expiration of the NAK timer, the receiver examines the record of missing PDUs. If a missing PDU still remains, the receiver generates another NAK and starts a NAK timer. This process continues until the receiver receives all necessary PDUs, which constitute the entire file content and the meta data PDU. Note that the receiver can request retransmission of a missing PDU before receiving EOF. However, the receiver does not request retransmission of the same PDU more than once before receiving the EOF. Also, note the peculiarity that according to [6] the receiver always starts the NAK timer upon EOF reception if there is still a missing PDU—that is, even if it does not generate a new NAK in response to EOF reception. (Note that a NAK is generated in response to EOF reception if and only if the first transmission of the very last file data PDU is unsuccessful.) After receiving all necessary PDUs, the receiver generates a FIN PDU. Upon receiving the FIN PDU, the sender acknowledges it with an ACK(FIN) and closes the transaction. The delivery of the FIN PDU is guaranteed in the same way (through acknowledgement) as the EOF PDU. The receiver closes the transaction when the ACK(FIN) is successfully received. See Figure 1 for an illustration of the operation of the Immediate NAK mode.

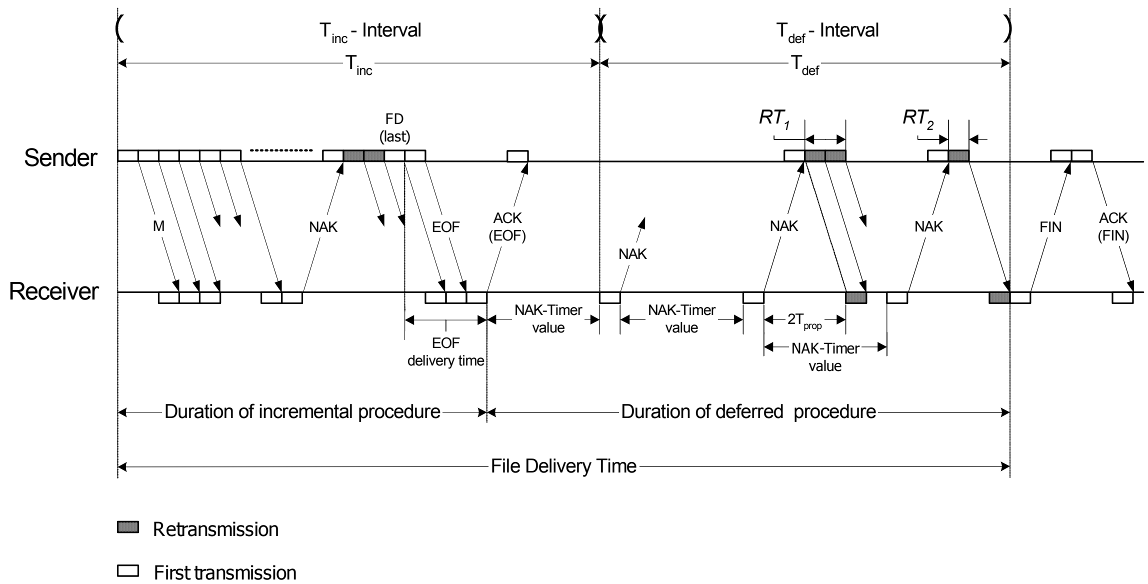

Figure 1. Immediate NAK mode.

M stands for the meta data PDU, and FD(k) stands for the k^{th} file data PDU. The protocol specification in [6] defines the meaning and format of these PDUs in detail. T_{prop} stands for one-way propagation delay, and RT_k stands for the duration of the k^{th} retransmission spurt in T_{def}. ("Transmission spurt" refers to consecutive transmissions of PDUs back to back.) T_{inc}-interval, T_{inc}, T_{def} -interval, T_{def} , and file delivery time are defined in Section 2.

In this paper, we present modeling and analysis of the CFDP Immediate NAK mode. We consider the single-hop file transfer operation. With regard to performance measures, this paper is mainly concerned with the time taken to transfer a file (expected file delivery time) and throughput efficiency resulting from the protocol specification. We define throughput efficiency as the average ratio of the total information data amount delivered to the amount of data transmitted. The more likely the retransmissions of the same PDU are, the less the throughput efficiency is. This definition of throughput efficiency is directly related to transmission power efficiency, which is extremely important in space communication.

The two different NAK mechanisms of the Immediate NAK mode make the mathematical derivation of the expected file delivery time quite challenging, and we will use an elaborately designed strategy of derivation to meet this challenge. For example, breaking the file delivery time into T_{inc} and T_{def} is an artifice designed to facilitate the derivation. In Section 2, we precisely define the file delivery time and break it into T_{inc} and T_{def} . We also introduce some assumptions for the simplicity of analysis. In Section 3, we address the problem of timer setting and propose our timer-setting rule under the constraint that throughput efficiency is maximized. In Section 4, we derive upper and lower bounds of the expression for the expected file delivery time in Immediate NAK mode for our timer-setting rule. In Section 5, on the basis of our derivation, we present the numerical results and a comparison of performance with the Deferred NAK mode.

2. Preliminaries

Because much of our analysis concerns the file delivery time of a transaction, we first define the "file delivery time" precisely as the time from the beginning of the transmission (the first bit of the meta data PDU) until the first instant when all file data, meta data, and the EOF PDU have been successfully received by the receiver. Similarly, "EOF delivery time" is defined as the time from the beginning of transmission of the EOF PDU until the first instant when the EOF PDU has been successfully received and an ACK(EOF) has been transmitted by the receiver.

For the convenience of mathematical derivation, we first define T_{inc}**-interval** to be the interval from the beginning of the transmission (the first bit of the meta data PDU) until the NAK timer set upon the first error-free EOF reception is expired, as illustrated in Figure 1. We denote by T_{inc} the length of the T_{inc}**-interval**. Thus, in Immediate NAK mode T_{inc} is the duration of the incremental procedure plus the time-out period of the NAK timer set on the first error-free EOF reception. The T_{def}**-interval** is defined as the interval from the issuance of the first NAK that requests all missing PDUs until the first moment when all file data, meta data, and the EOF PDU have been successfully received by the receiver, as illustrated in Figure 1. We denote by T_{def} the length of the T_{def}**-interval**. (If there is no NAK that requests all missing PDUs, then $T_{def} = 0$.) Thus, T_{def} consists of the duration of the retransmission spurts and time gaps between retransmission spurts. ("Transmission spurt" refers to consecutive transmissions of PDUs back to back. See Figure 1 for an illustration.) From the next sections, we derive the expected file delivery time for the Immediate NAK mode of CFDP by examining T_{inc} and T_{def}, respectively. Note that T_{inc} (T_{def}) does not exactly coincide with the duration of the "incremental loss segment detection procedure" ("deferred loss segment detection procedure").

For simplicity, we make the following assumptions.

- The meta data PDU and all file data PDUs have identical lengths, identical transmission times, and identical probabilities of failed delivery (PDU error or loss). The meta data

PDU is usually shorter than a data PDU and thus has lower probability of failed delivery. However, the length of the meta data is so small in comparison with the length of the file data that the assumption of equal lengths has relatively a negligible effect on the total file delivery time.

- All NAKs have identical lengths and identical probabilities of failed delivery. (The length of a NAK depends on the number of PDUs that it requests. However, the differences are small, and the lengths of NAKs are all small, so this assumption should not significantly affect the performance measure.)
- PDU error events in forward and backward links are statistically independent.
- Compared with the file data PDU, the size of NAK, ACK, and EOF PDUs is much smaller, so we ignore the transmission time of these PDUs.

Note the table below for the notations used in the analysis.

Symbol	**Definition**
P_{ef}	Probability of PDU error in forward link
$P_{ef(EOF)}$	Probability of error in delivering EOF PDU
P_{er}	Probability of error in delivering NAK
T_{prop}	One-way propagation delay
T_{PDU}	Transmission time of meta data or file data PDU

For performance analysis, we also have to specify the action of the sender at the instance of expiration of the EOF timer if a PDU to be retransmitted is waiting in the queue. Note that at the instance of expiration of the EOF timer, the sender may be retransmitting a PDU, and another PDU to be retransmitted may be waiting. We can consider two implementations of the sender's action in such a case.[1] First, after finishing the transmission of the rest of the PDU being transmitted, the sender could pause retransmission of the PDUs that have been requested for retransmission. The purpose of this pause is "rapid response to the expiration" of the EOF timer. While pausing the retransmission of PDUs previously requested, the sender sends the EOF PDU again. Then, the sender resumes retransmission of the PDUs requested. (EOF retransmission is inserted between the retransmission of file data PDUs previously requested.) Second, the sender could complete retransmission of all PDUs previously requested before retransmitting the EOF

[1] The protocol specification does not address this case, and the two implementations presented here are not found in existing literature.

PDU in response to the EOF timer expiration. In this paper, we assume the former implementation of CFDP.[2]

3. Proposal for timer implementation

Before analysis, we address the problem of timer setting. As can be deduced from the protocol description in the Introduction (illustrated in Figure 1), the expected file delivery time depends upon certain parameter values that an implementer can freely choose—for example, the time-out value of the EOF timer and the time-out value of the NAK timer. In this paper, we assume that the parameter values are set to minimize the expected file delivery time under the constraint that the throughput efficiency is maximized. Note that there is a trade-off between the throughput efficiency and file delivery time. For example, if the sender retransmits PDUs requested for retransmission multiple times in a row, the expected number of retransmission spurts will decrease, and so will the expected file delivery time. However, such a practice will result in decrease of the throughput efficiency and thus require more power consumption. In this paper, we place high priority on the throughput efficiency because power is extremely limited in space networks. In this section, we address the timer setting problem related to throughput efficiency and propose a timer setting rule that does not compromise throughput efficiency while minimizing the file delivery time.

In an environment of long propagation delays, throughput efficiency can be compromised by unnecessary duplicate retransmission of identical PDUs. For example, if the time-out value of the EOF timer, which we refer to as time-out-EOF in this paper, is set too small, the sender retransmits EOF PDU before receiving ACK(EOF) due to timer expiration, even in the case that the first EOF PDU and ACK(EOF) are successfully delivered. See Figure 2 for an illustration. Unnecessary duplicate retransmission of file data PDUs can occur if the time-out value of the NAK timer in the T_{def} **-interval** is set too small, as illustrated in Figure 3. For such parameter values, we suggest the following and use them for our mathematical derivation:

- "*time-out-EOF*," denoting the time-out value of the EOF timer, should be $2T_{prop}$. (In a real implementation, the value of $2T_{prop}$ should be estimated or computed by the sender. Also, the time-out value should also include node processing delay, the transmission time of ACK(EOF), the delay prior to transmitting ACK(EOF) if a NAK transmission is being completed at the instance of EOF arrival, etc. As mentioned in section 2, we assume that these

[2] A different implementation of the sender's action can result in a different expected file delivery time.

quantities are negligible. The actual implementation can add slack to a time-out value for these quantities and estimation error.)

- $T^{k}_{timer(NAK)}$, denoting the time-out value of the NAK timer set upon issuance of the NAK that causes the k^{th} retransmission spurt in the T_{def} **-interval**, should be $2T_{prop} + RT_k$, where RT_k denotes the transmission time of the PDUs requested by the receiver for the k^{th} retransmission spurt in the T_{def} **-interval**.

It can be intuitively argued that this timer setting does not compromise the throughput efficiency while minimizing the file delivery time.

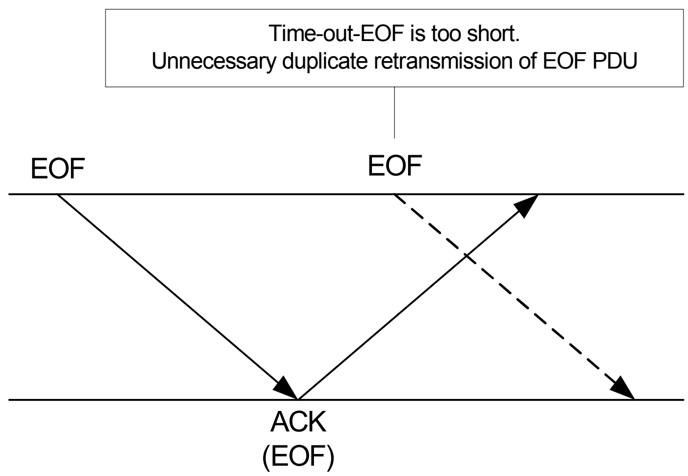

Figure 2. Too-small value of EOF timer.

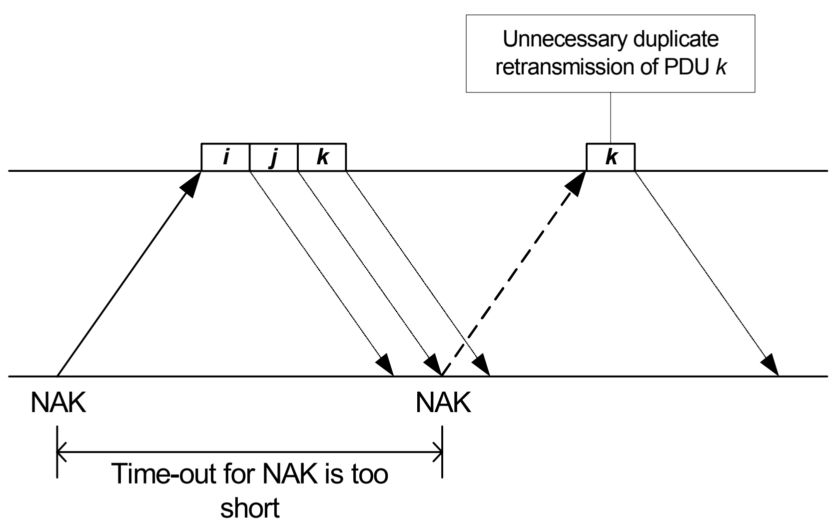

Figure 3. Effect of too-small value of the NAK timer during the T_{def}-interval.

