Variable-Length Resolvability for Mixed Sources and its Application to Variable-Length Source Coding

Hideki Yagi
Dept. of Computer & Network Engineering
University of Electro-Communications
Tokyo, Japan
Email: h.yagi@uec.ac.jp

Te Sun Han
National Institute of Information and Communications Technology (NICT)
Tokyo, Japan
Email: han@is.uec.ac.jp

Abstract—In the problem of variable-length δ-channel resolvability, the channel output is approximated by encoding a variable-length uniform random number under the constraint that the variational distance between the target and approximated distributions should be within a given constant δ asymptotically. In this paper, we assume that the given channel input is a mixed source whose components may be general sources. To analyze the minimum achievable length rate of the uniform random number, called the δ-resolvability, we introduce a variant problem of the variable-length δ-channel resolvability. A general formula for the δ-resolvability in this variant problem is established for a general channel. When the channel is an identity mapping, it is shown that the δ-resolvability in the original and variant problems coincide. This relation leads to a direct derivation of a single-letter formula for the δ-resolvability when the given source is a mixed memoryless source. We extend the result to the second-order case. As a byproduct, we obtain the first-order and second-order formulas for fixed-to-variable length source coding allowing error probability up to δ.

I. INTRODUCTION

In the problem of variable-length δ-channel resolvability, the channel output is approximated by encoding a variable-length uniform random number under the constraint that the distance (e.g. variational distance) between the target and approximated distribution should be within a given constant δ asymptotically. This problem, introduced by Yagi and Han [1], is a generalized form of the fixed-length δ-channel resolvability [3, 4] in which the fixed-length uniform random number is used as a coin distribution. The minimum achievable length rate of the uniform random number, referred to as the δ-resolvability, is the subject of analysis. In [1], a general formula for the δ-resolvability has been established for any given source and channel. Recently, a single-letter formula for the δ-resolvability has been given in [2] when the source and the channel are stationary and memoryless. An interesting next step may be a mixed memoryless channels and/or a mixed memoryless channel [2], which are stationary but non-ergodic stochastic processes.

In this paper, we assume that the given channel input is a mixed source with components which may be general sources. To establish a general formula of the δ-resolvability for a general channel, we introduce a variant problem of the variable-length δ-channel resolvability. When the channel is an identity mapping, it is shown that the δ-resolvability in the original and variant problems coincide. This relationship is of use to derive a single-letter formula for the δ-resolvability when the given source is a mixed memoryless source. We also extend the result to the second-order case. It is known that the δ-resolvability coincides with the minimum achievable coding rate of the weak fixed-to-variable length (FV) source coding allowing error probability up to δ. As a byproduct, we obtain the first-order and second-order formulas for this minimum achievable coding rate.

II. PROBLEM OF VARIABLE-LENGTH CHANNEL RESOLVABILITY

In this section, we review the problem of channel resolvability in the variable-length setting.

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite or countably infinite alphabets. Let $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ be a general channel, where $W^n: \mathcal{X}^n \to \mathcal{Y}^n$ denotes a stochastic mapping. We denote by $\mathbf{Y} = \{Y^n\}_{n=1}^{\infty}$ the output process via $\mathbf{W}$ due to the input process $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, where $X^n$ and $Y^n$ take values in $\mathcal{X}$ and $\mathcal{Y}$ respectively. The probability distributions of $X^n$ and $Y^n$ are denoted by $P_{X^n}$ and $P_{Y^n}$, respectively, and these symbols are used interchangeably.

Consider the problem of variable-length channel resolvability. Let $\mathcal{U}$ denote the set of all sequences $u \in \mathcal{U}^m$ over $m = 0, 1, 2, \ldots$, where $\mathcal{U}^0 = \{\lambda\}$ ($\lambda$ is the null string). Let $L_n$ denote a random variable which takes values in $\{0, 1, 2, \ldots\}$. We define the variable-length uniform random number $U^{(L_n)}$ so that $U^{(m)}$ is uniformly distributed over $\mathcal{U}^m$ given $L_n = m$. In other words, for $u \in \mathcal{U}^m$,

$$P_{U^{(L_n)}}(u, m) := Pr\{U^{(L_n)} = u, L_n = m\} = \frac{Pr\{L_n = m\}}{K^m},$$

(1)
where $K = |\mathcal{U}|$. It should be noticed that variable-length sequences $u \in \mathcal{U}^n$ are generated with joint probability $P_{U^n}(u, m)$. Consider the problem of approximating the target output distribution $P_{V^n}$ via $W^n$ under $\mathcal{U}^n$ by using another input $X^n = \varphi_n(U^{(L_n)})$ with a deterministic mapping (encoder) $\varphi_n : \mathcal{U}^* \to \mathcal{X}^n$. Let $d(P_{Y^n}, P_{Y^n}) := \frac{1}{n} \sum_{y} |P_{Y^n}(y) - P_{Y^n}(y)|$ be the variational distance between $P_{Y^n}$ and $P_{Y^n}$.

**Definition 1:** Let $\delta \in [0, 1)$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be $\delta$-variable-length achievable or simply $\nu(\delta)$-achievable for $X$ (under the variational distance) if there exists a variable-length uniform random number $U^{(L_n)}$ and a deterministic mapping $\varphi_n : \mathcal{U}^* \to \mathcal{X}^n$ satisfying

$$\limsup_{n \to \infty} \frac{1}{n} \mathbb{E}[L_n] \leq R,$$

$$\limsup_{n \to \infty} d(P_{Y^n}, P_{Y^n}) \leq \delta,$$

where $\mathbb{E}[\cdot]$ denotes the expected value and $\hat{Y}^n$ denotes the output via $W^n$ due to the input $X^n = \varphi_n(U^{(L_n)})$. The infimum of all $\nu(\delta)$-achievable rates for $X$:

$$S_\nu(\delta | X, W) := \inf \{ R : R \text{ is } \nu(\delta) \text{-achievable for } X \}$$

is called the $\delta$-variable-length channel resolvability or simply $\nu(\delta)$-channel resolvability for $X$.

When the channel $W^n$ is an identity mapping, the addressed problem reduces to that of source resolvability.

