Private Pliable Index Coding

Tang Liu and Daniela Tuninetti
University of Illinois at Chicago, Chicago, IL 60607 USA,
Email: tliu44, danielat@uic.edu

***Abstract*—The Pliable Index CODing (PICOD) problem is a variant of the Index Coding (IC) problem, where the desired messages by the users, who are equipped with message side information, are part of the optimization. This paper studies the PICOD problem where users are subject to a *privacy constraint*. In particular, the following special class of private PICODs is investigated: 1) the side information structure is circular, and 2) each user can decode one and only one message. The first condition is a special case of the "circular-arc network topology hypergraph" class of PICOD studied in [6], for which an optimal solution was given without the privacy constraint. The second condition was first studied in [9] and was motivated by the need to keep content private in some distribution networks.**

This paper proposes both converse and achievable bounds. The proposed achievable scheme not only strictly outperforms the existing one for some values of the system parameters, but it is also information theoretically optimal in some settings. For the remaining cases, the proposed linear code is shown to require at most one more transmission than the best possible linear code.

I. Introduction

a) Pliable Index Coding (PICOD): PICOD is a variant of the Index Coding (IC) problem and was first introduced in [2]. In PICOD, the messages to be decoded by the users, who have message side information, are not part of the problem definition. Instead, in PICOD, the sender assigns to the users the messages they need to decode so that (i) the assigned messages were not already present in the local side information, and (ii) the length of the code that allows every user to recover the assigned messages has the shortest possible length. The PICOD problem formulation captures the nature of some content delivery applications, where there is flexibility in the choice of the desired messages to be delivered to the users. This flexibility allows to reduce the number of transmissions compared to an IC with the same side information structure.

The IC problem in its general form is known to be challenging [1]. The general PICOD problem is not simpler than the IC problem in terms of complexity. For instance, the linear PICOD (here the sender is restricted to use linear codes) is still NP-hard [10]. Some efficient algorithms to solve the general PICOD were proposed in [11]. For the case where the side information structure of the PICOD has "symmetry," we found the optimal code length (under no restriction on the encoding scheme that the sender can use) in [6]. However, the general PICOD problem is open.

b) Private PICOD: The problem of security and privacy in IC has been studied from different perspectives. In [3], the Authors proposed an IC model where an eavesdropper has a limited access to the side information sets and to the transmitted codeword; the goal here is to prevent the eavesdropper from obtaining any new information. In [5], the Authors considered an IC model where the sender must design a code that allows each user to decode its desired message, but at the same time prevents him from obtaining any information about the side information or the desired messages of the other users. This latter model has the flavor of the private information retrieval problem [12], where a user wants to hide its desired message and/or side information from the other users and the server. The Authors formulated the *private Index Coding* problem in [8], where the privacy is defined as the condition that a message should only be decoded at the use who has it as its desired message. Recently, in [9], the Authors extended the private IC in [8] to the PICOD framework. Only the case where the side information structure is "circular" and each user can decode one and only one message was considered in [9]. Several schemes were given in [9] and shown to provide the desired level of privacy, but the optimality was discussed only under the linear encoding constraint for some cases.

c) Contributions and Paper Organization: In this paper we study a generalization (in terms of the form of the side information sets) of the private PICOD model from [9], as formally described in Section II. We provide both achievable and the converse bounds, where past work only focused on linear achievable schemes. The main result of this paper is presented and discussed in Section III. In Section IV we derive both information theoretic and linear-code restricted converse bounds. We also provide linear achievable schemes and show they are either information theoretically optimal, or differ from the linear-code restricted converse by at most one transmission. Section V concludes the paper. Some proofs are in the Appendix.

II. System Model

A private $(n, m, \mathcal{A})$ PICOD(t) is defined as follows. There are $n \in \mathbb{N}$ users and one central transmitter. The user set is denoted as $U := \{u_1, u_2, \ldots, u_n\}$. There are $m \in \mathbb{N}$ independent and uniformly distributed binary messages of $\kappa \in \mathbb{N}$ bits each. The message set is denoted as $\mathcal{W} := \{w_1, w_2, \ldots, w_m\}$.

