Finite-State Wiretap Channels: Secrecy Under Memory Constraints

Yogesh Sankarasubramaniam[*], Andrew Thangaraj[†] and Kapali Viswanathan[*]
[*]Hewlett-Packard Research Labs, Bangalore, India
[†]Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, India
Email:yogesh,kapali@hp.com; andrew@iitm.ac.in

***Abstract*—Information-theoretic security offered by the wiretap channel model has been extensively studied for various scenarios recently. One scenario that has not received much attention is secrecy for systems with memory in the form of input constraints or inter-symbol interference (ISI). In this work, we consider finite state wiretap channels (FSWCs), which model the scenario of secrecy with memory. Using results on secrecy capacity for arbitrary wiretap channels, we first arrive at the secrecy capacity of a FSWC. Then, we develop a stochastic algorithm for computing tight lower bounds on the secrecy capacity of a less-noisy FSWC, and illustrate the computation through examples. Our results provide numerical comparisons between secrecy capacities with and without memory, and provide specific targets for code design.**

I. Introduction

Information-theoretic security in practical physical systems has been an area of intense recent research. The wiretap channel model, originally introduced by Wyner [1], has been studied in several scenarios. The secrecy capacity of a wiretap channel is of primary importance in such studies. Whenever the secrecy capacity is positive, secret information can be sent from a legitimate transmitter to the legitimate receiver over the main channel in the presence of an eavesdropper observing over the wiretapper's channel. Thus, deriving the secrecy capacity and computing its value numerically are important first steps in designing practical codes for security.

In this paper, we introduce the novel scenario of a finite-state wiretap channel (FSWC), establish its secrecy capacity, and provide a method for computing lower bounds on the secrecy capacity. A study of FSWC could potentially open up several possibilities, including embedding secret messages on noisy media such as printed hardcopies, storage of secret messages on magnetic and optical devices, and secret transmission over fiber-optic cables. In most, if not all these cases, a suitable model for the main channel and/or the wiretapper's channel is the finite state channel. Examples of finite state channels include partial response models used in magnetic and optical recording, input-constrained inter-symbol interference (ISI) models, and input runlength constraints.

Computing information-theoretic quantities for finite-state channels with memory is a non-trivial problem [2], [3] with some important recent progress in [4], [5]. The algorithm proposed in this paper for computing lower bounds on the secrecy capacity of a FSWC, is similar in spirit to the stochastic algorithm of [4]. However, the algorithm of [4] cannot be directly used for secrecy capacity, which involves the difference of two mutual information terms. By a series of careful modifications and manipulations, we develop a new stochastic algorithm for computing estimates of secrecy capacity. We study the convergence rate of the algorithm through simulations, and develop lower and upper bounds for the estimates to establish convergence.

The specific contributions of this work are the following: In Section II-B, we characterize the secrecy capacity of a finite-state wiretap channel (FSWC), *i.e.*, a wiretap channel whose component channels are finite-state channels. We provide a stochastic algorithm in Section III for computing lower bounds on the secrecy capacity. To fix ideas, we first consider in Section III-A the example of a $(0,1)$-constrained input, noiseless main channel, and a BSC wiretapper's channel, and then generalize to the less-noisy FSWC in Section III-B.

II. Finite-state Wiretap Channels

The generalized wiretap channel framework, introduced in [6] for discrete memoryless channels (DMCs) has been extended to arbitrary channels recently in [7]. In a wiretap channel setting, the legitimate transmitter and receiver are connected by a main channel, while an eavesdropper observes the channel input over the wiretapper's channel with the same input alphabet as the main channel. In this work, we consider the setting where the main channel and wiretapper's channel are discrete finite-state channels [8], which is a general model for channels with memory. We refer to such a setting as the finite-state wiretap channel (FSWC). Precise definitions and notation are introduced below.

