

# Weakly Secure Network Coding

Kapil Bhattad, *Student Member, IEEE* and Krishna R. Narayanan, *Member, IEEE*  
Department of Electrical Engineering, Texas A&M University, College Station, USA

**Abstract**— In this work we consider the problem of secure data transmission on an acyclic multicast network. A new information theoretic model for security is proposed that defines the system as secure if an eavesdropper is unable to get any “meaningful” information about the source. The “wiretap” network model by Cai and Yeung in which no information of the source is made available to the eavesdropper is a special case in the new model.

We consider the case when the number of independent messages available to the eavesdropper is less than the multicast capacity of the network. We show that under the new security requirements communication is possible at the multicast capacity. A linear transformation is provided for networks with a given linear code to make the system secure. The transformation needs to be done only at the source and the operations at the intermediate nodes remain unchanged. We also show that if random coding scheme is used the probability of the system being secure can be made arbitrarily close to one by coding over a large enough field.

**Index Terms**— Secure Network Coding

## I. INTRODUCTION

IN their pioneering work, Ahlswede *et al* [1] showed that multicast rates could be increased by allowing for network coding instead of just routing. They showed that the multicast capacity is equal to the minimum of the max flows from the source to the destinations. Li and Yeung [8] showed that linear codes are sufficient to achieve the multicast capacity. Ho *et al* [4], [5] proposed a random coding scheme in which the messages on outgoing edges of a node are chosen to be a random linear combination of the messages on its incoming edges. They showed that the probability of the resulting network code being a valid multicast code can be made arbitrarily close to one by coding over a large enough field.

In many applications that require multicasting, it is necessary to make sure that only the intended destinations receive the transmitted data. This problem of secure data transmission on a multicast network was first studied by Cai and Yeung [2]. They considered a rate  $h$  multicast code used in an acyclic multicast network and they showed that if an eavesdropper had access to at most  $k$  independent messages ( $k < h$ ) then it is possible to modify the given linear code and transmit securely at a rate  $h - k$ . They imposed an information theoretic security requirement in which a system was said to be secure if and only if the mutual information between the messages available to the eavesdropper and the source symbols is zero.

In practice, the security requirement may not be this strict. For example, consider an eavesdropper having access to  $b_1 \oplus b_2$ . Although the eavesdropper has one bit of information about the source he is unable to recover any of the source bits. This may be secure enough for the application but is not information theoretically secure. If the security requirement is

weakened to suit practical requirements the loss in multicast rate can be reduced.

In [7], Jain derived a necessary and sufficient conditions for secure unicasting in cyclic networks at a rate of one symbol per unit time. An interesting observation made in [7] was that for a computationally limited eavesdropper with the use of one way functions it is possible to transmit at a higher rate without the eavesdropper getting any “meaningful” information about the source. This is another example where the system is secure for practical purposes although information theoretically it is not secure.

In other related work Feldman *et al* [3] consider the same set up as Cai and Yeung [2]. They show that the problem of finding a secure network code is same as that of finding a block code with some distance properties. They use that to show a tradeoff between the required field size and the multicast rate and show that the required field size could be unbounded if no loss in multicast rate is allowed. Ho *et al* [6] consider another aspect of security, security against Byzantine attack in which an attacker modifies some packets.

In this paper we give a new information theoretic model for security which accommodates a lot more practical requirements on security. The model proposed by Cai and Yeung is a special case in the new model. We give an information theoretic definition for meaningful information which is suited for many practical systems. We then give a constructive proof to show that it is possible to multicast without revealing any meaningful information and without any loss in rate when the number of independent messages available to the eavesdropper is less than the multicast capacity of the network.

We then study the robustness of the security scheme against an eavesdropper who is able to procure some side information. We show that it is possible to multicast at the multicast capacity using a secure network code that is maximally secure against guessing. i.e. if the eavesdropper has access to  $k$  independent messages and the multicast capacity is  $h$  then it is possible to construct a rate  $h$  multicast code such that for the first  $h - k - 1$  guesses the eavesdropper recovers at most one symbol per guess but with  $h - k$  guesses he recovers all  $h$  symbols.

We then consider the case of computationally limited eavesdropper. With the use of one way functions, we show that the system can be made secure against a computationally limited eavesdropper without any loss in rate when the number of independent messages available to the eavesdropper is less than the multicast capacity.

We compute bounds on the probability that a random code is not secure and show that the probability can be made arbitrarily close to zero by increasing the field size.

This paper is organized as follows. In the next section we

give the notation used in this paper. In section III we give a precise definition for meaningful information and describe our security model. In section IV we give a constructive proof to show that weakly secure communication is possible without any loss in rate. In section V we study the security provided by the proposed scheme when an eavesdropper is able to procure side information. In section VI we discuss the case of computationally limited eavesdropper. In section VII we extend our results to general security requirements. In section VIII we show that random codes are secure with high probability and finally we summarize our results in section IX.

## II. NETWORK MODEL AND NOTATION

We represent a network by a directed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the set of vertices (nodes) and  $\mathcal{E}$  is the set of edges (links). Each edge is assumed to have a unit capacity. In the multicast problem a source node  $\mathcal{S}$  sends information  $\mathbf{X}(t) = (x_1(t), x_2(t), \dots, x_r(t))^T$  at time  $t$  to destination nodes  $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_d$ . When we consider a particular time instant we represent the input by just  $\mathbf{X} = (x_1, x_2, \dots, x_r)^T$ . We denote the multicast capacity of this network by  $h$ .