Finally, special attention should be paid to the value of the NAK timer that the receiver sets in response to receiving the EOF. Let us recall that in the Immediate NAK mode a NAK timer is set regardless of whether or not a NAK is generated in response to receiving the EOF. We will refer to this special NAK timer as the "NAK_{EOF} timer." Without proper setting of the time-out value of the NAK_{EOF} timer, duplicate retransmissions may occur. In this section, we discuss the value to be set for the NAK_{EOF} timer. Let us illustrate two cases of a duplicate retransmission that is caused by improper setting of the time-out value of the NAK_{EOF} timer. First, we consider the case in which the first transmission of PDU *N* (the last file data PDU in the file) from the sender is unsuccessful. In this scenario, the reception of the EOF by the receiver spawns an actual issuance of a NAK message, and the NAK_{EOF} timer associated with the NAK message is set. We note that this NAK message requests retransmission of PDU *N* and possibly retransmission of other PDUs newly found missing, according to the protocol specification [6]. If the NAK_{EOF} timer were set, like other NAK timers set during the T_{def} **-interval**, to be two times the propagation delay plus the transmission time of PDUs requested by the NAK, then duplicate retransmissions could occur. Figure 4 illustrates such duplicate retransmissions.

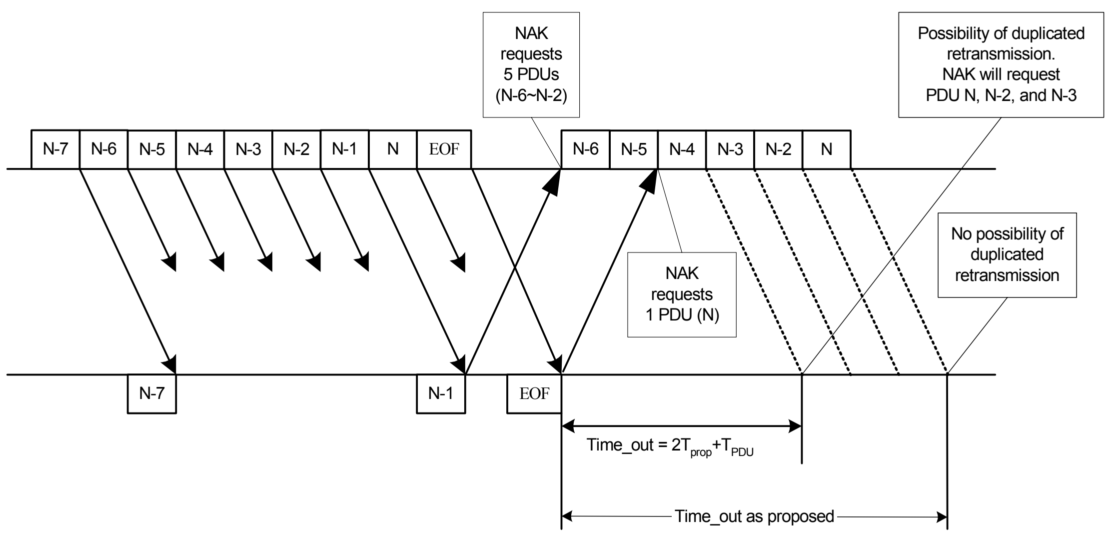

Figure 4. Timer setting problem for the NAK generated on successful reception of the EOF PDU.

Second, we consider the case in which the first transmission of PDU *N* by the sender is successful; that is, the receiver receives PDU *N* prior to receiving the EOF. In this case, the reception of EOF does not spawn a NAK generation. However, if there is still a missing PDU at the time of EOF reception, the NAK_{EOF} timer is set in accordance with the specification [6]. If the NAK_{EOF} timer is set at too small a value, there could again be an unnecessary duplicate retransmission.

In determining the time-out value of the NAK_{EOF} timer to be set upon EOF reception, there is a trade-off between the possibility of unnecessary duplicate retransmission and the delay in the file delivery time. Therefore, we propose a rule that chooses the time-out value of the NAK_{EOF} timer to be as small as possible while guaranteeing that there will be no unnecessary duplicate retransmission.

3.1. Rule to determine the NAK_{EOF} timer value

We propose that a timer be set upon every generation of a NAK, including even the NAKs generated prior to the reception of EOF (in the incremental procedure).[3] However, before successful EOF reception, the receiver does nothing at the instant of NAK timer expiration. The purpose of the receiver's setting of NAK timers prior to receiving EOF is to have an account of the system's pending retransmission activities at the instant of EOF reception. From now on, we refer to those NAK timers used before successful EOF reception as *NAK timer-inc* for clarity of description. The following conceptual mechanism describes the rules determining the value of each *NAK timer-inc*:

NAK timer-inc setting rules:

1) At the generation of the very first NAK, the time-out value of the *NAK timer-inc* is set to be two times the propagation delay plus transmission time of the requested PDUs.
2) At the generation of subsequent NAKs, the following rules are applied:
    - If there is no active *NAK timer-inc*, (remaining time-out = 0), the new *NAK timer-inc* value is again set to be two times the propagation delay plus transmission time of the requested PDUs.
    - If there is an active *NAK timer-inc* (remaining time-out > 0), the receiver compares the remaining time-out value of this active timer with two times the propagation delay.
        - If the remaining time-out value of the active timer is greater than two times the propagation delay, the receiver starts a new *NAK timer-inc* and sets its value to be the remaining time-out value of the active timer plus transmission time of the requested PDUs. The receiver stops the old timer.

[3] Even though a timer is set upon every generation of a NAK, the number of timers that the receiver maintains at the same time is at most two. More details are provided in the *NAK timer-inc* setting rule in this section. Also, a good tuning of system parameters should keep the packet error rate small, so the number of NAKs and the the *NAK timer-inc* setting events should be small.

- If the remaining time-out value of the active timer is less than two times the propagation delay, the receiver starts a new *NAK timer-inc* and sets its time-out value to be two times the propagation delay plus the transmission time of requested PDUs. The receiver stops the old timer. ■■

Upon the successful reception of EOF, the receiver takes one of the three actions stated below:

a) If reception of EOF spawns a NAK, the NAK_{EOF} timer is set as if it were a *NAK timer-inc*, *in* accordance with the *NAK timer-inc* setting rules (1) and (2) above.
b) If the reception of EOF does not spawn a NAK and there is an active *NAK timer-inc*, the receiver starts a NAK_{EOF} timer the time-out value of which is the remaining time-out value of the active *NAK timer-inc*, and the receiver stops the *NAK timer-inc*.
c) If the reception of EOF does not spawn a NAK and there is no active *NAK timer-inc*, NAK_{EOF} timer is set to be zero. If there are still missing PDUs at this time, the receiver starts deferred procedure by issuing a NAK that requests all missing PDUs. Otherwise, this implies that meta data, all file data PDUs, and EOF have been received successfully. The receiver transmits FIN PDU.

Our proposal for the NAK_{EOF} timer-setting rule, which relies on the *NAK timer-inc* setting rule, eliminates the ambiguity in how to determine the time-out value of the NAK_{EOF} timer. (The specification [6] gives freedom on this value to implementers.) As previously explained, NAK_{EOF} timer is set upon the first successful reception of EOF. However, an improper setting of the NAK_{EOF} timer value can cause unnecessary duplicate retransmissions. (Refer to Figure 4.) Without proper consideration of potential retransmission activities at the sender, it is hard to determine the smallest time-out value of the NAK_{EOF} timer that prevents any unnecessary duplicate retransmission. Our NAK_{EOF} timer setting rule uses *NAK timer-inc* so that the receiver, upon receiving an EOF, may have accurate estimation of the sender's potential retransmission activities. Section 3.2 rigorously shows how the proposed NAK_{EOF} timer-setting rule minimizes the NAK_{EOF} timer while preventing unnecessary duplicate retransmissions. From now on, we assume implementation of our timer-setting rules and the "rapid response to the expiration of EOF timer" that was explained in Section 2.

3.2. Mathematical description of timer-setting rules

Let us denote by R_n^{inc} the time-out value of the *n*th *NAK timer-inc* set during the T_{inc}-**interval**. According to the timer-setting rule, the time-out value of any *NAK timer-inc* is greater than two times the propagation delay. That is,

$$R_n^{inc} = 2T_{prop} + W_n^{inc} \ , \tag{1}$$

where $W_n^{inc} > 0$. To define W_n^{inc} more rigorously, we define two more random variables as follows:

- N_n^{inc} is the number of PDUs requested by the *n*th NAK in the T_{inc}-**interval**.
- T_n is the inter-issuance time between the *n*th NAK and the *n*-1th NAK in the T_{inc}-**interval**, where $T_1 \equiv 0$.

For the first *NAK timer-inc*, we have

$$W_1^{inc} = N_1^{inc} \cdot T_{PDU} \ , \tag{2}$$

so

$$R_1^{inc} = 2T_{prop} + W_1^{inc} = 2T_{prop} + N_1^{inc} \cdot T_{PDU} \ . \tag{3}$$

In other words, according to rule (1) in Section 3.1, the time required in addition to $2T_{prop}$ in the time-out value of the 1*st* *NAK timer-inc* is the time as that required to retransmit all the PDUs requested by the 1st NAK for retransmission. For the *n*th NAK in the T_{inc}-**interval** $(n \geq 2)$, we have

$$R_n^{inc} = \begin{cases} R_{n-1}^{inc} - T_n + N_n^{inc} \cdot T_{PDU} & \text{if } R_{n-1}^{inc} - T_n \geq 2T_{prop} \\ 2T_{prop} + N_n^{inc} \cdot T_{PDU} & \text{if } R_{n-1}^{inc} - T_n < 2T_{prop} \end{cases} . \tag{4}$$

This relation mathematically describes rule (2) in Section 3.1. (Note that $R_{n-1}^{inc} - T_n$ is the remaining time-out value at the issuance of the *n*th NAK.) From (1), we can re-write (4) into

$$R_n^{inc} = \begin{cases} 2T_{prop} + W_{n-1}^{inc} - T_n + N_n^{inc} \cdot T_{PDU} & \text{if } W_{n-1}^{inc} \geq T_n \\ 2T_{prop} + N_n^{inc} \cdot T_{PDU} & \text{if } W_{n-1}^{inc} < T_n \end{cases} . \tag{5}$$

Therefore, the *n*th *NAK timer-inc* in (1) can be described by recursion of W_n^{inc}:

$$W_n^{inc} = \begin{cases} W_{n-1}^{inc} + N_n^{inc} \cdot T_{PDU} - T_n & \text{if } W_{n-1}^{inc} \geq T_n \\ N_n^{inc} \cdot T_{PDU} & \text{if } W_{n-1}^{inc} < T_n \end{cases} . \tag{6}$$

Now, we interpret W_n^{inc}, which is given to the *n*th *NAK timer-inc* in addition to two times the propagation delay in order to avoid unnecessary duplicate retransmissions: W_n^{inc} is defined based on the receiver's estimation, under the hypothetical assumption that NAK delivery never fails, of the number of backlogged PDUs that the sender has at the time of its reception of the n^{th} NAK. The physical meaning of W_n^{inc} is the duration the sender needs from its nominal reception time of the n^{th} NAK issued by the receiver (the time at which the n^{th} NAK is issued by the receiver plus the propagation delay) to finish retransmitting all of the backlogged PDUs at that time (assuming no failure in NAK delivery) plus the duration required to retransmit all the PDUs requested in the n^{th} NAK for retransmission. See Figure 5 for an illustration of W_n^{inc}. In Figure 5 (a), when the receiver issues the n^{th} NAK it estimates the number of backlogged PDUs as 3 at the time of arrival of the n^{th} NAK. It is correct because the $(n\text{-}1)^{th}$ NAK is successfully delivered to the sender. However, the estimation would be incorrect if the $(n\text{-}1)^{th}$ NAK fails to be delivered to the sender. Figure 5 (b) illustrates such a case; when the receiver issues the n^{th} NAK it estimates the number of backlogged PDUs as 3 at the time of arrival of the n^{th} NAK, but there are no backlogged PDUs at the sender at the time of arrival of the n^{th} NAK because the $(n\text{-}1)^{th}$ NAK fails to be delivered to the sender. As a result, the time-out value of the n^{th} *NAK timer-inc* might be set to be longer than what it is actually required.

