APPENDIX

A. Proof of Theorem 1

We start with an auxiliary Lemma.

Lemma 1: For any $i \in [t-\tau]$, all $j \in \mathbb{B}_i^{(s)}$ where $\sum_{r=j}^{j+\tau-1} \zeta_r = 0$, and any $l \in \{j, \ldots, i\}$, $w^{(i,j,l)} \geq (1-\ell_l)u_l$.

Proof: First, $(1-\ell_l)n_l \geq (1-\ell_l)k_l \geq (1-\ell_l)u_l$.

For $l = j$, combining Equation 7 with $k_l \geq u_l$ yields

$$w^{(i,j,l)} \geq (k_l - u_l\ell_l) \geq (u_l - u_l\ell_l) = u_l(1-\ell_l) \geq 0. \quad (23)$$

For $l \in \{j+1, \ldots, i\}$, the case necessitates that $j < i$. First, by the inductive hypothesis on $(l-1)$,

$$w^{(i,j,l-1)} \leq k_{l-1} - u_{l-1}\ell_{l-1} + \sum_{r=j}^{l-2} k_r - u_r\ell_r - w^{(i,j,r)} \quad (24)$$

$$\sum_{r=j}^{l-1} w^{(i,j,r)} \leq \sum_{r=j}^{l-1} k_r - u_r\ell_r \quad (25)$$

where Equation 25 follows from rearranging terms. Thus, $w^{(i,j,l)} \geq (1-\ell_l)u_l$ by Equation 7 and the logic for $l = j$. ■

For any $i \in [t]$, the lossless-delay is met as $S[i]$ is in $X[i]$.

Next, to show satisfaction of the worst-case-delay, we consider any burst starting in $i \leq [t-\tau]$. We need not consider $i > (t-\tau)$, as the final τ message packets are known to be 0. If $\sum_{l=i+1}^{i+\tau-1} \zeta_l > 0$ then $S[i : i+b_i-1]$ need not be recovered, and the proof is concluded. Otherwise, $\sum_{l=i+1}^{i+\tau-1} \zeta_l = 0$. We show in two steps that each $S[i : i+b_i-1]$ is recovered within τ timeslots. First, the received symbols of $Y[i : i+b_i-1]$ and $P[i+b_i : i+\tau-1]$ are used to recover $V[i : i+b_i-1]$. Second, for $j \in \{i, \ldots, i+b_i-1\}$ where $\sum_{l=i+1}^{j+\tau-1} \zeta_l = 0$, $U[j]$ is recovered in time slot $(i+\tau)$ with $P[i+\tau]$.

First, for $j \in \{i, \ldots, i+\tau-1\}$, $U[j-\tau]$ is used to compute $P^{(*)}[j]$ to determine $P'[j] = (P[j] - P^{(*)}[j])$ (by Equation 13). Recall from Equation 9 (and $v_l = (k_l - u_l)$)

$$\sum_{l=i+b_i}^{i+\tau-1} p'[l] + \sum_{l=i}^{i+b_i-1} w^{(j,i,l)} \geq \sum_{l=i}^{i+b_i-1} k_l - u_l\ell_l \quad (26)$$

$$\sum_{l=i+b_i}^{i+\tau-1} p'[l] + \sum_{l=i}^{i+b_i-1} w^{(j,i,l)} - (1-\ell_l)u_l \geq \sum_{l=i}^{i+b_i-1} v_l \quad (27)$$

where for $l \in \{i, \ldots, i+b_i-1\}$ the $(1-\ell_l)u_l$ symbols of $U[l]$ that are received are subtracted out (valid by Lemma 1).

Without loss of generality, we pretend that each $P[j]$ is padded with extra parity symbols to be $P^{(pad)}[j]$ of size m but these extra $(m - p[j])$ symbols are all lost. Then each $(V^*[j], P^{(pad)}[j])$ comprises $2m$ symbols. The number of received parity symbols exceeds the number of missing message symbols. Thus, combining Equation 27 with Lemma 1 L1 of [3] shows $V[i]$ is recovered by time slot $(i+\tau-1)$ (e.g., by solving a system of linear equations). For $r = (i+1), \ldots, (i+b_i-1)$, by Lemma 1 and Equation 7,

$$\sum_{l=i}^{r-1} w^{(i,i+b_i-1,l)} \leq \sum_{l=i}^{r-1} k_l - u_l\ell_l \quad (28)$$

$$\sum_{l=i}^{r-1} w^{(i,i+b_i-1,l)} - u_l(1-\ell_l) \leq \sum_{l=i}^{r-1} v_l \quad (29)$$

Combining Equations 27 and 29 shows

$$\sum_{l=i+b_i}^{i+\tau-1} p'[l] + \sum_{l=r}^{i+b_i-1} (w^{(j,i,l)} - u_l(1-\ell_l) \geq \sum_{l=r}^{i+b_i-1} v_l\ell_l \quad (30)$$

Combining Equation 30 with Lemma 1 L1 of [3], $V[r]$ can be recovered by time slot $(i+\tau-1)$ (e.g., by solving a system of linear equations). After finishing iteration $r = (i+b_i-1)$, $V[i : i+b_i-1]$ have been recovered.

For $l = i, \ldots, (i+b_i-1)$ where $\sum_{r=i+1}^{l+\tau-1} \zeta_r = 0$, we now show that $U[l]$ is recovered by time slot $(i+\tau)$. During time slot $(l+\tau)$, $V[l : l+\tau]$ are available and are used to compute $P'[l+\tau]$, yielding $P^{(*)}[l+\tau] = (P[l+\tau] - P'[l+\tau])$. Then $P^{(*)}[l+\tau]$ comprises $p[l+\tau] \geq U[l]\ell_l$ linearly independent linear equations of the symbols of $U[l]$. Combining $P^{(*)}[l+\tau]$, the $(1-\ell_i)u[i]$ received symbols of $U[l]$, and the $(m - u[l])$ zeroes of $U^*[l]$ (padding) provides at least m symbols of the $[2m, m]$ maximum distance separable linear code. Thus, $U[l]$, is obtained by solving a system of linear equations. Both $V[l]$ and $U[l]$ are recovered by time slot $(l+\tau)$ for any $l \in [i, \ldots, i+b_i-1]$, so the worst-case-delay constraint is satisfied.

B. Properties of a relaxed offline code

Throughout Section B, we assume for all $i \in [t]$ that $\zeta_i = 0$. Let code construction, C, be any offline construction that satisfies the lossless-delay and worst-case-delay constraints. Under C, $X^{(C)}[i]$ is sent during time slot $i \in [t]$ of size $n_{C,i}$ comprising $(c_{C,i} + 1)$ packets.