The central transmitter has knowledge of all messages $\mathcal{W}$. User u_i has the messages indexed by its side information set $A_i \subset [m]$, $i \in [n]$. The messages index by A_i are denoted as W_{A_i}. The collection of all side information sets is denoted as $\mathcal{A} := \{A_1, A_2, \ldots, A_n\}$, which is assumed globally known at all users and the transmitter.

The sender and the users are connected by an error-free broadcast link. The sender transmits the codeword

$$x^{\kappa\ell} := \mathsf{ENC}(\mathcal{W}, \mathcal{A}), \tag{1}$$

where ENC is the encoding function.

The decoding function for user u_j is

$$\{\widehat{w}_1^{(j)}, \ldots, \widehat{w}_t^{(j)}\} := \mathsf{DEC}_j(W_{A_j}, x^{\kappa\ell}, \mathcal{A}), \ \forall j \in [n], \tag{2}$$

where t is the number of messages desired by user u_j that are not included in A_j. The decoding functions $\mathsf{DEC}_j, \ j \in [n]$, are such that decoding error probability

$$\begin{aligned}\Pr[\exists j \in [n], \forall \{d_{j,1}, \ldots, d_{j,t}\} \cap A_j = \varnothing : \\ \{\widehat{w}_1^{(j)}, \ldots, \widehat{w}_t^{(j)}\} \neq \{w_{d_{j,1}}, \ldots, w_{d_{j,t}}\}] \leqslant \epsilon,\end{aligned} \tag{3}$$

for some $\epsilon \in (0, 1)$. For a successful decoding at u_j we have $D_j := \{d_{j,1}, \ldots, d_{j,t}\} \subseteq [m] \backslash A_i$, i.e., D_j contains the indices of the t desired messages decoded by u_j.

Up to this point, the system definition is the same as the classical PICOD problem. We introduce now the privacy constraint. Privacy is modeled here as follows: user u_j can not decode any messages other than the t messages indexed by D_j. Specifically, we impose that for all $j \in [n]$

$$\begin{aligned}&H(W_{[m]\backslash(D_j \cup A_j)} | x^{\kappa\ell}, W_{A_j}, \mathcal{A}) \\ \geqslant &H(W_{[m]\backslash(D_j \cup A_j)}) - (m - t - |A_j|)\kappa\epsilon.\end{aligned} \tag{4}$$

A code is called *valid* for the private $(n, m, \mathcal{A})$ PICOD(t) if and only if it satisfies the conditions in (3) and (4). The goal is to find a valid code that result in the smallest possible codelength, i.e.,

$$\ell^\star := \min\{\ell : \exists \text{ a valid } x^{\kappa\ell} \text{ such that } \lim_{\kappa \to \infty} \epsilon = 0\}. \tag{5}$$

Finally, if the encoding function at the sender is restricted to be a linear map from the message set, the length of shortest possible such valid codes is denoted as $\ell^\star_{\text{lin}}$.

A. Network Topology Hypergraph (NTH) and size-s circular-h shift Side Information

In the rest of the paper we shall consider a class of $(n, m, \mathcal{A})$ private PICOD(t) problems with a specific structure on $\mathcal{A}$. Such class is a generalization of the one studied in [9], which is a special case of the circular-arc NTH in [6], where we fully solved the case $t = 1$ for the circular-arc NTH without the privacy constraint. The rest of the section contains graph definition that will be used later on.

Let $H = (V, \mathcal{E})$ denote a *hypergraph* with vertex set V and edge set $\mathcal{E}$, where an edge $E \in \mathcal{E}$ is a subset of V. The NTH, first introduced in [6], is a generalization of the network topology graph for the IC problem [4]. In a NTH, the hyperedges denote the messages, while the vertices denotes the users. A user does *NOT* have a message in its side information set if and only if its corresponding vertex is incident to the hyperedge that represents the message. A 1-*factor* of H is a spanning edge induced subgraph of H that is 1-regular. A hypergraph H is called an *circular-arc hypergraph* if there exists an ordering of the vertices $v_1, v_2, \ldots, v_n$ such that if $v_i, v_j, i \leqslant j$ are both incident to an edge E, then either $v_q, \forall q \in [i : j]$ are incident to E or $v_q, \forall q \in [m] \backslash [i : j]$ are incident to E.