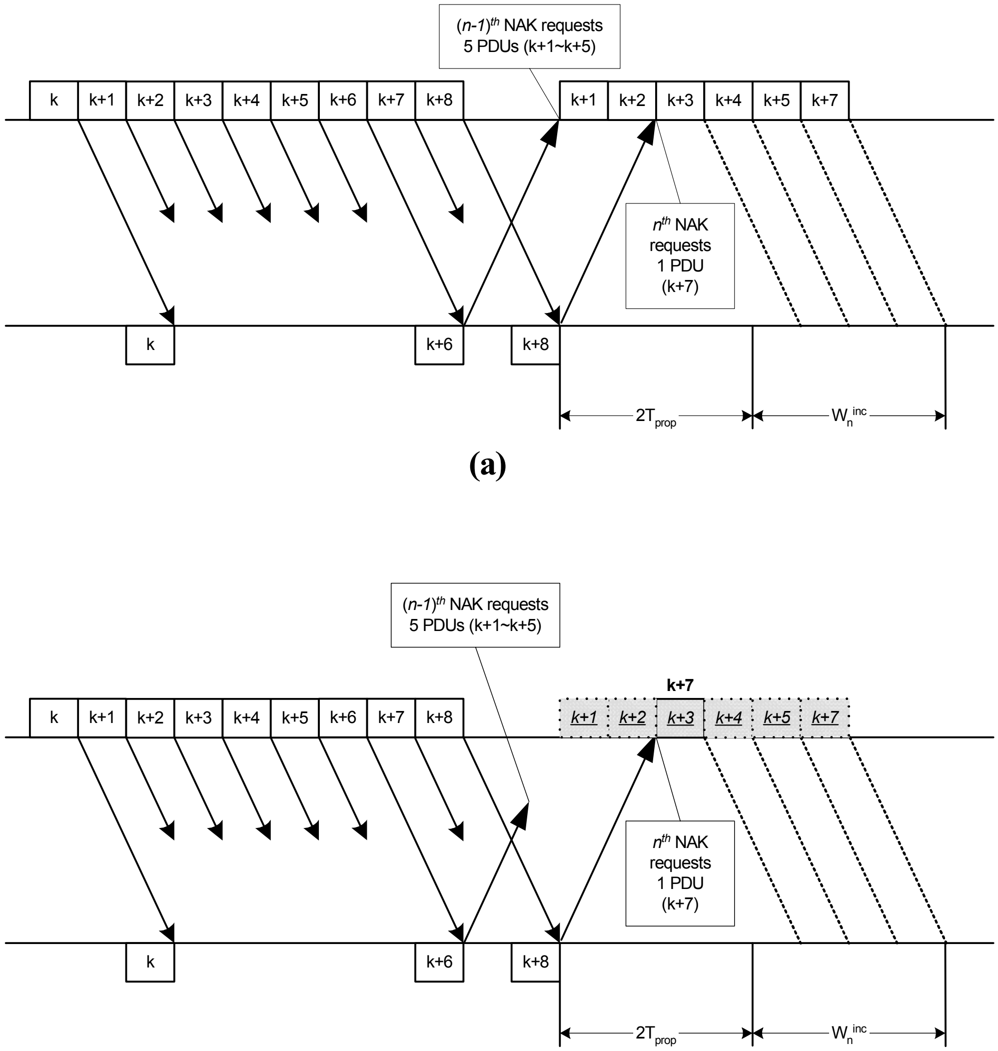

Figure 5. Illustration of W_n^{inc}.

Let us denote by *R* the time-out value set for NAK_{EOF}. Let *L* be the number of NAKs issued in the T_{inc}**-interval**. We also define *A* to be the time from the issuance of the last (L^{th}) NAK that is generated by the receiver in the T_{inc}**-interval** until the reception of *EOF*. (See Figure 7 for illustration of *A*.) Then, we can write:

$$R = \left[R_L^{inc} - A \right]^+ \quad . \tag{7}$$

where R_n^{inc} is the time-out value of the n^{th} *NAK timer-inc* set during the T_{inc}**-interval** (1). In the event that the reception of EOF spawns a NAK, this NAK is the last NAK in the T_{inc}**-interval** and is, by the definition of *L*, the L^{th} NAK in the T_{inc}**-interval,** so we set *A* at zero. In such an event, (7) is reduced to:

$$R = R_L^{inc} = 2T_{prop} + W_L^{inc} \ . \tag{8}$$

4. Analysis of Immediate NAK mode

As described in Section 2, the file delivery time of CFDP in Immediate NAK mode can be divided into two parts: T_{inc} and T_{def}. In the Immediate NAK operation, the receiver can request retransmission of a missing PDU before receiving EOF. However, the receiver does not request retransmission for the same PDU more than once before receiving the EOF. Thus, the Immediate NAK mode allows one retransmission for each PDU during the T_{inc}**-interval**. Taking this property into account, we examine and derive the expected value of T_{def} and T_{inc}. We denote by PDU 1, PDU 2,…, PDU *N* the consecutive PDUs in the original order of occurrence in the file to be delivered to the receiver.

4.1 Bounds on the expected value of T*def*

Recall that the time-out value of NAK timer used in T_{def} **-interval** for the *k*th retransmission spurt is set to be $T_{timer(NAK)}^{k} = 2T_{prop} + RT_k$, where RT_k denotes the duration of *k*th retransmission spurt. We denote by H_N the number of retransmission spurts for *N* PDUs during the T_{def} **-interval** in the Immediate NAK mode. Then, taking into account the case that NAK may be lost (and thus the NAK timer expires), the expected value of T_{def} is obtained as

$$\begin{aligned} & E\left(T_{def}\right) \\ & = E\left\{\frac{2T_{prop} + RT_1}{1-P_{er}} + \frac{2T_{prop} + RT_2}{1-P_{er}} + \cdots + \frac{2T_{prop} + RT_{H_N}}{1-P_{er}}\right\} \ . \\ & = \frac{E(H_N)2T_{prop}}{1-P_{er}} + \frac{E\left(RT_1 + RT_2 + \cdots + RT_{H_N}\right)}{1-P_{er}} \end{aligned} \tag{9}$$

In(9), $RT_1 + RT_2 + ... + RT_{H_N}$ is the time that the sender's transmitter spends in retransmitting the PDUs in the T_{def} **-interval.** $E\left(RT_1 + RT_2 + ... + RT_{H_N}\right)$ is given as the following:

Proposition 1

$$E\left(RT_1 + RT_2 + ... + RT_{H_N}\right) = \frac{N \cdot T_{PDU} \cdot P_{ef}}{1-P_{ef}} - N \cdot P_{ef}\left(1-P_{er}\right)T_{PDU} \ . \tag{10}$$

Proof: See Appendix A. ■

From (9) and Proposition 1, we have

$$E\left(T_{def}\right) = \frac{E(H_N)2T_{prop}}{1-P_{er}} + \frac{N \cdot T_{PDU} \cdot P_{ef}}{\left(1-P_{ef}\right)\left(1-P_{er}\right)} - N \cdot P_{ef} \cdot T_{PDU} \ . \tag{11}$$

For $E(H_N)$, we have the following upper and lower bounds, which are derived in Appendix B.

Proposition 2

$$E(M_N)-1-\frac{1-\left(1-P_{ef}\right)^N}{1+P_{er}} \le E(H_N) \le E(M_N)-1-\left[1-\left(1-P_{ef}\right)^N\right](1-P_{er})^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^N}}, \quad (12)$$

where the random variable M_N is defined as $\max(K_1, K_2, \cdots K_N)$ and random variable K_i represents the number of transmissions of the i^{th} PDU up to and including its first successful reception [19]. ■

We note that $K_1, K_2, \cdots K_N$ are I.I.D. and geometrically distributed random variables. In fact, we have [19]

$$E(M_N)=\sum_{m=1}^{\infty} P(M_N \ge m)=1+\sum_{m=1}^{\infty}\left[1-\left(1-P_{ef}{}^m\right)^N\right].$$

Then, $E(M_N)$ can be expressed as a finite summation as follows [19]:

$$E(M_N)=1+\sum_{k=1}^{N}\binom{N}{k}\frac{P_{ef}{}^k}{1-P_{ef}{}^k}(-1)^{k+1}.$$

Thus, in theory we can compute the exact value of $E(M_N)$ in a finite number of computational operations. However, we face difficulties in numerical evaluation for a large value of *N*. As a result, we can use finite summation $1+\sum_{m=1}^{s^*}\left[1-\left(1-P_{ef}{}^m\right)^N\right]$ as both an approximation and a lower bound. As we increase the number of addition *s**, the evaluation becomes more accurate. For the numerical inaccuracy (the remainder, $R_{s^*} \equiv \sum_{m=s^*+1}^{\infty}\left[1-\left(1-P_{ef}{}^m\right)^N\right]$), we have the following upper bound, which is derived in [19].

Proposition 3

$$\begin{aligned} R_{s^*} < \; & \frac{2P_{ef}{}^{s^*+1}}{1-P_{ef}}-\frac{P_{ef}{}^{2(s^*+1)}}{1-P_{ef}^2}+\sum_{n=1}^{N-2}\left[\frac{-\ln\left(1+nP_{ef}{}^{s^*+1}\right)}{n\ln P_{ef}}+\frac{P_{ef}{}^{s^*+1}}{1+nP_{ef}{}^{s^*+1}}\right] \\ & -\sum_{n=1}^{N-2}\left[\frac{-P_{ef}{}^{s^*+1}}{n\ln P_{ef}}+\frac{\ln\left(1+nP_{ef}{}^{s^*+1}\right)}{n^2\ln P_{ef}}\right] \end{aligned}.$$

■

For a desired level of accuracy, we can use Proposition 3 to determine an appropriate value of s^* for computing $E(M_N)$.

From (11) and Proposition 2, the upper bound and lower bound on $E(T_{def})$ are obtained, respectively, as

$$E(T_{def}) \le [E(M_N)-1]\left[\frac{2T_{prop}}{1-P_{er}}\right] - \left[1-(1-P_{ef})^N\right](1-P_{er})^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^N}}\left[\frac{2T_{prop}}{1-P_{er}}\right] + \frac{N\cdot T_{PDU}\cdot P_{ef}}{(1-P_{ef})(1-P_{er})} - N\cdot P_{ef}\cdot T_{PDU} \tag{13}$$

and

$$E(T_{def}) \ge [E(M_N)-1]\left[\frac{2T_{prop}}{1-P_{er}}\right] - \left[\frac{1-(1-P_{ef})^N}{1+P_{er}}\right]\left[\frac{2T_{prop}}{1-P_{er}}\right] + \frac{N\cdot T_{PDU}\cdot P_{ef}}{(1-P_{ef})(1-P_{er})} - N\cdot P_{ef}\cdot T_{PDU}. \tag{14}$$

4.2 Bounds on the expected value of T_{inc}

According to our definition of T_{inc} in the Immediate NAK mode, T_{inc} is the duration of the incremental procedure plus the time-out period of the NAK_{EOF} timer set upon the first error-free EOF reception (see Figure 1). In this section, we focus on $E(T_{inc})$. A major idea to utilize in deriving $E(T_{inc})$ is that the Immediate NAK mode allows no more than one retransmission of a PDU during the T_{inc}**-interval**. It turns out that the form of $E(T_{inc})$ depends upon whether $k \equiv 2T_{prop}/T_{PDU}$ is larger than *N*-2, where *N* is the number of PDUs in the file, including the meta data. Case $k+2>N$ applies to the configuration in which the propagation delay is long relative to the time taken to transmit the file. The other case, $k+2\le N$, applies to the configuration with short propagation delay relative to the transmission time of the entire file. The expected value of T_{inc} is given as

$$E(T_{inc}) = \begin{cases} N\cdot T_{PDU} + E(X) + E(\text{EOF delivery time}) + E(R) & \text{if } N \ge k+2 \\ N\cdot T_{PDU} + E(\text{EOF delivery time}) + E(R) & \text{if } N < k+2 \end{cases}, \tag{15}$$

where the random variables *X* and *R* are defined as follows:

1. Random variable *R* denotes the time-out value of the NAK_{EOF} timer, as defined in section 3.2.
2. Random variable *X* denotes the time spent by the sender in retransmitting PDUs before the first EOF transmission.

Figure 6 and Figure 7 illustrate (15).

