Information Leakage in Index Coding

Yucheng Liu†, Lawrence Ong‡, Phee Lep Yeoh§, Parastoo Sadeghi†, Joerg Kliewer∗, and Sarah Johnson†,
†The University of Newcastle, Australia (emails: {yucheng.liu, lawrence.ong, sarah.johnson}@newcastle.edu.au)
‡University of Sydney, Australia (email: phee.yeoh@sydney.edu.au)
§University of New South Wales, Canberra, Australia (email: p.sadeghi@unsw.edu.au)
∗New Jersey Institute of Technology, USA (email: jkliwer@njit.edu)

Abstract—We study the information leakage to a guessing adversary in index coding with a general message distribution. Under both vanishing-error and zero-error decoding assumptions, we develop lower and upper bounds on the optimal leakage rate, which are based on the broadcast rate of the subproblem induced by the set of messages the adversary tries to guess. When the messages are independent and uniformly distributed, the lower and upper bounds match, establishing an equivalence between the two rates.

I. INTRODUCTION

Index coding [1], [2] studies the communication problem where a server broadcasts messages via a noiseless channel to multiple receivers with side information. Due to its simple yet fundamental model, index coding has been recognized as a canonical problem in network information theory, and is closely connected with many other problems such as network coding, distributed storage, and coded caching. Despite substantial progress achieved so far (see [3] and the references therein), the index coding problem remains open in general.

In secure index coding [4]–[7], the server must simultaneously satisfy the legitimate receivers’ decoding requirements and protect the content of some messages from being obtained by an eavesdropping adversary. A variant of this setup puts security constraints on the receivers themselves against some messages [4], [8], [9]. Instead of protecting the messages, another variant of index coding has also been studied from a privacy-preserving perspective, where the goal is to limit the information that a receiver can infer about the identities of the requests of other receivers [10]. The privacy-utility tradeoff in a multi-terminal data publishing problem inspired by index coding was investigated in [11].

In this work, we study the information leakage to a guessing adversary in index coding, which, to the best of our knowledge, has not been considered in the literature. The adversary eavesdrops the broadcast codeword and tries to guess the message tuple via maximum likelihood estimation within a certain number of trials. Our aim is to characterize the information leakage to the adversary, which is defined as the ratio between the adversary’s probability of successful guessing after and before observing the codeword [12]–[14]. For a visualization of the problem setup, see Figure 1.

Recently we have studied [15] information leakage to a guessing adversary in zero-error source coding defined by a family of confusion graphs [16]. While the index coding problem can also be characterized by a confusion graph family [17], the study of information leakage in index coding is intrinsically different from that of source coding in the following aspects. The most significant difference comes from the different internal structures within the two confusion graph families. More specifically, for the source coding model we considered [15], the relationship among the confusion graphs of different sequence lengths is characterized by the disjoins product [18]. On the other hand, for the index coding problem, the relationship among the confusion graphs cannot be characterized by any previously defined graph product. Another difference is that while our previous work [15] requires zero-error decoding at the legitimate receiver assuming worst-case source distribution, this paper considers both zero-error and vanishing-error scenarios and assume a general message distribution. Furthermore, in this work we take into account the adversary’s side information which can include any message in the system.

Our main contribution is developing lower and upper bounds (i.e., converse and achievability results) on the optimal information leakage rate, for both vanishing-error and zero-error scenarios. The converse bound is derived using graph-theoretic techniques based on the notion of confusion identity.
graphs for index coding. The achievability result is established by constructing a deterministic coding scheme as a composite of the coding schemes for two subproblems, one induced by the messages the adversary knows as side information and the other induced by the messages the adversary does not know and thus tries to guess.

Moreover, we show that when the messages are uniformly distributed and independent of each other (as in most existing works for index coding), the lower and upper bounds developed match. This establishes an equivalence between the optimal leakage rate of the problem and the optimal compression rate of the subproblem induced by the messages the adversary tries to guess.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Notation: For any \( a \in \mathbb{Z}^+ \), \( [a] \equiv \{ 1, 2, \ldots, a \} \). For any discrete random variable \( Z \) with probability distribution \( P_Z \), we denote its alphabet by \( Z \) with realizations \( z \in Z \).

