Network Security with Network Coding

Salim El Rouayheb
ECE epartment, Texas A&M University

joint work with Emina Soljanin

Paris

March 1, 2007

Outline

1. Introduction
   - Secure Network Codes: Example
   - The Wiretap Multicast Network
2. The Wiretap Channel of type II
   - Review
   - Coset Codes for the Wiretap Network
3. Code Alphabet
   - Bound on the Field Size
4. Conclusion

Outline

1. Introduction
   - Secure Network Codes: Example
   - The Wiretap Multicast Network
2. The Wiretap Channel of type II
   - Review
   - Coset Codes for the Wiretap Network
3. Code Alphabet
   - Bound on the Field Size
4. Conclusion

What Is Network Coding?

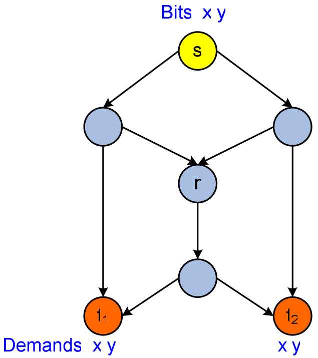


Figure: The butterfly network

- Source node s
- Two destination nodes t_1 and t_2
- Two bits per sec, x and y, are available at s
- All links are of capacity 1 bit/s

Problem

Multicast the information available at the source to all the destinations

Classical Approach

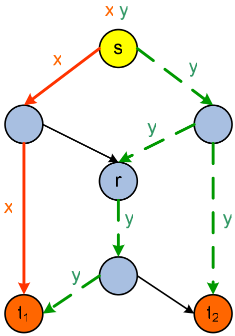


Figure: Routing (Steiner trees)

- Route the information along trees in the network
- we want to maximize the multicast rate to each destination
- $\Rightarrow$ Problem of Packing Steiner trees in graphs. NP-hard!
- average rate here 1.5 bits/s

Network Coding

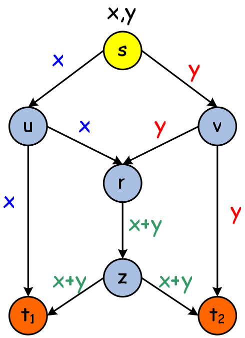


Figure: Network coding solution

- Network coding
  - Is an extension of routing
  - Allows "mixing" of packets at intermediate nodes
  - Achieves higher throughput

Linear Network Codes

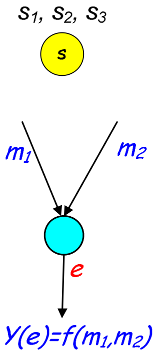


Figure: Linear Network Codes

Definition (Linear Network Code)

It is the collection of all the local encoding vectors for all the edges of a network.

- $s_1, s_2, s_3 \in GF(q)$
- $Y(e) = am_1 + bm_2$
- (a, b) is the local encoding vector of edge e
- $Y(e) = \alpha s_1 + \beta s_2 + \gamma s_3$
- (α, β, γ) global encoding vector of edge e

Wiretapped Butterfly

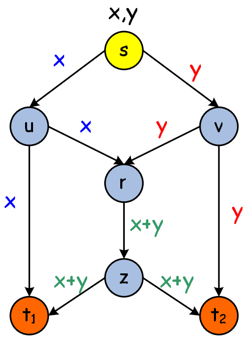


Figure: The butterfly network

- Assume the existence of a wiretapper
  - who can intercept the packets on any single edge of his choice
- How can we send data securely to both destinations?
  - i.e. without letting the wiretapper gain any information about the data

Secure Butterfly

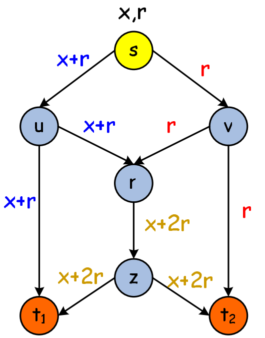


Figure: Secure Solution

- $x, r \in GF(3)$
- r randomly generated
- One symbol (x) is sent securely to t_1 and t_2
- The data is "hidden" from the wiretapper
- Note the need to use a field with a size larger than 2

The Wiretap Network Model

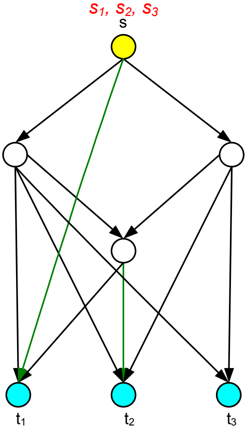


Figure: multicast network $(h, k) = (3, 3)$.