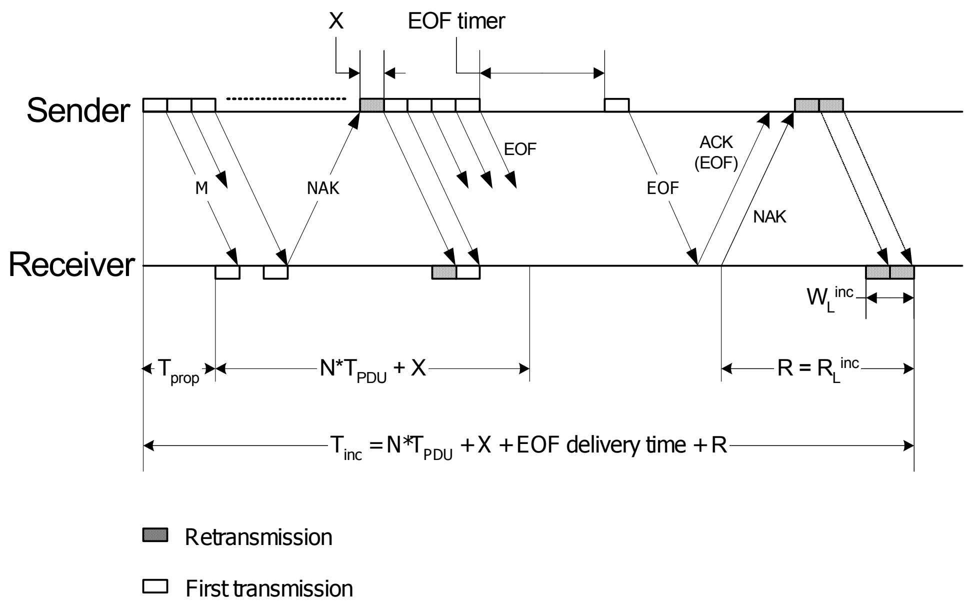

Figure 6. Incremental procedure of Immediate NAK mode $(N \geq k+2)$ **.**

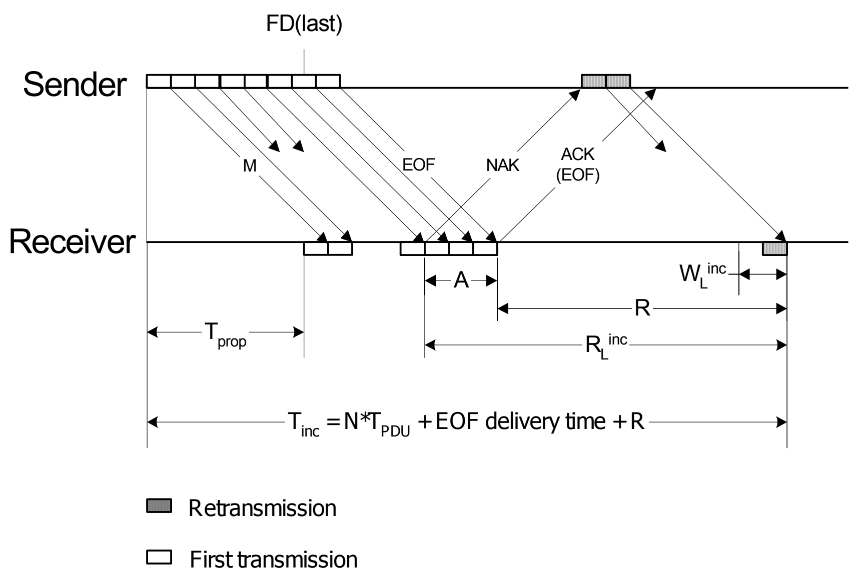

Figure 7. Incremental procedure of Immediate NAK mode $(N < k+2)$**.**

The expected EOF delivery time in (15) is obtained as follows. Taking into account the fact that the EOF may be lost/errored (and thus the EOF timer expires) and that time-out-EOF is set to $2T_{prop}$, the number of EOF transmissions up to and including the first successful delivery, which we denote by G_{EOF}, is the geometrically distributed random variable. Then, the EOF delivery time can be expressed as

$$(G_{EOF} - 1) \cdot \text{time-out-EOF} + T_{prop}.$$

Thus, its expected value is

$$\begin{aligned} & E[(G_{EOF}-1)] \cdot \text{time-out-EOF} + T_{prop} \\ & = \frac{P_{ef(EOF)}}{1-P_{ef(EOF)}} \cdot \text{time-out-EOF} + T_{prop} \end{aligned}.$$

For $\text{time-out-EOF} = 2T_{prop}$, the expected EOF delivery time is

$$\frac{P_{ef(EOF)} \cdot 2T_{prop}}{1-P_{ef(EOF)}} + T_{prop} = \frac{\left(1+P_{ef(EOF)}\right)T_{prop}}{1-P_{ef(EOF)}}.$$

Thus, we have

$$E(\text{EOF delivery time}) = \frac{\left(1+P_{ef(EOF)}\right)T_{prop}}{1-P_{ef(EOF)}} \quad . \tag{16}$$

The exact expression of $E(X)$ and $E(R)$ in (15) seems difficult to derive, so we will obtain their bounds. In doing so, we consider bounds on their conditional expected values conditioned on three mutually exclusive and collectively exhaustive events and take their properly weighted sum. (For the representation of R, the time-out value of the NAK_{EOF} timer, let us recall the definition of W_n^{inc} and A in Section 3.2.)

Event 1: The first EOF reception spawns issuance of a NAK. For example, if the first transmission of PDU N fails to be delivered to the receiver, the receiver notices the failed delivery of PDU N upon receiving EOF and issues a NAK:

$$R = R_L^{inc} = 2T_{prop} + W_L^{inc}.$$

Note that R_n^{inc} is the time-out value of the n^{th} *NAK timer-inc* set during the T_{inc}**-interval** and L is the number of NAKs issued in the T_{inc}**-interval**. Thus,

$$E(R \mid \text{event } 1) = 2T_{prop} + E\left(W_L^{inc} \mid \text{event } 1\right). \tag{17}$$

Event 2: The first EOF reception does not spawn issuance of a NAK, and the sender's first transmission of EOF results in successful EOF delivery to the receiver:

$$\begin{aligned} R &= \left[R_L^{inc} - A\right]^+ \\ &= \left[2T_{prop} + W_L^{inc} - A\right]^+ \end{aligned}.$$

Thus,

$$E(R \mid \text{event } 2) = E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 2\right). \tag{18}$$

Note that A is the time from the issuance of the last (L^{th}) NAK that is generated by the receiver in the T_{inc}**-interval** until the reception of *EOF* and $A > 0$ in Event 2.

Event 3: The first EOF reception does not spawn issuance of a NAK and the sender's first transmission of EOF fails to deliver EOF to the receiver:

$$\begin{aligned} R &= \left[R_L^{inc} - A\right]^+ \\ &= \left[2T_{prop} + W_L^{inc} - A\right]^+ \end{aligned}.$$

Thus,

$$E(R \mid \text{event } 3) = E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 3\right). \tag{19}$$

Note that $A > 2T_{prop}$ in Event 3 because the sender's first transmission of EOF fails to deliver EOF to the receiver.

The probability of each event is as follows:

1. ***P*(event 1) =** P_{ef}. Event 1 happens if and only if the first transmission of PDU N is unsuccessful.
2. ***P*(event 2)** = $(1 - P_{ef})(1 - P_{ef(EOF)})$. Note that the first EOF reception does not spawn a NAK if and only if the first transmission of PDU N is successful, the probability of which is $(1 - P_{ef})$.
3. ***P*(event 3)** = $(1 - P_{ef})P_{ef(EOF)}$.

Note that $E(R + X) = \sum_{i=1}^{3} P(\text{event } i)E(R + X \mid \text{event } i)$ where X denotes the time spent by the sender in retransmitting PDUs before the first EOF transmission and R denotes the time-out value of the NAK_{EOF} timer. The derivation found in Appendix C leads to the following upper and lower bounds.

Case (1) : $N \geq k + 2$

$$E(R + X) < N \cdot P_{ef} \cdot T_{PDU} + \left[1 - (1 - P_{ef})P_{ef(EOF)}\right]2T_{prop}, \tag{20}$$

$$E(R + X) \geq 2T_{prop} \cdot P_{ef} + \left[(N - k - 1)P_{ef}(1 - P_{ef})(1 - P_{er}) + P_{ef}\right]T_{PDU}. \tag{21}$$

Case (2) : $N < k + 2$

$$E(R) < 2T_{prop}\left[1 - (1 - P_{ef})P_{ef(EOF)}\right] + \left\{(N - 1)P_{ef}\left[1 - (1 - P_{ef})P_{ef(EOF)}^2\right] + P_{ef}\right\}T_{PDU}, \tag{22}$$

$$\begin{aligned} E(R) \geq\ & (2T_{prop} + T_{PDU})P_{ef} \\ & + \left[2T_{prop} - (N - 1)T_{PDU}\right]\left[1 - (1 - P_{ef})^{N-1}\right](1 - P_{ef})(1 - P_{ef(EOF)}) \\ & + \left[N \cdot P_{ef}(2 - P_{ef}) - 2P_{ef}(1 - P_{ef})\right]T_{PDU}(1 - P_{ef})(1 - P_{ef(EOF)}) \end{aligned}. \tag{23}$$

4.3 Bounds on the expected file delivery time of the Immediate NAK mode

Case (1) : $N \geq k+2$

By inserting (20) and (16) into (15) and adding (13), the upper bound on the expected file delivery time of the Immediate NAK mode is obtained as

$$\begin{aligned}
&\text{Expected file delivery time} \equiv E\left(T_{inc} + T_{def}\right) \\
&< N\left(1+P_{ef}\right)T_{PDU} + \left[1-\left(1-P_{ef}\right)P_{ef(EOF)}\right]2T_{prop} + \frac{\left(1+P_{ef(EOF)}\right)T_{prop}}{1-P_{ef(EOF)}} \\
&\quad +\left[E\left(M_N\right)-1\right]\left[\frac{2T_{prop}}{1-P_{er}}\right] - \left[1-\left(1-P_{ef}\right)^N\right]\left(1-P_{er}\right)^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^N}}\left[\frac{2T_{prop}}{1-P_{er}}\right] \\
&\quad + \frac{N\cdot T_{PDU}\cdot P_{ef}}{\left(1-P_{ef}\right)\left(1-P_{er}\right)} - N\cdot P_{ef}\cdot T_{PDU}
\end{aligned} \quad . \tag{24}$$

By inserting (21) and (16) into (15) and adding (14), the lower bound on the expected file delivery time of the Immediate NAK mode is obtained as

$$\begin{aligned}
&\text{Expected file delivery time} \equiv E\left(T_{inc} + T_{def}\right) \\
&\geq N\cdot T_{PDU} + 2T_{prop}\cdot P_{ef} + \left[(N-k-1)P_{ef}\left(1-P_{ef}\right)\left(1-P_{er}\right)+P_{ef}\right]T_{PDU} \\
&\quad + \frac{\left(1+P_{ef(EOF)}\right)T_{prop}}{1-P_{ef(EOF)}} \\
&\quad +\left[E\left(M_N\right)-1\right]\left[\frac{2T_{prop}}{1-P_{er}}\right] - \left[\frac{1-\left(1-P_{ef}\right)^N}{1+P_{er}}\right]\left[\frac{2T_{prop}}{1-P_{er}}\right] \\
&\quad + \frac{N\cdot T_{PDU}\cdot P_{ef}}{\left(1-P_{ef}\right)\left(1-P_{er}\right)} - N\cdot P_{ef}\cdot T_{PDU}
\end{aligned} \quad . \tag{25}$$

Case (2) : $N < k+2$

By inserting (22) and (16) into (15) and adding (13), the upper bound of the expected file delivery time of the Immediate NAK mode is obtained as

$$\begin{aligned}
&\text{Expected file delivery time} \equiv E\left(T_{inc} + T_{def}\right) \\
&< N\cdot T_{PDU} + 2T_{prop}\left[1-\left(1-P_{ef}\right)P_{ef(EOF)}\right] \\
&\quad + \left\{(N-1)P_{ef}\left[1-\left(1-P_{ef}\right)P^2_{ef(EOF)}\right]+P_{ef}\right\}T_{PDU} + \frac{\left(1+P_{ef(EOF)}\right)T_{prop}}{1-P_{ef(EOF)}} \\
&\quad +\left[E\left(M_N\right)-1\right]\left[\frac{2T_{prop}}{1-P_{er}}\right] - \left[1-\left(1-P_{ef}\right)^N\right]\left(1-P_{er}\right)^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^N}}\left[\frac{2T_{prop}}{1-P_{er}}\right] \\
&\quad + \frac{N\cdot T_{PDU}\cdot P_{ef}}{\left(1-P_{ef}\right)\left(1-P_{er}\right)} - N\cdot P_{ef}\cdot T_{PDU}
\end{aligned} \quad . \tag{26}$$

By inserting (23) and (16) into (15) and adding (14), the lower bound of the expected file delivery time of the Immediate NAK mode is obtained as

$$
\begin{aligned}
&\text{Expected file delivery time} \equiv E\left(T_{inc} + T_{def}\right) \\
&\geq N \cdot T_{PDU} + \left(2T_{prop} + T_{PDU}\right)P_{ef} \\
&\quad + \left[2T_{prop} - (N-1)T_{PDU}\right]\left[1 - \left(1 - P_{ef}\right)^{N-1}\right]\left(1 - P_{ef}\right)\left(1 - P_{ef(EOF)}\right) \\
&\quad + \left[N \cdot P_{ef}\left(2 - P_{ef}\right) - 2P_{ef}\left(1 - P_{ef}\right)\right]T_{PDU}\left(1 - P_{ef}\right)\left(1 - P_{ef(EOF)}\right) \\
&\quad + \frac{\left(1 + P_{ef(EOF)}\right)T_{prop}}{1 - P_{ef(EOF)}} \\
&\quad + \left[E\left(M_N\right) - 1\right]\left[\frac{2T_{prop}}{1 - P_{er}}\right] - \left[\frac{1 - \left(1 - P_{ef}\right)^N}{1 + P_{er}}\right]\left[\frac{2T_{prop}}{1 - P_{er}}\right] \\
&\quad + \frac{N \cdot T_{PDU} \cdot P_{ef}}{\left(1 - P_{ef}\right)\left(1 - P_{er}\right)} - N \cdot P_{ef} \cdot T_{PDU}
\end{aligned} \qquad . \tag{27}
$$

5. Numerical Evaluation and Discussions

5.1 Numerical results for the Immediate NAK mode

The mathematical expression derived in previous sections for the expected file delivery time in the Immediate NAK mode is numerically presented in Figures 8-11. Note that the bounds obtained in this paper in relation to the expected file delivery time become tighter as the transmission rate or propagation delay increases. This is predicted from the formulas (20)–(23). Note that the considered region of BER without FEC is between 10^{-5} and 10^{-7} since the achievable BER without FEC is between 10^{-5} and 10^{-7} in typical space communication. Also, note that the astronomical unit (a.u.) is used in case of long propagation delay (1 a.u. = 480 s).