There are \( n \) discrete memoryless stationary messages (sources), \( X_i, i \in [n] \), of some common finite alphabet \( \mathcal{X} \). For any \( S \subseteq [n] \), set \( X_S = (X_i, i \in S) \), \( x_S = (x_i, i \in S) \), and \( \mathcal{X}_S = \mathcal{X}^{|S|} \). Thus \( X_{[n]} \) denotes the tuple of all \( n \) messages, and \( x_{[n]} \in X_{[n]} \) denotes a realization of the message \( n \)-tuple. By convention, \( X_0 = x_0 = \emptyset \). We consider an arbitrary, but fixed distribution \( P_{X_{[n]}} \) on \( X_{[n]} \), assuming without loss of generality that it has full support.

There is a server containing all messages. It encodes the tuple of \( n \) message sequences \( X_{[n]}^t = (X_i^t, i \in [n]) \) according to some (possibly randomized) encoding function \( f \) to some codeword \( Y \) that takes values in the common finite alphabet \( \mathcal{Y} = \{ 1, 2, \ldots, M \} \). Each message sequence \( X_i^t = (X_{i,1}, X_{i,2}, \ldots, X_{i,t}) \) is of length \( t \) symbols. The server then transmits the codeword to \( n \) receivers via a noiseless broadcast channel of normalized unit capacity. Let \( P_{Y|X_{[n]}^t} \) denote the joint distribution of the message sequence tuple \( X_{[n]}^t \) and the codeword \( Y \). For any \( S \subseteq [n] \), we define the following notation for message sequence tuples.

- \( X_S^t = \mathcal{X}^{|S|} \).
- \( X_S = (X_i^t, i \in S) = (X_{S,1}, X_{S,2}, \ldots, X_{S,t}) \), where \( X_{S,j} = (X_{i,j}, i \in S) \) for every \( j \in [t] \). Note that \( X_{i,j} \) denotes the \( j \)-th symbol of message sequence \( X_i^t \).
- Similarly, \( X_i^t = (x_i^t, i \in S) = (x_{S,j}, j \in [t]) \), where \( x_{S,j} = (x_{i,j}, i \in S) \) for every \( j \in [t] \).

Also, as the messages are memoryless, for any \( x_{[n]}^t = (x_{[n],1}, x_{[n],2}, \ldots, x_{[n],t}) \), \( P_{X_{[n]}^t}(x_{[n]}^t) = \prod_{j \in [t]} P_{X_{[n],j}}(x_{[n],j}) \).

On the receiver side, we assume that receiver \( i \in [n] \) wishes to obtain message \( X_i^t \) and knows \( X_{A_i}^t \) as side information for some \( A_i \subseteq [n] \setminus \{ i \} \).

More formally, a \( (t, M, f, g) \) index code can be defined by

- One stochastic encoder \( f : \mathcal{X}^n \rightarrow \{ 1, 2, \ldots, M \} \) at the server that maps each message sequence tuple \( x_{[n]}^t \in \mathcal{X}^n \) to a codeword \( y \in \{ 1, 2, \ldots, M \} \), and

\[ R = \lim_{t \to \infty} \inf_{\epsilon \to 0} \frac{\log M}{t} \]
Most existing results in the literature on the optimal compression rate (vanishing or zero error) of index coding were established assuming deterministic encoding functions. Lemma 1 indicates that those results can be directly applied to characterizing $R$ and $\rho$.

Since we are considering fixed-length codes (rather than variable-length codes), the zero-error broadcast rate $\rho$ does not depend on $P_{X[n]}$ and can be characterized solely by the confusion graphs $\{\Gamma_t, t \in \mathbb{Z}^+\}$ as
\[
\rho = \lim_{t \to \infty} \frac{1}{t} \log \chi(\Gamma_t) = \lim_{t \to \infty} \frac{1}{t} \log \chi_t(\Gamma_t),
\]
where $\chi(\cdot)$ and $\chi_t(\cdot)$ respectively denote the chromatic number and fractional chromatic number of a graph, and the proof of (a) can be found in [3] Section 3.2.

It has been shown [19] that, with the messages $X[i]$ being uniformly distributed and independent of each other, the vanishing-error broadcast rate $R$ equals to the zero-error broadcast rate $\rho$. Such equivalence does not hold for a general distribution $P_{X[n]}$, as it has been shown in [20] that the (vanishing-error) broadcast rate $R$ can be strictly smaller than its zero-error counterpart $\rho$.