- Network represented by an acyclic directed graph $G(V, E)$; $|E| = N$
- Source s where h packets $S = (s_1, \dots, s_h)^T \in GF^h(q)$ are available
- k destination nodes $t_1, \dots, t_k$ that demand all the packets
- Min-cut between s and any t_i is n

Model - The Wiretapper

The wiretapper

- Can access the packets carried by any μ<n edges of his choice
- Knows the network code used
- Knows any keys shared between the s and the t_i's. (Key Cryptography is not possible here)

Problem Definition

Definition (Secure Multicast Problem)

Given a wiretap network, find, if possible, a linear network code that will

- deliver h packets to all the destinations
- achieve perfect secrecy by hiding all the information data from the wiretapper

Classical Security

Classical security

- Based on the conjectured hardness of some known problems such as *factoring* large integers
- Assume a computationally bounded adversary

Categories

- Symmetric Key Cryptography: DES, AES, hash functions
- Public Key Cryptography: RSA algorithm

Information-Theoretic Security

Information-theoretical security

- Based on concepts from information theory such as *entropy*
- No assumption on the strength of the adversary

Shannon, C. E., "A Communication Theory of Secrecy Systems," 1949, Bell Labs

Network Security

- Let W the set of μ edges observed by the wiretapper
- $Z_w = (z_1, \ldots, z_\mu)$ the observed packets
- $H(S|Z_w)$= how much information we are hiding from the wiretapper when he observes the edges in W
- $H(S)$= how much information we are sending to the destinations

Definition (Security Condition)

We say that the network is secure iff

$$H(S|Z_w) = H(S) \quad \forall W \subset E$$

Network Security

- Let W the set of μ edges observed by the wiretapper
- $Z_w = (z_1, \ldots, z_\mu)$ the observed packets
- $H(S|Z_w)$= how much information we are hiding from the wiretapper when he observes the edges in W
- $H(S)$= how much information we are sending to the destinations

Definition (Security Condition)

We say that the network is secure iff

$$H(S|Z_w) = H(S) \quad \forall W \subset E$$

Network Security

- Let W the set of μ edges observed by the wiretapper
- $Z_w = (z_1, \ldots, z_\mu)$ the observed packets
- $H(S|Z_w)$= how much information we are hiding from the wiretapper when he observes the edges in W
- $H(S)$= how much information we are sending to the destinations

Definition (Security Condition)

We say that the network is secure iff

$$H(S|Z_w) = H(S) \quad \forall W \subset E$$

Network Security

- Let W the set of μ edges observed by the wiretapper
- $Z_w = (z_1, \ldots, z_\mu)$ the observed packets
- $H(S|Z_w)$= how much information we are hiding from the wiretapper when he observes the edges in W
- $H(S)$= how much information we are sending to the destinations

Definition (Security Condition)

We say that the network is secure iff

$$H(S|Z_w) = H(S) \quad \forall W \subset E$$

Related Work (1)

- Cai & Yeung were first to define this problem (2002).
- They constructed multicast codes that can securely send $h = n - \mu$ packets ($S = (s_1, \ldots, s_h)^T$) to all the destinations
  1. multicast informatitonpackets ($S = (s_1, \ldots, s_h)^T$) to all the destinations
  2. at at rate $= \frac{h}{n}$ (i.e. by adding $\mu = n - h$ redundant packets)
  3. achieve perfect secrecy