Figures 8-11 also show the average file delivery time obtained through random simulation. From the numerical results, we observe that the average file delivery time obtained through random simulation numerically lies between the upper bound and the lower bound obtained from the mathematical derivation for most points. On a small number of occasions, we obtained results of random simulation outside the bounds. We must note that the results of random simulation are not fully accurate. The random simulation repeats and averages the computation results for different trials in which parameters are randomly generated. Therefore, the confidence level increases with the number of trials run. There can always be an anomaly for any finite number of trials.

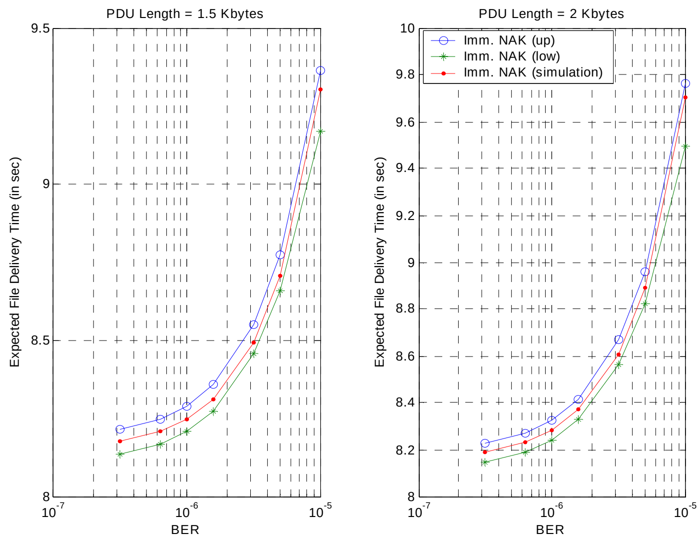

Figure 8. Immediate NAK: Analytic vs. simulation results.

Expected file delivery time of Immediate NAK mode with BER variation. File size = 1 Mbyte, transmission rate = 1 Mbps in both directions, and propagation delay = 40 ms.

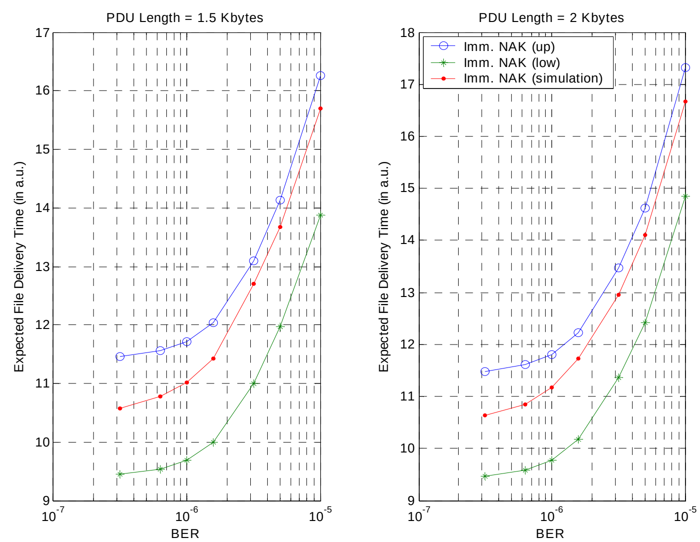

Figure 9. Immediate NAK: Analytic vs. simulation results.

Expected file delivery time of Immediate NAK mode with BER variation. File size = 1 Mbyte, transmission rate = 2 Kbps in both directions, and propagation delay = 480 s.

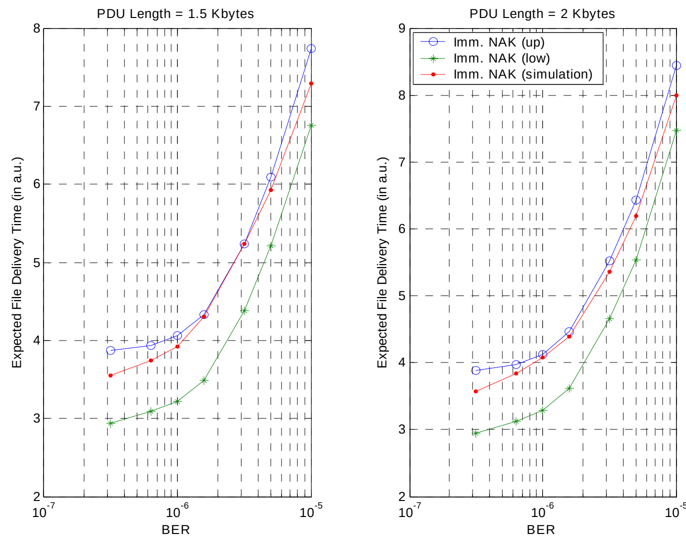

Figure 10. Immediate NAK: Analytic vs. simulation results.

Expected file delivery time of Immediate NAK mode with BER variation. File size = 1 Mbyte, transmission rate = 20 Kbps in both directions, and propagation delay = 480 s.

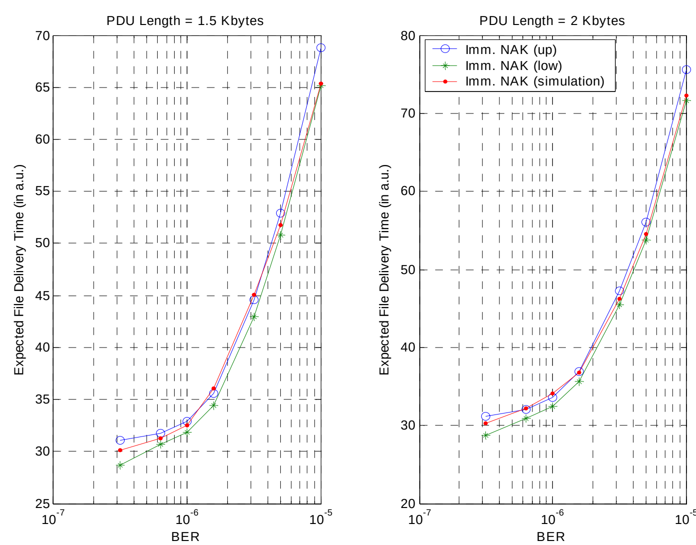

Figure 11. Immediate NAK: Analytic vs. simulation results.

Expected file delivery time of Immediate NAK mode with BER variation. File size = 1 Mbyte, Transmission rate = 20 Kbps in both directions, and propagation delay = 4,800 s.

5.2. Performance comparison of the Immediate and Deferred NAK modes

In Figures 12–15, the performance of the Immediate NAK mode is compared with that of the Deferred NAK mode. In Section 4.2, the analysis of the Immediate NAK mode is conducted differently for two cases. In case (1), the total number of PDUs, N, is greater than or equal to the number of PDUs that can be transmitted during two times the propagation delay, which we denote by num_in_2prop, plus 2. In case (2), $N <$ num_in_2prop + 2. In Figures 12–15 we compare the performance of the Deferred NAK mode and the Immediate NAK mode for both cases of $N \geq$ num_in_2prop + 2 and $N <$ num_in_2prop + 2.

When the sender and receiver are very close to each other, such as between a lander and an orbiter or between a rover and a lander in space networks deployed on other planets, we face smaller propagation delay than the transmission time of a file ($N \geq$ num_in_2prop + 2). To compare the performance of the Deferred and Immediate NAK modes in such an environment, we used the scenario of transferring a 1-Mbyte file through a full duplex link with link speed at 1 Mbps in both directions. We set the propagation delay at 40 ms. Numerical results for expected file delivery time of Deferred and Immediate NAK mode are shown in Figure 12. We also considered the case wherein $N \geq$ num_in_2prop + 2 results from very low transmission rate. For example, the transmission rate in deep space networks might be a few Kbps. Suppose that the transmission rate is 2 Kbps and file size is 1 Mbyte, and that propagation delay is 1 a.u. (1 a.u. = 480 s). Consider the example cases, N = 667 for PDU size 1.5 Kbytes and N = 500 for PDU size 2 Kbytes. In these cases, N is larger than num_in_2prop + 2. The numerical results are shown in Figure 13 for such a case.

To numerically compare the performance of the Deferred and Immediate NAK modes in $N <$ num_in_2prop + 2 environment, we take an example of a full duplex link in which the link speed is 20 Kbps in both directions. We set the file size to be 1 Mbyte. We take two cases: (1) the propagation delay = 1 a.u., and (2) the propagation delay = 10 a.u.. For 1 a.u. of propagation delay, numerical results for expected file delivery time of the Deferred and Immediate NAK modes are shown in Figure 14. For the other scenario (10 a.u. propagation delay), numerical results for expected file delivery times of the Deferred and Immediate NAK modes are shown in Figure 15.

From the numerical results (Figure 12, Figure 13, Figure 14, and Figure 15), we notice that no significant performance difference is observed comparing Immediate NAK mode to Deferred NAK mode. We observed some performance improvement by Immediate NAK mode in the case of the $N \approx$ num_in_2prop and low BER regions, but it is less than 10% improvement. We can deduce that no significant performance difference is expected in the range of parameters considered. (For other range of parameters, again data can be generated either through random simulation or the performance bounds analytically derived in this paper to compare Deferred NAK and Immediate NAK modes. With performance bound derived in this paper, we can compute the upper bound on the performance gain of the Immediate NAK mode. Use of the analytical performance bound has the advantages of low computational load and certainty, although this approach many not indicate the exact gain for a parameter that makes the numerical bound loose.) As a result, it might be favorable to operate CFDP in Deferred NAK mode, since many fewer NAKs are required in this mode. The power consumed to generate NAKs frequently may have to be a significant design consideration in the deep space environment where power is extremely limited.

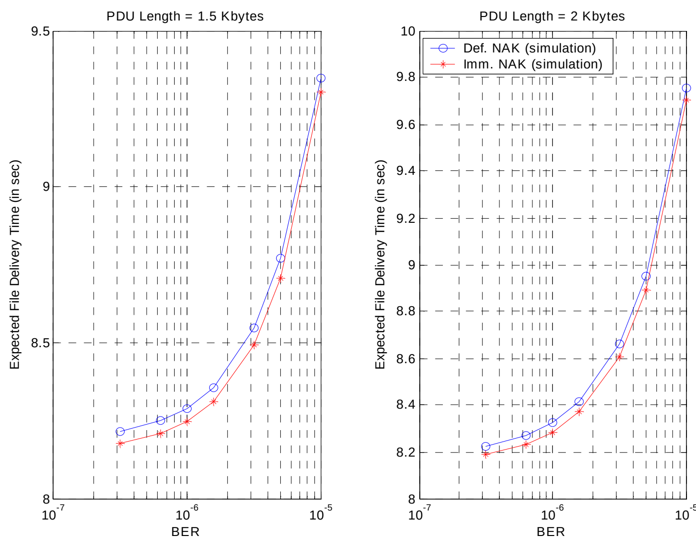

Figure 12. Simulation results. Expected file delivery time of Deferred and Immediate NAK modes with BER variation. File size = 1 Mbyte, transmission rate = 1 Mbps in both directions, and propagation delay = 40 ms.

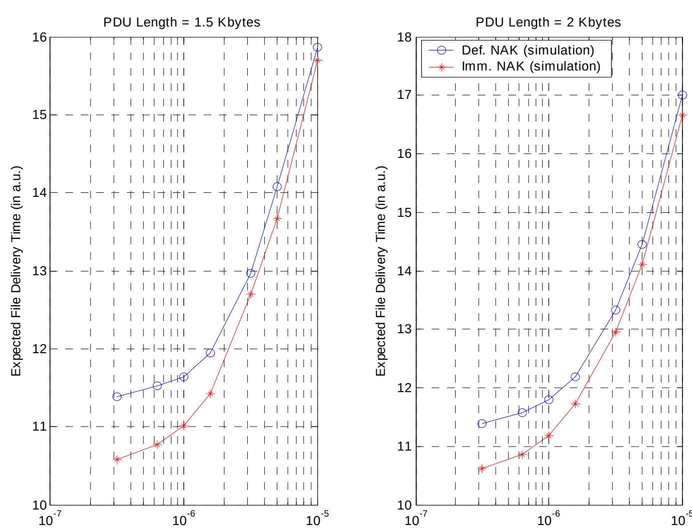

Figure 13. Simulation results. Expected file delivery time of the Deferred and Immediate NAK modes with BER variation. File size = 1 Mbyte, transmission rate = 2 Kbps in both directions, and propagation delay = 480 s.