1) Preliminaries: We begin with defining a few terms for time slot $i \in [t-\tau]$. Will use the following terms:

$$u_i^{(C)} = \min_{j \in \mathbb{B}_i^{(s)}, Y^{(C)}[j:j+b_j-1] \in \mathcal{Y}_j} \big(\mathcal{H}^{(e)}(S[i] | S[0 : i-1], Y^{(C)}[i], X^{(C)}[i+1 : i+\tau-1])\big) \quad (31)$$

$$v_i^{(C)} = k_i - u_i^{(C)} \quad (32)$$

$$p_i^{(C)} = n_i - k_i. \quad (33)$$

2) Useful identities:

Lemma 2: For any $i \in [t-\tau]$ such that $\mathbb{B}_i^{(s)} \neq \emptyset, \ell_i > 0$ the size of the $(i+\tau)$th channel packet is at least

$$n_{C,i+\tau} \geq \left\lceil u_i^{(C)} + k_{i+\tau} \right\rceil.$$

Proof: At a high level, by Equation 31, at least $u^{(C)}[i]\mathcal{H}(e)$ bits of information are needed to recover $S[i]$ that are unavailable prior to time slot $(i+\tau)$. They must be recovered during time slot $(i+\tau)$ due to the worst-case-delay, leading to at least $\lceil u^{(C)}[i] \rceil$ symbols being sent in $X^{(C)}[i+\tau]$. The lossless-delay constraint for $S[i+\tau]$ necessitates an additional $k_{i+\tau}$ symbols be sent in $X^{(C)}[i+\tau]$.

By Equations 32 and 31, there is a $i' \in \mathbb{B}_i^{(s)}$ and a burst $Y^{(C)}[i' : i' + b_{i'} - 1] \in \mathcal{Y}_{i'}$ so that

$$\left(\mathcal{H}^{(e)}(S[i]|S[0:i-1], Y^{(C)}[i:i+\tau-1])\right) = u_i^{(C)}. \quad (34)$$

Recall that $S[0:i-1], Y^{(C)}[i:i+\tau-1]$ are available by time slot $(i+\tau-1)$. By Equation 3 and the chain rule,

$$\mathcal{H}^{(e)}(S[i]|S[0:i-1], Y^{(C)}[i:i+\tau]) = 0. \quad (35)$$

Thus,

$$n_{C,i+\tau} \geq \mathcal{H}^{(e)}(Y^{(C)}[i+\tau]) \geq \quad (36)$$
$$\mathcal{H}^{(e)}(Y^{(C)}[i+\tau]|S[0:i-1], Y^{(C)}[i:i+\tau-1]) = \quad (37)$$
$$\mathcal{H}^{(e)}(Y^{(C)}[i+\tau]|S[0:i-1], Y^{(C)}[i:i+\tau-1]) + \mathcal{H}^{(e)}(S[i]|S[0:i-1], Y^{(C)}[i:i+\tau]) = \quad (38)$$
$$\mathcal{H}^{(e)}(S[i], Y^{(C)}[i+\tau]|S[0:i-1], Y^{(C)}[i:i+\tau-1]) \geq \quad (39)$$
$$\mathcal{H}^{(e)}(S[i]|S[0:i-1], Y^{(C)}[i:i+\tau-1]) + \mathcal{H}^{(e)}(Y^{(C)}[i+\tau]|S[0:i], Y^{(C)}[i:i+\tau-1]) = \quad (40)$$
$$u_i^{(C)} + \mathcal{H}^{(e)}(Y^{(C)}[i+\tau]|S[0:i], Y^{(C)}[i:i+\tau-1]) \geq \quad (41)$$
$$u_i^{(C)} + \mathcal{H}^{(e)}(Y^{(C)}[i+\tau]|S[0:i+\tau-1]) \geq \quad (42)$$
$$u_i^{(C)} + k_{i+\tau} \quad (43)$$

Equation 37 comes from conditioning reducing entropy; in Equation 38, the extra added term is 0 due to Equation 35; Equation 39 comes from applying the chain rule to Equation 38; Equation 40 comes from applying the chain rule to Equation 39; Equation 41 comes from Equation 34; Equation 42 comes from $Y^{(C)}[i:i+\tau-1]$ being a function of $S[0:i+\tau-1]$ and conditioning reducing entropy; Equation 43 comes from Equation 4. Finally, combining Equations 36 and 43 establishes the result.

■

We begin by bounding the amount of information the parity symbols of each channel packet provide about earlier message packets.

Lemma 3: For any $i \in [t], j \in [i-1]$, the amount of information a channel packet, $X^{(C)}[i]$, can provide about dropped symbols of earlier message packets is bounded by

$$\mathcal{I}^{(e)}(X^{(C)}[i]; S[j:\min(j+b_j, i)-1]|S[0:j-1]) \leq (n_{C,i} - k_i)$$

Proof:

$$\mathcal{I}^{(e)}(X^{(C)}[i]; S[j:\min(j+b_j, i)-1]|S[0:j-1]) = \quad (44)$$
$$\mathcal{H}^{(e)}(X^{(C)}[i]|S[0:j-1]) - \mathcal{H}^{(e)}(X^{(C)}[i]|\min(j+b_j, i)-1) \leq \quad (45)$$
$$\mathcal{H}^{(e)}(X^{(C)}[i]|S[0:j-1]) - \mathcal{H}^{(e)}(X^{(C)}[i]|S[0:i-1]) \leq \quad (46)$$
$$(n_{C,i} - k_i) \quad (47)$$

where Equation 45 comes from the definition of Mutual Information; Equation 46 comes from conditioning reducing information; Equation 47 comes from $X^{(C)}[i]$ having at most $n_{C,i}$ symbols and Equation 4. ■

Lemma 3 will later be used to show for a burst starting in time slot j how much redundancy channel packets received after the burst can provide to help in loss recovery.

Next, we bound the amount of information later channel packets in a burst can provide about earlier message packets in Lemma 4

Lemma 4: $\forall i \in [t-\tau] \forall j \in \mathbb{B}_i^{(s)} \forall l \in \{j, \ldots, i-1\}$, the amount of information about the missing message symbols of $S[j:l]$ provided in $Y^{(C)}[i]$ is at most

$$\mathcal{I}^{(e)}(Y^{(C)}[i]; S[j:l]|S[0:j-1], Y^{(C)}[j:i-1]) \leq (n_{C,i} - k_i) - \lceil \ell_i(n_{C,i} - k_i) \rceil.$$

Proof:

First, we provide a lower bound on the mutual information between $Y^{(C)}[i]$ and $S[i]$, as this information does not help recover $S[j:l]$. To do so, we break Equation 4 into an expression about each packet sent in $X^{(C)}[i]$ as follows

$$k_i = \mathcal{H}^{(e)}(X^{(C)}[i]|S[0:i-1]) = \sum_{r \in [c_{C,i}]} \mathcal{H}^{(e)}(X_{(r)}^{(C)}[i]|S[0:i-1], X_{(0)}^{(C)}[i], \ldots, X_{(r-1)}^{(C)}[i]).$$

