Mismatched Side Information in Wireless Network Compression via Overhearing Helpers

Mohsen Sardari, Ahmad Beirami, Faramarz Fekri
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA
Email:{mohsen.sardari, beirami, fekri}@ece.gatech.edu

***Abstract*—Recently, we proposed wireless network compression via memory-enabled overhearing helpers as an endeavor to reduce the traffic load on the wireless gateway via elimination of the redundant data in the network. In this setup, each memory-enabled helper overhears the data packets previously sent by the wireless gateway to various mobile clients within its coverage and uses them toward forming a model about the content of the packets from the traffic. The resulting model is then used as side information by the wireless network compression module in a two-part code to reduce the overall cost of delivering a packet to a client over links with asymmetric cost (where the helper-client link is far less costly than the gateway-client link). One main challenge in this scenario is the fact that memory-enabled overhearing helpers do not receive all of the sequences sent to the mobile clients (as there is no feedback in place in the overhearing link), resulting in mismatched side information between the encoder (i.e., gateway) and the helper. In this paper, we present an information theoretic formulation for the mismatched side information problem. We study this problem in the context of universal lossless compression and derive bounds on the average minimax redundancy of encoding each packet. Our results also lead to construction of coding schemes for the mismatched side information using two-part codes.**

***Index Terms*—Wireless Networks, Mismatched Side Information, Memory-Assisted Compression, Redundancy Elimination.**

I. Introduction

Mobile data efficiency is very important in wireless communication. Hence, mechanisms that reduce the amount of transmitted data can considerably impact wireless networks. Redundancy elimination at the packet-level is a promising approach for reducing wireless network traffic since experimental studies over real-world wireless traffic demonstrate 50% inter-client repetition among the segments of the packets [1], [2]. This repetition is due to the fact that packets that two mobile clients download from a server (source) can be highly correlated. It is important to note that existing solutions, such as application-layer caching and even deduplication within one data flow are ineffective to cope with the redundancy that exists mostly at the packet level and across several clients. This motivates the study of solutions that aim at learning the source statistics in the network and use that learning toward compression of new packets from all flows.

In [3]–[6], we took the first steps towards characterizing the achievable benefits of exploiting the packet redundancies beyond simple repetition suppression (i.e., deduplication). Data compression and source coding are natural candidates for this task. However, traditional compression techniques would be ineffective in the elimination of the redundancy within a small network packet [7]. Further, traditional compression methods cannot leverage the redundancy across clients; as they compress each packet independent of the other packets. In [3]–[6], we formulated the redundancy elimination as *network compression via network memory* and introduced a new framework for compression of network data called *memory-assisted compression* in wired networks.

In [8], [9], we proposed *wireless network packet compression via memory-enabled overhearing helpers*, where the gain of memorization and memory-assisted compression is more spelled out. The memory-enabled helpers are small, possibly cooperative nodes with sufficiently large storage space, which is used for the memorization of the overheard packets previously transmitted from the wireless gateway to mobile clients. The overhearing capability of helper nodes comes at no cost (in terms of bandwidth usage) and eliminates the need for backhaul connectivity while offering throughput enhancement by off-loading the gateway. In a nutshell, the overhearing helper forms the source model by learning from the overheard traffic. Then, in the compression of a new packet, a two-part coding strategy is employed, where the source model is sent from the overhearing (helper) node to a mobile client to supplement (as side information) the compressed data that is transmitted from the wireless gateway to the mobile client; enabling the client to decompress the codeword and recover the packet. Since the communication in the link between the helper and the client is by far less costly than that of the wireless gateway and the client, the proposed network compression via overhearing nodes, by design, reduces traffic on the latter link.

An important challenge that was not considered in [8], [9] is the impact of mismatch between the decoder side and the encoder side information. The error-prone wireless environment makes it difficult to guarantee that the sender node and the helper node (which has overheard the previous communication of the source with other mobile clients) share the same model of the information source. This is because the error recovery mechanism is implemented between the source and the client to which the packet is destined to, not the helper that only overhears the communication. This results in a mismatch between the source model at the encoder and the decoder, which in turn makes the memory-assisted compression challenging. In this paper, we view the overhearing in the network as an erasure channel, where a fraction $\mathcal{E}$ of the packets from the entire previously sent packets to other clients are erased due to overhearing at the helper. We analytically study the impact of the mismatched side information on the wireless network compression via memory-enabled overhearing nodes and provide theoretical results on the performance of memory-assisted compression in this scenario.

