Coding vs. Packet Retransmission over Noisy Channels

Dror Baron, Shriram Sarvotham, and Richard G. Baraniuk
Dept. of Electrical and Computer Engineering
Rice University, Houston, TX 77005

***Abstract*— In many packet-based communication systems such as TCP/IP-based systems, packets are communicated over a noisy physical layer (a channel), and if a packet cannot be decoded correctly, then the transport layer retransmits it. Of course, retransmissions consume significant resources and their use should be limited. However, decreasing the likelihood of retransmission requires to encode the packets with strong channel codes in the physical layer, which also requires additional channel resources. In this paper, we study the cross-layer tradeoff between coding and packet retransmissions, and optimize over the total channel resource consumption. We show that as the packet length k increases, the redundancy r beyond the k/C channel uses implied by Shannon's channel capacity C is $\Theta(\sqrt{k \ln(k)})$ extra channel uses. Moreover, as k increases we must use stronger channel codes. We then apply these results to universal coding over a piecewise memoryless channel with transitions between unknown i.i.d. statistics. Our constructive universal algorithm has redundancy $r = O(k^{2/3}\sqrt{\ln(k)})$ using packets of polynomially increasing lengths while accounting for possible packet drops caused by transitions in the statistics.**

Keywords: Channel coding, cross-layer design, non-asymptotic information theory, packet networks, piecewise memoryless channels, universal channel coding.

I. Introduction

Many communication systems rely on *packets*. Input data is partitioned into packets, each packet is *encoded* and transmitted separately over a lossy physical layer (a *channel*), and the channel output is then used to *decode* the packet. If the packet is decoded successfully, then all is well. But if the decoding process fails, then we have a *packet drop*. In this case, the packet can be *retransmitted*.[1] Therefore, as long as the communication designer is willing to admit some delay in case of a necessary retransmission, reliable communication is assured. For example, in TCP-IP the physical layer may incorporate channel coding, whereas the transport layer uses retransmissions to ensure reliability [1].

If a strong channel code is used, then the likelihood of a packet drop is small, and less channel resources are consumed for retransmissions. However, such powerful coding techniques also consume more resources. In this paper, we study the cross-layer tradeoff between resource expenditures on channel coding and on packet retransmissions.

Let us quickly discuss various factors that affect this tradeoff. First, if the delays in the retransmission process are significant, for example in space communications, then retransmissions are even more undesirable, hence there is a need to use stronger coding techniques. For applications such as email that are not very sensitive to delays, retransmission strategies are appropriate and frequently used [1]. Second, non-stationary channels are conducive to retransmission policies, because we prefer not to design the code for the worst-case channel [2]. Third, in multicast and broadcast applications, retransmissions are bad because each receiver may have lost or received a different set of packets, and so the acknowledgment packets add up and consume substantial resources [1]. Fourth, in many communication systems there are various queues, which may also cause packet drops. No matter how powerful our channel code may be, some packets will still be dropped, and these effects will bias the design to use slightly weaker channel coding techniques. We leave the study of such issues for future work. Finally, in *rateless codes* [3] the channel encoder

This work was supported by NSF, AFOSR, ONR, and the Texas Instruments Leadership University Program. Email: {drorb, shri, richb}@rice.edu; Web: dsp.rice.edu.

[1] The encoder decides whether to retransmit or not based on a small amount of feedback (such as acknowledgment packets in TCP [1]) received from the decoder. The decoder can determine whether the decoding process is successful using error detection mechanisms. Although such details are essential in any implementation, they may obscure the essence of our problem, and therefore our presentation is purposefully more abstract.

continues transmitting until the decoder has received enough information to recover the message. The decoder then acknowledges correct reception of the message. For a comparison of rateless codes to retransmission techniques, see Section V.

In this paper, we study the total expected channel uses required to transmit a packet in a communication system that uses retransmissions. We ignore additional important issues such as delay, design complexity, sophisticated source and channel modeling, packet drops caused by queueing, and so on. Yet despite our simple formulation, we will see that it provides useful insights on effective design of packet drop probabilities.

We then apply these results to a setup where a piecewise memoryless channel has transitions between unknown i.i.d. statistics. We provide a constructive universal algorithm that uses packets of polynomially increasing lengths up to a transition in the statistics. The design of the block lengths accounts for the additional tradeoff between packet drops caused by transitions in the statistics and how close we can approach capacity while using short packets.

