Message Fragmentation for Disrupted Links

Philip Ginzboorg
Nokia Research Center, Finland
Email: philip.ginzboorg@nokia.com

Valtteri Niemi
Nokia Research Center, Switzerland
Email: valtteri.niemi@nokia.com

Jörg Ott
Aalto University, Finland
Email: joerg.ott@tkk.com

***Abstract*—In networks with unstable links the connection between the sender and the receiver may be cut before the entire message has been transmitted. Quantization of the message into blocks before its transmission, enables message fragments consisting of one or more such blocks to be transmitted successfully, even if the connection time is too short for transmitting the entire message. But fragmentation may also reduce the chances of delivering the entire message to its destination, due to lost or misrouted fragments.**

In this paper we (1) formulate the problem of in-time transmission of fragmented messages in disrupted networks; and (2) develop methods to analytically estimate the mean transmission time of fragmented message within a basic system model for the case of transmission over a single disrupted link.

I. Introduction

In Delay-tolerant Networks (DTNs) created by opportunistic interactions among mobile nodes, communication links will go on and off frequently and the achievable transmission rate will vary. The transmission capacity depends on the quality of the wireless channel during a contact and the mobility patterns that determine the contact durations and inter-contact times between the nodes. Given that transmission capacity during a contact is limited and that incompletely transmitted messages are lost, short contact durations may hinder message passing and thus limit the connectivity of the DTN. Message fragmentation is one way to improve connectivity through increased utilization of contact durations [1].

Theoretical studies of fragmentation in packet networks that have veen motivated by IP include [2], [3] and [4]. In IP networks, any node along the path is allowed to fragment packets which are then re-assembled at the destination. However, IP fragmentation is discouraged in practice, among other reasons, due to the increase in packet loss probability [5]. To avoid fragmentation, senders may dynamically determine the maximum transmission unit (MTU) for a given path, by probing with different packet sizes and setting the *don't fragment* flag in the IP header. It also occurs that senders choose an MTU, e.g., based upon conservative expectations on the operating environment and the application demands.

Path MTU discovery relies on the communication path being stable for some time beyond the MTU measurement and on that measurement to conclude within a reasonably short period of time. It will not work in networks with large variation of communication delays, variable link availability and changing paths between sender and receiver as we may find in Delay-tolerant Networks. This variability may also make it tricky to estimate a suitable MTU.

The Delay-tolerant Networking architecture [6] and protocol [7] define two types of fragmentation for DTNs: pro-active and reactive. In the former, the source node for this hop divides application data into blocks and sends each block in a separate fragment (message). In the latter, the data is split only when the transmission between two nodes on any link of the message path is interrupted; yielding one fragment with data that made it to the receiver and one containing the yet-to-be-transmitted remainder at the sender. In both cases the fragmented data is re-assembled only at its destination.

Authentication and integrity protection of fragments with methods in [8], [9], [10] implies that *quantization* of data, i.e. its division into blocks, must be done proactively, before its transmission over the first link of the message path. This is because the secret key needed for signatures over message data (or its parts) is assumed to be known only by the originating node. In both fragmentation approaches, it is the responsibility of the originating node to perform the necessary data quantization. Splitting the message data into separate fragments according to the quantization boundaries may happen anywhere along the message path. Thus, for the quantization choice, there is virtually no difference between pro-active and reactive fragmentation—with the small exception that the originating node, when fragmenting, could change the quantization after partial transmission, a possibility that we do not consider further in this paper.

If all fragments need to reach the destination for successful message reassembly, and any message has a non-zero loss probability p (due to dropping or mis-routing), then the payload delivery probability reduces from $1-p$ without fragmentation to $(1-p)^k$ when using k fragments. This motivates minimizing the number of fragments created and hence large quantization blocks.

The work on [1] confirmed the above: several fragmentation strategies in DTNs were evaluated by simulations, finding for a set of scenarios that reactive fragmentation with data quantization at the message source improves performance. It became evident that a model of fragmented transmission over disrupted links would help to design fragmentation strategies and interpret simulation results. This paper presents elements of such model.

We consider a node A intending to send a message of a given size m to another node B via a mobile DTN. A message will be tagged with a time-to-live (TTL) indicating the deadline of its usefulness to B. The communicating nodes (A, B and intermediate nodes) of the mobile DTN encounter each other opportunistically, the contact durations and inter-contact times following mobility-dependent distributions. The resulting contact patterns define the "connectedness" of the network with respect to message size m. Node A can choose to transmit m in a single packet, thus requiring sufficiently long contact durations for m to fit. Or A may split m into blocks of size f, thus allowing transmission of message fragments consisting of one or more such blocks during shorter contacts.[1] f may vary between messages, and we shall call it "message fragmentation unit".

How can A reduce the number of fragments by setting f large, while still allowing in-time delivery of the message to B? We make two main assumptions to make this problem trackable. First, we focus on a single link: the more complicated multi-hop and multi-path cases cannot be unpacked without understanding the single-hop one.[2] Second, we assume that disruptions in the (single) link are the only obstacle to in-time delivery. Other obstacles, like additional queuing delays, could be added once the effect of link disruptions is understood.

The main contribution of this paper are methods to estimate the mean delivery time of a message of size m through a disrupted link in DTN, as a function of fragmentation unit size f, and the locally measured contact pattern characteristics. From that estimate a network node can locally determine an optimal f that will still guarantee in-time delivery of the message.

We emphasize that our calculations of the mean delivery times are done within a simplified system model; they are thus approximations to the real transmission times. We have used those theoretical calculations to understand better the fundamentals of fragmented message transmission, and as a source of approximate estimates in fragmentation algorithms. The algorithms and their comparative performance evaluation in simulated environments will be described in another paper [13].

A contribution of independent interest from the fragmentation problem is the closed form expression, given in Eq. (V.20), for the mean of the minimum number of throws with a n-faceted die such that their sum is at least $x \geq 0$.

We present our system model in section II and formulate the fragmentation problem in section III. Section IV contains additional definitions and notation. Ways to estimate the mean transmission time are given in section V. The paper ends with summary and future work topics in section VI.

Part of the theoretical results in this paper and preliminary experimental results were published in our technical report [14].

II. Basic model of transmission over disrupted link

Network node A sends messages over an intermittent communication link to node B. The observations by A of link state changes, and the acknowledgments by B of successful transmissions do not incur a delay. The link speed V during its ON state is constant. This arrangement is depicted in Figure 1.[3]

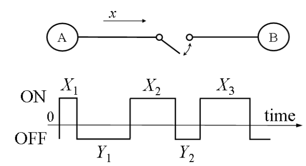

Fig. 1. System model.

The transitions from ON to OFF state may occur due to nodes mobility, link capacity being temporary alloted for other purpose, or unrecoverable transmission errors.