II. Background Review on Universal Compression

Let S be a multinomial (memoryless) source over alphabet $\mathcal{A}$, with a $(|\mathcal{A}|-1)$-dimensional parametric vector θ which takes values in $\Theta \subset \Re^d$, where $d = (|\mathcal{A}|-1)$ denotes the dimension of the unknown source parameter vector. One may extend this model to a more realistic setup for real-world sources by considering a mixture of parametric sources (with memory) as studied in [6], [10]. Please note that the side information (through memorized packets) primarily helps to remove the universal compression overhead, which is already significant for short memoryless sources.

Let $x^n = (x_1, \ldots, x_n)$ be a sequence generated by the source with probability $\mu_\theta(x^n)$. In the absence of side information, x^n is universally coded by $c : \mathcal{A}^n \to \{0,1\}^*$ with the length function denoted by $l(x^n)$ that satisfies Kraft's inequality. Let $H_n(\theta)$ be the entropy of the parametric source induced by μ_θ as given by

$$H_n(\theta) = \sum_{x^n} \mu_\theta(x^n) \log \frac{1}{\mu_\theta(x^n)}.^1 \quad (1)$$

The performance of the employed compression is measured in terms of the average code redundancy, which is given by $R(l,\theta) = \mathbf{E}[l(X^n)] - H_n(\theta)$. If the parameter vector $\theta \in \Theta$ was known, the ideal code length of a sequence x^n, obtained from the optimal code (ignoring the integer code length requirement), would be $-\log \mu_\theta(x^n)$. On the other hand, without the knowledge of θ, one has to encode the sequence with a penalty term that is characterized by the code redundancy. The average minimax redundancy, defined as

$$\bar{R}(n,\Theta) = \min_l \max_{\theta\in\Theta} R(l,\theta),$$

is a performance measure for universal lossless coding schemes. It is shown in [11], [12] that for a memoryless source with d unknown parameters, we have

$$\bar{R}(n,\Theta) = \frac{d}{2}\log\left(\frac{n}{2\pi e}\right) + \log\int_{\theta\in\Theta} |\mathcal{I}(\theta)|^{\frac{1}{2}}\, \mathrm{d}\theta + O\left(\frac{1}{n}\right), \quad (2)$$

where $|\mathcal{I}(\theta)|$ is the determinant of the Fisher information matrix evaluated at θ.

Denote by $\bar{R}(w,\theta)$ the expected redundancy of a universal compression scheme with the prior $w(\theta)$ on $\Theta_0 \subset \Theta$. A result related to (2) is the following [13], [14]:

$$\bar{R}(w,\theta) = \frac{d}{2}\log\left(\frac{n}{2\pi e}\right) - \log w(\theta) + \log|\mathcal{I}(\theta)|^{\frac{1}{2}} + o(1), \quad (3)$$

where the convergence is uniform in $\theta \in \Theta_0$. Accordingly, Jeffreys' prior, defined as

$$w_{\mathrm{J}}(\theta) = \frac{|\mathcal{I}(\theta)|^{\frac{1}{2}}}{\int_{\theta\in\Theta_0} |\mathcal{I}(\theta)|^{\frac{1}{2}}\, \mathrm{d}\theta}, \quad (4)$$

is maximin optimal. Note that Jeffreys' prior is also minimax optimal [15].

Finally, another important relationship that we use in this paper is the following result by Gallager [16] which shows that if μ_θ is a measurable function of θ, then

$$\bar{R}(n,\Theta) = \sup_{w(\theta)} I(X^n;\theta), \quad (5)$$

[1] In this paper, $\mathbf{E}$ denotes the expectation operation using the probability measure μ_θ. Further, $\log(\cdot)$'s are taken at base 2.