In this paper we study the $(n, m, \mathcal{A})$ private PICOD(1) with a special side information set structure: the sets in $\mathcal{A}$ are size-s circular-h shift of the message set. More precisely, The side information set of user u_i is of the form

$$A_i = \{(i-1)h + 1, \ldots, (i-1)h + s\}, \tag{6}$$

for $i \in [n]$ where all indices are intended modulo the size of the message set, i.e., denoted as $(\text{mod } m)$ when needed, where $0 \leqslant s \leqslant m - t$ and $h \geqslant 1$, here $t = 1$.

Let $g := \gcd(m, h)$. In this private PICOD(1) there are $n = m/g$ users, since all users have distinct side information sets. Note that the size-s circular-h shift side information setup is a special case of the side information structure with *circular-arc* we introduced in [6]. Also, the model studied in [9] is the special case when $g = 1$ (and thus $n = m$).

III. Main Result

For the size-s circular-h shift side information private PICOD(1) problem, we have the following main result.

Theorem 1. *For the private PICOD(1) where the side information sets are as in* (6) *we have the following.*

Impossibility: when m is odd, $g = 1$, and either $s = m - 2$ or $s = 1$, it is not possible to satisfy the privacy constraint.

For the remaining possible cases, we have:

- *For $s \geqslant m/2$, and either $1 \leqslant s < m/2, g \geqslant 3$, or $1 \leqslant s < m/2, s \neq 2, g = 2$*

$$\ell^\star = \begin{cases} 1, & \text{if the NTH has a 1-factor,} \\ 2, & \text{otherwise.} \end{cases} \tag{7}$$

- *For $1 \leqslant s < m/2$, and either $g = 1$ or $s = g = 2$*

$$\lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil \leqslant \ell^\star_{\text{lin}} \leqslant \begin{cases} \lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil, & \frac{m}{s} \in \mathbb{Z}, \\ \lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil + 1, & \frac{m}{s} \notin \mathbb{Z}. \end{cases} \tag{8}$$

When $s \geqslant m/2$, the achievable scheme provided in [9] is indeed information theoretically optimal as it matches the converse bound in (7); this converse bound was derived in [6, Theorem 3] for the case without privacy constraint. Therefore, our main contribution in Theorem 1 is three-fold compared to [9]: 1) for $s \geqslant m/2$ we provide information theoretic optimality of the scheme in [9]; 2) for $s < m/2$ we provide a new achievable scheme, and show it is almost linear optimal; 3) we generalize the side information structure to any $g > 1$.

Remark: In (8), if we fix s and g, $\lfloor \frac{m}{s} \rfloor$ is monotonic in the message set size m. One interesting observation is that, although the lower bound on $\ell^\star_{\text{lin}}$ is monotonic in m, the upper bound is not. For instance, consider the case $s = 2, g = 1$; when $m = 10$ or $m = 12$, we have $\ell^\star_{\text{lin}} \leqslant 3$, while when $m = 11$ we have $\ell^\star_{\text{lin}} \leqslant 4$. In other words, from the point of $m = 11$ (here the upper and the lower bounds differ), both increasing and decreasing the message set size may result

in an increase of the required number of transmissions in our achievable scheme. It is not clear at this point whether this means the achievable scheme here is not optimal, or the optimal private *linear* PICOD solution is not monotonic in m.

IV. Proof of Theorem 1

We divide the proof of Theorem 1 into various cases. In this paper we prove Theorem 1 except for the achievability of (7). Specifically, the impossibility result is proved in Section IV-A, the case $s < m/2, g = 1$ in Section IV-B, and the case $s < m/2, g = s = 2$ in Section IV-C. The schemes that achieve (7) are sketched in Section IV-D, while the full proof can be found in [7, Appendix D].