Hence, on average over $r \in [c_{C,i}]$,

$$\mathcal{H}^{(e)}(X_{(r)}^{(C)}[i]|S[0:i-1], X_{(0)}^{(C)}[i], \ldots, X_{(r-1)}^{(C)}[i]) = \frac{k_i}{c_{C,i}+1}.$$

So there exists a subset L of size $l = \lceil \ell_i(c_{C,i}+1) \rceil$ smallest elements so that

$$\sum_{r \in L} \mathcal{H}^{(e)}(X_{(r)}^{(C)}[i]|S[0:i-1], X_{(0)}^{(C)}[i], \ldots, X_{(r-1)}^{(C)}[i]) \leq k_i l/(C_i+1) \leq \ell_i k_i. \quad (48)$$

Let $L' = \{l_0, \ldots, l_{l'-1}\} = [c_{C,i}] \setminus L$ where $l_{r_1} < l_{r_2}$ for any $r_1 < r_2 \in [l'-1]$. Conditioning reduces entropy, leading to

$$\sum_{r \in [l'-1]} \mathcal{H}^{(e)}(X_{(l_r)}^{(C)}[i]|S[0:i-1], X_{(l_0)}^{(C)}[i], \ldots, X_{(l_{r-1})}^{(C)}[i]) \geq \quad (49)$$
$$\sum_{r \in L'} \mathcal{H}^{(e)}(X_{(r)}^{(C)}[i]|S[0:i-1], X_{(0)}^{(C)}[i], \ldots, X_{(r-1)}^{(C)}[i]) = \quad (50)$$
$$k_i - \sum_{r \in L} \mathcal{H}^{(e)}(X_{(r)}^{(C)}[i]|S[0:i-1], X_{(0)}^{(C)}[i], \ldots, X_{(r-1)}^{(C)}[i]) \geq \quad (51)$$
$$k_i(1-\ell_i) \quad (52)$$

where Equation 50 follows from conditioning reducing entropy; Equation 51 follows from $[c_{C,i}] = L \cup L'$; Equation 52 follows from Equation 48. Finally, recall that $k_i \ell_i$ is an integer. We note that there exists a lossy channel leading to $Y^{(C)}[i]$ reflecting losing all packets corresponding to L. In this case,

$$\mathcal{H}^{(e)}(Y^{(C)}[i]|S[0:i-1]) \geq k_i(1-\ell_i). \quad (53)$$

Next, we use the above identities to provide an upper bound on the mutual information between $Y^{(C)}[i]$ and $S[j:i-1]$ as

$$\mathcal{I}^{(e)}(Y^{(C)}[i]; S[j:i-1] | S[0:j-1], Y^{(C)}[j:i-1]) = \tag{54}$$

$$\begin{aligned}&\mathcal{H}^{(e)}(Y^{(C)}[i] | S[0:j-1], Y^{(C)}[j:i-1]) - \\ &\mathcal{H}^{(e)}(Y^{(C)}[i] | S[0:i-1]) \leq\end{aligned} \tag{55}$$

$$(n_{C,i} - \lceil \ell_i n_{C,i} \rceil) - k_i(1-\ell_i) \leq \tag{56}$$

$$(n_{C,i} - k_i) - \lceil \ell_i (n_{C,i} - k_i) \rceil \tag{57}$$

where Equation 55 comes from the definition of mutual information and the fact that $Y^{(C)}[j:i-1]$ is a function of $S[0:i-1]$; Equation 56 comes from bounding the size of $Y^{(C)}[i]$ given a worst-case loss and Equation 53; and Equation 57 comes from rearranging terms and noting that $\ell_i k_i$ is an integer. Finally, we note that this upper bound also applies to the mutual information between $Y^{(C)}[i]$ and $S[j:l]$ due to the identity $l < i$ and non-negativity of mutual information:

$$\begin{aligned}&\mathcal{I}^{(e)}(Y^{(C)}[i]; S[j:l] | S[0:j-1], Y^{(C)}[j:i-1]) \leq \\ &\mathcal{I}^{(e)}(Y^{(C)}[i]; S[j:i-1] | S[0:j-1], Y^{(C)}[j:i-1]).\end{aligned}$$

■

For a burst starting in time slot j of length b_j and $i \leq (j + b_j - 1)$, Lemma 4 upper bounds how much information $Y^{(C)}[i]$ provides for recovering $S[j:i-1]$.

We show for any $i \in [t-\tau]$ there is as a burst starting in time slot $j \in \mathbb{B}_i^{(s)}$ such that all symbols of $S[j:i-1]$ and all but $u_i^{(C)}$ symbols of $S[i]$ must be recovered by time slot $(i+\tau-1)$.

Lemma 5: Consider any $i \in [t-\tau]$ and any $j \in \mathbb{B}_i^{(s)}$. Then

$$\min_{Y^{(C)}[j:j+b_j-1] \in \mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i]; Y^{(C)}[j:j+\tau-1] | S[0:j-1])\big) \leq \sum_{l=j}^{i} k_l - u_l^{(C)}.$$

Proof:

$$\mathcal{H}^{(e)}(S[j:i] | S[0:j-1], Y^{(C)}[j:j+\tau-1]) = \tag{58}$$

$$\sum_{l=j}^{i} \mathcal{H}^{(e)}(S[l] | S[0:l-1], Y^{(C)}[j:j+\tau-1]) \geq \tag{59}$$

$$\sum_{l=j}^{i} \mathcal{H}^{(e)}(S[l] | S[0:l-1], Y^{(C)}[l:l+\tau-1]) \geq \tag{60}$$

$$\sum_{l=j}^{i} u_l^{(C)}. \tag{61}$$

where Equation 59 follows from the chain rule; Equation 60 comes from conditioning reducing entropy; Equation 61 follows from the choice of losses Equation 31.

We note that

$$\mathcal{I}^{(e)}(S[j:i]; Y^{(C)}[j:i+\tau-1] | S[0:j-1]) = \tag{62}$$

$$\begin{aligned}&\mathcal{H}^{(e)}(S[j:i] | S[0:j-1]) - \\ &\mathcal{H}^{(e)}(S[j:i] | S[0:j-1], Y^{(C)}[j:i+\tau-1])\end{aligned} \tag{63}$$

$$= \sum_{l=j}^{i} k_l - u_l^{(C)} \tag{64}$$

where Equation 63 follows from the definition of Mutual Information; Equation 64 follows from the sizes of $S[j:i]$/independence of message packets and Equation 61. ■

We note another form for the worst-case-delay constraint (e.g., Equation 3).