The durations of ON and OFF epochs are positive random variables X and Y, respectively. We denote by $F_X(\cdot)$ and $F_Y(\cdot)$ the distributions, and by $dF_X(\cdot)$ and $dF_Y(\cdot)$ the densities of those random variables. We will assume that X and Y have finite mean and variance.[4]

We assume in our model that the distributions of X and Y are stationary. We also assume in our model that X (and Y) are independent and identically distributed (i.i.d.).

We divide all message sizes by V, measuring message sizes in seconds. For instance, if the link speed is 1 Kb/s, we say that size of a 1 Kb message is one second. We denote the size of A's message by m, and by x the size of same message after quantization and addition of extra headers. Thus x is a function of m, the message fragmentation unit f, and of the header size h: $x = x(m, f, h)$.

If the link fails during transmission, node A will attempt to retransmit this message (fragment) during the next ON epoch. We denote the number of ON epochs needed for successful transmission of x seconds by $n(x)$ and by $N(x)$ the mean of $n(x)$: $N(x) = \mathrm{E}[n(x)]$. The sending node always starts to transmit (or retransmit) when the link changes state from OFF to ON. The number of OFF epochs during successful transmission is $n(x) - 1$ by this assumption. Let random variable $t(x)$ be the time needed for successful transmission of x seconds. We denote by $T(x)$ the mean of $t(x)$: $T(x) = \mathrm{E}[t(x)]$.

The mean transmission time $T(x)$ is the sum of $T_X(x)$, the mean of the total time spent in the ON state, and $T_Y(x)$, the mean of the total time spent in the OFF state after transmission has started:

$$T(x) = T_X(x) + T_Y(x). \tag{II.1}$$

Note that since $T_X(x)$ does not depend on Y, the quantity $T_X(x)$ will not change when OFF epochs are very brief. Thus, the simplest system for studying experimentally the effects of link disruptions and data quantization on transmission time is the one in which $Y \to 0$.

When $f = 0$, we assume that $h = 0$. In this case $T_X(x)$ is at its theoretical lower bound of $T_X(x) = x$: There is no waste

[1] Note that the last piece of the message may be smaller than f.

[2] Note also that some DTN routing protocols use direct transmission from A to B: Direct Delivery [11] and, for the last hop, Spray-and-Wait [12].

[3] We note that a similar model was used also in other studies of fragmentation issues. (See [15] and [16] and the references therein.)

[4] A conjecture implying that in many cases $\mathrm{E}[Y] > \infty$ have been proposed in [17], and then refuted in [18].

of transmission time due to quantization of the message into indivisible blocks; and transmission time is wasted due to link disruptions only.

We will assume that the sending node A has a lower bound f_0 on a packet size and an upper bound $L > f_0$ on the total transmission time $t(x)$ of the message. Clearly, the message may be transmitted in time only if $L \geq x$. The values of f_0 and L can be set by the sending node, based on the properties of the link and, e.g., the validity time of the message that is chosen by the sending application.

Our message transmission model is simple enough so that within it we can compute the mean transmission time $T(x)$ and thus understand better the fundamentals of fragmented message transmission over disrupted link. Naturally, $T(x)$ so computed is only an approximation to the transmission time within more elaborate models, or in a real network. The model transmission over disrupted link can be refined and elaborated, e.g., by adding details of the physical radio channel (propagation delay, fading of the radio signal, and interference effects). While our basic model neglects those, we believe that it still captures something essential for understanding the delivery of fragmented messages.

III. Problem statement

If X is an ON epoch's sample, such that $X = kf + r$, where k is a non-negative integer, then node A can transfer k packets of size f or a submessage of size kf within X. The remainder r, which unused for transmission, decreases on the average as f gets smaller. So, on the one hand, finer subdivision of the message should reduce total transmission time over a single link.

But on the other hand, if node B is not the end receiver of A's messages—i.e. in the multi-hop and multi-path cases, then finer subdivision will increase the chance of losing that message, e.g., due to a misrouted or dropped fragment [1]. Coarser subdivision (i.e. larger f) may increase the chance of a message reaching its destination.

Jelenković and Tan showed in [15] that the mean transmission time of unfragmented messages over a disrupted link may become arbitrarily large in some cases, and introduced a fragmentation algorithm to mitigate this problem. Nair et al. study in [16] the asymptotics of message transmission time; they also show how to choose f so that, under certain restrictions on the distribution of ON epochs, the mean transmission time of a message is minimized.

In this paper, we investigate a related, but different problem: We will assume that the cost of transmitting a message of size m, typically increases with the number of fragments into which that message may be split. Our target is to minimize this cost, provided that the message transmission time $t(x)$ does not exceed L. Therefore, we seek the largest possible fragmentation unit f for a given message size m and the constraint L on the message transmission time.

The fragmentation question can be approached in two ways: First, what is the largest f, for which inequality

$$T(x) \leq L, \tag{III.1}$$

is true? This is the problem of in-time delivery in the mean. Estimating $T(x)$, given f and the contact pattern characteristics, is needed in solving this problem.

Second, what is the largest f, for which inequality

$$\mathrm{P}\{t(x) \leq L\} > 1 - \epsilon \tag{III.2}$$

is true? The lower bound $1 - \epsilon$ on the probability of in-time delivery could be, e.g., 0.95. This is the problem of in-time delivery in probability.[5]

In this paper we will deal only with the problem of in-time delivery in the mean.

IV. Additional definitions and notation

The values related to transmission duration, $t(x)$, $T(x)$, $n(x)$, and $N(x)$, depend on the size f of transmitted packets. Therefore, more accurate notation would be to use symbols $t_f(x), T_f(x), n_f(x)$ and $N_f(x)$, but we use the simpler notation in the rest of this paper, because the intended f is clear from the context.

When a message of size m is quantized into blocks of size f its size grows due to headers and, e.g., possibly padding the last piece of data so that its size is f. We denote by x the message size after it has been so processed, and by h the size of the header: $0 \leq h \leq f_0$. For $f = 0$, we set $h = 0$ and define $x = m$ in that case. For $f \geq f_0$, $m = k(f - h) + r$, where k is a non-negative integer and $0 \leq r < f - h$. The quantized message will contain k blocks of size f, where $k = \lfloor \frac{m}{f-h} \rfloor$ if $f \geq f_0$; its total size x depends on how the remainder r is quantized.

As an example, the remaining part may be packaged into a single block of size f:

$$x = kf + f. \tag{IV.1}$$

This quantization simplifies analysis, but is unlikely in practice, unless $f = f_0$. In a more realistic example, the remaining part will be packaged into a single block of size $r + h$:

$$x = kf + r + h. \tag{IV.2}$$

We denote by $j(x)$ the number of blocks in a message of size x: $j(x) = \lceil x/f \rceil$. If the quantizing is according to Eq. (IV.1), then the size of the last block is f; if the quantizing is according to Eq. (IV.2) then the size of the last block is $r + h$.

