B.S.T.J. BRIEF

On Source Networks with Minimal Breakdown Degradation

By H. S. WITSENHAUSEN

(Manuscript received February 22, 1980)

I. INTRODUCTION

A source is encoded into two data streams for transmission to a receiver over two noiseless (or error-corrected) channels. This receiver is able to reproduce the source stream without error until there is a breakdown of one of the channels.

If such a breakdown can be sensed both at the transmitter and the receiver, then they can prearrange that, in case of breakdown, they will switch to a different encoder and decoder designed to achieve the minimum distortion possible for the capacity of the remaining channel. However, it is assumed that the transmitter will be, at least for some time, unaware of the breakdown.

If one channel is highly reliable and only breakdowns of the other need be considered, then one can use an encoding over the reliable channel as if it were the only one, achieving the rate distortion bound for the capacity of the reliable channel. The theory of side information shows that, if the total capacity is sufficient for reconstruction of the source output, which we assume, then there is an encoding for the unreliable channel that provides the "complementary" data such that, when both channels are up, reconstruction is still possible, at least in the Shannon sense.

However, we assume that both channels are susceptible to breakdowns, and this produces a new type of problem, which is of interest also in connection with packet transmission schemes.

II. AN INEQUALITY

Suppose a block of \( N = N_1 + N_2 \) bits from a memoryless binary symmetric source is encoded into two signals \( U \) and \( V \) with respective
alphabet sizes $2^N$ and $2^{N'}$. A receiver of the pair $(U, V)$ is able to reconstruct the source block $X^N$ without error. There are two other receivers, one receiving $U$ only and producing a binary block $Y^N = F(U)$, and the other receiving $V$ only and producing the binary block $Z^N = G(V)$.

For each bit position $k (1 \leq k \leq N)$, the source bit $X_k$ is compared with the decoded bits $Y_k$ and $Z_k$. Define the error probabilities

$$p_k = \Pr(X_k \neq Y_k), \quad p_k = \Pr(X_k \neq Z_k).$$

**Theorem:** For all $k$, the point $(p_k, p_k)$ lies in the region of the $(p_u, p_v)$ plane defined by $0 \leq p_u \leq 1$, $0 \leq p_v \leq 1$ and

$$\left(p_u + \frac{1}{2}\right)\left(p_v + \frac{1}{2}\right) \geq \frac{1}{2}.$$

To establish this theorem, use is made of the following lemma, of which we omit the proof, as it is a special case of results that will appear elsewhere.

Suppose $U_0$ and $V_0$ are two independent random variables (their values could be in any two measurable spaces). Let $f(U_0), g(V_0)$, and $h(U_0, V_0)$ be measurable functions with values in $(0, 1)$.

Define

$$p_u = \Pr\{f(U_0) \neq h(U_0, V_0)\},$$

$$p_v = \Pr\{g(V_0) \neq h(U_0, V_0)\},$$

and assume that

$$\Pr\{h(U_0, V_0) = 0\} = \frac{1}{2}.$$

**Lemma:** Under the above assumptions, one has

$$\left(p_u + \frac{1}{2}\right)\left(p_v + \frac{1}{2}\right) \geq \frac{1}{2},$$

and this inequality is the best possible.

Returning to the discrete situation of the theorem, observe that there are exactly as many $(U, V)$ pairs, $2^N 2^{N'} = 2^N$, as there are source blocks $X^N$. The condition of exact reconstructibility of $X^N$ from $(U, V)$ implies that each pair $(U, V)$ corresponds to one, and only one, distinct block. As the blocks all have the same probability $2^{-N}$, the variables $U, V$ are independent and uniformly distributed over their alphabets.

Consider the $k$th bit position. Exactly half the blocks, thus half the $(U, V)$ pairs, have $X_k = 0$. Now we can let $h_k(U, V)$ be $X_k$, and take $f_k(U), g_k(V)$ to be the $k$th position in $F(U)$, respectively, $G(V)$. Then $U, V, f_k, g_k, h_k$ satisfy the assumptions of the lemma, so that the corresponding error probabilities $p^u_k$ and $p^v_k$ lie in the region claimed.

