A Unified Framework for Key Agreement Over Wireless Fading Channels

Conference Paper *in* IEEE Transactions on Information Forensics and Security · November 2009

DOI: 10.1109/ITW.2009.5351257 · Source: IEEE Xplore

H. Vincent Poor
Princeton University

A Unified Framework for Key Agreement over Wireless Fading Channels

Lifeng Lai, Yingbin Liang and H. Vincent Poor

Abstract—**The problem of key generation over wireless fading channels is investigated. First, a joint source-channel approach that combines existing source and channel models for key agreement over wireless fading channels is developed. It is shown that, in general, to fully exploit the resources provided by time-varying channel gains, one needs to combine both the channel model, in which Alice sends a key to Bob over wireless channels, and the source model, in which Alice and Bob generate a key by exploiting the correlated observations obtained from the wireless fading channels. Asymptotic analyses suggest that in the long coherence time regime, the channel model is asymptotically optimal. On the other hand, in the high power regime, the source model is asymptotically optimal. Second, the framework is extended to the scenario with an active attacker whose goal is to minimize the key rate that can be generated using our protocol. The attacker's optimal attack strategy is identified and the key rate under this attack model is characterized.**

I. Introduction

Recently, the study of security from an information theoretic perspective has attracted considerable attention. (See [3] for a recent review of results in this area.) In this paper, we focus on the problem of key agreement over wireless fading channels, in which two terminals, Alice and Bob, connected by a wireless fading channel wish to establish a key through the wireless channel while keeping the key secret from an eavesdropper Eve. The goal is to establish a key with a rate as large as possible under the constraint that the observations at Eve do not provide any information about the generated key.

There are two lines of previous work relating to key agreement over fading channels: that concerned with the channel model, and that concerned with the source model. In the channel model studied in [4] and [5][1], the time-varying channel gain from Alice to Bob is assumed to be known by all parties, namely Alice, Bob and Eve. The ability to transmit information securely relies on a non-zero probability that the channel gain from Alice to Bob is larger than the channel gain from Alice to Eve. In the source model studied in [6]–[11], the channel gain from Alice to Bob is assumed to be unknown everywhere *a priori*. Alice and Bob each estimates the unknown channel gain. In this way, Alice and Bob obtain correlated observations that can then be used to generate keys using the key generation from common randomness method introduced in [12].

There are two main limitations of the existing studies. First, each of the channel model and the source model successfully exploits only one aspect of the resources provided by the varying channel gains. More specifically, the channel model exploits the possibility of a larger channel gain at the receiver while the source model exploits the fact that Eve does not know the channel gain from the source to the destination. However, the channel model does not exploit the opportunity provided by the fact that Eve does not know the channel gain from Alice to Bob. As a result, the key rate generated using the channel model saturates even if the available transmit power goes to infinity [4], [5]. On the other hand, the source model does not exploit the possibility that the channel gain from Alice to Bob might be better than the channel gain from Alice to Eve. Hence, the key rate generated using the source model goes to zero when the coherence time of the channel increases [6].

Second, in all these studies, it is assumed that the attacker is *passive*, meaning that it only overhears (does not transmit over) the channel and tries to infer information about the generated key. This assumption implies that the messages exchanged between Alice and Bob are authenticated and will not be modified by the attacker. In reality, an *active* attacker might modify the messages exchanged between Alice and Bob. For example, when Alice and Bob try to learn the channel gain, Eve can send attack signals to make the channel estimation imprecise. Similarly, when Alice and Bob exchange information over the channel, Eve can modify the message exchanged over the wireless channel. The problem of key generation over an unauthenticated channel has been studied in [13]–[15]. These papers assumed that the attacker can *completely* block the communication link between Alice and Bob. Under this assumption, these papers developed a key agreement protocol that allows these two terminals to achieve the following two goals: 1) In the time slots when the active attack occurs, the two terminals can detect the presence of the attack with a probability close to 1; 2) In the time slots when the active attack does not occur, the two terminals can establish a key with a rate equal to the rate that one can achieve

L. Lai and H. V. Poor's work was supported by the National Science Foundation under Grants CNS-09-05398 and CCF-10-16671. Y. Liang's work was supported by the National Science Foundation CAREER Award under Grant CCF-10-26565 and by the National Science Foundation under Grant CCF-10-26566.

Lifeng Lai is with Department of Systems Engineering, University of Arkansas, Little Rock, AR 72204. Email: lxlai@ualr.edu. Yingbin Liang is with Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244. Email: yliang06@syr.edu. H. Vincent Poor is with Department of Electrical Engineering, Princeton University, Princeton, NJ 08544. Email: poor@princeton.edu.

Parts of this paper have been presented at the IEEE Information Theory Workshop, Taormina, Sicily, Italy, Oct. 2009 [1] and at the Forty-Eighth Annual Allerton Conference on Communication, Control, and Computing Sept./Oct. 2010, Monticello, Illinois [2].

[1]These papers consider the transmission of a secret message from Alice to Bob. If Alice uses this secret message as a secret key, then the schemes in these papers can be used for key agreement purposes.

as if the attacker is passive. Obviously, if the attacker chooses to attack all the time, these two terminals will not be able to establish a key under this model. The main reason for this *pessimistic* result lies in the assumption that the attacker can completely block the communication link between Alice and Bob. Hence when an active attack occurs, what the receiver receives comes purely from the attacker. However, in wireless communications it is difficult, if not impossible, to completely block a communication link. Hence, even if the attack occurs, the receiver will still be able to receive signals from the transmitter (although the received signal will be corrupted by signals from the attacker).

In this paper, we develop key agreement algorithms that address these two issues. We first develop a joint source-channel approach that combines the existing channel model and source model for the key generation. As a result, one can design a scheme that can exploit the advantages provided by both of these two models. Our key agreement protocol has two phases. In the first phase, Alice and Bob send training signals over the channel alternately, and obtain an estimate of their respective channel gains. In the second phase, Alice sends an auxiliary message, which will be used to distill a key from the correlated observations obtained in the first phase, and a new randomly generated key to Bob. The total key rate is the sum of the key rate that can be generated from the correlated observations and the rate of the newly generated key. Our asymptotic analysis suggests that the channel model is asymptotically optimal as the coherence time of the channel becomes long. On the other hand, in the high power regime, the source model is asymptotically optimal. We note that the idea of sending artificial noise can also be incorporated into our work. However, it is more suitable to send artificial noise if there are relays [16] or if we consider feedback [17], or if we consider multiple antennas [18], [19]. In the single antenna case as considered in this paper, sending artificial noise may lead to performance loss.

We then extend this approach to the case of an active attacker, whose goal is to minimize the key rate that can be generated using our key agreement protocol. The attacker can design the signal it transmits based on the signal overheard over the channel. We first characterize the attacker's optimal attack strategy to our protocol. In this paper, we assume that the attacker uses an independently and identically distributed attack strategy. We note that the active attacker considered in this work is more benign than those considered in the arbitrary varying channels [20], [21]. The study of more advanced attack models is an interesting topic for future work. We show that during the first phase, the optimal attack strategy is to send correlated Gaussian random signals. During the second phase, the optimal attack strategy is to send a Gaussian jamming signal. We then characterize the key rate that can be generated from the fading wireless channel in the presence of an attacker that employs the optimal attack strategy. With this approach, Alice and Bob can establish a key over the wireless fading channels even in the presence of an active attacker under certain circumstances.

The remainder of the paper is organized as follows. In Section II, we introduce the model under study. In Section III, we develop our joint source-channel approach for the key generation. In Section IV, we extend our protocol to the case of an active attacker and study the corresponding performance. Finally, we present concluding remarks and point out possible future directions in Section V.