In most cases[6] only the part $f\lfloor X/f \rfloor$ of an ON epoch may be used for transmission, unless $f = 0$: If $f = 0$, then the whole of an ON epoch X may be used for transmission. We denote this "useful" part by X_f:

$$X_f = \begin{cases} \lfloor \frac{X}{f} \rfloor f & \text{if } f > 0, \\ X & \text{if } f = 0. \end{cases} \tag{IV.3}$$

[5] Our simple modeling of fragmentation penalties is not well-suited for answering minimization questions, e.g., "What f minimizes $t(x)$ with probability $1 - \epsilon$?" and "What f minimizes $T(x)$?" Therefore, we do not address these minimization problems in this paper.

[6] Possible exceptions are the ON epochs in which the remainder $r + h$ is transmitted.

Assuming that all blocks (including the last one) are of equal size $f > 0$, the distribution of X_f is:

$$\begin{aligned} \mathrm{P}\{X_f \le x\} &= \mathrm{P}\{X < (k+1)f\} \\ &= F_X((k+1)f) - \mathrm{P}\{X = (k+1)f\}, \end{aligned} \tag{IV.4}$$

where $kf \le x < (k+1)f$, and $k = 0, 1, 2, \cdots$;[7] its probability mass function is:

$$\begin{aligned} \mathrm{P}\{X_f = kf\} =& F_X((k+1)f) - \mathrm{P}\{X = (k+1)f\} \\ &- (F_X(kf) - \mathrm{P}\{X = kf\}). \end{aligned} \tag{IV.5}$$

We denote by $S_k(X)$ the sum of k independent and identically distributed random variables X_i. For example, $S_0(X)$ is zero, $S_k(Y)$ is the sum $Y_1 + Y_2 + \cdots + Y_k$, and $S_k(X_f)$ is the sum $X_{f,1} + X_{f,2} + \cdots + X_{f,k}$. The variable $n(x)$ can be defined in terms of $S_k(X_f)$: $n(x) = \min\{k : S_k(X_f) \ge x\}$.

If k is large, then by the central limit theorem, $S_k(X_f)$ is distributed approximately normally with mean E$[X_f]k$ and variance Var$[X_f]k$. It can be shown that if $x \gg$ E$[X_f]$, then also $n(x)$ is distributed approximately normally with mean $x/$E$[X_f]$ and variance Var$[X_f]x/$E$[X_f]^3$. (See [19], p. 61 for an outline of a proof.)

But in scenarios where k is small, this result does not apply. This motivates further investigations into the behavior of $n(x)$.

Note also that since x depends on the header size h, all functions of x depend also on the constant h.

V. Estimating the mean transmission time

A. Elementary considerations

When $f = 0$ there is no waste of contact time due to data quantization, and

$$T_X(x) = x. \tag{V.1}$$

Thus, $T_X(x)$ is at least x seconds; it depends on x and on the probability mass function of X_f.

If a value of $N(x)$ is given, then $T_X(x)$ can be approximated as E$[X] \cdot N(x)$. Suppose that the given $N(x)$ is exact. Since the transmission of the entire message will be typically completed somewhere inside the last ON epoch, i.e. before the last ON epoch ends, that approximation overestimates $T_X(x)$ by at most E$[X]$. ($T_X(x)$ can be estimated better in special cases. See section V-E.)

Conversely, if a value of $T_X(x)$ is given, then $N(x)$ can be approximated as $T_X(x)/$E$[X]$. (See section V-B below.)

The expression for $T_Y(x)$ is

$$T_Y(x) = \mathrm{E}[Y]\,(N(x) - 1)\,. \tag{V.2}$$

As mentioned in section II, the number of OFF epochs during successful transmission is one less than the number of ON epochs, due to our assumption that new message transmission starts at the beginning of ON epoch.

The fact that $T_X(x)$ is at least x seconds, together with Eq. (II.1) and (V.2) define the lower bound on $T(x)$. A message whose uninterrupted transmission within a single ON epoch will take x seconds cannot be transmitted over disrupted link in less than $N(x)$ ON epochs and $N(x) - 1$ OFF epochs on the average:

$$T(x) \ge x + \mathrm{E}[Y](N(x) - 1). \tag{V.3}$$

B. Direct approximation of $T_X(x)$ in the general case

A linear approximation for $T_X(x)$ can be derived by beginning from Eq. (V.1); $N(x)$ is then estimated as $T_X(x)/$E$[X]$ and an approximate value of $T(x)$ obtained from Eq. (V.2) and (II.1):

An ON epoch that is less than f cannot be used for transmission; when the message is divided into fragments of size f, the proportion of the total ON time $S_{n(x)} = X_1 + X_2 + \cdots + X_{n(x)}$ that is used for transmitting data is approximately P$\{X_i \ge f\}$, i.e. $1 - F_X(f) +$ P$\{X = f\}$. Therefore, $T_X(x)$ (which equals x when $f = 0$), grows by a factor of $1/$P$\{X \ge f\}$.

If the last fragment of a message is less than f, it may be transmitted in an ON epoch that is less than f: $X_{n(x)} < f$. But as the number of fragments gets larger, the effect of an event $\{X_{n(x)} < f\}$ on $T_X(x)$ gets smaller. Thus, also in this case, $T_X(x)$ grows by a factor of about $1/$P$\{X \ge f\}$ compared to when $f = 0$.

Based on the above, $T_X(x)$ can be estimated as

$$T_X(x) \approx \frac{x}{\mathrm{P}\{X \ge f\}}. \tag{V.4}$$

Dividing the estimate of $T_X(x)$ by the mean length of an ON epoch, we obtain

$$N(x) \approx \frac{x}{\mathrm{E}[X]\,\mathrm{P}\{X \ge f\}}. \tag{V.5}$$

It should be noted that there are two systematic errors in the above approximation which fortunately impact in opposite directions. First, not all of the time of ON epochs longer than f can be used for transmission; in the end of each epoch X_i there is some remainder time $X_i - X_{f,i}$ that is wasted. Neglecting this wasted time leads towards a too high denominator in Eq. (V.4) (and thus a too low estimate of $T_X(x)$).

Second, P$\{X_i \ge f\}$ actually gives the proportion of the *number* of ON epochs that may be used for transmission, not the proportion of the *time* that may be used for transmission. Since an ON epoch that may be used for transmission must be at least f, it is longer than just an arbitrary X_i; the proportion of the time that may be used for transmission is more than P$\{X_i \ge f\}$. This error leads towards a too low denominator in Eq. (V.4) (and thus a too high estimate of $T_X(x)$).

We discuss next how to decrease the effect of those systematic errors. The first error increases the amount of transmission time in any "useful" ON epoch: $X_i \ge f$, by an amount of wasted time $X_i - X_{f,i} \in [0, f)$. The exact amount added depends of course on f and the distribution of X, but because we deal with an approximation here, a good candidate for reducing the systematic error is to substract $f/2$ from each $X_i \ge f$.