**Leakage to a guessing adversary:**

We assume the adversary knows messages $X_P$ and tries to guess the remaining messages $X_Q$, where $Q = [n] \setminus P$, via maximum likelihood estimation within a number of trials. In other words, the adversary generates a list of certain size of guesses, and is satisfied iff the true message sequence is in the list. We characterize the number of guesses the adversary can make by a function of sequence length, $c : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$, namely, the guessing capability function. We assume $c(t)$ to be non-decreasing and upper-bounded2 by $\alpha(\Gamma_t(Q))$, where $\alpha(\cdot)$ denotes the independence number of a graph.

Consider any valid $(t, M, f, g)$ index code. Before eavesdropping the codeword $y$, the expected probability of the adversary successfully guessing $x_Q^t$ with $c(t)$ number of guesses is
\[
P_s(x_Q^t) = \mathbb{E}_{x_Q^t} \left[ \max_{K \subseteq X_Q^t : |K| \leq c(t)} \sum_{x_K} P_{X_K^t|x_Q^t}(x_K|x_Q^t) \right],
\]
and the expected successful guessing probability after observing $y$ is
\[
P_s(x_Q^t, Y) = \mathbb{E}_{y,x_Q^t} \left[ \max_{K \subseteq X_Q^t : |K| \leq c(t)} \sum_{x_K} P_{X_K^t|Y,x_Q^t}(x_K|y,x_Q^t) \right].
\]
The leakage to the adversary, denoted by $L$, is defined as the logarithm of the ratio between the expected probabilities of the adversary successfully guessing $x_Q$ after and before observing the transmitted codeword $y$. That is,
\[
L = \log \frac{P_s(x_Q^t)}{P_s(x_Q^t,Y)}.
\]
The (optimal) leakage rate can then be defined as
\[
L = \lim_{t \to \infty} t^{-1} \inf_{\epsilon > 0} \mathbb{E}_{(t, M, f, g) \text{ code}} \inf_{\text{zero-error decoding}} L.
\]
By definition, we always have $L \leq \lambda$.

**III. INFORMATION LEAKAGE IN INDEX CODING**

**A. Leakage Under A General Message Distribution**

Consider any index coding problem $(ij \in A_i, i \in [n])$ with confusion graphs $\{\Gamma_t, t \in \mathbb{Z}^+\}$ and distribution $P_{X[n]}$. Our main result is the following theorem.

**Theorem 1:** For the vanishing-error leakage rate $L$, we have
\[
\rho(Q) - |Q| + \log \frac{1}{\max_{x_Q} \sum_{x_Q} P_{X_Q}(x_Q)} \leq L \leq R(Q).
\]

For the zero-error leakage rate $\lambda$, we have
\[
\rho(Q) - |Q| + \log \frac{1}{\max_{x_Q} \sum_{x_Q} P_{X_Q}(x_Q)} \leq \lambda \leq \rho(Q).
\]

In the following, we prove the lower and upper bounds in (7). As for (8), the lower bound follows directly from the lower bound in (7) and the fact that $L \leq \lambda$, and the upper bound can be shown using similar techniques to the proof of the upper bound in (7).

**Proof of the lower bound in (7):** Consider any $\epsilon > 0$ and any valid $(t, M, f, g)$ index code for which $P_e \leq \epsilon$.

Consider any codeword $y \in \mathcal{Y}$ and any realization $x_Q^t \in X_Q^t$. Let $G_{X_Q^t}(y, x_Q^t)$ denote the collection of realizations

![Figure 2. The confusion graph $\Gamma_1$ with $t = 1$ for the 3-message index coding instance (1–), (2[3], 3[2]). Note that, for example, $x_1 = \{0, 0, 0\}$ and $z_1 = \{0, 0, 1\}$ are confluence because $x_1 = 0 \neq z_1 = 1$ and $x_{A_3} = x_2 = z_2 = z_{A_3}$. Suppose $(0, 0, 0)$ and $(0, 0, 1)$ are mapped to the same codeword $y$ with certain nonzero probabilities. Then upon receiving this $y$, receiver 3 will not be able to tell whether the value for $X_2$ is 0 or 1 based on its side information of $X_2 = 0$. For this graph, it can be easily verified that the independence number is 2, and that the chromatic number equals to the fractional chromatic number, both of which equal to 4. We have drawn an optimal coloring scheme with 4 colors in the graph.](image-url)
which can be shown by contradiction as follows. Assume
independence number of $i$

