Secure Network Coding with Erasures and Feedback

László Czap
EPFL, Switzerland
laszlo.czap@epfl.ch

Christina Fragouli
EPFL, Switzerland
UCLA, USA
christina.fragouli@epfl.ch

Vinod M. Prabhakaran
TIFR, India
vinodmp@tifr.res.in

Suhas Diggavi
UCLA, USA
suhas@ee.ucla.edu

***Abstract*— Secure network coding assumes that the underlying network links are lossless, thus it can be applied over lossy networks after channel error correction. Yet it is well known that channel losses, such as packet erasures, can be constructively used for secrecy over a link. We address here the challenge of extending these results for arbitrary networks. We provide achievability schemes over erasure networks with feedback, that outperform the alternative approach of channel error correction followed by secure message transmission separation. We derive outer bounds on the securely achievable rate and as a consequence we show optimality of our proposed scheme in some special cases.**

I. Introduction

Secure network coding assumes that the underlying network links are lossless; thus if a wiretapper, Eve, observes a link, she gets access to all the information that flows through it. Since most practical channels introduce errors, underlying the lossless assumption is an implicit "channel coding" followed by "security coding" separation: if our links introduce errors, we need first apply a channel code to correct them, and then build security on top of the resulting lossless networks. But as a result, we can convey zero rate securely through the links that Eve observes.

It is well known that channel losses, such as erasures, can be constructively used to enable non-zero secrecy rate over a link. Assume for example that Eve observes a node's transmission independently and with a larger erasure probability than the legitimate next hop node; then, by applying a wiretap code [1] we can convey through this link a message at a nonzero rate. Moreover, if we allow channel state feedback, i.e., the next hop node to acknowledge packet reception as is the case in most network protocols today, we can convey a nonzero rate even if the eavesdropper has a better channel than the legitimate receiver [2].

The challenge is how to extend these results from single links to arbitrary networks. For a single link, we have an exact characterization of the secret message capacity with feedback, yet as soon as we go to a network with more than one hop and multiple nodes, the complexity of the problem increases exponentially, as there exists an exponential number of subsets of nodes that can generate randomness, create secret keys, and cooperate for secrecy. Finding secrecy capacity of a general network is as hard as determining the capacity region of multiple unicast network coding, which is a long-standing open problem [3], [4].

This paper provides achievability schemes over arbitrary erasure networks with link-by-link state feedback. We start from the simplest case, independent erasure networks with the same erasure probability δ in each link. We assume that there exist h edge-disjoint paths from the source to each receiver, and that our eavesdropper, Eve, observes any z links in the network. When eavesdropping transmissions on a link that connects node u to node v, Eve also receives the transmissions of node u with erasure probability δ_E, independently from node v. We also make the assumption that only the source node can generate randomness; this assumption clearly reduces the rates we can achieve, but at the same time simplifies the problem, and fits well with the current networking philosophy of having the intelligence at the edge of the network and keeping intermediate node operations simple.

Our proposed scheme can achieve secrecy rates that consistently outperform a separate channel error correction and secure message transmission approach. Our scheme applies a separation of two phases, where at a first stage we generate shared randomness (key) between the source and the receivers, we use the key for encryption and at the second phase we reliably send the encrypted message. This approach is known to achieve secrecy capacity over a single link [2].

We first consider lossless networks and establish connection between the secure network coding scheme [5], [6] and the two phase approach. We propose a modified, two phase secure network coding scheme showing that there is no fundamental difference between these approaches.

We then examine how we can take advantage of erasures and feedback in each of the two phases to achieve higher secret rates. Over a lossy network the advantage of separating the two phases and using feedback becomes clear.

Finally, we give outer bounds on the securely achievable rate and prove optimality of our scheme in some special cases.

A. Related Work

Secure network coding by Cai and Yeung [5], [6], [7] is a seminal work in the field of secret communication over networks. The secrecy capacity of an error-free network is established and linear achievability schemes are proposed.

This work was supported by the ERC Starting Grant Project NOWIRE ERC-2009-StG-240317. V. M. Prabhakaran's research was supported in part by a Ramanujan Fellowship from the Department of Science and Technology, Government of India. The work of S. Diggavi was supported in part by NSF award 1136174 and MURI award AFOSR FA9550-09-064.

This work was followed by a number of alternative constructions and extensions [8], [9], [10], yet as far as we know ours is the first work that looks at network coding secrecy over erasure networks with feedback.

Secret communication over a noisy channel was investigated by Wyner [1] and the result was extended for networks by Cui [11]. Using a similar approach capacity results for broadcast erasure networks were derived [12]. However, none of these work take advantage of feedback and thus they offer any nonzero secrecy rate over a wiretapped channel only if the eavesdropper's observation is more noisy than the legitimate receiver's. By exploiting a limited rate feedback significantly higher rates are achievable [13], [14], [2]. These results were extended for a broadcast setting with multiple receivers [15], but the generalization for a multihop network is a new and challenging problem that we start investigating here.

II. Network Model and Background

A. Network model

Communication takes place over a network which is represented by a directed acyclic multigraph $\mathcal{G}(V, E)$, where V is the set of network nodes and E is the (multi-)set of edges. An eavesdropper Eve can select arbitrarily up to z edges of the network to wiretap. $A \subseteq E$ denotes the subset of wiretapped edges, where $|A| \leq z$. We assume that z is known as a design parameter, but A is known only by the eavesdropper.

Every link $e = (u, v) \in E$ is an erasure channel with parameters δ, δ_E. The input alphabet of the channel is $\mathbb{F}_q^L$, length L vectors of a finite field $\mathbb{F}_q$. We often call such vectors packets. The network node v receives transmissions on e with an erasure probability δ, while in case $\delta_E \in A$, Eve receives packets sent over e with an erasure probability δ_E. All erasures are assumed to be independent. For a given channel $e = (u, v)$ let X_e denote the channel input while Y_e is the receptions of v and Z_e is the potential receptions of Eve. Then,

$$\Pr\{Y_e = X_e | X_e\} = 1 - \delta, \quad \Pr\{Y_e = \perp | X_e\} = \delta$$
$$\Pr\{Z_e = X_e | X_e\} = 1 - \delta_E, \quad \Pr\{Y_e = \perp | X_e\} = \delta_E$$
$$\Pr\{Y_e, Z_e | X_e\} = \Pr\{Y_e | X_e\} \Pr\{Z_e | X_e\},$$

where $\perp$ is the symbol of erasure.

After every transmission over a link (u, v) node v acknowledges its receptions (whether it received correctly or an erasure happened). The acknowledgments are available to all the nodes causally. We also assume that all the acknowledgments are public to the eavesdropper regardless of whether or not the given link is wiretapped.

A source node $s \in V$ has a message $W \in \mathbb{F}_q^{LN}$ to send securely to a set of destination nodes $D \subset V$. Source s can further generate independent randomness Θ. We assume that Θ is uniformly distributed and it can be generated without rate constraint. We will treat both W and Θ as a row vector of packets. The length of W is N, while the length of Θ, i.e. the amount of randomness needed is a property of the communication scheme.

The multicast capacity of $\mathcal{G}$ with source s and destination nodes D is $h(1 - \delta)$, where h denotes the number of edges in the smallest value min-cut between s and any $d \in D$. We introduce parameters $h = t + \ell$, where t is the number of multihop paths between s and d while ℓ is the number of direct s-d links in the smallest value min-cut.

1) Notation: The set of incoming and outgoing edges of $v \in V$ are denoted by I_v and O_v.

If $\mathcal{E} \subseteq E$ then $Y_{i,\mathcal{E}}$ denotes the set of received packets by the network nodes in the ith time slot on the set of edges $\mathcal{E}$. Similarly for $\mathcal{V} \subseteq V$ the notation $Y_{i,\mathcal{V}}$ denotes the set of packets that the set of nodes $\mathcal{V}$ receives in the ith time slot. In case there are parallel edges the notation (u, v) means the set of edges starting from u and ending at v.

We use $Y_{\mathcal{E}}^i$ and $Y_{\mathcal{V}}^i$ as a shorthand for $Y_{1,\mathcal{E}} \dots Y_{i,\mathcal{E}}$ and $Y_{1,\mathcal{V}} \dots Y_{i,\mathcal{V}}$. We apply the same notation for other vectors also. E.g. the source node transmits $X_{i,s}$ in the ith time slot, while Eve's observation after n time slots is Z_A^n.

F_i denotes the acknowledgments of the ith time slot.

To simplify notation, we express entropy and rate in terms of packets. This allows us to omit the constant factor $L \log q$, which is the size of one packet.

B. Security and rate

Definition 1: A secure scheme that uses every channel $e \in E$ of the network n times has parameters (n, ϵ, N). For all $1 \leq i \leq n$ it defines encoding maps $\phi_{i,e}$ for all $e \in E$:

$$X_{i,(s,v)} = \phi_{i,(s,v)}(W, \Theta, F^{i-1})$$
$$X_{i,(u,v)} = \phi_{i,(u,v)}(Y_{I_u}^{i-1}, F^{i-1}), \ \forall u \neq s.$$

For all $d \in D$ it defines decoding maps ψ_d such that

$$\Pr\left\{\psi_d(Y_{I_d}^n) \neq W\right\} < \epsilon.$$

Further, Eve learns negligible information about W:

$$I(W; Z_A^n, F^n) < \epsilon. \tag{1}$$

In all our cases we consider linear schemes and hence linear encoding and decoding maps. It follows that $X_{i,e}$ can be written in the following form:

$$X_{i,e} = [\Theta \quad W] \begin{bmatrix} f_{i,e}^{\Theta T} \\ f_{i,e}^{WT} \end{bmatrix}$$

where $f_{i,e} = [f_{i,e}^{\Theta} \ f_{i,e}^{W}]$ is the global encoding vector of edge e in the ith time slot. Here we explicitly separate in notation the coefficients of packets from Θ and packets from W.

