Interactive Single-Group Secure Network Coding

Shao-Heng Ko

Department of Electrical Engineering

National Taiwan University

b00901169@ntu.edu.tw

Abstract— **This report introduces the idea of key recycling and key distribution in single-group two-way and multi-way wiretap networks, which increase the throughput without loss of security in such networks.**

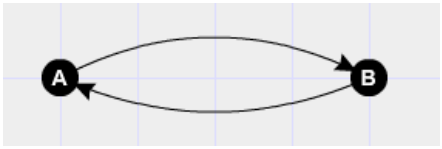
I. Introduction

A network is represented by a directed graph $G=(V,E)$ with nodes correspond to communication units and edges correspond to channels. We focus on security issues on such model, and thus all channels are assumed to be noiseless. Also, it is reasonable to time-divide any channel and partition a channel into channels with fractional capacity; hence we assume that all channels have unit capacity.

A wiretap network introduced in [1] is a network whose channels are vulnerable to mischievous eavesdroppers. In such a network, an eavesdropper may attack any possible edge sets; the sender need to randomize the message in order to protect it to assure that the eavesdropper gains no information about the sensitive messages after wiretapping. The model introduced in [1] has 2 assumptions: (1) all channels are vulnerable to be eavesdropped, (2) all the random secret keys are generated at source.

In [1], under a single-source, single-destination acyclic wiretap network, the possible throughput from the source to the destination through secure network coding is found. However, the network in practice is seldom acyclic. Taking into consideration the interaction of user nodes, it may be able to make better use of the secure network coding tool and further improve the performance of a cyclic network.

The idea of key-recycling and key-distribution is rather simple. In a wiretap network, the network resource is divided, conceptually, into those we need to deliver the necessary randomness (secure keys), and the rest to actually deliver the sensitive information. From the setting of a wiretap network, we know that adding randomness is inevitable; but once we can make the randomness function more efficiently, we then reduce the resource needed to generate and deliver them, and finally make more resource available for the data.



To make a simple example for key-recycling, consider a wiretap network with only two user nodes A and B and two single unit channel between them, with the opposite direction. Suppose a wiretapper can compromise at most one channel a time. With the view of secure network coding in [1], we may see the model as two disjoint acyclic network, but we will also find that no possible secure information flow can happen between the two nodes this way. However, there is a quite naïve method to improve from that: in time slot T, a secure key K_T is generated at A and passed from A to B. A message packet M_T is encoded by $M'_T = M_T + K_{T-1}$ and is passed from B to A. By utilizing the interaction between the two nodes, we obtain a better result.

In a multi-way network, there are more than two user nodes. When we are desired to recycle the keys and using them for traffic among the nodes, we will encounter the need to pass the keys around the network. Then there comes the idea of key-distribution, which is indeed a distribution problem itself.

In this work, an improved throughput region for a single-group, two-way wiretap network is found explicitly, with a conceptual method to implement the work. Some partial result for a single-group, multi-way network is also found, with a proposal of a protocol that can be used by the user nodes to address the problem.

II. Interactive Secure Network Coding on Single-Group, Two-Way Network

A. Model

A single-group two-way wiretap network consists of the following components:

1) *Directed multigraph* $G=(V,E)$, where V and E are respectively the node set and the edge set of G.
2) *User node* $A,B \subseteq V$, where the random messages are generated.
3) *Compromised channels sets* C : $\forall c \in C, c \subseteq E$, where c is a set of compromised channels that a wiretapper can listen to. It is not known which c is the particular one.
4) *Secrecy constraint* R : $\forall c \in C, |c| \le R$.

When the secrecy constraint is set to be R, a wiretapper gathers information from at most R channels. In the following text, every model is of secrecy constraint R, and C consists of merely every subset of size R of E. This equals to that the wiretapper simply chooses R channels to access.

