ON THE DERIVATION OF THE OPTIMAL PAYLOAD SIZE FOR PACKET BASED TRANSMISSION OVER A BINARY SYMMETRICAL COMMUNICATION CHANNEL

Djemal H. Kolonic[1], *Danilo P. Mandic*[2], *Ben Milner*[2], *and Richard Harvey*[2]

1. Faculty of Electrical Engineering, University of Banjaluka, Banjaluka, Bosnia–Herzegovina.
E–mail: kolonic@etf-bl.rstel.net
2. School of Information Systems, University of East Anglia, Norwich, UK.
E-mail: {mandic,bpm,rwh}@sys.uea.ac.uk

ABSTRACT

A criterion function for selecting the optimal payload size in data transmission over a binary symmetrical communication channel is presented. The work assumes that the packets comprise a fixed size header and variable length payload. The conventional criterion function is based on the ratio between payload size and mean number of transmitted bits per packet, which includes re-transmissions of corrupted packets. Applying a logarithm to the inverse of this criterion function makes the resulting function very suitable for mathematical analysis. This also facilitates parameter sensitivity studies without recourse to numerical methods. As such, this helps to visualise and understand the problem of optimal payload length selection when teaching courses in signal processing for communications and multimedia.

1. INTRODUCTION

The rapid popularisation of multimedia communications [1, 2], over both IP and wireless networks [3], means that there is now an increasing requirement to discuss packet-based data transmission techniques as part of communications courses. An important issue within this is to determine the optimal size of payload allocated to a packet for a given probability of bit error.

Packet transmission at the physical layer can be modelled by assuming that packets comprise $(x + h)$ bits. The information part of the packet is termed the payload and can be set to varying lengths of x bits - as determined by the application or developer. The header is assumed to be of fixed size (h bits - determined by network protocols) and contains error protection, address information, etc. The transmission channel can be modelled by a memoryless binary symmetrical channel (BSC) (although in practice bit errors typically occur in bursts [4]) with channel matrix [5, 6, 7]

$$P_{BSC} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}, \quad 0 < p < 0.5 \qquad (1)$$

where $p(1/0) = p(0/1) = p$, which denotes the probability of bit error for a given channel assuming that any error correction has been applied. When the received packet contains uncorrectable bit errors a request is made to higher layer protocols, such as TCP (Transmission Control Protocol [3]), to re–transmit those packets. This continues until confirmation is given by the destination that the packets have been received without error. In practise other factors, such as packet loss due to network congestion, will also result in packet re–transmission requests.

To maximise packet efficiency it is desirable to make the payload (x bits) as large as possible in comparison to the fixed size packet header (h bits). This results in an increase in the overall packet size ($(x + h)$ bits) which consequently increases the likelihood of a bit error occurring within the packet. In this instance re–transmission in necessary which directly reduces the proportion of information bits which are transmitted. Therefore the aim is to find the optimal size of the payload (x bits) in the packet ($(x + h)$ bits) for a given probability of bit error, p, with respect to the maximum ratio between the payload size and mean number of transmitted bits per packet - including any re–transmission of packets.

The problem itself is not new and has been extensively considered [6]. This is a constrained optimisation problem, which is usually solved by computer simulation for a given channel model and channel quality. Ultimately the final payload size depends on a number of factors which include the network bandwidth, the application, the maximum permissible delay and the processing power of the terminal device. In the physical layer the parameter h contains error correction codes and is, for modelling purposes, assumed to also comprise header information of higher layer protocols. Here, we present a new analytical approach to the problem of deriving the optimum packet length for the channel defined above, which despite its simplicity can point to solutions for real conditions of data transfer. In addition, due to its convenient graphical interpretation, this approach is suitable for teaching the problem, and for determining sub-optimal solutions, when needed.

2. DEFINITION OF THE CRITERION FUNCTION

The probability, p_c, of correctly receiving a packet comprising $(x+h)$ bits, for the binary symmetric channel (after error correction) is

$$p_c = (1-p)^{x+h} \tag{2}$$

The probability of the packet being in error and hence requiring a re–transmission is therefore $(1-p_c)$. The mean number of re–transmissions per packet $\bar{n}$, can be therefore expressed as

$$\begin{aligned} \bar{n} &= p_c + 2p_c(1-p_c) + 3p_c(1-p_c)^2 + \cdots \\ &= \frac{1}{p_c} = \frac{1}{(1-p)^{x+h}} \end{aligned} \tag{3}$$