Definition 2: A secure communication rate $\mathcal{R}$ is achievable over $\mathcal{G}$ if for any $\epsilon > 0$ there exists a secure scheme with parameters (n, ϵ, N) such that

$$\mathcal{R} - \epsilon < \frac{N}{n}.$$

We call the highest achievable secure communication rate the secure capacity of the network.

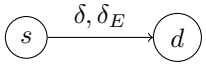


Fig. 1. Single link network

C. Secure network coding over lossless networks

In the special case of lossless channels $\delta = \delta_E = 0$ our model becomes the same as seen in [5]. Here we shortly summarize the work of Cai and Yeung [5], [6]. A linear coding scheme known as the secure network coding scheme is proposed with – in our notation – parameters $(1, \epsilon, (h-z)^+)$. The scheme uses source randomness of size z and ensures that all destination nodes receive both the message and the additional randomness. It is shown that the secure network coding scheme is optimal in terms of the achieved rate and it uses the minimum amount of additional randomness any optimal scheme might use.

Let us assume that Eve simply discards any packet that she receives more than once. We denote $z' \leq z$ the number of innovative packets Eve observes. Then, we can write Eve's observation in the following form:

$$Z_A = [\Theta \quad W] \begin{bmatrix} f_A^{\Theta T} \\ f_A^{WT} \end{bmatrix}.$$

Here $f_A^{\Theta T}$ is a $z \times z'$ matrix and f_A^{WT} is a $(h-z) \times z'$ matrix. The secure network coding scheme has the property that the matrix $f_A^{\Theta T}$ has rank z'. Thus, $\Theta f_A^{\Theta T}$ is a set of $z' \leq z$ independent uniform random packets, while $W f_A^{WT}$ is a set of z' linear combinations of the message packets, hence from Eve's perspective what she observes is some data $W f_A^{WT}$ encrypted using one time pad with key $\Theta f_A^{\Theta T}$.

One possible intuitive interpretation of these results is the following: to give perfect security against Eve who has access to at most z innovative packets, we need to send z packets of additional randomness and hence the secure capacity of the network is reduced by z compared to its multicast capacity. We use this intuition when we design our scheme for lossy networks.

D. Secure message sending over a single link

Consider the simplest possible network shown in Fig. 1 consisting of a single channel with parameters δ, δ_E, and $z = 1$.

1) Direct solution: One straightforward approach to deal with lossy networks is to use a forward error correction code at all the transmitting nodes and by this turn the noisy channels into error free channels. We can then treat the network a lossless network with links of capacity $(1-\delta)$ where we can apply the secure network coding scheme and achieve a rate $(h-z)(1-\delta)$. However, having a single wiretapped link we cannot achieve any secrecy this way. This approach is too pessimistic though, because it implicitly assumes that the eavesdropper does not experience any erasures after applying coding.

2) Exploiting Eve's erasures: As observed by Wyner [1] and applied for networks in [11], if $\delta_E > \delta$ despite of the error correction coding Eve still does not receive everything, which allows to achieve a secrecy rate $(\delta_E - \delta)^+$. Applying this result in a network, for $z \leq h$ a secrecy rate $(h-z)(1-\delta) + z(\delta_E - \delta)^+$ is achievable. In other words, as opposed to the direct solution a nonzero rate is potentially achieved also over the wiretapped edges [11].

3) Using feedback: By exploiting the acknowledging feedback we can do even better. Instead of a forward error correction code the source can send packets using an ARQ strategy, i.e. it repeats every packet until it is acknowledged. Using this strategy the next node experiences no erasures, while for Eve there is still a probability $\frac{\delta_E(1-\delta)}{1-\delta\delta_E}$ that she does not receive a certain packet. This situation is equivalent to a channel with capacity $(1-\delta)$ and parameters $\delta' = 0$, $\delta'_E = \frac{\delta_E(1-\delta)}{1-\delta\delta_E}$. Hence, we can apply Wyner's scheme on this logical channel and achieve a secrecy rate $(1-\delta)\frac{\delta_E(1-\delta)}{1-\delta\delta_E}$.

4) Separation of phases: None of the previous strategies achieves capacity in this setting. An optimal two-phase coding scheme was proposed in [2], which we summarize here.

The scheme has a key generation phase and an encrypted message sending phase. In the first phase s and d agrees securely in a shared key, which key is used for encryption in the second phase.

a) Key generation phase: Source s sends n_1 independent uniform random packets. For the moment, let us assume that d receives exactly $(1-\delta)n_1$ packets, while Eve does not receive $\delta_E(1-\delta)n_1$ out of these. Let M denote the $1 \times (1-\delta)n_1$ vector of the received packets. Let H be a $(1-\delta)n_1 \times \delta_E(1-\delta)n_1$ matrix, and let H be a parity check matrix of an MDS code. Then both s and d can compute

$$K = MH,$$

where K is a uniformly distributed random key of size $n_1\delta_E(1-\delta)$, for which $I(KF^{n_1}; Z^{n_1}) = 0$.

Of course, we have a probabilistic channel, hence we cannot assume that d and Eve receive exactly as many packets as they are expected. Still a secret key can be established at the same rate with the following change of parameters.

Let s send $n_1 = n'_1 + n_1'^{\frac{3}{4}}$ independent and uniform random packets. If d does not acknowledge $n'_1(1-\delta)$ packets, then an error is declared. Let M denote the first $n'_1(1-\delta)$ packets that d acknowledges. Let $k = \delta_E(1-\delta)n'_1 - n_1'^{\frac{3}{4}}$ and H be a $(n'_1(1-\delta) \times k)$ MDS parity check matrix. Both s and d calculates $K = MH$.

Theorem 1: [2] For any $\epsilon > 0$ there exists a large enough n_1 for which K is computable with an error probability smaller than ϵ, further K is uniformly distributed and

$$I(K; F^{n_1} Z^{n_1}) < \epsilon$$
$$\frac{|K|}{n_1} > \delta_E(1-\delta) - \epsilon.$$

In other words, the same key generation rate $\delta_E(1-\delta)$ is achievable as if d and Eve received exactly as many packets as they are expected.

b) Encryption and encrypted message sending: Again, let us assume for a moment that d and Eve receive as many packets as they are expected. Then let $N = |W| = n_2(1-\delta)$.

Given a secret key K of size $|K| = n_2(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$ established between s and d, the encrypted message W_E is calculated as follows:

$$W_E = W + KG,$$

where G is a $(|K| \times N)$ and is a generator matrix of an MDS code. The packets of the encrypted message W_E are then sent using ARQ. Note that Eve is expected to receive $n_2(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$ different packets of W_E, which is exactly the size of the key we use. Like in the case of secure network coding scheme Eve hence observes a one time pad encryption, from which the secrecy of the message follows.

In our probabilistic model we need to modify the parameters as follows:

$$N = n_2(1-\delta) - n_2^{\frac{3}{4}}; \quad |K| = n_2(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E} + n_2^{\frac{3}{4}}.$$

Given these, G has to be of size $(|K| \times N)$.

Theorem 2: [2] Given a secret key K as generated in the first phase, for any $\epsilon > 0$ there exist large enough n_1, n_2 for which the probability that d receives W_E and can decode W is larger than $1-\epsilon$, and

$$I(Z^n F^n; W) < \epsilon$$
$$\frac{N}{n} > \delta_E(1-\delta)\frac{1-\delta\delta_E}{1-\delta\delta_E^2} - \epsilon.$$

Further, no other scheme can achieve a rate larger than $\delta_E(1-\delta)\frac{1-\delta\delta_E}{1-\delta\delta_E^2}$.

Theorems 1-2 are stated in a slightly different form in [2] without explicitly separating the security properties of the two phases. Theorems 3-4 that we prove in Section IV generalize this result for a network setting.

III. Two-phase secure network coding scheme

In this section, we only examine lossless networks and simply make the point that the secure network coding scheme [5] – with a slight modification – can be cast as a two-phase scheme. We show that the modification does not effect the achieved rate, hence the two phase secure network coding scheme is also optimal. That is, we provide a new, alternative achievability scheme achieving the secure capacity of a lossless network. In Section IV we will derive a unified achievability scheme that will accept as special cases the achievability scheme we provide next for lossless networks, as well as the achievability scheme for the single channel erasure network we described in Section II-D.

A. Example

For simplicity in this example we assume unicast traffic. Consider the following simple network (Fig. 2). Source s and destination d are connected through two parallel unit capacity lossless links out of which any one is being wiretapped by Eve ($h = 2, z = 1$). The secrecy capacity of the network is 1, hence s can send securely a unit size message w. To apply a secure network coding scheme it generates a unit size randomness r. As shown in the figure, on one of the links s sends r while on the other link it sends $r + w$. The eavesdropper either sees r or $r + w$, in either case no information about w is leaked.