The second systematic error can be reduced as follows. The probability that an ON epoch is "useful" is indeed P$\{X_i \ge f\}$;

[7] Note that if F_X is continuous at $(k+1)f$, then $dF_X((k+1)f)$ is finite. Therefore, the term P$\{X = (k+1)f\}$ above, which equals to $\lim_{\Delta x \to 0} \Delta x \cdot dF_X((k+1)f)$, tends to zero and can be neglected.

and a typical $X_i \geq f$ is longer than a typical X_i by a factor of $\mathrm{E}[X \mid X \geq f]/\mathrm{E}[X]$. These two values should now be multiplied in order to get the proportion of the time that may be used for transmitting data in the total time of all ON epochs.

If we now combine the two corrections then the corrective multiplier to the term $\mathrm{P}\{X \geq f\}$ is the ratio of $\mathrm{E}[X \mid X \geq f] - f/2$ and $\mathrm{E}[X]$. Thus, the slightly better approximations are

$$T_X(x) \approx \frac{x}{\mathrm{P}\{X \geq f\} \cdot \frac{\mathrm{E}[X|X \geq f] - f/2}{\mathrm{E}[X]}} \tag{V.6}$$

and

$$N(x) \approx \frac{x}{\mathrm{P}\{X \geq f\}(\mathrm{E}[X \mid X \geq f] - f/2)}. \tag{V.7}$$

Let us illustrate the behavior of the corrective multiplier in Eq. (V.6) with three different distributions of X.

First, when X is uniformly distributed in $(0,1)$, the conditional expected value $\mathrm{E}[X \mid X \geq f]$ is $(f+1)/2$ and the value of $T_X(x)$ estimated with Eq. (6.6) is exactly the same as $T_X(x)$ estimated with Eq. (V.4):

$$T_X(x) \approx \frac{x}{(1-f) \cdot \frac{(f+1)/2 - f/2}{1/2}} = \frac{x}{1-f} = \frac{x}{\mathrm{P}\{X \geq f\}}. \tag{V.8}$$

Second, when X is exponentially distributed with parameter λ, $P\{X \geq x\} = e^{-\lambda x}$ and $\mathrm{E}[X] = 1/\lambda$. After being ON for f seconds, the link will continue to be ON for another $1/\lambda$ seconds on the average, due to the memoryless property of exponential distribution. Thus, $\mathrm{E}[X \mid X \geq f] = f + 1/\lambda$ and

$$T_X(x) \approx \frac{x}{e^{-\lambda f} \frac{f/2 + 1/\lambda}{1/\lambda}} = \frac{x}{e^{-\lambda f}(1 + \frac{\lambda f}{2})} \leq \frac{x}{\mathrm{P}\{X \geq f\}}. \tag{V.9}$$

Third, consider the distribution where $F_X(x) = x^2$, if $x \in [0,1]$ and $F_X(x) = 0$, otherwise. In this case $\mathrm{E}[X] = 2/3$, $\mathrm{P}\{X \geq f\} = 1 - f^2$, and $\mathrm{E}\left[X \mid X \geq f\right\} = (2 + 2f + 2f^2)/(3 + 3f)$.

It can be verified that the corrective multiplier:

$$\frac{\mathrm{E}\left[X \mid X \geq f\right] - f/2)}{\mathrm{E}[X]} = \frac{\frac{2+2f+2f^2}{3+3f} - \frac{f}{2}}{2/3} = \frac{4 + f + f^2}{4 + 4f}, \tag{V.10}$$

monotonically decreases from 1 to $3/4$, as f increases from 0 to 1. So, in this case, $T_X(x)$ estimated with Eq. (V.6) will be more than or equal to $T_X(x)$ estimated with Eq. (V.4).

C. Computation of $N(x)$ in the general case

The discrete *renewal equation*:

$$N(x) = \frac{1 + \sum_{k=1}^{j(x)} N(x - kf)\mathrm{P}\{X_f = kf\}}{1 - \mathrm{P}\{X_f = 0\}}, \tag{V.11}$$

where $N(x) = 0$, if $x \leq 0$, provides a way for estimating $N(x)$ recursively for any $x \geq 0$. It can be derived based on $X_{f,1}$, $X_{f,2}$, etc., being i.i.d.; e.g., by the following method from [20]:

The expectation of $n(x)$ conditioned on the event "the suitable for transmission part of the first ON epoch equals kf", where k is some nonnegative integer, is:

$$\mathrm{E}[n(x) \mid X_f = kf] = 1 + \mathrm{E}[n(x - kf)] = 1 + N(x - kf).$$

Multiplying both sides by the probability $\mathrm{P}\{X_f = kf\}$ of that event and summing for all $k \geq 0$, leads to:

$$\begin{aligned} N(x) &= \sum_{k=0}^{\infty} (1 + N(x - kf))\, \mathrm{P}\{X_f = kf\} \\ &= \sum_{k=0}^{\infty} \mathrm{P}\{X_f = kf\} + \sum_{k=0}^{\infty} N(x - kf)\mathrm{P}\{X_f = kf\} \\ &= 1 + N(x)\mathrm{P}\{X_f = 0\} + \sum_{k=1}^{\infty} N(x - kf)\mathrm{P}\{X_f = kf\}. \end{aligned}$$

Since $N(x) = 0$, if $x \leq 0$, the upper limit of the sum on the right hand side can be truncated to the number of fragments in x: $j(x)$; and Eq. (V.11) results after rearrangement of terms.

When $x \gg \mathrm{E}[X]$ and $f > 0$, it can be shown that $N(x)$ obeys the following asymptotic relation:

$$N(x) \to \frac{d \cdot x}{\mathrm{E}[X_f]} + \frac{\mathrm{Var}[X_f] - \mathrm{E}[X_f]f + \mathrm{E}[X_f]^2}{2\mathrm{E}[X_f]^2}, \tag{V.12}$$

where d is the "period" of X_f. (This relation corresponds to Eq. (12.2), p. 340, in [21].) The period d is a positive integer derived from X_f taking its values from the sequence 0, df, $2df$, $3df$, $\cdots$. The period $d = 1$ if the range of X includes f. If $d > 1$, then there will be periodic gaps between the values of X_f, caused by the corresponding gaps in the values of X. For example, if the range of X is restricted to $(0, 0.9f]$ and $[2.1f, 2.9f]$, then $X_f \in \{0, 2f\}$ and $d = 2$.

$N(x)$ is a non-decreasing function of x when f is fixed. Therefore, a value of $N(x)$ computed assuming that the message is quantized into $j(x)$ fragments of equal size f, according to Eq. (IV.1), provides an upper bound on $N(x)$ in the case that the last message part, that is less than $f - h$, is quantized into a smaller packet.