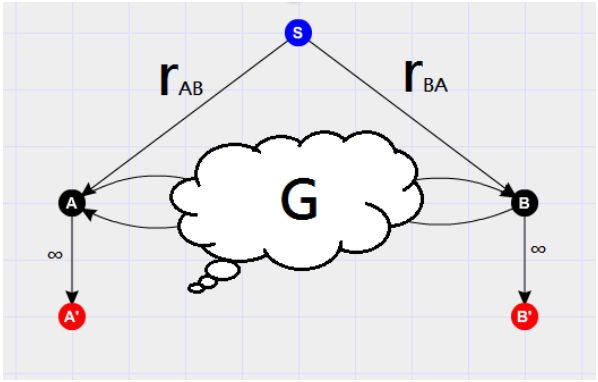
B. Equivalent model

In this section, we conceptually claim that every arbitrary network in this model is equivalent to a much simpler network with

$V = \{A, B\}$

$E = \{E_{AB}i, E_{BA}j \mid i = 1,2,...c(A;B),\ j = 1,2,...c(B;A)\}$.

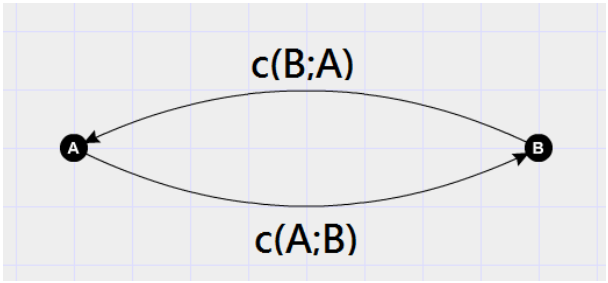
We start from an arbitrary single-group, two-way network, setting $R = 0$, thereby ignoring the security issue.



Without altering the already existing network structure, we add a virtual source S and two virtual destinations A', B'. Connect $(S, A), (S, B), (A, A'), (B, B')$, all with infinite capacity. This qualifies as a single-source multicast problem. Suppose when the multicast reaches its maximum, the traffic between S and A is r_{AB}, and that between S and B is r_{BA}. We can say that r_{AB} is the maximum flow size from A to B. If it is not, S can multicast more information through A, since the capacity of (S, A) and (A, A') are infinity. Likewise, r_{BA} is the maximum flow size from B to A.

Now, from the theorem of single-source multicast problem in [2], we obtain the following:

$$r_{AB} \le \min\{c(A;A'), c(A;B')\} = c(A;B)$$
$$r_{BA} \le \min\{c(B;A'), c(B;B')\} = c(B;A)$$
$$r_{AB} + r_{BA} \le c(S;A',B') = c(A,B;A',B') = \infty$$



A simple network like this satisfies the result. With network coding, we can achieve everything as in this simple network, in the case $R = 0$. Also, from [1], when it comes to secure coding, it suffices to precode the message symbols with random keys before the phase of network coding; that is, a wiretap network with secrecy constraint r can be viewed as a coding network with r default message to send.

Thus, any arbitrary network in our model is "equivalent" to the simple network with only direct links with capacity of those Min-cuts in the original network.

C. *Addressing security issues*

Given a secrecy constraint r, we are interested in the probable region of r_{AB} and r_{BA}, which stand for the secure throughput in each direction, respectively. We are also interested in the range of the sum, $r_{total} = r_{AB} + r_{BA}$.

It is clear that

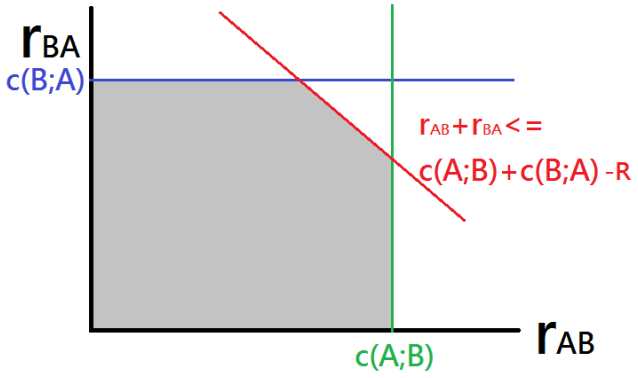
$$r_{AB} \le c(A;B)$$
$$r_{BA} \le c(B;A)$$

From the view of information amount, the mutual information of the wiretapper to the message base is 0. This suggests that the dimension of secret key base is at least R. Since all keys in this network need to flow between the user nodes, this leads to

$$r_{AB} + r_{BA} \le (c(A;B) + c(B;A) - R)^+ .$$

If $R \ge c(A;B) + c(B;A)$, the right side of the equation shrinks to 0, and we have a totally useless network. We shall ignore the case and write simply

$$r_{AB} + r_{BA} \le c(A;B) + c(B;A) - R$$



The area in grey in the figure is the probable throughput for our model. Observe two points on the boundary, $(c(A;B), (c(B;A) - R)^+)$, $((c(A;B) - R)^+, c(B;A))$; if we can achieve these two particular set of throughputs, we will be able to achieve anywhere in the probable area, by utilizing the time fraction method.

