Network Coding for Facilitating Secrecy in Large Wireless Networks

Çağatay Çapar, Dennis Goeckel, Chee Yen Leow, Kin K. Leung, and Don Towsley

***Abstract*—We study the wireless secrecy capacity scaling problem where the question of interest is how much information can be shared among n randomly located nodes such that the throughput is kept information-theoretically secure from m eavesdroppers also present in the network. We present achievable scaling results for both one-dimensional and two-dimensional networks. We show that in a 1-D network, n nodes can share a per-node throughput that scales as $1/n$ which can be kept secure from m randomly located eavesdroppers of unknown location as long as m grows more slowly than $n/\log n$. For a 2-D network, the per-node secure throughput scales as $1/\sqrt{n \log n}$ for any number of eavesdroppers of unknown location which could be arbitrarily located inside this network. These results provide a significant improvement over previous work which either assumed known eavesdropper locations or the number of eavesdroppers that could be tolerated were very limited. The key technique realizing these improvements is the application of simple network coding methods, which were known to help secrecy in a network but their extension to wireless physical-layer secrecy had been limited.**

I. Introduction

Consider the transmission of a message from one party (Alice) to another (Bob), such that it is kept secret from an eavesdropping adversary (Eve). Cryptographic solutions assume that Eve will intercept the transmitted signal cleanly but impose a hard mathematical problem on Eve that is beyond her computational power to solve. On the other hand, information-theoretic solutions exploit the relative signal quality achieved at Bob compared to Eve. Specifically, if the signal quality is better at Bob than it is at Eve, a number of secret bits can be delivered to Bob [1] that is a function of the difference in signal qualities. Hence, physical-layer based information-theoretic security allows a secrecy guarantee without making assumptions on the current and future computational capabilities of the adversary. However, the advantage required for the transmitter-receiver channel versus the transmitter-eavesdropper channel can be difficult to guarantee in a wireless communication network, where the eavesdropper might be very near to the source, might employ a highly-directive antenna to obtain a signal-to-noise ratio (SNR) gain, etc. Hence, one could argue that information-theoretic security in the wireless network has simply traded assumptions on the (long-term) computational capabilities of the eavesdropper for assumptions on the (short-term) operating environment, which may not always match the common physical-layer assumptions. Here, we study information-theoretic secrecy for wireless networks, although it will be apparent that the techniques introduced here can also be used to help traditional cryptographic security to obscure the message from an eavesdropping adversary.

An example where the difficulty in satisfying the physical layer assumptions has made information-theoretic secrecy challenging is in the works on "secrecy capacity scaling", which is the subject of this paper. Here, the question of interest is how much secret information can be shared among nodes in a large wireless network that also includes eavesdroppers. This problem can be seen as a security extension of the original problem of capacity scaling in large wireless networks [2], where an achievable per-node throughput that scales as $1/\sqrt{n \log n}$ is shown for a network of n nodes randomly placed in a two-dimensional region. The studies in this area have shown that it is in fact possible to achieve this throughput without compromising security; however, the assumptions under which security is maintained can be onerous. For example, in [3], it is shown that the optimal throughput can be achieved securely if one assumes (i) the locations of the eavesdroppers are known, (ii) no eavesdropper is very close to any of the legitimate nodes. Obviously, both assumptions are better avoided in a more realistic scenario, especially considering that the eavesdroppers are assumed to be passive.

In this work, we study the secrecy capacity scaling problem under the assumption that the locations of the eavesdroppers are unknown. The first solution to this problem was presented in [4] by employing artificial noise generation by legitimate nodes to enforce some minimum noise floor at the potential locations of the eavesdroppers. This technique is also known as cooperative jamming [5], [6], and has been considered in other scenarios related to wireless secrecy [7]. The use of artificial noise generation addresses the unknown eavesdropper location problem; however, it introduces added interference and requires an excessive amount of transmit energy. Furthermore, this solution is still not able to address the problem of near eavesdroppers, which severely limits the number of uniformly distributed eavesdroppers that could be tolerated. For example,

C. Capar, D. Goeckel, and D. Towsley are with the University of Massachusetts Amherst, Amherst, MA, 01003 USA. Email: {ccapar,goeckel}@ecs.umass.edu, towsley@cs.umass.edu. Kin K. Leung is with the Imperial College, London, SW7 2AZ, UK. Email: kin.leung@imperial.ac.uk. Chee Yen Leow was with the Imperial College and is now with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia. Email: bruceleow@fke.utm.my.

This work was presented in part at the INFOCOM 2012 Conference, March 2012, Orlando, FL, and at the Conference on Information Sciences and Systems, March 2012, Princeton, NJ.

This research was sponsored by the National Science Foundation under grants CNS-0905349 and CNS-1018464, and by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

the scheme in [4] requires the number of eavesdroppers to grow more slowly than $\log n$ in order to achieve a per-node secure throughput that scales as $1/\sqrt{n \log n}$.

Our work of [8] significantly improves upon the result in [4] by presenting a secrecy scheme which achieves a per-node throughput scaling of $1/\sqrt{n \log n}$, while tolerating up to $n/\log n$ eavesdroppers of unknown location. The core of the solution in [8] is a method called "secret sharing" [9]. Here the secret message is divided into a number of packets by the source node such that no information can be obtained about the message unless all packets are received. After generating these packets, the source sends them over separate distant paths to the destination. This way, although the location of the eavesdroppers are unknown, it is known that no eavesdropper can be located close to all paths at once, hence cannot receive all packets. It is interesting to note that this significant improvement comes with a simple coding operation, and it does not require cooperative jamming.