Similarly, the value of $N(x)$, computed assuming that the message is quantized into $j(x) - 1$ fragments of equal size f, provides a lower bound on $N(x)$ in the above case.[8]

D. Computation of $N(x)$ in special cases

1) Case of $f = 0$: no quantization: The integral renewal equation for the case $f = 0$ that follows, can be derived similarly to the Eq. (V.11) above,

$$N(x) = 1 + \int_{y=0}^{x} N(x - y)dF_X(y). \tag{V.13}$$

[8]In section V-D non-recursive equations for $N(x)$ are given for uniformly and exponentially distributed X, again assuming that the message is quantized into fragments of equal size. It can be verified that a tighter lower bound may be obtained by replacing the integer $j(x) = \lceil x/f \rceil$ with the fraction x/f in those equations.

When $x \gg \mathrm{E}[X]$, it can be shown that $N(x)$ obeys the following asymptotic relation:

$$N(x) \to \frac{x}{\mathrm{E}[X]} + \frac{\mathrm{Var}[X] + \mathrm{E}[X]^2}{2\mathrm{E}[X]^2}. \tag{V.14}$$

(This relation corresponds to Eq. (3.1), p. 366, in [22].)

The linear form of (V.14) can be explained intuitively as follows. When $x \gg \mathrm{E}[X]$, $n(x)$ is large. Therefore, first, the sum of the ON epochs that are needed to transmit a (large) message is quite close to the size of that message: $S_{n(x)}(X) \approx x$, because the remainder after the end of transmission of the last ON epoch, $X_{n(x)}$, is insignificant compared to x. Second, the many random fluctuations in X_i around its mean $\mathrm{E}[X]$, during transmission of a single large message, cancel each other. This results in that the mean difference between $S_{n(x)+1}(X)$ and $S_{n(x)}(X)$ is $\mathrm{E}[X]$, which, in turn, implies that the mean growth rate of $n(X)$ as a function of $S_{n(x)}(X)$ is $1/\mathrm{E}[X]$.

2) Case of $f = m + h$: no fragmentation: When $f = m + h$, i.e. there is no fragmentation and $x = f$, an ON epoch X cannot be used for transmission if $X < x$. Setting $j(x) = 1$ in Eq. (V.11) we obtain immediately:

$$N(x) = \frac{1 + N(0)\mathrm{P}\{X_f = f\}}{1 - \mathrm{P}\{X_f = 0\}} = \frac{1}{\mathrm{P}\{X \geq x\}}. \tag{V.15}$$

For example, if X is distributed uniformly between zero and one, then $F_X(x) = x$ for $0 < x < 1$, and is 1 for $x \geq 1$. Thus, by Eq. (V.15),

$$N(x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < 1, \\ \infty & \text{if } x \geq 1. \end{cases} \tag{V.16}$$

3) Case of exponentially distributed X: It can be verified that if $F_X(x) = 1 - e^{-\lambda x}$, then

$$N(x) = j(x)(e^{\lambda f} - 1) + 1. \tag{V.17}$$

Taking the limit at $f \to 0$ of $N(x)$, we get

$$N(x) = \lambda x + 1. \tag{V.18}$$

Eq. (V.17) can be derived as follows. By (IV.5), the probability $\mathrm{P}\{X_f = kf\}$ is $(1-p)^k p$, where $p = 1 - e^{-\lambda f}$: the number of successfully transmitted blocks of size f within a single ON epoch is distributed geometrically with parameter p. Thus, the count of failures in transmitting a block before successful transmission of $j(x)$ blocks (in a sequence of ON epochs) has a negative binomial distribution with parameters $j(x)$ and p. Its expectation $j(x)p/(1-p)$ equals to the mean number of OFF epochs during a successful transmission of $j(x)$ blocks. Therefore, $N(x) = j(x)p/(1-p) + 1$.

4) Case of uniformly distributed X: We will give below an expression for $N(x)$ in two sub-cases. In both of them $X \sim \mathrm{U}(0, 1)$ and the message is quantized into fragments of equal size, according to Eq. (IV.1).

Sub-case of $1/2 < f < 1$: In this case at most one fragment may be transmitted within any ON epoch; a message requires $j(x)$ uninterrupted transmissions of f seconds each. Therefore,

$$N(x) = j(x)N(f) = j(x)/(1-f). \tag{V.19}$$

Sub-case of $f = 1/n$, and $n = 2, 3, \cdots$: In this case,

$$N(x) = \sum_{k=0}^{q} (-1)^k \left(\frac{n}{n-1}\right)^{j(x)-kn} \frac{(j(x)-kn)^{\overline{k}}}{k!(n-1)^k}, \tag{V.20}$$

where $N(x) = 0$ if $x \leq 0$, $(x)^{\overline{k}}$ is the rising factorial $x(x+1)\cdots(x+k-1)$, and q is the integer part of x: $q = \lfloor x \rfloor$. This formula will be proved by induction on $j(x)$ in Appendix B.

$N(x)$ in this case is related to the mean of the minimum number of throws with a n-faceted die such that their sum is at least x. The distribution of the minimum number of throws is given, e.g., in ex. 19, p. 285 of [21]. But the closed form expression for the mean seems to be new; so far, we have not found it in the literature.

As f gets smaller, the values of $N(x)$ computed with Eq. (V.20) approach from above those of $N(x)$ with $f = 0$:

$$N(x) = \sum_{k=0}^{q} (-1)^k e^{x-k} \frac{(x-k)^k}{k!}. \tag{V.21}$$

(Eq. (V.21) is given in [22] p. 385.)

E. Computation of $T_X(x)$ in special cases

In general, the mean of the total time spent in ON state $T_X(x)$ can be approximated as $\mathrm{E}[X] \cdot N(x)$. We will see below that under conditions that enforce transmission of at most one fragment of size f per ON epoch, a better estimate of $T_X(x)$ may be obtained.

1) Case of $f = m+h$: no fragmentation: We can get a better estimate by noticing that in this case (i) the whole message must be transmitted in x seconds during the last ON epoch and that (ii) none of the $n(x) - 1$ ON epochs preceding successful transmission may equal or exceed x. Therefore,

$$T_X(x) = x + (N(x) - 1)\mathrm{E}[X \mid X < x]. \tag{V.22}$$

For example, if $0 < x < 1$ and X is distributed uniformly in $(0, 1]$, then inserting $\mathrm{E}[X \mid X < x] = x/2$ and the expression for $N(x)$ from Eq. (V.16) into Eq. (V.22) we obtain

$$T_X(x) = x + (N(x) - 1)\frac{x}{2} = \frac{1}{2}\left(\frac{1}{1-x} - (1-x)\right). \tag{V.23}$$

As another example, if X is distributed exponentially, then

$$\mathrm{E}[X \mid X < x] = \frac{\int_0^x y\lambda e^{-\lambda y} dy}{\mathrm{P}\{X < x\}} = \frac{1}{\lambda} - \frac{x}{e^{\lambda x} - 1}.$$