II. Model

Two terminals Alice (A) and Bob (B) wish to agree on a key through a wireless fading channel in the presence of an active attacker Eve (E). All three terminals can transmit over the wireless channel. We assume that Alice and Bob are half-duplex nodes, while the attacker is a full-duplex node. In this paper, we assume that the goal of the attacker is to minimize the key rate generated by Alice and Bob from the wireless channel. The attacker can receive a noisy version of the signal transmitted by the legitimate terminals. In addition, it can transmit signals to contaminate the signal transmitted by the legitimate users. In particular, if Alice transmits X_A in a given channel use, then Bob and Eve receive

$$\begin{aligned} Y_B &= h_{AB}X_A + X_{E1} + N_B \quad (1)\\ \text{and} \quad Y_E &= h_{AE}X_A + N_E, \quad (2)\end{aligned}$$

respectively, in which h_{AB} is the channel gain from Alice to Bob, X_E is the signal transmitted by the Eve, N_B is zero mean Gaussian noise with variance σ^2, h_{AE} is the channel gain from Alice to Eve, and N_E is zero mean Gaussian noise with variance σ^2. We assume that there is no fading in the links from Eve to Alice and Bob. Alternatively, if Bob transmits X_B in a given channel use, then Alice and Eve receive

$$\begin{aligned} Y_A &= h_{BA}X_B + X_{E2} + N_A \quad (3)\\ \text{and} \quad Y_E &= h_{BE}X_B + N_E, \quad (4)\end{aligned}$$

respectively, in which h_{BA} is the channel gain from Bob to Alice, N_A is zero mean Gaussian noise with variance σ^2, and h_{BE} is the channel gain from Bob to Eve. We note that the analysis can be easily carried out to the case in which the noise variance of N_A is different from that of N_B. Again, X_E is the attack signal from the attacker. We assume that N_A, N_B and N_E are independent of each other. We note that in the model considered in [13]–[15], $Y_B = X_{E1}$ and $Y_A = X_{E2}$ (i.e., if there is an active attack, the receiver receives a signal only from the attacker).

We assume that the channel is reciprocal, that is $h_{AB} = h_{BA}$. Due to different transmission paths, h_{AB} is independent of h_{AE} and h_{BE}. We consider an ergodic block fading model, in which the channel gains are fixed for a block of T symbols and change to other values at the beginning of the next block. In this paper, we assume $h_{AB} \sim \mathcal{N}(0, \sigma_h^2)$ and $h_{AE} \sim \mathcal{N}(0, \sigma_{AE}^2)$. We assume that none of the terminals knows the value of the fading gains. The noise processes are assumed to be independent and identically distributed (i.i.d.) over channel uses and terminals.

Let $\mathbf{X}_A = [X_A(1), \cdots, X_A(N)]^T$ and $\mathbf{X}_B = [X_B(1), \cdots, X_B(N)]^T$ denote codewords sent by Alice and Bob respectively, and $\mathbf{X}_E$ be the attack signal sent by Eve, over N uses of the channel. Here N could be larger than the channel coherence time T; that is a codeword can span multiple coherence blocks. Let $\mathbf{Y}_A = [Y_A(1), \cdots, Y_A(N)]^T$, $\mathbf{Y}_B = [Y_B(1), \cdots, Y_B(N)]^T$ and $\mathbf{Y}_E = [Y_E(1), \cdots, Y_E(N)]^T$ denote corresponding observations at Alice, Bob and Eve, respectively. Since we have a half-duplex constraint on the legitimate users, $Y_A(i) = \phi$ when $X_A(i) \neq \phi$. Here ϕ denotes either no observation or no transmission. Similarly, $Y_B(i) = \phi$ when $X_B(i) \neq \phi$. To make a fair comparison to schemes in which only one terminal transmits, we have a total power constraint, that is

$$\frac{1}{N}\mathbb{E}\{\mathbf{X}_A^T\mathbf{X}_A + \mathbf{X}_B^T\mathbf{X}_B\} \leq P. \tag{5}$$

We also assume that the attacker has an average power constraint P_E, that is

$$\frac{1}{N}\mathbb{E}\{\mathbf{X}_E^T\mathbf{X}_E\} \leq P_E. \tag{6}$$

Both Alice and Bob will generate a key based on the sequence it sends and signals it receives from the wireless channel. Let f_A and f_B denote the key generation functions at Alice and Bob, respectively, so that $K_A = f_A(\mathbf{X}_A, \mathbf{Y}_A)$ and $K_B = f_B(\mathbf{X}_B, \mathbf{Y}_B)$. A key rate R_{key} is said to be achievable if for each $\epsilon > 0$, there exists n_0 such that for each $N \geq n_0$ we have that

$$\Pr(K_A \neq K_B) \leq \epsilon, \tag{7}$$

$$\frac{1}{N}H(K_A) \geq R_{key} - \epsilon, \tag{8}$$

$$\frac{1}{N}I(K_A;\mathbf{Y}_E,\mathbf{X}_E) \leq \epsilon, \quad \text{and} \tag{9}$$

$$H(K_A) \geq \log|K_A| - \epsilon. \tag{10}$$

III. A Joint Source-Channel Key Agreement Protocol

In this section, we develop a joint source-channel key agreement protocol. Here, we assume that the eavesdropper is passive, i.e., $\mathbf{X}_E = \mathbf{0}$. We first consider a scenario in which there exists a public channel, through which both Alice and Bob can exchange messages. All messages transmitted over the public channel will be overheard by Eve noiselessly. The scheme developed in this scenario provides insights for a more realistic scenario in which there is no public channel available. We then consider this more realistic model. In both cases, key agreement schemes that benefit from both the source model and the channel model are developed. In both scenarios, asymptotic analyses suggest that the channel model is asymptotically optimal as the coherence time of the channel becomes long. On the other hand, in the high power regime, the source model is asymptotically optimal. We also find that, in the asymptotic regime, either in long coherence time or high power, the achievable key rate without the public channel is the same as that we can achieve when there is a public channel.

A. Key Agreement With a Public Channel

To assist in the presentation, we first consider a scenario in which, in addition to the wireless channel, there is a public channel with infinite capacity. This scenario will provide insights for a more realistic scenario in which there is no public channel available. Both Alice and Bob can transmit over this public channel, and Eve can overhear any messages exchanged over this public channel. In this scenario, the key generation functions at Alice and Bob can also depend on the communications that have taken place over the public channel. Let $\mathbf{C}$ be the collection of messages exchanged over the public channel; then $K_A = f_A(\mathbf{X}_A, \mathbf{Y}_A, \mathbf{C})$ and $K_B = f_B(\mathbf{X}_B, \mathbf{Y}_B, \mathbf{C})$. Now, Eve observes both $\mathbf{Y}_E$ and $\mathbf{C}$, and hence we require that the mutual information between the generated key and $(\mathbf{Y}_E, \mathbf{C})$ should be small; that is

$$\frac{1}{N}I(K_A;\mathbf{Y}_E,\mathbf{C}) \leq \epsilon.$$

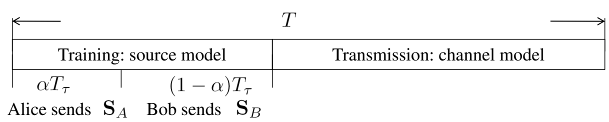

Fig. 1: Training based scheme.

We consider a training based scheme as shown in Figure 1. In this training based scheme, Alice and Bob first obtain an estimate of their channel gain through training. That is, at the beginning of each block, Alice sends a known training sequence to the wireless channel, Bob obtains an estimate of the channel gain, and then Bob sends a known training sequence to the wireless channel from which Alice obtains an estimate of the channel gain. These two estimates will not be the same, but will be correlated. Eve can also estimate its channel, but the observations at Eve will be independent of the observations at both Alice and Bob because of the independence of the noise processes and fading gains. Then Alice and Bob generate a key from these correlated observations with the assistance of the public channel. After the training phase, Alice also sends another randomly generated key using the noisy wireless channel. Let T_τ denote the amount of time spent on training, and let $T - T_\tau$ denote the amount of time that is used in the second stage.

Suppose Alice sends a known sequence $\mathbf{S}_A$ of size $1 \times \alpha T_\tau$, with $0 < \alpha < 1$. Bob receives

$$\mathbf{Y}_{B,\tau} = h_{AB}\mathbf{S}_A + \mathbf{N}_B, \tag{11}$$

where $\mathbf{N}_B = [N_B(1), \cdots, N_B(\alpha T_\tau)]^T$. After that, Bob sends a known sequence $\mathbf{S}_B$ of size $1 \times (1-\alpha)T_\tau$ over the wireless channel, and Alice receives

$$\mathbf{Y}_{A,\tau} = h_{AB}\mathbf{S}_B + \mathbf{N}_A, \tag{12}$$

where $\mathbf{N}_A = [N_A(1), \cdots, N_A((1-\alpha)T_\tau)]^T$.