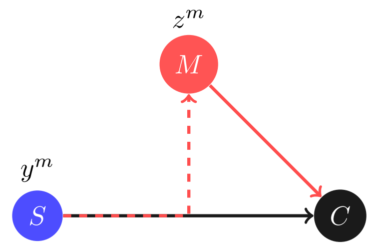

Fig. 1. Illustration of network compression via overhearing helper M. The memorized sequence y^m represents the total past data from S to the clients overheard by M. The overhearing channel is represented by dotted line.

where $I(X^n;\theta)$ is the mutual information between X^n and θ, and $w(\theta)$ is the prior distribution on θ.

III. Wireless Network Compression via Memory-Enabled Overhearing Helpers

The idea of wireless network compression via overhearing helpers is to deploy memory-enabled (non-mobile) helpers that are capable of overhearing communication from the wireless gateway to all the mobile clients inside the coverage area of the wireless gateway. The overhearing comes at no extra cost due to the broadcast nature of the wireless communication. Although this can be applied to every cellular or WiFi access network, one realization of such memory-enabled helpers can be in femto-cell network design combined with traditional macro-cell networks, as in [17]. Note that the backbone connectivity of the helpers are not included in the problem setup, first because the learning process of helper nodes is performed only based on the overheard data which is available for free and secondly, solutions that rely on helper connectivity should include provisions to deal with intermittent connectivity and the impact of the extra load imposed on the backbone.

Wireless network compression via memory-enabled overhearing helpers works as follows. Consider an example scenario involving a single wireless gateway S, a mobile client C and a helper M as in Fig. 1. In a real-world scenario, the node S is wireless gateway (or tower) that is connected to the Internet or the network backbone and transmits the packets to clients in unicast sessions. In our abstraction of the problem, node S may be viewed as a parametric source that sends independent sequences of length n to the clients in the cell. However, the source parameter is unknown to S and clients. Now, assume that several sequences (packets) have already been destined to some other clients via unicast from S, but due to the broadcast nature of the wireless environment, the helper M also overheard a subset of these sequences. Let y^m denote a sequence of length m, which is formed by the concatenation of all previously sent sequences to the other clients by S. Throughout this paper, we assume that $\frac{n}{m} = o(1)$,[2] i.e., the length of the memorized sequence is sufficiently large.

Next, S wishes to send a new sequence x^n to C. Recall that the traffic (i.e., the packets) destined to different mobile clients from the gateway S are highly correlated. Therefore, the memory-enabled overhearing helper M can learn the source model by estimating the unknown source parameter by using the overheard packets from the past communication between the cell tower (or the WiFi access-point) and mobile nodes.

[2] $f(n) = o(g(n))$ if and only if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

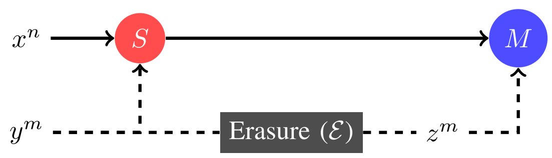

Fig. 2. The model abstraction of compression problem with mismatched side information.

This extracted source model can then be used as a side information (if provided to the client) improving the compression performance on the future traffic from the gateway S to any new mobile client C. In other words, the memory-enabled helpers can possibly help to reduce the transmission load of the wireless gateway by transmitting the side-information about the data traffic to the clients using a *less costly* memory-client M-C link.

In absence of erasure caused by overhearing, S and M would share a common side information y^m. Let z^m be the resulting sequence from overhearing y^m by the overhearing helper node M in Fig. 2. Due to channel erasures, a subset of the symbols in z^m are marked as erased. In this paper, we assume that both encoder (at S) and decoder (at M) know the length of side information, i.e., m, and the number of erased symbols r, hence, $\mathcal{E} = \frac{r}{m}$ is also known to both the encoder and the decoder. Please note that S just knows the number of erased symbols and does not know which symbols in z^m are marked as erasures. The purpose of the present paper is to analytically derive the impact of the mismatched side information y^m (available at S) and z^m (available at M) in the compression performance.

