データ長が確率分布に従う場合の最適ブロック長の考察 第2報

3N−3

小池 慎一　　安井 一民　　中川 覃夫
愛知工業大学

1．はしがき

通信回線を介してデータ伝送を行う方法である基本型データ伝送手順[1]（以下ベーシック手順と略す）において，伝送データ長の確率分布が与えられた場合のブロック長の決定方法について論ずる．

データ伝送におけるビット誤り率BER(bit error rate)の値は10^{-4}から10^{-6}程度である[2]．ブロック長との関係についは、BER値の大きい場合にはブロック長を短くしてブロック誤り確率を小さくすし，BER値が小さい場合には，ブロック誤り確率が小さいのでブロック長を長くした方が効率は良いことが知られている．

2．ベーシック手順

ベーシック手順における伝送の基本単位は文字であり，データは通常ブロックに細分され，ブロック毎にブロックの先頭を示す文字や付加的な情報，データ本体，誤り制御文字，およびブロックの終わりを表す文字等からなる．さらに、誤りなく受信したか否かを返送する文字を必要とする。

これを，簡単のためにデータ本体とその他の部分に分けて次のようにモデル化する．

ブロック＝データ＋制御・応答文字

データx文字，制御・応答文字などの付加的な文字数をεとする．1個のブロックを送信するための文字数は$x+\varepsilon$となる．

伝送中に文字を構成するビット値が反転するなどの誤りが発生することがある．その確率を誤字率という．誤字率は文字ごとに互いに独立であると仮定する．

ベーシック手順では，制御・応答文字部に生じた誤りはブロック長の誤認など致命的な障害を引き起こし，再送では回復できない．タイムアウトなど別の観点から対応する．データ部の誤りについては再送により回復可能である．ここでは，後者の誤りのみを仮定する．

1ブロックの伝送におけるブロック誤り確率をqとすると，データ長はxであることから

$$q = 1 - (1-p)^{x} \tag{1}$$

を得る．再送は誤りが検出されなくなるまで何回でも実行されると仮定すると平均伝送ブロック長は

$$\begin{aligned} E(x) &= (x+\varepsilon)(1-q)[1+2q+3q^2+4q^3+\ldots] \\ &= (x+\varepsilon))(1-p)^{-x} \end{aligned} \tag{2}$$

となる．

3．データ長が一様分布に従う場合[4]

伝送すべきデータがいろいろな長さからなる非定型のテキスト型文書等の場合には，その長さは不定であるが，一般的にはある確率分布に従うと考えられる．本報では伝送データが区間$[L_0-\sqrt{3}\sigma, L_0+\sqrt{3}\sigma]$の一様分布に従う場合について検討する．ここに，分布の平均はL_0，分散はσ^2である．

簡単のため，$\sigma=L_0/\sqrt{3}$とする．すなわち，区間$[0, 2L_0]$の一様分布を考える．

3．1　方式1　分割しない方法

与えられたデータをそのまま1ブロックとして伝送する．これを，方式1とする．

データ長を確率変数Lで表す．1ブロック伝送する場合の平均長に必要とする平均文字数を平均伝送データ長と呼び，それを$E_t(L;p)$とおくと

$$\begin{aligned} E_t(L;p) &= \int_{L_1}^{L_2} (L+\varepsilon)(1-p)^{-L} dL/2\sqrt{3}\sigma \\ &= [(L_2+\varepsilon-1/a)\exp(L_2 a) - (L_1+\varepsilon-1/a)\exp(L_1 a)] \\ &\qquad /(2\sqrt{3}\sigma a) \end{aligned} \tag{3}$$

ここに

$$a = -\ln(1-p) \tag{3)'$$

$$L_1 = L_0-\sqrt{3}\sigma,\ L_2 = L_0+\sqrt{3}\sigma \tag{3)''$$

とおいた．

この場合の$E_t(L;p)$の値を表1に示す。誤字率pと平均データ長L_0が大きくなると，$E_t(L;p)$の値は極端に大きくなることがわかる．すなわち、再送の繰り返し数が大きくなり、現実には伝送が不可能であることを現している。

3．2　方式2　分割して伝送する

一般にデータ長が大きい場合には，データを細分して，ブロックサイズを小さくして伝送した方が平均伝送データ長が短くなることが知られている[3]．そこで，前節で示されたことを避けるために，データをN等分して伝送する．これを，方式2とする．

伝送すべきデータ長をLとするとブロック当たりの文字数はL/Nであるので

$$\begin{aligned} E_t(L;N,p) &= \int_{L_1}^{L_2} N(L/N+\varepsilon)(1-p)^{-L/N} dL/2\sqrt{3}\sigma \\ &= N^2[T_1-T_2]/(2\sqrt{3}\sigma a) \end{aligned} \tag{4}$$

ここに

Optimal Block Length for Basic Mode with uniformly distributed data

Shin-ichi Koike, Kazumi Yasui ,Toshio Nakagawa
Aichi Institute of Tecnology

T_1= $[L_2/N+\varepsilon-1/a]\exp(L_2a/N)$,
T_2= $[L_1/N+\varepsilon-1/a]\exp(L_1a/N)$ (4)'
とおいた．a,L1,L2は(3)',(3)"式で与えられたものである．

なお，N=1とおけば，方式１の式(3)となる．

最適分割数Nを数値的に求めたものを，表２に示す。分割したことによる改善効果は明らかであり，誤字率が10^{-5}から10^{-3}と悪化しても、平均伝送長$E_t(L;N,p)$は20-30%しか増加しない。もちろん伝送可能である。