Secure Multicast Scheme

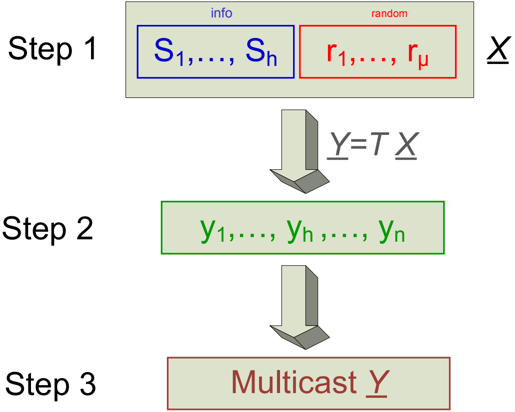


Figure: Secure Solution

Related Work

Theorem (1)

Let $Y = TX$. The resulting code is secure iff any set of vectors consisting of

1. *at most μ l.i. global encoding vectors*
2. *vectors from the <u>first h rows of T^{-1}</u>*

is linearly independent. [Feldman et al. 2004, Cai&Yeung 2002]

Related Work (2)

Theorem (2)

Secure linear network codes $(h = n - \mu)$ *exist over* $GF(q), \forall q > \binom{N}{\mu}$

- *N is the number of edges in the Network*
- μ *is the number of edges that the wiretapper can observe*

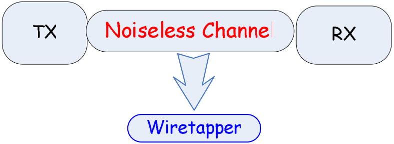
[Cai&Yeung 2002]

Contributions

- We look at the wiretap multicast network as an *generaliztion* of the wiretap channel of type II (WTCII) [Ozarow & Wyner 1984]
- We study a secure multicast scheme using coset codes, originally proposed for the WTCII
- We show that this scheme is *equivalent* to the previously described one. Nevertheless it allows to recover very easily Theorem 1 (and therefore Theorem 2)
- We also improve on the bound on the field size, by showing that secure multicast codes exist over $GF(q)$, $\forall q \geq \binom{h^3 k^2 + \delta}{\mu - 1} + k$; δ is the source node degree

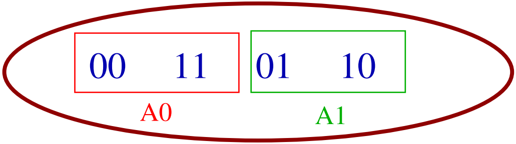
Outline

WTCII- Ozarow & Wyner 84



- A transmitter can send n symbols $X = (x_1, \ldots, x_n)$ to a receiver through a noiseless channel
- A wiretapper can observe $Z_w = (z_1, \ldots, z_\mu)$, a subset W of his choice of size $\mu < n$ of the transmitted symbols
- The transmitter wants to communicate h symbols $S = (s_1, \ldots, s_h)$ securely to the receiver
- The problem is to find an encoding of S into X that maximizes $\Delta = \min_W H(S/Z_w)$

Example of a Code for the WTCII



Set of all possible transmitted symbols

- $n = 2; h = \mu = 1$
- Code
  - $s = 0 \longrightarrow A_0 = \{00, 11\}$
  - $s = 1 \longrightarrow A_1 = \{01, 10\}$
- If the input to the encoder is s, then the output is a random element of A_s
- It can be shown that this code achieves perfect secrecy, i.e. $\Delta = 1$

Codes for the WTCII

Figure: Partition Σ^n

1. Partition
2. Toss a die
3. Send

Coset Codes

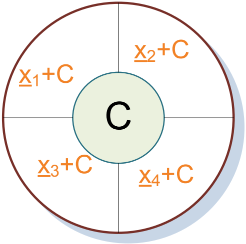
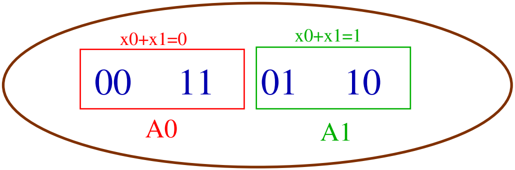