IV. Two-Part Code Design

In this section, we present the code design based on the adaptation of the two-part coding (c.f. [7], [13] and the references therein). The benefit of using two-part coding strategy for wireless network compression problem is two-fold. First, the compressed codeword describing x^n is consisted of two parts that can be separately sent to the end-user, i.e., the client node C; one part from the source S and the other from the memory-enabled helper M in Fig. 1. Secondly, the memorized sequence y^m is assumed to be longer than x^n and can be used to obtain a more accurate estimate of the source parameter for more efficient compression. First, let us consider the matched memory case (i.e., $\mathcal{E} = 0$). We defer the impact of the mismatched side information to Section V. Hence, the first part of the code describes the best estimation $\hat{\theta}(y^m) \in \Theta$ of the unknown source parameter vector θ by using the statistics of the source extracted from y^m. This part is extracted by M and is sent by M to C. The estimate $\hat{\theta}(y^m)$ is also extracted by S, which is then used in the second part of the code by S to encode the packet x^n and send it to C.

Note that the main objective of the wireless network compression is to minimize the total communication cost and hence support more clients. As such, we can define virtual costs for S-C and M-C links in Fig. 1. Let κ denote the ratio of the cost of communicating one bit in the M-C link to that of the S-C link. In practical settings, it is rational to assume that the S-C link is much more costly than the M-C link. This is because S serves several femto-cells but a helper node only serves the clients within a single femto-cell. Hence, κ is much smaller than unity. Let $\mathcal{L}_S(0)$ be the expected number of bits sent by S and $\mathcal{L}_M(0)$ be the expected number of bits sent by M to the client C in Fig. 1 when $\mathcal{E} = 0$. The following theorem determines the communication cost in each link in the case of network compression via overhearing helper [8].

Theorem 1 *Given $\mathcal{E} = 0$ and a memory of size m, we have*

$$\begin{cases} \mathcal{L}_S(0) &= H_n(\theta) + \frac{d}{2}\log\left(1+\frac{n}{m}\right) + o(1) \\ \mathcal{L}_M(0) &= \frac{d}{2}\log m + O(1) \end{cases}.$$

Proof: Please see [8] for the proof. ■

Remark: According to Theorem 1, when m is sufficiently large and $\mathcal{E} = 0$, the number of bits sent by the wireless gateway ($\mathcal{L}_S(0)$) is close to the entropy of the sequence which is the information-theoretic lower bound on the average number of bits sent by S. Further, the aggregate expected communication cost for both links in Fig. 1 is $\mathbf{E}[\mathcal{C}_0(X^n)] = \mathcal{L}_S(0) + \kappa\mathcal{L}_M(0)$. Hence, we observe that when $\kappa \ll 1$ for sufficiently large m (asymmetric communication cost), larger memorized sequences can be employed while the total communication cost is still dominated by the bits sent from the gateway to the client, which is close to the entropy.

V. Impact of Mismatched Side Information

Thus far, we demonstrated that significant performance improvement is expected from memory-enabled overhearing helpers by adapting two-part codes to the problem setup when $\mathcal{E} = 0$ (no mismatch). Please note that $\mathcal{L}_S(0)$ in Theorem 1 is indeed the entropy plus the average minimax redundancy given a random side information sequence y^m of length m is available to both the encoder and the decoder. In this section, we consider the impact of the mismatched side information on the achievable benefits.

Let $\bar{R}_{\mathcal{E}}(n, m, \Theta)$ be the average minimax redundancy of the compression of a random sequence x^n of length m using a random side information sequence z^m of length m in Fig. 2, where $\mathcal{E} = \frac{r}{m}$ is the fraction of erased packets in z^m compared with y^m (which is also the fraction of erased symbols). Please note that we have assumed that the source is memoryless and hence only the number of erased symbols becomes relevant in the analysis of the erasure process. It is straightforward to extend Theorem 1 to verify that

$$\begin{cases} \mathcal{L}_S(\mathcal{E}) &= H_n(\theta) + \bar{R}_{\mathcal{E}}(n,m,\Theta) + o(1) \\ \mathcal{L}_M(\mathcal{E}) &= \frac{d}{2}\log m + O(1) \end{cases}. \tag{6}$$

Thus, we only need to characterize $\bar{R}_{\mathcal{E}}(n, m, \Theta)$ to analyze the cost of communication. Next, we obtain a lower limit on this average minimax redundancy by analyzing the case that the encoder knows the location of erased symbols.