A. Definitions and Notation

A discrete finite-state channel is defined by the transition probability $\Pr\{Y_l S'_l | X_l S'_{l-1}\}$, where $X_l \in \mathcal{X}$ is the input, $S'_l \in \mathcal{S}'$ is the state, and $Y_l \in \mathcal{Y}$ is the output at time l. The alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}'$ are assumed to be finite, and the transition probability is independent of the time index l. We denote such a finite-state channel as $X \rightarrow_{S'} Y$. To simplify notation, let $\mathbf{V}^N = [V_1 \ V_2 \ \cdots V_N] \in \mathcal{V}^N$ for an arbitrary variable $V \in \mathcal{V}$, and let $|\mathcal{V}|$ denote the size of $\mathcal{V}$. The conditional probability $\Pr(\mathbf{Y}^N | \mathbf{X}^N, S'_0 = s'_0)$ across $X \rightarrow_{S'} Y$ can be computed from the transition probability. As usual, $I(X;Y)$ denotes the average mutual information between X and Y.

978-1-4244-4983-5/09/$25.00 © 2009 IEEE

For the finite-state wiretap channel (FSWC), the main channel is characterized by the transition probability $P_1(Y_lS'_l|X_lS'_{l-1}\}$, input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and state alphabet $\mathcal{S}'$, while the wiretapper's channel is characterized by the transition probability $P_2(Z_lS''_l|X_lS''_{l-1}\}$, input alphabet $\mathcal{X}$, output alphabet $\mathcal{Z}$ and state alphabet $\mathcal{S}''$. The common input X_l to the main channel and the wiretapper's channel could, in general, be the output of a finite-state source characterized by the transition probability $P(X_lS_l|S_{l-1}\}$, state alphabet $\mathcal{S}$, and output alphabet $\mathcal{X}$. In such a case, the finite-state joint source-channel model for the wiretap channel (FSJSCW) is defined by the transition probability $\Pr\{X_lY_lZ_l\tilde{S}_l|\tilde{S}_{l-1}\}$, where $\tilde{S} \in \tilde{\mathcal{S}} \triangleq \mathcal{S} \times \mathcal{S}' \times \mathcal{S}''$ is the redefined state variable[1].

B. Secrecy Capacity

Suppose that the legitimate transmitter needs to send a secret message $M \in \mathcal{M}$ over the main channel, while keeping the eavesdropper ignorant rate-wise. A secrecy rate $R_s = \log_2|\mathcal{M}|/N$ is said to be achievable if there exists a series of encodings at the transmitter from $\mathcal{M}$ into $\mathcal{X}^N$ with corresponding decodings from $\mathcal{Y}^N$ into $\mathcal{M}$ producing $\hat{M}$ such that (1) $\Pr\{\hat{M} \neq M\} \to 0$ and (2) $I(M;\mathbf{Z}^N)/N \to 0$ as $N \to \infty$. The secrecy capacity of a wiretap channel is the supremum of all achievable secrecy rates.

The secrecy capacity of an arbitrary wiretap channel with several applications has been determined in [7]. One such application is to the mixed channel, which exactly matches the definition of the FSWC. The main result (adapted from Theorem 3 in [7]) is that the secrecy capacity of a FSWC is given by

$$C_s = \max_{P(\mathbf{V}^N,\mathbf{X}^N)} \left(\min_{s'_0} \lim_{N\to\infty} \frac{1}{N} I(\mathbf{V}^N;\mathbf{Y}^N|S'_0 = s'_0) - \max_{s''_0} \lim_{N\to\infty} \frac{1}{N} I(\mathbf{V}^N;\mathbf{Z}^N|S''_0 = s''_0), \right) \quad (1)$$

where $\mathbf{V}^N$ is a sequence of auxiliary random variables such that $\mathbf{V}^N \to \mathbf{X}^N \to (\mathbf{Y}^N,\mathbf{Z}^N)$ is a Markov chain. The maximization is over possible joint distributions of $\mathbf{V}^N$ and $\mathbf{X}^N$ denoted $P(\mathbf{V}^N,\mathbf{X}^N)$.

If the finite-state channels C1 and C2 are *indecomposable* (see [8] for the exact definition, but most practical finite state channels are indecomposable), the maximization and minimization with respect to the initial states is not necessary. Hence, for indecomposable channels, the secrecy capacity becomes

$$C_s = \max_{P(\mathbf{V}^N,\mathbf{X}^N)} \lim_{N\to\infty} \frac{1}{N} \left(I(\mathbf{V}^N;\mathbf{Y}^N|S'_0 = s'_0) - I(\mathbf{V}^N;\mathbf{Z}^N|S''_0 = s''_0) \right) \quad (2)$$

for arbitrary initial states s'_0 and s''_0 keeping $\mathbf{V}^N \to \mathbf{X}^N \to (\mathbf{Y}^N,\mathbf{Z}^N)$ as a Markov chain.