Another coding solution to a wireless secrecy problem was introduced by us in [10], [11] for a different single source-destination pair problem. Here, the basic idea is to employ "two-way communication" to transfer a secret message from Alice to Bob in two steps. In the first step Bob sends a randomly generated message to Alice, and in the second step Alice replies by XORing the random key with the message, i.e., Alice uses Bob's message as a one-time pad. This forces Eve to be able decode both messages, and [10], [11] exploits the fact that the fading channel between Alice and Bob has the same gain in both directions due to reciprocity; however, Eve essentially draws two independent channel gain values to Alice and Bob. This solution is a second example of how a simple coding operation in the upper layers can help with physical-layer security.

It is important to note that the methods of secret sharing and two-way communication have long been known to be useful for achieving secrecy in a communication network. These methods are nothing but special cases of "network coding", and "secure network coding" is a well-understood field [12], [13]. However, its application to wireless secrecy has been limited, as secure network coding is a graph-based approach in which eavesdroppers tap edges (or not), which does not map well to the wireless environment where there are no edges but rather there is a continuum of SNRs. Motivated by the results in [8], [10], [11], here we take a more general look at the secrecy capacity scaling problem with a better understanding of the full potential of network coding which helps us further improve on the result of [8]. In particular, for a 2-D network with n randomly placed nodes, we present a secrecy scheme which achieves secure per-node throughput on the order of $1/\sqrt{n \log n}$ for *any* number of eavesdroppers of unknown location, which could be *arbitrarily distributed* inside the network. As in [8], our scheme uses secret sharing at the source, and routes the packets over distant paths to address unknown eavesdropper locations. However, this does not prevent an eavesdropper located close to the source to receive all packets, which is the major reason restricting the number of (uniformly-placed) eavesdroppers that can be tolerated. The physical challenge in addressing this "near eavesdropper" problem is the fact that whatever the source transmits, a near eavesdropper has a big SNR advantage. Here, we use the two-way scheme to *even out* this SNR gap. In particular, at the start of a route, the source delivers each packet by first receiving a random key from the corresponding relay, and then by replying with XORing the key with the packet. A near eavesdropper has the SNR advantage for the second (outgoing) message, but not the first (incoming) message, hence the packet can be protected from a near eavesdropper. In short, the combination of these two coding methods allows an arbitrary number of eavesdroppers of unknown location to be tolerated.

We also consider 1-D networks, where n legitimate nodes are randomly placed in an interval. We present a construction which achieves per-node secure throughput of order $1/n$ as long as the number of randomly distributed eavesdroppers grows more slowly than $n/\log n$. A fundamental challenge in 1-D is the fact that there is a single path for each source-destination pair. This means even if the location of the eavesdroppers were known, it would not be possible to route around them to reach the destination. This leaves cooperative jamming as the only option to connect source destination pairs and we use it in our construction. In order to handle unknown eavesdropper locations, secret sharing is used, where the 1-D interval is partitioned into a number of regions and each packet is protected from a certain region. However, the construction in 1-D is still limited by the possibility of a nearby eavesdropper, which results in the stated number of eavesdroppers that can be tolerated.

In the rest of the paper, the network and channel models are given in the next section, which are used in our main results presented in Section III for 1-D networks, and in Section IV for the 2-D case. Section V is the conclusion.

II. Model

A. Network and Channel Model

The wireless network is composed of legitimate nodes and eavesdroppers inside the interval $[0, n]$ in the one-dimensional case, and inside the square region $[0, \sqrt{n}] \times [0, \sqrt{n}]$ in the two-dimensional case. Legitimate nodes are distributed according to a homogeneous Poisson point process with intensity $\lambda = 1$. All nodes are assumed to be static. Legitimate nodes are matched into source-destination pairs uniformly at random, such that each node is the destination of exactly one source node, and the source for exactly one destination node. For each pair, we associate a *stream* of information that needs to flow from the source to the destination. Eavesdroppers are assumed to be passive and operating independently of each other, i.e., they do not collaborate by sharing their observations.

Only path loss is assumed for the wireless channels between transmitter and receiver nodes. Hence, whenever a node A transmits with some transmit power P, the received power at node B is modeled as

$$P_{\text{rcv},B} = P/d_{AB}^{\alpha},$$

where d_{AB} is the distance between nodes A, B, and $\alpha > 1$ in 1-D, $\alpha > 2$ in 2-D, is the path loss exponent. The received signal-to-interference-plus-noise ratio (SINR) at B is then

$$\text{SINR}_B = \frac{P_{\text{rcv},B}}{N_0 + I_B}, \tag{1}$$

where N_0 is the power in the additive white Gaussian noise (AWGN) at the receiver, and I_B is the interference received at node B due to other transmissions in the network. In our case, this interference may be due to other legitimate signal transmissions and (for the 1-D case) artificial noise generated by legitimate nodes.

B. Physical-Layer Secrecy Scheme

For all transmissions in the wireless network, the sender node A employs a physical-layer secrecy scheme to deliver the message to the receiver node B at some fixed rate. Motivated by the results in the "wireless wiretap channel" area [14], [15], we assume this secrecy scheme is designed to guarantee secrecy from any eavesdropper E that has roughly the same signal quality with B (or worse). More precisely, for some decoding threshold $\gamma > 0$ for the signal-to-interference-and-noise ratio (SINR), and some (small) δ such that $0 < \delta < 1$, A sends bits to B at some fixed rate R bits per second, which is kept secret from any eavesdropper E if 1) $\text{SINR}_B \geqslant \gamma$, 2) $\text{SINR}_E \leqslant (1+\delta)\text{SINR}_B$. One example of such a secrecy scheme is the low-complexity on-off method in [16], which utilizes fading by sending only at instants when the gain of the channel from A to B is larger than a certain threshold in a given transmission period, and is shown to achieve a positive secrecy rate even when the eavesdropper channel is more capable than the main channel. Many other methods are available (e.g., see [17]).