Lemma 6: Consider any $i \in [t-\tau]$ and a burst starting in time slot $j \in \mathbb{B}_i^{(s)}$. Then for any $l \in \{j+\tau, \ldots, i+\tau\}$

$$\min_{Y^{(C)}[j:j+b_j-1] \in \mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i]; Y^{(C)}[j:l] \mid S[0:j-1])\big) \leq \sum_{r=j}^{i} k_r - \sum_{r=l-\tau+1}^{i} u_r^{(C)},$$

with equality when $l = (i+\tau)$ or $l = (i+\tau-1)$.

Proof:

$$\mathcal{I}^{(e)}(S[j:i]; Y^{(C)}[j:l] \mid S[0:j-1]) = \tag{65}$$

$$\begin{aligned}&\mathcal{H}^{(e)}(S[j:i] \mid S[0:j-1]) - \\ &\mathcal{H}^{(e)}(S[j:i] \mid S[0:j-1], Y^{(C)}[j:l]) =\end{aligned} \tag{66}$$

$$\Big(\sum_{r=j}^{i} k_r\Big) - \sum_{r=j}^{i} \mathcal{H}^{(e)}(S[r] \mid S[0:r-1], Y^{(C)}[j:l]) = \tag{67}$$

$$\begin{aligned}&\Big(\sum_{r=j}^{i} k_r\Big) - \sum_{r=j}^{l-\tau} \mathcal{H}^{(e)}(S[r] \mid S[0:r-1], Y^{(C)}[r:r+\tau]) - \\ &\sum_{r=l-\tau+1}^{i} \mathcal{H}^{(e)}(S[r] \mid S[0:r-1], Y^{(C)}[r:l]) \leq\end{aligned} \tag{68}$$

$$\begin{aligned}&\Big(\sum_{r=j}^{i} k_r\Big) - \\ &\sum_{r=l-\tau+1}^{i} \mathcal{H}^{(e)}(S[r] \mid S[0:r-1], Y^{(C)}[r:r+\tau-1])\end{aligned} \tag{69}$$

$$= \Big(\sum_{r=j}^{i} k_r\Big) - \sum_{r=l-\tau+1}^{i} u_r^{(C)} \tag{70}$$

where Equation 66 follows from the definition of mutual information; Equation 67 follows from independence of message packets and the chain rule; Equation 68 follows the first $(l-\tau-j+1)$ subtracted terms being 0 and $Y^{(C)}[j:r-1]$ being functions of $S[0:r-1]$ for the remaining subtracted terms; Equation 69 follows from Equation 3 and conditioning reducing entropy where Equation 69 equals Equation 68 whenever $l = (i+\tau)$ or $l = (i+\tau-1)$ since then conditioning was not used; Equation 70 follows from Equation 31. ■

We rewrite the worst-case-delay constraint in terms of mutual information as follows.

Lemma 7: Consider any $i \in [t-\tau]$ and any $j \in \mathbb{B}_i^{(s)}$. Then

$$\min_{Y^{(C)}[j:j+b_j-1]\in\mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:i+\tau] \mid S[0:j-1])\big) = \sum_{l=j}^{i} k_l.$$

Proof:

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:i+\tau] \mid S[0:j-1]) = \tag{71}$$

$$\mathcal{H}^{(e)}(S[j:i] \mid S[0:j-1]) - \mathcal{H}^{(e)}(S[j:i], \mid S[0:j-1],Y^{(C)}[j:i+\tau]) = \tag{72}$$

$$\sum_{l=j}^{i} k_l - \sum_{l=j}^{i} \mathcal{H}^{(e)}(S[l], \mid S[0:j-1],Y^{(C)}[l:l+\tau]) = \tag{73}$$

$$\sum_{l=j}^{i} k_l. \tag{74}$$

Where Equation 72 follows from the definition of mutual information; Equation 73 follows from independence of message packets and the fact that the Equation 3 shows for any $l \in \{j,\ldots,i\}$, $\mathcal{H}^{(e)}(S[l], \mid S[0:j-1],Y^{(C)}[l:l+\tau]) = 0$, leading to Equation 74. ■

3) Relaxations: We will use the following relaxations which may increase (but never decrease) the mutual information between received symbols under C and missing information. Consider any $i \in [t-\tau]$ and any $j \in \mathbb{B}_i^{(s)}$.

Relaxation 1.

$$\min_{Y^{(C)}[j:j+b_j-1]\in\mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[i] \mid S[0:j-1],Y^{(C)}[j:i-1])\big) =$$

$$\min\big(n_{C,i,Y^{(C)}}, k_i - u_i^{(C)} + \sum_{r=j}^{i-1} k_r - u_r^{(C)} - \mathcal{I}^{(e)}(S[j:r];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1])\big).$$

Relaxation 2. For any $l \in \{i+1,\ldots,j+b_j-1\}$,

$$\min_{Y^{(C)}[j:j+b_j-1]\in\mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])\big) =$$

$$\min\big(\lfloor p_l^{(C)} - \ell_l p_l^{(C)}\rfloor, \sum_{r=j}^{i} k_r - u_r^{(C)} - \sum_{r=j}^{l-1} \mathcal{I}^{(e)}(S[j:\min(r,i)];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1])\big).$$

Relaxation 3. For any $l \in \{j+b_j,\ldots,j+\tau-1\}$,

$$\min_{Y^{(C)}[j:j+b_j-1]\in\mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i];X^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])\big) =$$

$$\min\big(n_{C,l} - k_l, \sum_{r=j}^{i} k_r - u_r^{(C)} - \sum_{r=j}^{l-1} \mathcal{I}^{(e)}(S[j:\min(r,i)];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1])\big).$$

Relaxation 4. For any $l \in \{j+\tau,\ldots,i+\tau\}$,

$$\min_{Y^{(C)}[j:j+b_j-1]\in\mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i];X^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])\big) =$$

$$\min\big(n_{C,l} - k_l, k_{l-\tau} + \sum_{r=l-\tau+1}^{i} k_r - u_r^{(C)} - \sum_{r=l-\tau}^{l-1} \mathcal{I}^{(e)}(S[l-\tau:\min(r,i)];Y^{(C)}[r] \mid S[0:r-1],Y^{(C)}[l-\tau:r-1])\big).$$

Lemma 8: Relaxations 1,2,3, and 4 do not cause C to send extra symbols or violate any constraints.

Proof:

Relaxation 1 . We note

$$\sum_{l=j}^{i} \big(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[l] \mid S[0:l-1],Y^{(C)}[j:l-1]\big) = \tag{75}$$

$$\sum_{l=j}^{i} \big(\mathcal{I}^{(e)}(S[j:l];Y^{(C)}[l] \mid S[0:l-1],Y^{(C)}[j:l-1]\big) = \tag{76}$$

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:i] \mid S[0:j-1]) \leq \tag{77}$$

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:j+\tau-1] \mid S[0:j-1]) \leq \tag{78}$$

$$\left(\sum_{l=j}^{i} k_l - u_i^{(C)}\right) \tag{79}$$

where Equation 76 comes from the independence of $S[l+1:i]$ from $S[0:l],Y^{(C)}[j:l]$; Equation 77 comes from the chain rule for mutual information; Equation 78 comes from $(j+\tau-1) \geq i$, the chain rule for mutual information, and non-negativity of mutual information; Equation 79 comes from Lemma 5. Combining Equation 75 with Equation 79 and the fact that $\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[i] \mid S[0:j-1],Y^{(C)}[j:i-1])$ is at most the size of $Y^{(C)}[i]$ (i.e., $n_{C,i,Y^{(C)}}$) establishes that the relaxation only maintains or increases the mutual information.