Proposition 2 *Given that $\mathcal{E} < 1 - \delta$ for some $\delta > 0$, we have*

$$\bar{R}_{\mathcal{E}}(n,m,\Theta) \geq \frac{d}{2}\log\left(1 + \frac{n}{m(1-\mathcal{E})}\right) + O\left(\frac{1}{n}\right). \tag{7}$$

Proof: Let $\bar{R}(n, t, \Theta)$ be the average minimax redundancy of a universal scheme compressing a sequence of length n given a side information v^t sequence, such that $\frac{n}{t} = o(1)$ known to both encoder and decoder. According to [18] and using (5), the average minimax redundancy of a memory-assisted compression scheme with a side information of size t can be obtained as

$$\begin{aligned}\bar{R}(n,t,\Theta) &= \max_{p(\theta)} I(X^n;\theta|V^t) \\ &= \max_{p(\theta)} \left[I(X^n,V^t;\theta) - I(V^t;\theta)\right] \\ &= \bar{R}(n+t,\Theta) - \bar{R}(t,\Theta). \end{aligned} \tag{8}$$

From (2), we have

$$\bar{R}(n,t,\Theta) = \frac{d}{2}\log\left(1+\frac{n}{t}\right) + O\left(\frac{1}{n}\right).$$

The proof is completed by using (8) and the fact that the destination has only access to a memory of size $t = m-r$ and no strictly lossless compression scheme can benefit from a side information longer than the one available at the destination. ■

Please note that the bound in Proposition 2 is achieved if M can encode and transmit the location of the erased packets to S (through uplink). On the other hand, if we assume no uplink communication between M and S then this bound is clearly not achievable.

Next, we state a trivial upper limit, which is tight when the erasure $\mathcal{E} \to 1$, i.e., there is no memory shared between S and M. This bound is obtained by ignoring the available memory at the encoder and the decoder. Then, $\bar{R}_{\mathcal{E}}(n,m,\Theta)$ is bounded from above by

$$\bar{R}_{\mathcal{E}}(n,m,\Theta) \leq \bar{R}(n,\Theta). \tag{9}$$

In the following, we provide a constructive approach which leads to a non-trivial upper bound on the average minimax redundancy for $m \to \infty$. The bound is given for binary case for the simplicity of the presentation. Extension to discrete non-binary alphabet requires a more careful investigation of the simplex of the source parameter vectors.

Theorem 3 *For sufficiently large side information size m, and $\delta < \mathcal{E} < 1-\delta$ for some $\delta > 0$, we have*

$$\bar{R}_{\mathcal{E}}(n,\infty,\Theta) \leq \sum_{i=1}^{\min(\lceil \frac{1}{\mathcal{E}} \rceil, 1)} \frac{2}{\pi}\left(\sin^{-1}\sqrt{i\mathcal{E}} - \sin^{-1}\sqrt{(i-1)\mathcal{E}}\right) \left(\frac{1}{2}\log\frac{2n}{\pi e} + \log \mathcal{C}_{(i-1)\mathcal{E}}^{i\mathcal{E}} + o(1)\right), \tag{10}$$

where

$$\mathcal{C}_{\alpha_1}^{\alpha_2} = \int_{\alpha_1}^{\alpha_2} \frac{1}{\sqrt{x(1-x)}}\, dx.$$

Sketch of the proof: Consider a memoryless source with parameter space $\Theta = (0,1)$ and alphabet $\mathcal{A} = \{a,b\}$. The Jeffreys' prior for this source, defined in (4), is $w(\theta) = \frac{1}{\pi\sqrt{x(1-x)}}$. If we use this prior for coding a sequence x^n, the resulting redundancy would be

$$\bar{R}(n,\Theta) = \frac{1}{2}\log(\frac{n}{2\pi e}) + \log(\pi) + o(1).$$

However, the side information will induce another prior on the parameter space that reduces the redundancy of the Jeffreys' prior. Consider a sequence y^m at S with $m_S^{(a)}$ number of a's. Likewise, let $m_D^{(a)}$ be the number of a's in z^m. Let $\hat{\theta}_S$ denote the ML estimate of θ at S and $\hat{\theta}_D$ be the ML estimate at M. We have

$$\begin{aligned}\hat{\theta}_S &= \frac{m_S^{(a)}}{m} \\ \frac{m_S^{(a)}-r}{m} \leq \hat{\theta}_D &= \frac{m_D^{(a)}}{m} \leq \frac{m_S^{(a)}}{m}. \end{aligned} \tag{11}$$