C. Less-noisy Condition and Simplifications

The formula for secrecy capacity given in (2) involves an auxiliary random variable in the general case. An interesting scenario that simplifies the computation of (2) is the less-noisy characterization. A DMC $X \to Y$ is less-noisy than a DMC $X \to Z$, if for all random variables V such that $V \to X \to (Y,Z)$ is a Markov chain, $I(V;Y) \geq I(V;Z)$. If the main channel is less-noisy than the wiretapper's channel in a DMC wiretap scenario, the auxiliary random variable drops out of the formula for secrecy capacity.

In a similar way, an indecomposable finite-state channel $X \to_{S'} Y$ can be said to be less-noisy than another indecomposable finite-state channel $X \to_{S''} Z$, if there exists a positive integer N_0 so that, for $N > N_0$ and all initial states $(s'_0,\ s''_0)$, $I(\mathbf{V}^N;\mathbf{Y}^N|s'_0) \geq I(\mathbf{V}^N;\mathbf{Z}^N|s''_0)$ for all sequences of auxiliary random variables $\mathbf{V}^N$ such that $\mathbf{V}^N \to \mathbf{X}^N \to (\mathbf{Y}^N,\mathbf{Z}^N)$ is a Markov chain.

Under the above definition for less-noisy finite-state channels, the secrecy capacity formula simplifies in a manner similar to that of the DMC. Therefore, by the same proof as for DMCs, we have that the secrecy capacity of indecomposable FSWCs, whose main channel is less-noisy than the wiretapper's channel, is given by

$$C_s = \max_{P(\mathbf{X}^N)} \lim_{N\to\infty} \frac{1}{N} \left(I(\mathbf{X}^N;\mathbf{Y}^N|S'_0 = s'_0) - I(\mathbf{X}^N;\mathbf{Z}^N|S''_0 = s''_0) \right) \quad (3)$$

for arbitrary initial states s'_0 and s''_0. We do not explicitly write-out the state initialization in the rest of this paper.

III. Algorithm for Computing Tight Lower Bounds on Secrecy Capacity

In most cases, an analytical solution for equation (3) cannot be obtained. This is true even in seemingly simple settings where the main channel is noiseless and the input $\mathbf{X}^N$ is constrained. Thus, a general numeric procedure for computing the secrecy capacity is of much interest here. Though Arimoto-Blahut-like algorithms have recently been proposed for the DMC wiretap setting [9], these algorithms are not applicable for our present case of finite-state wiretap channels. Thus, the remainder of this paper is focused on deriving a numerical procedure to compute tight lower bounds on the secrecy capacity given by (3).

Let us start by observing that the maximization in (3) is over all possible input distributions $P(\mathbf{X}^N)$. Our first step is to restrict this search space by only considering Markov sequences $\mathbf{X}^N$, *i.e.*, $\mathbf{X}^N$ is generated by a Markov source S_M of some memory order M. Next, we provide a stochastic algorithm for optimizing the transition probabilities of the Markov source S_M, so as to maximize the secrecy rate over a given less-noisy FSWC. Let us denote this optimized secrecy rate by $C_s(\mathsf{S}_M)$. Then, $C_s(\mathsf{S}_M)$ is a lower bound on the secrecy capacity C_s in (3).

The basic idea behind our stochastic algorithm is best understood using the following example.

[1] In certain cases, $|\tilde{\mathcal{S}}| = \max\{|\mathcal{S}|,|\mathcal{S}'|,|\mathcal{S}''|\}$ suffices

A. *Example:* $(0, 1)$*-constrained source, noiseless* $C1$*, noisy* $C2$

A (d, k)-constrained sequence is defined as a binary sequence where successive “1”s are separated by at least d and at most k consecutive “0”s. When $d = 0$ and $k = 1$, we obtain a $(0, 1)$-constrained sequence, where runs of consecutive “0”s are prohibited. For example, 1010111 satisfies the $(0, 1)$ constraint, whereas 1001011 violates this constraint.