A. Impossible Cases

First we show that in some cases the privacy constraint can not be satisfied. The proof of the same under a linear encoding constraint was provided in [9]. Here we provide a simple information theoretic proof of the same. The main idea is to proof the existence of a "decoding chain" (as defined in [6]) regardless of the choices of the desired messages at the users. This "decoding chain" technique was used in [6] for the converse proof of so called consecutive complete–S PICOD(t). Since this argument does not rely on any assumption on the encoding function at the server, the resulting bound is truly information theoretical (as opposed to a form of 'restricted converse').

1) Case m is odd, $s = m - 2$, and $g = 1$: Without loss of generality (Wlog) assume $h = 1$. User u_i has two possible choices for its desired message (because all the others are in its side information set); these messages are $d_i = (i+s) \pmod m$ or $d_i = (i - 1) \pmod m$. If $d_i = (i + s) \pmod m$, by decoding w_{d_i}, user u_i can mimic $u_{(i-1) \pmod m}$ since $A_{(i-1) \pmod m} \subset \{(i + s) \pmod m\} \cup A_i$. Therefore, user u_i can decode $w_{d_{(i-1) \pmod m}}$. To make sure user u_i can decode only one message, we need $d_{(i-1) \pmod m} \in A_i$ so that user u_i does not decode another message that is not in its side information set. We thus have $u_i \in A_{(i-1) \pmod m}$ and $d_{(i-1) \pmod m} \in A_i$. User u_i and $u_{(i-1) \pmod m}$ can mimic each other. We say that two user mimicking each other form a "loop". The same argument holds for the other choice of d_i as well. To make sure all users can decode one message only, every user must be in a loop. However, one user can be in only one loop. Thus, there must be one user that is not contained in any loop because here we have taken m to be odd. Therefore, there exists one user that can mimic another user and thus decode two messages, which violates the privacy constraint.

2) Case m is odd, $s = 1$, and $g = 1$: Wlog assume $h = 1$. User u_i, by decoding its desired message $d_i = j, j \neq i$, can mimic user u_j and thus also decode d_j. Sine user u_i can decode only one message, we have $d_j \in A_i$, i.e., $d_j = i$. User u_i and u_j form a loop. Similarly, every user can be in only one loop. We need all users to be in a loop to make sure that every user can decode at most one message. Since m is odd, this is impossible. Thus, there must exist one user that can decode two messages, which violates the privacy constraint.

B. Case $s < m/2$ and $g = 1$ (here $m = n$)

1) Achievability: Let $m = 2sq + r$ for some $q, r \in \mathbb{Z}$ such that $0 \leqslant r < 2s$, i.e., r is the remainder of m modulo $2s$, and q is the maximum number of users who can have disjoint side information sets. We can have $2q + \lfloor \frac{r}{s} \rfloor$ groups of s users such that the users in each group have at least one message in common in their side information sets. Also, $r - s\lfloor \frac{r}{s} \rfloor$ is the number of users that are not contained in any of these groups.

The intuition of our achievable scheme is as follows. Under the privacy constraint, we can satisfy the users in two groups with one transmission, therefore $2sq$ users can be satisfied by q transmissions. If $r = 0$, q transmissions suffice; if $0 < r \leqslant s$, we can satisfy the remaining r users by one transmission; and if $s < r < 2s$, we can satisfy the remaining r users by two transmissions. Therefore the total number of transmissions is $q + \lceil \frac{r}{s} \rceil$. Based on this intuition, we distinguish three sub-cases: a) $r = 0$; b) $0 < r \leqslant s$; and c) $s < r < 2s$.

Case $r = 0$: This is the case where m is divisible by $2s$, therefore is divisible by s. We partition the users into groups $G_1, G_2, \ldots, G_{2q}$, such that all users in G_i have message w_{is} in their side information. Set the desired message of the users in $G_{2i}, i \in [q]$, to be $w_{(2i-1)s}$, and the desired message of the users in $G_{2i-1}, i \in [q]$ to be w_{2is} There are q transmissions, each of them is $w_{2is} + w_{(2i-1)s}, i \in [q]$, that satisfies the users in G_i and G_{i+1} while it does not provide any useful information for the users in other groups. Therefore, $q = \frac{m}{2s}$ transmissions suffice to satisfy all the m users.