A network code is said to be linear if the message on any outgoing edge of any node is a linear combination of the messages on the incoming edges of the node. We represent a linear code over graph  $\mathcal{G}$  by  $(\mathcal{G}, \Phi, r, F_q)$  where  $\Phi$  defines the input output relationship for each node,  $r$  is the rate of transmission and  $F_q$  is a field of size  $q$  over which the code is defined. In a linear code if the source node  $\mathcal{S}$  also transmits a linear combination of the source symbols in each of its outgoing edges then it is easy to see that the message on any edge will be a linear combination of the source symbols. Therefore for a linear code the message on edge  $e_j \in \mathcal{E}$  can be written as  $\Gamma_{e_j} \mathbf{X}$  where  $\Gamma_{e_j}$  is a length  $r$  vector over  $F_q$ . We note that if the source transmits a linear transformation of  $\mathbf{X}$ ,  $\mathbf{C}\mathbf{X}$ , instead of  $\mathbf{X}$  then the message transmitted on edge  $e_j$  would be  $\Gamma_{e_j} \mathbf{C}\mathbf{X}$ .

A wiretap network is specified by a collection  $\mathcal{A}$  of sets of edges.  $\mathcal{A} = \{A_1, A_2, \dots, A_{|\mathcal{A}|}\}$ ;  $A_i \subset \mathcal{E}$ . An eavesdropper selects a particular set  $A_i \in \mathcal{A}$  and listens to all messages transmitted on edges in  $A_i$  to get some information. We assume that the set doesn't change with time. For convenience we will use the name Eve for the eavesdropper. When we are specified a linear code and a wiretap network we use  $\mathbf{A}_i$  to represent a matrix whose rows contain all linearly independent  $\Gamma_{e_j}$ 's corresponding to edge  $e_j \in A_i$ . In this case the messages available to Eve is  $\mathbf{A}_i \mathbf{X}$ . The number of rows in  $\mathbf{A}_i$  is represented by  $k_i$ . We define  $k$  as  $\max_i k_i$ . We use the notation row space of  $A_i \in \mathcal{A}$  to represent the space spanned by the rows of  $\mathbf{A}_i$ .

We use  $\mathbf{b}_j$  to denote the  $j^{\text{th}}$  row of matrix  $\mathbf{B}$ . The  $j^{\text{th}}$  row of matrix  $\mathbf{A}_i$  is represented by  $\mathbf{a}_{i,j}$ . We represent rows  $j$  to  $j+r$  of matrix  $\mathbf{B}$  by  $\mathbf{B}_j^{j+r}$ .  $\mathbf{I}_r$  is used to denote an  $r \times r$  identity matrix. We use the notation  $\cup \mathbf{B}$  to represent the set that contains rows of  $\mathbf{B}$  as its elements. We use  $\text{span}\{\mathbf{B}\}$  or  $\text{span}\{\cup \mathbf{B}\}$  to represent row space of  $\mathbf{B}$ .

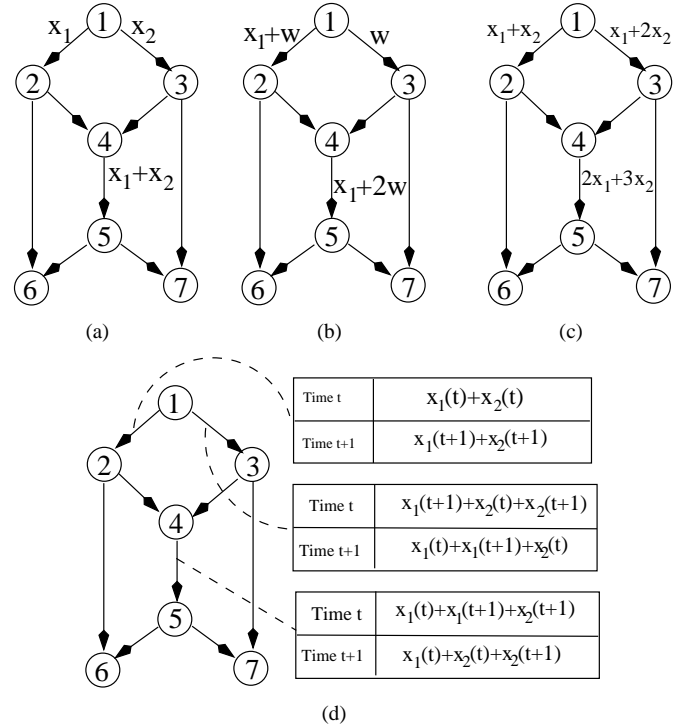


Fig. 1. Network 1

## III. MEANINGFUL INFORMATION

Consider a set of messages  $M$ . Let  $U$  and  $G$  be subsets of the set containing the multicast information symbols. We say that  $M$  has no information about  $U$  given  $G$  if  $I(U; M|G)$  is zero. We say that  $M$  has no *meaningful information* about  $U$  given  $G$  if  $I(x_i(t); M|G) = 0 \forall x_i(t) \in U$ . In this paper we consider the case when we have  $r$  streams of data that need to be multicast and the sets  $U$  and  $G$  are such that if  $x_i(t_1) \in U$  and  $x_j(t_2) \in G$  then  $x_i(t) \in U$  and  $x_j(t) \in G \forall t$ . The sets  $U$  and  $G$  are then subsets of the set of streams. We use  $|U|$  to represent the number of streams in  $U$ . In a general secure network coding problem a network, a collection of wiretap sets  $\mathcal{A}$ , and  $P$  pairs of sets  $(U_p, G_p)$  will be specified. We will need to design a code such that  $I(M; U_p|G_p) = 0 \forall p$ .