In particular, if $p_u = p_v$, then $p_u \geq (\sqrt{2} - 1)/2$.

**III. INTERPRETATION**

It is important to realize that the above results are extremely strong assumptions stated in Section II, generally unachievable, on the resulting Hamming distance.

One could easily show the same bound as above, in the block reconstruction is correct with probability $1$, the alphabet sizes are $2^N (1 + \delta_1), 2^{N'} (1 + \delta_2)$ which are small. This, however, is still far from the Shannon limits, for one would have to prove that the average Hamming distance

$$p_u = \frac{1}{N} \sum_{k=1}^{N} p^u_k, \quad p_v = \frac{1}{N} \sum_{k=1}^{N} p^v_k,$$

(as opposed to each individual $p^u_k, p^v_k$ pair) satisfy the expected number of erroneous positions in $X^N$ is $\leq \delta_0$, and the $U, V$ alphabet sizes are sufficiently small $\delta_1, \delta_2$.

Thus, a slightly weaker conclusion has to be stated as a weaker assumption.

The best known bound under the Shannon assumptions is by the tangents to the hyperbola at the two points of the coordinate axis. This bound was first obtained by Cover and El Gamal, and is given by $p_u = p_v \geq \frac{1}{2}$ in the symmetric case.

On the other hand, work by Cover and El Gamal, and Ozarow, and Kaspi (private communication) has shown the Shannon assumptions, all points above the hyperbola.

It is an open conjecture that the hyperbola and the region of the achievable region in the Shannon sense is not. The conjecture is sustained by the fact that, in practice, the error probabilities for Gaussian sources with square law distortion are below the corresponding converse.

**IV. RELATED PROBLEMS**

In a problem of transmission of sampled images, Gersho proposed a scheme for graceful breakdown on the available redundancy. His scheme does not minimize the rate of transmission, but achieves the rate distortion, which led to the work above. In the general case, a memoryless source...
A receiver of the pair \((U, V)\) is able to block \(X_{1}^{N}\) without error. There are two other \(U\) only and producing a binary block \(Y_{1}^{N}\) =
receiving \(V\) only and producing the binary block

\[ p_{i} = Pr\{X_{i} \neq 0\}, \quad p_{i} = Pr\{X_{i} \neq Z_{i}\}. \]

The point \((p_{u}, p_{v})\) lies in the region of the

\[ p_{u} + \frac{1}{2} \left( p_{v} + \frac{1}{2} \right) \geq \frac{1}{2}. \]

Theorem, use is made of the following lemma, of

If, as it is a special case of results that will

cise two independent random variables (their

to measurable spaces). Let \(f(U_{0}), g(V_{0})\), and

\[ Pr\{f(U_{0}) \neq h(U_{0}, V_{0})\}, \quad Pr\{g(V_{0}) \neq h(U_{0}, V_{0})\}, \]

\[ Pr\{h(U_{0}, V_{0}) = 0\} = \frac{1}{2}. \]

Under assumptions, one has

\[ p_{u} + \frac{1}{2} \left( p_{v} + \frac{1}{2} \right) \geq \frac{1}{2}. \]

The best possible.

Concrete situation of the theorem, observe that

any \((U, V)\) pairs, \(2^{N}/2^{N} = 2^{N}\), as there are

condition of exact reconstructibility of \(X_{1}^{N}\) from

each pair \((U, V)\) corresponds to one, and only

the blocks all have the same probability \(2^{-N}\),

independent and uniformly distributed over

position. Exactly half the blocks, thus half the

0. Now we can let \(h_{s}(U, V)\) be \(X_{s}\), and take

with position in \(F(U)\), respectively, \(G(V)\). Then

the assumptions of the lemma, so that the

corresponding error probabilities \(p_{u}^{*}\) and \(p_{v}^{*}\) lie in the region which was

claimed.