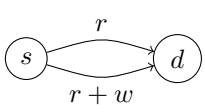


Fig. 2. Secure network coding

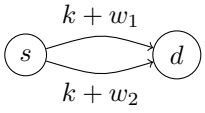
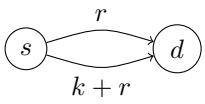
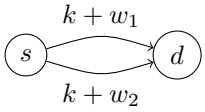


Fig. 3. Coding with shared key



(a) First time slot



(b) Second time slot

Fig. 4. Two-phase scheme example

Assume now, that s and d already share a unit size random key k, which is not known by Eve. Then, as shown in Fig. 3 s can securely send two unit size messages w_1 and w_2 using k for encryption on both links. Hence, in this case a unit size shared key allows us to exploit the min-cut capacity of the network. One might ask, how s and d can set up a shared key. The secure network coding scheme offers a way to send any message w securely to d, this message can equally well be a key k. Consider the example in Fig. 4, where in two time slots two messages are sent securely to d. In the first slot a key is set up, while in the second slot this key is used for encryption. Note that the achieved rate is 1, the same as what the secure network coding scheme achieves. Also the amount of additional randomness remains the same.

To avoid confusion we note here that the randomness used in the secure network coding scheme is often called a key. Indeed, this randomness is used to encrypt a message, however this randomness is not known by d at the moment of encryption. Further, this randomness does not necessarily remain secret from Eve. E.g. in our example in Fig. 2 if Eve selects the top link to wiretap, she learns r. To distinguish source randomness from keys we call a key a shared randomness between s and d which randomness is secret from Eve at the moment of encryption. We also have the property that the key remains secret from Eve given the message is uniformly distributed.

B. Scheme description

The properties that we have seen through our example can be generalized as follows. We call the scheme described below the two-phase secure network coding scheme. As opposed to the secure network coding scheme our scheme uses every link $n = n_1 + n_2$ times, where n_1 and n_2 are the number of time slots used for the two phases respectively. We use the secure network coding scheme as a building block,

we select one such code at the outset and then in each of the n time slots we use the same code on different inputs. Hence, we have that $f_{i,e} = f_{j,e} = f_e, \forall i, j$.

In our scheme the size of our message is $N = n_2 h$. To securely send a message of this size, we need a shared key K of size $n_2 z$ between s and the destination nodes.

1. Key generation: The sender generates a uniformly random K of size $n_2 z$. It also generates additional randomness Θ of size $n_2 \frac{z^2}{h-z}$. The key generation phase consists of $n_1 = n_2 \frac{z}{h-z}$ time slots, in each slot s securely sends $h - z$ packets from K. On edge e in the ith slot we send thus

$$[\Theta(i) \quad K(i)] f_e,$$

where $K(i)$ is the ith $h - z$ length fraction of K: $K(i) = K_{(i-1)(h-z)+1 \ldots i(h-z)}$. Similarly, $\Theta(i)$ is the ith z length fraction of Θ: $\Theta(i) = \Theta_{(i-1)z+1 \ldots iz}$.

2. Encrypted message sending: In the second phase we use K for encryption and in each slot h message packets are sent securely. We use again the same secure network code n_2 times. We denote $W(i)$ the first $h-z$ elements of the ith h length fraction of W and $W'(i)$ the last z elements of the same fraction. On edge e in the ith slot of the second phase we then send

$$[W'(i) + K(i) \quad W(i)] f_e.$$

It directly follows from the properties of the secure network coding scheme that all destination nodes know K and hence can decode W.

Building on the security of the secure network code we use we show that the scheme is secure. We delegate the proof of security to Appendix I.

Achieved rate: Our scheme conveys a message of size $n_2 h$ using $n_1 + n_2$ transmissions, thus our rate is

$$\mathcal{R} = \frac{n_2 h}{n_1 + n_2} = \frac{n_2 h}{n_2 \frac{z}{h-z} + n_2} = h - z,$$

which is the same as the rate of the secure network coding scheme. We further note that the amount of randomness we use is $|K| + |\Theta| = n_2 \frac{hz}{h-z}$, which is also the same as the secure network coding scheme uses to securely send a message of size $n_2 h$. By selecting $n_2 = h - z$ the rate $h - z$ is achieved in a finite block length.

C. Discussion

The two-phase secure network coding scheme gives us a way to isolate two different problems, the key generation and the message communication problem. We have to note that our security argument holds for any key generation phase that satisfies (4) (or possibly instead of equality with zero one that makes the same mutual information term arbitrarily small). In the case of lossless networks the separation does not make any difference in the achieved rates, since the rate of key generation is the same as the achievable rate of secret message sending. However, in some cases this might not hold and a higher key generation rate is possible. In those cases designing the two phases separately results in an improved secure communication rate.

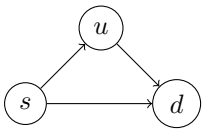


Fig. 5. Example network

IV. Our achievability scheme

We first describe our scheme for a single receiver node $D = d$, then in Section IV-C we generalize our description for multicast. We also assume here that $z \leq h$. This assumption was necessary in case of an error free network to achieve any nonzero rate securely, however in an erasure network we can achieve some rate even if $z > h$. We discuss this case in a subsequent section.

A. Example

Before a detailed description, we explain ideas through the example network in Fig. 5. Let $z = 1$. For simplicity, in the example when we calculate the number of received packets we work with expected values instead of random variables.

1) Key generation: Source s sends independent random packets over all its outgoing links. Both the destination node d and the intermediate node u receive $n_1(1 - \delta)$ packets. On the link between u and d the packets that u received are then sent to d using ARQ. To complete this task u needs n_1 transmissions.

The achievable key rate corresponds to the number of packets that d receives but Eve does not. Eve has three possible choices to select a wiretapped link, and when generating the key we need to consider her worst-case selection.

Case 1: Eve selects the s-d link. In this case the number of packets that both d and Eve receive is $n_1(1-\delta)(1-\delta_E)$.

Case 2: Eve selects the u-d link. Since d eventually receives every packet that u has and every packet is repeated potentially several times, the probability that Eve overhears a certain packet of d is increased to $\frac{1-\delta_E}{1-\delta\delta_E}$. Node u sends $n_1(1-\delta)$ different packets, hence Eve has $n_1(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$ packets in common with d.

Case 3: Eve selects the s-u link. We know that all the packets that u receives d also receives. Eve and u have $n_1(1-\delta)(1-\delta_E)$ packets that they both receive, we get the same result as in the first case.

We conclude that Eve's best choice (from her perspective) is the u-d link. Destination d has

$$|K| = 2n_1(1-\delta) - n_1(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$$

packets not received by Eve, hence a key rate of $1 - \delta + \frac{(1-\delta)^2\delta_E}{1-\delta\delta_E}$ is achievable.

2) Encrypted message transmission: There are two edge disjoint paths between s and d. Let n_2 be the number of transmissions in the second phase. The message is encrypted in the form that we have already seen: $W_E = W + KG$, where K is the key and G is an MDS generator matrix. W_E is split into two parts and each half of the message is assigned to one of the paths. The message packets are then forwarded towards d using ARQ on each link.

The size of the key K we use has to equal the number of packets Eve receives in the second phase. In this case, the MDS property of G ensures that Eve receives every packet with an independent linear combination of K, thus the security of the scheme follows.

Since the same forwarding strategy is applied on each link, regardless of which link Eve selects, she receives a certain packet with probability $\frac{1-\delta_E}{1-\delta\delta_E}$, thus she receives overall $n_2(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$ different packets. Hence n_1 and n_2 are chosen such that $|K| = n_2(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$.

B. Algorithm

As a first step we select h edge disjoint paths between s and d. We ignore all other edges of $\mathcal{G}$. The example in the previous section suggests that the achievable rate depends not only on h, z, δ and δ_E, but also on the number of direct s-d links. Let $h = \ell + t$, where ℓ denotes the number of direct s-d links and t denotes the number of multihop paths.

1) Key generation: We define

$$\begin{aligned} n_1' &= n_1 - n_1^{\frac{3}{4}} \\ \zeta_1 &= n_1'(z-t)^+(1-\delta)(1-\delta_E) \\ &\quad + n_1' \min\{z,t\}(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E} \\ |K| &= hn_1'(1-\delta) - \zeta_1 - n_1'^{\frac{3}{4}}. \end{aligned}$$

Source s sends at most n_1 random packets on all its h outgoing edges. It stops transmission on each link as soon as $n_1'(1-\delta)$ packets are acknowledged on the given link. Intermediate nodes on each path forward the $n_1'(1-\delta)$ packets that they receive to the next node on the path towards d using ARQ.

If d does not receive $hn_1'(1-\delta)$ packets, then an error is declared. Otherwise, let M denote the vector of all the packets that d receives. Both s and d compute

$$K = MH,$$

where H is a $(hn_1'(1-\delta) \times |K|)$ matrix and it is a parity check matrix of an MDS code.

2) Encryption and message sending: We find N, n_2 and n_2' such that

$$\begin{aligned} \zeta_2 &= n_2' z(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}; & |K| &= \zeta_2 + n_2'^{\frac{3}{4}} \\ n_2' &= n_2 - n_2^{\frac{3}{4}}; & N &= hn_2'(1-\delta) \end{aligned}$$

The encrypted message W_E is computed as

$$W_E = W + KG,$$

where K is the key from the first phase and G is a $(|K| \times N)$ matrix and it is a generator of an MDS code.

We assign $n_2'(1-\delta)$ packets to each of our paths. These packets are then forwarded on their assigned path to d using ARQ over each link. If d does not receive all the packets of W_E after n_2 transmissions, then an error is declared.

C. Multicast

In this section we present our scheme for the multicast problem, where there are more than one destination nodes and all of them have to receive the same message securely. Compared to the unicast scheme only a few modifications are needed. To avoid repetition, below we highlight only the differences.

Instead of h edge disjoint paths, first we need to find a network code for multicasting at rate $(1-\delta)h$. Again, we can ignore all edges that are not used by the network code.