We also have

which can be shown by contradiction as follows. Assume there exists two different $x_i^{[\mathcal{D}]}, z_i^{[\mathcal{D}]} \in X_i^{[\mathcal{D}]}$, such that $x_i^{[\mathcal{D}]} = x_i^{[\mathcal{D}]}$, $y_i^{[\mathcal{D}]} \in G_{X_i^{[\mathcal{D}]}}(y_i^{[\mathcal{D}]})$, $z_i^{[\mathcal{D}]} \in G_{X_i^{[\mathcal{D}]}}(y_i^{[\mathcal{D}]})$, and $x_i^{[\mathcal{D}]}$ and $z_i^{[\mathcal{D}]}$ are adjacent (i.e., confusable) in $\Gamma_i(Q)$. Hence, there exists some receiver $i \in Q$ such that $x_i^{[\mathcal{D}]} \neq x_i^{[\mathcal{D}]}$ and $x_i^{[\mathcal{D}]} \cap Q = z_i^{[\mathcal{D}]} \cap Q$. Then considering $x_i^{[\mathcal{D}]}$ and $z_i^{[\mathcal{D}]}$, since they have the same realizations for messages in $P$, we have $x_i^{[\mathcal{D}]} = (x_i^{[\mathcal{D}]} \cap Q, x_i^{[\mathcal{D}]} \cap Q) = (z_i^{[\mathcal{D}]} \cap Q, z_i^{[\mathcal{D}]} \cap Q).$ From the perspective of receiver $i$, upon receiving codeword $y_i$ and observing side information $x_i^{[\mathcal{D}]}$, it cannot tell whether the true sequence for message $i$ is $x_i^{[\mathcal{D}]}$ or $z_i^{[\mathcal{D}]}$. Therefore, with the transmitted codeword being $y_i$, either $x_i^{[\mathcal{D}]}$ or $z_i^{[\mathcal{D}]}$ being the true realization will lead to an erroneous decoding at receiver $i$, which contradicts the assumption that both $x_i^{[\mathcal{D}]}$ and $z_i^{[\mathcal{D}]}$ belong to $G_{X_i^{[\mathcal{D}]}}(y_i^{[\mathcal{D}]})$. Therefore, any realizations $x_i^{[\mathcal{D}]}$, $z_i^{[\mathcal{D}]} \in G_{X_i^{[\mathcal{D}]}}(y_i^{[\mathcal{D}]})$ must be not confusable and thus not adjacent to each other in $\Gamma_i(Q)$. In other words, the vertex subset $G_{X_i^{[\mathcal{D}]}}(y_i^{[\mathcal{D}]}) \subseteq V(\Gamma_i(Q))$ must be an independent set in $\Gamma_i(Q)$ and thus its cardinality is upper bounded by the independence number of $\Gamma_i(Q)$.

We lower bound $P_s(X_i^{[\mathcal{D}]}, Y)$, i.e., the adversary’s expected successful guessing probability after observing $Y$, as

\[
\sum_{y, x^{[\mathcal{D}]}} \max_{K} \sum_{x^{[\mathcal{D}]} \in K} P_{Y, X^{[\mathcal{D}]}}(y_i, x^{[\mathcal{D}]}) \geq \sum_{y, x^{[\mathcal{D}]}} \max_{K} \sum_{x^{[\mathcal{D}]} \in K} P_{Y, X^{[\mathcal{D}]}}(y_i, x^{[\mathcal{D}]}) \geq \sum_{y, x^{[\mathcal{D}]}} \sum_{x^{[\mathcal{D}]} \in K} P_{Y, X^{[\mathcal{D}]}}(y_i, x^{[\mathcal{D}]}) \geq \min(\frac{c(t)}{\alpha(\Gamma_i(Q))}, 1).
\]

where $c(t) = \min\{c(t), |G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})|\}$, and

- (a) follows from the fact that each $x^{[\mathcal{D}]} \in G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})}$ appears in exactly $\left(\frac{|G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})|-1}{c(t)}\right)$ subsets of $G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})}$. of size $c(t)$

- (b) follows from \((9)\), \((10)\), and that if $c(t) \leq |G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})}|$, then $\frac{c(t)}{\alpha(\Gamma_i(Q))} \geq \frac{c(t)}{\alpha(\Gamma_i(Q))}$, otherwise we have $c(t) > |G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})}|$ and thus $\frac{c(t)}{\alpha(\Gamma_i(Q))} \geq 1 \geq \frac{c(t)}{\alpha(\Gamma_i(Q))}$, where the inequality is due to the assumption that $c(t) \leq \alpha(\Gamma_i(Q))$.

