

Bounds on the Capacity of Deletion Channels

Suhas N. Diggavi and Matthias Grossglauser
 AT&T Shannon Laboratories
 Florham Park, NJ 07932
 e-mail: {suhas,mgross}@research.att.com

I. SUMMARY

In this paper, we develop bounds on the achievable rate for deletion channels. Deletion channels occur when symbols are randomly dropped, and a subsequence of the transmitted symbols is received. Our initial motivation for studying these channels arose in the context of information transmission over finite-buffer queues in packet-switched networks, where the receiver does not have access to side-information on which packets were dropped. This also motivated our study on the effect of large alphabet sizes on the achievable rate.

In deletion channels, unlike erasure channels, there is no side-information about which subsequence is received. Clearly, the capacity of the erasure channel is therefore a simple upper bound for the capacity of the deletion channel. The natural question which arises is about how much worse (in terms of information transmission rate) a deletion channel is compared to an erasure channel. This might shed light on the amount of “redundancy” actually needed for transmission over packet loss channels.

The deletion channel is a special case of insertion/deletion/substitution channels¹ which model the effect of synchronization errors and have a long history ([1, 2] and references therein.) Even in the presence of memoryless deletions, there is no single-letter characterization for achievable rates. All the published literature deals with the binary alphabet case. For memoryless deletion channels [1, 2] showed that

$$C_{del} \geq 1 - H_0(p_d), \quad p_d \leq 0.5 \quad (1)$$

where $H_0(p_d) = -(1-p_d)\log(1-p_d) - p_d\log(p_d)$ is the binary entropy function. We provide an alternative proof for this result which also yields lower bounds for larger (non-binary) alphabet sizes and when the deletion process is stationary and ergodic. We also derive bounds that improve (1) by using codebooks with memory.

Our main result is that the achievable rate in deletion channels differs from that of erasure channels by at most $H_0(p_d) - p_d \log \frac{K}{K-1}$ bits, where p_d is the deletion probability, K is the alphabet size and $H_0(\cdot)$ is the binary entropy function. We sharpen these bounds by giving a characterization of achievable rates using input codebooks with memory for the non-binary deletion channel. These lower bounds, when specialized for the binary deletion channel, improve the bounds reported in (1).

II. PROBLEM STATEMENT AND MAIN RESULT

The K -ary deletion channel is defined as follows. Let $x = (x_1, \dots, x_n)$ be a codeword, where $x_i \in \{1, \dots, K\}$. A deletion pattern D is a binary vector (D_1, \dots, D_n) , where $D_i = 1$ indicates that the i -th symbol of x is deleted, and $D_i = 0$ indicates that the i -th symbol is received at the output. We are mainly interested in i.i.d. distributions (with $\Pr\{D_i = 1\} = p_d$) for the binary sequence D_i , but Theorem II.2 also applies when D is stationary and ergodic. Note that the deletion channel has memory in that $p(y|x)$ does not become a product distribution even for an i.i.d. deletion process.

We first prove lower bounds for the capacity of the deletion channel. We do this using random codebooks generated with a first-order Markov process. Intuitively, we expect such

¹In an insertion channel, additional symbols can randomly be inserted into the codeword. Substitutions are the familiar symbol errors for noisy channels.

codebooks to perform better than memoryless codebooks because its codewords tend to contain runs, i.e., sequences of identical symbols. The information can then be viewed as being encoded as a sequence of runlengths. If no run is deleted completely in the channel, the channel acts as a DMC for the runlengths. This motivates the use of first-order Markov chains, which generate i.i.d. runlengths. Our main result is the following.

Theorem II.1 *Given an i.i.d. deletion pattern with deletion probability p_d , and a K -ary input alphabet, the capacity of this channel is lower bounded as*

$$C_{del} \geq \sup_{\substack{\gamma > 0 \\ 0 < p < 1}} [(p_d - 1) \log\{(1-q)A + qB\} - \gamma] \text{ nats} \quad (2)$$

$$\text{where } A = A(p, \gamma) = e^{-\gamma} \frac{(1-p)}{(K-1)[1 - e^{-\gamma} \{1 - \frac{1-p}{K-1}\}]}, \quad B = B(p, \gamma) = e^{-\gamma} ((1-p)A + p) \quad \text{and} \quad q = q(p) = \frac{1}{K} \left[1 + \frac{(1-p_d)(K-1)(pK-1)}{(K-1)-p_d(pK-1)} \right].$$

The proof requires calculating the probability that a random subsequence of a codeword generated by the first-order Markov chain is a subsequence of another random codeword. The parameter p of Theorem II.1 essentially controls the tradeoff between small probability of runs being completely deleted (long runs are better) and the length of the sequence of runs (short runs are better). Note that the optimization in (2) for a given p can be accomplished in closed form as it results in a simple quadratic equation in γ . In the binary case, the bound (2) is sharper than (1).

We can specialize the result for i.i.d. codebooks by making $p = \frac{1}{K}$, which results in the following theorem. Note that for i.i.d. codebooks, we can actually prove a stronger result for stationary and ergodic deletion patterns.

Theorem II.2 *Given a stationary and ergodic deletion pattern with long-term deletion probability given by p_d (with $p_d < 1 - 1/K$), and an input alphabet size K , the capacity of this channel is lower bounded as*

$$C_{del} \geq \log \left(\frac{K}{K-1} \right) + (1-p_d) \log(K-1) - H_0(p_d). \quad (3)$$

This result when specialized to the binary case ($K = 2$) results in the Gallager-Zigangirov result given in (1). As mentioned earlier, the erasure channel provides an obvious upper bound of $(1-p_d)\log(K)$, which therefore shows that we have sandwiched the deletion channel capacity to within $H_0(p_d) - p_d \log \frac{K}{K-1}$ bits per (input) symbol.

One question that we are pursuing is that of tighter upper bounds for the deletion channel. For example, one promising approach is a channel that only conveys side-information about runs (sequences of identical symbols) that are completely deleted.

REFERENCES

- [1] R. G. Gallager, “Sequential decoding for binary channels with noise and synchronization errors”, *Lincoln Lab. Group Report*, October 1961.
- [2] K. Sh. Zigangirov, “Sequential decoding for a binary channel with drop-outs and insertions”, *Problems Information Transmission*, vol 5., no. 2, pp 17–22, 1969. *Translated from Problemy Peredachi Informatsii*, vol. 5, no. 2, pp 23–30, 1969