Due to the symmetry of our simplified network, we can assume without loss of generality $c(A;B) \ge c(B;A)$. Three cases follows.

(i) $c(A;B) \ge c(B;A) \ge R$

This is the easiest case to address. In any given time slot, generating R secret keys at A, B respectively and letting them flow through, we achieve those two throughput points. For an example, user node A generates secret keys K_{t1}

to K_{tR} at time slot t, and encode messages m_{ABt1} to m_{ABti} with $k_{(t-1)1}$ to $k_{(t-1)R}$, $i = c(A;B) - R$. Those keys and encoded messages are network coded and flow to B. In the meantime, B encodes messages m_{BAt1} to m_{BAtj}, $j = c(\mathrm{B};A)$ with the same set of secret keys, which have already arrived at B one time slot before! This achieves the throughput set $((c(A;B)-R)^+, c(B;A))$.

(ii) $c(A;B) \geq R \geq c(B;A)$

(iii) $R \geq c(A;B) \geq c(B;A)$

The condition in these two other cases are somewhat similar, with just a little alteration. When R is big enough, the secret keys cannot be entirely generated by one node. The point here is, no matter who generates the keys, they flow through the network, and are actually put into use in the next time slot. We will see later that these cases are actually special cases of the protocol to address single-group multi-way interactive secure network coding problem.

After discussion, we conclude that the two specific throughput sets are really achievable, therefore any point in the grey area is feasible. The region is bounded by

$$r_{AB} \leq c(A;B)$$
$$r_{BA} \leq c(B;A) \quad .$$
$$r_{AB} + r_{BA} \leq c(A;B) + c(B;A) - R$$

D. Stronger setting: time-variant compromised channels

A slightly different setting is to change c into $c(t)$; that is, the compromised channel set may change from time to time. Apply our previous method on this setting, one possible scheme is that a wiretapper happens to get a set of secret keys at time slot k and those messages encrypted by them at time slot $k+1$. The wiretapper is suddenly able to decrypt the messages! We will need to slightly alter our protocol, adding pieces of secret keys which I call "shared secrets".

To illustrate the idea, suppose that the user nodes have some previous shared secret information s_1 to s_R. This is R dimension of secrets. In every time slot, the user nodes use these secrets in addition to the secret keys received to encrypt the messages. This results in R more dimensions in the encrypted messages then the wiretapper can decode each time. The wiretapper can never decode s_1 to s_R fully, therefore impossible to obtain information of the messages.

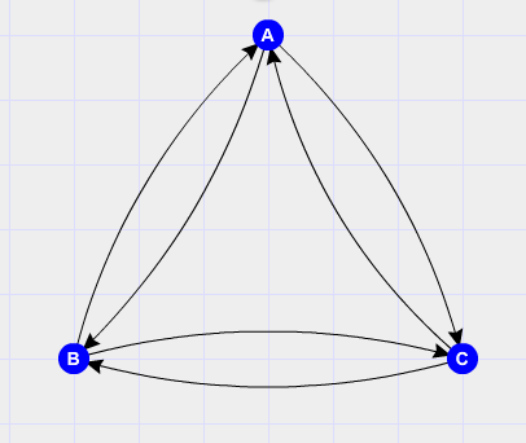
Such secrets can be previously shared, say, in the same network but setting the secrecy constraint to $2R$. Once we assure the shared secrets to be secure, the interaction coding phase may begin. The overhead to deal with the shared secrets will be small enough to not change the ideal throughput rate.

III. Interactive Secure Network Coding on Single-Group, Multi-Way Network

A single-group multi-way network is slightly different from the single-group two-way network on user nodes. In here there are more than 2 user nodes, $A_1, A_2, \ldots A_k$. It is also assumed that the network is strongly connected.

We cannot directly apply the equivalent model here, since we cannot apply the multicast theorem to a multi-way network. The network structure now has some effect on the throughput itself, with issues like bottlenecks may step into the picture.