In this work we concentrate on two special cases and generalize the results towards the end. We say that Eve has no information about the source if  $I(\mathbf{X}(t); M) = 0 \forall t$  where  $M$  is the set of messages that Eve chooses to observe. This corresponds to the case when  $U$  contains all the source streams and  $G$  is empty. The security condition considered by Cai and Yeung [2] falls in this category. We will use Shannon security to refer to this security requirement. The second case we consider is when Eve gets no meaningful information about the source i.e.,  $I(\{x_i(t)\}_{t=1}^{\infty}; M) = 0 \forall i$  for messages  $M$  observed by Eve. This corresponds to the case when we are given pairs of sets  $(U_1, G_1), \dots, (U_r, G_r)$  with  $U_i = \{x_i(\cdot)\}$  and  $G_i$  is null. We call this type of security as weak security. For example, if Eve accesses an edge containing  $x_1 \oplus x_2$ , where  $x_1, x_2$  are i.i.d. bits from source she gets one bit of information about the source but she doesn't get any meaningful information. The system is then weakly secure but

not Shannon secure.

These two notions of security have different capacities as is shown in the following example. Consider the network shown in Fig. 1. The edges have unit capacity. The multicast capacity for this network is 2 and a capacity achieving coding scheme that multicasts two symbols,  $x_1$  and  $x_2$ , is shown in Fig. 1(a). Let us assume that Eve can listen to any one edge on this network. Consider the coding scheme shown in Fig. 1(b).  $x_1$  represents a sequence of source symbol and  $w$  is a uniform random sequence independent of the message. Each edge in the coding scheme shown gets message of the form  $ax + bw$ ,  $b \neq 0$  and its easy to see  $I(x_1; ax_1 + bw)$  is zero when  $b \neq 0$ . Therefore the coding scheme is Shannon secure. It can be shown that this is the maximum multicast rate supported when this system has to be Shannon secure. When the security condition is relaxed to weak security then using the scheme shown in Fig. 1(c) two symbols can be multicast without Eve getting any meaningful information about the source. With the data that is available on the edges she cannot determine any of the source symbols. In (Fig. 1(d)) another scheme that is weakly secure is shown, but here the coding is done over multiple time slots.

The codes shown in the example have two interesting properties. In the examples, it can be seen that the secure code (Fig. 1(b), 1(c), 1(d)) can be derived from the original code (Fig. 1(a)) by just changing the outputs of the source node. The encoding functions at the intermediate nodes do not require any change. The second property is that sometimes, to get a secure code, it may be necessary to increase the field size (Fig. 1(b), 1(c)) or the number of time slots (Fig. 1(d)) over which the coding is done. For example the code shown in Fig. 1(a) is a valid code in  $F_2$  but it is not possible to construct a code over  $F_2$  which is Shannon secure or weakly secure. Note that by increasing the number of time slots over which coding is done we are able to even keep the operating field in the intermediate nodes the same. The proofs for the results presented in this paper are for the case when the operating field size is changed. Very similar proofs can be used to show that the results hold when coding is done over multiple time slots. In the second case the intermediate nodes do not have to change their operations.

#### IV. MULTICAST CAPACITY OF WEAKLY SECURE NETWORK

In this section we find the multicast capacity of a wiretap network under weak security requirements when the number of independent messages available to the eavesdropper is less than the multicast capacity.

*Theorem 1:* Consider a network code  $(\mathcal{G}, \Phi, h, F_q)$  and a collection of sets of wiretap edges  $\mathcal{A}$ . If  $k = \max_{A \in \mathcal{A}} \text{rank}(\mathbf{A}) < h$  then there exist a transformation matrix  $\mathbf{C}$ ,  $h \times h$ , over  $F_{q^m}$ ,  $q^{mh} > |\mathcal{A}|q^{mk} + q^{m(h-1)}$ , which when applied to the source makes the network code weakly secure.

*Proof:* Let  $\mathbf{X} = (x_1, x_2, \dots, x_h)^T$  be the input vector that is sent over the network. Since a linear code is being used the message on each edge  $e_j$  can be written  $\Gamma_{e_j} \mathbf{X}$ . With

each set  $A_i \in \mathcal{A}$ ,  $A_i = (e_{j_1}, e_{j_2}, \dots, e_{j_{k_i}})$ , we associate a matrix  $\mathbf{A}_i$ ,  $k_i \times h$ , such that the  $l^{\text{th}}$  row of  $\mathbf{A}_i$  is  $\Gamma_{e_{j_l}}$ . The message obtained by Eve by accessing edges in  $A_i$  is  $\mathbf{A}_i \mathbf{X}$ . Eve tries to recover any information she can by performing linear operations on  $\mathbf{A}_i \mathbf{X}$ .