Therefore, for some region $\mathcal{R}$ in the network, a message is decoded by B while being secret from eavesdroppers inside $\mathcal{R}$ if node B and the eavesdroppers inside $\mathcal{R}$ satisfy the above SINR condition. For multi-hop transmission from a source node to a destination node, if this SINR condition is satisfied at every hop, we refer to the rate of information as the *secure throughput* achieved by this pair (secure from eavesdroppers in $\mathcal{R}$). Note that secrecy at each hop over a multi-hop path is shown to be sufficient for end-to-end secrecy in [3].

III. One-dimensional Networks

The following theorem establishes security in the absence of eavesdroppers near the source-destination nodes, which is then used in Theorem 2 to establish the number of eavesdroppers that can be tolerated.

Theorem 1: Consider the one-dimensional network inside the interval $[0, n]$, where the eavesdroppers are arbitrarily distributed. The locations of the eavesdroppers are unknown, and they are assumed not to collaborate. Legitimate nodes can maintain a throughput of $\Theta\big(1/n\big)$[1] w.h.p. for all source-destination pairs, for any number of eavesdroppers. For some fixed positive constant r, the throughput achieved is secure for the source-destination pairs where both the source and the destination are free from eavesdroppers within a distance $r \log n$.

[1] $f(n) = O(g(n))$ w.h.p. if there exists a constant k such that $P(f(n) \leqslant kg(n)) \to 1$. $f(n) = \Theta(g(n))$ if $f(n) = O(g(n))$, and $g(n) = O(f(n))$ w.h.p.

Overview of the Proof

We prove Theorem 1 by providing a construction summarized by the following steps:

1) In order to handle unknown eavesdropper locations, we partition the network into a finite number t of interlaced regions $\{\Gamma_i, i = 1, \cdots, t\}$. We refer to this partitioning as "coloring" the network, and treat each region (color) one by one, assuming each time that eavesdroppers are all confined to that particular region.
2) For each message to be delivered to the destination node, the source node generates t "packets" corresponding to the t regions. These packets are generated in a way that ensures the message cannot be decoded by a node unless all t packets are successfully received. Packets are delivered in separate transmissions such that the i-th packet is protected from eavesdroppers in Γ_i, thus guaranteeing that an eavesdropper located anywhere in the network misses at least one of the t packets.
3) We provide an algorithm that routes packets from a source to a destination in a multi-hop fashion, and ensures each packet is kept secure from potential eavesdroppers inside its corresponding region at each hop. This is achieved by legitimate nodes inside the region acting as "jammers" by transmitting random noise to prevent eavesdroppers in that region from decoding the packet.
4) We use time division multiplexing, where time is considered as a sequence of "periods". Each period consists of t "frames", and packets corresponding to color $i, i \in \{1, 2, \cdots, t\}$, are transmitted in the i-th frame. Each frame is further divided into slots, where a standard spatial reuse scheme (as in [2]) is employed.

The proof is completed by showing that this construction achieves the stated throughput properties w.h.p.

Proof: Our construction is given in detail in the following. This construction is then proved to achieve a per-node secure throughput of $\Theta\big(1/n\big)$ w.h.p. for the source-destination pairs with no very nearby eavesdroppers.

A. Coloring the Network

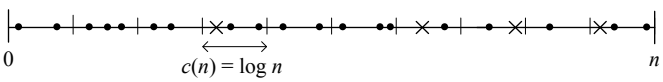

Fig. 1. The one-dimensional network consists of legitimate nodes (represented by dots) and eavesdroppers (represented by crosses) placed in the interval $[0, n]$, divided into cells of length $c(n) = \log n$, as part of the signaling construction.

We divide $[0, n]$ into sub-intervals referred to as "cells", each of length $c(n) = \log n$ (Fig. 1); hence, each cell contains

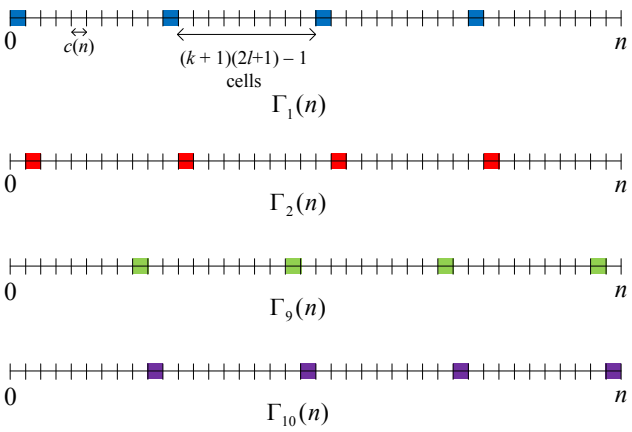

Fig. 2. The network is partitioned into regions (colors), where each region is a collection of cells regularly sampled in the linear grid. Cells in a region are spaced $(k+1)(2l+1)-1$ cells apart ($k=1$, $l=2$ in the figure). Hence, the network consists of $t=(k+1)(2l+1)$ regions ($t=10$ in the figure). The network is shown here with four of these ten regions highlighted.

a legitimate node w.h.p. Let $s_i(n)$ denote the i-th cell, $i = 1, \cdots, n/\log n$, with $s_1(n) = [0, \log n]$.