In the key generation phase we need the following modification. Instead of sending new random packets on the outgoing edges, s selects in advance $n_1'h(1-\delta)$ random packets that are sent reliably to all destination nodes using ARQ on each link and applying the network code that we have chosen. The same network code is used in each time slot. This ensures that all $d \in D$ receive the same set of packets and hence they all can compute the same key. According to this we modify parameter ζ_1:

$$\zeta_1 = n_1' z(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}.$$

Note that this change implies a change of parameters $|K|, n_2, n_2', \zeta_2$ and N, however all formulas remain the same as defined for unicast.

In the second phase the only difference is that instead of forwarding through h edge-disjoint paths we use the network code (together with ARQ) to reliably send the encrypted packets to all destinations.

Another modification is needed in the selection of matrices H and G. Note that in the unicast case intermediate network nodes do not perform any coding, hence Eve might only receive packets that s produces. This property enables to code only at the source using any H and G matrices that have MDS property. In case of multicast, intermediate nodes might produce new linear combinations, hence Eve might receive combined packets as well.

As for matrix H, consider the $hn_1'(1-\delta)$ packets that s sends in the key generation phase and all their different linear combinations that the prescribed network code produces. Let $f_A^{n_1}$ denote a coefficient matrix of size $hn_1'(1-\delta) \times hn_1'(1-\delta) - |K|$ that describes a $hn_1'(1-\delta) - |K|$ size subset of these packets. This subset corresponds to a set of packets that Eve might receive. We will see during the analysis that the probability that Eve receives a larger subset of packets is negligible. We select H such that $[H \quad f_A^{n_1}]$ is a full rank (in fact invertible) matrix for all possible $f_A^{n_1}$. This property ensures the security of the generated keys. Later we show that such H exists and that generated keys are secure.

As for matrix G, we consider a $|K|$ size subset of the different encoded packets that Eve might receive during the second phase. Let $f_A^{n_2 K}$ denote the $|K| \times |K|$ coefficient matrix of Eve's possible receptions that contain the coefficients of packets from K. We select G such that all possible such $f_A^{n_2 K}$ matrix is invertible. As shown in our analysis this property ensures security of the message.

The existence of such matrices H and G is a direct consequence of known results. The conditions we pose for H and G are the same conditions that need to be satisfied when finding a secure network code. In fact G itself gives a secure network code in a network with parameters $h = |E|n_2'(1-\delta), z = |K|$. Our condition for H can be satisfied as proved by Lemma 3 of [5]. Hence, the existence of such H and G matrices over a sufficiently large field is shown by Lemma 3 and Theorem 2 from [5]. A worst case estimate for the required field size is $q \leq \max\{\binom{|E|}{z}n_1' \choose hn_1'(1-\delta)-|K|}, \binom{\binom{|E|}{z}n_2'(1-\delta)}{|K|}\}$, however any construction proposed for secure network coding (e.g. [16], [17]) can be used to find H and G.

D. Analysis

In case of unicast, the key generation phase achieves a key rate

$$\kappa = h(1-\delta) - (z-t)^+(1-\delta)(1-\delta_E) - \min\{z,t\}(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}, \quad (2)$$

while in case of multicast a key rate

$$\kappa = h(1-\delta) - z(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E} \quad (3)$$

is achieved according to the following theorem.

Theorem 3: For any $\epsilon > 0$ there exists a large enough n_1 such that the key generation runs without error with probability at least $1-\epsilon$, K is uniformly distributed, and the following inequalities hold for any $A \subseteq E$, $|A| = z \leq h$,

$$I(Z_A^{n_1}, F^{n_1}; K) < \epsilon, \text{ and}$$
$$\frac{|K|}{n_1} > \kappa - \epsilon,$$

where in case of unicast κ is as defined by (2), while in case of multicast (3) applies for κ.

The scheme is secure and achieves a rate $\mathcal{R}$ as given by the next theorem.

Theorem 4: For any $\epsilon > 0$ there exists a large enough $n = n_1 + n_2$ such that the above scheme is secure as defined in Definition 1 and achieves a rate

$$\mathcal{R} = \frac{h}{z\frac{1-\delta_E}{\kappa(1-\delta\delta_E)} + \frac{1}{1-\delta}},$$

where κ is as defined by (2) for unicast and by (3) for multicast.

We provide the proofs of Theorems 3-4 in Appendix II-III.

With $\delta = \delta_E = 0$ we see that $\kappa = h - z$ and $\mathcal{R} = h - z$, hence in this special case the scheme achieves the same rate as the secure network coding scheme. Also, for $h = z = \ell = 1$ we have $\kappa = \delta_E(1-\delta)$ and get back $\mathcal{R} = \delta_E(1-\delta)\frac{1-\delta\delta_E}{1-\delta\delta_E^2}$, the optimal rate of a single channel network.

Beside these two, we show optimality for some further cases, see Section V, Theorems 5-6 for outer bounds and Corollary 1 for the optimality result.

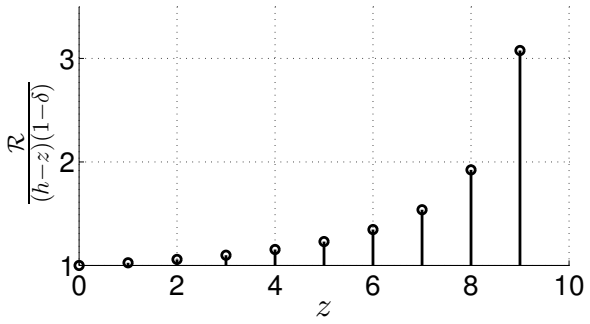


Fig. 6. Advantage of using feedback as a function of the number of eavesdropped edges z, when $h = t = 10, \delta = \delta_E = 0.3$. In this case multicast and unicast rates are the same.

1) Qualitative comparison: The direct application of secure network coding in an erasure network – as we described in Section II-D – allows a rate $(h-z)(1-\delta)$. In case $\delta \geq \delta_E$, taking into account Eve's erasures, but not using feedback does not allow any better rates [11]. The advantage of exploiting feedback is twofold. First, it allows a higher key generation rate $\kappa \geq (h-z)(1-\delta)$. Second, it allows to reduce the size of the key we need in the second phase from $n_2 z(1-\delta)$ to $n_2 z(1-\delta)\frac{1-\delta_E}{1-\delta\delta_E}$. In this section we illustrate qualitatively how large this advantage is.

One can immediately see that the larger δ_E and z are the larger the advantage of using feedback is. In particular if $z = h$ our scheme still achieves a nonzero rate, which is not possible without feedback (assuming $\delta_E \leq \delta$).

In our example we consider the case when $\delta = \delta_E$. In this case the best achievable rate without feedback is $(h-z)(1-\delta)$. We consider a network with parameters $h = t = 10$ and $\delta = 0.3$. We plot in Fig. 6 the advantage of our scheme as the ratio between $\mathcal{R}$ and $(h-z)(1-\delta)$. We see that in this case our scheme achieves a rate up to 3 times higher than the scheme without feedback. With the increase of the network size or $\delta = \delta_E$ the advantage becomes even larger. Note that we have selected the parameter values for the example such that there is no difference between the unicast and the multicast rate of our scheme.

V. Outer bounds

In this section we provide outer bounds on the securely achievable rates in our network model. In some cases (see Corollary 1) the outer bound and the rate achieved by our scheme match, thus the presented scheme is optimal. For other cases we perform numerical evaluations to compare the achieved rates with the upper bound.

When deriving our upper bounds we make the following two assumptions which can only increase the achievable rates: (a) The set of eavesdropped edges are known, hence we restrict Eve to one particular selection of edges. (b) The state of the eavesdropper's channel is also known to every node in the network. In particular we give two bounds. The first bound is valid for any network and depends on h and z, while the other is valid for networks where $O_s = I_d = h$ and beside h and z parameter t also plays a role.

Theorem 5: Assuming $z \leq h$, for the securely achievable rate over $\mathcal{G}$ it holds that

$$\mathcal{R} \leq (1-\delta)(h-z) + z\delta_E(1-\delta)\frac{1-\delta\delta_E}{1-\delta\delta_E^2}.$$

We note here that when $z > h$ we can substitute $z = h$ to get a valid upper bound for all cases. (The secure capacity cannot decrease by decreasing z.)

Theorem 6: Assuming $O_s = I_d = h$, for the securely achievable rate over $\mathcal{G}$ it holds that

$$\mathcal{R} \leq (1-\delta)h - \min\{t,z\}\frac{(1-\delta_E)(1-\delta)}{1-\delta\delta_E}.$$

We provide proofs of Theorems 5-6 in Appendix IV-V. As a corollary of Theorems 5-6 we have the following optimality result.

Corollary 1: Our scheme presented in Section IV achieves secure capacity in the following cases:

1) $h = \ell = z$,
2) $O_s = I_d = h$ and $t \geq z$,
3) $\delta_E = 0$ or $\delta_E = 1$.

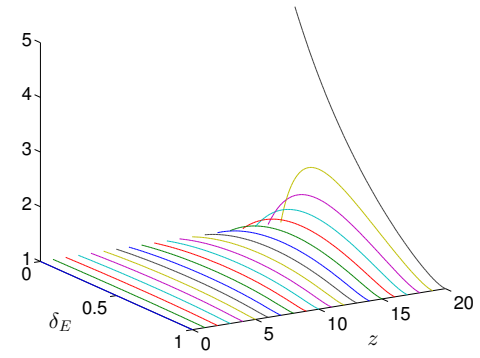
A. Numerical examples

In order to see how the gap between our outer bound and achieved rate behaves we give a few numerical examples for the cases when there is a gap. Theorem 5 holds for any network, while Theorem 6 offers a potentially better bound when $O_s = I_d = h$. For cases when our achieved rate for unicast and for multicast differ we evaluate for the unicast rate.