II. Problem formulation

Consider a length-k input packet that is communicated over an independent and identically distributed (i.i.d.) channel with capacity C.[2] Information theory [4] suggests that the number n of channel uses necessary is approximately k/C. However, for many channels of practical interest (for example a binary symmetric channel), using the channel n times to communicate the packet incurs a strictly positive probability of packet drop, even if $n \gg k/C$. This discrepancy is caused by the infinitely long codewords used in the proofs of the capacity theorems [4]. Although information theory has set the theoretical limits on communication systems, these limits are asymptotic in nature and may not apply to our specific problem. To deal with input packets of finite length k, we adapt the following theorem from our recent work [5] (see also Wolfowitz [6] and references therein).

Theorem 1: [5] For a binary symmetric channel with crossover probability p, there exists a constant Q_1 such that, if the number n of channel uses in the first transmission satisfies

$$n = k/C + Q_1\sqrt{k}\delta, \tag{1}$$

then the probability ϵ of packet drop satisfies

$$\epsilon = \Phi(\delta) + O(1/\sqrt{k}).$$

[2]An i.i.d. model may be too simple for some wireless systems, yet may be appropriate in other scenarios, especially for some wired channels. We consider non-stationary channels in Section IV.

In our theorem, we use $\Phi(\cdot)$ to denote the error function[3] and $O(\sqrt{k})$ to denote a function that is upper bounded by a term proportional to $\sqrt{k}$.[4] The Berry-Esséen inequality [7, 8] enables to upper bound this $O(\sqrt{k})$ term.

The main intuition that may be gleaned from Theorem 1 is that backing off δ standard deviations from the channel capacity buys us $\Phi(\delta)$ probability of error. Our previous results [5] show how to compute Q_1 in closed form for the binary symmetric channel (BSC). Similar forms are also available for other types of channels [6, 9, 10] and for the dual problem of distributed source coding [6, 11].

If the first attempt to transmit the packet succeeds, then the total resource expenditure is n channel uses. However, if a packet drop occurred, then we retransmit the packet. How many extra channel uses must we allocate for this unfortunate occurrence? Because there may be additional packet drops, it could be tedious to examine all possible cases. Instead, we assume that the *expected* resource expenditure in case of a packet drop is Q_2k channel uses, where Q_2 is constant. We justify this assumption by noting that we can use a stronger channel code in a second transmission round to ensure an extremely high probability of success (c.f. Birk et al. [12, 13]). Finally, we emphasize again that packet drops caused by queueing or additional effects will bias the design to use slightly weaker channel coding techniques. We leave the study of such issues for future work.

Let t be the *total expected resource expenditure* in terms of channel uses. Combining the number n of channel uses during the first transmission attempt (1) and the resource expenditures in case of packet drop, we have

$$\begin{aligned} t &= k/C + Q_1\sqrt{k}\delta + Q_2k\overbrace{\left[\Phi(\delta) + O(1/\sqrt{k})\right]}^{\epsilon} \\ &= k/C + Q_1\sqrt{k}\delta + Q_2k\Phi(\delta) + O(\sqrt{k}). \end{aligned} \tag{2}$$

We also define the *redundancy* r, which is the extra expected number of channel uses beyond capacity, i.e.,

$$r = Q_1\sqrt{k}\delta + Q_2k\Phi(\delta) + O(\sqrt{k}). \tag{3}$$

This formula captures the essential tradeoff between coding and packet retransmission. If a strong channel code is used for the first transmission (small ϵ), then $Q_2k\epsilon$ is small, but then δ must be large, and so extra resources are expended for that first transmission. In contrast, large ϵ wastes substantial resources on retransmissions.

[3]More formally, $\Phi(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$.

[4]For two functions $f(n)$ and $g(n)$, $f(n) = O(g(n))$ if $\exists c, n_0 \in \mathbb{R}^+,\ 0 \le f(n) \le cg(n)$ for all $n > n_0$. Similarly, $f(n) = \Theta(g(n))$ if $\exists c_1, c_2, n_0 \in \mathbb{R}^+,\ 0 \le c_1g(n) \le f(n) \le c_2g(n)$ for all $n > n_0$.

III. Main result

The following theorem describes the optimal theoretical tradeoff between coding and packet retransmission.