We partition these cells into non-overlapping subsets, which we refer to as "coloring" the network. Specifically, we divide the network into $t = (k+1)(2l+1)$ regions (colors), where $k \geqslant 1$ and $l \geqslant 2$ are integers to be defined later. Denote the collection of regions as:

$$\{\Gamma_i(n), i = 1, 2, \cdots, t\}$$

Each region is a collection of non-contiguous cells regularly sampled in the grid as shown in Fig. 2. Specifically, cells in $\Gamma_i(n)$ are spaced $t-1$ cells apart. In other words,

$$\Gamma_i(n) \triangleq \bigcup_{j=1}^{\frac{n/\log n}{t}} s_{i+(j-1)t}(n). \tag{2}$$

For convenience, we denote the j-th cell of region $\Gamma_i(n)$ as $\mathcal{C}_i^j(n)$. In other words,

$$\mathcal{C}_i^j(n) \triangleq s_{i+(j-1)t}(n).$$

The whole network is the union of the t regions:

$$[0, n] = \bigcup_{i=1}^{t} \Gamma_i(n)$$

Note that the number of regions t is independent of the size n of the network.

We refer to $\Gamma_i(n)$ and each of its cells $\mathcal{C}_i^j(n)$ as belonging to the i-th color. Also, we use the notation $\Gamma_i, \mathcal{C}_i^j$ in what follows, keeping in mind that the number of cells in a region, and the cell sizes depend on n. As will be clear in the description of the routing algorithm, the cells in a region can be thought of as potential locations of eavesdroppers corresponding to that region. For each cell $\mathcal{C}_i^j$, we define an interval called the "neighborhood" of this cell, and denote it by $N(\mathcal{C}_i^j)$. This neighborhood consists of $(2l+1)$ cells, with $\mathcal{C}_i^j$ being the middle cell (Fig. 3). These neighborhoods are separated by $k(2l+1)$ cells.

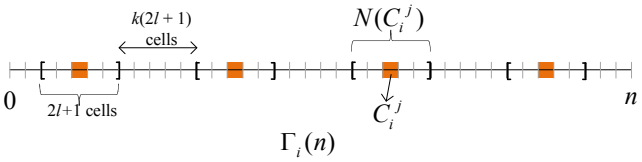

Fig. 3. The network is shown with one region Γ_i highlighted as done in Fig. 2. $\mathcal{C}_i^j$ denotes the j-th cell in region Γ_i. Around each cell, the "neighborhood" of that cell $N(\mathcal{C}_i^j)$ is defined as the interval with $(2l+1)$ cells ($l=2$ above). So, neighborhoods are separated by $k(2l+1)$ cells ($k=1$ above).

It is also useful to define the "periphery" and the "interior" of a neighborhood. We define the *periphery* of $N(\mathcal{C}_i^j)$ as the two cells at the two ends of the neighborhood, and the *interior* of $N(\mathcal{C}_i^j)$ as the smaller interval that consists of $(2l-1)$ cells centered at $\mathcal{C}_i^j$.

For any source-destination pair S-D, let x be the b-bit message to be delivered from S to D. S generates $(t-1)$ random b-bit packets $w_1, \cdots, w_{t-1}$ and then sets w_t such that the message x satisfies

$$w = w_1 \oplus w_2 \oplus \cdots \oplus w_t, \tag{3}$$

where $\oplus$ denotes bit-wise XOR operation. We refer to packet w_i as belonging to the i-th color. The basic idea is that w_i is transmitted such that it is *protected* from eavesdroppers located in Γ_i. Note that any node that receives all t packets can compute x, while any node that misses one or more packets acquires no information about x.

B. Routing Algorithm

For the transmission of a packet of any color i from a source node S to its destination node D, S transmits the packet to a relay in the next cell on the route. Each relay that receives the packet does the same until the packet reaches the first neighborhood $N(\mathcal{C}_i^j)$ on the route (Fig. 4 (a)). Inside $N(\mathcal{C}_i^j)$, we assign two nodes to act as relays, and one node to act as a jammer: A relay node A is selected from the cell where the route enters $N(\mathcal{C}_i^j)$, a jammer node J is selected from $\mathcal{C}_i^j$, and a relay node B is selected from the cell at the end of the neighborhood (Fig. 4 (b)). A receives the message from outside the neighborhood, and then transmits to B while J transmits random noise. Therefore, inside a neighborhood, the message is transmitted across a number of cells in one slot. A jammer is only active when there is a transmission inside its corresponding neighborhood. Therefore, each packet is carried in a repeating sequence of single-cell hops followed by one multi-cell hop until it reaches D (Fig.4 (a)). When D receives all t packets, it decodes the message by performing the operation in (3).

Note that packets of color i are routed in a way that prevents it from entering the interiors of neighborhoods $N(\mathcal{C}_i^j)$. The only exception is possibly at the start or the end of the route.

To see this, consider a source node S inside $\mathcal{C}_i^j$. S will generate a packet w_i of color i. This packet is first routed in single-cell hops, and follows the above scheme only after it reaches outside $N(\mathcal{C}_i^j)$. Similarly, deliveries to destination nodes inside neighborhoods are also done in a sequence of single-cell hops (see Fig. 4 (a)).