Alice and Bob use $\mathbf{Y}_{A,\tau}$ and $\mathbf{Y}_{B,\tau}$ in the following two ways: (1) to generate a key from these two correlated observations using the source model through the public channel; and (2) to generate an estimate of the channel gain h_{AB} in the given coherence block, which will be used for the key generation using the channel model.

1) Key generation from the training phase: We first look at the key generation using the source model. Alice computes a sufficient statistic $\tilde{Y}_A$ for $\mathbf{Y}_{A,\tau}$ via

$$\tilde{Y}_A = \frac{\mathbf{S}_B^T}{||\mathbf{S}_B||}\mathbf{Y}_{A,\tau} = h_{AB} + \frac{\mathbf{S}_B^T}{||\mathbf{S}_B||}\mathbf{N}_A, \quad (13)$$

in which $||\cdot||$ denotes the norm of its argument. Similarly, Bob computes a sufficient statistic $\tilde{Y}_B$ for $\mathbf{Y}_{B,\tau}$ via

$$\tilde{Y}_B = \frac{\mathbf{S}_A^T}{||\mathbf{S}_A||}\mathbf{Y}_{B,\tau} = h_{AB} + \frac{\mathbf{S}_A^T}{||\mathbf{S}_A||}\mathbf{N}_B. \quad (14)$$

Note that $\tilde{Y}_A$ is a zero mean Gaussian random variable with variance $\sigma_h^2 + \frac{\sigma^2}{||S_B||}$, and similarly $\tilde{Y}_B$ is a zero mean Gaussian random variable with variance $\sigma_h^2 + \frac{\sigma^2}{||S_A||}$. Assuming that Alice and Bob transmit with power P_τ during the training period, we have $||S_B|| = (1-\alpha)T_\tau P_\tau$ and $||S_A|| = \alpha T_\tau P_\tau$.

We first have the following observation showing that $\tilde{Y}_A$ and $\tilde{Y}_B$ retain the mutual information between $\mathbf{Y}_{A,\tau}$ and $\mathbf{Y}_{B,\tau}$; i.e., they are sufficient for the key generation purpose.

Lemma 3.1:

$$I(\tilde{Y}_A;\tilde{Y}_B) = I(\mathbf{Y}_{A,\tau};\mathbf{Y}_{B,\tau}). \quad (15)$$

Proof: It is easy to see that the following Markovian relationship is true:

$$\tilde{Y}_A \longleftrightarrow \mathbf{Y}_{A,\tau} \longleftrightarrow h_{AB} \longleftrightarrow \mathbf{Y}_{B,\tau} \longleftrightarrow \tilde{Y}_B, \quad (16)$$

which implies $I(\tilde{Y}_A;\tilde{Y}_B) \le I(\mathbf{Y}_{A,\tau};\mathbf{Y}_{B,\tau})$. Similarly, from the Markovian relationship

$$\mathbf{Y}_{A,\tau} \longleftrightarrow \tilde{Y}_A \longleftrightarrow h_{AB} \longleftrightarrow \tilde{Y}_B \longleftrightarrow \mathbf{Y}_{B,\tau}, \quad (17)$$

we have $I(\tilde{Y}_A;\tilde{Y}_B) \ge I(\mathbf{Y}_{A,\tau};\mathbf{Y}_{B,\tau})$. Hence, $I(\tilde{Y}_A;\tilde{Y}_B) = I(\mathbf{Y}_{A,\tau};\mathbf{Y}_{B,\tau})$. ∎

From $(\tilde{Y}_A,\tilde{Y}_B)$ one can generate a key with rate [12]

$$\begin{aligned} R_s &= \frac{1}{T}I(\tilde{Y}_A;\tilde{Y}_B) & (18)\\ &= \frac{1}{2T}\log\left(\frac{(\sigma^2+\sigma_h^2\alpha P_\tau T_\tau)(\sigma^2+\sigma_h^2(1-\alpha)P_\tau T_\tau)}{\sigma^4+\sigma^2\sigma_h^2 P_\tau T_\tau}\right), & (19)\end{aligned}$$

in which the normalization factor $1/T$ comes from the fact that the channel gain is fixed for T symbols, meaning that we can observe only one value of $(\tilde{Y}_A,\tilde{Y}_B)$ for every T symbols. To generate a key with such a rate, one can use the standard Slepian-Wolf coding scheme [12]. More precisely, for every N symbol times, Alice has $m = \lfloor N/T \rfloor$ observations of the random variable $\tilde{Y}_A$. Here $\lfloor\cdot\rfloor$ is the largest integer that is smaller than its argument. These observations are collected into a vector $\tilde{\mathbf{Y}}_A = [\tilde{Y}_A(1),\cdots,\tilde{Y}_A(m)]^T$. Here, the $\tilde{Y}_A(i)$s are independent of each other. Similarly, Bob has a vector of observations $\tilde{\mathbf{Y}}_B = [\tilde{Y}_B(1),\cdots,\tilde{Y}_B(m)]^T$. Alice randomly divides the typical $\tilde{Y}_A$ sequences into non-overlapping bins, with each bin having $2^{mI(\tilde{Y}_A;\tilde{Y}_B)}$ typical $\tilde{Y}_A$ sequences. Hence, each sequence has two indices: bin number and index within the bin. Now, after observing the vector $\tilde{\mathbf{Y}}_A$, Alice sets the key value as the index of this sequence within each bin and sends the bin number to Bob through the public channel. That is, Alice needs to send $H(\tilde{Y}_A|\tilde{Y}_B)$ bits of information through the public channel. After combining the information observed from the public channel with $\tilde{\mathbf{Y}}_B$, it can be shown that Bob can recover the value of $\tilde{\mathbf{Y}}_A$ with probability arbitrarily close to 1. Then Bob can recover the value of the key. At the same time, it can be shown that the bin number and index within each bin are independent of each other. Hence, even though the eavesdropper can observe the bin number transmitted over the public channel, it learns nothing about the generated key. We note here that the codebook information is public, i.e., everyone including the attacker knows the codebook information.

2) Key generation after the training phase: After the training period of T_τ symbols, Alice can send another randomly generated key to Bob using the scheme developed for the fading eavesdropper channel [4]. More specifically, Bob obtains a Minimum Mean Square Error (MMSE) estimate $\hat{h}_{AB}$ of the channel gain h_{AB} in the given coherence block,

$$\hat{h}_{AB} = \frac{\sigma_h^2}{\sigma^2+\alpha P_\tau T_\tau \sigma_h^2}\mathbf{S}_A^T\mathbf{Y}_{B,\tau}, \quad (20)$$

and treats this as the true value of the channel gain. We can write

$$h_{AB} = \hat{h}_{AB} + \bar{h}_{AB},$$

in which $\bar{h}_{AB}$ is the estimation error. It is easy to verify that $\bar{h}_{AB}$ is a zero mean Gaussian random variable with variance $\sigma_h^2/(\sigma_h^2\alpha P_\tau T_\tau + \sigma^2)$.

We consider a simple scheme in which Alice does not perform power control or rate control. Clearly, one can improve this rate by allowing Alice to adapt her transmission scheme based on its estimate of the channel. But this simple strategy allows us to decouple the key generation problem in these two stages. If Alice adapts her transmission scheme based on its estimated channel gain, the eavesdropper might be able to learn some information about the channel gain h_{AB} during the second stage, which complicates the key generation from the source model. Alice sends a key to Bob, using a constant power P_d. Then the following secrecy rate is achievable [4]:

$$\begin{aligned} R_{ch} &= \frac{T-T_\tau}{T}[I(X_A;Y_B|\hat{h}_{AB}) - I(X_A;Y_E|h_{AE})]^+ & (21)\\ &= \frac{T-T_\tau}{2T}\Bigg[\mathbb{E}\Bigg\{\log\Bigg(1+\frac{\hat{h}^2P_d}{\sigma^2+\frac{\sigma_h^2P_d}{\sigma_h^2\alpha P_\tau T_\tau+\sigma^2}}\Bigg) \\ &\qquad -\log\left(1+\frac{h_{AE}^2P_d}{\sigma^2}\right)\Bigg\}\Bigg]^+, & (22)\end{aligned}$$

in which $[x]^+ = \max\{x,0\}$. Here, the first term is the rate that Bob can decode using a mismatched decoder [22], [23]. The second term is an upper-bound on the mutual information that Eve can accumulate. We obtain this upper-bound by assuming that Eve has perfect knowledge of h_{AE}. We note here that Alice and Bob do not need to know the instantaneous value of h_{AE}.