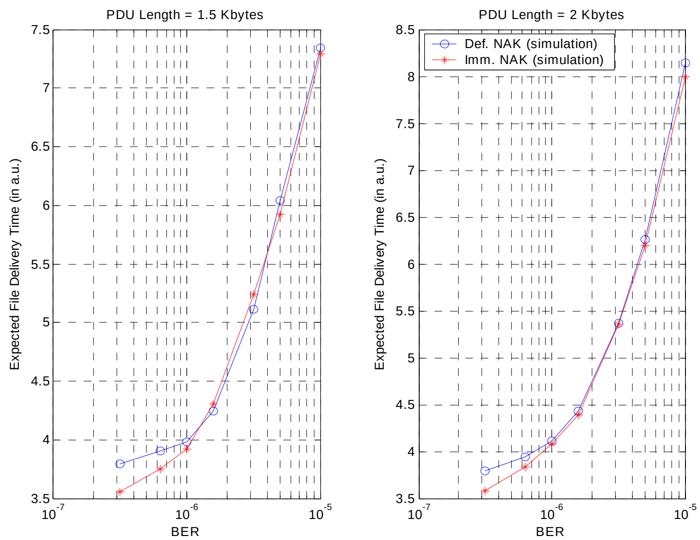

Figure 14. Simulation results. Expected file delivery time of the Deferred and Immediate NAK modes with BER variation. File size = 1 Mbyte, transmission rate = 20 Kbps in both directions, and propagation delay = 480 s.

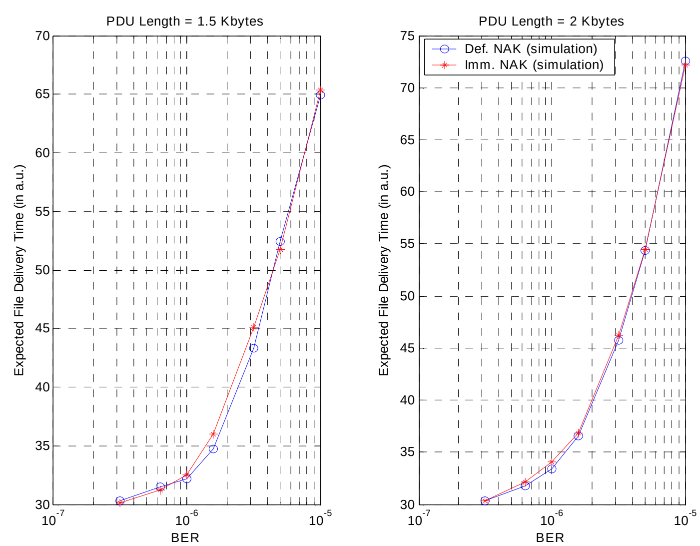

Figure 15. Simulation results. Expected file delivery time of the Deferred and Immediate NAK modes with BER variation. File size = 1 Mbyte, transmission rate = 20 Kbps in both directions, and propagation delay = 4,800 s.

6. Conclusion

In this paper, we derived performance bounds on the expected file delivery time of the CFDP *Immediate NAK* mode. We observe that the throughput efficiency can be compromised in the form of unnecessary duplicate retransmissions of an identical PDU. Due to very limited power budget in the deep space networking, in our performance analysis we consider the operational constraint that the throughput efficiency should not be compromised. To maximize the throughput efficiency (in other words, to avoid unnecessary duplicate retransmissions) and to minimize the expected file delivery time at the same time, we proposed a timer control scheme to be used in the *Immediate NAK* of CFDP. Based on our proposal for the timer control scheme, we derived bounds on the expected number of retransmission spurts, which is the most crucial part in determining the expected file delivery time. Then, we further derived bounds on the expected file delivery time. We provided numerical evaluation of our performance bounds and results of random simulation for comparison.

Appendix A

The total time that the sender's transmitter spends in transmitting meta data and file data PDUs (total N PDUs) until all of them are successfully delivered during the whole duration of the transaction can be expressed as,

$$N \cdot T_{PDU} + Z + RT_1 + RT_2 + ... + RT_{H_N},$$

where Z denotes the total time that the sender's transmitter spends in retransmitting PDUs in the T_{inc}**-interval** and $RT_1 + RT_2 + ... + RT_{H_N}$ is the time that the sender's transmitter spends in retransmitting the PDUs in the T_{def} **-interval.** (For an illustration, see Figure 1.) We have a simple expression for the expected value of the total time that the sender's transmitter spends in transmitting meta data and file data PDUs—namely,

$$E[N \cdot T_{PDU} + Z + RT_1 + RT_2 + ... + RT_{H_N}] = \frac{N \cdot T_{PDU}}{1 - P_{ef}} . \quad \text{A.1}$$

Thus, we have

$$E(RT_1 + RT_2 + ... + RT_{H_N}) = \frac{N \cdot T_{PDU}}{1 - P_{ef}} - N \cdot T_{PDU} - E(Z) = \frac{N \cdot T_{PDU} \cdot P_{ef}}{1 - P_{ef}} - E(Z) . \quad \text{A.2}$$

To obtain $E(Z)$, we observe that PDU n is retransmitted during the T_{inc}**-interval** if and only if (i) the first transmission of PDU n is unsuccessful, and (ii) the NAK requesting PDU n for retransmission is successfully delivered to the sender. Therefore, the probability that PDU n is retransmitted during the T_{inc}**-interval** is $P_{ef}(1-P_{er})$. Because there are N PDUs in the file, the expected value of Z is given as

$$E(Z) = N \cdot P_{ef}(1-P_{er})T_{PDU} \; . \qquad \text{A.3}$$

From A.2 and A.3, we have

$$E\left(RT_1 + RT_2 + \ldots + RT_{H_N}\right) = \frac{N \cdot T_{PDU} \cdot P_{ef}}{1-P_{ef}} - N \cdot P_{ef}(1-P_{er})T_{PDU} \; . \qquad \text{A.4}$$

Appendix B

The random variable M_N is defined as $\max(K_1, K_2, \cdots K_N)$, where random variable K_i represents the number of transmissions of the i^{th} PDU up to and including its first successful reception [19]. We again note that $K_1, K_2, \cdots K_N$ are I.I.D. and geometrically distributed random variables. We observe a certain relationship between M_N and H_N, the number of retransmission spurts during the T_{def}**-interval** in the Immediate NAK mode. First, we note that if all N PDUs are successfully received by the receiver at their first transmission (i.e., $M_N = 1$), then $H_N = 0$; therefore,

$$E(H_N \mid M_N = 1) = 0 \, .$$

For event $M_N \equiv \max(K_1, K_2, \cdots, K_N) > 1$, we consider the following two sub-events:

Event a : $M_N > 1$ and

$\exists$ PDU i such that $K_i = M_N$ and such that PDU i is transmitted only once (no retransmission) during the T_{inc}**-interval.** (This event happens due to the error/loss of NAKs requesting such PDUs.)

Event b: $M_N > 1$ and

$\forall$ PDU i such that $K_i = M_N$ PDU i is transmitted twice during the T_{inc}**-interval.**

Event $M_N = 1$, Event a, and Event b are mutually exclusive and collectively exhaustive. Thus,

$$\Pr[M_N = 1] + \Pr[\text{Event a}] + \Pr[\text{Event b}] = 1 \, .$$

Also, note that

$$E(H_N \mid M_N = 1) = E(M_N - 1 \mid M_N = 1) = 0$$
$$E(H_N \mid \text{Event a}) = E(M_N - 1 \mid \text{Event a}) \quad .$$
$$E(H_N \mid \text{Event b}) = E(M_N - 2 \mid \text{Event b})$$

Therefore, the unconditional expectation of H_N is given as

$$\begin{aligned} E(H_N) &= E(M_N - 1 \mid M_N = 1)\Pr[M_N = 1] \\ &\quad + E(M_N - 1 \mid \text{Event a})\Pr[\text{Event a}] \\ &\quad + E(M_N - 2 \mid \text{Event b})\Pr[\text{Event b}] \\ &= E(M_N) - 1 - \Pr[\text{Event b}] \end{aligned} \quad . \qquad \text{B.1}$$

As a result, we can have a simple relation between $E(M_N)$ and $E(H_N)$:

$$E(M_N) - 2 \le E(H_N) \le E(M_N) - 1 \ . \qquad \text{B.2}$$

Note that using B.2 to obtain the bounds on the expected value of T_{def} will result in loose bounds if propagation delay is very long. Now, we make the bounds on $E(H_N)$ tighter by finding the bounds on $\Pr[\text{Event b}]$. First, we consider a set, Π_M, of PDUs that are transmitted M_N times in the transaction—i.e., these are the PDUs that prolong the file delivery time. (This set, Π_M, is a random set because exactly which *N* PDUs will be in the set is random.) Then, we consider the NAKs that are generated in the T_{inc}**-interval** and request a PDU belonging to set Π_M. We denote by random variable *S* the number of such NAKs. Then, we have $\Pr[S = 0] = (1 - P_{ef})^N$ because $S = 0$ if and only if all *N* PDUs are successfully delivered to the receiver without retransmission. Thus, we have $\Pr[S > 0] = 1 - (1 - P_{ef})^N$. We now express $\Pr[\text{Event b}]$ as:

$$\Pr[\text{Event b}] = \Pr[\text{Event b} \mid S = 0]\Pr[S = 0] + \Pr[\text{Event b} \mid S > 0]\Pr[S > 0].$$

We obviously have $\Pr[\text{Event b} \mid S = 0] = 0$ because $S = 0$ implies that all *N* PDUs are successfully delivered to the receiver without retransmission. Thus, we have

$$\Pr[\text{Event b}] = \Pr[\text{Event b} \mid S > 0]\Pr[S > 0].$$

For each $s > 0$, we have

$$\Pr[\text{Event b} \mid S = s] = (1 - P_{er})^s \ . \qquad \text{B.3}$$

(Given that *S* takes a positive integer *s*, event b happens if and only if those *s* NAKs reach the sender successfully.) We observe that

$$\begin{aligned}\Pr[\text{Event b} \mid S>0] &= \sum_{s>0}\Pr[S=s \mid S>0](1-P_{er})^{s} \\ &= E\left[(1-P_{er})^{S} \mid S>0\right] \\ &\geq (1-P_{er})^{E(S|S>0)}\end{aligned} \quad . \qquad \text{B.4}$$

(The last inequality follows from Jansen's inequality [20].) From the fact that S is less than or equal to the total number of NAKs generated in the T_{inc}**-interval**, we have

$$E(L) \geq E(S) = E(S \mid S>0)\Pr[S>0] \ , \qquad \text{B.5}$$

where random variable L denotes the total number of NAKs generated in the T_{inc}**-interval** as defined in Section 3.2. Relations in B.5 provide a bound on $E(S \mid S>0)$,

$$E(S \mid S>0) \leq \frac{E(L)}{\Pr[S>0]} \ , \qquad \text{B.6}$$

and $E(L)$ can be simply derived.

Lemma

$$E(L) = (N-1)P_{ef}\left(1-P_{ef}\right) + P_{ef} \ . \qquad \text{B.7}$$

Proof: For the reception of PDU *i* to generate a NAK, PDU *i* must be delivered successfully at the first transmission, and also the first transmission of PDU *i*-1 must be unsuccessful. Term $(N-1)P_{ef}\left(1-P_{ef}\right)$ accounts for the expected number of NAKs generated in response to successful reception of file data PDUs in the T_{inc}**-interval**. For the reception of EOF to generate a NAK, the first transmission of PDU N must be unsuccessful. Term P_{ef} accounts for the expected number of NAKs generated in response to the reception of EOF. ■

Thus, from B.4-B.7 we have

$$\Pr[\text{Event b} \mid S>0] \geq (1-P_{er})^{E(S|S>0)} \geq (1-P_{er})^{\frac{E(L)}{\Pr[S>0]}} = (1-P_{er})^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^{N}}} \ . \qquad \text{B.8}$$

Thus, from B.8, we obtain the following lower bound of $\Pr[\text{Event b}]$:

$$\begin{aligned}\Pr[\text{Event b}] &= \Pr[S>0]\Pr[\text{Event b}|\text{S}>0] \\ &\geq \Pr[S>0](1-P_{er})^{\frac{E(L)}{\Pr[S>0]}} = \left[1-\left(1-P_{ef}\right)^{N}\right](1-P_{er})^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^{N}}}\end{aligned} \ . \qquad \text{B.9}$$

For an upper bound, we note $(1-P_{er})^{s} \leq \frac{1}{1+sP_{er}} \leq \frac{1}{1+P_{er}}$ for $s>0$ from Bernoulli's inequality [21]. Thus, from B.3

$$\Pr[\text{Event b} \mid S>0] \leq \frac{1}{1+P_{er}} \, .$$

Therefore, we have

$$\Pr[\text{Event b}] = \Pr[S>0]\Pr[\text{Event b}|\text{S>0}] \le \frac{\Pr[S>0]}{1+P_{er}} = \frac{1-\left(1-P_{ef}\right)^N}{1+P_{er}} . \qquad \text{B.10}$$