Case $0 < r \leqslant s$: We partition the users into $2q + 1$ groups. As for to the case $r = 0$, the first $2q$ groups contain s users. The users in $G_i, i \in [2q]$, all have w_{is} in their side information. Group G_{2q+1} has r users. The first q transmissions are $w_{2is} + w_{(2i-1)s}, i \in [q]$, and satisfy the users in groups $G_i, i \in [2q]$. We next satisfy the users in G_{2q+1}.

If $r = 1$, we have $G_{2q+1} = \{u_m\}$. Let $d_m = s + 1$ and the $(q + 1)$-th transmission be $w_{s+1} + \sum_{j \in A_m} w_j$. Note that $s \geqslant r + 1 = 2$, therefore user u_m can decode w_{s+1} while the other users can not decode any new messages once they receive the last transmission.

If $r \geqslant 2$, the users in G_{2q+1} all have $W_{[1:s-r] \cup \{m\}}$ in their side information. Let $d_{2sq+1} = s - r + 1$ and $d_j = 2sq + 1, j \in [2sq + 2 : m]$. The $(q + 1)$-th transmission is $w_{2sq+1} + w_m + \sum_{j=1}^{s-r+1} w_j$. Since user u_{2sq+1} can compute $w_{2sq+1} + w_m + \sum_{j=1}^{s-r} w_j$ and users $u_j, j \in [2sq + 2 : m]$, can compute $w_m + \sum_{j=1}^{s-r+1} w_j$, these users have the message that is not in their side information set as their desired message. All the other users who are not in G_{2q+1} have at least two messages unknown in the transmission and thus cannot decode it. Therefore, each user can decode only one message by the achievable scheme with $q + 1$ transmissions. If m is divisible by s, then $r = s$ and $q + 1 = \lceil \frac{m}{2s} \rceil$; if m is not divisible by s, $q + 1 = \lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil + 1$.

Case $s < r < 2s$: We partition the users into $2q + 2$ groups. The users in group $G_i, i \in [2q + 1]$, all have message $w_{(is)}$, while the users in group G_{2q+2} all have $W_{[1:2s-r] \cup \{m\}}$. We satisfy the first $2q$ groups by sending

$w_{2is} + w_{(2i-1)s}, i \in [q]$. We satisfy all users in G_{2q+1} by sending $w_{2sq+1} + w_{2sq+s} + w_{2sq+s+1}$. If $r = s+1$, $G_{2q+2} = \{u_m\}$ and we let $d_m = s+1$ and send as last transmission $w_{s+1} + \sum_{j \in A_m}$; otherwise, we let $d_{2sq+s+1} = 2s - r + 1$ and $d_j = 2sq + s + 1, j \in [2sq + s + 1 : m]$ and send $w_{2sq+s+1} + w_m + \sum_{i=1}^{2s-r+1} w_i$. One can verify that all users can decode one and only one message by using this code of length $q + 2 = \lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil + 1$.

2) Converse: Messages are bit vectors of length κ, for some κ; we thus see each message as an element in $\mathbb{F}_{2^\kappa}$. When the sender uses a linear code (on $\mathbb{F}_{2^\kappa}$), we can write the transmitted codeword as $x^\ell = Ew^m$, where $w^m = (w_1, w_2, \ldots, w_m)^T$ is the vector containing all the messages, and where $E \in \mathbb{F}_{2^\kappa}^{\ell \times m}$ is the generator matrix of the code. We denote the linear span of the row vectors of E as Span(E). Recall that in this setting, user $u_i, i \in [n]$, must to be able to decode one and only one message outside its side information set A_i; the index of the decoded message is d_i. Let $v_{i,j}$ be a vector whose j-th element is non-zero and all elements with index not in A_i are zeros.

A valid generator matrix E must satisfy the following two conditions:

1) *Decodability*: $v_{i,d_i} \in$ Span(E), for all $i \in [m]$;
2) *Privacy*: $v_{i,j} \notin$ Span(E), $\forall i \in [m], j \in [m] \backslash (A_i \cup \{d_i\})$.