In summary, we have the following result.

Theorem 3.2: In a wireless fading channel with a public channel, the following secret key rate is achievable using the training based scheme:

$$R_{key} = \max_{\alpha, P_\tau, T_\tau} \{R_s + R_{ch}\} \quad (23)$$
$$\text{s.t.} \quad T_\tau P_\tau + (T - T_\tau) P_d \leq TP, \quad (24)$$

in which R_s and R_{ch} are given by (18) and (21), respectively.

One can optimize the key rate by choosing appropriate values of α, P_τ and T_τ. If T_τ is small, one has more time left for transmitting a key using the channel model. But the estimates of channel gain at Alice and Bob will be coarse, which will affect both key generation processes using the source model and the channel model. On the other hand, if T_τ is large, one can generate a larger key rate using the source model, since the estimates of the channel at Alice and Bob are more precise. But, in this case the time left for sending a key from Alice to Bob is reduced. For general values of the available power P and the coherence length T, it is difficult to obtain closed form expressions for the optimal values of α, P_τ and T_τ. In the following, we consider two asymptotic regimes to gather insight into the behavior of these quantities.

1) *Long coherence time regime*, in which $T \to \infty$.

We have the following inequalities, which can be verified easily:

$$\begin{aligned} & R_s \\ & \leq \max_{\alpha, P_\tau, T_\tau} \frac{1}{2T} \log \left(\frac{(\sigma^2 + \sigma_h^2 \alpha P_\tau T_\tau)(\sigma^2 + \sigma_h^2 (1-\alpha) P_\tau T_\tau)}{\sigma^4 + \sigma^2 \sigma_h^2 P_\tau T_\tau} \right) \\ & \leq \frac{1}{2T} \log \left(\frac{(\sigma^2 + \frac{1}{2}\sigma_h^2 PT)^2}{\sigma^4 + \sigma^2 \sigma_h^2 PT} \right) \quad (25) \\ & \leq \frac{1}{2T} \log \left(\frac{\sigma^2}{\sigma_h^2 PT} + 1 + \frac{1}{4\sigma^2} \sigma_h^2 PT \right). \quad (26) \end{aligned}$$

Thus, as $T \to \infty$, $R_s \to 0$. That is, in this regime, the channel model is asymptotically optimal. As a result, to maximize R_{key}, we can choose α, P_τ and T_τ to maximize R_{ch}. It easy to see that we should set $\alpha = 1$; that is, only Alice sends a training sequence, since even if Bob sends a training sequence, the key rate that we can generate from the correlated observations will be zero.

2) *High power regime*, in which $P \to \infty$.

Let us examine the R_{ch} term:

$$\begin{aligned} & R_{ch} \quad (27) \\ & = \max_{\alpha, T_\tau, P_\tau} \frac{T - T_\tau}{2T} \Bigg[\mathbb{E} \Bigg\{ \log \left(1 + \frac{\hat{h}^2 P_d}{\sigma^2 + \frac{\sigma_h^2 P_d}{\sigma_h^2 \alpha P_\tau T_\tau + \sigma^2}} \right) \\ & \qquad - \log \left(1 + \frac{h_{AE}^2 P_d}{\sigma^2} \right) \Bigg\} \Bigg]^+ \quad (28) \\ & \leq \max_{P_d} \frac{1}{2} \Bigg[\mathbb{E} \Bigg\{ \log \left(1 + \frac{h_{AB}^2 P_d}{\sigma^2} \right) \\ & \qquad - \log \left(1 + \frac{h_{AE}^2 P_d}{\sigma^2} \right) \Bigg\} \Bigg]^+ \quad (29) \\ & \leq \max_{P_d} \mathbb{E}_{\{h_{AB}^2 \geq h_{AE}^2\}} \Bigg\{ \log \left(1 + \frac{h_{AB}^2 P_d}{\sigma^2} \right) \\ & \qquad - \log \left(1 + \frac{h_{AE}^2 P_d}{\sigma^2} \right) \Bigg\} \quad (30) \\ & \leq \mathbb{E}_{\{h_{AB}^2 \geq h_{AE}^2\}} \left\{ \log \left(\frac{h_{AB}^2}{h_{AE}^2} \right) \right\} \quad (31) \\ & = \int_0^\infty \log(h_{AB}^2) f(h_{AB}^2) dh_{AB}^2 \\ & \quad - \int_0^\infty \int_0^{h_{AB}^2} \log(h_{AE}^2) f(h_{AE}^2) f(h_{AB}^2) dh_{AB}^2 dh_{AE}^2 \quad (32) \\ & \leq \int_0^\infty h_{AB}^2 f(h_{AB}^2) dh_{AB}^2 \\ & \quad - C_1 \int_0^1 \log(h_{AE}^2) f(h_{AE}^2) dh_{AE}^2 \quad (33) \\ & \leq \mathbb{E}\{h_{AB}^2\} + C_1 C_2, \quad (34) \end{aligned}$$

in which $C_1 = \sup f(h_{AB}^2)$ and $C_2 = \sup f(h_{AE}^2)$, and the last equation is due to the fact that

$$\left| \int_0^1 \log x dx \right| = 1.$$

Hence, the R_{ch} term is bounded by a constant when P increases. On the other hand, it is easy to see that the R_s term increases with P. Thus, in the high power regime, the source model is asymptotically optimal. As a result, in order to maximize the key rate, we choose the parameters to maximize R_s. Simple calculation shows that the optimal value parameters are $\alpha = 1/2$, $P_\tau = P$ and $T_\tau = T$. As a result,

$$R_{key} \sim \frac{1}{2T} \log P.$$

Hence, in the high power regime, if the coherence time is fixed, the secrecy rate increases logarithmically with P.

B. Key Agreement Without a Public Channel

In this section, we study a more realistic scenario in which there is no public channel available. Similarly to the development in Section III-A, we consider a training based scheme, in which both Alice and Bob send training sequences over the wireless channel during the training period. Then, Alice and Bob generate a key from the correlated observations using the source model. Alice also sends another randomly generated key to Bob after the training period using the channel model.

Hence, the total key rate that can be generated from the wireless channel is the sum of the two key rates.

If there is no public channel, the key generation problem using the channel model is the same as that of Section III-A, since no public resources were used. On the other hand, due to the absence of the public channel, the key generation process from the correlated observations should be revised. As discussed in Section III-A, to generate a key with a rate of $I(\tilde{Y}_A;\tilde{Y}_B)/T$ from the correlated estimates of the channel gain, Alice needs to send $H(\tilde{Y}_A|\tilde{Y}_B)$ bits of information (more precisely, the bin number of its observations) to Bob. Since $\tilde{Y}_A$ and $\tilde{Y}_B$ are continuous random variables, $H(\tilde{Y}_A|\tilde{Y}_B)$ is infinite. If there is a public channel with infinite capacity, this is not an issue. If there is no public channel, one has to send the bin number over the wireless channel. Since the wireless channel has limited capacity, the key rate that one can generate from these correlated observations is less than $I(\tilde{Y}_A;\tilde{Y}_B)/T$.

The problem of key generation from correlated sources through a public channel with limited capacity has been studied in [24]. More precisely, if the public channel has a rate constraint R, then the following secret key rate can be generated from the correlated observations $(\tilde{Y}_A,\tilde{Y}_B)$:

$$R_s = I(U;\tilde{Y}_B) \tag{35}$$

$$\text{s.t.} \quad U \to \tilde{Y}_A \to \tilde{Y}_B, \tag{36}$$

$$\text{and} \quad I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B) \le R, \tag{37}$$

where U is an auxiliary random variable subject to the Markov chain relationship given to it in (16).

Furthermore, this rate can be achieved by sending from Alice only. Roughly speaking, we generate $2^{mI(U;\tilde{Y}_A)}$ typical U sequences. We then divide these typical sequences into bins, each bin containing $2^{mI(U;\tilde{Y}_B)}$ sequences. Hence, each U^m sequence can be specified by two indices: the bin number (ranging from 1 to $2^{m(I(U;\tilde{Y}_A)-I(U;\tilde{Y}_B))}$), and the index of the sequence within each bin. Now, after observing $\tilde{\mathbf{Y}}_A = [\tilde{Y}_A(1),\cdots,\tilde{Y}_A(m)]^T$, Alice finds a U^m sequence that is jointly-typical with $\tilde{\mathbf{Y}}_A$. (This step will be successful with very high probability.) Alice sets the key value as the index of the sequence in the bin and sends the bin number to Bob, which requires a rate of $I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)$. This rate can be accommodated by the public channel since the capacity of the public channel is larger than this rate requirement. After receiving the bin number, Bob obtains an estimate $\hat{U}^m$ by looking for a unique sequence in the bin specified by the bin number that is jointly typical with its observation $\tilde{\mathbf{Y}}_B$. $\hat{U}^m$ will be equal to U^m with probability 1, thus Bob can then recover the key value.