In particular, if \(p_{u} = p_{v}\), then \(p_{u} \geq (\sqrt{2} - 1)/2.\)

III. INTERPRETATION

It is important to realize that the above result is derived under the

extremely strong assumptions stated in Section II. It is a lower bound,

generally unachievable, on the resulting Hamming distortion.

One could easily show the same bound as holding within \(\epsilon\) when the

block reconstruction is correct with probability \(\geq 1 - \delta_{0}\) and the \(U, V\)

alphabet sizes are \(2^{N/2}, 2^{N/2}, (1 + \delta_{0})\), \(2^{N}, \cdots, (1 + \delta_{0})\) where the \(\delta_{0}\) are suitably

small. This, however, is still far from the Shannon set-up, for which

one would have to prove that the average Hamming distortions

\[ p_{u} = \frac{1}{N} \sum_{k=1}^{N} p_{u}^{*}, \quad p_{v} = \frac{1}{N} \sum_{k=1}^{N} p_{v}^{*} \]

(as opposed to each individual \(p_{u}^{*}, p_{v}^{*}\) pair) satisfy the inequality within

\(\epsilon\), when the expected number of erroneous positions in the

reconstruction of \(X_{1}^{N}\) is \(\leq \delta_{0}\), and the \(U, V\) alphabet sizes are \(2^{N/2}, 2^{N/2}, (1 + \delta_{0})\) for

sufficiently small \(\delta_{0}\), \(i = 0, 1, 2\).

Thus, a slightly weaker conclusion has to be derived from a much

weaker assumption.

The best known bound under the Shannon assumptions is defined by

the tangents to the hyperbola at the two points where it cuts the

cordinate axis. This bound was first obtained by Wolf, Wyner, and

Ziv and gives \(p_{u} = p_{v} \geq \%\) in the symmetric case.

On the other hand, work by Cover and El Gamal and by Wyner,

Ozarow, and Kaspi (private communication) has shown that, under

the Shannon assumptions, all points above the hyperbola are

achievable.

It is an open conjecture that the hyperbola actually is the boundary

of the achievable region, in the Shannon sense. Belief in the validity

of this conjecture is sustained by the fact that, in an analogous situation

for Gaussian sources with square law distortion, Ozarow has obtained

the corresponding converse.

IV. RELATED PROBLEMS

In a problem of transmission of sampled speech waveforms, A.

Gersho proposed a scheme for graceful breakdown degradation based

on the available redundancy. His scheme does not work for i.i.d.

sources. This writer then proposed the breakdown problem as a source

network with rate distortion, which led to the work and results reported

above. In the general case, a memoryless source is encoded over \(n\)

SOURCE NETWORKS WITH MINIMAL DEGRADATION 1085
channels at rates $R_i$, $(i = 1, \ldots, n)$. There are $2^n - 1$ decoders, one for each nonvoid subset of channels. For a given distortion measure, the problem is to find the feasible combinations of distortions and rates. Above, only the case $n = 2$ was touched upon. There are several interesting questions for $n > 2$. One is the generalization of the key lemma, which will be taken up in another paper. Another is the question of the rates required to obtain error-free operation. This is the subject of the next section.

V. ERROR-FREE OPERATION AND REED-SOLOMON CODES

A discrete memoryless binary symmetric source is encoded over $n$ channels with equal rates $R$. Breakdowns can occur and will be sensed at the receiving end only. It is required that, for a certain value of $k$, $0 < k < n$, if any $k$ (or fewer) channels break down, the source will yet be reproduced without error. This is to be done with the smallest possible value of $R$.

One has $R \geq 1/(n-k)$, since a unit rate source must be accommodated by the remaining $n-k$ channels each at rate $R$. For $k = 1$, it is obvious that this bound is achievable. One need only take a block of $n-1$ source bits and assign one of them to each of the first $n-1$ channels; the last channel carries a parity check bit. This gives a rate of $(n-1)^{-1}$ and permits the recipient of any $n-1$ channels to reconstruct the missing channel by the parity condition.

For $k > 1$, the bound can also be achieved, using (truncated) Reed-Solomon codes, as follows.