If the source transmits a linear transformation of the input, i.e., source transmits  $\mathbf{C}\mathbf{X}$  instead of  $\mathbf{X}$ , then the message available to Eve is  $\mathbf{A}_i \mathbf{C}\mathbf{X}$ . We note that in this case the destination nodes can solve for  $\mathbf{C}\mathbf{X}$  using the same network code and when  $\mathbf{C}$  is full rank they can recover  $\mathbf{X}$  from  $\mathbf{C}\mathbf{X}$ . If  $\mathbf{C}$  is not full rank some portion of the information will be lost.

We will now give a constructive proof to show that a linear transformation of the source is sufficient to achieve weak security. In a weakly secure system Eve cannot recover any information symbols from the observed messages. By taking linear combinations of the observed symbols  $\mathbf{A}_i \mathbf{C}\mathbf{X}$  Eve shouldn't be able to recover any  $x_j$  which implies

$$\mathbf{b}_i \mathbf{A}_i \mathbf{C} \neq \mathbf{I}_{h,n} \quad \forall \mathbf{b}_i, n, i \quad (1)$$

where  $\mathbf{I}_{h,n}$  is the  $n^{\text{th}}$  row of an  $h \times h$  identity matrix and  $\mathbf{b}_i$  is a  $k_i \times n$  vector in  $F_{q^m}$ . Multiplying both sides by  $\mathbf{C}^{-1}$ , the necessary condition for the system to be weakly secure becomes

$$\mathbf{b}_i \mathbf{A}_i \neq \mathbf{I}_{h,n} \mathbf{C}^{-1} \quad \forall \mathbf{b}_i, n, i \quad (2)$$

The above condition is satisfied if we find a full rank matrix  $\mathbf{C}^{-1}$  such that each row in  $\mathbf{C}^{-1}$  is not in the row space of each  $A_i \in \mathcal{A}$ . The matrix  $\mathbf{C}^{-1}$  is constructed as follows The  $(l+1)^{\text{th}}$  row in  $\mathbf{C}^{-1}$ ,  $\mathbf{c}_{l+1}^{-1}$ , is selected as a length  $h$  vector with elements in  $F_{q^m}$  that is not in  $\text{span}\{\mathbf{A}_i\}$  for  $i = 1 \dots L$  and not in  $\text{span}\{\{\mathbf{C}^{-1}\}_1^l\}$ , the previous  $l$  selected rows in  $\mathbf{C}^{-1}$ . Such a vector can definitely be found if the total number of vectors in the  $h$ -dimensional space is more than the sum of the number of vectors in the subspaces that it shouldn't lie in. i.e.  $\mathbf{c}_{l+1}^{-1}$  can be found if

$$q^{mh} > \sum_{A_i \in \mathcal{A}} q^{mk_i} + q^{ml} \quad \text{for } l = 0, 1, \dots, h-1 \quad (3)$$

Using the fact that  $k_i \leq k$  the sufficient condition in Eq. 3 is satisfied if

$$q^{mh} > |\mathcal{A}|q^{mk} + q^{m(h-1)} \quad (4)$$

The sufficient condition in Eq. 4 can be satisfied by choosing a sufficiently large  $q$  and hence a linear transformation  $\mathbf{C}$  can be found that transforms the given code into a weakly secure code. ■

*Corollary 1:* In a network that supports a multicast rate of  $h$ , if at most  $k$  ( $k < h$ ) edges can be tapped simultaneously then the multicast capacity under weak security requirements is  $h$ .

*Proof:* The network supports a multicast rate of  $h$  so a linear code can be found to multicast  $h$  symbols [8]. From theorem 1 if  $k < h$  a transformation at the source can be applied to make it weakly secure. ■

### A. Shannon Secure Network Coding

In this section we give a slightly different proof to some of the previous results in secure network coding using the same formulation as in the proof above. For this problem, a network code  $(\mathcal{G}, \Phi, h, F_q)$  and a collection of sets of edges  $\mathcal{A}$  is specified. The input vector is divided into two portions  $\mathbf{X} = ((x_1, \dots, x_r), (x_{r+1}, \dots, x_h))^T$ . The first  $r$  symbols are information symbols and the remaining symbols are chosen uniformly from  $F_{q^m}$ .

When Eve accesses edges in the set  $A_i$ , she sees messages  $\mathbf{A}_i \mathbf{X}$ . To make the system secure we have to find a transformation matrix  $\mathbf{C}$  such that Eve doesn't get any information about the source from the observed messages  $\mathbf{A}_i \mathbf{C} \mathbf{X}$ . Let  $k = \max \text{rank}(\mathbf{A}_i)$ . Eve can get information from the messages she observes only if she is able to get some linear combinations of the message symbols. Therefore the system will be secure if we are able to find a full rank matrix  $\mathbf{C}$  such that

$$\mathbf{b} \mathbf{A}_i \mathbf{C} \mathbf{X} \neq [\alpha_1 \alpha_2 \dots \alpha_r 0 \dots 0] \mathbf{X} \quad \forall i, \mathbf{b}, (\alpha_1 \dots \alpha_r) \neq 0 \quad (5)$$

Clearly Eq. 5 can be satisfied if we find a  $\mathbf{C}$  such that the row space of the first  $r$  rows of  $\mathbf{C}^{-1}$  has no vectors in common with the row space of any of the  $\mathbf{A}_i$ 's apart from of course the zero vector. The procedure for constructing  $\mathbf{C}^{-1}$  is as follows. For  $l = 0$  to  $r - 1$  we select  $\mathbf{c}_{l+1}^{-1}$  as a vector not in  $\text{span}\{\cup \mathbf{A}_i \cup \{\mathbf{C}^{-1}\}_1^l\}$  for all  $A_i \in \mathcal{A}$ .