Now, if we do not have a public channel at our disposal, we can use the wireless channel after the training stage to send the bin number needed for the key generation from the correlated observations. In Section III-A, we use the wireless channel after the training stage to send another randomly generated key from Alice to Bob using the wiretap channel model. One important observation is that in a code for the wiretap channel, one needs to use randomization. Roughly speaking the randomization rate is the same as the mutual information between Alice and Eve. In the coding scheme used in Section III-A, this randomization rate does not convey any information, although Bob is able to decode these randomization bits. Hence, the basic idea here is that instead of randomly generating randomization bits, we use the bin number to specify the random bits. In this way, we can use the wireless channel after the training phase to send a new key and the bin number simultaneously.

In our scheme, we set $U = \tilde{Y}_A + Z$, in which Z is a zero mean Gaussian random variable with variance σ_z^2 and is independent of other random variables considered in this paper. The variance is chosen to satisfy the condition that the wireless channel is able to support the rate of the helper data necessary for the key generation from the correlated noisy observations. It is easy to check that $U \to \tilde{Y}_A \to \tilde{Y}_B$. In this case, the key rate one can generate from the correlated observations is

$$2TR_s = 2I(U;\tilde{Y}_B) \tag{38}$$

$$= \log\left(\frac{(\sigma_h^2 + \frac{\sigma^2}{(1-\alpha)P_\tau T_\tau} + \sigma_z^2)(\sigma_h^2 + \frac{\sigma^2}{\alpha P_\tau T_\tau})}{(\sigma_h^2 + \frac{\sigma^2}{(1-\alpha)P_\tau T_\tau} + \sigma_z^2)(\sigma_h^2 + \frac{\sigma^2}{\alpha P_\tau T_\tau}) - \sigma_h^4}\right) \tag{39}$$

To achieve this rate, one needs to transmit at rate

$$\frac{1}{T}(I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)) \tag{40}$$

$$= \frac{1}{2T}\log\left(1 + \frac{\sigma_h^2\sigma^2}{\sigma_z^2(\sigma^2 + \sigma_h^2\alpha P_\tau T_\tau)} + \frac{\sigma^2}{\sigma_z^2(1-\alpha)P_\tau T_\tau}\right) \tag{41}$$

over the wireless channel. Hence, the value of σ_z^2 should be chosen carefully.

Theorem 3.3: Using a fading wireless channel without a public channel, a key rate of

$$R_{key} = \max_{\alpha,P_\tau,T_\tau}\{R_s + R_{ch}\}, \tag{42}$$

is achievable. Here, we require that

$$P_\tau T_\tau + (T - T_\tau)P_d \le PT. \tag{43}$$

At the same time, R_{ch} and R_s are given in (21) and (38), respectively, and σ_z^2 should be chosen to satisfy the following condition:

$$\frac{I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)}{T - T_\tau} \le \tag{44}$$

$$\min\left\{\mathbb{E}\left\{\log\left(1 + \frac{\hat{h}^2 P_d}{\sigma^2 + \frac{\sigma_h^2 P_d}{\sigma_h^2\alpha P_\tau T_\tau + \sigma^2}}\right)\right\},\right. \tag{45}$$

$$\left.\mathbb{E}\left\{\log\left(1 + \frac{h_{AE}^2 P_d}{\sigma^2}\right)\right\}\right\}. \tag{46}$$

Similarly to the situation in Section III-A, for general values of the available power P and the coherence length T, it is difficult to obtain closed form expressions for the optimal values of these parameters. In the following, we again consider two asymptotic regimes to gather insight.

1) *Long coherence time regime*, in which $T \to \infty$.

We first look at the R_s term. For any values of P_τ, T_τ and α, a simple calculation shows that

$$\frac{dR_s}{d\sigma_z^2} < 0. \tag{47}$$

Hence

$$
\begin{aligned}
&2TR_s \leq \\
&\max_{\alpha,T_\tau,P_\tau} \log\left(\frac{(\sigma_h^2+\frac{\sigma^2}{(1-\alpha)P_\tau T_\tau}+\sigma_z^2)(\sigma_h^2+\frac{\sigma^2}{\alpha P_\tau T_\tau})}{(\sigma_h^2+\frac{\sigma^2}{(1-\alpha)P_\tau T_\tau}+\sigma_z^2)(\sigma_h^2+\frac{\sigma^2}{\alpha P_\tau T_\tau})-\sigma_h^4}\right) \\
&\leq \max_{\alpha,T_\tau,P_\tau} \log\left(\frac{(\sigma_h^2+\frac{\sigma^2}{(1-\alpha)P_\tau T_\tau})(\sigma_h^2+\frac{\sigma^2}{\alpha P_\tau T_\tau})}{(\sigma_h^2+\frac{\sigma^2}{(1-\alpha)P_\tau T_\tau})(\sigma_h^2+\frac{\sigma^2}{\alpha P_\tau T_\tau})-\sigma_h^4}\right) \\
&\leq \log\left(\frac{\sigma^2}{\sigma_h^2 PT}+1+\frac{1}{4\sigma^2}\sigma_h^2 PT\right). \qquad (48)
\end{aligned}
$$

Thus, as $T\to\infty$, $R_s\to 0$. As a result, in this regime, the channel model is asymptotically optimal. The R_{ch} term is the same as that of the scenario with a public channel. Hence in the long coherence time regime, the key rate is the same as that of the scenario with a public channel.

2) *High power regime*, in which $P\to\infty$.

We can bound the R_{ch} term in the same manner as that of Section III-A. Hence, in the high power regime, the source model is asymptotically optimal. In the following, we study how R_s scales as P increases. From Section III-A, we know that if there is a public channel with infinite capacity, R_s scales logarithmically with P. Hence, in the absence of the public channel, R_s scales at most logarithmically with P. In the following, we show that R_s indeed scales logarithmically with P. We set $P_\tau=P$, $P_d=P$, $T_\tau=T/2$ and $\alpha=1/2$. Note that these parameters are not necessarily optimal.

Note that in the high power regime,

$$
\mathbb{E}\left\{\log\left(1+\frac{h_{AE}^2 P_d}{\sigma^2}\right)\right\} \doteq \log P, \quad (49)
$$

$$
\mathbb{E}\left\{\log\left(1+\frac{\hat{h}^2 P_d}{\sigma^2+\frac{\sigma_h^2 P_d}{\sigma_h^2\alpha P_\tau T_\tau+\sigma^2}}\right)\right\} \doteq \log P. \quad (50)
$$

Hence, if we choose $\sigma_z^2=P^{-2}$, (44) will be satisfied. Now, we substitute these choices of parameters into (38) and obtain

$$
\begin{aligned}
R_s &= \frac{1}{2T}\log\left(\frac{(\sigma_h^2+\frac{4\sigma^2}{PT}+P^{-2})(\sigma_h^2+\frac{4\sigma^2}{PT})}{(\sigma_h^2+\frac{4\sigma^2}{PT}+P^{-2})(\sigma_h^2+\frac{4\sigma^2}{PT})-\sigma_h^4}\right) \quad (51)\\
&\sim \frac{1}{2T}\log P. \quad (52)
\end{aligned}
$$

Hence, $R_{key}\sim\frac{1}{2T}\log P$ in the high power regime, which is the same as that in the case with a public channel.

IV. Key Agreement with the Presence of An Active Attacker

In this section, we extend the key generation approach developed in Section III to the case of an active attacker who can send attack signals to minimize the key rate. We first investigate the attacker's optimal attack strategy to this protocol. We then evaluate the key rate that can be generated under this active attack model. In this section, we consider only the more practical model in which there is no public channel.

A. Training Phase

As shown in Figure 1, our key generation protocol has two phases: a training phase and a transmission phase. The active attacker can initiate an attack during both these two phases. We first characterize the attacker's optimal strategy for the training phase.