Consider such a $(0, 1)$-constrained sequence $\mathbf{X}^N$, generated by the 2-state stationary Markov source shown in Fig. 1. Let us denote this source as S_1, since it has memory order 1. We now ask: what is the maximum secrecy rate using S_1, *i.e.*, what is the maximum rate at which Alice can transmit source sequences to Bob while keeping Eve ignorant rate-wise? Here, we assume that the channel between Alice and Bob C1 is noise-free, and that Eve can overhear only a noisy version of Alice’s transmission. For illustration, let us further assume that C2 is a binary symmetric channel (BSC) with cross-over probability p, denoted as BSC(p). The above assumptions will be relaxed in subsequent, more general discussions.

With this set-up, we obtain following (3)

$$C_s(\mathsf{S}_1) = \max_{P_{11}} \lim_{N\to\infty} \frac{1}{N}\left[H(\mathbf{X}^N) - I(\mathbf{X}^N;\mathbf{Z}^N)\right], \qquad (4)$$

where $C_s(\mathsf{S}_1)$ denotes the maximum secrecy rate using Markov source S_1 with state-transition probability P_{11}, as shown in Fig. 1. Clearly, an analytical closed-form solution for (4) is infeasible. Our goal now is to provide a stochastic algorithm for performing the optimization in (4).

To this end, we would like to express (4) in the following form, which would then allow us to use the noisy adjacency matrix technique [4].

$$\begin{aligned} C_s(\mathsf{S}_1) &= \max_{P_{11}} \lim_{N\to\infty} \frac{1}{N}\left[H(\mathbf{X}^N) - I(\mathbf{X}^N;\mathbf{Z}^N)\right] && (5)\\ &= \max_{P_{11}} \lim_{N\to\infty} \left[\frac{1}{N}H(\mathbf{X}^N) - \frac{1}{N}H(\mathbf{Z}^N)\right. \\ &\qquad\qquad \left. + \frac{1}{N}H(\mathbf{Z}^N|\mathbf{X}^N)\right] && (6)\\ &= \max_{P_{11}} \sum_{i,j:(i,j)\in\mathcal{T}} \mu_i P_{ij}\left[\log\frac{1}{P_{ij}} + T_{ij}^{(1)} + T_{ij}^{(2)}\right] && (7) \end{aligned}$$

where μ_i denotes the stationary probability of state i, P_{ij} denotes the transition probability from state i to state j, and $\mathcal{T} = \{(1, 1), (1, 2), (2, 1)\}$ is the set of all valid state transitions for the Markov source S_1 shown in Fig. 1. Note that $\sum_j P_{ij} = 1$ for $i = 1, 2$ so that S_1 is specified by P_{11} alone, and $\mu_j = \sum_i \mu_i P_{ij}$ by the stationarity requirement.

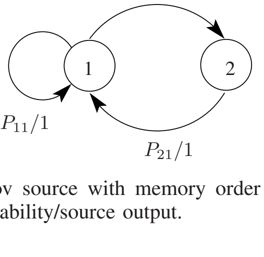

Fig. 1. 2-state Markov source with memory order 1, with branch labels denoting transition probability/source output.

Since C2 is BSC(p), the term $T_{ij}^{(2)}$ in (7) is simply the binary entropy $h(p) = -p\log p - (1-p)\log(1-p)$. The term $T_{ij}^{(1)}$ in (7) can be accurately estimated as follows

$$\hat{T}_{ij}^{(1)} = \frac{\log \Pr(z^N)}{N\mu_i P_{ij}}\left(\frac{1}{N}\sum_{l=1}^{N}\Pr_l(i, j|z^N)\right), \qquad (8)$$

where $\Pr_l(i, j|z^N)$ is used to denote the conditional probability $\Pr(S_{l-1} = i, S_l = j|z^N)$. Using Bayes rule, (8) can be viewed as

$$\hat{T}_{ij}^{(1)} = \frac{1}{N}\left(\frac{1}{N}\sum_{l=1}^{N}\log\Pr(z^N|S_{l-1} = i, S_l = j)\right). \qquad (9)$$

By invoking the law of large numbers along with the ergodicity assumption, we see that $\lim_{N\to\infty}\hat{T}_{ij}^{(1)} = T_{ij}^{(1)}$ and $\lim_{N\to\infty}\sum_{i,j:(i,j)\in\mathcal{T}}\mu_i P_{ij}\hat{T}_{ij}^{(1)} = -\mathcal{H}(Z)$, with probability 1, where $\mathcal{H}(Z) = \lim_{N\to\infty}\frac{1}{N}H(\mathbf{Z}^N)$ is the entropy-rate of the hidden Markov process Z_l. Thus, the estimate $\hat{T}_{ij}^{(1)}$ can be obtained as in (8) by a single long simulation. The quantities $\Pr_l(i, j|z^N)$ and $\log\Pr(z^N)$ can be computed using the Arnold-Loeliger sum-product approach [10], which is a variant of the well-known BCJR algorithm [11].