From (3), we can find the mean number of bits transmitted per message

$$\bar{n}(x+h) = \frac{x+h}{(1-p)^{x+h}} \tag{4}$$

As a measure of transfer efficiency, we can define the criterion function f as the ratio between the number of information bits in a data frame x, and the mean number of transmitted bits per message $\bar{n}(x+h)$ [6]

$$f = \frac{x(1-p)^{x+h}}{x+h} \tag{5}$$

Although both x and h belong to the set of natural numbers $\mathbb{N}$, $(x, h \in \mathbb{N})$, we will assume $0 < x < \infty$. From (5), it is obvious that function f satisfies the conditions

$$\begin{gathered} 0 < f < 1 \\ \lim_{x \to 0^+} f = 0 \\ \lim_{x \to \infty} f = 1 \end{gathered} \tag{6}$$

The optimal length of the payload x_0 can be determined by finding the maximum of the function f in equation (5) with respect to x, for given p and h, i.e.

$$\max(f) = f(x_0) = f_{max} \tag{7}$$

If a relatively good channel is considered, namely small p, then the function f can be approximated by

$$f^* = \frac{x\left[1 - p(x+h)\right]}{x+h}, \quad p << 1 \tag{8}$$

In the case of an ideal channel $(p = 0)$, we have

$$f^{**} = \frac{x}{x+h} \tag{9}$$

Differentiating (5) with respect to x, we have

$$f' = \left[\frac{h}{(x+h)x} + \log_e(1-p)\right] f \tag{10}$$

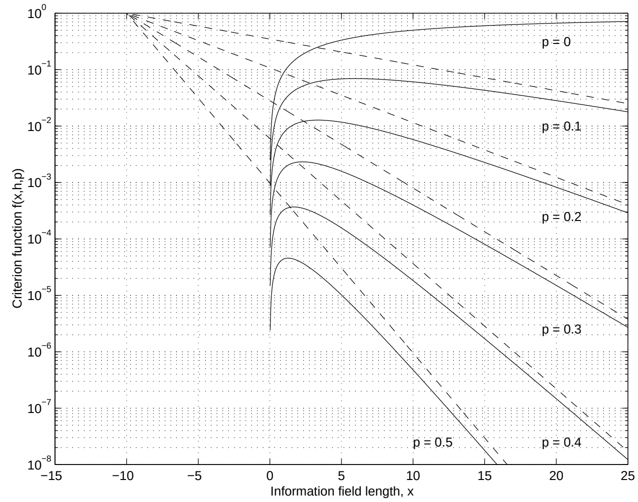

Fig. 1. Criterion function $f(x,h,p)$ for $h = 10$ and various p

Equating (10) to zero, and utilising the nature of the variable x, the root of (10) becomes

$$\begin{aligned} x_0 &= \frac{-ah + \sqrt{a^2h^2 + 4ah}}{2a} \\ a &= \log_e \frac{1}{1-p} \end{aligned} \tag{11}$$

which is the optimal size of the payload. Alternatively, from (8), we have

$$x_0^* = \sqrt{\frac{h}{p}} - h \tag{12}$$

If we proceed with the formal analysis of the criterion function, the second differentiation gives the following analytic expression

$$\begin{aligned} f'' &= \frac{d}{dx}\left[\frac{h}{(x+h)x} + \log_e(1-p)\right] f \\ &= \frac{-h(2x+h)}{x^2(x+h)^2} f + \left[\frac{h}{x(x+h)} + \log_e(1-p)\right] f' \end{aligned} \tag{13}$$

with f and f' given respectively in (5) and (10). Figure 1 shows the criterion function $f = f(x,h,p)$ for fixed $h = 10$ and various values of the error probability of the channel p. Figure 1 also shows the asymptotes of f as $x \to \infty$. Because of its large dynamics, function f is not very suitable for strict analysis, and is sensitive to parameter perturbations. Therefore we look for an alternative way of representing the criterion function f, which would be more suitable for analysis and graphical representation.

3. THE PROPOSED CRITERION FUNCTION

The above analysis can be significantly simplified by using the logarithm of the criterion function f (5). Knowing that the logarithm function is a monotonically increasing function, the result of the optimisation with respect to x is identical. As the function f lies within the $(0, 1)$ interval, the function $\log_e f$ lies within the $(-\infty, 0)$ interval. Hence, it is more convenient to consider a new criterion function g in the form of a logarithm of the reciprocal value of function f, i.e.

$$g = \log_e \frac{1}{f} = \log_e \frac{x+h}{x} + (x+h)\log_e \frac{1}{1-p} \quad (14)$$