Figure: Partition by cosets $GF(q)^n/C$

- $\Sigma = GF(q)$
- Let $\mathbb{C}$ be $(n, n-h)$ q-ary code (i.e. a linear subspace of dim $n-h$)
- The partitions A_i's are the cosets of $\mathbb{C}$ $(x+\mathbb{C})$
- Let H be the $n \times h$ parity check matrix of $\mathbb{C}$
- Choose the subset A_S associated with $S \in GF^h(q)$ to be

$$A_S = \{X \in GF^n(q); HX = S\}$$

Example Revisited



Set of all possible transmitted symbols

The previous code is a coset code where $H = [1 \quad 1]$

Performance of Coset Codes

Theorem (Ozarow& Wyner)

Consider a coset code of parity check matrix H of columns h_i ($1 \leq i \leq n$). Let $W \subseteq \{1, 2, \ldots, n\}$. Then,

$$\Delta = \min_{|W|=n-\mu} rank\{h_j; j \in W\}$$

Coset Codes for the Wiretap Network

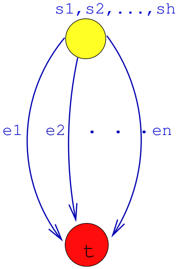


Figure: Wiretap network equivalent to the WTCII

- WTCII can be regarded as an instance of the wiretap network
- Use coset codes to achieve security for the wiretap multicast network:
  - Encode $S = (s_1, \ldots, s_h)^T$ into $Y = (y_1, \ldots, y_n)^T$ using a coset code of parity check matrix H
  - Use a network code to multicast Y to all the destinations

Questions

1. How is the coset code scheme different than the one proposed by Cai & Yeung?
2. How should the parity check matrix H be chosen to ensure the security condition?

Equivalence

C&Y secure network code ($\mu = n - h$)

1. Randomly generate $R = (r_1, \ldots, r_{n-h})^T$
2. $Y = T * (s_1, \ldots, s_h, r_1, \ldots, r_{n-h})^T$
3. Use a network code to multicast $Y = (y_1, \ldots, y_n)^T$

Lemma

C&Y secure network code is equivalent to a coset code of parity check matrix $H = T^*$, where T^* is formed by taking the first h rows of T^{-1}

The Parity Check Matrix

- Let $W \subset E$ be the set of the $\mu = n - h$ wiretapped edges
- Z_w the packets observed on these edges
- WLOG, the global encoding vectors of these edges are linearly independent
- C_w an $(n-h) \times n$ matrix formed by these vectors

Theorem (III)

A coset code with a parity check matrix H satisfies the security condition iff $M_w = \begin{bmatrix} H \\ C_w \end{bmatrix}$ *is invertible* $\forall W$

Theorem 3 is equivalent to Theorem 1.

The Parity Check Matrix

- Let $W \subset E$ be the set of the $\mu = n - h$ wiretapped edges
- Z_w the packets observed on these edges
- WLOG, the global encoding vectors of these edges are linearly independent
- C_w an $(n-h) \times n$ matrix formed by these vectors

Theorem (III)

A coset code with a parity check matrix H satisfies the security condition iff $M_w = \begin{bmatrix} H \\ C_w \end{bmatrix}$ *is invertible* $\forall W$

Theorem 3 is equivalent to Theorem 1.

Simple Proof

- Remember
  - Data vector $S = (s_1, \dots, s_h)$
  - Transmitted vector $Y = (y_1, \dots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \dots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n - h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Simple Proof

- Remember
  - Data vector $S = (s_1, \ldots, s_h)$
  - Transmitted vector $Y = (y_1, \ldots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \ldots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n - h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Simple Proof

- Remember
  - Data vector $S = (s_1, \ldots, s_h)$
  - Transmitted vector $Y = (y_1, \ldots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \ldots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n - h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Simple Proof

- Remember
  - Data vector $S = (s_1, \ldots, s_h)$
  - Transmitted vector $Y = (y_1, \ldots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \ldots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n-h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Simple Proof

- Remember
  - Data vector $S = (s_1, \ldots, s_h)$
  - Transmitted vector $Y = (y_1, \ldots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \ldots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n - h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Simple Proof