**Definition 2:** Assume that the channel $W^n$ is an identity mapping. The infimum of all $\nu(\delta)$-achievable rates for $X$:

$$S_\nu(\delta | X) := \inf \{ R : R \text{ is } \nu(\delta) \text{-achievable for } X \}$$

is called the $\delta$-variable-length source resolvability or simply $\nu(\delta)$-source resolvability for $X$.

Let $\mathcal{P}(X^n)$ denote the set of all probability distributions on $\mathcal{X}^n$. For $\delta \in [0, 1]$, defining the $\delta$-ball using the variational distance as

$$B_\delta(X^n) = \{ P_{V^n} \in \mathcal{P}(X^n) : d(P_{V^n}, P_{V^n}) \leq \delta \},$$

we introduce the smooth entropy:

$$H_\delta(X^n) := \inf_{P_{V^n} \in B_\delta(X^n)} H(V^n),$$

where $H(V^n)$ denotes the Shannon entropy of $P_{V^n}$. The $H_\delta(X^n)$ is a nonincreasing monotone function of $\delta$. Based on this quantity for a general source $X = \{ X^n \}_{n=1}^\infty$, we define

$$H_\delta(X) = \limsup_{n \to \infty} \frac{1}{n} H_\delta(X^n).$$

The following theorem indicates that the $\nu(\delta)$-resolvability $S_\nu(\delta | X)$ can be characterized by the smooth entropy for $X$.

**Theorem 1 (172):** For any general target source $X$,

$$S_\nu(\delta | X) = \lim_{\gamma \downarrow 0} H_{[\delta + \gamma]}(X) \quad (\delta \in [0, 1)).$$

**III. Resolvability for Mixed Sources and Non-Mixed Channels**

**A. Definitions**

In this section, the source $X = \{ X^n \}_{n=1}^\infty$ is a mixed source with general component sources. Let $\Theta := \{ 1, 2, \cdots \}$ be the index set of component sources $X_i = \{ X_i^n \}_{n=1}^\infty$, $i \in \Theta$, which may be a finite or countably infinite set. The probability distribution of mixed source $X^n$ is given by

$$P_{X^n}(x) = \sum_{i \in \Theta} \alpha_i P_{X^n_i}(x) \quad (\forall n = 1, 2, \cdots ; \forall x \in \mathcal{X}^n),$$

where $\alpha_i \geq 0$ with $\sum_{i \in \Theta} \alpha_i = 1$. Let $Y = \{ Y^n \}_{n=1}^\infty$ be the channel output via $W$ due to input $X$. It is easily verified that the output distribution is given as a mixture of output distributions:

$$P_{Y^n}(y) = \sum_{i \in \Theta} \alpha_i P_{Y^n_i}(y) \quad (\forall y \in \mathcal{Y}^n),$$

where $Y^n$ denotes the output via $W^n$ due to input $X^n$. The mixed source is formally denoted by $\{ (X_i, \alpha_i) \}_{i \in \Theta}$. Hereafter, the mixing ratio $\{ \alpha_i \}_{i \in \Theta}$ is omitted if it is clear from the context, and we occasionally denote the mixed source simply by $\{ X_i \}$.

In this section, we consider a variant of the channel resolvability problems for mixed sources. Let $L_n^{(i)}$ denote a variable-length uniform random number for $i \in \Theta$. Let the random variable of length $L_n$ be specified by

$$\Pr\{ L_n = m \} = \sum_{i \in \Theta} \alpha_i \Pr\{ L_n^{(i)} = m \} \quad (\forall m = 0, 1, 2, \cdots).$$

In other words, the length of a variable-length uniform random number $L_n^{(i)}$ obeys a mixture of the probability distributions for the lengths of component uniform random numbers $U^{(L_n^{(i)})}$. The average length of the uniform random number $U^{(L_n^{(i)})}$ is given by

$$\mathbb{E}[L_n] = \sum_{i \in \Theta} \alpha_i \mathbb{E}[L_n^{(i)}].$$

In the following problem, there are component encoders $\varphi_n^{(i)} : \mathcal{U}^* \to \mathcal{X}^n$, each of which approximates the channel output $Y^n_i$ via $W^n$ due to the $i$-th component source $X^n_i$.

**Definition 3:** Let $\delta \in [0, 1)$ be fixed arbitrarily. A resolution rate $R \geq 0$ is said to be $\delta$-variable-length achievable or simply $\nu(\delta)$-achievable for mixed source $\{ (X_i, \alpha_i) \}_{i \in \Theta}$ (under the variational distance) if there exists a set of variable-length uniform random numbers $U^{(L_n^{(i)})}$ and a deterministic mapping

\[1\]More generally, all results provided in this section hold for any mixed source with a general mixture. Any stationary process can be characterized as a mixed source with general mixture whose components are ergodic processes.
\( \varphi_n^{(i)} : U^* \to X^n \) satisfying
\[
\lim_{n \to \infty} \frac{1}{n} E[L_n] \leq R, 
\]
\[
\lim_{n \to \infty} \sum_{i \in \Theta} \alpha_i d(P_{Y^n}, P_{Y^n}) \leq \delta, 
\]
where \( Y^n \) denotes the output via \( W^n \) due to the input \( X^n = \varphi_n^{(i)}(U(L_{\alpha_i})) \). The infimum of all \( v(\delta) \)-achievable rates is
\[
S^*_v(\delta|\{X_i\}, W) := \inf \{R : R \text{ is } v(\delta) \text{-achievable for } \{X_i\}\} 
\]
is called the \( \delta \)-variable-length channel resolvability or simply \( v(\delta) \)-channel resolvability for \( \{X_i, \alpha_i\} \in \Theta \).

**Remark 1:** In this problem, the condition for the approximation measure \( \{15\} \) is changed from \( \{3\} \). It is well-known that the variational distance is jointly convex in its arguments, and in general it holds that
\[
d(P_{Y^n}, P_{Y^n}) \leq \sum_{i \in \Theta} \alpha_i d(P_{Y^n}, P_{Y^n}), 
\]
where
\[
P_{Y^n}(y) = \sum_{i \in \Theta} \alpha_i P_{Y^n}(y) \quad (\forall y \in \mathcal{Y}^n). 
\]
Equation \( \{15\} \) imposes a more stringent condition than the one in \( \{3\} \). Since \( S_v(\delta|X, W) \) coincides with the \( \delta \)-mean channel resolvability \( \{11\} \), for which the coin distribution may be any general source, in general we have
\[
S_v(\delta|X, W) \leq S^*_v(\delta|\{X_i\}, W). 
\]

When the channel \( W^n \) is an identity mapping, the addressed problem reduces to that of source resolvability for \( \{X_i\} \).