*Lemma 1:* The space spanned by the first  $r$  rows of  $\mathbf{C}^{-1}$  obtained by using the procedure given above has no intersection with the space spanned by  $\mathbf{A}_i$  apart from the zero vector.

*Proof:* We prove by contradiction. Let us assume that

$$\sum_{i=1}^{k_i} \lambda_i \mathbf{a}_{i,j} = \sum_{l=1}^{r_1} \alpha_l \mathbf{c}_l^{-1} \quad (6)$$

We assume that  $r_1 \leq r$  is the largest  $i$  such that  $\alpha_i \neq 0$

$$\mathbf{c}_{r_1}^{-1} = \alpha_i^{-1} \left( \sum_{i=1}^{k_i} \lambda_i \mathbf{a}_{i,j} - \sum_{l=1}^{r_1-1} \alpha_l \mathbf{c}_l^{-1} \right) \quad (7)$$

i.e.  $\mathbf{c}_{r_1}^{-1} \in \text{span}\{\cup \mathbf{A}_i \cup_{j=1}^{r_1-1} \mathbf{c}_j^{-1}\}$  which is a contradiction. Hence proved. ■

Using Lemma 1 and the construction procedure if we are able to find  $r$  vectors then the remaining vectors in  $\mathbf{C}^{-1}$  can be chosen as basis vectors of the space orthogonal to the first  $r$  vectors. So if the  $r$  vectors are found the remaining  $h - r$  can be found. A sufficient condition for finding the  $r$  vectors is as follows. It is possible to find  $\mathbf{c}_{l+1}^{-1}$  if

$$q^{mh} > \sum_{A_i \in \mathcal{A}} q^{mk_i + ml} \quad \text{for } l = 0, 1, 2, \dots, r - 1. \quad (8)$$

Eq. 8 is satisfied if

$$q^{mh} > |\mathcal{A}| q^{mk + ml} \quad \text{for } l = 0, 1, 2, \dots, r - 1. \quad (9)$$

if  $(r > h - k)$  Eq. 8 cannot be satisfied. If  $r \leq h - k$  then matrix  $\mathbf{C}^{-1}$  can be found and the condition reduces to

$$q^{m(h+1-k-r)} > |\mathcal{A}| \quad (10)$$

if  $r = h - k$  then if  $q > |\mathcal{A}|$  a secure network code can be found [2]. if  $r = h - (1 + \epsilon)k$  then if  $q > |\mathcal{A}|^{\frac{k}{1+k\epsilon}}$  a secure

network code can be found [3]. Although the construction procedure for the matrix  $\mathbf{C}$  is the same as that in [2] the proof differs in the sense that in [2] the construction procedure is described and then it is proved that the code constructed is secure. Here, the construction procedure is motivated from the security condition and hence the resulting code is guaranteed to be secure. It also directly gives a sufficient condition on the required field size when some loss in rate is allowed.

### V. SECURITY AGAINST GUESSING

In this section we look at the security aspects when the eavesdropper is able to guess some information. We assume that Eve is able to perfectly guess some linear combinations of the message symbols. Eve is also allowed to choose the linear combinations. We study the amount of information she can recover per bit of information she guesses.

Consider a system that is Shannon secure. Then  $I(M; S) = 0$ . Let  $G$  represent the guesses.  $I(M; S, G) = I(M; S) + I(M; G|S) \leq H(G)$ . Therefore Eve can recover only one bit of information per bit of information that she guesses. Now consider the following example. In a system that multicasts 5 bits lets say that the eavesdropper has access to two observations  $x_1 \oplus x_5$  and  $x_2 \oplus x_5$ . By guessing one bit say  $x_5$  Eve will be able to get three bits of information. We will now show that it is possible to make the system weakly secure and maximally secure against guessing without any loss in rate.

*Theorem 2:* If an eavesdropper listens to messages on  $A_i$  and guesses information then it is possible to find a transformation  $\mathbf{C}$  over  $F_{q^m}$  such that the number of symbols recovered using  $g$  guesses is at most  $g$  when  $g < h - k_i$ . When  $g = h - k_i$  the eavesdropper can recover all  $h$  information symbols.

*Proof:* Before we prove the theorem we prove the following lemma's.

*Lemma 2:* Given a set of  $k$  linearly independent equations  $\mathbf{A} \mathbf{X} = \mathbf{b}$  for  $h$  variables such that none of the  $h$  variables can be solved, it is possible to solve for  $l$  variables say  $x_{j_1}, \dots, x_{j_l}$  given  $g$  linear equations  $\mathbf{b}_1 \mathbf{X} = b_1, \mathbf{b}_2 \mathbf{X} = b_2, \dots, \mathbf{b}_g \mathbf{X} = b_g$  if and only if the dimension of  $\text{span}\{\cup \mathbf{A} \cup_{i=1}^l I_{h,j_i}\} \leq k + g$ .