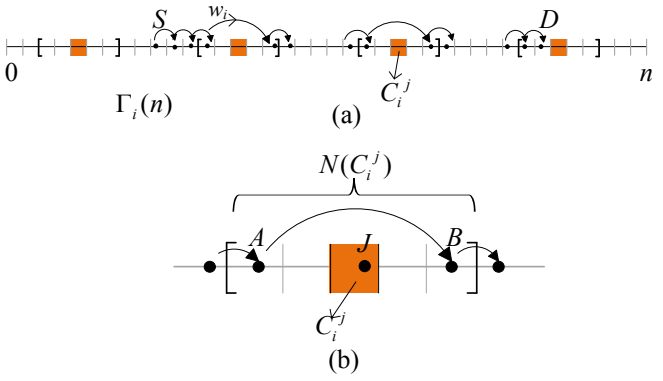

Fig. 4. (a) The route followed by a packet w_i from a source node S to a destination node D is shown. At each hop, the packet is delivered to the next cell on the route except inside the neighborhoods, $N(\mathcal{C}_i^j)$, where the packet is transmitted such that it reaches over multiple cells at once. (b) Inside $N(\mathcal{C}_i^j)$, a transmitting relay A in the first cell transmits to a receiving relay B in the last cell, while a jammer node J in $\mathcal{C}_i^j$ transmits artificial noise. Hence, packets of color i are routed in a way that avoids entering the interiors of the neighborhoods $N(\mathcal{C}_i^j)$. The only exception is possibly at the start or the end of the route, as the source or the destination node may be located inside the interior of a neighborhood (e.g., the destination node D is inside the interior of a neighborhood in (a)).

C. Time Division Multiplexing Scheme

Time is considered as a sequence of "periods". Each period consists of t "frames". In the i-th frame, only packets belonging to the i-th color are transmitted. In each frame, a spatial reuse scheme is employed such that in the i-th frame, every cell in the network transmits a packet of color i once. This is done by further dividing each frame into t time slots. In each slot, transmitting cells are $t-1$ cells apart (see Fig. 5). During the i-th frame, jammer nodes inside Γ_i become active only in the time slots where multiple-cell hops take place (see Fig. 2 (b)).

The throughput achieved per stream is constrained due to the fact that the streams arriving to a cell take turns being relayed. Each cell has to relay information for at most a constant factor of n streams w.h.p., hence a throughput of $\Theta\big(1/n\big)$ per stream is achieved w.h.p.

In order to consider the secrecy of the achieved throughput, note that the route of a packet contains the following types of hops: (1) single-cell hop outside the neighborhoods, (2) multi-cell hop inside a neighborhood, (3) single-cell hop inside a neighborhood if it contains either the source or the destination (see Fig. 4 (a)). It can be shown that (see Appendix I in [8]) the first two types of hops are achieved securely by showing that there exist constants k, l for coloring the network, and transmit power values for relays and jammer nodes, such that for any stream of color i, the destination node and the eavesdroppers inside Γ_i satisfy the SINR requirement for secure transmission for these hops. Hence, the only possible insecure transmissions are in the close proximity of the source and destination nodes (i.e., the third type above). For a source-destination pair, if no eavesdropper is within a distance $rc(n)$, with $r = l$, to the source and to the destination, then these hops will also be secure, hence the result follows. ■

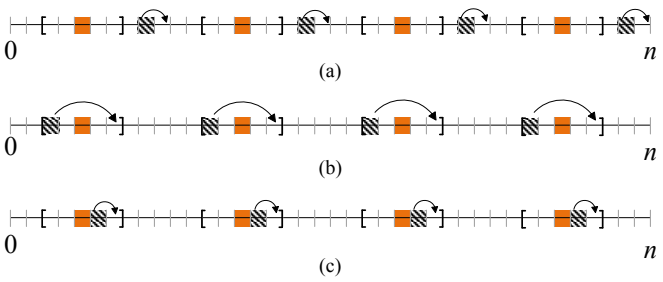

Fig. 5. One period is divided into t frames. In the i-th frame, packets of color i are transmitted according to the routing protocol corresponding to Γ_i (see Fig. 4). Each frame further consists of t time slots ($t = 10$ in the figure). Cells transmitting simultaneously (dashed cells) in one slot are $t-1$ cells apart. For the i-th frame, three time slots are shown above: (a) shows a time slot with single-cell transmissions outside neighborhoods, (b) shows a time slot with multi-cell hops over the neighborhoods (with all jammers active), (c) shows a time slot with transmissions inside the neighborhoods. Note that cells in the periphery of the neighborhoods also do "occasional" single-cell hops for deliveries to destinations inside neighborhoods, with all jammers passive (not shown above).

Theorem 2: Consider the one-dimensional network inside the interval $[0, n]$, where the eavesdroppers are placed according to a Poisson point process with some density $\lambda_e > 0$, independent of the placement of the legitimate nodes. The locations of the eavesdroppers are unknown, and they are assumed not to collaborate. Then, the fraction of source-destination pairs that can maintain a per-node secure throughput of $\Theta\big(1/n\big)$ is arbitrarily close to one w.h.p. if $\lambda_e = o(1/\log n)$.

Proof: We use the same construction as that used to prove Theorem 1. For source-destination pairs free from any nearby eavesdroppers, this construction achieves w.h.p. the stated secure throughput. Hence, the proof follows by showing that for $\lambda_e = o(1/\log n)$, the fraction of source-destination pairs that do not have any nearby eavesdroppers is arbitrarily close to one w.h.p.