**Definition 4:** Assume that the channel \( W^n \) is an identity mapping. The infimum of all \( v(\delta) \)-achievable rates for \( \{X_i, \alpha_i\} \in \Theta \):
\[
S^*_v(\delta|\{X_i\}, W) := \inf \{R : R \text{ is } v(\delta) \text{-achievable for } \{X_i\}\} 
\]
is called the \( \delta \)-variable-length source resolvability or simply \( v(\delta) \)-source resolvability for \( \{X_i, \alpha_i\} \in \Theta \).

**B. Theorems**

To characterize \( S^*_v(\delta|\{X_i\}, W) \), we define
\[
H^v_{\delta},W^n(\{X^n_i\}) := \inf_{\{P_{Y^n_i}\} \in B^v_H(\{X^n_i\}, W^n)} \sum_{i \in \Theta} \alpha_i H(V^n_i). 
\]
where \( B^v_H(\{X^n_i\}, W^n) \) is the set of distributions \( \{P_{Y^n_i}\} \in \mathcal{P}(\mathcal{X}^n) : \sum_{i \in \Theta} \alpha_i d(P_{Y^n_i}, P_{Z^n_i}) \leq \delta \).

where \( Z^n_i \) denotes the output random variable due to \( W^n \) with \( \delta \)-satisfying \( \{14\} \) and \( \{29\} \).

where \( H^v_{\delta},W^n(\{X^n_i\}) \) and \( H^v_{\delta},W^n(\{X^n_i\}) \) are nonincreasing monotone functions in \( \delta \). When the channel \( W^n \) is an identity mapping, \( H^v_{\delta},W^n(\{X^n_i\}) \) and \( H^v_{\delta},W^n(\{X^n_i\}) \) are denoted simply by \( H^v_{\delta}(\{X^n_i\}) \) and \( H^v_{\delta}(\{X^n_i\}) \), respectively. We establish the following theorem:

**Theorem 2:** For any mixed source \( X = \{(X_i, \alpha_i)\} \in \Theta \), it holds that
\[
S^*_v(\delta|\{X_i\}, W) = \lim_{\gamma \to 0} H^v_{\delta+\gamma}(im|\{X_i\}) \quad (\forall \delta \in [0,1)]. 
\]
(Proof) The proof is described in Sect. **IV.**

When \( W^n \) is an identity mapping, we have the following corollary.

**Corollary 1:** For any mixed source \( X = \{(X_i, \alpha_i)\} \in \Theta \), it holds that
\[
S^*_v(\delta|\{X_i\}, W) = \lim_{\gamma \to 0} H^v_{\delta+\gamma}(im|\{X_i\}) \quad (\forall \delta \in [0,1]). 
\]
(Proof) The proof is described in Sect. **IV.**

As is noted in Remark **1,** we have \( \{29\} \) in general. It is not clear if \( S^*_v(\delta|\{X_i\}, W) \) is equal to \( S_v(\delta|X, W) \). The following theorem provides an interesting relationship between the two \( v(\delta) \)-source resolvability problems for mixed sources.

**Theorem 3:** For any mixed source \( X = \{(X_i, \alpha_i)\} \in \Theta \), it holds that
\[
S_v(\delta|X) = S^*_v(\delta|\{X_i\}) \quad (\forall \delta \in [0,1]). 
\]
(Proof) The proof is described in Sect. **IV.**

**Remark 2:** The \( v(\delta) \)-source resolvability \( S_v(\delta|X) \) is equal to the minimum rate of the FV source coding achieving the decoding error probability asymptotically not greater than \( \delta \in [0,1] \). We denote by \( R^*_v(\delta|X) \) this minimum rate, and then from Theorem 2 we obtain
\[
R^*_v(\delta|X) = S_v(\delta|X) = S^*_v(\delta|\{X_i\}) \quad (\forall \delta \in [0,1)). 
\]
for any mixed source \( X = \{(X_i, \alpha_i)\} \in \Theta \). To characterize \( R^*_v(\delta|X) \), it suffices to analyze \( S^*_v(\delta|\{X_i\}) \), which may be easier for some mixed sources. In the succeeding sections, we demonstrate this claim for mixed memoryless sources.

**IV. PROOF OF THEOREMS 2 AND 3**

**A. Proof of Theorem 2**

1) **Converse Part:** Let \( R \) be \( v(\delta) \)-achievable for \( \{X_i\} \).

Then, there exist \( U(L_{\alpha_i}) \) and \( \varphi_n^{(i)} \) satisfying \( \{14\} \) and
\[
\limsup_{n \to \infty} \delta_n \leq \delta, 
\]
where we define
\[
\delta_n = \sum_{i \in \Theta} \alpha_i d(P_{Y^n_i}, P_{\tilde{Y}^n_i}). 
\]
and $\tilde{Y}_i^n$ is the output via $W^n$ due to the input $\tilde{X}_i^n = \varphi_n(i) (U(L_i^{(i)}))$. Equation (28) implies that for any given $\gamma > 0$, $\delta_n \leq \delta + \gamma$ for all $n \geq n_0$ with some $n_0 > 0$, and therefore

$$H_{[\delta+\gamma],W^n}(\{X_i^n\}) \leq H_{[\delta],W^n}(\{X_i^n\}) \quad (\forall n \geq n_0) \quad (30)$$

because $H_{[\delta],W^n}(\{X_i^n\})$ is a nonincreasing monotone function of $\delta$. Since $\{P_{\tilde{X}_i^n}\} \subset B_{\delta_n}(\{X_i^n\}, W^n)$, we have

$$H_{[\delta],W^n}(\{X_i^n\}) \leq \sum_{i \in \Theta} \alpha_i H(\tilde{X}_i^n). \quad (31)$$