A strictly lossless compression scheme requires both the encoder and the decoder use the same parameter estimate or prior. To overcome the mismatch in (11) between $\hat{\theta}_S$ and $\hat{\theta}_D$, we consider the following scheme: both the encoder and the decoder divide the interval $(0,1)$ into sub-intervals of size $\frac{r}{m}$. Since $\hat{\theta}_D \leq \hat{\theta}_S$ and $|\hat{\theta}_D - \hat{\theta}_S| < \mathcal{E}$, the estimated parameter at the encoder and the decoder are either in the same sub-interval or in two adjacent sub-intervals. This discrepancy can be resolved with one extra bit sent by the encoder.

Let $\Theta_i = ((i-1)\mathcal{E}, i\mathcal{E})$ be the i-th sub-interval. Since, $w_J(\theta)$ is Jeffreys' prior,

$$\begin{aligned}\mathbf{P}[\theta \in \Theta_i] &= \int_{\theta\in\Theta_i} w_J(\theta)\, \mathrm{d}\theta \\ &= \frac{2}{\pi}\left(\sin^{-1}\sqrt{i\mathcal{E}} - \sin^{-1}\sqrt{(i-1)\mathcal{E}}\right). \end{aligned}$$

Further, for binary memoryless sources, $\mathcal{I}^{-1}(\theta) = \theta(1-\theta)$. Hence, according to (3), the redundancy of a compression scheme, with the side information that the source parameter is chosen from Θ_i, can be obtained as

$$\begin{aligned}\bar{R}(n,\Theta_i) &= \frac{1}{2}\log\left(\frac{n}{2\pi e}\right) + \log\int_{\theta\in\Theta_i} |\mathcal{I}(\theta)|^{\frac{1}{2}} + o(1) \\ &= \frac{1}{2}\log\left(\frac{n}{2\pi e}\right) + \log \mathcal{C}_{(i-1)\mathcal{E}}^{i\mathcal{E}} + o(1), \end{aligned}$$

which completes the proof of the theorem. ■

It is straightforward to verify that the bound provided by Theorem 3 reduces to the trivial bound stated (2) when $\mathcal{E} \to 1$. This is because $\mathcal{C}_0^1$ reaches its maximum of π and the total number of sub-intervals within $(0,1)$ is one and hence no extra information is needed from the source.

For the special case of $\mathcal{E} = O(\frac{1}{\sqrt{n}})$, we can provide a stronger result that exactly characterizes $\bar{R}_{\mathcal{E}}(n,m,\Theta)$.

Proposition 4 *If $\mathcal{E} = o(\frac{1}{\sqrt{n}})$, as $m \to \infty$, we have*

$$\bar{R}_{\mathcal{E}}(n,\infty,\Theta) = o(1). \tag{12}$$

Sketch of proof: Since $\mathcal{E} = O(\frac{1}{\sqrt{n}})$, the size of sub-interval Θ_i is also $o(\frac{1}{\sqrt{n}})$. Let $\theta^\star \in \Theta_i$, then,

$$\begin{aligned}\bar{R}(n,\Theta_i) &= \mathbf{E}[\log\mu_\theta(X^n) - \log\mu_{\theta^\star}(X^n)] \\ &= nD(\mu_\theta||\mu_{\theta^\star}) \\ &\overset{(i)}{=} \frac{n}{2}(\theta-\theta^\star)^2\mathcal{I}(\theta) + o(1) \\ &\overset{(ii)}{=} o(1), \end{aligned} \tag{13}$$

where $D(.||.)$ is the KL divergence. In (13), equality (i) follows from the second order approximation of KL divergence and (ii) follows from the fact that $(\theta-\theta^\star)^2 < \frac{1}{n}$. ■