The decodability condition guarantees successful decoding of the desired message w_{d_i} by user u_i as argued in [1]. The privacy condition must hold because the existence of a vector $v_{i,j} \in$ Span(E) for some $j \in [m] \backslash (A_i \cup \{d_i\})$ implies that user u_i is able to decode message w_j in addition to its desired message w_{d_i}.

The optimal linear code length $\ell^\star_{\text{lin}}$ is the smallest rank of the generator matrix E, which by definition is the maximum number of pairwise linearly independent vectors in Span(E). We prove the linear converse bound by giving a lower bound on the maximum number of pairwise linearly independent vectors in Span(E), i.e., the rank of E. To do so, we need the following two propositions, proved in Appendices A and B, respectively. These propositions are the key technical novelty of this work.

Proposition 1. *In a working system (where every user can decode without violating the privacy condition) with $g = 1$ we must have $e_i \notin$ Span(E) for all $i \in [m]$, where e_i are standard bases of m-dimensional linear space.*

Proposition 2. *For a working system with $g = 1$, among all n users, consider k users whose side information sets are pairwise disjoint. The number of transmissions of any linear code that satisfies these k users must be $\ell_{\text{lin}} \geqslant \lceil k/2 \rceil$.*

Proposition 1 states that in this case, a trivial 'uncoded scheme' (that consists of sending $\ell^\star_{\text{lin}}$ messages one by one) always violates the privacy constraint. In other words, no user is allowed to decode without using its side information.

Proposition 2 provides a lower bound on the code-length of a linear code for a subset of the users in the system (those with pairwise disjoint side information sets), thus for all users. Therefore, among all m users in the system, there are $\lfloor \frac{m}{s} \rfloor$ users with pairwise disjoint side information sets. By Proposition 2, we need at least $\lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil$ transmissions to satisfy these users. Therefore, in order to satisfy all the users in the system, we must have $\ell^\star_{\text{lin}} \geqslant \lceil \lfloor \frac{m}{s} \rfloor / 2 \rceil$. This provides the claimed lower bound.

C. Case $s < m/2$ and $g = s = 2$ (here $n = m/2$)

1) Achievability: In this case we show $\ell^\star_{\text{lin}} = \lceil m/4 \rceil$. We use the achievable scheme for case $s = 2 < m/2$ and $g = 1$ from Section IV-B1, where we need $\lceil m/4 \rceil$ transmissions to satisfy all $n = m$ users. The users we have here are a proper subset of the users in the case $g = 1$. The achievable scheme for $g = 1$ still satisfies all users and meets the privacy constraint. We have $\ell \leqslant \lceil m/4 \rceil$ in this case.

2) Converse: The converse proof in Section IV-B2 does not directly apply in this case, mainly because the proof of Proposition 1 requires $g = 1$. In Appendix C we show that it holds for $g = 2$, stated as Proposition 3.

Hence the converse follows the same argument in Section IV-B2 by replacing Proposition 1 with Proposition 3 in Appendix C. We show that for k users with pairwise disjoint side information sets, $\lceil k/2 \rceil$ transmissions are needed for this case under the linear encoding restriction. Note that in this case all $n = m/2$ users are with pairwise disjoint side information sets. Therefore, the total number of transmissions that satisfy all users is at least $\lceil m/4 \rceil$.

D. Remaining Cases

We aim to prove (7). Here we provide the converse proof, and a sketch of the achievability proofs. The detailed proofs can be found in [7, Appendix D].

1) Converse: By the converse bound in [6, Theorem 3] for the circular-arc PICOD(1) without the privacy constraint, we have $\ell^* \geqslant 1$ when the NTH has 1-factor, and $\ell^\star \geqslant 2$ when the NTH has no 1-factor. This converse bound holds also when we impose an additional privacy constraint.

2) Achievability for $s < m/2$, either $g = 2, s \neq 2$, or $g \geqslant 3$: We show how to find the first message to transmit. Then, all the users that do not have this message in their side information sets must be satisfied by a second transmission. We show how to find this second transmission in such a way that the privacy constraint is met.