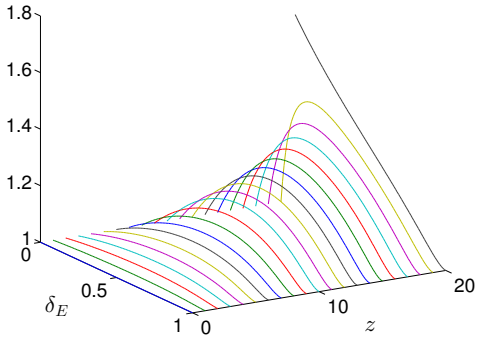
First we consider the bound given by Theorem 5. We express the gap between the outer bound and the achieved rate as the ratio between the two (i.e. 1 means no gap). We evaluate for values $z \leq h$. We can observe that the gap takes its largest value if $h = t$, i.e. all paths between s and d are multihop paths. Further, we get the largest gap for $z = h$ and δ_E close to 0. In this case with $\delta_E \to 0$ the gap tends to $\frac{1}{1-\delta}$, which is the largest possible gap we might get. For other cases the gap is more moderate, see Fig. 7 for a few examples.

For cases when $O_s = I_d = h$ holds, we can take the minimum of our two outer bounds. Note that in this case when $z \leq t$ we do not have any gap, our scheme is optimal. Beside giving this optimality result, Theorem 6 offers an improvement over Theorem 5 for some cases also when $t > z$. As an example, on Fig. 8 we compare the gap that the bound of Theorem 5 gives and the gap we get when taking the minimum of the two bounds. We plot the gap for a few specific values and for $z \geq t$ (for $z < t$ there is no gap in the second case).

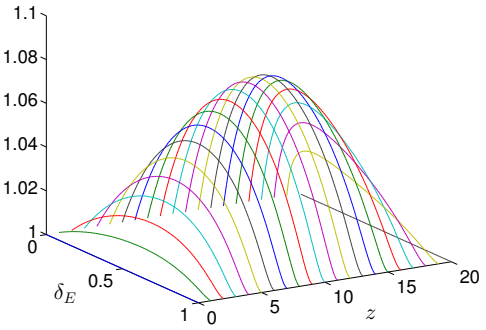
Both our scheme and our upper bounds are general in the sense that beside the min-cut value and the number of direct s-d channels they do not depend on the topology of the network. We conjecture that a more sophisticated network specific analysis could result both in higher achieved rates and in lower upper bounds. However, we see that for most parameter values the rate of the general scheme is already reasonably close to the upper bound.



(a) Upper bound/achieved rate for $h = t = 20$, $\delta = 0.8$



(b) Upper bound/achieved rate for $h = 20$, $t = 10$, $\delta = 0.8$



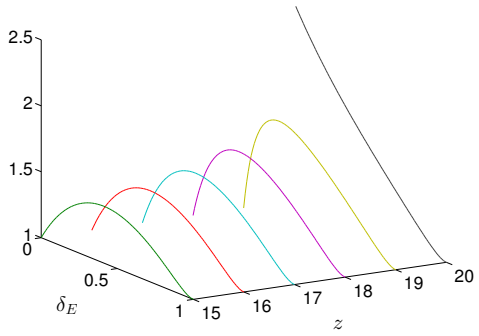
(c) Upper bound/achieved rate for $h = 20$, $t = 0$, $\delta = 0.8$

Fig. 7. Gap between upper bound (Theorem 5) and achieved rate

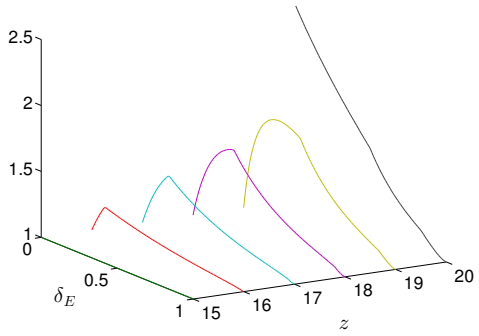
VI. Discussion

A. Code constructions

Theorems 3-4 show the achievability of the claimed secret message rate, however no practical code constructions are given. In case of a unicast problem we need codes with MDS property, for which efficient constructions exist. In case of a multicast problem the codes need to satisfy more constraints, which makes the associated coding problem significantly harder. Giving a practical algorithm for finding codes with the required property for the multicast problem remains an open question.



(a) Upper bound/achieved rate for $h = 20$, $t = 15$, $\delta = 0.8$



(b) Upper bound/achieved rate for $O_s = I_d = h = 20$, $t = 15$, $\delta = 0.8$

Fig. 8. Gap between upper bound (Theorems 5-6) and achieved rate

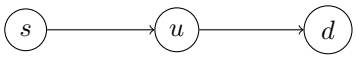


Fig. 9. Two-hop line network

B. Extension for $z > h$

Assume $z = 2$ and consider the two-hop line network shown in Fig. 9. Against this stronger Eve, we can run our scheme as presented in Section IV, but with different parameters. We need to calculate how many packets Eve might receive in each phase. We give the calculation in expectation.

In the message sending phase Eve has two independent chances to overhear a certain packet, on each link she receives a given packet with probability $\frac{1-\delta_E}{1-\delta\delta_E}$, hence the number of different packets she receives (in expectation) is:

$$n_2\frac{1-\delta_E}{1-\delta\delta_E} + n_2\frac{1-\delta_E}{1-\delta\delta_E}\left(1-\frac{1-\delta_E}{1-\delta\delta_E}\right).$$

In the key generation phase she gets $n_1(1-\delta)(1-\delta_E)$ packets in common with u on the first link, while she receives a packet with probability $\frac{1-\delta_E}{1-\delta\delta_E}$ on the second link, hence she is expected to get

$$n_1(1-\delta)(1-\delta_E) + n_1(1-(1-\delta)(1-\delta_E))\frac{1-\delta_E}{1-\delta\delta_E}$$

packets in common with d, which allows a key rate

$$\kappa = \delta_E(1-\delta) - (1-(1-\delta)(1-\delta_E))\frac{1-\delta_E}{1-\delta\delta_E}.$$

To calculate the achievable rate $\frac{(1-\delta)n_2}{n_1+n_2}$ we need to consider n_1 and n_2 such that

$$n_1\kappa = n_2\frac{1-\delta_E}{1-\delta\delta_E} + n_2\frac{1-\delta_E}{1-\delta\delta_E}\left(1-\frac{1-\delta_E}{1-\delta\delta_E}\right).$$

Note that for any given network and any given set of wiretapped edges a similar analysis is feasible. After investigating all the $\binom{|E|}{z}$ possible sets of wiretapped edges, we can design our code such that it provides secrecy against all possible eavesdropped sets. However, the worst-case selection of eavesdropped edges and thus the actual rates achieved highly depends on the topology of our network.

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *The Bell system Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[2] L. Czap, V. Prabhakaran, C. Fragouli, and S. Diggavi, "Secret message capacity of erasure broadcast channels with feedback," in *Information Theory Workshop (ITW)*, 2011, pp. 65–69.

[3] M. Langberg and M. Médard, "On the multiple unicast network coding, conjecture," in *47th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2009, pp. 222–227.

[4] W. Huang, T. Ho, M. Langberg, and J. Kliewer, "On secure network coding with uniform wiretap sets," in *International Symposium on Network Coding (NetCod)*. IEEE, 2013.

[5] N. Cai and R. Yeung, "Secure network coding on a wiretap network," *IEEE Transactions on Information Theory,*, vol. 57, no. 1, pp. 424–435, 2011.

[6] ——, "Secure network coding," in *International Symposium on Information Theory (ISIT)*. IEEE, 2005, p. 323.

[7] R. W. Yeung and N. Cai, "On the optimality of a construction of secure network codes." in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2008, pp. 166–170.

[8] K. Jain, "Security based on network topology against the wiretapping attack," *Wireless Communications, IEEE*, vol. 11, no. 1, pp. 68–71, 2004.

[9] J. Feldman, T. Malkin, C. Stein, and R. Servedio, "On the capacity of secure network coding," in *Allerton Conference on Communication, Control, and Computing*, 2004.

[10] S. Rouayheb and E. Soljanin, "On wiretap networks II," in *International Symposium on Information Theory (ISIT)*, 2007.

[11] T. Cui, "Coding for wireless broadcast and network secrecy," Ph.D. dissertation, California Institute of Technology, 2010.

[12] A. Mills, B. Smith, T. Clancy, E. Soljanin, and S. Vishwanath, "On secure communication over wireless erasure networks," in *IEEE International Symposium on Information Theory (ISIT)*, 2008, pp. 161–165.

[13] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[14] M. Jafari Siavoshani, S. Diggavi, C. Fragouli, U. K. Pulleti, and K. Argyraki, "Group Secret Key Generation over Broadcast Erasure Channels," in *Asilomar Conference on Signals, Systems, and Computers*, 2010, pp. 719–723.

[15] L. Czap, V. Prabhakaran, S. Diggavi, and C. Fragouli, "Broadcasting private messages securely," in *International Symposium on Information Theory (ISIT)*. IEEE, 2012, pp. 428–432.

[16] S. E. Rouayheb, E. Soljanin, and A. Sprintson, "Secure network coding for wiretap networks of type II," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1361–1371, 2012.