We now give the stochastic algorithm to compute $C_s(\mathsf{S}_1)$ for the present example.

```
Initialization
  Pick any arbitrary $P_{11}$, $0 < P_{11} < 1$.
  Set $P_{12} = 1 - P_{11}$, $P_{21} = 1$.
Repeat until convergence
    Step 1: For $N$ large, generate $x^N$ as the output of source $\mathsf{S}_1$ (according to the transition probabilities $P_{ij}$) and pass them through C2 to get $z^N$
    Step 2: Run the Arnold-Loeliger modified sum-product algorithm to compute the estimate $\hat{T}_{ij}^{(1)}$
    Step 3: Estimate the noisy adjacency matrix as
        $\hat{A}_{ij} = \begin{cases} 2^{\hat{T}_{ij}^{(1)}} & \text{if } (i,j)\in\mathcal{T} \\ 0 & \text{otherwise,} \end{cases}$
    and find its maximal eigenvalue $\hat{W}_{max}$ and the corresponding eigenvector $\hat{\mathbf{b}} = \left[\hat{b}_1\ \hat{b}_2\right]^T$
    Step 4: Compute the entries for the new transition probability matrix for $(i,j)\in\mathcal{T}$ as $P_{ij} = \frac{\hat{b}_j}{\hat{b}_i}\cdot\frac{\hat{A}_{ij}}{\hat{W}_{max}}$
  end
Compute the secrecy capacity estimate $\hat{C}_s(\mathsf{S}_1) = \log(\hat{W}_{max}) + h(p)$
```

Thus, $C_s(\mathsf{S}_1)$ can be estimated using the given stochastic algorithm. Let us denote by $C_s(0, 1)$, the maximum achievable secrecy rate for noiseless C1, BSC(p) C2, and $(0, 1)$-constrained input. Then, $C_s(\mathsf{S}_1)$ gives a lower bound on $C_s(0, 1)$. For comparison, another lower bound $C_s^l(0, 1)$ can be obtained as follows. Let us assume Alice transmits max-entropic $(0, 1)$ sequences to Bob. Then $H(\mathbf{X}^N)/N$ in (4) is

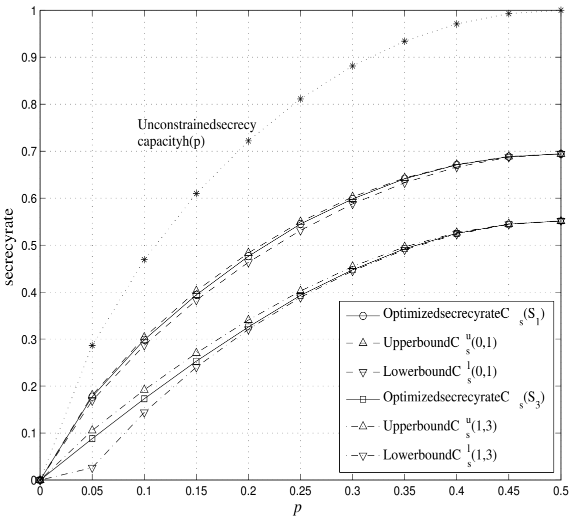

Fig. 2. Optimized secrecy rate for $(0, 1)$ and $(1, 3)$ input constraints as a function of the BSC crossover probability p. Also shown are corresponding upper bounds, lower bounds, and the unconstrained secrecy capacity.

simply the noiseless $(0,1)$-constrained capacity C_{01}. Next, using upper bounds given in [3], we obtain an upper bound on the second term as $I(\mathbf{X}^N;\mathbf{Z}^N)/N \leq C_{ub}(p)$. For the exact computation of $C_{ub}(p)$, see [3]. Since the secrecy capacity is at least as large as the RHS of (4) evaluated for the maxentropic distribution, we obtain $C_s(0,1) \geq C_s^l(0,1) = C_{01} - C_{ub}(p)$.