This gives $f_{max} \Leftrightarrow g_{min}$ with respect to x, and hence, the criterion function g becomes a measure of the ratio between the mean number of transmitted bits per message and the number of bits in the information field of the message. Following the same approach as for the criterion function f, from equation (14) we have

$$g' = \log_e \frac{1}{1-p} - \frac{h}{x(x+h)} \quad (15)$$

Solving (15) with respect to variable x, we obtain x_0 as given in (11). The second derivative

$$g'' = \frac{h(2x+h)}{x^2(x+h)^2} \quad (16)$$

is positive along the interval of interest $(0 < x < \infty)$.
From equation (16), we see that g is convex, since its second derivative is strictly greater than zero, and thus it reaches its minimum at x_0. In other words, function f (5) indeed reaches its maximum at x_0 (11).
Figure 2 shows simulation results for the criterion function $g = g(x, h, p)$ for various values of the error probability of the channel p and fixed $h = 10$. Using (11) and (14) we can find an expression for the curve g_0 that represents the minima of function g (trajectory connecting optimal payloads) as

$$g_0 = \log_e \frac{x_0+h}{x_0} + \frac{h}{x_0} \quad (17)$$

which is shown by the broken line in Figure 2.
For most applications the payload is significantly larger than the header, hence $h \ll x$. For a given p, function g now becomes

$$g(x, h, p) \approx (x+h)\log_e \frac{1}{1-p} \quad (18)$$

since $\log_e \frac{x+h}{x} \approx 0$ for $h \ll x$. The set of curves (18), are the asymptotes of curves $\{g\}_{p=0.1,\ldots,0.5}$ shown in Figure 2. All the asymptotes intersect in the point $(-h, 0)$. A further approximation, for a fixed p would be

$$g(x, h) \approx c\,(x+h)\,, \quad C = \log_e \frac{1}{1-p} \quad (19)$$

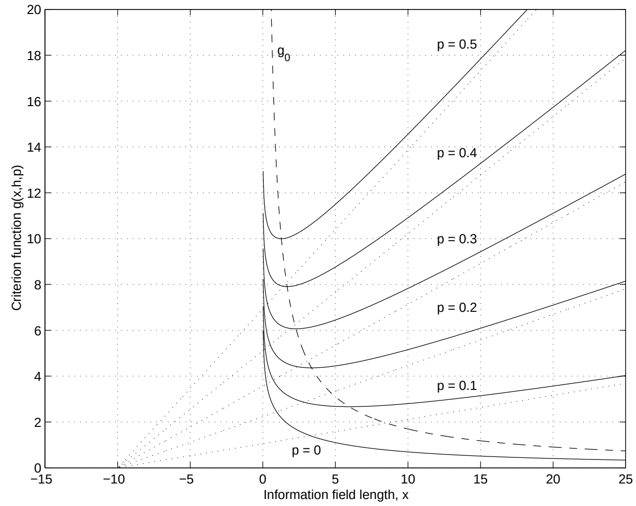

Fig. 2. Criterion function $g(x, h, p)$ for $h = 10$ and various p

Due to the simplicity of functions g, g', and g'', it is easy to show that the function g rises slower on the right from x_0 than on the left from this point (optimal location), i.e. for $\forall \epsilon > 0$

$$g(x_0 + \epsilon) < g(x_0 - \epsilon) \quad (20)$$

which is equivalent to

$$f(x_0 + \epsilon) > f(x_0 - \epsilon) \quad (21)$$

This means, since x_0 in practice has to be a positive integer, if we take an approximate value for x_0, then it is better, having in mind other limitations, to take greater rather than lower values of x with respect to x_0.

4. SIMULATION RESULTS

Table 1 shows several results for the optimal size of the payload x_0, together with appropriate values of functions f and g for $h = 40$ and various p. A header size of 40 bits is not unrealistic in the physical layer in which the packet headers consist of between two and four bytes. Allowing another byte for redundancy within the payload gives 40 bits.
The maximum value of the criterion function, denoted by f_{max}, is shown in the fifth column of Table 1 and its approximation f^*_{max} is shown in the sixth column of Table 1. The approximation x^*_0 is useable for many practical systems in which the error rate is mostly less than 10^{-3}. Due to the approximations made, the values of f^*_{max} and x^*_0 appear negative for a poor channel. Figure 3 illustrates the optimum packet length x_0 and its approximation x^*_0 for a