For given $n$ and $k$, choose $r$ such that $n \leq 2^r - 1$. Then there exists a Reed-Solomon code$^6$ (a special BCH code), over $GF(2^r)$ of length $2^r - 1$ with prescribed minimum Hamming distance $d = k + 1$. This code has $2^r - d = 2^r - k - 1$ information symbols; that is, the code words form a subspace of dimension $2^r - d$ in the $(2^r - 1)$-dimensional vector space over $GF(2^r)$. If $n < 2^r - 1$, take the subset of code words having their first $2^r - n - 1$ symbols equal to zero. Dropping the zeros, one is left with a code of length $n$ with the same minimum distance $d = k + 1$. By the group property of BCH codes, the remaining code words fill a subspace of dimension $(2^r - d) - (2^r - n - 1) = n - d + 1 = n - k$ in the $n$ dimensional vector space over $GF(2^r)$. Thus, there are $2^{r(n-k)}$ code words.

This code is used as follows. Take a block of $r(n-k)$ binary source bits and assign to each of the possible blocks a distinct one of the $2^{r(n-k)}$ code words. The $i$th symbol in this length $n$ code word is an element of $GF(2^r)$; it can be viewed as a block of $r$ binary bits, and these bits are sent over the $i$th channel.

The receiver will know the $n-k$ symbols from the surviving channels, the others being erased. Reconstruction is possible, as the code word actually sent is the only one compatible with the data, for if a second one were such, the Hamming distance of the words would be at most $k$, and this is only if there are $n$ channels.

In this way, $r$ bits are sent over each channel, and $r(n-k)$ source bits, which achieves the rate of $(n-k)/(n-1)$.

Remark that the crucial property of the construction is the post-office channel,$^6$ where truncation is also made. This formulation is asymptotic so that $n$ can always be made great.

REFERENCES


SOURCE NETWORKS WITH MINIMAL CHANNEL DISTORTION
There are $2^n - 1$ decoders, one for each channel. For a given distortion measure, the feasible combinations of distortions and rates. $n = 2$ was touched upon. There are several for $n > 2$. One is the generalization of the key taken up in another paper. Another is the required to obtain error-free operation. This is the section.

REED-SOLOMON CODES

A binary symmetric source is encoded over $n$ symbols $R$. Breakdowns can occur and will be sensed by $n - k$ channels each of rate $R$. For $k = 1$, it is is achievable. One need only take a block of $n - 1$ channel carries a parity check bit. This gives a rate and can also be achieved, using (truncated) Reed-Solomons.

Choose $r$ such that $n \leq 2^r - 1$. Then there exists a special BCH code, over $GF(2^r)$ of length $n$ and minimum Hamming distance $d = k + 1$. This $k - 1$ information symbols; that is, the code of dimension $2^r - d$ in the $(2^r - 1)$-dimensional $GF(2^r)$. If $n < 2^r - 1$, take the subset of code words $n - 1$ symbols equal to zero. Dropping the zeros, the length $n$ with the same minimum distance and, property of BCH codes, the remaining code of dimension $(2^r - d) - (2^r - n - 1) = n - d + d$ dimensional vector space over $GF(2^r)$. Thus, there follows. Take a block of $r(n - k)$ binary source of the possible blocks a distinct one of the $i$th symbol in this length $n$ code word is an the $i$th channel.

know the $n - k$ symbols from the surviving being erased. Reconstruction is possible, as the code word actually sent is the only one compatible with the received data, for if a second one was such, the Hamming distance of these words would be at most $k$, and this is one less than the minimum distance $d$ of the code. (See Ref. 5 for decoding algorithms.)

In this way, $r$ bits are sent over each channel to transmit $r(n - k)$ source bits, which achieves the rate of $(n - k)^{-1}$ as claimed.

Remark that the crucial property of the codes used is that they are MDS codes. Note also that Reed-Solomon codes achieve capacity for the post-office channel, where truncation is not required because the formulation is asymptotic so that $n$ can always be increased.

REFERENCES