*Proof:* If we are able to solve for  $x_{j_1}, \dots, x_{j_l}$  using the  $g$  equations then we know that  $\{I_{h,j_1}, I_{h,j_2}, \dots, I_{h,j_l}\} \subset \text{span}\{\cup \mathbf{A} \cup \mathbf{b}_1, \dots, \cup \mathbf{b}_g\}$  and therefore  $\dim(\text{span}\{\cup \mathbf{A} \cup_{i=1}^l I_{h,j_i}\}) \leq k + g$ .

If  $\dim(\text{span}\{\cup \mathbf{A} \cup I_{h,j_1} \cup I_{h,j_2} \dots \cup I_{h,j_l}\}) \leq k + g$ , we can find  $g$  vectors  $\mathbf{b}_1, \dots, \mathbf{b}_g$  not in  $\text{span}(\mathbf{A})$  such that the span  $\{\cup \mathbf{A} \cup_{i=1}^g \mathbf{b}_i\}$  covers the  $k + g$  dimensional space that includes vectors  $I_{h,j_1}, I_{h,j_2}, \dots, I_{h,j_l}$ . ■

*Lemma 3:* Given a set of  $k$  linearly independent equations  $\mathbf{A} \mathbf{X} = \mathbf{b}$  for  $h$  variables such that none of the  $h$  variables can be solved, it is possible to solve for  $l$  ( $l > g$ ) variables given  $g$  equations ( $g + k < h$ ) if and only if the dimension of span of  $\mathbf{A}$  and some  $h - k$  rows of identity matrix  $\mathbf{I}_h$ , is less than  $h$ .

*Proof:* If we are able to find  $h - k$  rows of identity matrix  $\mathbf{I}_h$ ,  $(I_{h,j_1}, \dots, I_{h,j_{h-k}})$  such that  $D = \dim(\text{span}\{\cup \mathbf{A} \cup_{i=1}^{h-k} I_{h,j_i}\}) < h$ , then from Lemma 2 we can find  $g = D - k < h - k$  equations that can be used along with the  $k$  equations  $\mathbf{A} \mathbf{X} = \mathbf{b}$  to solve for the  $h - k$  variables.

Now consider the case when we are given  $g$  equations such that we are able to solve for  $l$  variables say  $x_{j_1}, \dots, x_{j_l}$ . Let  $x_{r_1}, \dots, x_{r_{h-l}}$  be variables that aren't in the set  $x_{j_1}, \dots, x_{j_l}$ . If  $l > (h - k)$  consider the space  $S_1 = \text{span}\{\cup_{i=1}^{h-k} I_{h,j_i}\}$ . By lemma 2 dimension of  $S_1$  is  $g + k < h$ . If  $l < h - k$  consider the space  $S_2 = \text{span}\{\cup_{i=1}^l I_{h,j_i} \cup_{i=1}^{(h-k-l)} I_{h,r_i}\}$ . Since the  $l$  variables can be solved with  $g$  equations, from lemma 2  $\dim(\text{span}\{\cup_{i=1}^l I_{h,j_i}\})$  is at most  $k + g$ . Space  $S_2$  is the span of these vectors in the  $k + g$  dimensional space and  $h - k - l$  more vectors. Therefore  $\dim(S_2) \leq k + g + (h - k - l) = h + g - l < h$ . ■

From Lemma 3 to ensure that the eavesdropper recovers only one symbol per guess up to  $h - k_i - 1$  guesses we should find a transformation  $\mathbf{C}$  such that the span of  $\mathbf{A}_i \mathbf{C}$  and any selections of  $h - k_i$  rows of  $I_{h \times h}$  is always the entire space. In other words we need to make sure that  $\mathbf{C}$  is such that span of  $A_i$  and any  $h - k_i$  rows of  $\mathbf{C}^{-1}$  is a  $h$ -dimensional vector space. Such a  $\mathbf{C}$  is constructed using the following procedure.

$\mathbf{c}_{1+1}^{-1}$  is chosen as a vector not in the span of  $\{\mathbf{C}^{-1} \mathbf{1}\}$  and not in the span of  $\mathbf{A}_i$  and all combination of  $\min((h - k_i - 1), l)$  vectors from  $(\mathbf{c}_1^{-1}, \dots, \mathbf{c}_l^{-1})$ . Using arguments similar to the ones used before we can show that the such a matrix can definitely be found if the field  $F_{q^m}$  is chosen such that

$$q^{mh} > \sum_{A_i \in \mathcal{A}} \binom{h-1}{h-k_i} q^{m(h-1)} + q^{m(h-1)} \quad (11)$$

We can find a large enough  $m$  such that the sufficient condition 11 is satisfied and hence a transformation  $\mathbf{C}$  can be found that ensures that the eavesdropper gets only one symbol per guess till  $h - k_i - 1$  guess. ■

## VI. COMPUTATIONALLY LIMITED EAVESDROPPER

In [7], Jain had shown that it is possible to increase the multicast capacity for secure communication when the eavesdropper is computationally bounded. He assumed the existence of one way function  $f$  such that computing  $f(x)$  from  $x$  is feasible but recovering  $x$  from  $f(x)$  is computationally infeasible. Use of one way functions in a weakly secure network code does not increase capacity when  $k < h$  but it could give significant reduction in the required field size. We assume that the one way function is such that  $f(x)+f(y)$  is random and if  $x$  and  $y$  are independent it does not give any information about any linear combination of  $x$  and  $y$ .