On the other hand, it follows that

$$\sum_{i \in \Theta} \alpha_i H(\tilde{X}_i^n) \leq \sum_{i \in \Theta} \alpha_i H(U(\{L_i^{(i)}\}))$$

$$\leq \sum_{i \in \Theta} \alpha_i E[L_i^{(i)}] + \sum_{i \in \Theta} \alpha_i H(L_i^{(i)}), \quad (32)$$

where the inequality is due to the fact that $\varphi_n(i)$ is a deterministic mapping and $\tilde{X}_i^n = \varphi_n(i) (U(L_i^{(i)}))$. By invoking the well-known relation (cf. [1, Corollary 3.12]) it holds that

$$H(L_i^{(i)}) \leq \log(e \cdot E[L_i^{(i)}]). \quad (33)$$

In view of (14), (33) leads to

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i \in \Theta} \alpha_i H(L_i^{(i)})$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{i \in \Theta} \alpha_i \log(e \cdot E[L_i^{(i)}])$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \log \left( e \cdot \sum_{i \in \Theta} \alpha_i E[L_i^{(i)}] \right) = 0. \quad (34)$$

Combining (30), (32) yields

$$H_{[\delta+\gamma],W^n}(\{X_i^n\})$$

$$= \limsup_{n \to \infty} \frac{1}{n} H_{[\delta+\gamma],W^n}(\{X_i^n\})$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} E[L_i^n] + \limsup_{n \to \infty} \frac{1}{n} \sum_{i \in \Theta} \alpha_i H(L_i^{(i)}) \leq R,$n \to \infty

where we have used (13) for the first inequality and (14) and (34) for the second inequality. Since $\gamma > 0$ is arbitrary, we obtain

$$\lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W^n}(\{X_i^n\}) \leq R. \quad (35)$$

2) Direct Part: By the analogous argument to the proof of the direct part of Theorem [12], we can show that the rate $R := H^* + 3\gamma$ is $\nu(\delta)$-achievable for $\{X_i^n\}$, where $H^* = \lim_{\gamma \downarrow 0} H_{[\delta+\gamma],W^n}(\{X_i^n\})$ and $\gamma > 0$ is an arbitrarily small constant. The proof sketch is as follows:

(i) We choose some $\{P_{Y_i^n}\} \subset B_{[\delta+\gamma]}(\{X_i^n\}, W^n)$ satisfying

$$\sum_{i} \alpha_i H(V_i^n) \leq H_{[\delta+\gamma],W^n}(\{X_i^n\}) + \gamma. \quad (36)$$

By definition, we have

$$\sum_{i \in \Theta} \alpha_i d(P_{Y_i^n}, P_{Z_i^n}) \leq \delta + \gamma, \quad (37)$$

where $Z_i^n$ denotes the output via $W^n$ due to the input $V_i^n$.

(ii) Define

$$S_i^{(i)}(m) := \left\{ x \in X^n : \log \frac{1}{P_{Y_i^n}(x)} + n\gamma = m \right\}. \quad (38)$$

For each $i \in \Theta$, we set

$$\Pr[L_i^{(i)} = m] := \Pr[V_i^n \in S_i^{(i)}(m)]. \quad (39)$$

In the same way as in the proof of Theorem [11], we arrange an encoder $\varphi_n(i)$ to generate $X_i^n = \varphi_n(i) (U(L_i^{(i)}))$.

(iii) The average length rate can be evaluated as

$$E[L_i^{(i)}] \leq \left( 1 + \frac{1}{K n^\gamma} \right) (H(V_i^n) + n\gamma + 1), \quad (40)$$

whereas the variational distance satisfies

$$d(P_{Z_i^n}, P_{Y_i^n}) \leq d(P_{V_i^n}, P_{X_i^n}) \leq \frac{1}{2} K^{-n_\gamma} + \gamma. \quad (41)$$

From (40) and (41), we obtain

$$\limsup_{n \to \infty} \frac{1}{n} E[L_i^n] = \limsup_{n \to \infty} \frac{1}{n} \sum_{i \in \Theta} \alpha_i E[L_i^{(i)}]$$

$$\leq \limsup_{n \to \infty} \frac{1}{n} \sum_{i \in \Theta} \alpha_i (H(V_i^n) + 2\gamma)$$

$$\leq H^* + 3\gamma = R \quad (42)$$

and

$$\limsup_{n \to \infty} \sum_{i \in \Theta} \alpha_i d(P_{V_i^n}, P_{Y_i^n})$$

$$\leq \limsup_{n \to \infty} \sum_{i \in \Theta} \alpha_i (d(P_{V_i^n}, P_{Z_i^n}) + d(P_{Z_i^n}, P_{Y_i^n}))$$

$$\leq \limsup_{n \to \infty} \sum_{i \in \Theta} \alpha_i (d(P_{V_i^n}, P_{Z_i^n}) + \gamma \leq \delta + 2\gamma,$n \to \infty

where the first inequality is due to the triangle inequality and the third inequality follows from (37). Since $\gamma > 0$ is an arbitrary small constant, we conclude that $R$ is $\nu(\delta)$-achievable for $\{X_i^n\}$.

B. Proof of Theorem [5]

Assume, without loss of generality, that the elements of $X^n$ are indexed as $x_1, x_2, \ldots \in X^n$ so that

$$P_{X^n}(x_j) \geq P_{X^n}(x_{j+1}) \quad (\forall j = 1, 2, \ldots). \quad (43)$$

For a given $\delta \in [0, 1)$, let $j^*$ denote the integer satisfying

$$\sum_{j=1}^{j^*-1} P_{X^n}(x_j) < 1 - \delta, \quad \sum_{j=1}^{j^*} P_{X^n}(x_j) \geq 1 - \delta. \quad (44)$$


Let \( V^n \) be a random variable taking values in \( \mathcal{X}^n \) whose probability distribution is given by

\[
P_{V^n}(x_j) = \begin{cases} 
P_{X^n}(x_j) + \delta & \text{for } j = 1, 2, \ldots, j^* - 1 \\
\eta_{j^*} & \text{for } j = j^* \\
0 & \text{otherwise,} \end{cases}
\]

(45)

where we define \( \varepsilon = \delta - \sum_{j=j^*+1} P_{X^n}(x_j) \). It is easily checked that \( 0 \leq \varepsilon \leq P_{X^n}(x_j) \) and the probability distribution \( P_{V^n} \) majorizes any \( P_{V^n} \) in \( B_3(\mathcal{X}^n) \). Since the Shannon entropy is a Schur concave distribution, we have the following lemma, which provides a characterization of \( H_{[\delta]}(X^n) \).