For bounding $P_s(X_i^{[\mathcal{D}]}, Y)$, consider any two disjoint subsets $A, B \subseteq [n]$. Note that any realization $x^{[\mathcal{D}]}$ can be explicitly denoted as $(x_{A,1}, x_{A,2}, \ldots, x_{A,t})$. We have

\[
\sum_{x^{[\mathcal{D}]}} \max_{x^{[\mathcal{D}]} \in \mathcal{X}^{[\mathcal{D}]}} P_{X_{A,B}}(x^{[\mathcal{D}]}) = \sum_{x^{[\mathcal{D}]} \in \mathcal{X}^{[\mathcal{D}]}} \prod_{j \in [m]} P_{X_{A,j}}(x_{A,j}) \sum_{x^{[\mathcal{D}]}} \max_{x^{[\mathcal{D}]} \in \mathcal{X}^{[\mathcal{D}]}} P_{X_{A,B}}(x^{[\mathcal{D}]}, x_{B,j})
\]

where the last equality can be shown via induction. Based on \((11)\) and \((12)\), we have

\[
\frac{c(t)}{\alpha(\Gamma_i(Q))} \geq \min\{c(t), |G_{X_i^{[\mathcal{D}]}(y_i^{[\mathcal{D}]})|\}.
\]

Proof of the upper bound in \((7)\): Consider any decoding error $\epsilon > 0$. Construct a deterministic encoding function $f$ that maps messages $X_i^{[\mathcal{D}]}$ to codeword $Y = (Y_1, Y_2)$ according to the following rules.

1) Codeword $Y_1$ is generated from $X_i^{[\mathcal{D}]}$ according to some deterministic encoding function $f_1 : \mathcal{X}^{[\mathcal{D}]} \rightarrow \{1, 2, \ldots, |\mathcal{Y}_1|\}$ such that there exist some decoding functions $g_i, i \in P$ allowing zero-error decoding for all receivers $i \in P$ and that $t^{-1} \log |\mathcal{Y}_1| = \rho(P)$.

2) Codeword $Y_2$ is generated from $X_i^{[\mathcal{Q}]}$ according to some deterministic encoding function $f_2 : \mathcal{X}^{[\mathcal{Q}]} \rightarrow \{1, 2, \ldots, |\mathcal{Y}_2|\}$ such that there exist some decoding functions $g_i, i \in Q$ allowing $\epsilon$-error decoding for all receivers $i \in Q$ and that $t^{-1} \log |\mathcal{Y}_2| = R(Q)$. 

Such encoding functions \( f_1 \) and \( f_2 \) exist for sufficiently large \( t \). We further verify that the coding scheme described above leads to an average probability of error \( P_e \) no more than \( \epsilon \) and thus is valid. Note that \( f_1 \) and \( f_2 \) are all deterministic. Define \( B_{X_Q^t} = \{ x_Q^t : \text{there exists some } i \in Q \text{ such that } g_i(f_2(x_Q^t)) \neq x_i^t \} \). That is, \( B_{X_Q^t} \) denotes the set of \( x_Q^t \) for which there is at least one receiver \( i \in Q \) that decodes erroneously. We have
\[
\epsilon \geq \sum_{x_Q^t \in B_{X_Q^t}} P_{X_Q^t}(x_Q^t).
\]
Hence, we have
\[
P_e = \sum_{x_P^t} \sum_{x_Q^t \in B_{X_Q^t}} P_{X_Q^t}(x_Q^t) = \sum_{x_Q^t \in B_{X_Q^t}} P_{X_Q^t}(x_Q^t) \sum_{x_P^t} P_{X_P^t|X_Q^t}(x_P^t|x_Q^t) = \sum_{x_Q^t \in B_{X_Q^t}} P_{X_Q^t}(x_Q^t) \cdot 1 = \epsilon,
\]
where the last inequality follows from (13). Now we have shown that the proposed coding scheme is valid.