Theorem 2: The optimal redundancy r satisfies

$$r = \Theta(\sqrt{k\ln(k)}).$$

Proof: Equation (3) for the expected resource expenditure redundancy includes an uncertainty of $O(\sqrt{k})$ channel uses. Therefore, we cannot determine the δ that minimizes r, because δ may affect the $O(\sqrt{k})$ term. Instead, we assign

$$\delta^* = \sqrt{\ln(k/\ln^2(k))},$$

where $\ln(\cdot)$ denotes the natural logarithm. We then incorporate the well known approximation that $\Phi(x) \propto e^{-x^2/2}/x$ for large x, which yields

$$\begin{aligned}\Phi(\delta^*) &\approx \frac{e^{-[\ln(k/\ln^2(k))]/2}}{\sqrt{\ln(k/\ln^2(k))}} \\ &= \frac{\sqrt{\ln^2(k)/k}}{\sqrt{\ln(k/\ln^2(k))}} \\ &= O(\sqrt{\ln(k)/k}).\end{aligned}$$

Therefore, the total expected redundancy is

$$\begin{aligned}r &= Q_1\sqrt{k\ln(k/\ln^2(k))} \\ &\quad +Q_2 k\cdot O(\sqrt{\ln(k)/k}) + O(\sqrt{k}).\end{aligned}$$

The first and second terms on the right hand side are both of order $\sqrt{k\ln(k)}$, and the third Berry-Esséen term [7, 8] is of smaller order, and thus no longer affects the performance. Therefore, we have obtained a *tight* $\Theta(\cdot)$ order term bound for the redundancy, because a reduction in the order of either of the first two terms would increase the order of the other term. □

We provide numerical results in Figure 1. Our results use a binary symmetric channel with crossover probability $p = 0.1$. This value for p is reasonable in applications where the physical layer is moderately noisy; the channel capacity in this case is $C(p) = 0.53$. We also used a retransmission factor $Q_2 = 5$, which is much larger than $1/C$, and thus enables to use powerful coding techniques. As the packet length k increases, the redundancy r increases as $O(\sqrt{k\ln(k)})$ and the packet drop probability decays. For packet lengths of practical interest, the redundancy is quite significant.

IV. Application to channels with transitions

We now consider communication over a piecewise memoryless binary channel whose crossover probability transitions between different values. In this universal setting, neither the times when these transitions occur nor the channel statistics (crossover probabilities) are known. Our goal is to provide universal channel coding techniques over non-stationary channels. Whereas earlier we wanted to minimize the amount of resources required for communication in a setting where a transport layer allowed us to compensate for errors over a physical layer, here the goal is to operate in a universal manner. Our approach is that the minimization of resource usage (channel uses in our problem) requires to use variable rate codes. However, the channel statistics are unknown, and so we must estimate them. To do so, feedback from the channel decoder must be used. In recent work [14–17], universal channel coding approaches were provided for a stationary BSC. The main idea there was to use packets of increasing length, where at the end of each packet the decoder relays the current estimate of the crossover probability to the encoder via feedback; as more packets are processed, the estimation quality improves and the encoder can use rates closer to the channel capacity. We will show that the setting with transitions between crossover probabilities requires to consider the tradeoffs between coding and retransmission.

Let us assume that at some deterministic (yet unknown) time the channel transitions between crossover probability p to a different p'. Before the transition, we use a channel code designed for p (or our current estimate for p). The transition occurs during the transmission of some packet. If $p' < p$, then we have a *good transition* – the packet will encounter less channel errors than usual, and therefore the likelihood of success is extremely high. In contrast, if $p' > p$ then we have a *bad transition* – there will be more channel errors than usual, and the packet will be dropped.

Obviously the possibility that a packet is dropped because of a bad transition requires us to retransmit and is wasteful of resources. But even good transitions cause difficulties, because the number of channel uses may greatly exceed the resource expenditure required for the smaller crossover probability p'. Furthermore, although the packet most likely succeeded, we are incapable of estimating p'. We conclude that (*i*) a transition during a packet that required n channel uses will usually cause $\Theta(n)$ resource loss (even for a good transition) and (*ii*) after each such transition the process of estimating p' must begin from scratch. Consequently, our analysis considers how to transmit as many bits as possible over