Relaxation 2 . By Lemma 4, $\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])$ is no more than $(\lfloor p_i^{(C)} - \ell_i p_i^{(C)}\rfloor$. Additionally,

$$\sum_{l=j}^{l} \big(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[l] \mid S[0:l-1],Y^{(C)}[j:l-1]\big) = \tag{80}$$

$$\sum_{l=j}^{i}\left(\mathcal{I}^{(e)}(S[j:l];Y^{(C)}[l] \mid S[0:l-1],Y^{(C)}[j:l-1])+\right.$$

$$\sum_{l=i+1}^{l}\left(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[l] \mid S[0:l-1],Y^{(C)}[j:l-1]\right) = \tag{81}$$

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:i] \mid S[0:j-1]) \leq \tag{82}$$

$$\left(\sum_{l=j}^{i} k_l - u_i^{(C)}\right) \tag{83}$$

where Equation 81 comes from the independence of $S[l+1:i]$ from $S[0:l],Y^{(C)}[j:l]$; Equation 82 comes from the chain rule for mutual information; Equation 83 comes from Lemma 5 and $i \leq (j+\tau-1)$.

Relaxation 3 . By Equations 3, $\mathcal{I}^{(e)}(S[j:i];X^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])$ is at most $(n_{C,l}-k_l)=p_l^{(C)}$. Also,

$$\sum_{r=j}^{l}\left(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1]\right) = \tag{84}$$

$$\sum_{r=j}^{i}\left(\mathcal{I}^{(e)}(S[j:r];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1])+\right.$$

$$\sum_{r=i+1}^{l}\left(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1]\right) = \tag{85}$$

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:l] \mid S[0:j-1]) \leq \tag{86}$$

$$\left(\sum_{r=j}^{i} k_r - u_r^{(C)}\right) \tag{87}$$

where Equation 85 comes from the independence of $S[l+1:i]$ from $S[0:l],Y^{(C)}[j:l]$; Equation 86 comes from the chain rule for mutual information; Equation 87 comes from Lemma 5 and $l \leq (j+\tau-1)$.

Relaxation 4 . By Equations 3, $\mathcal{I}^{(e)}(S[j:i];X^{(C)}[l] \mid S[0:j-1],Y^{(C)}[j:l-1])$ is at most $(n_{C,l}-k_l)=p_l^{(C)}$. Additionally,

$$\sum_{r=j}^{l}\left(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1]\right) = \tag{88}$$

$$\sum_{r=j}^{i}\left(\mathcal{I}^{(e)}(S[j:r];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1])+\right. \tag{89}$$

$$\sum_{r=i+1}^{l}\left(\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[r] \mid S[0:j-1],Y^{(C)}[j:r-1]\right) = \tag{90}$$

$$\mathcal{I}^{(e)}(S[j:i];Y^{(C)}[j:l] \mid S[0:j-1]) \leq \tag{91}$$

$$\sum_{r=j}^{i} k_r - \sum_{r=l-\tau+1}^{i} u_r^{(C)} \tag{92}$$

where the logic through Equation 91 is the same as in the proof for Relaxation 3; Equation 92 follows from Lemma 6. ■

The relaxations lead to a mutual information that depends on the sizes of message packets and channel packets, not the symbols that are sent under C themselves.

Relaxation 5. For any $i \in [t], (c_{C,i}+1) = n_{C,i}$.

Lemma 9: Relaxation 5 does not cause C to send extra symbols or violate any constraints.

Proof: All considered bounds in the proof (under the relaxation) apply to any choice for packetization. The only way that packetization effects the bounds is if it changes the total number of received symbols, with the greedy choice being to receive as many symbols as possible for each channel packet.

Relaxation 5 has no change for $i > (t-\tau)$. We apply the change for $i = 0,\ldots,(t-\tau)$; this does not alter the total number of symbols sent (or even the total number sent in any channel packet $X^{(C)}[i]$).

Next, we show that this change does not effect decoding. When $\ell_i n_{C,i}$ is an integer, then the minimum possible number of symbols lost over $X^{(C)}[i]$ is $\ell_i n_{C,i}$ and that is what is lost in the worst case.

Otherwise, exactly $\lceil \ell_i n_{C,i} \rceil$ symbols are lost in the worst case. For any $(c_{C,i}+1)$ packets sent in $X^{(C)}[i]$, the largest $l = \lceil \ell_i(c_{C,i}+1) \rceil$ could be lost, which contain in total at least $\lceil l\frac{n_{C,i}}{(c_{C,i}+1)} \rceil \geq \lceil \ell_i n_{C,i} \rceil$ symbols by the pigeonhole principle and packets containing an integral number of symbols. Thus, one symbol per packet leads to as many symbols being received as possible, which maximizes their utility. ■

Relaxation 6 . $\forall l \in [\tau-1]$, each symbol of $X[l]$ has $\mathcal{H}(e)$ bits of information about $S[0:l]$. The correctness of this is immediate.

Corollary 2: Adjusting C so that $p_j^{(C)} = 0$ for $j < \tau$ and for any $i \in [t-\tau]$,

$$p_{i+\tau}^{(C)} = u_i^{(C)} \tag{93}$$

leads to at most $2(t-\tau)(\tau-1)$ extra symbols being sent.

Proof: We start by showing without loss of generality, $p_l^{(C)} = 0$ for $l < \tau$. When a burst of length b_0 starts in $X^{(C)}[0]$, it is recovered by time slot $(\tau+b_0-1)$ then $S[0:\tau+b_0-1]$ are available by the same time slot $(\tau+b_0-1)$ by Lemma 6 and Equation 4. In addition, $n_{C,l} \geq k_l$ for $l < b_0$ by Equation 4, so a worst case burst drops at least $k_l\ell_l$ symbols of $X^{(C)}[l]$. Thus,

$$\sum_{l=b}^{\tau+b_0-1} n_{C,l} \geq \sum_{l=0}^{b_0-1} \ell_l k_l + \sum_{l=b_0}^{\tau+b_0-1} k_l.$$

For each $l \in [b_0-1]$ while $n_{C,l+\tau} < (k_{l+\tau}+\ell_l k_l)$, for $j \in [\tau-1]$ we move all but k_j symbols of $X^{(C)}[j]$ to $X^{(C)}[l+\tau]$.