Lemma 1 (45):

\[
H_{[\delta]}(X^n) = H(V^n) \quad (\forall \delta \in [0, 1]).
\]

(46)

Let \( j^* \) be the integer satisfying (44). Let \( V^n \) be a random variable taking values in \( \mathcal{X}^n \) whose probability distribution is given by

\[
P_{V^n}(x_j) = \begin{cases} 
P_{X^n}(x_j) & \text{for } j = 1, 2, \ldots, j^* - 1 \\
\eta & \text{for } j = j^* \\
0 & \text{otherwise,} \end{cases}
\]

(47)

where we define \( \eta = \sum_{j=j^*} P_{X^n}(x_j) \). To prove Theorem 2, we immediately obtain the following lemma, which is of use.

Lemma 2: Let \( X^n = \{(X^n_i, \alpha_i)\}_{i \in \Theta} \) be a mixed source. Then,

\[
H(V^n) \leq H_{[\delta]}(X^n) + \frac{2 \log e}{e} \quad (\forall \delta \in [0, 1]).
\]

(48)

\[(\text{Proof})\] Let \( V^n \) be defined as in (45). From Lemma 1 we have

\[
H(V^n) - H_{[\delta]}(X^n)
= H(V^n) - H(V^n) \\
\leq P_{V^n}(x_1) \log \frac{1}{P_{V^n}(x_1)} + P_{V^n}(x_{j^*}) \log \frac{1}{P_{V^n}(x_{j^*})} \\
\leq \frac{2 \log e}{e},
\]

(49)

where the last inequality is due to \( x \log x \geq \frac{-\log e}{e} \) for all \( x > 0 \). For every \( i \in \Theta \), let \( P_{V^n_i} \) be the probability distribution satisfying

\[
P_{V^n_i}(x_j) = \begin{cases} 
P_{X^n_i}(x_j) & \text{for } j = 1, 2, \cdots, j^* - 1 \\
\eta_i & \text{for } j = j^* \\
0 & \text{otherwise,} \end{cases}
\]

(50)

where we define \( \eta_i = \sum_{j \geq j^*} P_{X^n_i}(x_j) \). Then, we can easily verify that

\[
P_{V^n}(x) = \sum_{i \in \Theta} \alpha_i P_{V^n_i}(x) \quad (\forall x \in \mathcal{X}^n).
\]

(51)

That is, \( \{V^n_i, \alpha_i\}_{i \in \Theta} \) is a mixed source. Defining

\[
D_n(i) = \left\{ x \in \mathcal{X}^n : P_{V^n_i}(x) > P_{X^n}(x) \right\},
\]

(52)

the average variational distance can be evaluated as

\[
\sum_{i \in \Theta} \alpha_i d(P_{X^n_i}, P_{V^n_i}) = \sum_{i \in \Theta} \alpha_i \sum_{x \in D_n(i)} (P_{V^n_i}(x) - P_{X^n_i}(x)) \\
= \sum_{i \in \Theta} \alpha_i \sum_{j \geq j^*} (P_{V^n_i}(x_j) - P_{X^n_i}(x_j)) \\
= \sum_{i \in \Theta} \alpha_i \sum_{j \geq j^*} P_{X^n_i}(x_j) \\
= \sum_{j \geq j^*} P_{X^n}(x_j) \leq \epsilon,
\]

(53)

where the inequality is due to (44). Since the Shannon entropy is a concave function, (51) and (53) imply that

\[
H(V^n) \geq \sum_{i \in \Theta} \alpha_i H(V^n_i) \geq H_{[\delta]}(\{X^n_i\}).
\]

(54)

Combining (48) and (54) with Theorem 1 and Corollary 1 we obtain

\[
S_c(\delta|X, W) \geq S_c(\delta|\{X_i\}, W).
\]

(55)

The reverse inequality obviously holds (cf. (19)), and hence we obtain the claim.

Remark 3: As is seen from the above proof arguments, Theorems 2 and 3 hold even with general probability space \( \Theta \).

V. RESOLVABILITY FOR MIXED MEMORYLESS SOURCES

In this section, we assume that the source \( X = \{X^n\}_{n=1}^\infty \) is a mixed memoryless source and the channel \( W \) is an identity mapping. Each component source \( X_i = \{X^n_i\}_{n=1}^\infty \), \( i \in \Theta \), is stationary and memoryless, which is specified by a source \( X_i \) over \( \mathcal{X} \) as

\[
P_{X^n_i}(x) = \prod_{j=1}^n P_{X_i}(x_j) \quad (\forall x = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n).
\]

(56)

Without loss of essential generality, we assume that

\[
+\infty > H(X_1) \geq H(X_2) \geq \cdots,
\]

(57)

where the component sources \( \{X_i\}_{i \in \Theta} \) are indexed in the decreasing order of \( H(X_i) \).

For given \( \delta \in (0, 1) \), we define the positive integer \( i^* \) satisfying

\[
\sum_{i < i^*} \alpha_i \leq \epsilon, \quad A_{i^*} := \sum_{i \leq i^*} \alpha_i > \epsilon.
\]

(58)
We demonstrate an application of the general relationship (26) between the two variable-length resolvability problems to establish a single-letter formula for the \(v(\delta)\)-source resolvability.