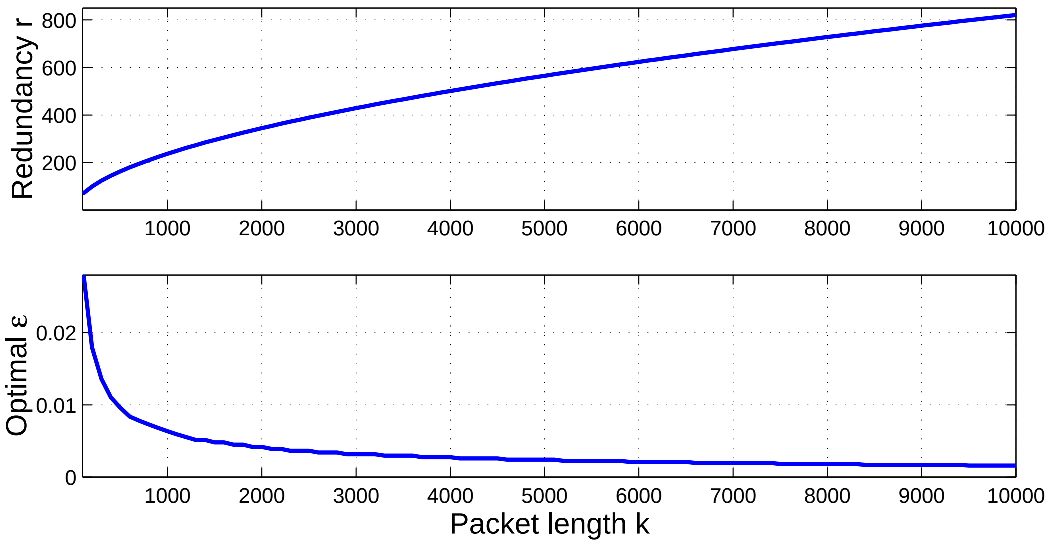

Fig. 1. **Numerical tradeoff between coding and packet retransmission**: As the packet length k increases, the redundancy r increases and packet drop probability ϵ decays. We used a BSC with crossover probability $p = 0.1$ and chose a retransmission factor $Q_2 = 5$.

a channel that has crossover probability p for some time and later transitions to p'. Note, however, that in contrast to the result of Section III, here the packet drop occurs because of the channel transition, and over-design for several standard deviations of channel noise (Theorem 1) does not suffice. Unless we prefer to operate well below capacity, packet drops are inevitable in this setting.

In recent work on universal coding for a stationary BSC [14–17], packets of geometrically increasing lengths enabled to approach the channel capacity quickly. Suppose that we have processed i blocks so far, and in block $j \in \{1, \dots, i\}$ we conveyed k_j bits via n_i channel uses. Let $N_i \triangleq \sum_{j=1}^{i} n_i$ be the total number of channel uses so far. If n_i are increasing geometrically then the $n_{i+1} = \Theta(N_i)$. If a transition occurs in block $i + 1$, then the $O(n)$ penalty will significantly increase the redundancy. Instead, we suggest to use $n_{i+1} = O((N_i)^{2/3})$. With this choice of packet lengths, the results of Section III can be used to show that the aggregate redundancy during the previous i blocks satisfies $r = O((N_i)^{2/3}\sqrt{\ln(N_i)})$, whereas the penalty for a transition is $O((N_i)^{2/3})$. Because the order of the aggregate number of channel uses during the first i blocks is similar to the order of the aggregate number of bits conveyed, i.e., $N_i = O(K_i)$, we have the following result. The detailed proof has been omitted for brevity.

Theorem 3: Using packets of length $n_{i+1} = O((N_i)^{2/3})$ for a transmission scheme over a piecewise memoryless channel with transitions between i.i.d. segments, the redundancy r required while conveying k aggregate bits satisfies

$$r = O(k^{2/3}\sqrt{\ln(k)}).$$

We can also provide a somewhat smaller order term to lower bound the redundancy. In words, this lower bound states that any packet-based feedback scheme for a piecewise memoryless channel cannot achieve a significantly smaller redundancy. The disparity between our converse and achievable bounds lies in the extra $\sqrt{\ln(\cdot)}$ term required to support possible retransmissions in Section III. Noting that LDPC codes approach capacity as quickly as indicated in Theorem 1 (with a somewhat larger constant Q_1 [18]), the results of this section are significant because they identify a research direction where the potential gains to be had are much larger than in traditional channel coding.