Suppose Alice sends a known sequence $\mathbf{S}_A$ of size $1\times\alpha T_\tau$, with $0<\alpha<1$ during the training stage, then Bob receives

$$
\mathbf{Y}_{B,\tau}=h_{AB}\mathbf{S}_A+\mathbf{X}_{E1}+\mathbf{N}_B, \quad (53)
$$

where $\mathbf{N}_B=[N_B(1),\cdots,N_B(\alpha T_\tau)]^T$. After that, Bob sends a known sequence $\mathbf{S}_B$ of size $1\times(1-\alpha)T_\tau$ over the wireless channel, and Alice receives

$$
\mathbf{Y}_{A,\tau}=h_{AB}\mathbf{S}_B+\mathbf{X}_{E2}+\mathbf{N}_A, \quad (54)
$$

where $\mathbf{N}_A=[N_A(1),\cdots,N_A((1-\alpha)T_\tau)]^T$.

Following the protocol discussed in Section III, Alice computes a statistic $\tilde{Y}_A$ for $\mathbf{Y}_{A,\tau}$ via

$$
\tilde{Y}_A=\frac{\mathbf{S}_B^T}{||\mathbf{S}_B||}\mathbf{Y}_{A,\tau}=h_{AB}+\frac{\mathbf{S}_B^T}{||\mathbf{S}_B||}(\mathbf{X}_{E1}+\mathbf{N}_A), \quad (55)
$$

in which $||\cdot||$ denotes the norm of its argument. Similarly, Bob computes a statistic $\tilde{Y}_B$ for $\mathbf{Y}_{B,\tau}$ via

$$
\tilde{Y}_B=\frac{\mathbf{S}_A^T}{||\mathbf{S}_A||}\mathbf{Y}_{B,\tau}=h_{AB}+\frac{\mathbf{S}_A^T}{||\mathbf{S}_A||}(\mathbf{X}_{E2}+\mathbf{N}_B). \quad (56)
$$

We use Γ_1 to denote $\mathbf{S}_B^T\mathbf{X}_{E1}/||\mathbf{S}_B||$, N_1 to denote $\mathbf{S}_B^T\mathbf{N}_A/||\mathbf{S}_B||$, Γ_2 to denote $\mathbf{S}_A^T\mathbf{X}_{E1}/||\mathbf{S}_A||$, and N_2 to denote $\mathbf{S}_A^T\mathbf{N}_B/||\mathbf{S}_A||$ respectively. Hence, (55) and (56) can be rewritten as

$$
\begin{aligned}
\tilde{Y}_A &= h_{AB}+\Gamma_1+N_1, \quad (57)\\
\tilde{Y}_B &= h_{AB}+\Gamma_2+N_2. \quad (58)
\end{aligned}
$$

If the attacker is passive, as discussed in Section III, $\tilde{Y}_A$ and $\tilde{Y}_B$ are jointly Gaussian random variables. However, when the attacker is active, the statistics of these two random variables depend on the attacker's strategy. Alice and Bob will generate a key from these two correlated observations. As will be clear in the sequel, our protocol will generate a key from $(\tilde{Y}_A,\tilde{Y}_B)$ with a rate

$$
R_s=\frac{1}{T}(I(\tilde{Y}_A+Z;\tilde{Y}_B)-I(\tilde{Y}_A+Z;\Gamma_1,\Gamma_2)). \quad (59)
$$

Here Z is a zero mean Gaussian random variable with variance σ_z^2, and is independent of other random variables of interest in the paper. The normalization factor $1/T$ comes from the fact that the channel gain is fixed for T symbols, meaning that we can observe only one value of $(\tilde{Y}_A,\tilde{Y}_B)$ for every T symbols. Roughly speaking, $I(\tilde{Y}_A+Z;\tilde{Y}_B)$ is the common randomness that both Alice and Bob share, and $I(\tilde{Y}_A+Z;\Gamma_1,\Gamma_2)$ is the amount of information that Eve knows about the value of $\tilde{Y}_A+Z$. This is due to the fact that both $\tilde{Y}_A$ and $\tilde{Y}_B$ are related to the signal transmitted by Eve. Hence, the attacker will design its attack signal such that the mutual information between the observations at Alice and Bob is small, while the mutual information between the observations at Alice and the attack signal at Eve is large.

At the same time, Bob obtains a Minimum Mean Square Error (MMSE) estimate $\hat{h}_{AB}$ of the channel gain h_{AB} in the given coherence block. $\hat{h}_{AB}$ will be treated as the true value of the channel gain in the second phase of the key agreement protocol. We can write $h_{AB} = \hat{h}_{AB} + \bar{h}_{AB}$, in which $\bar{h}_{AB}$ is the estimation error. As will be clear in the sequel, the rate of the key that can be generated using our protocol depends on the variance of $\bar{h}_{AB}$, which will be denoted by σ_{est}^2. The larger the variance, the smaller the rate of the key.

Hence, the attacker needs to design its attack signal $\mathbf{X}_{E1}$ and $\mathbf{X}_{E2}$ to simultaneously maximize σ_{est}^2 and minimize R_s. First, it is clear that the attacker should set $\mathbb{E}\{\Gamma_1\} = \mathbb{E}\{\Gamma_2\} = 0$. Assuming that Alice and Bob transmit with power P_τ during the training period, we have $||S_B|| = (1-\alpha)T_\tau P_\tau$ and $||S_A|| = \alpha T_\tau P_\tau$. Also, assuming that the attacker transmits at a power P_{E1} for $\mathbf{X}_{E1}$ and P_{E2} for $\mathbf{X}_{E2}$ respectively, then $\text{Var}\{\Gamma_1\} = \sigma_1^2 = P_{E2}/P_\tau$ and $\text{Var}\{\Gamma_2\} = \sigma_2^2 = P_{E1}/P_\tau$. Assuming that the correlation coefficient between Γ_1 and Γ_2 is ρ, we need to characterize the distribution of (Γ_1, Γ_2) that the attacker will adopt to maximize σ_{est}^2 and minimize R_s.

Theorem 4.1: Choosing (Γ_1, Γ_2) to be jointly Gaussian simultaneously minimizes R_s and maximizes σ_{est}^2. Furthermore, the optimal correlation coefficient between Γ_1 and Γ_2 is given by

$$\rho_{opt} = \begin{cases} -\frac{\sigma_h^2}{\sigma_1\sigma_2}, & \text{if} \quad \sigma_h^2 \leq \sigma_1\sigma_2 \\ -1, & \text{otherwise.} \end{cases} \tag{60}$$

Proof: First, from [25], we know that to maximize σ_{est}^2, one should use the Gaussian distribution. That is choosing the probability density function (PDF) of Γ_2 to be $\mathcal{N}(0, \sigma_2^2)$ maximizes σ_{est}^2.

Next, we characterize the optimal distribution of (Γ_1, Γ_2) that minimizes R_s. We can rewrite R_s as follows:

$$\begin{aligned} &TR_s \\ &= I(\tilde{Y}_A + Z; \tilde{Y}_B) - I(\tilde{Y}_A + Z; \Gamma_1, \Gamma_2) && (61) \\ &= h(\tilde{Y}_A + Z) - h(\tilde{Y}_A + Z|\tilde{Y}_B) - h(\tilde{Y}_A + Z) \\ &\qquad + h(\tilde{Y}_A + Z|\Gamma_1, \Gamma_2) && (62) \\ &= -h(\tilde{Y}_A + Z|\tilde{Y}_B) + h(h_{AB} + \Gamma_1 + N_1 + Z|\Gamma_1, \Gamma_2) && (63) \\ &= -h(\tilde{Y}_A + Z|\tilde{Y}_B) + h(h_{AB} + N_1 + Z). && (64) \end{aligned}$$

The only term in (64) that the attacker can control is the conditional entropy $h(\tilde{Y}_A + Z|\tilde{Y}_B)$. Hence, to minimize R_s, the attacker will choose its attack strategy to maximize $h(\tilde{Y}_A + Z|\tilde{Y}_B)$. Similar to [26], we have

$$\begin{aligned} h(\tilde{Y}_A + Z|\tilde{Y}_B) &= h(\tilde{Y}_A + Z - c\tilde{Y}_B|\tilde{Y}_B) && (65) \\ &\overset{(a)}{\leq} h(\tilde{Y}_A + Z - c\tilde{Y}_B) && (66) \\ &\overset{(b)}{\leq} \frac{1}{2}\log(2\pi e\sigma_e^2). && (67) \end{aligned}$$