For $l \in [\tau-1]$, move all but k_l symbols of $S[l]$ to $X^{(C)}[\tau+b_0-1]$ and spread all symbols of $X^{(C)}[l]$ evenly over the packets of $X^{(C)}[l]$. By Relaxation 6 and $\ell_l | k_l$ for $l \in [t]$, loss recovery is no worse. The same number of symbols are sent.

We prove by induction that we can alter C to obey Equation 93 at the cost of 2 symbols per time slot.

Suppose for the inductive hypothesis that

$$\forall j \in [i_*], p_{j+\tau}^{(C)} = u_j^{(C)}. \tag{94}$$

The inductive hypothesis has been shown to hold for $i_* = (b_0 - 1)$.

Now we apply induction for $i = (i_* + 1)$. We know by Lemma 2 that $p_{i+\tau}^{(C)} \geq u_i^{(C)}$ before we altered C. Only moving parity symbols from $X^{(C)}[i+\tau-1]$ to $X^{(C)}[i+\tau]$ may have changed the definition of $u_i^{(C)}$, but doing so would maintain the inequality $p_{i+\tau}^{(C)} \geq u_i^{(C)}$. Let $\delta_i = (p_{i+\tau}^{(C)} - u_i^{(C)})$ We now move δ symbols from $X^{(C)}[i+\tau]$ to $X^{(C)}[i+\tau+1]$, which increases $p_{i+\tau+1}^{(C)}$ by δ and decreases $p_{i+\tau}^{(C)}$ by δ.

Loss recovery for $S[0 : i-1]$ is the same because the changes occur after the deadline of $(i-1+\tau)$. By Equation 31, for any $j \in \mathbb{B}_i^{(s)}$ and $Y^{(C)}[j : j+b_j-1] \in \mathcal{Y}_j$ that $\mathcal{H}^{(e)}(S[i] | Y^{(C)}[j : i+\tau-1], S[0 : j-1]) \leq u_i^{(C)}$. After the change, the $p_{i+\tau}^{(C)} = u_i^{(C)}$ symbols of $X^{(C)}[i+\tau]$ are still available to be used to recover $S[i]$.

For $j = (i+1)$, the total amount of information available to recover $S[j]$ by time slot $(j+\tau)$ has perhaps increased but not decreased (because the same symbols are received). It is possible that $u_j^{(C)}$ has increased as a result; however, the at most δ symbols of $X^{(C)}[i+\tau]$ that would have been used to recover $S[j]$ are now available in $X^{(C)}[j+\tau]$ to recover $S[j]$. So $S[j]$ is still recovered within τ time slots, and $p_{j+\tau}^{(C)} \geq u_j^{(C)}$.

For a burst starting in $j \in \mathbb{B}_i^{(s)}$ (or more generally ending by $(i+\tau)$) the total number of symbols received over $X^{(C)}[i+\tau : i+\tau+1]$ is unchanged, so loss recovery of $S[i+1 : j+b_j-1]$ is not negatively impacted. For a burst starting during or after $(i+\tau+1)$, the symbols of $X[i+\tau]$ would not be used anyway because $S[0 : i+\tau]$ are available. For bursts that include $X[i+\tau : i+\tau+1]$—therefore, starting no sooner than $X[i+2]$—up to 2 extra symbols may be lost due to a rounding issue of $\lceil \ell_l n_{C,l} \rceil$ symbols being lost in $X[l]$ for $l \in \{i+2, i+\tau+1\}$. This can be mitigated by sending two extra symbols in $X[r+\tau]$ for $r \in \{i+2, \ldots, i+\tau\}$ that is used to recover $S[r]$ and increasing $u_r^{(C)}$ by two (up to a max of k_r). in other words, all information about up to two symbols of $S[r]$ are removed from the transmission, C pretends k_r was two symbols smaller, and the extra two symbols are sent via replication.

Iterating over $i_* = b_0, \ldots, (t-\tau)$, we add at most $2(\tau-1)$ symbols for each value of i_*. ■

Corollary 3: Consider any $i \in [t-\tau]$ and any $j \in \mathbb{B}_i^{(s)}$. Then

$$\min_{Y^{(C)}[j:j+b_j-1] \in \mathcal{Y}_j} \big(\mathcal{I}^{(e)}(S[j:i]; Y^{(C)}[j : j+\tau-1] | S[0 : j-1])\big) = \sum_{l=j}^{i} k_l - u_l^{(C)}.$$

Proof: Follows from Lemma 5, Lemma 7, Corollary 2, and Equation 31. ■

C. Proof of Theorem 2

Consider any rate optimal construction, C. To do so, we apply the relaxations of Appendix B to show. This means that without loss of generality, for all $i \in [t-\tau]$ then $u_i^{(C)} = p_{i+\tau}^{(C)}$ by Corollary 2. Thus, C sends at least

$$2(t-\tau)(\tau+1) + \sum_{l=0}^{t} p_i^{(LP)}$$

symbols compared to the algorithm (without the Constraint 7) by Corollary 2. To show this, we illustrate how the values of $u_i^{(C)}$ for $i \in [t-\tau]$ of C can be used as $p_{i+\tau}^{(LP)} = u_i^{(C)}$ in a solution that satisfies all constraints. Essentially, we show that Constraint 6 is analogous to the worst-case-delay constraint for a burst starting in time slot $i \in [t-\tau]$ where (a) for $j \in \{i, \ldots, i+b_i-1\}$, $u_j^{(C)}$ symbols of $S[j]$ are recovered during time slot $(j+\tau)$ and (b) for all $l \in \{i, \ldots, i+\tau-1\}$, $w_{i,j,l}$ reflects the number of useful symbols of $Y^{(C)}[l]$ for recovering $S[i : j]$. Ultimately, we prove that C sends at least

$$2(t-\tau)(\tau+1) + \sum_{l=0}^{t} p_i^{(LP)}$$

symbols compared to the algorithm without the Constraint 7.

We begin by noting that Constraint 1 holds by the proof of Corollary 2. Also, Constraint 2 holds by Equation 33 and Equation 4.