**Theorem 4:** For any mixed memoryless source \(X = \{(X_i, \alpha_i)\}_{i \in \Theta}\), it holds that
\[
S_v(\delta|X) = S^v_1(\delta|\{X_i\}) = (A_i - \delta)H(X_i) + \sum_{i > i^*} \alpha_i H(X_i). \tag{59}
\]

for all \(\delta \in [0,1]\). \(\square\)

**Remark 4:** As was mentioned in Remark 2 we have \(S_v(\delta|X) = R^*_v(\delta|X)\) for all \(\delta \in [0,1]\) for any general source \(X\), where \(R^*_v(\delta|X)\) denotes the minimum rate of the FV source coding achieving the decoding error probability asymptotically not greater than \(\delta \in [0,1]\). For mixed memoryless source \(X = \{X_i\}\), Koga and Yamamoto \(6\) (for \(\Theta = 2\)) and Kuzuoka \(8\) (for any finite \(\Theta\)) have shown that \(R^*_v(\delta|X)\) is characterized as
\[
R^*_v(\delta|X) = (A_i - \delta)H(X_i) + \sum_{i > i^*} \alpha_i H(X_i) \tag{60}
\]
for all \(\delta \in [0,1]\) if the source alphabet \(X\) is finite. Since formula \(59\) holds for any countably infinite \(\Theta\) and \(X\), the relation \(S_v(\delta|X) = R^*_v(\delta|X)\) implies that formula \(60\) actually holds for a wider class of mixed memoryless sources. \(\square\)

Since any stationary memoryless source is a mixed source with a singleton set \(\Theta\), we immediately obtain the following corollary.

**Corollary 2 \((6), (12)\):** Let \(X\) be a stationary memoryless source \(X\). Then, it holds that
\[
S_v(\delta|X) = (1 - \delta)H(X). \tag{61}
\]

for all \(\delta \in [0,1]\). \(\square\)

**(Proof of Theorem 4)**

The following argument demonstrates the usefulness of the general relationship \(26\) between the two variable-length resolvability problems. Since it holds that
\[
S_v(\delta|X) = S^v_1(\delta|\{X_i\}) = \lim_{\gamma \downarrow 0} H^\dagger_{[\delta + \gamma]}(\{X_i\}) \tag{62}
\]
as is shown in Corollary 1, we first focus on the quantity \(H^\dagger_{[\delta]}(\{X_i^n\})\). The \(\delta\)-ball \(B_{[\delta]}(\{X_i^n\}, W^n)\), which is defined as \(B_{[\delta]}(\{X_i^n\}, W^n)\) with an identity mapping \(W^n\), can be written as
\[
B_{[\delta]}(\{X_i^n\}) = \{P_{X^n} : \exists \delta_i \geq 0 \text{ s.t. } \sum_i \alpha_i \delta_i = \delta, \quad d(P_{X^n}, P_{X^n}^* \leq \delta_i, \forall i \in \Theta) = \sum_{i \geq 0: \alpha_i \delta_i = \delta} B_{\delta_i}(X_i^n). \tag{63}
\]

Then, it obviously holds that
\[
H^\dagger_{[\delta]}(\{X_i^n\}) = \inf_{\{P_{X^n}\} \in B_{[\delta]}(\{X_i^n\})} \sum_i \alpha_i H(V^n_i) \geq \inf_{\{\delta_i \geq 0: \sum_i \alpha_i \delta_i = \delta\}} \inf_{P_{X^n} \in B_{\delta_i}(X_i^n)} \sum_i \alpha_i H(V^n_i) \tag{64}
\]
\[
\geq \inf_{\{\delta_i \geq 0: \sum_i \alpha_i \delta_i = \delta\}} \inf_{P_{X^n} \in B_{\delta_i}(X_i^n)} H(V^n_i) = \inf_{\{\delta_i \geq 0: \sum_i \alpha_i \delta_i = \delta\}} \sum_i \alpha_i H(\delta_i|X_i^n). \tag{65}
\]

It is known (cf. \(14\)) that
\[
\lim_{n \to \infty} \frac{1}{n} H_{[\delta]}(X_i^n) = (1 - \delta)H(X_i) \quad (\forall \delta \in [0,1]) \tag{66}
\]
for any stationary memoryless source \(X = \{X_i^n\}_{n=1}\), and thus
\[
H^\dagger_{[\delta]}(\{X_i^n\}) = \limsup_{n \to \infty} \frac{1}{n} H^\dagger_{[\delta]}(\{X_i^n\}) \geq \liminf_{n \to \infty} \frac{1}{n} H^\dagger_{[\delta]}(\{X_i^n\}) \geq \inf_{\{\delta_i \geq 0: \sum_i \alpha_i \delta_i = \delta\}} \inf_{P_{X^n} \in B_{\delta_i}(X_i^n)} H(V^n_i) = \inf_{\{\delta_i \geq 0: \sum_i \alpha_i \delta_i = \delta\}} \sum_i \alpha_i (1 - \delta_i)H(X_i) \tag{67}
\]
where the second inequality is due to Fatou’s lemma. Noticing that the \(\inf\) in \(67\) is a linear program and in view of \(57\), we find that the solution is given by
\[
\varepsilon_i = \begin{cases}
\alpha_i & \text{for } i < i^* \\
\delta - \sum_{i < i^*} \alpha_i & \text{for } i = i^* \\
0 & \text{for } i > i^*.
\end{cases} \tag{68}
\]
yielding
\[
H^\dagger_{[\delta]}(\{X_i^n\}) = (A_i - \delta)H(X_i) + \sum_{i > i^*} \alpha_i H(X_i). \tag{69}
\]
The right-hand side is right-continuous in \(\delta \geq 0\), and thus it follows from Corollary 1 that
\[
S_v(\delta|X) = S^v_1(\delta|\{X_i\}) \geq (A_i - \delta)H(X_i) + \sum_{i > i^*} \alpha_i H(X_i), \tag{70}
\]
where it should be noted that \(i^* = i^*(\delta)\) is right-continuous in \(\delta\).