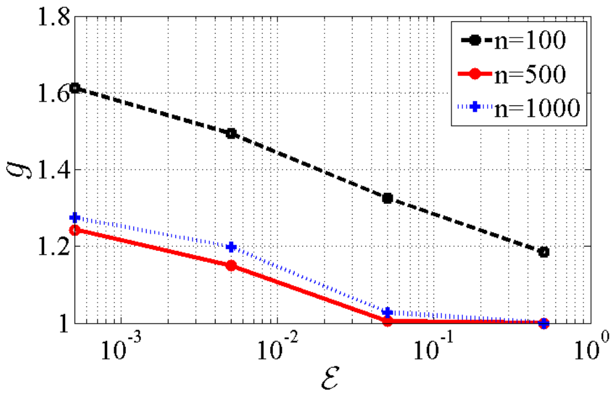

Fig. 3. Gain of memory-assisted compression over the end-to-end compression. The memory size is $m = 10^6$.

VI. Code Design for Mismatched Side Information

Construction of a memory-assisted compression scheme with mismatched side information follows from the proof of Theorem 3. As the proof suggests, we should first construct a code that would compress a sequence with the side information that the parameter is from a sub-interval Θ_i. Let $x_1^t = x_1 x_2 \ldots x_t$ be a sequence with binary symbols. Clearly,

$$\mathbf{P}\left[x_1^t|\theta\right] = \theta^{n_a(t)}\theta^{t-n_a(t)},$$

where $n_a(t)$ is the number of symbols a in x_1^t. The prior probability on the sub interval Θ_i is the normalized Jefferys' distribution, i.e.,

$$w_J(\theta) = \frac{|\mathcal{I}(\theta)|^{\frac{1}{2}}}{\mathcal{C}_{(i-1)\mathcal{E}}^{i\mathcal{E}}}.$$

Therefore, the probability of the sequence x_1^t is equal to

$$\mathbf{P}\left[x_1^t\right] = \int_{(i-1)\mathcal{E}}^{i\mathcal{E}} \frac{1}{\mathcal{C}_{(i-1)\mathcal{E}}^{i\mathcal{E}}\sqrt{\theta(1-\theta)}}\theta^{n_a(t)}\theta^{t-n_a(t)}\, \mathrm{d}\theta. \quad (14)$$

Now, the sequence x^n along with its probability can be passed to an arithmetic encoder. However, a better compression scheme, from practical point of view, can be used that evaluates the probability in (14) sequentially. As such, the sequential probability estimates of x_1^{t+1} can be evaluated as follows [19]:

$$\begin{aligned}\mathbf{P}\left[x_1^{t+1}\right] &= \mathbf{P}\left[x_1^t\right]\frac{n_{x_{t+1}}(t)+\frac{1}{2}}{t+1}\\ &+\beta\times\frac{{\alpha_1}^{n_a(t)+\frac{1}{2}}(1-\alpha_1)^{n_b(t)+\frac{1}{2}}}{\mathcal{C}_{\alpha_1}^{\alpha_2}(1+t)}\\ &-\beta\times\frac{{\alpha_2}^{n_a(t)+\frac{1}{2}}(1-\alpha_2)^{n_b(t)+\frac{1}{2}}}{\mathcal{C}_{\alpha_1}^{\alpha_2}(1+t)},\end{aligned}$$

where $\alpha_1 = (i-1)\mathcal{E}$, $\alpha_2 = i\mathcal{E}$, and

$$\beta = \begin{cases} 1 & x_{t+1} = a, \\ -1 & x_{t+1} = b. \end{cases}$$

The results in Fig. 3 show the performance of the proposed sequential compression scheme for different sequence lengths, fraction of erased symbols, and side information of length $m = 10^6$. The quantity g in Fig. 3 is defined as the gain of memory-assisted compression (with side information) over compression with no side information, i.e., $g \triangleq \frac{\mathbf{E}l(X^n)}{\mathbf{E}l(X^n|Z^m)}$. We observe that for simple binary sequences of length 100, a compression gain of 1.6 on top of the end-to-end compression of x^n is achieved, on the average. This gain is expected to be higher if we consider non-binary sources, as observed in [4].