Inserting the above and the expression for $N(x)$ from Eq. (V.17) with $j(x)$ set to 1 into Eq. (V.22) we obtain

$$T_X(x) = \frac{1}{\lambda}(e^{\lambda x} - 1). \tag{V.24}$$

2) Case of $X \sim U(0,1)$ and $1/2 < f < 1$: The value of $N(x)$ in this case is given by Eq. (V.19). The value of $T_X(x)$ is given by

$$T_X(x) = \frac{1}{2}\left(\frac{j(x)}{1-f} - (1-f)\right). \qquad \text{(V.25)}$$

This equation can be derived from the more general one:

$$\begin{aligned} T_X(x) &= j(x)\mathrm{E}[(X \mid X < f)](N(f)-1) \\ &\quad + (j(x)-1)\mathrm{E}[X \mid X \geq f] + f, \end{aligned} \qquad \text{(V.26)}$$

that is valid when the distribution of ON epochs $F_X(x)$ vanishes outside an interval $[a,b]$, and f exceeds $b/2$. (Those conditions enforce transmission of at most one fragment of size f per ON epoch.) Eq. (V.26) can be constructed using the following observations: (i) the mean length of the $N(f-h)-1$ ON epochs preceding successful transmission of a single fragment is $\mathrm{E}[X \mid X < f]$; (ii) $\mathrm{E}[X \mid X \geq f]$ is the mean length of the ON epoch during which f is successfully transmitted; and (iii) the transmission ends after f seconds of $X_{n(x)}$.

F. Summary of results on mean transmission time

The answer to the basic question "What is the smallest message size that has to be fragmented to be delivered in-time on the average?" can be computed in our model using the combination of Eq. (II.1), (V.2), (V.15), and (V.22).

Intuitively, if the message fits into a "typical" contact time, then fragmentation of that message is not necessary for its in-time delivery; and if the message is much larger than a "typical" contact time, then that message must be fragmented to be delivered in-time. Those simple facts are confirmed by our theoretical calculations of the mean transmission time of messages that have been quantized into blocks of size f. (Cf. Eq. (V.16) and (V.20)).

We have also found that if f is held constant, then for large messages the mean delivery time grows approximately linearly with the message size (see Eq. (V.12)), while for smaller messages that growth is non-linear. The speed of that growth depends on the constant f: as f decreases, so does the speed, until at $f = 0$ the mean transmission time approaches its theoretical lower limit, where no transmission time is lost due to quantization of data into discrete blocks; and link disruptions are the only remaining obstacle to in-time delivery.

It is plain that if we keep the message size constant while decreasing f, then the gross message size will increase due to addition of extra headers. We have also found that even if we neglect the effect of extra headers by setting $h = 0$, constant independent of f, then the decrease in transmission time as we decrease f is not necessarily monotonic: for some message sizes, decreasing f may actually increase the transmission time (see Appendix A).

For reference, we summarize methods for estimating the mean transmission time in Table I. The leftmost column in the summary contains the number of a section in which the corresponding expression has been introduced.

In subtable on uniformly distributed X, formulae for X uniformly distributed in $(0,a]$, where $a > 0$, rather than in $(0,1]$, are given.

VI. Summary and future work

In this paper we have investigated how to quantize a message into blocks of size f before its transmission in networks with unstable links. It is assumed that in-time delivery of a message is what matters to applications, and that the cost of transmitting a message grows with the number of pieces into which it is split.

Thus, we seek the largest f that will (if that is possible) guarantee in-time delivery of a message. The guarantee can be either in terms of in-time delivery in the mean, or in terms of lower bound on the probability of in-time delivery. We deal with the first of those subjects in this paper and the question of how to compute the probability of in-time delivery from link statistics for a given message size and f could be a topic of furture work.

In this paper we have shown how to analytically estimate the mean delivery time of fragmented message within a basic model of transmission over disrupted link. We use those theoretical results to understand better the fundamentals of fragmented message transmission (see § V-F), and as a source of approximations in our fragmentation algorithms [13]. Estimating message transmission times within more refined and elaborate models could be a topic of future work. Other topics for future work include fragmentation in multi-link and multi-path networks. Further studies could also consider the impact that quantization of a message has on the in-time delivery of subsequent (queued) messages.

In conclusion, there are inherent non-linear dependencies (some obvious and some not, see § V-F) of transmission time of a message that has been quantized into blocks of size f, on the size of that message and on f. Those non-linear dependencies make the design of fragmentation algorithms interesting and challenging.

Acknowledgments

The authors thank J. Virtamo and S. Aalto for help, and P. Thiran, M. Grossglauser and M. Rinne for comments.

This work was supported by TEKES as part of the Future Internet program of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT).

References

[1] M. Pitkänen, A. Keränen, and J. Ott, "Message fragmentation in opportunistic DTNs," in *Proceedings of the Second WoWMoM Workshop on Autonomic and Opportunistic Communications (AOC)*, 2008.

[2] C. J. Bennett, "The overheads of transnetwork fragmentation," *Computer Networks*, vol. 6, no. 1, pp. 21–36, 1982.

[3] M. Bonuccelli, "Minimum fragmentation internetwork routing," *INFOCOM '91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s., IEEE*, pp. 289–294 vol.1, Apr 1991.

[4] A. Orda and R. Rom, "Optimal packet fragmentation and routing in computer networks," *Networks*, vol. 29, no. 1, pp. 11–28, 1997.

[5] C. A. Kent and J. C. Mogul, "Fragmentation considered harmful," in *SIGCOMM '87: Proceedings of the ACM workshop on Frontiers in computer communications technology*. New York, NY, USA: ACM, 1988, pp. 390–401.

[6] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H.Weiss, "Delay-Tolerant Network Architecture," RFC 4838, April 2007.

[7] K. Scott and S. Burleigh, "Bundle Protocol Specification," RFC 5050, November 2007.

Key notation

f	the size of message fragmentation unit.
x	the size of the message with additional headers (due to quantization) included.
X, (Y)	length of an ON (respectively OFF) epoch.
X_f	the part of an ON epoch that may be used for transmission when the fragmentation unit's size is f; see Eq. (IV.3).
$T(x)$	mean time needed to transmit x; $T_X(x)$ (resp. $T_Y(x)$) is the part of $T(x)$ spent in ON (resp. OFF) state.
$N(x)$	mean number of ON epochs during transmission of x.