Our results are shown in Fig. 2. We see that $C_s(\mathsf{S}_1)$ is a better lower bound than $C_s^l(0,1)$. Still better lower bounds can be derived by increasing the memory order of the Markov source, though we observed only marginal improvements for this example.

An upper bound on secrecy capacity, denoted $C_s^u(0,1)$, is also shown in Fig. 2 for comparison. Rewriting (4), we see that the RHS becomes $H(\mathbf{X}^N|\mathbf{Z}^N)/N$, which is upper bounded by $C_s^u(0,1) = H(S_2|S_1, Z_2, Z_3)$ (S_i: state at time i) evaluated at the stationary distribution as shown in [12]. It is seen that $C_s^u(0,1)$ is numerically very close to the lower bound $C_s(\mathsf{S}_1)$, thus suggesting the tightness of our estimate.

The above computation and bounds are readily extended to other (d,k)-constrained inputs. Plots for $d = 1$ and $k = 3$ are shown in Fig. 2. In this case, an order-3 Markov source was used to generate the $(1,3)$-constrained sequences. As before, the secrecy rate $C_s(\mathsf{S}_3)$ was estimated using our stochastic algorithm by optimizing the transition probabilities of this Markov source. The corresponding upper bound $C_s^u(1,3)$ is also shown for comparison. Once again, $C_s(\mathsf{S}_3)$ is seen to be numerically very close to the upper bound, and hence it provides a tight lower bound on the secrecy capacity $C_s(1,3)$.

B. The general less-noisy case

Having understood the basic idea, let us now extend the stochastic algorithm of Section III-A to the general case of indecomposable finite-state channels C1 and C2 which satisfy the less-noisy condition (see Section II-C). For this purpose, we use the finite-state joint source-channel model for the wiretap channel (FSJSCW). As an example, let us consider C1 to be the dicode channel with impulse response 1-D, and C2 to be the EPR4 channel with impulse response 1+D-D^2-D^3. We allow a Markov source of memory order 1. Trellis sections for the Markov source and the dicode main channel C1 are shown in Fig. 3. Trellis sections for the EPR4 wiretapper's channel and the FSJSCW model are shown in Fig. 4.

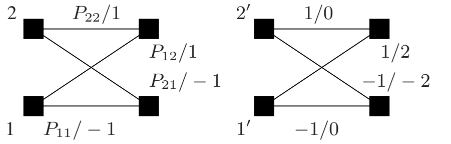

Fig. 3. On the left is order-1 Markov source with $\mathcal{X} = \{-1, 1\}$ and branch labels denoting transition probability/source output; on the right is dicode main channel with branch labels denoting channel input/noiseless output.

Let us now denote the Markov source of Fig. 3(a) by S_1. Proceeding similar to Section III-A, we obtain

$$\begin{aligned}
C_s(\mathsf{S}) &= \max_{P_{ij}:(i,j)\in\mathcal{T}} \lim_{N\to\infty} \frac{1}{N}\left[I(\mathbf{X}^N;\mathbf{Y}^N) - I(\mathbf{X}^N;\mathbf{Z}^N)\right] \\
&= \max_{P_{ij}:(i,j)\in\mathcal{T}} \lim_{N\to\infty} \frac{1}{N}\Big[H(\mathbf{X}^N) - H(\mathbf{X}^N|\mathbf{Y}^N) \\
&\qquad - H(\mathbf{Z}^N) + H(\mathbf{Z}^N|\mathbf{X}^N)\Big] \qquad (10) \\
&= \max_{P_{ij}:(i,j)\in\mathcal{T}} \sum_{i,j:(i,j)\in\mathcal{T}} \mu_i P_{ij}\Big[\log\frac{1}{P_{ij}} + T_{ij}^{(1)} \\
&\qquad + T_{ij}^{(2)} + T_{ij}^{(3)}\Big], \qquad (11)
\end{aligned}$$

where, as before, μ_i denotes the stationary probability of state i, P_{ij} denotes the transition probability from state i to state j, and $\mathcal{T}$ denotes the set of all valid state transitions, but the model now is the FSJSCW. In (11), $T_{ij}^{(1)}$, $T_{ij}^{(2)}$, and $T_{ij}^{(3)}$ are now estimated as follows