Let the random variable $m(n)$ be the number of eavesdroppers in the network, which has an expected value of $\lambda_e n$. Let $y_i \in [0, n]$ be the location of the i-th eavesdropper, and define $A_i(n) = [y_i - l\log n, y_i + l\log n]$, with length $\ell(n) = |A_i(n)| = 2l\log n, \forall i$. Let $A(n)$ be the total region covered by the eavesdroppers, i.e., any source or destination node inside $A(n)$ will not be able to communicate secretly.

$$A(n) \triangleq \bigcup_{i=1}^{m(n)} A_i(n) \tag{4}$$

Let $N_i(n), N_o(n)$ be the random variables denoting the number of legitimate nodes inside and outside $A(n)$, respec-

tively. For some $\varepsilon > 0$, let the event $C^{\varepsilon}(n)$ be defined as

$$C^{\varepsilon}(n) \triangleq \left\{ \frac{N_i(n)}{N_i(n) + N_o(n)} < \varepsilon \right\}. \tag{5}$$

We can write $P\left(C^{\varepsilon}(n)\right)$ as

$$\begin{aligned} &P\left(C^{\varepsilon}(n)\right) = \\ &P(C^{\varepsilon}(n) \,|\, \{|A(n)| \le 2\lambda_e n\ell(n)\}) P(\{|A(n)| \le 2\lambda_e n\ell(n)\}) \\ &+P(C^{\varepsilon}(n) \,|\, \{|A(n)| > 2\lambda_e n\ell(n)\}) P(\{|A(n)| > 2\lambda_e n\ell(n)\}). \end{aligned}$$

Define the random variable $X(n)$ as

$$X(n) \triangleq \frac{N_i(n)/n}{N_o(n)/n}.$$

Given $\lambda_e = o(1/\log n)$, and $|A(n)| \le 2\lambda_e n\ell(n)$,

$$N_i(n)/n \to 0, N_o(n)/n \to 1, \text{ and } X(n) \to 0, \text{ a.s.}$$

Then,

$$\begin{aligned} &P\big(C^{\varepsilon}(n) \,|\, \{|A(n)| \le 2\lambda_e n\ell(n)\}\big) = \\ &P\left(\frac{X(n)}{1+X(n)} < \varepsilon \,|\, \{|A(n)| \le 2\lambda_e n\ell(n)\} \right) \to 1, \text{ as } n \to \infty. \end{aligned} \tag{6}$$

Finally,

$$\begin{aligned} P(|A(n)| \le 2\lambda_e n\ell(n)) &\ge P(m \le 2\lambda_e n) \\ &\to 1, \text{ as } n \to \infty. \end{aligned}$$

Thus, for any $\varepsilon > 0$, $P(C^{\varepsilon}(n)) \to 1$, as $n \to \infty$. This shows the fraction of nodes inside $A(n)$ is arbitrarily close to zero w.h.p., which readily implies that the fraction of source-destination pairs inside $A(n)$ is arbitrarily close to zero w.h.p ■

IV. Two-Dimensional Networks

The following theorem states our main result for a 2-D network.

Theorem 3: Consider an extended two-dimensional network, where legitimate nodes are placed according to a Poisson point process with density 1 over a torus formed by wrapping around a square region of size $[0, \sqrt{n}] \times [0, \sqrt{n}]$ at the edges. Legitimate nodes are matched into source-destination pairs uniformly at random. In addition to the legitimate nodes, eavesdroppers are arbitrarily distributed with their location unknown. Eavesdroppers are assumed not to collaborate. Each source-destination pair can achieve a throughput that scales as $1/\sqrt{n \log n}$ with probability one as $n \to \infty$. The throughput achieved is secure for any number of eavesdroppers.

Overview of the Proof

We prove Theorem 3 by providing a construction summarized by the following steps. The most important differences compared to the 1-D construction are: (i) here the packets generated are carried over different paths utilizing the fact that in 2-D many paths are available between a single-source destination pair, (ii) at the start of a path, instead of simply giving the packet from the source to the first node on the path, the initiation is done in a special way using a two-step protocol.

1) For each source-destination pair S–D, S generates four "packets" for each secret message x to be conveyed from S to D. The first three packets w_1, w_2, w_3 are generated randomly, and the last packet w_4 is set such that $x = w_1 \oplus w_2 \oplus w_3 \oplus w_4$.
2) For each S–D, we define regions surrounding S and D called the "source base", and the "destination base", respectively. We define four paths between S–D and each packet is carried on a different path. The paths keep a certain minimum distance outside the source and the destination bases (see Fig. 7). This ensures that eavesdroppers *outside the bases* cannot be close to all paths at once to decode all four packets.
3) To initiate the flow from S, four relays R_1, R_2, R_3, R_4 are selected, where packet w_i is delivered from S to R_i (see Fig. 6). Each packet w_i is conveyed from S to R_i using a two-way scheme: R_i first generates a random key k_i and sends to S, after which S replies with $w_i \oplus k_i$. Then, R_i extracts the packet w_i. The locations of the four relays are selected such that no eavesdropper can be located in a position to decode all four packets (an eavesdropper needs to hear all eight transmissions for this). This ensures the message is protected from any eavesdropper *inside the source base*.
4) The delivery to D is done by sending the four packets to D from four directions (see Fig. 6). This ensures that no eavesdropper *inside the destination base* can receive all four packets.
5) A standard time division multiplexing and spatial reuse scheme is employed where cells take turns transmitting as in the 1-D case. The only difference is that there are three distinct phases: 1) the draining phase, where the packets are delivered to the relays from the sources, 2) the routing phase, where the packets are carried on the paths outside the bases, 3) the delivery phase, where the relays deliver the four packets to the destinations.

The proof is completed by showing that this construction achieves the stated throughput properties w.h.p.

Proof: The construction achieving the secrecy result consists of a routing algorithm, and a time-division multiplexing scheme.

A. Routing Algorithm

The routing algorithm consists of three stages: draining, routing, and delivery.

Draining: For each source node S, we define a square region of size 7×7 cells with the source cell at its center as the "source base" (see Fig. 6). The four corner cells of the source base are designated as the "relay cells". Four legitimate nodes R_1, R_2, R_3, R_4 are selected from these four relay cells, and the packets are conveyed to the relays using a two-way scheme. For example, the node R_1 generates a random key k_1 and sends it to S, and S replies with $c_1 = w_1 \oplus k_1$, and R_1 extracts the packet w_1 (Fig. 6).