Summary for generally distributed X (and Y)

§	f	Results	Notes
II V-A		$T(x) = T_X(x) + T_Y(x)$, $T_X(x) \geq x$.	
		$T_X(x) \approx \mathrm{E}[X]N(x)$, $T_Y(x) \approx \mathrm{E}[Y]\,(N(x)-1)$.	Approximation of $T_X(x)$ and $T_Y(x)$ from $N(x)$.
V-B		$T_X(x) \approx x/\mathrm{P}\{X \geq f\}$, $N(x) \approx T_X(x)/\mathrm{E}[X]$.	Direct approximation of $T_X(x)$.
V-B		$T_X(x) \approx x \Big/ \left(\mathrm{P}\{X \geq f\} \cdot \frac{\mathrm{E}[X \mid X \geq f] - f/2}{\mathrm{E}[X]}\right)$.	Direct approximation of $T_X(x)$ with corrective multiplier.
V-C	$f > 0$	$N(x) = \frac{1+\sum_{k=1}^{j(x)} N(x-kf)\mathrm{P}\{X_f = kf\}}{1-\mathrm{P}\{X_f=0\}}$. $N(x) \to \frac{d \cdot x}{\mathrm{E}[X_f]} + \frac{\mathrm{Var}[X_f] - \mathrm{E}[X_f]f + \mathrm{E}[X_f]^2}{2\mathrm{E}[X_f]^2}$.	Discrete renewal equation, $j(x) = \lceil \frac{x}{f} \rceil$, $N(x) = 0$, if $x \leq 0$. $x \to \infty$. $X_f \in \{0, df, 2df, \cdots\}$. Corresponds to Eq. (12.2) in [21], p. 340.
V-D1	$f = 0$	$T_X(x) = x$ $N(x) = 1 + \int_{y=0}^{x} N(x-y)dF_X(y)$, $N(x) \to \frac{x}{\mathrm{E}[X]} + \frac{\mathrm{Var}[X] + \mathrm{E}[X]^2}{2\mathrm{E}[X]^2}$.	No quantization. Renewal equation. $x \to \infty$. See [22], p. 366, Eq. (3.1).
V-D2 V-E	$f = x$	$N(x) = 1/\mathrm{P}\{X \geq x\}$, $T_X(x) = x + (N(x)-1)\,\mathrm{E}[X \mid X < x]$.	No fragmentation.

Summary for exponentially distributed X

§	f	Results	Notes
V-D3	$f > 0$	$N(x) = j(x)(e^{\lambda f} - 1) + 1$.	$j(x) = \lceil \frac{x}{f} \rceil$.
V-D3	$f = 0$	$N(x) = \lambda x + 1$.	No quantization
V-D2 V-E1	$f = x$	$N(x) = e^{\lambda x}$, $T_X(x) = \frac{1}{\lambda}(e^{\lambda x} - 1)$.	No fragmentation.

Summary for X uniformly distributed in $[0, a]$

§	f	Results	Notes
V-D4	$f = \frac{a}{n}$, $n \geq 2$	$N(x) = \sum_{k=0}^{q} (-1)^k \left(\frac{n}{n-1}\right)^{j(x)-kn} \frac{(j(x)-kn)^{\overline{k}}}{k!(n-1)^k}$.	$N(x) = 0$ if $x \leq 0$, $q = \lfloor x/a \rfloor$, $j(x) = \lceil \frac{x}{f} \rceil$, $(x)^{\overline{k}} = x(x+1)\cdots(x+k-1)$.
V-D4	$f = 0$	$N(x) = \sum_{k=0}^{q} (-1)^k e^{x-k} \frac{(x-k)^k}{k!}$.	No quantization, see [22], p. 385.
V-D2 V-E1	$f = x$	$N(x) = 1/(1 - x/a)$, $T_X(x) = \frac{1}{2}\left(\frac{1}{1-x/a} - (1 - x/a)\right)$.	No fragmentation.
V-D4 V-E2	$\frac{a}{2} < f < a$	$N(x) = j(x)/(1 - f/a)$, $T_X(x) = \frac{1}{2}\left(\frac{j(x)}{1-f/a} - (1 - f/a)\right)$.	At most one fragment per ON epoch, $j(x) = \lceil \frac{x}{f} \rceil$.

TABLE I

SUMMARY OF RESULTS ON MEAN TRANSMISSION TIME.

[8] S. Farrell, S. Symington, H. Weiss, and P. Lovell, "Delay-Tolerant Networking Security Overview," Internet Draft draft-irtf-dtnrg-sec-overview-06.txt, Work in progress, May 2009.
[9] C. Partridge, "Authentication for fragments," in *Proc. ACM SIGCOMM HotNets-IV workshop*, 2005.
[10] N. Asokan, K. Kostiainen, P. Ginzboorg, J. Ott, and C. Luo, "Applicability of identity-based cryptography for disruption-tolerant networking," Nokia Research Center, Helsinki, Tech. Rep., March 2007, available at: http://research.nokia.com/files/NRC-TR-2007-007.pdf.
[11] S. Jain, K. Fall, and R. Patra, "Routing in a Delay Tolerant Network," in *Proceedings of the ACM SIGCOMM*, 2004.
[12] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and Wait: An Efficient Routing Scheme for Intermittently Connected Mobile Networks," in *Proceedings of ACM SIGCOMM Workshop on Delay-Tolerant Networking (WDTN)*, 2005.
[13] P. Ginzboorg, V. Niemi, and J. Ott, "Message fragmentation algorithms for disrupted links," in preparation.
[14] ——, "Message fragmentation in disruptive networks," Nokia Research Center, Tech. Rep., July 2009. http://research.nokia.com/files/NRCTR2009003.pdf.
[15] P. R. Jelenković and J. Tan, "Dynamic packet fragmentation for wireless channels with failures," *MobiHoc'08*, pp. 73–82, May 2008.
[16] J. Nair, M. Andreasson, L. Andrew, S. Low, and J. Doyle, "File fragmentation over an unreliable channel," in *INFOCOM, 2010 Proceedings IEEE*, 14-19 2010, pp. 1 –9.
[17] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms," in *Proc. IEEE INFOCOM*, April 2006.
[18] T. Karagiannis, J.-Y. L. Boudec, and M. Vojnovic, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. ACM MobiCom*, E. Kranakis, J. C. Hou, and R. Ramanathan, Eds., September 2007, pp. 183–194.
[19] R. G. Gallager, *Discrete Stochastic Processes*. Kluwer, 1996.
[20] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*. Academic Press, 1984.
[21] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. Wiley, 1968, vol. 1.
[22] ——, *An Introduction to Probability Theory and Its Applications*, 2nd ed. Wiley, 1971, vol. 2.

Appendix A
Non-monotonicity of transmission time

If we neglect the effect of extra headers, and assume that the gross size of transmitted data is a constant x independent of f, then $t(x)$ will in general decrease as we decrease f. But even under this assumption, the decrease in transmission time $t(x)$ is not necessarily monotonic.

We give a toy example to illustrate this. Assume that $x = 30$ and the first five ON epochs are 10, 7, 5, 11, and 10 seconds with OFF epochs of 4 seconds in between them. We tabulate below the values of $t(x)$ for several f in this example:

f	2	3	4	5	6
$t(x)$	44	52	57	44	> 59

It would be an interesting combinatorial problem to characterize the cases where $t(x)$ decreases monotonically when f decreases.