$$\hat{T}_{ij}^{(1)} = \frac{1}{N}\sum_{l=1}^{N}\left(\log\frac{\Pr_l(i,j|y^N)^{\frac{\Pr_l(i,j|y^N)}{\mu_i P_{ij}}}}{\Pr_l(i|y^N)^{\frac{\Pr_l(i|y^N)}{\mu_i}}}\right) \qquad (12)$$

$$\hat{T}_{ij}^{(2)} = \frac{\log\Pr(z^N)}{N\mu_i P_{ij}}\left(\frac{1}{N}\sum_{l=1}^{N}\Pr_l(i,j|z^N)\right) \qquad (13)$$

$$\hat{T}_{ij}^{(3)} = -\frac{\log\Pr(z^N|x^N)}{N\mu_i P_{ij}}\left(\frac{1}{N}\sum_{l=1}^{N}\Pr_l(i,j|z^N)\right), \qquad (14)$$

where $\Pr_l(i,j|z^N)$ now denotes the conditional probability $\Pr(\tilde{S}_{l-1} = i, \tilde{S}_l = j|z^N)$, and $\Pr_l(i|y^N)$ denotes the conditional probability $\Pr(\tilde{S}_{l-1} = i|y^N)$. For the derivation of (12), we refer the reader to [4]. Estimates $\hat{T}_{ij}^{(2)}$ and $\hat{T}_{ij}^{(3)}$ are obtained similar to (8) earlier.

By invoking the law of large numbers along with the ergodicity assumption, we see that the above estimates converge to $T_{ij}^{(1)}$, $T_{ij}^{(2)}$, and $T_{ij}^{(3)}$, respectively, with probability 1 as $N \to \infty$. Thus, $\hat{T}_{ij}^{(1)}$ $\hat{T}_{ij}^{(2)}$ and $\hat{T}_{ij}^{(3)}$ can be obtained using a single long simulation of the FSJSCW model. The quantities $\Pr_l(i,j|y^N)$, $\Pr_l(i|y^N)$, $\Pr_l(i,j|z^N)$, and $\log\Pr(z^N)$ can be computed using the Arnold-Loeliger sum-product approach, while the quantity $\Pr(z^N|x^N)$ can be computed using the

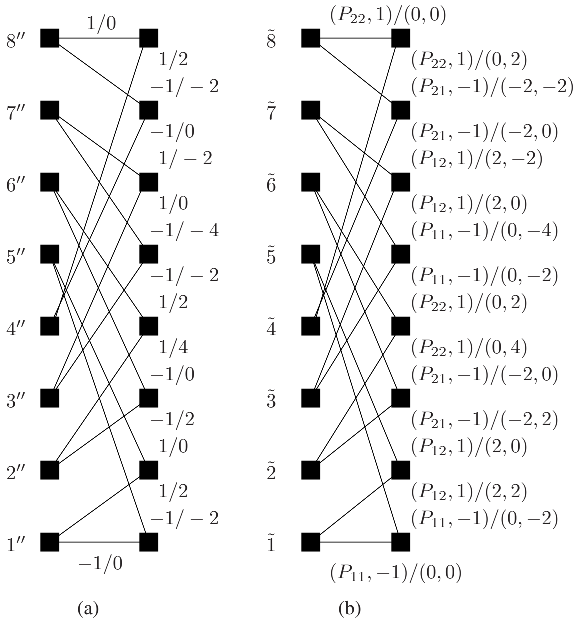

Fig. 4. (a) EPR4 wiretapper's channel with branch labels denoting channel input/noiseless output; (b) FSJSCW model with branch labels denoting (transition probability, source output)/(noiseless main channel output, noiseless wiretapper's channel output).