A simple scheme using one way functions is described below. For the network we first find a linear network code. Instead of transmitting symbols  $\mathbf{X} = (x_1, x_2, \dots, x_h)$  we transmit  $\mathbf{X} = (x_1, x_2 + f(\alpha_{11}x_1), x_3 + f(\alpha_{21}x_1 + \alpha_{22}x_2), \dots, x_h + f(\sum_{j=1}^{h-1} \alpha_{hj}x_j))$ . A linear transformation is done to the new source symbols so that the eavesdropper cannot recover the first symbol  $x_1$ . A sufficient condition on the field size security security against computationally limited eavesdropper is

$$q^{mh} > |\mathcal{A}|q^{mk} \quad (12)$$

Like before the system can also be made maximally secure against guessing.

## VII. MULTICAST CAPACITY UNDER ARBITRARY SECURITY CONSTRAINTS

In this section we consider the problem when we have data that requires different types of protection. We are given  $P$  pairs of sets  $U_p$  and  $G_p$  and the security requirement is that the eavesdropper should get no information about  $U_p$  given  $G_p$ .

*Theorem 3:* Consider a network code  $(\mathcal{G}, \Phi, h, F_q)$  and a collection of sets of wiretap edges  $\mathcal{A}$ . Let  $k = \max_{A \in \mathcal{A}} |A|$ . If the security requirements specified by sets  $(U_p, G_p)$  are such that  $\max(|U_p| + |G_p|) \leq h - k$  then there exists a linear transformation of the source symbols that makes the network code secure.

*Proof:* We first note that the information about message symbols in  $U_p$  given  $G_p$  is zero if and only if the span of vectors corresponding to streams in  $U_p$  has a non trivial intersection with the span of vectors in  $\mathbf{A}_i \mathbf{C} \cup G_p$ . If we make sure that  $\text{span } U_p \cup G_p$  has no nontrivial intersection with  $\mathbf{A}_i \mathbf{C}$  then  $\text{span}\{U_p\} \cap \text{span}\{\mathbf{A}_i \mathbf{C} \cup G_p\} = \{\mathbf{0}\}$ .

A matrix  $\mathbf{C}$  that satisfies these conditions is constructed as follows. The  $(l + 1)^{\text{th}}$  row in matrix  $\mathbf{C}^{-1}$  is constructed as follows.  $\mathbf{c}_{1+1}^{-1}$  is chosen as a vector not in the span of  $\{\{\mathbf{C}^{-1} \mathbf{1}\}_1\}$  and not in span of  $\{\cup_{i=1}^l \mathbf{c}_{j_i(p)}^{-1} \dots \cup \mathbf{c}_{j_z(p)}^{-1}\}$  for all  $i$  and for all  $p$  such that  $x_{l+1} \in U_p \cup G_p$  where  $\{x_{j_1(p)}, x_{j_2(p)}, \dots, x_{j_z(p)}\}$  is the largest set in  $U_p \cup G_p$  such that  $j_1(p), \dots, j_z(p)$  are all less than  $l + 1$ . If  $|U_p| + |G_p| < h - k$  then using the techniques used before we can show that we can find a valid  $\mathbf{C}$  in a sufficiently large field. ■

## VIII. SECURITY AND RANDOM CODING

Ho et al [4], [5] showed that if each node sets its outputs to be a random linear combination of its inputs then the network code obtained is a valid multicast code with probability  $(1 - \eta/q)^d$  where  $\eta$  is the number of times a random linear combination is done,  $q$  is the field size and  $d$  is the number of destination node. In this section we find bounds on the probability of a random code being secure.

*Theorem 4:* Given a network that employs random coding, the probability that an eavesdropper gets meaningful information about the source is less than  $\frac{|\mathcal{A}|hk}{q^{(h-k)}}$ . The probability that the eavesdropper is able to get more than one symbol per guess for any  $g < (n - k)$  guesses is bounded by  $\binom{h}{k} \frac{|\mathcal{A}|k}{q}$ .

The probability that a random code is not Shannon secure is bounded by  $\frac{|\mathcal{A}|k}{q}$ .

*Proof:* Consider an eavesdropper, Eve, listening to messages on edges in a set  $A_i \in \mathcal{A}$ . Eve recovers a particular symbol  $x_j$  iff row space of  $\mathbf{A}_i$  includes  $I_{h,j}$ . The only restriction imposed by the network is that the output message on any edge is a linear combination of the messages on the incoming edges. This translates to the fact that some  $k_i$  rows of  $A_i$  are independent and random and the others rows are a linear combination of these rows. An upper bound on the probability of Eve decoding any bit  $x_j$  by listening to edges in  $A_i$  when  $k_i < h$  and  $q > h$  is derived below.