Also $N(x)$ (and hence the mean transmission time $T(x)$), is not necessarily monotonic in f, even if we neglect the effect of extra headers. For example, if X is distributed uniformly in [0,1], $f = 1/k$ and $x \leq 1$, then by Eq. (V.20):

$$N(x) = (k/(k-1))^{j(x)}.$$

Assuming that x is slightly less than $1/2$, e.g., $1/2 - 1/100$, we tabulate below the values of $j(x)$ and $N(x)$ for several f:

f	$1/2$	$1/3$	$1/4$	$1/5$	$1/6$	$1/7$
$j(x) = \lceil x/f \rceil$	1	2	2	3	3	4
$N(x)$	2	2.25	1.78	1.95	1.73	1.85

We see that a similar non-monotonicity phenomenon appears again. However, the transmission time is monotonic in x when the size of f is fixed in our model: as we increase message size while keeping f fixed, the transmission time never gets lower.

The effect of data quantization described above implies that even if $T(x)$ could be estimated without error as a function of f, the difference between the optimal f (i.e. the largest f such that $T(x) \leq L$), and f found by an algorithm that examines the possible values of f with search step Δ, may exceed Δ.

Appendix B
$N(x)$ in the discrete uniform case

In this section we provide an inductive proof for the Eq. (V.20). Our proof uses the fact that a valid formula for $N(x)$ must also satisfy the recursion of Eq. (V.11).

Base case: Let us begin by denoting $N(x) = u_j$ when $x = jf$ (and thus $j(x) = j$). Because $\mathrm{P}\{X_f = kf\} = 1/n$ for each $k = 0, 1, \cdots, n-1$ and is zero for other values of k, the formula (V.11) becomes

$$u_j = \frac{1 + (u_{j-1} + u_{j-2} + \cdots + u_{j-(n-1)})\frac{1}{n}}{1 - 1/n}, \tag{B.1}$$

with $u_j = 0$ if $j \leq 0$. Next, the corresponding formula for u_{j-1} can be formally derived by change of indexes. The difference between u_j and u_{j-1} is

$$u_j - u_{j-1} = \frac{(u_{j-1} - u_{j-n})\frac{1}{n}}{1 - 1/n}.$$

Moving u_{j-1} to the right hand side and simplifying gives

$$u_j = \frac{nu_{j-1} - u_{j-n}}{n-1}. \tag{B.2}$$

Now, $u_{j-n} = 0$ if $0 < j \leq n$; for this range of j (B.2) reduces to $u_j = (n/(n-1))u_{j-1}$. Also, $u_1 = n/(n-1)$ by (B.1). Therefore, $u_j = (n/(n-1))^j$ if $0 < j \leq n$; and since Eq. (V.20) agrees with these values when $q = 0$, we have a starting point for our induction.

Inductive step: In order to make our subsequent derivations a little bit simpler we first re-write (V.20) as

$$N(x) = u_j = u_{qn+i} = \sum_{k=0}^{q} (-1)^k \left(\frac{n}{n-1}\right)^{(q-k)n+i} \frac{((q-k)n+i)^{\overline{k}}}{k!(n-1)^k}, \tag{B.3}$$

where $x = (qn+i)/n$ with $0 \leq i < n$.

Let us next assume that (B.3) is valid whenever $j = qn + i$ is replaced by a smaller value (inductive hypothesis) and we try to prove that (B.3) is also true for the value $j = qn + i$.

Now we have $x = (qn+i)/n$ where $0 \leq i < n$. Let us first consider the case where $0 < i < n$. Our strategy is to use both the inductive hypothesis and the recurrence equation (B.2).

First, by applying the inductive hypothesis twice, we get

$$\frac{n}{n-1}u_{qn+i-1} = \sum_{k=0}^{q}(-1)^k\left(\frac{n}{n-1}\right)^{(q-k)n+i}\frac{((q-k)n+i-1)^{\overline{k}}}{k!(n-1)^k}, \tag{B.4}$$

and

$$\frac{u_{(q-1)n+i}}{n-1} = \sum_{k=0}^{q-1}(-1)^k\left(\frac{n}{n-1}\right)^{(q-1-k)n+i}\frac{((q-1-k)n+i)^{\overline{k}}}{k!(n-1)^{k+1}}. \tag{B.5}$$

By replacing k with $c-1$ (i.e. $k+1=c$) we can re-write the last formula as

$$\frac{u_{(q-1)n+i}}{n-1} = \sum_{c=1}^{q}(-1)^{c-1}\left(\frac{n}{n-1}\right)^{(q-c)n+i}\frac{((q-c)n+i)^{\overline{c-1}}}{(c-1)!(n-1)^c}.$$

Renaming the index c now by k and substituting to the recurrence equation (B.2) we obtain

$$\begin{aligned} u_{qn+i} &= \frac{nu_{qn+i-1}-u_{(q-1)n+i}}{n-1} \\ &= \left(\frac{n}{n-1}\right)^{qn+i} + \sum_{k=1}^{q}(-1)^k\left(\frac{n}{n-1}\right)^{(q-k)n+i}\cdot\alpha, \end{aligned} \tag{B.6}$$

where

$$\alpha = \left(\frac{((q-k)n+i-1)^{\overline{k}}}{k!(n-1)^k} + \frac{((q-k)n+i)^{\overline{k-1}}}{(k-1)!(n-1)^k}\right).$$

The α multiplier can be further simplified, firstly by observing that

$$\alpha = \frac{((q-k)n+i-1)^{\overline{k}} + k((q-k)n+i)^{\overline{k-1}}}{k!(n-1)^k}.$$

Secondly, denoting $(q-k)n+i$ by β, we can reduce the numerator of α: $(\beta-1)^{\overline{k}} + k\beta^{\overline{k-1}} = (\beta-1+k)\beta^{\overline{k-1}} = \beta^{\overline{k}}$, i.e. the numerator is simply $((q-k)n+i)^{\overline{k}}$.

Combining these derivations into Eq. (B.6) we get the wanted formula

$$\begin{aligned} u_{qn+i} &= \left(\frac{n}{n-1}\right)^{qn+i} \\ &\quad + \sum_{k=1}^{q}(-1)^k\left(\frac{n}{n-1}\right)^{(q-k)n+i}\frac{((q-k)n+i)^{\overline{k}}}{k!(n-1)^k} \\ &= \sum_{k=0}^{q}(-1)^k\left(\frac{n}{n-1}\right)^{(q-k)n+i}\frac{((q-k)n+i)^{\overline{k}}}{k!(n-1)^k}; \end{aligned}$$

and so Eq. (B.3) is valid in the case where $0 < i < n$.

The case of $i = 0$ is only slightly different. Eq. (B.3) and (B.5) remain the same even in case of $i = 0$, but in the Eq. (B.4) the summation is only for $k = 0, \ldots, q-1$. However, when $i = 0$ and $k = q$ the term $((q-k)n+i-1)^{\overline{k}} = (-1)^{\overline{k}}$ in (B.4) equals zero. Therefore, the term for $k = q$ may be added to the sum, and the case of $i = 0$ can be covered by exactly the same derivation steps.

Hence, our proof is completed.