By Corollary 3, for any burst starting in time slot $i \in [t-\tau]$

$$\min_{Y^{(C)}[i:i+b_i-1] \in \mathcal{Y}_i} \big(\mathcal{I}^{(e)}\big(S[i : i+b_i-1]; Y^{(C)}[i : i+\tau-1] | S[0 : i-1]\big)\big) = \sum_{l=i}^{i+b_i-1} k_l - u_l^{(C)}. \tag{95}$$

For any $j \in \{i, \ldots, i+b_i-1\}$

$$\mathcal{I}^{(e)}\big(S[i : i+b_i-1]; Y^{(C)}[i : i+\tau-1] | S[0 : i-1]\big) = \tag{96}$$

$$\mathcal{I}^{(e)}\big(S[i : j]; Y^{(C)}[i : i+\tau-1] | S[0 : i-1]\big) + \mathcal{I}^{(e)}\big(S[j : i+b_i-1]; Y^{(C)}[i : i+\tau-1] | S[0 : j-1]\big) \leq \tag{97}$$

$$\mathcal{I}^{(e)}\big(S[i : j]; Y^{(C)}[i : i+\tau-1] | S[0 : i-1]\big) + \sum_{r=j}^{i+b_i-1} k_r - u_r^{(C)} \tag{98}$$

$$\sum_{r=i}^{j} k_r - u_r^{(C)} \leq \mathcal{I}^{(e)}\big(S[i : j]; Y^{(C)}[i : i+\tau-1] | S[0 : i-1]\big). \tag{99}$$

Equation 97 follows from the chain rule for MI; Equation 98 follows from Equation 31; Equation 99 follows from combining Equations 95 and 31.

Consequently,

$$\mathcal{I}^{(e)}\big(S[i:j];Y^{(C)}[i:i+\tau-1]|S[0:i-1]\big)\big) = \tag{100}$$
$$\mathcal{I}^{(e)}\big(S[i:j];Y^{(C)}[i:j]|S[0:i-1]\big)+$$
$$\mathcal{I}^{(e)}\big(S[i:j];Y^{(C)}[j+1:i+b_i-1-1]|$$
$$S[0:i-1],Y^{(C)}[i:j]\big)+ \tag{101}$$
$$\mathcal{I}^{(e)}\big(S[i:j];Y^{(C)}[i+b_i:i+\tau-1]|$$
$$S[0:i-1],Y^{(C)}[i:i+b_i-1]\big)\big) \leq$$
$$\sum_{l=i}^{j}\min\big(\ell_l(p_l^{(LP)}+k_l),k_l-u_l^{(C)}+\sum_{r=i}^{l-1}k_r-u_r^{(C)}-$$
$$\mathcal{I}^{(e)}(S[i:r];Y^{(C)}[r]\mid S[0:i-1],Y^{(C)}[i:r-1])\big)+$$
$$\sum_{l=j+1}^{i+b_i-1}\min\big(p_l^{(C)}(1-\ell_l),\sum_{r=i}^{j}k_r-u_r^{(C)}-\sum_{r=i}^{l-1}$$
$$\mathcal{I}^{(e)}(S[i:\min(r,j)];Y^{(C)}[r]\mid S[0:i-1],Y^{(C)}[i:r-1])\big)$$
$$+\sum_{l=i+b_i}^{i+\tau}\min(p_i^{(C)},\sum_{r=i}^{j}k_r-u_r^{(C)}-\sum_{r=i}^{l-1}$$
$$\mathcal{I}^{(e)}(S[i:j];Y^{(C)}[r]\mid S[0:i-1],Y^{(C)}[i:r-1])\big) \tag{102}$$

where Equation 101 follows form the chain rule for MI; Equations 102 follows from relaxations 1,2,3 and 4.

Therefore, for any $i\in[t-\tau], j\in\{i,\dots,i+b_i-1\}$ we note $w_{i,j,l}$ reflects $\mathcal{I}^{(e)}(S[i:j];Y^{(C)}[l]|S[0:i-1],Y^{(C)}[i:l-1])$ for any $l\in\{i,\dots,i+b_i-1+\tau\}$ in Constraint 6; so constraints 3, 4, and 5 are satisfied. By Equation 99, Constraint 6 is satisfied.

Therefore, the minimization of Algorithm 1, the values for $p_i^{(LP)}$ lead to

$$\sum_{i=0}^{t-\tau}p_{i+\tau}^{(C)} \geq \sum_{i=0}^{t-\tau}p_{i+\tau}^{(LP)}$$

Thus, the rate of C must be lower, as the rate of C is at most

$$\sum_{i=0}^{t-\tau}k_i\Big/\left(\sum_{i=0}^{t-\tau}k_i+p_{i+\tau}^{(C)}\right).$$

Now we show that Constraint 7 does not increase the value of the objective function. Let us impose the constraint for $i_*=0,\dots,(t-\tau)$. For any burst starting earlier than the minimum value of $\mathbb{B}_{i_*}^{(s)}$, the change is irrelevant. For $i=i_*$, decreasing $p_{i+\tau}^{(LP)}$ to equal $\ell_i k_i$ increases $w_{i,j,i}$ be the same amount that the RHS of Constraint 6 is increased so it is still satisfied. For $i<i_*$ recall that for $j=(i_*-1)$ Constraint 6 is satisfied with w_{i,j,i_*} providing at most $p_{i_*+\tau}^{(LP)}\ell_{i_*}$ symbols (the change has not yet come into effect). Thus, for $j\geq i_*$ where $j\in\{i+1,\dots,i+b_i-1\}$, decreasing $p_{i+\tau}^{(LP)}$ to equal $\ell_i k_i$ increases $w_{i,j,i}$ be the same amount that the RHS of Constraint 6 is increased so it is still satisfied. For a burst starting during or after time slot $(i+1)$ we can always increase a future $p_j^{(LP)}$ for $j>i_*$ by the amount that $p_{i_*}^{(LP)}$ was reduced to ensure that Constraint 6 is still satisfied. The objective function's value does not increase.

D. Proof of Theorem 3

We prove by induction on $\zeta=\sum_{l=0}^{t}\zeta_l$. For the base case, $\zeta=0$, and the output is simply that of applying Algorithm 1. So correctness follows from Theorem 2.

For the inductive step, we prove the result for ζ assuming that it is been proven for all $j\in\{0,\dots,\zeta-1\}$. Let r be the smallest value so that $\zeta_r=1$. If $r<\tau$ then only the lossless-delay constraint is imposed for $S[0:r-1]$ so $\sum_{l=0}^{r-1}k_l$ symbols are sent and no parity need to be sent. Otherwise, $S[r-\tau:r-1]$ need not be recovered under lossy conditions. By considering $k_{r-\tau},\dots,k_{r-1}$ to all be 0, the output of the algorithm through time slot $(r-1)$ can be viewed as an approximately rate-optimal code for the transmission of $S[0:r-1]$ where the final τ time slots are 0. Then an extra $\sum_{l=r-\tau}^{r-1}k_l$ symbols must also be sent under C for the lossless-delay constraint.

Either way, for time slot r and above, we can view Algorithm 1 as being applied a second time on the remainder of the transmission given $p_l^{(LP)}=0$ for $l\in\{r,\dots,r+\tau-1\}$ and the worst-case-delay is not imposed for $S[r-\min(r,\tau):r-1]$. Then $\sum_{l=r}^{t}\zeta_l=(\zeta-1)$, so the correctness of Algorithm 2 holds by the inductive hypothesis.