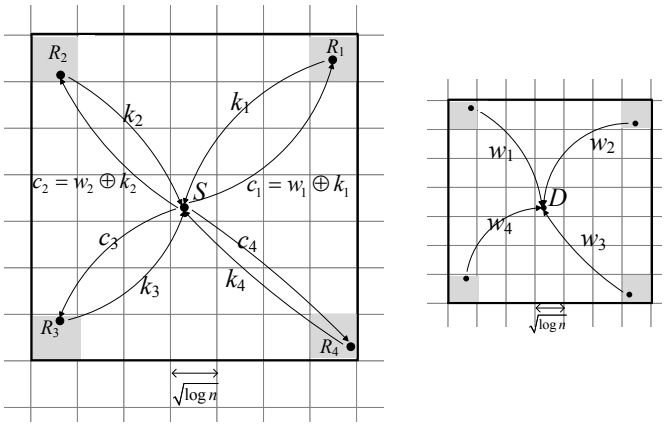

Fig. 6. (Left) Around each source s, a "source base" is defined, which is a square region of size 7×7 cells. The four (shaded) corner cells are the relay cells, where nodes are selected to help initiate the transmission. The four relays do two-way exchanges with the source to receive four packets that form the secret message. The locations of the relays ensure that (compared to the source) no eavesdropper can be located closer to all relays at once, i.e., for any given eavesdropper E, $d(S, R_i) \leqslant d(E, R_i)$ for some $i \in \{1, 2, 3, 4\}$. (Right) The delivery of the four packets to the destination is shown. As is the case for the draining phase, due to the location of the relays, no eavesdropper can be close enough to all relays at once to collect all four packets.

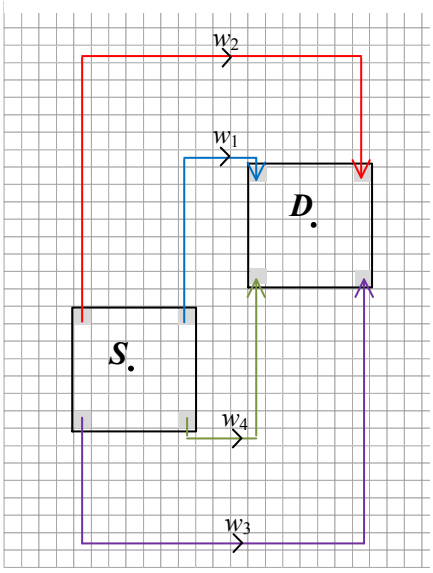

Fig. 7. The source and the destination bases are connected with four paths, each carrying one of the packets. The paths have the same minimum spacing throughout the route; hence, no eavesdropper can be close enough to all four paths at once.

Routing: We define four paths between the source and the destination bases (Fig. 7). Each packet w_i is carried on a different path. The paths consist of vertical or horizontal lines, which are traversed by the packets in single-cell hops, where the packet is delivered to a node in the next cell on the path. Two paths leave the top two relay cells on a vertical line, and arrive to the correponding relay cells in the destination base on a vertical line while keeping the same spacing (Fig. 7). The same is true for the paths leaving the bottom relay cells.

Delivery: For each destination node D, a "destination base" is defined in the same way as the source base (see Fig. 6). Again, the four corners are labeled as relay cells. After a packet reaches a relay cell in the destination base, the packet is delivered from the relay directly to D by reaching over multiple cells as done for the draining case. Once all four packets arrive to D, it decodes the secret message x by XORing the packets.

Remark: Some special cases need to be considered: (i) Source and destination bases which are roughly vertically aligned: the paths leave the source base and arrive at the destination base on horizontal lines. (ii) The source and the destination bases overlap: the secret message is delivered via a helper node. In particular, a helper node is selected, and the secret message is delivered first to this helper node, and then from the helper node to the destination node using the routing algorithm described above for both stages. The helper node is selected from a cell that is far enough away from the source and the destination bases to allow employing the routing algorithm as described above. Details are omitted due to space constraints.

B. Time Division Multiplexing Scheme

Time is divided into three phases corresponding to the draining, routing, and delivery stages.

Draining: The draining phase is divided into eight frames. The first four frames are for the transmissions of the keys from the relays to the sources, and the last four frames are for the responses of the sources. Each frame consists of a constant number of time slots, where cells take turns for signal transmissions employing a standard spatial reuse scheme where cells transmitting in the same time slot are regularly spaced in the network (e.g., see [18]). Hence, at the end of each frame, it is ensured that each cell has transmitted once.

Routing: The routing phase is divided into four frames for each type of packet. In the ith frame, packets of color i are routed. Each frame is further divided into time slots again employing a spatial reuse scheme. In each time slot, relaying nodes from the active cells deliver their packets to the next cell on the path.

Delivery: The delivery phase consists of four frames for the transmission of the packets of four colors. Again, each frame is divided into time slots, and transmissions are done as in the draining phase.

The proof completes by showing that: (i) this construction is feasible, (ii) it achieves a per-node throughput on the order of $1/\sqrt{n \log n}$, and (iii) the achieved throughput is secure.

The first two statements can be shown by standard arguments used in similar works (see e.g., [8], [4], [18]), and we omit the detailed proof here. Basically, the throughput achieved by the construction is found by considering the throughput constraint imposed by each phase. For the draining

and delivery phases, the difference in our construction compared to a standard construction is that transmissions require multi-cell hops, and that these phases complete in more than one transmissions. However, these both bring only a constant factor to the throughput achieved and do not affect the scaling. The difference in the routing phase is that each message requires four packets to be carried, which again does not affect the order. It can be shown that the performance bottleneck is due to the routing phase, and since the relaying load in each cell grows with $\sqrt{n \log n}$ (see e.g., Appendix II in [8]), the overall per-node throughput scales as $1/\sqrt{n \log n}$. Finally, note that the construction requires nodes to transmit with power that is proportional to $(\log n)^{\alpha/2}$, where $\alpha > 2$ is the path loss exponent of the medium.