3) Achievability for $s \geqslant m/2$: The achievable scheme in this case is the one proposed in [9], where only the case $g = 1$ was considered. For the cases where $g > 1$, the set of users in the system is a proper subset of the set of users when $g = 1$. Therefore the scheme for $g = 1$ is still valid for any g in that both decoding and privacy constraints are met.

V. Conclusion

In this paper we gave both achievable and converse bounds for the private PICOD(1) problem with circular side information sets. We proposed a linear achievable scheme is information theoretically optimal for some parameters, or it requires at most one more transmission compared to a converse developed under the constraint that the sender is restricted

to use linear codes. Proving, or disproving, that our linear codes are actually information theoretically optimal is subject of current investigation.

This work was supported in part by NSF Award number 1527059. The opinion expressed in this paper are of the authors and do not necessarily reflect those of the NSF.

APPENDIX A
PROOF OF PROPOSITION 1

Recall that, for $g = 1$, the side information sets are $A_i = (i, \dots, i + s - 1 \pmod m)$ for all $i \in [m]$, as here $n = m$. The proof is by contradiction. Wlog assume that we have a working system with $e_1 \in \text{Span}(E)$, that is, every user can decode message w_1 without even using its side information. Then, all users $u_i, i \in [2 : m - s + 1]$ (who do not have w_1 in their side information sets) must have desired message w_1, in order to make sure that the privacy constraint is not violated. This implies Fact 1: user u_1 can only have $w_{d_1} = w_{s+1}$ as desired message. Fact 1 is true because u_2 desires w_1, therefore $A_2 \cup \{d_2\} \supset A_1$. After decoding w_1, user u_2 can mimic user u_1 and thus decode message d_2. Since user u_2 can decode only one message, $d_1 \in A_2 \backslash A_1 = \{s + 1\}$. Therefore $d_1 = s+1$. By taking $d_1 = s+1$, we conclude that there must exist vector $v_{1,d_1} = v_{1,s+1} = c + \alpha_{s+1} e_{s+1}$, where $\alpha \in \mathbb{F}_{2^\kappa}, \alpha \neq 0$ and $c \in \text{Span}(A_1)$, where with an abuse of notation we let $\text{Span}(A_i)$ denote $\text{Span}(\{e_j : j \in A_i\})$.

Given Fact 1, let j be the position of the fist non-zero element in the so found $v_{1,s+1}$. Clearly, $j \leqslant s + 1$ since the $(s + 1)$-th element of $v_{1,s+1}$ is $\alpha_{s+1} \neq 0$. We have the following cases:

1) If $j = s+1$, all the users who do not have w_{s+1} in their side information sets can decode w_{s+1}, since $v_{1,s+1} = \alpha e_{s+1}$ for this case. u_{s+2} can decode both w_1 and w_{s+1}.
2) If $1 < j < s + 1$, then user u_{j+1} can decode w_j since $s + 1 \in A_j$. But user u_{j+1} decodes w_1 by assumption. Therefore, user u_j can decode both w_1 and w_j.
3) If $j = 1$, user u_{s+2} can decode both w_{s+1} and w_1. Therefore, u_{s+2} can decode two messages.

In all the three above cases, there exists at least one user who can decode at least two messages, thus violating the privacy constraint. Therefore, the original assumption $e_1 \in \text{Span}(E)$ must be impossible in a working system. The same reasoning applies to any $e_j, j \in [m]$. This proves the claim.