４．誤字率が変動する場合

方式２で分割して伝送することにしても、現実の回線の状態は変動するものであるので、表２の結果をそのまま利用することは有用とは言えない．

たとえば、誤字率が10^{-5}であるして設計したが，実際には10^{-3}に変動していたとすると，やはり再送の繰り返しが生じ、実際には伝送できないであろう．したがって、誤字率の変化をも含めた方策が必要とされる．

回線の状態が良好な場合の誤字率をp_0，変動して悪化した場合の誤字率をp_1とする。rを誤字率がp_0をとる割合、1-rを誤字率がp_1をとる割合とする。この場合の平均伝送長を$E_r(L;N)$とおくと

$$E_r(L;N) = rE_t(L;N,p_0) + (1-r)E_t(L;N,p_1) \quad (6)$$

となる．

p_0,p_1とrを与えた場合の最適な分割数を数値的に求めたものを表３．１，３．２に示す．

５．数値的検討

前節で述べた方式について数値的な検討をしよう．

表１から，データ長が長い場合に細分しないで伝送すると、誤字率が10^{-3},L_0=2048においてさえ、平均伝送長が$L_0+\varepsilon$の22倍程度となる．すなわち、再送回数が20回を越し、実用的な方策とは言いがたい．

表２により長いデータを分割数Nで細分して送れば、誤字率が10^{-3},L_0=2048においてさえ，平均伝送長は$L_0+\varepsilon$の120%程度であり実用的な値を得る．

表３．１および３．２からは、設計品質として$p_0=10^{-5}$，1-rの割合で$p_1=10^{-3}$などに悪化することを想定した分割数を与える．L_0の増大、rの減少に従い分割数Nは増加するが，平均伝送長は$L_0+\varepsilon$の1.1倍程度に抑えられる．この場合には，誤字率の変動はおりこみ済みなので、この値が保証値と考えられる．

６．むすび

データ長が一様分布に従う場合のベーシック手順によるデータ伝送における最適なブロック長の定め方ついて，誤字率が変動を考慮して，分割数Nを求めた．

表３の値は誤字率が$p=p_0$と一定と考えた場合の値よりは悪くなるが、誤字率は変動するものであり、この方策が現実的である．

参 考 文 献

(1) JIS C 6362(1975) 基本型データ伝送制御手順
(2) 松崎武夫，坂井隆明：情報通信とシステム，東京電機大学 (1974)
(3) 小池、安井、中川：データ伝送における最適なブロック長について、電子情報通信学会、R90-49,pp47-51,(1991)
(4) 小池，安井，中川：ベーシック手順におけるブロック長決定の一考察，電子情報通信学会，R91-47,pp1-6,(1992)

表１ 方式１ 分割しない

p		10^{-5}	10^{-4}	10^{-3}
ε	L_0	$E_t(L;p)$	$E_t(L;p)$	$E_t(L;p)$
8	1,024	1,046	1,184	4,485
	2,048	2,113	2,713	45,864
	4,096	4,335	7,213	3,186,342
	8,192	9,153	26,180	$12\cdot10^9$
16	1,024	1,054	1,193	4,512
	2,048	2,121	2,723	45,980
	4,096	4,343	7,225	3,189,881
	8,192	9,162	26,201	$12\cdot10^9$

表２ 方式２ N分割する

p		10^{-5}		10^{-4}		10^{-3}	
ε	L_0	N	$E_t(L;N,p)$	N	$E_t(L;N,p)$	N	$E_t(L;N,p)$
8	1,024	1	1,046	4	1,092	14	1,250
	2,048	3	2,091	9	2,185	28	2,500
	4,096	5	4,181	17	4,369	56	5,001
	8,192	11	8,362	34	8,739	112	10,002
16	1,024	1	1,054	3	1,121	10	1,353
	2,048	2	2,109	6	2,243	20	2,706
	4,096	4	4,217	12	4,486	41	5,411
	8,192	8	8,434	24	8,972	81	10,822

表３．１ 誤字率に変動がある場合(1) $p_0=10^{-5}$，$p_1=10^{-3}$

r		0.9		0.95		0.99	
ε	L_0	N	$E_r(L;N)$	N	$E_r(L;N)$	N	$E_r(L;N)$
8	1,024	5	1,100	4	1,081	2	1,058
	2,048	10	2,201	8	2,163	5	2,115
	4,096	20	4,401	15	4,325	9	4,229
	8,192	40	8,803	31	8,650	18	8,459
16	1,024	4	1,136	3	1,108	2	1,074
	2,048	8	2,272	6	2,216	4	2,148
	4,096	15	4,543	12	4,433	7	4,293
	8,192	30	9,086	23	8,865	14	8,586

表３．２ 誤字率に変動がある場合(2) $p_0=10^{-5}$，$p_1=10^{-4}$

r		0.9		0.95		0.99	
ε	L_0	N	$E_r(L;N)$	N	$E_r(L;N)$	N	$E_r(L;N)$
8	1,024	2	1,054	2	1,050	1	1,048
	2,048	4	2,107	3	2,100	3	2,093
	4,096	7	4,215	7	4,200	6	4,185
	8,192	15	8,429	13	8,399	11	8,370
16	1,024	1	1,068	1	1,061	1	1,056
	2,048	3	2,133	2	2,122	2	2,111
	4,096	5	4,265	5	4,244	4	4,223
	8,192	11	8,530	9	8,486	8	8,445