A. 3-way complete unit-capacity network



A complete unit-capacity network is a multi-way network that has a direct unit-capacity channel from A_i to A_j for $i, j \in 1,2,\ldots k$. Since such a network is symmetric, we are interested in (1) the maximum throughput r_{Max} between arbitrary A_i to A_j, and (2) the maximum total throughput r_{total} among all possible communication pairs.

We would still like to utilize key-recycling. Suppose $R = 1$. Say a key K is first generated by A and delivered to B. (If a key is not delivered at all, it is useless.) Will it be wise to utilize K at C? If no secret key is used among all 3 user nodes, this network will reduce to 3 disjoint 2-way networks. Using the results before, we will have $r_{Max} = 2$ and $r_{total} = 3$.

Consider what happens if we use K among all 3 nodes. It is obvious that K would travel at least 2 times to arrive at all nodes, although at different time. We say the total cost to spread K is 2. Notice that K can be generated anywhere in the network and follows any path to traverse the network. In here, we notice that K traverses a spanning tree of the network.

Once K is delivered, it is to be used to encrypt messages. It turns out that one such K each time slot is enough for the entire network. One example is like below.

At time slot t, node A generates K_t and pass it to node B. Node B pass $K_{(t-1)}$ to C, which it received one slot before. Messages from B to A is encrypted with $K_{(t-1)}$, the

newest common acknowledgements between them. Messages from A to C, C to A or C to B uses $K_{(t-2)}$, the newest secret key that C knows in this moment. $r_{Max}=2$ and $r_{total}=4$, which is an improvement. Since the min-cut between any two user nodes is 2, there is no room for improvement on r_{Max}, but we saved one channel's capacity in total by recycling (and thus distributing) the keys.

B. Distributing Cost and the Newest Acknowledged Key Protocol

We shall limit our discussion here to a particular type of multi-way networks. Suppose user nodes are A_1 to A_k, and there are a sub-network between any two of them, all disjoint. This equals to say, if some node V is on a path from A_i to A_j, either (1) there are other user nodes on this path, or (2) if V is on another path from $A_{i'}$ to $A_{j'}$ which are different from A_i to A_j, there are other user nodes on that path.

Suppose secrecy constraint R, we introduce the distribution cost $C_{dist}=|T|R$, where $|T|$ is the cardinality of a spanning tree. It follows that $C_{dist}=(N-1)R$. This cost is the average throughput used to broadcast R secret keys per time slot to the whole network, and is the cost for key-distribution in the Newest Acknowledged Key Protocol.

In this protocol, a series (or consequence) of keys are generated anywhere in the network, and is broadcast through the network by a (directed) spanning tree. Keys come in bunches of R, and will arrive at any node in at most $N-1$ time slots later. Any traffic between two user nodes is encrypted by the newest shared secret key(s) between the two nodes. If node A_i has secret keys up to K_p, node A_j has secret keys up to K_q, they will use keys up to $K_{\min(p,q)}$ to communicate. It follows that

$$r_{total}=c_{total}-c_{dist}=c_{total}-(N-1)R.$$

Note that on a k-way complete unit-capacity network with $R=1$, $r_{total}=c_{total}-c_{dist}=2C_2^k-(k-1)=(k-1)^2$.

IV. Conclusions

This report concludes the behavior of single-group, two-way interactive wiretap network, and takes a briefly look at the multi-way case. Also, the concepts of key-distribution is introduced. It then follows that a two-way network is a special case of key-distribution network, with an additional property on network coding, which leads to hugely simplified results.

It is shown that the interactive wiretap network problem can be separated into the key-distribution problem and the secure network coding problem. Once the keys are distributed, the problem reduced to using known keys to encrypt the messages, and passing them through network coding. The major improvement lies in the key-distribution phase.

Acknowledgment

Thanks to Prof. I-Hsiang Wang for instruction and motivation again. My work this semester seems to be more systematic and in good order, thanks to your directions.

References

[1] N. Cai and R. W. Yeung, "Secure Network Coding on a Wiretap Network", *IEEE Transactions on Information Theory*, vol. 57, No. 1, Jan. 2011.

[2] S. R. Li and R. W. Yeung, "Linear Network Coding", *IEEE Transactions on Information Theory*, vol. 49, No. 2, Feb. 2003.