- Remember
  - Data vector $S = (s_1, \ldots, s_h)$
  - Transmitted vector $Y = (y_1, \ldots, y_n)$ chosen randomly among the solutions of $HY = S$
  - $Z_w = (z_1, \ldots, z_{n-h})$ wiretapped data; $Z_w = C_w Y$
- $H(Y|S, Z_w) = H(S|Y, Z_w) + H(Y|Z_w) - H(S|Z_w)$
- $H(S|Y, Z_w) = 0$ cosets are disjoint
- $H(S|Z_w) = H(S) = k$ security condition
- $H(Y|Z_w) = n - rank(C_w) = n - (n - h) = h$
- $\Rightarrow H(Y|S, Z_w) = 0$ and the system $HY = S$ & $C_w Y = Z_w$ has a unique solution

Outline

A Different Approach

- Previous approach
  - Use a secure code on top of an already designed network code to achieve security
- New approach
  - Incorporate the security condition in the algorithm that constructs the network code
- We gain better bound on the field size

Jaggi's Algorithm

Sketch of Jaggi's et al. algorithm for finding (not necessarily secure) network codes

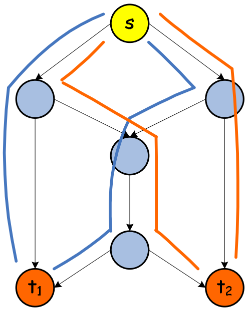


Figure: Flows

1. Find k (# of destinations) flows $(F_1, \ldots, F_k)$ each of h disjoint paths
2. Visit the network edges in topological order
3. Let B_{F_i} an $h \times h$ matrix formed by the h global encoding vectors of the last processed edges of flow F_i
4. Find the encoding vector for the currently visited edge such that all the k matrices B_{F_i} are invertible

Bound on the Field Size

Theorem

There is always a linear multicast network code over $GF(q), \forall q \geq k$ [Jaggi et al. 2003]

- The proof is by construction and follows from the correctness of Jaggi's algorithm
- A field of size $q \geq k$ is sufficient for keeping the k matrices B_{F_i} invertible

Theorem (I)

Secure linear network codes exist over $GF(q), \forall q > \binom{N}{\mu}$. [Cai&Yeung 2002]

Bound on the Field Size

Theorem

There is always a linear multicast network code over $GF(q), \forall q \geq k$ [Jaggi et al. 2003]

- The proof is by construction and follows from the correctness of Jaggi's algorithm
- A field of size $q \geq k$ is sufficient for keeping the k matrices B_{F_i} invertible

Theorem (I)

Secure linear network codes exist over $GF(q), \forall q > \binom{N}{\mu}$. [Cai&Yeung 2002]

Special Case

- Assume the wiretapper can intercept only one edge ($\mu = 1$)
- Coset code of parity check matrix H is to be used
- We modify Jaggi's algorithm so it outputs a secure network code in the following way
  - When looping over the edges of the network, choose the global encoding vector $U(e)$ such that
    1. The k matrices B_{F_i} invertible
    2. The matrix $\begin{bmatrix} H \\ U(e) \end{bmatrix}$ is invertible (Theorem 3)
- $k+1$ constraints $\Rightarrow$ A field of size $q \geq k+1$ is sufficient
- Compare to the $\binom{N}{1} = N$ bound (N is the number of edges in the network)

Bound for the General Case

Theorem

Secure linear network codes for a wiretap multicast network exist over $GF(q)$, $\forall q \geq \binom{N-1}{\mu-1} + k$.

Theorem

Secure linear network codes for a wiretap multicast network exist over $GF(q)$, $\forall q \geq \binom{h^3k^2+\delta}{\mu-1} + k$; δ is the source node degree

Bound for the General Case

Theorem

Secure linear network codes for a wiretap multicast network exist over $GF(q)$, $\forall q \geq \binom{N-1}{\mu-1} + k$.

Theorem

Secure linear network codes for a wiretap multicast network exist over $GF(q)$, $\forall q \geq \binom{h^3k^2+\delta}{\mu-1} + k$; δ *is the source node degree*

Outline

Summary

- We considered the problem of designing network code that will guaranty security in a network with multicast demands
- Building on an analogy with the wiretap channel of type II, we proposed using coset codes for the multicast channel
- We showed that coset codes are equivalent to other codes already studied in literature. Nevertheless, they permit an easy recovery of some important results.
- We also gave an improved lower bound on the field size sufficient for the existence of secure network code