Table 1. Criterion functions and optimal information field lengths for $h = 40$ and various p

p	x_0	x_0^*	g_0	$f_{\max}$	$f_{\max}^*$
10^{-8}	63226	63206	0.001	0.999	0.999
10^{-7}	19981	19960	0.004	0.996	0.996
10^{-6}	6305	6285	0.0127	0.987	0.987
10^{-5}	1981	1960	0.0402	0.961	0.960
10^{-4}	613	593	0.129	0.879	0.877
10^{-3}	181	160	0.421	0.657	0.638
10^{-2}	47	24	1.47	0.231	0.070
10^{-1}	8	-20	6.79	0.001	-0.633
0.1	8	-20	6.79	0.001	-0.633
0.2	5	-25	10.197	0.00	-0.889
0.4	2	-30	23.044	0.00	-0.752

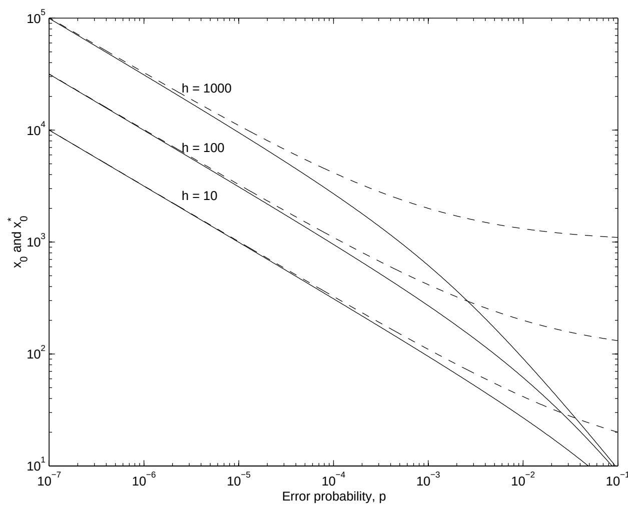

Fig. 3. Optimal payload size x_0 (solid line) and approximate payload size x_0^* (dashed line) for various h and p

variety of channel probabilities and header sizes. For realistic bit error probabilites and moderate header sizes the approximation x_0^* is perfectly satisfactory.

As we expected, in the case of higher quality channels, the length of the information field could be relatively longer, thus providing better transmission efficiency. So, for instance, for a high quality channel, with $p = 10^{-7}$ the calculated optimum data length is $x_0 = 19981$ bits. This is close to the TCP maximum payload size of 1500 bytes.

Considering a poor quality channel, such as a cellular link which may have a bit error rate of $p = 10^{-3}$, the calculated optimal payload size is $x_0 = 181$ bits. This is again close to typical packet sizes used in, for example, GSM networks. These calculated figures of the optimal payload size compare reasonably well to actual payload sizes used in various networks [8].

5. CONCLUSIONS

An analytical approach to the derivation of the optimal payload length for data transmission over a binary symmetrical channel has been presented. The ratio between the number of bits in the payload and the mean number of bits needed for the transmission of one packet has been used as a convenient criterion function. This definition is a constrained optimisation problem which has been traditionally solved by computer simulation.

It has been shown that the analysis becomes much simpler if the criterion function is defined as a logarithm of the reciprocal value of the above mentioned function. In the analysis, while varying the payload size, the header has been assumed to be of constant size as is specified by network protocols. We have shown, that when in a need of a suboptimal solution, it is always better to take greater rather than lower values of the optimal data field length x_0. Simulated results compare reasonably well to payload sizes used in practical applications.

The analysis presented is useful when teaching courses on signal processing for communications and dealing with the problem of deriving an optimal payload size for transmission.

6. REFERENCES

[1] J. D. Gibson, ed., *Muiltimedia Communications*. Academic Press, 2001.

[2] S. Gringeri, R. Egorov, K. Shuaib, A. Lewis, and B. Basch, "Robust compression and transmission of MPEG-4 video," in *Proceedings of the seventh ACM international conference on Multimedia*, pp. 113–120, 1999.

[3] F. Halsall, *Data Communications, Computer Networks and Open Systems*. Addison Wesley, third ed., 1992.

[4] I. Cidon, A. Khamisy, and M. Sidi, "Analysis of packet loss processes in high speech networks," *IEEE Transactions On Information Theory*, vol. 39, no. 1, pp. 98–108, 1993.

[5] M. Schwartz, *Computer Communication Network Design and Analysis*. Prentice–Hall, 1977.

[6] G. J. Marshall, *Principles of Digital Communications*. McGraw–Hill, 1980.

[7] D. Petersen, *Audio, Video and Data Telecommunications*. McGraw–Hill, 1992.

[8] P. Manzoni, D. Ghosal, and G. Serazzi, "Impact on mobility on TCP/IP: An integrated performance study," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 5, pp. 858–867, 1995.