From B.1, B.9, and B.10, the bounds on $E(H_N)$ are given as:

$$E(M_N)-1-\frac{1-\left(1-P_{ef}\right)^N}{1+P_{er}} \le E(H_N) \le E(M_N)-1-\left[1-\left(1-P_{ef}\right)^N\right](1-P_{er})^{\frac{(N-1)P_{ef}(1-P_{ef})+P_{ef}}{1-(1-P_{ef})^N}} .$$

Appendix C

C.1 ***<u>Case (1)</u>:*** $N \ge k+2$ ***(Short propagation delay relative to file transmission time)***

C.1.1 Upper bound of $E(X+R)$

C.1.1.1 Event 1

From (17), we have $E(X+R \mid \text{event 1}) = E(X \mid \text{event 1}) + E\left(W_L^{inc} \mid \text{event 1}\right) + 2T_{prop}$ where *X* denotes the time spent by the sender in retransmitting PDUs before the first EOF transmission and *R* denotes the time-out value of the NAK_{EOF} timer. For the definition of W_L^{inc}, refer to (6). To obtain an upper bound on $E(X \mid \text{event 1}) + E\left(W_L^{inc} \mid \text{event 1}\right)$, we observe that $X + W_L^{inc}$ cannot exceed the total transmission time of PDUs whose transmission is unsuccessful on the first attempt**.** Therefore,

$$E(X \mid \text{event 1}) + E\left(W_L^{inc} \mid \text{event 1}\right) \le E(Q_N \cdot T_{PDU} \mid \text{event 1}),$$

where Q_N is the number of PDUs whose transmission is unsuccessful on the first attempt**.** Event 1 happens if and only if the first transmission of PDU *N* (the last file data PDU) fails to be delivered to the receiver. Therefore, we have $E(Q_N \mid \text{event 1}) = (N-1)P_{ef}+1$, where the last term 1 accounts for the fact that the last file data PDU is unsuccessful in Event 1. Thus, we have

$$E(X \mid \text{event 1}) + E\left(W_L^{inc} \mid \text{event 1}\right) \le E(Q_N \mid \text{event 1})T_{PDU} = \left[(N-1)P_{ef}+1\right]T_{PDU} . \qquad \text{C.1}$$

As a result,

$$\begin{aligned} E(X+R \mid \text{event 1}) &= E(X \mid \text{event 1}) + E\left(W_L^{inc} \mid \text{event 1}\right) + 2T_{prop} \\ &\le \left[(N-1)P_{ef}+1\right]T_{PDU} + 2T_{prop} \end{aligned} . \qquad \text{C.2}$$

C.1.1.2 Event 2

From (18), we have

$$E(X+R \mid \text{event } 2) = E(X \mid \text{event } 2) + E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 2\right) . \qquad \text{C.3}$$

where A is the time from the issuance of the last (L^{th}) NAK that is generated by the receiver in the T_{inc} **-interval** until the reception of *EOF*. From the fact that $A > 0$ in Event 2,

$$\begin{aligned} &E(X \mid \text{event } 2) + E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 2\right) \\ &< E(X \mid \text{event } 2) + E\left(W_L^{inc} \mid \text{event } 2\right) + 2T_{prop} \end{aligned} . \qquad \text{C.4}$$

Also, note that $E(X \mid \text{event } 2) + E\left(W_L^{inc} \mid \text{event } 2\right)$ can be bounded in a way similar to that in Section C.1.1.1—namely,

$$E(X \mid \text{event } 2) + E\left(W_L^{inc} \mid \text{event } 2\right) \le E(Q_N \mid \text{event } 2) T_{PDU} = (N-1) P_{ef} \cdot T_{PDU} . \qquad \text{C.5}$$

(Note that event 2 implies that PDU N's first transmission is successful.)

Thus, we have

$$E(X+R \mid \text{event } 2) < (N-1) P_{ef} \cdot T_{PDU} + 2T_{prop} . \qquad \text{C.6}$$

C.1.1.3 Event 3

From (19), we have

$$E(X+R \mid \text{event } 3) = E(X \mid \text{event } 3) + E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 3\right) . \qquad \text{C.7}$$

From the fact that $A > 2T_{prop}$ in Event 3,

$$E(X \mid \text{event } 3) + E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 3\right) < E(X \mid \text{event } 3) + E\left(W_L^{inc} \mid \text{event } 3\right) . \qquad \text{C.8}$$

Note that $E(X \mid \text{event } 3) + E\left(W_L^{inc} \mid \text{event } 3\right)$ can be bounded in the same way as was done in Section C.1.1.2. As a result, we have

$$E(X+R \mid \text{event } 3) < (N-1) P_{ef} \cdot T_{PDU} . \qquad \text{C.9}$$

C.1.1.4 Resulting upper bound of $E(X+R)$

From the weighted sum of C.2, C.6, and C.9 with the probability for each event specified in Section 4.2, the upper bound on the expected value of $X+R$ is obtained as

$$E(X+R) < N \cdot P_{ef} \cdot T_{PDU} + \left[1 - (1 - P_{ef}) P_{ef(EOF)}\right] 2T_{prop} . \qquad \text{C.10}$$

C.1.2 Lower bound of $E(X+R)$

C.1.2.1 Event 1

To obtain a lower bound on *E(X)*, the expected value of *X* that denotes the time spent by the sender in retransmitting PDUs before the first EOF transmission, we observe the following. Let us denote by random variable L_{N-k} the total number of NAKs spawned by receiving any of PDUs 1, 2, …, (*N*-*k*) during the T_{inc}**-interval.** Then, the expected number of those NAKs successfully delivered to the sender is $E(L_{N-k})(1-P_{er})$. Each of these NAKs requests retransmission of at least one PDU, so we have a lower bound $E(X \mid \text{event } 1) \geq E(L_{N-k})(1-P_{er})T_{PDU}$. Regarding $E(L_{N-k})$, we have

$$E(L_{N-k}) = \sum_{i=2}^{N-k} P(\text{ PDU } i-1 \text{ fails }) \text{ P(PDU } i \text{ is successfully delivered)} = (N-k-1)P_{ef}\left(1-P_{ef}\right)$$

Thus, we have

$$E(X \mid \text{event } 1) \geq E(L_{N-k})(1-P_{er})T_{PDU} = (N-k-1)P_{ef}\left(1-P_{ef}\right)(1-P_{er})T_{PDU} \quad \text{. C.11}$$

With regard to the random variable *R*, the time-out value of the NAK_{EOF} timer, in Event 1 we observe that

$$R = 2T_{prop} + W_L^{inc} \geq 2T_{prop} + T_{PDU} \qquad \text{C.12}$$

because at least the last file data PDU is unsuccessful in Event 1. As a result,

$$\begin{aligned} E(X+R \mid \text{event } 1) &= E(X \mid \text{event } 1) + E\left(W_L^{inc}\right) + 2T_{prop} \\ &\geq \left[(N-k-1)P_{ef}\left(1-P_{ef}\right)(1-P_{er})+1\right]T_{PDU} + 2T_{prop} \end{aligned} \quad . \qquad \text{C.13}$$

C.1.2.2 Event 2

A lower bound on *E*(*X*) is obtained in the same way as in C.11, so we have

$$E(X+R \mid \text{event } 2) \geq E(X \mid \text{event } 2) \geq (N-k-1)P_{ef}\left(1-P_{ef}\right)(1-P_{er})T_{PDU} \quad . \qquad \text{C.14}$$

C.1.2.3 Event 3

A lower bound on *E*(*X*) is obtained in the same way as in C.11, so we have

$$E(X+R \mid \text{event } 3) \geq E(X \mid \text{event } 3) \geq (N-k-1)P_{ef}\left(1-P_{ef}\right)(1-P_{er})T_{PDU} \quad . \qquad \text{C.15}$$

C.1.2.4 Resulting lower bound of $E(X+R)$

From the weighted sum of C.13, C.14, and C.15 with the probability for each event specified in Section 4.2, the lower bound on the expected value of $X+R$ is obtained as

$$E(X+R) \geq 2T_{prop} \cdot P_{ef} + \left[(N-k-1)P_{ef}\left(1-P_{ef}\right)(1-P_{er}) + P_{ef}\right]T_{PDU} \quad . \qquad \text{C.16}$$

C.2 *Case (2)*: $N < k+2$ *(Long propagation delay relative to file transmission time)*

C.2.1 Upper bound of $E(R)$

Note that *X*, the time spent by the sender in retransmitting PDUs before the first EOF transmission, is zero ($X=0$) for *N* < *k+2*. Therefore, we obtain a bound on the expected value of *R*, the time-out value of the NAK_{EOF} timer, in the following sections.

C.2.1.1 Event 1

Note that in Event 1 W_L^{inc} is at most $Q_N \cdot T_{PDU}$, where Q_N is the number of PDUs whose transmission is unsuccessful at on the first attempt. Event 1 happens if and only if the first transmission of PDU *N* (the last file data PDU) fails to be delivered to the receiver. Therefore, we have $E(Q_N \mid \text{event } 1) = (N-1)P_{ef} + 1$. Thus,

$$E\left(W_L^{inc} \mid \text{event } 1\right) \leq E(Q_N \mid \text{event } 1) \cdot T_{PDU} = \left[(N-1)P_{ef} + 1\right] \cdot T_{PDU} \quad . \qquad \text{C.17}$$

As a result,

$$\begin{aligned} E(R \mid \text{event } 1) &= 2T_{prop} + E\left(W_L^{inc} \mid \text{event } 1\right) \\ &\leq 2T_{prop} + \left[(N-1)P_{ef} + 1\right] \cdot T_{PDU} \end{aligned} \quad . \qquad \text{C.18}$$

C.2.1.2 Event 2

From the fact that W_L^{inc} is at most $Q_N \cdot T_{PDU}$ and $E(Q_N \mid \text{event } 2) = (N-1)P_{ef}$, we have

$$E\left(W_L^{inc} \mid \text{event } 2\right) \leq E(Q_N \mid \text{event } 2) \cdot T_{PDU} = \left[(N-1)P_{ef}\right]T_{PDU} \quad . \qquad \text{C.19}$$

From C.19 and the fact that *A*, the time from the issuance of the last (L^{th}) NAK that is generated by the receiver in the T_{inc}**-interval** until the reception of *EOF*, is greater than zero ($A > 0$) in Event 2,

$$\begin{aligned} E(R \mid \text{event } 2) &= E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 2\right) \\ &< 2T_{prop} + E\left(W_L^{inc} \mid \text{event } 2\right) \\ &\leq 2T_{prop} + \left[(N-1)P_{ef}\right]T_{PDU} \end{aligned} \quad . \qquad \text{C.20}$$

C.2.1.3 Event 3

Note that if the second EOF trial is successful, the expectation of W_L^{inc} can be bounded the same as in C.19, $E\left(W_L^{inc} \mid \text{event 3, success of second EOF Tx}\right) \le \left[(N-1)P_{ef}\right]T_{PDU}$. Otherwise, $E\left(W_L^{inc} \mid \text{event 3, failure of second EOF Tx}\right) = 0$. This is due to the fact that if any NAK generated in the T_{inc}**-interval** is successfully delivered to the sender, it is before the sender's second EOF transmission. Therefore, any retransmission in response successful NAK receptions during the T_{inc}**-interval** is completed before the sender's third EOF transmission since $N < k+2$. As a result, we obtain

$$E\left(W_L^{inc} \mid \text{event 3}\right) \le \left[(N-1)P_{ef}\left(1-P_{ef(EOF)}\right)\right]T_{PDU} \quad , \qquad \text{C.21}$$

where $1-P_{ef(EOF)}$ is the probability of the success of the second EOF transmission.

From C.21 and the fact that $A > 2T_{prop}$ in Event 3,

$$\begin{aligned} E(R \mid \text{event 3}) &= E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event 3}\right) \\ &< E\left(W_L^{inc} \mid \text{event 3}\right) \\ &\le \left[(N-1)P_{ef}\left(1-P_{ef(EOF)}\right)\right]T_{PDU} \end{aligned} \quad . \qquad \text{C.22}$$

C.2.1.4 Resulting upper bound of $E(R)$

From the weighted sum of C.18, C.20, and C.22 with the probability for each event specified in Section 4.2, the upper bound on the expected value of R is obtained as

$$E(R) < 2T_{prop}\left[1-\left(1-P_{ef}\right)P_{ef(EOF)}\right] + \left\{(N-1)P_{ef}\left[1-\left(1-P_{ef}\right)P_{ef(EOF)}^2\right] + P_{ef}\right\}T_{PDU} \,. \quad \text{C.23}$$

C.2.2 Lower bound of $E(R)$

C.2.2.1 Event 1

C.12 Still holds. Thus,

$$E(R \mid \text{event 1}) = 2T_{prop} + E\left(W_L^{inc} \mid \text{event 1}\right) \ge 2T_{prop} + T_{PDU} \quad . \qquad \text{C.24}$$

where R denotes the time-out value of the NAK_{EOF} timer.