FSJSCW model and the noise model (AWGN for instance). Finally, the stochastic algorithm to compute $C_s(\mathsf{S})$ can be given as follows.

```
Initialization
  Pick an arbitrary distribution $P_{ij}$ which satisfies the following two conditions
    1) if $(i, j) \in \mathcal{T}$ then $0 < P_{ij} < 1$, else $P_{ij} = 0$
    2) $\sum_j P_{ij} = 1$ for any $i$

  Repeat until convergence
    Step 1: For $N$ large, generate $x^N, y^N, z^N$ according to the FSJSCW model.
    Step 2: Run the Arnold-Loeliger modified sum-product algorithm to compute estimates $\hat{T}_{ij}^{(1)}$, $\hat{T}_{ij}^{(2)}$, and $\hat{T}_{ij}^{(3)}$.
    Step 3: Estimate the noisy adjacency matrix as
      $$\hat{A}_{ij} = \begin{cases} 2^{\hat{T}_{ij}^{(1)} + \hat{T}_{ij}^{(2)} + \hat{T}_{ij}^{(3)}} & \text{if } (i, j) \in \mathcal{T} \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$
      and find its maximal eigenvalue $\hat{W}_{max}$ and the corresponding eigenvector $\hat{\mathrm{b}} = \left[\hat{b}_1 \ \hat{b}_2\right]^T$
    Step 4: Compute the entries for the new transition probability matrix for $(i, j) \in \mathcal{T}$ as $P_{ij} = \frac{\hat{b}_j}{\hat{b}_i} \cdot \frac{\hat{A}_{ij}}{\hat{W}_{max}}$
  end

  Compute the secrecy capacity estimate $\hat{C}_s(\mathsf{S}) = \log(\hat{W}_{max})$
```

To summarize, we have presented a general stochastic algorithm to optimize the transition probabilities of any Markov source S_M, so as to increase the secrecy rate over a given less-noisy FSWC. Let us denote the optimized secrecy rate estimated by our algorithm as $\hat{C}_s(\mathsf{S}_M)$. Our experimental results suggest that for large simulation block lengths N, the estimate $\hat{C}_s(\mathsf{S}_M)$ converges to $C_s(\mathsf{S}_M)$, the maximum secrecy rate achievable by S_M. We have also verified this behavior using brute-force optimization techniques on Markov sources of small memory order.

Further, with increasing M, $C_s(\mathsf{S}_M)$ provides a series of lower bounds on the secrecy capacity C_s in (3). Comparison with outer bounds suggests that these lower bounds are tight. An interesting question is whether $C_s(\mathsf{S}_M)$ actually converges to the secrecy capacity C_s as the memory order M of the Markov source goes to infinity. This behavior has been recently confirmed for finite-state channels [13], and we conjecture that it also holds for finite-state wiretap channels.

IV. Conclusion

In this work, we considered the finite-state wiretap channel (FSWC), which models information-theoretic secrecy scenarios for systems with memory. We characterized the secrecy capacity of a FSWC, and provided a stochastic algorithm for optimizing the secrecy rate of a given Markov source over the less-noisy FSWC. Our results provide accurate numerical estimates of secrecy capacity for several practical scenarios such as storage channels with memory. These estimates can be used as targets for code design for secrecy. On the theoretical side, we conjecture that the optimized secrecy rates actually approach the secrecy capacity as the memory order of the Markov source goes to infinity.

References

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

[2] S. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. on Info. Theory*, vol. 42, no. 5, pp. 1388–1404, September 1996.

[3] S. Shamai and Y. Kofman, "On the capacity of binary and gaussian channels with run-length-limited inputs," *IEEE Trans. on Communications*, vol. 38, no. 5, pp. 584–594, May 1990.

[4] A. Kavcic, "On the capacity of markov sources over noisy channels," in *Proc. of IEEE GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 2997–3001.

[5] P. O. Vontobel, A. Kavcic, D. M. Arnold, and H.-A. Loeliger, "A generalization of the blahut-arimoto algorithm to finite-state channels," *IEEE Trans. on Info Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[6] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. on Info. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[7] M. Bloch and J. N. Lanema, "On the secrecy capacity of arbitrary wiretap channels," in *Proc. of 46th Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2008.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley and Sons, 1968.

[9] K. Yasui, T. Suko, and T. Matsushima, "An algorithm for computing the secrecy capacity of broadcast channels with confidential messages," *ISIT 2007, IEEE International Symp. on Info. Theory*, pp. 936–940, June 2007.

[10] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. of ICC 2001*, 2001, pp. 2692–2695.

[11] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Info. Theory*, vol. 20, pp. 284–287, March 1974.

[12] E. Zehavi and J. Wolf, "On runlength codes," *Information Theory, IEEE Transactions on*, vol. 34, no. 1, pp. 45–54, Jan 1988.

[13] J. Chen and P. H. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *IEEE Trans. on Info. Theory*, vol. 54, no. 3, pp. 1295–1303, March 2008.