APPENDIX B
PROOF OF PROPOSITION 2

By Proposition 1, for all $i \in [k]$ there exists $v_{i,d_i} = \alpha_i e_{d_i} + c_i \in \text{Span}(E)$, where $c_i \in \text{Span}(A_i)$ and $\alpha_i \neq 0$. The side information sets A_i are assumed to be disjoint so the vectors c_i are linearly independent. v_{i,d_i} is linearly dependent on the vectors $v_{j,d_j}, \forall j \neq i$ only if $d_i \in A_j$ and $d_j \in A_i$ for some $i \neq j$. In other words, there exists a "loop" between u_i and u_j. Note that since the side information sets are disjoint, one user can be in at most one loop, and the number of loops is at most $\lfloor k/2 \rfloor$. Therefore the number of v_{i,d_i} that are linearly dependent is at most $\lfloor k/2 \rfloor$, and thus the number of *linearly independent* v_{i,d_i} is at least $k - \lfloor k/2 \rfloor = \lceil k/2 \rceil$. Therefore, the number of transmissions that is needed to satisfy k users with disjoint side information sets must satisfy $\ell = \text{rk}(E) \geqslant \lceil k/2 \rceil$.

APPENDIX C
PROOF OF PROPOSITION 3

Proposition 3. *In a working system (where every user can decode without violating the privacy condition) with $g = s = 2$ we must have $e_i \notin$ Span(E) for all $i \in [m]$, where e_i are standard bases of m-dimensional linear space.*

Similar to the proof of Proposition 1, Wlog assume e_1 is in $\text{Span}(E)$. All users $u_i, i \in [2 : m - s + 1]$ in this case need to desire message w_1. Let $d_1 \in A_j$, for some $j \neq 1$. There exists a vector $v_{1,d_1} \in \text{Span}(E)$ such that: 1) the d_1-th element is non-zero; 2) all elements with indices that are not $1, 2$ or d_1 are zeros. We check the first and second elements of v_{1,d_1} and have the following cases:

1) Both elements are zero, i.e., $v_{1,d_1} = e_{d_1}$. All users without w_{d_1} in their side information sets decode w_{d_1}.
2) The first is zero and the second is non-zero. The user u_j is able decode w_2 since u_j already decodes w_1 and has w_{d_1} in its side information set.
3) The first is non-zero and the second is zero. All users can decode w_1, then decode w_{d_1}.
4) Both are non-zero. u_j decodes w_1 by assumption. It also has w_{d_i} in its side information set. Therefore u_j can decode w_2.

All possible cases show that there exists at least one user that can decode at least two messages. The assumption that e_1 is in $\text{Span}(E)$ is impossible. The reasoning applies to all $e_j, j \in [m]$. Therefore we conclude that $e_i \notin \text{Span}(E)$ for all $i \in [m]$.

REFERENCES

[1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. on Information Theory*, vol. 57, no. 3, pp. 1479–1494, Mar 2011.

[2] S. Brahma and C. Fragouli, "Pliable index coding," *IEEE Trans. Information Theory*, vol. 61, no. 11, pp. 6192–6203, Nov 2015.

[3] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Trans. on Information Theory*, vol. 58, no. 6, pp. 3975 – 3988, June 2012.

[4] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. on Information Theory*, vol. 60, no. 1, pp. 529 – 568, Jan. 2014.

[5] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, "Private broadcasting: An index coding approach," *Proc. Int. Symp. Inf. Theory*, 2017. [Online]. Available: https://arxiv.org/abs/1701.04958.

[6] T. Liu and D. Tuninetti, "Tight information theoretic converse results for some pliable index coding problems," *ITW*, 2018. [Online]. Available: https://arxiv.org/abs/1810.02451.

[7] ——, "Private pliable index coding," *arXiv:1904.04468*, 2019.

[8] V. Narayanan, J. Ravi, V. K. Mishra, B. K. Dey, N. Karamchandani, and V. M. Prabhakaran, "Private index coding," *Proc. Int. Symp. Inf. Theory*, 2018.

[9] S. Sasi and B. S. Rajan, "On pliable index coding," *arXiv:1901.05809*, 2019.

[10] L. Song and C. Fragouli, "Content-type coding," *NetCod*, May 2015. [Online]. Available: https://arxiv.org/abs/1505.03561.

[11] ——, "A polynomial-time algorithm for pliable index coding," *IEEE Trans. on Information Theory*, vol. 64, no. 2, pp. 979 – 999, Feb 2018.

[12] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Trans. on Information Theory*, vol. 63, no. 7, pp. 4075 – 4088, July 2017.