Next we show that for each source-destination pair S–D, each message x is delivered from s to d securely. For secrecy, we show that an eavesdropper located anywhere in the network is guaranteed to miss at least one packet out of the four packets after listening to all the transmissions required for the delivery of x. First consider the draining phase. Consider any given eavesdropper E. Due to the relative locations of the relays with respect to the source, E is guaranteed to satisfy $d(E, R_i) \geqslant d(S, R_i)$ for some $i \in \{1, 2, 3, 4\}$ (Fig. 6). Hence, for the transmission of k_i from R_i to S, the received signal power at S is larger than the received signal power at E. In addition, the spatial reuse scheme can be designed such that the interference at S is low enough to ensure $\text{SINR}_E \leqslant (1+\delta)\text{SINR}_S$. Therefore, k_i is delivered to S but not to E; hence, E misses the packet w_i. Therefore, any eavesdropper is guaranteed to miss at least one packet, and the message x is not leaked during the draining phase. A similar argument can be made for the delivery phase. The four packets arrive to D from four directions and any given eavesdropper E satisfies $d(E, R_i) \geqslant d(D, R_i)$ for some i. Outside the bases, the packets are carried on paths with some minimum spacing; hence, no eavesdropper can be close enough to many paths at once, thus establishing secrecy during the routing phase. Finally, it can be easily verified that no eavesdropper can collect the four packets by listening to all three phases.

Therefore, using this construction, as n grows, each source-destination pair can share on the order of $1/\sqrt{n \log n}$ secret bits per second for any number of independent eavesdroppers arbitrarily distributed to the network. ■

V. Conclusion

Network coding techniques have the potential to improve information-theoretic secrecy in wireless networks, most notably by enabling the secure connection of one node to another in the presence of very nearby eavesdroppers. We address the secrecy capacity scaling problem using network coding techniques. Most notably, we show that, in a 2-D network, n randomly located nodes can share per-node secret information at a rate on the order of $1/\sqrt{n \log n}$, for any number of arbitrarily distributed eavesdroppers of unknown location.

This work partially completes a line of research that originated with the secrecy-capacity tradeoffs in asymptotically large networks of [4]. In [4], even when multi-user diversity and cooperative jamming were employed, the near eavesdropper problem severely limited the number of uniformly distributed eavesdroppers that could be tolerated in the network. In [10] and [8], we began to realize the utility of modifications at higher layers in resolving difficult secrecy problems, but we still were not able to address eavesdroppers very near the nodes originating messages. The work here addresses this last problem and thus allows for a secure per-session throughput of $O(1/\sqrt{n \log n})$ in the presence of an arbitrarily located set of non-collaborating eavesdroppers. The case of collaborating eavesdroppers is the subject of our future work in this area.

References

[1] A. D. Wyner, "The Wire-tap Channel," *Bell Systems Technical Journal*, vol. 54, no. 8, pp. 1355–1387, January 1975.

[2] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388 –404, March 2000.

[3] O. O. Koyluoglu, C. E. Koksal, and H. E. Gamal, "On secrecy capacity scaling in wireless networks," *IEEE Trans. Inf. Theory*, submitted for publication (eprint arXiv:0908.0898).

[4] S. Vasudevan, D. Goeckel, and D. F. Towsley, "Security-capacity trade-off in large wireless networks using keyless secrecy," in *Proc. of MobiHoc '10*. New York, NY, USA: ACM, 2010, pp. 21–30.

[5] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Communications*, vol. 7, no. 6, pp. 2180 –2189, June 2008.

[6] X. He and A. Yener, "Cooperative jamming: The tale of friendly interference for secrecy," in *Securing Wireless Communications at the Physical Layer*, R. Liu and W. Trappe, Eds. Springer, 2010, pp. 65–88.

[7] D. Goeckel, S. Vasudevan, D. Towsley, S. Adams, Z. Ding, and K. Leung, "Artificial noise generation from cooperative relays for everlasting secrecy in two-hop wireless networks," *IEEE J. Select. Areas Commun.*, vol. 29, no. 10, pp. 2067 –2076, December 2011.

[8] C. Capar, D. Goeckel, B. Liu, and D. Towsley, "Secret communication in large wireless networks without eavesdropper location information," in *Proc. of INFOCOM 2012*, March 2012.

[9] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, pp. 612–613, November 1979.

[10] C. Y. Leow, D. Goeckel, and K. K. Leung, "Two-way secrecy schemes for the broadcast channel with internal eavesdroppers," in *Proc. Annual Conference of the International Technology Alliance 2011*.

[11] C. Y. Leow, C. Capar, D. Goeckel, and K. K. Leung, "Two-way secrecy schemes for the broadcast channel with internal eavesdroppers," in *Proc. of Asilomar 2011*, Nov. 2011.

[12] N. Cai and R. Yeung, "Secure network coding," in *ISIT 2002*, p. 323.

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

[14] M. Bloch, J. Barros, M. Rodrigues, and S. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, June 2008.

[15] M. Bloch and J. Barros, *Physical-Layer Security*. Cambridge University Press, 1971.

[16] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[17] S. Xiao, W. Gong, and D. Towsley, "Secure wireless communication with dynamic secrets," in *INFOCOM 2010, IEEE*, 2010.

[18] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran, "Closing the gap in the capacity of wireless networks via percolation theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1009 –1018, March 2007.