The optimal leakage rate is upper bounded by the rate of the information leakage of the proposed coding scheme as \( \epsilon \) goes to 0. Let \( Y_t, X_{[n]}^t \) denote the joint distribution of \( Y = (Y_1, Y_2) \) and \( X_{[n]}^t \) according to the proposed coding scheme. For any \( x_P^t \in X_P^t \) and \( y_2 \in Y_2 \), define
\[
\mathcal{X}_Q^t(x_P^t, y_2) = \{ x_Q^t \in X_Q^t : P_{Y_2, X_{[n]}^t}(y_2, x_P^t, x_Q^t) > 0 \},
\]
Then we have
\[
\mathcal{L} \leq \lim_{t \to 0} \lim_{t \to \infty} \frac{1}{t} \log \sum_{x_P^t} \max_{y_2, K \leq l(t)} \sum_{x_Q^t \in \mathcal{X}_Q^t} P_{Y_2, X_{[n]}^t}(y_2, x_P^t, x_Q^t) = \lim_{t \to 0} \lim_{t \to \infty} \frac{1}{t} \log \sum_{x_P^t} \max_{y_2, K \leq l(t)} \sum_{x_Q^t \in \mathcal{X}_Q^t} P_{X_{[n]}^t}(x_Q^t) = \lim_{t \to 0} \lim_{t \to \infty} \frac{1}{t} \log \sum_{x_P^t} \max_{y_2, K \leq l(t)} |\mathcal{X}_Q^t(x_P^t)| = \mathcal{L}.
\]
Remark 2: An interesting observation is that the bounds in Theorem 1 are independent of the guessing capability function \( c(t) \). Whether \( \mathcal{L} \) and \( \lambda \) depend on \( c(t) \) remains unclear.

The upper and lower bounds in Theorem 1 do not match in general, as shown in the following example.

Consider the 4-message index coding problem (1|4), (2|3), (3|2), (4|1), where the messages are binary and independent of each other with \( P_{X_1}(0) = 1/4 \) and \( P_{X_1}(1) = 3/4 \), and \( X_2, X_3, \) and \( X_4 \) all follow a uniform distribution. Consider an adversary knowing \( X_P = X_4 \) as side information, and thus \( Q = \{1, 2, 3\} \). The broadcast rate for the subproblem induced by \( Q \) has been previously found [20] to be \( R(Q) = 3 - \frac{3}{4} \log 3 \). By Theorem 1, the leakage rate \( \mathcal{L} \) is upper bounded by \( R(Q) \), and lower bounded as
\[
\mathcal{L} \geq \rho(Q) - |Q| + \log \frac{1}{\sum_{x_P} \max x_Q P_{X_{[n]}^t}(x_{[n]}^t)} = 3 - 3 \log \frac{1}{3/32 + 3/32} = 3 - \log 3.
\]
Note that \( \rho(Q) = 2 \) can be easily verified (for example, see [3] Section 8.6). For the zero-error leakage rate \( \lambda \), by [8] in Theorem 1, we have
\[
3 - \log 3 \leq \lambda \leq 2.
\]

B. Leakage Under A Uniform Message Distribution

In most existing works for index coding, the messages \( X_{[n]} \) are assumed to be uniformly distributed and thus independent of each other. In such cases, Theorem 1 simplifies to the following corollary.

**Corollary 1:** If \( P_{X_{[n]}^t} \) follows a uniform distribution, then
\[
\mathcal{L} = \lambda = R(Q) = \rho(Q).
\]

**Proof:** We have
\[
\rho(Q) - |Q| + \log \frac{1}{\sum_{x_P} \max x_Q P_{X_{[n]}^t}(x_{[n]}^t)} = \rho(Q) - |Q| + \log \frac{1}{|X|^{|P|} \cdot (1/|X|^{|n|})} = \rho(Q) = R(Q),
\]
where the last equality follows from the fact that the vanishing-error and zero-error broadcast rates are equal when messages are uniformly distributed [19]. Combining Theorem 1 and the above result yields (14).

**Remark 3:** Even though we have established the equivalence between the leakage and broadcast rates under uniform message distribution, a computable single-letter characterization of the value in (14) is unknown. Nevertheless, the equivalence between the leakage and broadcast rates means that the extensive results on the broadcast rate of index coding established in the literature (such as single-letter lower and upper bounds, explicit characterization for special cases, and structural properties) can be directly used to determine or bound the leakage rate.