[17] D. Silva and F. R. Kschischang, "Security for wiretap networks via rank-metric codes," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2008, pp. 176–180.

APPENDIX I
SECURITY OF TWO-PHASE SECURE NETWORK CODING

In the first phase we use new independent randomness in each slot, hence we know from the secure network code that

$$I(Z_A^{n_1}; K, W) = 0. \tag{4}$$

In the second phase we use only randomness of which Eve has no information, hence her observation of the first phase does not help her learning about W. With a slight abuse of our notation we denote Eve's observation in the second phase by $Z_A^{n_2}$. Formally,

$$\begin{aligned} I(Z_A^n;W) &= I(Z_A^{n_1}Z_A^{n_2};W) = I(Z_A^{n_2};W) + I(Z_A^{n_1};W|Z_A^{n_2}) \\ &\le I(Z_A^{n_2};W) + I(Z_A^{n_1};W) + I(Z_A^{n_1};Z_A^{n_2}|W) \\ &= I(Z_A^{n_2};W) + I(Z_A^{n_1};Z_A^{n_2}|W) \\ &\le I(Z_A^{n_2};W) + I(Z_A^{n_1};K,W|W) \\ &= I(Z_A^{n_2};W) + I(Z_A^{n_1};K|W) = I(Z_A^{n_2};W), \end{aligned}$$

where we used (4) and that $Z_A^{n_2}$ is a function of (K,W). Further, in each slot we use independent randomness, thus

$$\begin{aligned} I(Z_A^{n_2};W) &= \sum_{i=n_1+1}^{n_2} I(Z_{i,A};W|Z_A^{i-1}) \\ &\le \sum_{i=n_1+1}^{n_2} I(Z_{i,A};W) + I(Z_{i,A};Z_A^{i-1}|W) \\ &= \sum_{i=n_1+1}^{n_2} I(Z_{i,A};W). \end{aligned}$$

Hence, we can focus on one time slot of the second phase. Again we can assume that Eve discards any packet received more than once. Let $z' \le z$ be the number of distinct packets she receives. We can write Eve's observation in the ith time slot of the second phase as

$$\begin{aligned} Z_{n_1+i,A} &= [W'(i) + K(i) \quad W(i)] f_A \\ &= [W'(i) + K(i) \quad W(i)] \begin{bmatrix} f_A^{\Theta T} \\ f_A^{WT} \end{bmatrix} \\ &= K(i) f_A^{\Theta T} + W\Phi_{i,A}, \end{aligned}$$

for some matrix $\Phi_{i,A}$. From the properties of the secure network code and K we see that $K(i)f_A^{\Theta T}$ is a set of uniform random packets and $W\Phi_{i,A}$ is at most z' linear combinations of packets from W. Hence, as we noted in the case of secure network coding, from Eve's perspective this is a one time pad encrypted data. From this observation

$$I(Z_A^n;W) \le I(Z_A^{n_2};W) \le \sum_{i=n_1+1}^{n_2} I(Z_{i,A};W) = 0$$

follows, hence our scheme is secure.

Appendix II
Proof of Theorem 3

We introduce some notation. We denote the size of the generated keys k_1,

$$k_1 = |K| = hn_1'(1-\delta) - \zeta_1 - n_1'^{\frac{3}{4}}.$$

We denote M the set of packets d receives, M^Z is the subset of these that Eve receives and M^d is the subset that only d receives. The corresponding rows of matrix H are H^Z and H^d. Hence K can be written as

$$K = MH = \begin{bmatrix} M^d & M^Z \end{bmatrix} \begin{bmatrix} H^d \\ H^Z \end{bmatrix}.$$

Note that in case the key generation is successful, then $|M|$ is $hn_1'(1-\delta)$, but $|M^d|, |M^Z|$ are not deterministic, they depend on the channel realizations F^{n_1}.

K does not depend on $|M^Z|$, hence

$$\begin{aligned} &I(Z_A^{n_1}F^{n_1};K) = I(Z_A^{n_1}F^{n_1}|M^Z|;K) \\ &= I(|M^Z|;K) + I(Z_A^{n_1}F^{n_1};K||M^Z|) \\ &= I(Z_A^{n_1}F^{n_1};K||M^Z|) \\ &= H(K||M^Z|) - H(K|Z_A^{n_1}F^{n_1}|M^Z|) \\ &= k_1 - \sum_{i=1}^{n_1} H(K|Z_A^{n_1}F^{n_1}|M^Z| = i) \Pr\left\{|M^Z| = i\right\} \end{aligned}$$

We have

$$\begin{aligned} &H(K|Z_A^{n_1}F^{n_1}|M^Z| = i) \\ &= H(M^dH^d + M^ZH^Z|M^ZF^{n_1}|M^Z| = i) \\ &= H(M^dH^d|M^ZF^{n_1}|M^Z| = i) \\ &= H(M^dH^d||M^Z| = i) = \min\{|M| - i, k_1\}, \end{aligned}$$

since H^d is full-rank and $|M^d| = |M| - |M^Z|$. Given this,

$$\begin{aligned} &I(Z_A^{n_1}F^{n_1};K) = I(Z_A^{n_1}F^{n_1}|M^Z|;K) \\ &= k_1 - \sum_{i=1}^{|M|-k_1} k_1 \Pr\left\{|M^Z| = i\right\} \\ &\quad - \sum_{i=|M|-k_1+1}^{n_1} (|M| - i) \Pr\left\{|M^Z| = i\right\} \\ &\le k_1 - k_1 \Pr\left\{|M^Z| \le |M| - k_1\right\} \\ &\quad + n_1 \Pr\left\{|M^Z| > |M| - k_1\right\} \\ &= (n_1 + k_1) \Pr\left\{|M^Z| > |M| - k_1\right\}. \end{aligned} \tag{5}$$

The probability that Eve receives more than $|M| - k_1$ packets can be bounded as follows:

$$\begin{aligned} &\Pr\left\{|M^Z| > |M| - k_1\right\} = \Pr\left\{|M^Z| > \zeta_1 + n_1'^{\frac{3}{4}}\right\} \\ &\le \Pr\left\{|M^Z| - \mathbb{E}\left\{|M^Z|\right\} > n_1'^{\frac{3}{4}}\right\} \\ &\le \Pr\left\{\left||M^Z| - \mathbb{E}\left\{|M^Z|\right\}\right| > n_1'^{\frac{3}{4}}\right\} \le e^{-c_1\sqrt{n_1'}}, \end{aligned} \tag{6}$$

for some constant $c_1 > 0$. We used that $\zeta_1 \ge \mathbb{E}\left\{|M^Z|\right\}$ irrespective of Eve's selection. The last inequality follows from the Chernoff-Hoeffding bound. We see from (6) and (5) that $I(Z_A^{n_1}F^{n_1};K)$ can be made arbitrarily small by choosing a large enough n_1. This proves the security of the key.

The key generation fails if d does not receive $hn_1'(1-\delta)$ packets. We calculate the probability of the event that a node who has received $n_1'(1-\delta)$ packet fails to forward all of these to the next node towards d. This event happens if out of n_1 transmissions more than $n_1 - n_1'(1-\delta)$ erasures occur. Let

η denote the number of erasures of n_1 transmissions. Then, the probability of the event equals

$$\begin{aligned}
&\Pr\{\eta > n_1 - n_1'(1-\delta)\} \\
&= \Pr\{\eta - \delta n_1 > (n_1 - n_1')(1-\delta)\} \\
&= \Pr\left\{\eta - \mathbb{E}\{\eta\} > n_1^{\frac{3}{4}}(1-\delta)\right\} \\
&\le \Pr\left\{|\eta - \mathbb{E}\{\eta\}| > n_1^{\frac{3}{4}}(1-\delta)\right\} = e^{-c_2\sqrt{n_1}},
\end{aligned}$$

where $c_2 > 0$ is some constant and we used the Chernoff-Hoeffding bound. This shows that the probability of successful forwarding of $n_1'(1-\delta)$ packets can be made arbitrarily close to 1 on each link, hence the probability that d receives $hn_1'(1-\delta)$ packets is also arbitrarily close to 1 by selecting a large enough n_1.

The last claim of the theorem $\lim_{n_1\to\infty} \frac{|K|}{n_1} = \kappa$ directly follows from the parameter values and from the fact that $\lim_{n_1\to\infty} \frac{n_1'}{n_1} = 1$.

Appendix III
Proof of Theorem 4

Destination d can decode the message if he receives all packets in the second phase. The probability of error has the same nature as in the first phase and thus can be made arbitrarily small by selecting a large enough n_2.

Similarly as in the two phase secure network coding scheme, observing the first phase does not help Eve to learn about W. Formally, if $I(Z_A^{n_1}F^{n_1}; K) = \epsilon'$, then

$$I(Z_A^n F^n; W) \le I(Z_A^{n_2}F^{n_2}; W) + \epsilon', \tag{7}$$

where $Z_A^{n_2}$ denotes the packets that Eve receives in the second phase.