To show the reverse inequality, we start with the characterization \(64\). We choose
\[
\delta_i = \begin{cases}
\frac{1}{\alpha_i} & \text{for } i < i^* \\
\frac{\delta - \sum_{i < i^*} \alpha_i}{\alpha_i} & \text{for } i = i^* \\
0 & \text{for } i > i^*. \tag{71}
\end{cases}
\]
We also set probability distributions \( \{P_{V^n}\} \) on \( \mathcal{X}^n \) by

\[
P_{V^n}(x) = \begin{cases} 
\Delta(x) & \text{for } i < i^* \\
(1 - \delta_i) P_{X^n}(x) + \delta_i \Delta(x) & \text{for } i = i^* \\
P_{X^n}(x) & \text{for } i > i^*,
\end{cases}
\]

(72)

where \( \Delta(x) = 1\{x = x_0\} \) is the delta distribution with some specific \( x_0 \in \mathcal{X}^n \). Then, it is easily verified that

\[
d_i \geq 0 \quad (\forall i \in \Theta) \quad \text{s.t.} \quad \sum_{i \in \Theta} \alpha_i d_i = \delta,
\]

(73)

\[
d(P_{X^n}, P_{V^n}) \leq d_i \quad (\forall i \in \Theta),
\]

(74)

meaning \( P_{V^n} \in B_{d_i}(X^n) \) for all \( i \in \Theta \). Also, \( H(V^n) \) can be evaluated as

\[
H(V^n) = \sum_{x \in \mathcal{X}^n \setminus \{x_0\}} P_{V^n}(x) \log \frac{1}{1 - \delta_i - \delta_i \Delta(x)} + \log e
\]

\[
\leq \sum_{x \in \mathcal{X}^n} (1 - \delta_i) P_{X^n}(x) \log \frac{1}{1 - \delta_i} + \log e
\]

\[
\leq (1 - \delta_i) H(X^n) + 2 \log e,
\]

(75)

where the inequalities are due to \( x \log x \geq -\log e \) for all \( x \geq 0 \). With these choices of \( \{d_i\} \) and \( \{P_{V^n}\} \) satisfying (72)–(74), it follows from (64) that

\[
\frac{1}{n} H_{d_i}^i(\{X_i^n\}) \leq \frac{1}{n} \sum_{i \in \Theta} \alpha_i H(V^n)
\]

\[
= \frac{1}{n} \sum_{i \geq i^*} \alpha_i H(V^n)
\]

\[
= \frac{\alpha_{i^*}}{n} H(V^n) + \sum_{i > i^*} \alpha_i H(X_i).
\]

(76)

Taking the limit superior in \( n \) on both sides, we obtain

\[
\limsup_{n \to \infty} \frac{1}{n} H_{d_i}^i(\{X_i^n\}) \leq \limsup_{n \to \infty} \frac{\alpha_{i^*}}{n} H(V^n) + \sum_{i > i^*} \alpha_i H(X_i)
\]

\[
\leq \alpha_{i^*} (1 - \delta_i) H(X_{i^*}) + \sum_{i > i^*} \alpha_i H(X_i)
\]

\[
= (A_{i^*} - \delta) H(X_{i^*}) + \sum_{i > i^*} \alpha_i H(X_i),
\]

(77)

where the second inequality follows from (75). Again, the right-hand side is right-continuous in \( \delta \geq 0 \), and Corollary II indicates that

\[
S_\nu(\delta | X) = \nu(A_{i^*} - \delta) H(X_{i^*}) + \sum_{i > i^*} \alpha_i H(X_i).
\]

(78)

We complete the proof.

VI. SECOND-ORDER RESOLVABILITY FOR MIXED SOURCES

A. Definitions

In this section, we generalize the addressed problems to the second order case. The first definition corresponds to Definition 1 in the first order [12].

Definition 5: A second-order rate \( L \in (-\infty, +\infty) \) is said to be \( \nu(\delta, R) \)-achievable (under the variational distance) for \( X \) with \( \delta \in [0, 1) \) and \( R \geq 0 \) if there exists a variable-length uniform random number \( U^{(L_n)} \) and a deterministic mapping \( \varphi_n : U^* \to \mathcal{X}^n \) satisfying

\[
\limsup_{n \to \infty} \frac{1}{n} (\mathbb{E}[L_n] - nR) \leq L,
\]

(79)

\[
\limsup_{n \to \infty} d(P_{X^n}, P_{\varphi_n}) \leq \delta,
\]

(80)

where \( \tilde{X}^n = \varphi_n(U^{(L_n)}) \) and \( \mathbb{E}[L_n] \) is specified as in (13). The infimum of all \( \nu(\delta, R) \)-achievable rates for \( X \) is denoted by

\[
T\nu(\delta, R|X) := \inf \{ L : L \text{ is } \nu(\delta, R) \text{-achievable for } X \}.
\]

(81)

We also consider a variant problem for mixed sources \( X = \{X_i\} \).

Definition 6: A second-order rate \( L \in (-\infty, +\infty) \) is said to be \( \nu(\delta, R) \)-achievable (under the variational distance) for mixed source \( \{(X_i, \alpha_i)\}_{i \in \Theta} \) with \( \delta \in [0, 1) \) and \( R \geq 0 \) if there exists a set of variable-length uniform random number \( U^{(L_n)} \) and a deterministic mapping \( \varphi_n^{(i)} : U^* \to \mathcal{X}^n \) satisfying

\[
\limsup_{n \to \infty} \frac{1}{n} (\mathbb{E}[L_n] - nR) \leq L,
\]

(82)

\[
\limsup_{n \to \infty} \sum_{i \in \Theta} \alpha_i d(P_{X_i^n}, P_{\varphi_n^{(i)}}) \leq \delta,
\]

(83)

where \( \tilde{X}^n_i = \varphi_n^{(i)}(U^{(L_n)}) \). The infimum of all \( \nu(\delta, R) \)-achievable rates for \( \{(X_i, \alpha_i)\}_{i \in \Theta} \) is denote by:

\[
T\nu(\delta, R|\{X_i\}) := \inf \{ L : L \text{ is } \nu(\delta, R) \text{-achievable for } \{X_i\} \}.
\]

(84)

Remark 5: It is easily verified that

\[
T\nu(\delta, R|X) = \begin{cases} 
+\infty & \text{for } R < S_\nu(\delta | X) \\
-\infty & \text{for } R > S_\nu(\delta | X).
\end{cases}
\]

(85)

Hence, only the case \( R = S_\nu(\delta | X) \) is of our interest. The same remark also applies to \( T\nu(\delta, R|\{X_i\}) \).

B. Theorems

The following theorems indicate that \( T\nu(\delta, R|X) \) and \( T\nu(\delta, R|\{X_i\}) \) can also be characterized by the smooth entropies.
Theorem 5: For any mixed source \( \{(X_i, \alpha_i)\}_{i \in \Theta} \),
\[
T_v(\delta, R|X) = \lim_{\gamma \downarrow 0} \limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left( H_{\{\delta + \gamma\}}(X^n) - nR \right), \tag{86}
\]
\[
T_v^t(\delta, R|\{X_i\}) = \lim_{\gamma \downarrow 0} \limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left( H_{\{\delta + \gamma\}}(\{X_i^n\}) - nR \right) \tag{87}
\]
for all \( \delta \in [0, 1) \) and \( R \geq 0 \).