V. Closing statements

This work has studied the cross-layer tradeoff between resource expenditures on channel coding in the physical layer and on packet retransmissions in the transport layer. Theorem 2 proves that a proper choice of δ yields a penalty for retransmissions on the same order as the number of channel uses expended beyond capacity. (The latter expenditure is necessary to combat non-asymptotic effects [5, 9, 11].) Furthermore, because δ^* is monotone increasing in k, as k increases we must use stronger channel codes. This can be explained by realizing that a packet drop is more costly as k increases, and so it is advantageous to back off more in order to reduce the retransmission penalty.

While appealing and insightful, these results are not the complete story, they are just the first chapter in our work. When piecewise memoryless channels are considered, retransmissions must be needed to combat against a bad transition from a small crossover probability p to a larger p'. In this case, in addition to the tradeoff between the probability of decoding error and the penalty for retransmission, we have an additional tradeoff between using longer packets (thus employing channel codes with rates closer to capacity) and using shorter packets that are not too costly when dropped. The intriguing result is that the redundancy in this case is of a much larger order. Therefore, universal coding for non-stationary channels appears to be a research direction where large potential gains are possible. Additional future work will consider effects such as packet drops induced by queueing.

Finally, rateless codes [3] provide reliable communication by encoding until the decoder has received enough information to recover the packet. Therefore, in some sense these methods enable to do away with packet drops. Unfortunately, the redundancy using that technique is $O(\sqrt{k}\log^2(k))$, whereas the penalty we have achieved using retransmissions is smaller. We conclude that judicious design of drop probabilities makes retransmission policies an effective tool for ensuring reliable communication over a lossy channel, and additional related tradeoffs appear in the design of packet lengths for universal schemes.

Acknowledgments. We thank K. Ramchandran and T. Weissman, who inspired our study of and approach to this problem. We also thank Y. Birk and M. A. Khojastepour for informative and inspiring conversations.

References

[1] J. F. Kurose and K. Ross, *Computer Networking: A Top-Down Approach Featuring the Internet*. Addison-Wesley Longman Publishing Co., Boston, 2002.

[2] T. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, 2001.

[3] M. Luby, "LT codes," *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, Vancouver, Canada, pp. 271–282, November 2002.

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, New York, 1991.

[5] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "How quickly can we approach channel capacity?," *Proc. 38th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2004.

[6] J. Wolfowitz, *Coding theorems of information theory*, 3rd ed. Springer-Verlag, Berlin; New York, 1978.

[7] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. American Math. Society*, vol. 49, pp. 122-136, 1941.

[8] C. G. Essèen, "On the Liapounoff limit of error in the theory of probability," *Ark. Mat. Astr. och Fys.* vol. 28A, no. 9, pp. 1-19, 1942.

[9] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Bounds on Non-Asymptotic Channel Capacity," *in preparation.*

[10] C. E. Shannon, "Probability of Error for Optimal Codes in a Gaussian Channel," *Bell System Tech. Journal*, vol. 38, pp. 611-656, 1959.

[11] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Redundancy rates of Slepian-Wolf coding," *Proc. 42nd Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2004.

[12] Y. Birk and Y. Keren, "Judicious use of redundant transmissions in multichannel ALOHA networks with deadlines," *IEEE J. Selected Areas in Communications*, vol. 17, no. 2, pp. 257–269, Feb. 1999.

[13] D. Baron and Y. Birk, "Multiple working points in multichannel ALOHA with deadlines," *Wireless Networks*, vol. 8, no. 1, pp. 5–11, Jan. 2002.

[14] M. Johnson, P. Ishwar, V. M. Prabhakaran, D. Schonberg, and K. Ramchandran "On compressing encrypted data," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2992–3006, Oct. 2004.

[15] D. Schonberg, S. C. Draper, and K. Ramchandran, "On blind compression of encrypted correlated data approaching the source entropy rate," *Proc. 43rd Annual Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sept. 2005.

[16] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Variable-rate coding with feedback for universal communication systems," *Proc. 43rd Annual Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, Sept. 2005.

[17] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Variable-rate universal Slepian-Wolf coding with feedback," *Proc. 39th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2005.

[18] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, "Finite-length scaling for iteratively decoded LDPC ensembles," *Proc. 41st Annual Allerton Conf. on Comm., Control, and Computing*, Oct. 2003.