$Z_A^{n_2}$can be written as W_E^Z, which denotes the subset of encrypted packets Eve receives. Let G^Z denote the corresponding rows of G, then

$$W_E^Z = W^Z + KG^Z.$$

In case $|W_E^Z| \le |K|$, then KG^Z is a set of uniformly distributed independent packets, thus W_E^Z is a set of one-time-pad encrypted message packets, hence

$$\begin{aligned}
&I(Z_A^{n_2}F^{n_2}; W) = I(W_E^Z F^{n_2}; W) \\
&= I(W_E^Z; W|F^{n_2}) = I(W_E^Z; W|F^{n_2}|W_E^Z|) \\
&= I(W_E^Z; W|F^{n_2}|W_E^Z| \le |K|)\Pr\{|W_E^Z| \le |K|\} \\
&\quad + I(W_E^Z; W|F^{n_2}|W_E^Z| > |K|)\Pr\{|W_E^Z| > |K|\} \\
&= I(W_E^Z; W|F^{n_2}|W_E^Z| > |K|)\Pr\{|W_E^Z| > |K|\} \\
&\le n_2 \Pr\{|W_E^Z| > |K|\} \le n_2 e^{-c_3\sqrt{n_2'}},
\end{aligned}$$

where $c_3 > 0$ is a constant. We omit the details of the last step, where we bound the probability of the event that Eve receives significantly more packets than she is expected to. We use the same technique as we have seen in the proof of Theorem 3. This together with (7) shows that for a sufficiently large $n = n_1 + n_2$ the scheme satisfies (1). The rate assertion follows directly from the parameter definitions and Theorem 3.

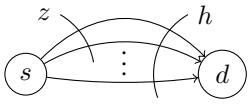


Fig. 10. Transformed network $\mathcal{G}'$

Appendix IV
Proof of Theorem 5

Given a network $\mathcal{G}$ consider the partitioning of the vertices such that (V_1, V_2) such that $s \in V_1, d \in V_2$ and it has the minimum cut value h. We create a new network $\mathcal{G}'(V', E')$ by merging all the nodes in V_1 with s and all the nodes in V_2 with d. We further remove all (d, s) edges. The resulting graph is depicted in Fig. 10. The secure capacity over $\mathcal{G}'$ cannot be smaller than over $\mathcal{G}$ (z remains the same for both). Clearly, merging nodes can only increase capacity. The removal of (d, s) edges does not affect the achievable rates, since no nodes in V_2 can generate randomness, and thus whatever a scheme could send through the (d, s) edges, s can also generate from its randomness and from the public acknowledgments. (Note that the channel states are known to Eve, thus the channel itself cannot be used to generate secure randomness.) We give an upper bound on the secure capacity over $\mathcal{G}'$ which is also a valid upper bound for $\mathcal{G}$.

Without loss of generality we assume that packets sent in the same time slot on different edges are always all independent. If any scheme does not satisfy this assumption, we can construct another scheme that achieves the same rate and satisfies the assumption as follows. We take h independent copies of the scheme (using independent messages and new independent additional randomness). In every h time slots we proceed one transmission of each copy such that on all the h edges a different copy of the scheme runs in each time slot. Clearly the rate does not change and also packets sent in the same time slot are independent.

Notation: Consistently with our previously introduced notation we use $Y_{i,d} = (Y_{i,A}, Y_{i,E'\setminus A})$, where $Y_{i,A}$ denotes the ith reception of d through the set of eavesdropped channels and $Y_{i,E'\setminus A}$ the reception over the not eavesdropped channels. We use $X_{i,s} = (X_{i,A}, X_{i,E'\setminus A})$ in the same fashion. Recall that for the outer bound we assume that A is known.

We start from the following inequality:

$$\begin{aligned}
hn &\ge \sum_{i=1}^{n} H(X_{i,s}) \ge \sum_{i=1}^{n} H(X_{i,s}|Y_d^{i-1}F^{i-1}) \\
&= \sum_{i=1}^{n} H(X_{i,s}|Y_d^{i-1}F^{i-1}W) + I(X_{i,s}; W|Y_d^{i-1}F^{i-1}) \\
&\ge \sum_{i=1}^{n} H(X_{i,s}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + I(X_{i,s}; W|Y_d^{i-1}F^{i-1})
\end{aligned} \tag{8}$$

Through a series of inequalities we prove the following two lemmas that provide bounds for the latter two terms in (8).

Lemma 1:

$$\sum_{i=1}^{n} I(X_{i,s}; W|Y_d^{i-1}F^{i-1}) \geq \frac{n\mathcal{R}}{1-\delta} - \mathcal{E}_1,$$

where $\mathcal{E}_1 = \frac{\epsilon N + n\epsilon + h_2(\epsilon)}{1-\delta}$.

Lemma 2:

$$n\mathcal{R} - n(h-z)(1-\delta) \leq \sum_{i=1}^{n} \frac{\delta_E(1-\delta)(1-\delta\delta_E)}{1-\delta_E} H(X_{i,s}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + \mathcal{E}_2,$$

where $\mathcal{E}_2 = \epsilon N + n\epsilon + h_2(\epsilon) + \frac{\epsilon(1-\delta\delta_E)}{1-\delta_E}$.

Applying these results in (8) and rearranging terms provide the claimed upper bound on $\mathcal{R}$.

A. Proof of Lemma 1

Proof: We know that d can decode with an error probability at most ϵ. Thus, from Fano's inequality:

$$\begin{aligned} n\mathcal{R} - n\epsilon - h_2(\epsilon) - \epsilon N &\leq I(Y_d^n F^n; W) \\ &= \sum_{i=1}^{n} I(Y_{i,d}F_i; W|Y_d^{i-1}F^{i-1}) \\ &= \sum_{i=1}^{n} (1-\delta) I(X_{i,s}; W|Y_d^{i-1}F^{i-1}), \end{aligned}$$

where we used that channel erasures are independent of the message and of each other and that packets sent in the same time slot are independent of each other. ∎

B. Proof of Lemma 2

Proof: We show the following two inequalities:

$$\sum_{i=1}^{n} I(X_{i,A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}) \geq \frac{n\mathcal{R} - n(1-\delta)(h-z)}{1-\delta\delta_E} - \mathcal{E}_{2,1} + \sum_{i=1}^{n} \frac{1-\delta}{1-\delta\delta_E} H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W), \quad (9)$$

where $\mathcal{E}_{2,1} = \frac{\epsilon N + n\epsilon + h_2(\epsilon)}{1-\delta\delta_E}$.

$$\sum_{i=1}^{n} I(X_{i,A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}) \leq \sum_{i=1}^{n} \frac{1-\delta}{1-\delta_E} H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + \frac{\delta_E(1-\delta)}{1-\delta_E} H(X_{i,A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + \frac{\epsilon}{1-\delta_E} \quad (10)$$

We combine these two and get that

$$\frac{n\mathcal{R} - n(1-\delta)(h-z)}{1-\delta\delta_E} - \mathcal{E}_{2,1} - \frac{\epsilon}{1-\delta_E} + \sum_{i=1}^{n} \frac{1-\delta}{1-\delta\delta_E} H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) \leq \sum_{i=1}^{n} \frac{1-\delta}{1-\delta_E} H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + \frac{\delta_E(1-\delta)}{1-\delta_E} H(X_{i,A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W).$$

We observe that $\frac{1-\delta}{1-\delta_E} - \frac{1-\delta}{1-\delta\delta_E} \leq \frac{\delta_E(1-\delta)}{1-\delta_E}$ and thus we can merge the entropy terms corresponding to A and $E'\setminus A$ without violating the inequality (we use again the independence property of parallel transmissions). We conclude that

$$\frac{n\mathcal{R} - n(1-\delta)(h-z)}{1-\delta\delta_E} - \mathcal{E}_{2,1} - \frac{\epsilon}{1-\delta_E} \leq \sum_{i=1}^{n} \frac{\delta_E(1-\delta)}{1-\delta_E} H(X_{i,s}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W).$$

What remains is to show (9) and (10). Consider first (9). We use again Fano's inequality:

$$\begin{aligned} n\mathcal{R} - n\epsilon - h_2(\epsilon) - \epsilon N &\leq I(Y_d^n Z_A^n F^n; W) \\ &= \sum_{i=1}^{n} (1-\delta) I(X_{i,E'\setminus A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}) \\ &\quad + (1-\delta\delta_E) I(X_{i,A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}) \\ &\leq \sum_{i=1}^{n} (1-\delta)(h-z) - (1-\delta) H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) \\ &\quad + (1-\delta\delta_E) I(X_{i,A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}), \end{aligned}$$

where we used the independence property of the channel erasures as well as independence of packets sent in the same time slot over different channels.

We derive (10) as follows.

$$\begin{aligned} 0 &\leq H(Y_d^n|Z_A^n F^n W) \\ &= H(Y_d^{n-1}|Z_A^n F^n W) + H(Y_{n,d}|Y_d^{n-1}Z_A^n F^n W) \\ &= H(Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad - I(Z_{n,A}F_n; Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad + H(Y_{n,d}|Y_d^n Z_A^n F^n W) \\ &= H(Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad - I(Z_{n,A}F_n; Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad + H(Y_{n,E'\setminus A}|Y_d^n Z_A^n F^n W) + H(Y_{n,A}|Y_d^n Z_A^n F^n W) \\ &= H(Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad - (1-\delta_E) I(X_{n,A}; Y_d^{n-1}|Z_A^{n-1}F^{n-1}W) \\ &\quad + (1-\delta) H(X_{n,E'\setminus A}|Y_d^{n-1}Z_A^{n-1}F^{n-1}W) \\ &\quad + \delta_E(1-\delta) H(X_{n,A}|Y_d^{n-1}Z_A^{n-1}F^{n-1}W) \\ &= \sum_{i=1}^{n} -(1-\delta_E) I(X_{i,A}; Y_d^{i-1}|Z_A^{i-1}F^{i-1}W) \\ &\quad + (1-\delta) H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) \\ &\quad + \delta_E(1-\delta) H(X_{i,A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) \\ &\leq \sum_{i=1}^{n} -(1-\delta_E) I(X_{i,A}; W|Y_d^{i-1}Z_A^{i-1}F^{i-1}) \\ &\quad + (1-\delta) H(X_{i,E'\setminus A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) \\ &\quad + \delta_E(1-\delta) H(X_{i,A}|Y_d^{i-1}Z_A^{i-1}F^{i-1}W) + \epsilon. \end{aligned}$$