**Remark 4:** As the leakage rate in (14) can be achieved by the proposed coding scheme in the achievability proof of Theorem 1 for any index coding instance with uniform message distribution satisfying \( R = R(P) + R(Q) \) (or equivalently, \( \rho = \rho(P) + \rho(Q) \)), we know that the broadcast rate and leakage rate can be simultaneously achieved by some deterministic index code.
REFERENCES

[1] Y. Birk and T. Kol, “Informed-source coding-on-demand (ISCOD) over broadcast channels,” in IEEE INFOCOM, Mar. 1998, pp. 1257–1264.
[2] Z. Bar-Yossef, Y. Birk, T. Jayram, and T. Kol, “Index coding with side information,” IEEE Trans. Inf. Theory, vol. 57, pp. 1479–1494, 2011.
[3] F. Arabbaji-Faei and Y.-H. Kim, “Fundamentals of index coding,” Foundations and Trends® in Communications and Information Theory, vol. 14, no. 3–4, pp. 163–346, 2018.
[4] S. H. Dau, V. Skachek, and Y. M. Chee, “On the security of index coding with side information,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3975–3988, 2012.
[5] L. Ong, B. N. Vellambi, P. L. Yeoh, J. Kliewer, and J. Yuan, “Secure index coding: Existence and construction,” in Proc. IEEE Int. Symp. on Information Theory (ISIT), Barcelona, Spain, 2016, pp. 2834–2838.
[6] Y. Liu, Y.-H. Kim, B. Vellambi, and P. Sadeghi, “On the capacity region for secure index coding,” in Proc. IEEE Information Theory Workshop (ITW), Guanzhou, China, Nov. 2018.
[7] L. Ong, B. N. Vellambi, J. Kliewer, and P. L. Yeoh, “A code and rate equivalence between secure network and index coding,” IEEE J. Sel. Areas Inf. Theory, vol. 2, no. 1, pp. 106–120, 2021.
[8] Y. Liu, P. Sadeghi, N. Aboutorab, and A. Sharififar, “Secure index coding with security constraints on receivers,” in Proc. IEEE Int. Symp. on Information Theory and its Applications (ISITA), Oct. 2020.
[9] V. Narayanan, J. Ravi, V. K. Mishra, B. K. Dey, N. Karamchandani, and V. M. Prabhakaran, “Private index coding,” arXiv preprint arXiv:2006.00257, 2020.
[10] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, “Privacy in index coding: k-limited-access schemes,” IEEE Trans. Inf. Theory, vol. 66, no. 5, pp. 2625–2641, 2019.
[11] Y. Liu, N. Ding, P. Sadeghi, and T. Rakotoarivelolo, “Privacy-utility tradeoff in a guessing framework inspired by index coding,” in Proc. IEEE Int. Symp. on Information Theory (ISIT), Jun. 2020, pp. 926–931.
[12] G. Smith, “On the foundations of quantitative information flow,” in International Conference on Foundations of Software Science and Computational Structures. Springer, 2009, pp. 288–302.
[13] C. Braun, K. Chatzikokolakis, and C. Palamidessi, “Quantitative notions of leakage for one-try attacks,” Electronic Notes in Theoretical Computer Science, vol. 249, pp. 75–91, 2009.
[14] I. Issa, A. B. Wagner, and S. Kamath, “An operational approach to information leakage,” IEEE Trans. Inf. Theory, vol. 66, no. 3, pp. 1625–1657, 2019.
[15] Y. Liu, L. Ong, S. Johnson, J. Kliewer, P. Sadeghi, and P. L. Yeoh, “Information leakage in zero-error source coding: A graph-theoretic perspective,” in Proc. IEEE Int. Symp. on Information Theory (ISIT), Melbourne, Australia, 2021.
[16] J. Körner, “Coding of an information source having ambiguous alphabet and the entropy of graphs,” in 6th Prague Conference on Information Theory, 1973, pp. 411–425.
[17] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, “Broadcasting with side information,” in 49th Annu. IEEE Symp. on Foundations of Computer Science (FOCS), Oct. 2008, pp. 823–832.
[18] E. R. Scheinerman and D. H. Ullman, Fractional Graph Theory: A Rational Approach to the Theory of Graphs. Courier Corporation, 2011.
[19] M. Langberg and M. Effros, “Network coding: Is zero error always possible?” in Proc. 49th Ann. Allerton Conf. Comm. Control Comput., 2011, pp. 1478–1485.
[20] S. Miyake and J. Muramatsu, “Index coding over correlated sources,” in Proc. IEEE Int. Symp. on Network Coding (NetCod), Sydney, Australia, 2015, pp. 36–40.