Let  $\zeta$  be the number of matrices  $\mathbf{A}_i$ ,  $k_i \times h$ , with elements in  $F_q$  such that row space of  $\mathbf{A}_i$  doesn't include any row of the identity matrix.

$$\zeta \geq (q^h - qh)(q^h - q^2h) \dots (q^h - q^{k_i}h) \quad (13)$$

where each term in the product is a lower bound for the number of choices of  $\mathbf{a}_{i,j}$  given  $\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,j-1}$  such that the span  $\{\mathbf{a}_{i,1}, \dots, \mathbf{a}_{i,j}\}$  doesn't include any row of  $I_h$ . The number of matrices  $\mathbf{A}_i$  is  $q^{hk_i}$ . If  $P_i$  is the probability that row space of  $A_i$  doesn't include any row of  $I_h$  then

$$P_i \geq \frac{\prod_{j=1}^{k_i} (q^h - hq^j)}{q^{hk_i}} = \prod_{j=1}^{k_i} \left(1 - \frac{h}{q^{(h-j)}}\right) \quad (14)$$

We can now find a lower bound on the probability  $P_{WS}$  that a random coding scheme is weakly secure.

$$\begin{aligned} 1 - P_{WS} &\leq \sum_i (1 - P_i) \leq \sum_i \left(1 - \prod_{j=1}^{k_i} \left(1 - \frac{h}{q^{(h-j)}}\right)\right) \\ &\leq \sum_i \left(1 - \left(1 - \frac{h}{q^{(h-k_i)}}\right)^{k_i}\right) \leq \sum_i \frac{hk_i}{q^{(h-k_i)}} \\ &\leq \frac{|\mathcal{A}|kh}{q^{(h-k)}} \end{aligned} \quad (15)$$

We now consider the case when Eve is able to guess  $g$  linear combinations of the message symbols perfectly while listening to edges in  $\mathbf{A}_i$ . We are interested in finding the probability  $P_g(i)$  that Eve is able to recover more than  $g$  symbols with  $g$  guesses when  $g < h - k_i$ . Eve will be able to recover more than  $g$  symbols using  $g$  guesses if the space spanned by  $\mathbf{A}_i$  and some  $g + 1$  rows of  $I_h$  has a dimension strictly less than  $k_i + g + 1$ . Let  $\eta_g$  be the number of matrices  $\mathbf{A}_i$  with elements in  $F_q$  such that space spanned by  $\mathbf{A}_i$  and  $g + 1$  rows of  $I_h$  has a dimension  $k_i + g + 1$ . We have

$$\eta_g \geq \prod_{j=1}^{k_i} \left(q^h - \binom{h}{g+1} q^{g+j}\right) \geq \left(q^h - \binom{h}{g+1} q^{g+k_i}\right)^{k_i} \quad (16)$$

We have

$$P_g(i) = \frac{\eta_g}{q^h} \geq 1 - \binom{h}{g+1} \frac{k_i}{q^{h-k_i-g}} \quad (17)$$

The probability  $P_{WSG}$  that with  $g$  guesses the eavesdropper recover more than  $g$  symbols ( $g < h - k$ ) is upper bounded by

$$P_{WSG} \leq \sum_i (1 - P_g(i)) \leq \binom{h}{g+1} \frac{|\mathcal{A}|k}{q^{h-k-g}} \quad (18)$$

There is a tradeoff between the field size and security against guessing. If maximum possible security is required then  $g = h - k - 1$  and the lower bound on the probability of the system being maximally secure reduces to  $1 - \binom{h}{k} \frac{|\mathcal{A}|k}{q}$ .

Note that for the Shannon security case the proof remains almost the same as the proof for case with guessing but here we need not consider all combinations of  $g + 1$  rows of  $I_h$ . As data is sent only in the first  $(h - k)$  rows we need to pick  $g + 1$  rows from the first  $h - k$  rows of  $I_h$ . For large field size the bound in this case becomes  $\frac{|\mathcal{A}|k}{q}$ . ■

## IX. CONCLUSION

We defined a new model for security that is more suitable for practical applications. We showed that for practical security requirements like computationally limited eavesdropper when

the number of independent messages available to the eavesdropper is less than the multicast capacity, secure communication is possible without any loss in rate. We also showed that the processing necessary to make a given code secure had to be done only at the source and destination nodes leaving the operations at the intermediate nodes unchanged. We showed that it is possible to make the system maximally secure against guessing. Finally we showed that random coding is secure with high probability if the coding is done over a large field.

## X. ACKNOWLEDGEMENTS

We would like to thank Dr. Ralf Koetter for valuable suggestions given throughout the course of this work.

## REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204-1216, 2000.
- [2] N. Cai and R. W. Yeung, "Secure network coding," *International Symposium on Information Theory (ISIT) 2002*, June 30 - July 5, Lausanne, Switzerland.
- [3] J. Feldman, T. Malkin, C. Stein, R. A. Servedio "On the capacity of secure network coding", *Proc. 42nd Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2004.
- [4] T. Ho, M. Medard, J. Shi, M. Effros and D. R. Karger, "On randomized network coding", *41st Annual Allerton Conference on Communication Control and Computing*, Oct. 2003.
- [5] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "Toward a random operation of networks", *IEEE Transactions on Information Theory*, submitted. (2004).
- [6] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, D. Karger, "Byzantine modification detection in multiast networks using randomized network coding", *International Symposium on Information Theory (ISIT) 2004*, June 27 - July 2, Chicago.
- [7] K. Jain, "Security based on network topology against the wiretapping attack," *IEEE Wireless Communications*, pp. 68-71, Feb 2004.
- [8] S.-Y. R. Li, R. W. Yeung, and N. Cai. "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 371-381, 2003.