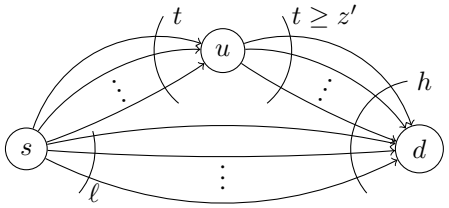


Fig. 11. Transformed network $\mathcal{G}''$

In the last step we used that

$$\sum_{i=1}^{n} I(X_{i,A}; Y_d^{i-1} | Z_A^{i-1} F^{i-1} W) \geq$$
$$\sum_{i=1}^{n} I(X_{i,A}; W | Y_d^{i-1} Z_A^{i-1} F^{i-1}) - I(X_{i,A}; W | Z_A^{i-1} F^{i-1})$$

and that from (1)

$$\epsilon > I(Z_A^n F^n; W) = \sum_{i=1}^{n} I(Z_{i,A} F_i; W | Z_A^{i-1} F^{i-1})$$
$$= \sum_{i=1}^{n} (1-\delta_E) I(X_{i,A}; W | Z_A^{i-1} F^{i-1}).$$

■

APPENDIX V
PROOF OF THEOREM 6

We proceed similarly as we did in the proof of Theorem 5. From a network $\mathcal{G}$ we construct a new graph $\mathcal{G}''$ such that the secure capacity over $\mathcal{G}''$ cannot be smaller than over $\mathcal{G}$. First, we delete all nodes u for which $(d,u) \in E$. Note that this step cannot decrease the secure capacity of network, because if $(d,u) \in E$, then there is no path between u and d, otherwise $\mathcal{G}$ would have a cycle. After this step d has only incoming edges. Next, we merge all intermediate nodes $u \notin \{s,d\}$ into one node. As a result, $\mathcal{G}''$ is a network with three nodes: s, d and u which represents all other nodes. By this we could only increase the achievable rates, hence the upper bound we derive is valid for $\mathcal{G}$. Note that $\mathcal{G}''$ might be cyclic, there might be some edges (u,s). We know that $(d,u) \notin E''$, since d does not have any outgoing edges. As a next step we delete all edges (u,s) from E''. This step cannot reduce the secure capacity of the network, because s knows exactly every packet that u has, hence it can produce any packet that u might send on the (u,s) link. We derive our bound for $z' = \min\{t,z\}$. In case $z > t$ using z' instead of z restricts Eve, hence cannot decrease the secure capacity. Our resulting graph $\mathcal{G}''$ looks as depicted in Fig. 11.

For the same reasons as seen in Theorem 5 we might assume that transmissions over different channels in the same time slot are independent.

We consider an eavesdropper who wiretaps on a known z' size subset of the u-d channels.

We have

$$hn \geq \sum_{i=1}^{n} H(X_{i,s}) \geq \sum_{i=1}^{n} H(X_{i,s} | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1})$$
$$= \sum_{i=1}^{n} H(X_{i,s} | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1} W)$$
$$+ I(X_{i,s}; W | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1})$$
$$\geq \sum_{i=1}^{n} H(X_{i,(s,u)} | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1} W)$$
$$+ I(X_{i,s}; W | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1})$$
$$\geq \sum_{i=1}^{n} H(X_{i,(s,u)} | Y_{(s,d)}^{i-1} Y_u^{i-1} Z_A^{i-1} F^{i-1} W)$$
$$+ I(X_{i,s}; W | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1}). \quad (11)$$

We give bounds on the last two terms seen in (11).

Lemma 3:

$$\sum_{i=1}^{n} I(X_{i,s}; W | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1}) \geq \frac{n\mathcal{R}}{1-\delta} - \mathcal{E}_3,$$

where $\mathcal{E}_3 = \frac{\epsilon N + n\epsilon + h_2(\epsilon)}{1-\delta}$.

Lemma 4:

$$\sum_{i=1}^{n} (1-\delta) H(X_{i,(s,u)} | Y_{(s,d)}^{i-1} Y_u^{i-1} Z_A^{i-1} F^{i-1} W) \geq$$
$$\sum_{i=1}^{n} (1-\delta\delta_E) H(X_{i,A} | Y_d^{i-1} Z_A^{i-1} F^{i-1} W).$$

Lemma 5:

$$n\mathcal{R} - n(h-z')(1-\delta) \leq$$
$$\sum_{i=1}^{n} \frac{\delta_E (1-\delta)(1-\delta\delta_E)}{1-\delta_E} H(X_{i,A} | Y_d^{i-1} Z_A^{i-1} F^{i-1} W) + \mathcal{E}_5,$$

where $\mathcal{E}_5 = \epsilon N + n\epsilon + h_2(\epsilon) + \frac{\epsilon(1-\delta\delta_E)}{1-\delta_E}$.

We give the proof of Lemmas 3-4 in the following subsections. The proof of Lemma 5 is a verbatim copy of the proof of Lemma 2 with $(s,d) \cup (u,d) \setminus A$ placed in the role of $E' \setminus A$. For this reason, we omit the proof.

We apply the results of Lemmas 3-5 in (11) and get the claim of the theorem after rearranging terms.

A. Proof of Lemma 3

Proof: We observe that Y_A^n is a function of (Y_u^n, F^n), and hence

$$I(W; Y_d^n F^n) \leq I(W; Y_{(s,d)}^n Y_u^n F^n)$$
$$= \sum_{i=1}^{n} (1-\delta) I(X_{i,s}; W | Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1})$$

The proof of Lemma 1 holds verbatim here, from which we know that

$$n\mathcal{R} - n\epsilon - h_2(\epsilon) - \epsilon N \leq I(Y_d^n F^n; W).$$

From these two inequalities the claim of the lemma follows.

■

B. Proof of Lemma 4

Proof: We introduce the notation $P = (s,d) \cup (u,d) \setminus A$, i.e. P denotes the set of not eavesdropped incoming edges of d. We use the fact that Y_A^n is a function of (Y_u^n, F^n). From this we have

$$H(Y_u^n | Z_A^n Y_P^n F^n W) \geq H(Y_A^n | Z_A^n Y_P^n F^n W). \tag{12}$$

We expand these terms as follows:

$$\begin{aligned}
&H(Y_u^n | Z_A^n Y_P^n F^n W) \\
&= H(Y_u^{n-1} | Z_A^n Y_P^n F^n W) \\
&\quad + H(Y_{n,u} | Z_A^n Y_P^n Y_u^{i-1} F^n W) \\
&= H(Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad + H(Y_{n,u} | Z_A^{n-1} Y_P^{n-1} Y_u^{i-1} F^{n-1} W) \\
&\quad - I(Z_{n,A}; Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad - I(Y_{n,P}; Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\overset{(a)}{=} H(Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad + H(Y_{n,u} | Z_A^{n-1} Y_P^{n-1} Y_u^{i-1} F^{n-1} W) \\
&\quad - H(Z_{n,A} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad - I(Y_{n,P}; Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&= H(Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad + (1-\delta) H(X_{n,(s,u)} | Z_A^{n-1} Y_P^{n-1} Y_u^{i-1} F^{n-1} W) \\
&\quad - (1-\delta_E) H(X_{n,A} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\quad - (1-\delta) I(X_{n,P}; Y_u^{n-1} | Z_A^{n-1} Y_P^{n-1} F^{n-1} W) \\
&\overset{(b)}{=} \sum_{i=1}^{n} (1-\delta) H(X_{i,(s,u)} | Z_A^{i-1} Y_{(s,d)}^{i-1} Y_u^{i-1} F^{i-1} W) \\
&\quad - (1-\delta_E) H(X_{i,A} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W) \\
&\quad - (1-\delta) I(X_{i,P}; Y_u^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W).
\end{aligned} \tag{13}$$

In (a) we used that $Z_{n,A}$ is a function of (Y_u^{n-1}, F_n) and F_n is independent of every other variable. In (b) we used recursion and that $Y_{(u,d)}^{i-1}$ is a function of (Y_u^{i-1}, F^{i-1}). With a similar derivation we get

$$\begin{aligned}
&H(Y_A^n | Z_A^n Y_P^n F^n W) \\
&= \sum_{i=1}^{n} \delta_E (1-\delta) H(X_{i,A} | Z_A^{i-1} Y_d^{i-1} F^{i-1} W) \\
&\quad - (1-\delta_E) I(X_{i,A}; Y_A^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W) \\
&\quad - (1-\delta) I(X_{i,P}; Y_A^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W) \\
&= \sum_{i=1}^{n} (1-\delta\delta_E) H(X_{i,A} | Z_A^{i-1} Y_d^{i-1} F^{i-1} W) \\
&\quad - (1-\delta_E) H(X_{i,A} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W) \\
&\quad - (1-\delta) I(X_{i,P}; Y_A^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W).
\end{aligned} \tag{14}$$

To get the statement of our lemma we combine (12), (13) and (14) as well as use the fact that Y_A^{i-1} is a function of (Y_u^{i-1}, F^{i-1}) and thus

$$\sum_{i=1}^{n} I(X_{i,P}; Y_A^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W) \leq \sum_{i=1}^{n} I(X_{i,P}; Y_u^{i-1} | Z_A^{i-1} Y_P^{i-1} F^{i-1} W).$$

■