C.2.2.2 Event 2

If L, the number of NAKs generated during the T_{inc}**-interval,** is 0, then $R = 0$. Now, let us consider case $L > 0$. The L^{th} NAK requests retransmission of at least one PDU if $L > 0$, so we have $W_L^{inc} \ge T_{PDU}$. Also, due to $N < k+2$, we have A, the time from the issuance of the last

(L^{th}) NAK that is generated by the receiver in the T_{inc}**-interval** until the reception of EOF, is greater than two times the propagation delay ($A < 2T_{prop}$) in Event 2. Thus,

$$\begin{aligned} E(R \mid \text{event } 2,\, L > 0) &= E\left(\left[2T_{prop} + W_L^{inc} - A\right]^+ \mid \text{event } 2,\, L > 0\right) \\ &= 2T_{prop} - E(A \mid \text{event } 2,\, L > 0) + E\left(W_L^{inc} \mid \text{event } 2,\, L > 0\right) \\ &\geq 2T_{prop} - E(A \mid \text{event } 2,\, L > 0) + T_{PDU} \end{aligned} \quad \text{C.25}$$

Now, we derive an upper bound of $E(A \mid \text{event } 2,\, L > 0)$. First, we observe that the receiver receives any retransmitted PDUs after the reception of EOF because any NAK generated in the T_{inc}**-interval** arrives at the sender (if successfully delivered to the sender) after the sender's transmission of EOF. (This is due to the fact that $N < k + 2$ and that EOF is successfully delivered to the receiver upon first trial in Event 2. See Figure 16 for an illustration.)

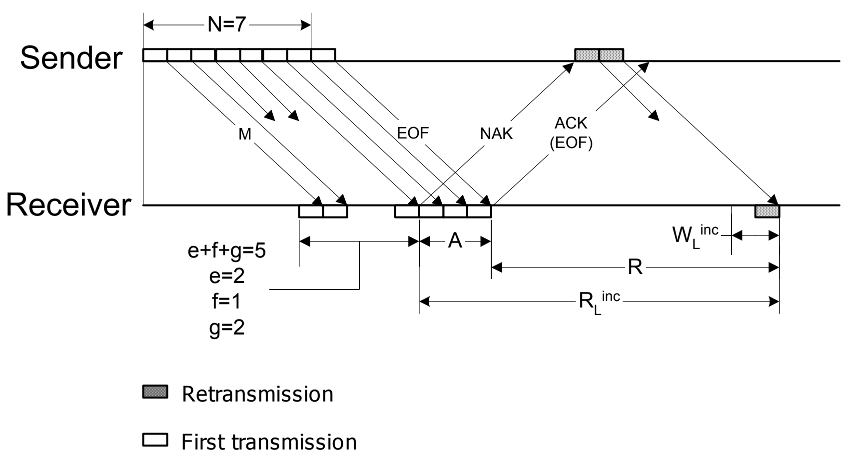

Figure 16.

Thus, it can be deduced that the time from the beginning of the reception of the meta data PDU until the reception of EOF is exactly $N \cdot T_{PDU}$. We also observe that the first transmission of each PDU results in one of the following three events:

1) Successful reception of the PDU and this does not spawn a NAK.
2) Successful reception of the PDU and this spawns a NAK.
3) Failed reception of the PDU.

Now, suppose that the L^{th} NAK (the last NAK in the T_{inc}**-interval**) is spawned by the reception of J^{th} PDU. (We are still assuming $L > 0$.) Let us denote by e, f, and g the number of received PDUs corresponding to events (1), (2), and (3), respectively, out of set {PDU 1, PDU 2, …, PDU J}. Then, the time from the beginning of the reception of the meta data PDU until and including the reception of J^{th} PDU is obtained as $J \cdot T_{PDU} = (e + f + g)T_{PDU} \geq (f + g)T_{PDU}$. Now, we note that $f = L$ and $g = Q_N$ where L is the number of NAKs generated during the T_{inc}**-interval** and Q_N is the number of PDUs whose transmission is unsuccessful at on the first attempt. (Regarding $g = Q_N$, from the definitions of L and Event 2 there is no NAK generated from the reception of the J^{th} PDU until the reception of EOF, so the first transmissions of $(J+1)^{\text{th}}$, $(J+2)^{\text{th}}$, …, N^{th} PDUs are all successful.) For example, in Figure 16, $J = e + f + g = 5$, $e = 2$, $f = L = 1$, and $g = Q_N = 2$.

From the definition of A, we have $A = (N - J)T_{PDU} \leq [N - (L + Q_N)]T_{PDU}$. As a result, we have

$$E(A \mid \text{event } 2,\, L > 0) \leq \{N - E(L + Q_N \mid \text{event } 2,\, L > 0)\}T_{PDU} \quad . \qquad \text{C.26}$$

Inserting C.26 into C.25, we have

$$E(R \mid \text{event } 2,\, L > 0) \geq 2T_{prop} - (N-1)T_{PDU} + E(L + Q_N \mid \text{event } 2,\, L > 0)T_{PDU} \quad . \qquad \text{C.27}$$

Note that $L = 0$ implies that R and Q_N are all zero, which means that we have

$$\begin{aligned} E(R \mid \text{event } 2) &= E(R \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \\ &\quad + E(R \mid \text{event } 2,\, L = 0)P(\ L = 0 \mid \text{event } 2)\ , \\ &= E(R \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \end{aligned} \qquad \text{C.28}$$

$$\begin{aligned} E(L \mid \text{event } 2) &= E(L \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \\ &\quad + E(L \mid \text{event } 2,\, L = 0)P(\ L = 0 \mid \text{event } 2)\ , \\ &= E(L \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \end{aligned} \qquad \text{C.29}$$

$$\begin{aligned} E(Q_N \mid \text{event } 2) &= E(Q_N \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \\ &\quad + E(Q_N \mid \text{event } 2,\, L = 0)P(\ L = 0 \mid \text{event } 2)\ . \\ &= E(Q_N \mid \text{event } 2,\, L > 0)P(\ L > 0 \mid \text{event } 2) \end{aligned} \qquad \text{C.30}$$

From C.27, C.28, C.29, and C.30 we have

$$\begin{aligned} &E(R \mid \text{event } 2) \\ &\geq [2T_{prop} - (N-1)T_{PDU}]P(L > 0 \mid \text{event } 2) + [E(L \mid \text{event } 2) + E(Q_N \mid \text{event } 2)]T_{PDU} \end{aligned} \ . \qquad \text{C.31}$$

Now, with regard to $P(\ L > 0 \mid \text{event } 2)$, $E(L \mid \text{event } 2)$, and $E(Q_N \mid \text{event } 2)$, we have the following proposition:

Proposition

a. $P(\ L > 0 \mid \text{event } 2) = 1 - (1 - P_{ef})^{N-1}$.

Proof: Event 2 implies that the first transmission of PDU N is successful. Therefore, in event 2, $L = 0$ if and only if PDU 1, PDU 2, …, PDU N-1 are in addition successful on their first transmission. Thus,

$$P(\ L = 0 \mid \text{event } 2) = \left(1 - P_{ef}\right)^{N-1}.$$

b. $E(Q_N \mid \text{event } 2) = (N-1)P_{ef}$.

Proof: In event 2, the first transmission of PDU N is successful, so Q_N is the number of PDUs in {PDU 1, PDU 2, … , PDU N-1} that fail on their first transmission. Therefore,

$$E(Q_N \mid \text{event } 2) = (N-1)P_{ef}.$$

c. $E(L \mid \text{event } 2) = (N-2)P_{ef}\left(1 - P_{ef}\right) + P_{ef}$.

Proof: PDU i for $2 \le i \le N-1$ spawns a NAK if and only if transmission of PDU i-1 is unsuccessful and the first transmission of PDU i is successful. Therefore, the expected number of NAKs spawned by PDU i for $2 \le i \le N-1$ is $P_{ef}\left(1 - P_{ef}\right)$. In event 2, PDU N's first transmission succeeds, so PDU N spawns a NAK if and only if PDU N-1 fails to be delivered on its first transmission. Therefore, the expected number of NAKs spawned by PDU N is P_{ef}. Thus,

$$E(L \mid \text{event } 2) = (N-2)P_{ef}\left(1 - P_{ef}\right) + P_{ef}.$$

From this proposition and C.31, we obtain

$$\begin{aligned} &E(R \mid \text{event } 2) \\ &\ge \left[2T_{prop} - (N-1)T_{PDU}\right]\left[1 - \left(1 - P_{ef}\right)^{N-1}\right] + \left[N \cdot P_{ef}\left(2 - P_{ef}\right) - 2P_{ef}\left(1 - P_{ef}\right)\right]T_{PDU} \end{aligned} \quad \text{. C.32}$$

C.2.2.3 Event 3

In this event, we use an obvious lower bound:

$$E(R \mid \text{event } 3) = E\left(\left[2T_{prop} + W_L^{inc} - A\right]^{+} \mid \text{event } 3\right) \ge 0\ . \qquad \text{C.33}$$

Note that $A > 2T_{prop}$, so this bound is not too loose.

C.2.2.4 Resulting lower bound of $E(R)$

From the weighted sum of C.24, C.32, and C.33 with the probability for each event specified in Section 4.2, the lower bound on the expected value of R is obtained as

$$
\begin{aligned}
E(R) \geq & \left(2T_{prop} + T_{PDU}\right)P_{ef} \\
& + \left[2T_{prop} - (N-1)T_{PDU}\right]\left[1-\left(1-P_{ef}\right)^{N-1}\right]\left(1-P_{ef}\right)\left(1-P_{ef(EOF)}\right) \\
& + \left[N \cdot P_{ef}\left(2-P_{ef}\right) - 2P_{ef}\left(1-P_{ef}\right)\right]T_{PDU}\left(1-P_{ef}\right)\left(1-P_{ef(EOF)}\right)
\end{aligned} \quad . \qquad \text{C.34}
$$

REFERENCES

[1] J. H. Yuen, M.K. Simon, W. Millar, F. Pollara, C.R. Ryan, D. Divsalar, and J.C. Morakis, "Modulation and Coding for Satellite and Space Communications," *Proc. IEEE*, vol. 78, Issue 7, July 1990.

[2] R. M. Taylor, "Space Communications," *IEEE Spectrum*, vol. 29, Issue 2, Feb. 1992.

[3] R. Alena, B. Gilbaugh, B. Glass, and S.P. Braham, "Communication System Architecture for Planetary Exploration," *IEEE Aerosp. Elctron. Syst. Mag.,* Nov. 2001.

[4] CCSDS, "CCSDS File Delivery Protocol (CFDP) Part 1: Introduction and Overview," CCSDS 720.1-G-1, January 2002.

[5] CCSDS, "CCSDS File Delivery Protocol (CFDP) Part 2: Implementers Guide," CCSDS 720.2-G-1, January 2002.

[6] CCSDS, "CCSDS File Delivery Protocol (CFDP) Recommendation for Space Data System Standards," CCSDS 727.0-B-2, October 2002.

[7] N. R. Kuo, "Mars Network Operation Concept," *Proc. IEEE Aerospace Conference*, March 2000.

[8] K. K. Choi, G. Maral, and R. Rumeau, "Space Link Simulator for Development of a New Standard CCSDS File Delivery Protocol," *Proc. IEEE Aerospace Conference,* March 1999.

[9] K. K. Choi, G. Maral, and R. Rumeau, "The Implementation and Validation of the New Standard CCSDS File Delivery Protocol for Multi-Hopped Space File Transfer," *Proc. IEEE Aerospace Conference,* March 1999

[10] S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, "Delay-Tolerant Networking: An Approach to Interplanetary Internet," *IEEE Commun. Mag.,* June 2003, pp.128-36.

[11] D. Towsley and J. K. Wolf, "On the Statistical Analysis of Queue Lengths and Waiting Times for Statistical Multiplexers with ARQ Retransmission Scheme," *IEEE Trans. Commun.,* vol. COM-27, no. 4, April 1979.

[12] Shu Lin, D. Costello, and M. Miller, "Automatic-Repeat-Request Error-Control Schemes," *IEEE Commun. Mag.,* vol. 22, no. 12, Dec. 1984.

[13] A. G. Konheim, "A Queueing Analysis of Two ARQ Protocols," *IEEE Trans. Commun.*, vol. COM-28, no. 7, July 1980.

[14] D.-L. Lu and J.-F. Chang, "Analysis of ARQ Protocols via Signal Flow Graphs," *IEEE Trans. Commun.,* vol. 37, no. 3, March 1989

[15] M. Schwartz, Telecommunication Networks: Protocols, Modeling and Analysis, Addison Wesley, 1987.

[16] D. Bertsekas and R. Gallager, Data Networks, Upper Saddle River, NJ: Prentice-Hall, 1992.

[17] T. N. Saadawi and M. H. Ammar, Fundamentals of Telecommunication Networks, Wiley-Interscience, 1994.

[18] W. Baek and D. C. Lee, "Caution for Initializing and Closing CFDP Transaction," *Proc. International Telemetering Conference*, Las Vegas, NV, Oct. 2001.

[19] D. C. Lee and W. Baek, "Expected File Delivery Time of Deferred NAK ARQ in CCSDS File Delivery Protocol", *IEEE Transactions on Communications,* vol. 52, no. 8, August, 2004, pp. 1408−1416.

[20] R. G. Gallager, Discrete Stochastic processes, Boston, MA: Kluwer Academic, 1996.

[21] D. S. Mitrinović, J. E. Pečarić, and A. M. Fink, *Classical and New Inequalities in Analysis*, Dordrecht, Netherlands: Kluwer Academic, 1993.