The equalities in (a) and (b) will hold, if $c = \sigma_{AB}/\sigma_{\tilde{Y}_B}^2$ and $(\tilde{Y}_A + Z, \tilde{Y}_B)$ are jointly Gaussian. Here $\sigma_{AB} = \mathbb{E}\{(\tilde{Y}_A + Z)\tilde{Y}_B\} = \sigma_h^2 + \rho\sigma_1\sigma_2$, and $\sigma_{\tilde{Y}_B}^2 = \sigma_h^2 + \sigma_2^2 + \sigma^2/(\alpha P_\tau T_\tau)$. This is due to the fact that if $(\tilde{Y}_A + Z, \tilde{Y}_B)$ are jointly Gaussian, then equality in (b) holds. Furthermore, if $(\tilde{Y}_A + Z, \tilde{Y}_B)$ are jointly Gaussian and c is chosen in this manner, $\tilde{Y}_A + Z - c\tilde{Y}_B$ will be independent of $\tilde{Y}_B$ and thus equality in (a) holds. In this case

$$\begin{aligned} \sigma_e^2 &= \mathbb{E}\{(\tilde{Y}_A + Z - c\tilde{Y}_B)^2\} && (68) \\ &= \left(\sigma_h^2 + \sigma_1^2 + \frac{\sigma^2}{(1-\alpha)P_\tau T_\tau}\right) - \frac{(\sigma_h^2 + \rho\sigma_1\sigma_2)^2}{\sigma_h^2 + \sigma_2^2 + \sigma^2/(\alpha P_\tau T_\tau)}. \end{aligned}$$

To make $(\tilde{Y}_A + Z, \tilde{Y}_B)$ jointly Gaussian, (Γ_1, Γ_2) should be jointly Gaussian. Combined with the fact that choosing Γ_2 to be Gaussian maximizes σ_{est}^2, we know that choosing (Γ_1, Γ_2) to be jointly Gaussian simultaneously minimizes R_s and maximizes the variance of $\bar{h}_{AB}$.

Since only R_s depends on ρ, the attacker should choose ρ to minimize R_s, which is equivalent to maximizing σ_e^2 in (68). It is easy to see from (68) that

$$\rho_{opt} = \begin{cases} -\frac{\sigma_h^2}{\sigma_1\sigma_2}, & \text{if} \quad \sigma_h^2 \leq \sigma_1\sigma_2 \\ -1, & \text{otherwise.} \end{cases} \tag{69}$$

Hence, during the training stage, the attacker should adopt a correlated jamming attack with ρ_{opt} given in (60).

■

B. Key Generation Phase

As discussed in Section III, after the training period of T_τ symbols, Alice will send two pieces of information to Bob via the wireless channel: 1) the information needed to distill a key from the correlated estimations $(\tilde{Y}_A, \tilde{Y}_B)$ obtained in the first phase, which is public information and does not need to be kept secure, and 2) a new randomly generated key with a rate R_{ch}, which needs to be kept secure from the attacker. The total key rate will be $R_{ch} + R_s$.

1) Key generation from the correlated observations: We first look at the key distillation part, in which we generate a key from the correlated observations $(\tilde{Y}_A, \tilde{Y}_B)$. Compared with the scenario discussed in Section III-B, the attacker now possesses observations (Γ_1, Γ_2) that are correlated with the observations $(\tilde{Y}_A, \tilde{Y}_B)$ at the legitimate users. The problem of key generation from correlated sources through a channel with limited capacity has been studied in [24]. More precisely, if the channel has a rate constraint R, then the following secret key rate can be generated from the correlated observations $(\tilde{Y}_A, \tilde{Y}_B)$ with Eve observing (Γ_1, Γ_2) [24]:

$$\begin{aligned} R_s^* &= \left[I(U; \tilde{Y}_B) - I(U; \Gamma_1, \Gamma_2)\right]^+ && (70) \\ &\text{s.t.} \quad U \to \tilde{Y}_A \to \tilde{Y}_B, && (71) \\ &\text{and} \quad I(U; \tilde{Y}_A) - I(U; \tilde{Y}_B) \leq R, && (72) \end{aligned}$$

where U is an auxiliary random variable subject to the Markov chain relationship given to it in (71).

More precisely, for every N symbol times, Alice has $m = \lfloor N/T \rfloor$ observations of the random variable $\tilde{Y}_A$. We call these N symbols a group. Here $\lfloor \cdot \rfloor$ is the largest integer that is smaller than its argument. These observations are collected into a vector $\tilde{\mathbf{Y}}_A = [\tilde{Y}_A(1), \cdots, \tilde{Y}_A(m)]^T$. Here, the $\tilde{Y}_A(i)$'s are independent of each other. Similarly, Bob has a vector of observations $\tilde{\mathbf{Y}}_B = [\tilde{Y}_B(1), \cdots, \tilde{Y}_B(m)]^T$. Furthermore, this rate can be achieved by sending from Alice only. Roughly speaking, we generate $2^{mI(U;\tilde{Y}_A)}$ typical U sequences. We then divide these typical sequences into bins, each bin containing $2^{mI(U;\tilde{Y}_B)}$ sequences. Hence, each U^m sequence can be specified by two indices (i, j) with i being the bin number (ranging from 1 to $2^{m(I(U;\tilde{Y}_A)-I(U;\tilde{Y}_B))}$), and j being the index of the sequence within each bin. Now, after observing $\tilde{\mathbf{Y}}_A = [\tilde{Y}_A(1), \cdots, \tilde{Y}_A(m)]^T$, Alice finds a U^m sequence that is jointly-typical with $\tilde{\mathbf{Y}}_A$. (This step will be successful with a probability very close to one.) Alice sets the key value as the $j \mod 2^{mI(U;\Gamma_1,\Gamma_2)}$, and sends the value of i to Bob, which requires a rate of $I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)$. After receiving the bin number i, Bob obtains an estimate $\hat{U}^m$ by looking for a unique sequence in the bin specified by the bin number that is jointly typical with its observation $\tilde{\mathbf{Y}}_B$. $\hat{U}^m$ will be equal to U^m with probability 1, thus Bob can then recover the key value by setting it as $\hat{j} \mod 2^{mI(U;\Gamma_1,\Gamma_2)}$. In our protocol, we adopt a simple strategy and set $U = \tilde{Y}_A + Z$, with Z being $\mathcal{N}(0, \sigma_z^2)$ and independent of other random variables of interest. Hence, the key rate that can be generated from the correlated observations is $R_s = R_s^*/T$. Again, the normalization term $1/T$ comes from the fact that we have one observation for every T seconds. To generate this, we need to transmit the bin number i over the wireless channel, which requires a rate of $R = [I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)]/T$.

2) Key generation from the channel: Now, we look at how to send a newly generated key over the wireless channel. There are two main difference from that of Section III-B: 1) the channel estimation is more coarse due to the attack in the channel estimation stage; and 2) the attacker will send attack signal in this stage.

More specifically, Bob still obtains an MMSE estimate $\hat{h}_{AB}$ of the channel gain h_{AB} in the given coherence block,

$$\hat{h}_{AB} = \frac{\sigma_h^2}{\sigma^2 + \sigma_2^2 + \alpha P_\tau T_\tau \sigma_h^2} \mathbf{S}_A^T \mathbf{Y}_{B,\tau}. \quad (73)$$

Bob will treat this as the true value of the channel gain. We can write $h_{AB} = \hat{h}_{AB} + \bar{h}_{AB}$, in which $\bar{h}_{AB}$ is the estimation error. $\bar{h}_{AB}$ is a zero mean Gaussian random variable with variance

$$\sigma_h^2/(\sigma_h^2 \alpha P_\tau T_\tau + \sigma^2 + \sigma_2^2).$$

Now, when Alice transmits, Bob and Eve receive

$$Y_B = \hat{h}_{AB} X_A + \bar{h}_{AB} X_A + X_E + N_B, \quad (74)$$

$$\text{and} \quad Y_E = h_{AE} X_A + N_E. \quad (75)$$

Eve will choose attack signal X_E to minimize R_{ch} specified by (21), which we reproduce here for the easiness of presentation:

$$R_{ch} = \frac{T - T_\tau}{T} [I(X_A; Y_B|\hat{h}_{AB}) - I(X_A; Y_E|h_{AE})]^+. \quad (76)$$

Obviously, the attacker will design X_E such that $I(X_A; Y_B|\hat{h}_{AB})$ is minimized. Since the attacker receives Y_E which is correlated with X_A, the attacker can design X_E based on its knowledge of X_A.