(Proof) For the proof of \( 86 \), see [12]. Formula \( 87 \) can be proven in a parallel way to Theorem 3.

As in the first order case, we have the equivalence between \( T_v(\delta, R|X) \) and \( T_v^t(\delta, R|\{X_i\}) \) for any mixed source \( \{X_i\} \).

Theorem 6: For any mixed source \( \{(X_i, \alpha_i)\}_{i \in \Theta} \),
\[
T_v(\delta, R|X) = T_v^t(\delta, R|\{X_i\}) \quad (\delta \in [0, 1), R \geq 0). \tag{88}
\]

(Proof) This theorem can be proven in a parallel way to Theorem 5.

We now turn to analyzing the \( v(\delta, R) \)-source resolvability for mixed memoryless sources. We assume the following properties:

(i) The index set \( \Theta \) is finite.

(ii) Each component source \( X_i \) has the finite third absolute moment of \( \log \frac{1}{P_{X_i}(x)} \).

(iii) Component sources \( \{X_i\} \) satisfy
\[
+\infty > H(X_1) > H(X_2) > \cdots. \tag{89}
\]

The following lemma is useful to establish a single-letter formula of the \( v(\delta, R) \)-source resolvability.

Lemma 3 ([7]): Assume that a stationary memoryless source \( X^n \) has a finite absolute moment of \( \log \frac{1}{P_{X^n}(x)} \). Then, it holds that
\[
H_{\{\delta\}}(X^n) = (1 - \delta)nH(X) - \sqrt{n} V(X) \left\{ \frac{e^{-\left(\frac{V(X)}{2}\right)^2}}{2\pi} \right\} + O(1), \tag{90}
\]
where \( V(X) \) denotes the variance of \( \log \frac{1}{P_{X^n}(x)} \) (varentropy) and \( Q^{-1} \) denotes the inverse of the complementary cumulative distribution function of the standard Gaussian distribution.

Theorem 7: Let \( X = \{(X_i, \alpha_i)\}_{i \in \Theta} \) be a mixed memoryless source satisfying (i)–(iii). For \( R = S_v(\delta|X) = S_v^t(\delta|\{X_i\}) \) given by (59), it holds that
\[
T_v(\delta, R|X) = T_v^t(\delta, R|\{X_i\}) = -\alpha_t \sqrt{\frac{V(X)}{2\pi} e^{-\left(\frac{V(X)}{2}\right)^2}}, \tag{91}
\]
where \( \alpha_t \) is the integer satisfying (58) and \( \delta_t \) is defined as in (71).

(Proof) The direct part is comparatively easy and we omit the proof due to the space limitation.

To prove the converse part, we define \( D(\delta) := \{|\delta_i| : \delta_i \geq 0, \sum_{i \in \Theta} \alpha_i \delta_i = \delta\} \). Using (65) and Lemma 3, we obtain
\[
\frac{1}{\sqrt{n}} H_{\{\delta\}}(X^n) \geq \inf_{(\delta_i) \in D(\delta)} \sum_{i \in \Theta} \alpha_i \left\{ (1 - \delta_i) \sqrt{n} H(X_i) - \sqrt{\frac{V(X)}{2\pi} e^{-\left(\frac{V(X)}{2}\right)^2}} + o(1) \right\}, \tag{92}
\]
and thus for all \( n > n_0 \) with some \( n_0 > 0 \) the minimizer \( \{\delta_i\} \in D(\delta) \) on the right-hand side is \( \{\delta_i\} \) given in (71): i.e.,
\[
\frac{1}{\sqrt{n}} H_{\{\delta\}}(X^n) \geq \sum_{i \geq t^*} \alpha_i (1 - \delta_i) \sqrt{n} H(X_i) - \alpha_t \sqrt{\frac{V(X)}{2\pi} e^{-\frac{1}{\pi} \left(\frac{V(X)}{2}\right)^2}} + o(1) \tag{93}
\]
for all \( n > n_0 \), where we have used the fact that \( e^{-\frac{1}{\pi} \left(\frac{V(X)}{2}\right)^2} = 0 \). Since \( R = S_v(\delta|X) \) is given by
\[
R = \sum_{i \geq t^*} \alpha_i (1 - \delta_i) H(X_i) \tag{94}
\]
due to Theorem 4, it follows that
\[
\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left( H_{\{\delta + \gamma\}}(X^n) - nR \right) \geq -\alpha_t \sqrt{\frac{V(X)}{2\pi} e^{-\frac{1}{\pi} \left(\frac{V(X)}{2}\right)^2}}. \tag{95}
\]
In view of Theorems 5 and 6 we complete the proof of the converse part.

Remark 6: As in the first order case, \( T_v(\delta, R|X) \) is equal to \( R_v^* (\delta, R|X) \), which denotes the minimum achievable rate of the \( v \)-source coding [14]. Theorem 7 also indicates that
\[
R_v^* (\delta, R|X) = -\alpha_t \sqrt{\frac{V(X)}{2\pi} e^{-\frac{1}{\pi} \left(\frac{V(X)}{2}\right)^2}}, \tag{96}
\]
for a mixed memoryless source \( X = \{(X_i, \alpha_i)\}_{i \in \Theta} \) satisfying (i)–(iii).
[11] H. Yagi and T. S. Han, “Variable-length resolvability for general sources and channels,” arXiv:1701.08712, Jan. 2017.

[12] H. Yagi and T. S. Han, “Variable-length resolvability for general sources,” Proc. 2017 IEEE Int. Symp. on Inf. Theory, Jun. 2017.

[13] H. Yagi and T. S. Han, “Variable-length resolvability for discrete memoryless channels,” submitted to Proc. 2018 IEEE Int. Symp. on Inf. Theory, Jun. 2018.

[14] H. Yagi and R. Nomura, “Variable-length coding with cost allowing non-vanishing error probability,” IEICE Trans. Fundamentals, vol. E100-A, no. 8, pp. 1683–1692, Aug. 2017.