VII. Conclusion

In this paper, we studied the impact of mismatched side information in wireless network compression via overhearing helpers. The mismatch occurs due to the erasures in the wireless overhearing link. We modeled mismatch by an erasure channel that erases a subset of the symbols from the side information in the decoder side. We provided theoretical results on the impact of erasure and memorization on the performance of universal compression and also verified the achievable gains through numerical simulations.

References

[1] S.-H. Shen, A. Gember, A. Anand, and A. Akella, “Refactor-ing content overhearing to improve wireless performance,” in *Mobicom*, Las Vegas, Nevada, US, 2011.
[2] S. Sanadhya, R. Sivakumar, K.-H. Kim, P. Congdon, S. Lakshmanan, and J. P. Singh, “Asymmetric caching: improved network deduplication for mobile devices,” in *Proc. of Mobicom '12*, 2012, pp. 161–172.
[3] M. Sardari, A. Beirami, and F. Fekri, “On the network-wide gain of memory-assisted source coding,” in *2011 IEEE Information Theory Workshop (ITW)*, October 2011, pp. 476–480.
[4] A. Beirami, M. Sardari, and F. Fekri, “Results on the fundamental gain of memory-assisted universal source coding,” in *2012 IEEE International Symposium on Information Theory (ISIT '2012)*, 2012, pp. 1092–1096.
[5] M. Sardari, A. Beirami, and F. Fekri, “Memory-assisted universal compression of network flows,” in *2012 Proceedings IEEE, INFOCOM*, Mar. 2012, pp. 91 –99.
[6] M. Sardari, A. Beirami, J. Zou, and F. Fekri, “Content-aware network data compression using joint memorization and clustering,” in *2013 IEEE Conference on Computer Networks (INFOCOM 2013)*, April 2013.
[7] A. Beirami and F. Fekri, “Results on the redundancy of universal compression for finite-length sequences,” in *IEEE Intl. Symp. Info. Theory (ISIT)*, Jul 31-Aug 5 2011, pp. 1504–1508.
[8] M. Sardari, A. Beirami, and F. Fekri, “Wireless network compression via memory-enabled overhearing helpers,” submitted to *IEEE Transactions on Wireless Communications*.
[9] ——, “Wireless network compression: Code design and trade offs,” in *Information Theory and Applications Workshop (ITA)*, February 2013.
[10] A. Beirami, M. Sardari, and F. Fekri, “Results on the optimal memory-assisted universal compression performance for mixture sources,” in *51st Annual Allerton Conference*, October 2013.
[11] A. Barron, J. Rissanen, and B. Yu, “The minimum description length principle in coding and modeling,” *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2743 –2760, oct. 1998.
[12] J. Rissanen, “Fisher information and stochastic complexity,” *IEEE Trans. Info. Theory*, vol. 42, no. 1, pp. 40 –47, Jan 1996.
[13] P. D. Grunwald, *The minimum description length principle*. The MIT Press, 2007.
[14] B. Clarke and A. Barron, “Information-theoretic asymptotics of bayes methods,” *IEEE Trans. Info. Theory*, vol. 36, no. 3, pp. 453 –471, 1990.
[15] N. Merhav and M. Feder, “A strong version of the redundancy-capacity theorem of universal coding,” *IEEE Trans. Info. Theory*, vol. 41, no. 3, pp. 714 –722, May 1995.
[16] R. G. Gallager, “Source coding with side information and universal coding,” *unpublished*.
[17] V. Chandrasekhar, J. Andrews, and A. Gatherer, “Femtocell networks: a survey,” *IEEE Comm. Magazine*, vol. 46, no. 9, pp. 59–67, 2008.
[18] A. Beirami and F. Fekri, “On lossless universal compression of distributed identical sources,” in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, july 2012, pp. 561 –565.
[19] G. I. Shamir, T. J. Tjalkens, and F. M. J. Willems, “Low-complexity sequential probability estimation and universal compression for binary sequences with constrained distributions,” in *IEEE Intl. Symp. Info. Theory (ISIT)*, Toronto, Canada, 2008, pp. 995–999.