To characterize the attacker's optimal attack strategy, we need a result from [27]. The result says that if h_{AB} is independent of X_A in the system and X_A is Gaussian, then even if Eve knows X_A completely, the optimal attack strategy of the Eve is to send i.i.d. Gaussian noise that is independent of X_A. When one tries to use this result, caution should be exercised to satisfy this condition. As discussed in Section III-B, X_A contains two pieces of information: the number of the bin to which the channel gain belongs, and the newly generated key. That is X_A is specified by the bin number i, which contains some information about the channel gain h_{AB}. We can overcome this issue by using the scheme illustrated in Figure 2. More specifically, as discussed in Section IV-B1, we divide the time into groups, each containing N symbol times (i.e., m fading blocks). In group k, Alice collects a vector of channel observations $\tilde{\mathbf{Y}}_A$, and determines the bin number i_k of this vector. Instead of transmitting i_k to Bob using the wireless channel during the kth group (which will introduce correlation between the channel gain and the codeword sent over the channel), we will transmit i_k over the $k+1$th block. With this idea, we can use the result of [27] and know that the optimal strategy of the attacker is to send i.i.d. Gaussian noise.

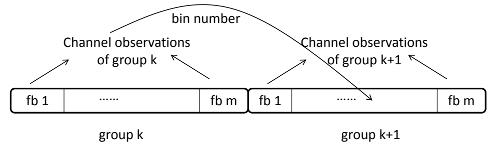

Fig. 2: A scheme to avoid correlation between the channel gain and the transmitted codeword.

Suppose the power used by Alice and Eve during this stage are P_d and P_{E3} respectively, then (76) is

$$R_{ch} = \frac{T - T_\tau}{2T} \left[\mathbb{E} \left\{ \log \left(1 + \frac{\hat{h}^2 P_d}{\sigma^2 + P_{E3} + \sigma_{est}^2} \right) \right. \right. \quad (77)$$

$$\left. \left. - \log \left(1 + \frac{h_{AE}^2 P_d}{\sigma^2} \right) \right\} \right]^+. \quad (78)$$

In summary, we have the following.

Theorem 4.2: Using a fading wireless channel, a key rate of

$$R_{key} = \min_{P_{E1}, P_{E2}, P_{E3}} \max_{\alpha, P_\tau, T_\tau} \{R_s + R_{ch}\}, \quad (79)$$

is achievable. Here, we require that

$$P_\tau T_\tau + P_d(T - T_\tau) \le PT, \quad (80)$$

$$P_{E1} \alpha T_\tau + P_{E2}(1 - \alpha) T_\tau + P_{E3}(T - T_\tau) \le P_E T. \quad (81)$$

At the same time, σ_z^2 should be chosen to satisfy the following condition:

$$\frac{I(U;\tilde{Y}_A) - I(U;\tilde{Y}_B)}{T - T_\tau} \le \quad (82)$$

$$\min \left\{ \mathbb{E} \left\{ \log \left(1 + \frac{\hat{h}^2 P_d}{\sigma^2 + P_{E3} + \sigma_{est}^2} \right) \right\}, \right. \quad (83)$$

$$\left. \mathbb{E} \left\{ \log \left(1 + \frac{h_{AE}^2 P_d}{\sigma^2} \right) \right\} \right\}. \quad (84)$$

V. Conclusions

In this paper, we have developed a joint source-channel approach for key agreement over wireless channels that combines benefits of existing models. We have shown that in general, one can increase the key rate by using both the channel model and the source model. We have further shown that in the long coherence time regime, the channel model is asymptotically optimal. On the other hand, we have shown that in the high power regime, the source model is asymptotically optimal. We have further extended the protocol to the scenario with an active attacker. We have characterized the attacker's optimal attack strategy to the adopted key agreement protocol. We have also quantified the rate of the key that can be generated under this attack strategy. We have shown that, unlike the situation in wireline communications, one can generate a key with a nonzero rate over unauthenticated wireless fading channels.

In terms of future research, it will be interesting to extend our study to the multiple antennas case. It is also important to study the arbitrary channel model in which the adversary is more powerful. It is also of interest to study the scenarios in which the attackers have objectives other than minimizing key rate.

References

[1] L. Lai and H. V. Poor, "A unified framework for key agreement over wireless fading channels," in *Proc. IEEE Inform. Theory Workshop*, (Taormina, Sicily, Italy), Oct. 2009.

[2] L. Lai, Y. Liang, and H. V. Poor, "Key agreement over wireless fading channels with an active attacker," in *Proc. 48th Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), Sept. 2010.

[3] Y. Liang, H. V. Poor, and S. Shamai (Shitz), *Information Theoretic Security*. Dordrecht, The Netherlands: Now Publishers, 2009.

[4] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 4687–4698, Oct. 2008.

[5] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Secure communication over fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2470–2492, June 2008.

[6] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: Secret sharing using reciprocity in ultrawideband channels," *IEEE Trans. Inform. Forensics and Security*, vol. 2, pp. 364–375, Sept. 2007.

[7] T.-H. Chou, A. M. Sayeed, and S. C. Draper, "Minimum energy per bit for secret key acquisition over multipath wireless channels," in *Proc. IEEE Intl. Symposium on Inform. Theory*, (Seoul, Korea), Jun. 2009.

[8] T.-H. Chou, A. M. Sayeed, and S. C. Draper, "Impact of channel sparsity and correlated eavesdropping on secret key generation from multipath channel randomness," in *Proc. IEEE Intl. Symposium on Inform. Theory*, (Austin, TX), Jun. 2010.

[9] A. Sayeed and A. Perrig, "Secure wireless communications: Secret keys through multipath," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, (Las Vegas, NV), Apr. 2008.

[10] C. Ye, S. Mathur, A. Reznik, W. Trappe, and N. Mandayam, "Information-theoretic key generation from wireless channels," *IEEE Trans. Inform. Forensics and Security*, vol. 5, pp. 240–254, Jun. 2010.

[11] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radiotelephathy: Extracting a secret key from an unauthenticated wireless channel," in *Proc. ACM International Conference on Mobile Computing and Networking*, 2008.

[12] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part I: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[13] U. M. Maurer and S. Wolf, "Secret key agreement over a non-authenticated channel - Part I: Definitions and bounds," *IEEE Trans. Inform. Theory*, vol. 49, pp. 822–831, Apr. 2003.

[14] U. M. Maurer and S. Wolf, "Secret key agreement over a non-authenticated channel - Part II: The simulatability condition," *IEEE Trans. Inform. Theory*, vol. 49, pp. 832–838, Apr. 2003.

[15] U. M. Maurer and S. Wolf, "Secret key agreement over a non-authenticated channel - Part III: Privacy amplification," *IEEE Trans. Inform. Theory*, vol. 49, pp. 839–851, Apr. 2003.

[16] L. Lai and H. El Gamal, "The relay-eavesdropper channel: Cooperation for secrecy," *IEEE Trans. Inform. Theory*, vol. 54, pp. 4005–4019, Sept. 2008.

[17] L. Lai, H. El Gamal, and H. V. Poor, "The wiretap channel with feedback: Encryption over the channel," *IEEE Trans. Inform. Theory*, vol. 54, pp. 5059 – 5067, Nov. 2008.

[18] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas-I: The MISOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, pp. 3088–3104, July 2010.

[19] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas-II: The MIMOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, pp. 5515–5532, Nov. 2010.

[20] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. 33, pp. 267–284, Mar. 1987.

[21] I. Csiszar and P. Narayan, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 37, pp. 18–26, Jan. 1991.

[22] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933–946, May 2000.

[23] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[24] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inform. Theory*, vol. 46, pp. 344–366, Mar. 2000.

[25] T. Basar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. Inform. Theory*, vol. 29, pp. 152–157, Jan. 1983.

[26] M. Medard, "Capacity of correlated jamming channels," in *Proc. 35th Allerton Conf. on Communication, Control, and Computing*, (Monticello, IL), Sept. 1997.

[27] A. Kashyap, T. Basar, and R. Srikant, "Correlated jamming on MIMO Gaussian fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2119–2123, Sept. 2004.