E. Proof of Corollary 1

By Constraints 3, 4, 5, 6 and Equations 7, 8 (and $p_l^{(LP)}=0$ for l at least τ time slots after the burst's start if $\sum_{r=l-\tau+1}^{l}\zeta_r>0$), Equation 9 is always satisfied. Thus, the construction's requirements are met.

For $j\in[t-\tau]$, if $\sum_{l=j+1}^{j+\tau}\zeta_l>0$ then $p_{j+\tau}=0$. Recall for any $u_j^{(LP)}$ that increasing $u_j^{(LP)}=p_{j+\tau}^{(LP)}$ to a quantity no more than $\ell_j k_j$ retains satisfaction of all constraints of the LP. So increasing $u_i^{(*)}$ from $p_{i+\tau}^{(LP)}$ for $i\in[t-\tau]$ (to ensure $q_i|u_i^{(*)}$) and likewise increasing $p_{i+\tau}$ (i.e., $p_{i+\tau}^{(LP)}$) is still a valid solution to the LP. Increasing $p_{i+\tau}$ (similarly $p_{i+\tau}^{(LP)}$) to be divisible by $h_{i+\tau}$ for $i\in[t-\tau]$ likewise retains the satisfaction of all constraints (with Constraint 7 removed). After these changes to the values of $p_j^{(LP)}$, $\big(\tau,t,K,\mathrm{Z},\mathcal{L},B,u^{(+,t-\tau)}\big)$-Split Code is fully specified and sends $\sum_{l=0}^{t-\tau}k_l+p_{l+\tau}^{(LP)}$. The total increase in $\sum_{l=0}^{t-\tau}p_{l+\tau}^{(LP)}$ due to the change is at most $\sum_{l=0}^{t-\tau}\mathbb{1}[p_{l+\tau}^{(LP)}\neq 0](q_l+h_{l+\tau}-2)$.

F. Proof of Theorem 4

At a high-level, the proof is divided into three steps. First, we bound how many extra symbols are modeled as being sent under Algorithm 2 in terms of $\mathcal{R}_0,\dots,\mathcal{R}_{t-\tau}$ by adding constraints for $p_\tau^{(LP)},\dots,p_t^{(LP)}$ to equal $O_0,\dots,O_{t-\tau}$, respectively (Appendix F1). We then bound the probability that $\sum_{l=0}^{t-\tau}\mathcal{R}_i$ exceeds its mean by a significant amount (Appendix F2). Finally, we establish the rate in terms of these quantities (Appendix F3).

1) Extra symbols sent under Algorithm 2: First, we show that the increase in $\sum_{l=0}^{t}p_l^{(LP)}$ due to adding the constraint $p_{i+\tau}^{(LP)}=O_i$ for $i\in[t-\tau]$ where $\sum_{l=i+1}^{i+\tau}\zeta_l>0$ is at most $\mathcal{R}_i$.

Let $u_i^{(Opt)} \in U_i^{(Opt)}$ be the value that minimizes $|u_i^{(Opt)} - O_i|$.

Suppose $O_i \geq u_i^{(Opt)}$. Then using the values of $p_j^{(LP)}$ for all $j > (i+\tau)$ still satisfy all constraints if $p_{i+\tau}^{(LP)}$ is set to equal O_i and $p_{i+\tau,opt}^{(LP)}$

Otherwise, suppose $O_i < u_i^{(Opt)}$. Let $\delta = \delta' = u_i^{(Opt)} - O_i$. Let us set $u_i^{(Opt)} = O_i$. While $\delta > 0$ let $j = \min_{l \in \{i,\ldots,i+b_i-1 | p_{l+\tau}^{(LP)} < \ell_l k_l, \sum_{r=l+1}^{l+\tau} \zeta_r = 0\}}(l)$. At least one such j exists, since otherwise $p_{i+\tau}^{(LP)}$ could be reduced, violating the minimizing the objective function. Increase $p_{j+\tau}^{(LP)}$ by $\min(\ell_j k_j - u_j^{(Opt)}, \delta)$ and decrement δ by the changed amount. The changes ensure that all constraints are satisfied. The total number of extra symbols sent is at most δ'.

2) Bounding the regret: Next, we show for any $\tau, t, K, \mathrm{Z}, \mathcal{L}, B, O_0, \ldots, O_{t-\tau}$,

$$\mathbb{P}[N_{\tau,t,K,\mathrm{Z},\mathcal{L},B} - N_{\tau,t,K,\mathrm{Z},\mathcal{L},B,\mathcal{R}_{[t]}} \geq (t+1)\epsilon] \leq e^{-2t^2\epsilon^2/(\sum_{l=0}^{t-\tau}(\ell_l k_l)^2)} \leq e^{-2t\epsilon^2/(m^2)}$$

First, suppose we start with $(\tau, t, K, \mathrm{Z}, \mathcal{L}, B)$-Split ML Code and then incrementally for $j = 0, \ldots, (t-\tau)$ switch to $(\tau, t, K, \mathrm{Z}, \mathcal{L}, B, u^{(+,O,j)})$-Split ML Code. With each switch, the total number of extra symbols sent is at most $\mathcal{R}_j$ (Appendix F1). In total, the number of extra symbols sent compared to $(\tau, t, K, \mathrm{Z}, \mathcal{L}, B)$-Split ML Code is $\sum_{l=0}^{t-\tau} \mathcal{R}_j$. The proof follows from the Hoeffding Bound [34].

3) Online approximately optimality: With probability at least $(1 - \delta' - \delta)$, $K^{(+)} \geq 1/2\mathrm{E}[K^{(+)}]$ and by Appendix F1 and F2 the rate obeys

$$R^{(opt)} - R^{(on)} \leq \tag{103}$$

$$K^{(+)}/(N_{\tau,t,K,\mathrm{Z},\mathcal{L},B} - 2(t-\tau)(\tau-1) - 2\sum_{l=0}^{t} h_i) - \tag{104}$$

$$K^{(+)}/(N_{\tau,t,K,\mathrm{Z},\mathcal{L},B,\mathcal{R}_{[t]}} + (t+1)\epsilon) \leq$$

$$\big(K^{(+)}((t+1)\epsilon + \sum_{l=0}^{t-\tau}(2h_i + 2(\tau-1)\mathrm{E}[\mathcal{R}_l])\big)/(K^{(+)}K^{(+)}) \tag{105}$$

$$\leq 2\epsilon + 4(h' + \tau - 1)(t + 1 - \tau)/\mathrm{E}[K^{(+)}] \tag{106}$$

Equation 105 follows from combining fractions in Equation 104 to have the denominator as the multiple of their two denominators, canceling terms, and noting each denominator being at least $K^{(+)}$; Equation 106 follows from $\mathrm{E}[\mathcal{R}_i] \leq \epsilon k_i, h_i \leq h'$ for all $i \in [t]$, $K^{(+)} \geq 1/2\mathrm{E}[K^{(+)}] > t$.