Pilot Contamination Attacks in Massive MIMO Systems

Berk Akgun, Marwan Krunz, and O. Ozan Koyluoglu
Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721
Email: {berkakgun, krunz, ozan}@email.arizona.edu

***Abstract*—We consider a single-cell massive multiple-input multiple-output (MIMO) system in which a base station (BS) with a large number of antennas simultaneously transmits to K single-antenna users in the presence of an attacker. Massive MIMO systems often operate in a time division duplexing (TDD) fashion. The BS estimates the channel state information (CSI) at receivers based on their uplink pilot transmissions. Downlink transmission rates are highly dependent on these estimates, as the BS utilizes the CSI to exploit the beamforming gain offered by massive MIMO. However, this CSI estimation phase is vulnerable to malicious attacks. Specifically, an attacker can contaminate the uplink pilot sequences by generating identical pilot signals to those of legitimate users. We formulate a denial of service (DoS) attack in which the attacker aims to minimize the sum-rate of downlink transmissions by contaminating the uplink pilots. We also consider another attack model where the attacker generates jamming signals in both the CSI estimation and data transmission phases by exploiting in-band full-duplex techniques. We study these attacks under two power allocation strategies for downlink transmissions. Our analysis is conducted when the attacker knows or does not know the locations of the BS and users. When the attacker does not have perfect location information, stochastic optimization techniques are utilized to assess the impact of the attack. The formulated problems are solved using interior-point, Lagrangian minimization, and game-theoretic methods. We obtain a closed-form solution for a special case of the problem. Our results indicate that even though the attacker does not have the perfect location information, proposed pilot contamination attacks degrade the throughput of a massive MIMO system by more than 50%, and reduce fairness among users significantly. In addition, we show that increasing the number of pilot symbols does not prevent the proposed attacks, if the BS uniformly allocates powers for downlink transmissions.**

***Index terms*—Massive MIMO, pilot contamination attack, physical layer security, stochastic optimization, game theory.**

I. Introduction

Massive multiple-input multiple-output (MIMO) is one of the key technologies in the upcoming 5G systems. It is envisioned that a cellular base station (BS) in 5G systems will be equipped with a very large antenna array, e.g., hundreds of antennas or more, boosting the transmission rate by orders of magnitude compared to conventional MIMO systems. Even though MIMO is a well-studied concept in wireless communications, massive MIMO requires novel techniques to overcome new design challenges, and as such it has received significant attention from researchers over the last few years (see, for example, [1], [2], [3], and the references therein).

One of the important issues in massive MIMO systems is pilot contamination (PC) [4]. Because of the large number of antennas at the BS and the relatively short channel coherence time, the channel state information (CSI) between the BS and various users must be estimated frequently using uplink pilot transmissions. The BS utilizes these CSI estimates for downlink data transmissions, assuming channel reciprocity. However, due to the limited number of orthogonal pilot sequences (e.g., 42 [4]), users in neighboring cells may share the same pilots. Interference among these pilots causes erroneous channel estimates at the BS, which lead to poor system performance.

In [5], the authors studied an attack that exploits vulnerabilities in time division duplexing (TDD) systems during the channel training phase. The idea behind this attack is to contaminate uplink pilot transmissions and cause an erroneous uplink channel estimation. Typically, if the CSI is available, the BS would use MIMO beamforming techniques such as maximum-ratio transmission (MRT) to maximize the signal-to-noise-ratio (SNR) at users. However, the benefits of these techniques vanish rapidly if the CSI estimates are erroneous. A self-contamination technique was proposed in [6] to detect this type of attack. The authors in [7] also proposed several attack-detection methods. Secure transmissions for TDD-based massive MIMO systems was studied in [8] in the presence of an active eavesdropper. The authors derived the optimal power allocation for the information and artificial noise (AN) signals at the BS such that secrecy is guaranteed asymptotically, i.e., as the number of BS antennas (M) tends to infinity. Notably, none of above works proposed countermeasures to prevent PC attacks. In [9], the authors proposed providing secrecy against PC attacks by keeping pilot assignments hidden and using a pilot set that scales with M. However, there are two main problems with this scheme. First, it requires a longer pilot transmission phase, which increases the overhead and decreases the throughput. Second, computational cryptographic methods are required to keep pilot assignments hidden. All of the papers discussed above consider an attacker that targets a

This research was supported in part by the National Science Foundation (grants # CNS-1409172, CNS-1513649, IIP-1265960, and CNS-1617335) and by Qatar Foundation (grant # NPRP 8-052-2-029). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the NSF and QF.

single user. Even when a multiuser system is considered, the attacker randomly selects one user and contaminates its pilot sequence. Given that one of the key aspects of massive MIMO systems is to serve tens of users simultaneously, the vulnerabilities of these systems to a multiuser pilot contamination attack should be investigated.

In this paper, we consider a single-cell multiuser massive MIMO network in the presence of an attacker. We study an attack model in which the attacker aims at minimizing the sum-rate of downlink transmissions, i.e., a denial of service (DoS) attack, by contaminating uplink pilot transmissions. We derive the downlink transmission rates, with and without the PC attack, exploiting the *channel hardening effect* (effect of small-scale fading on channel gains vanishes as M tends to infinity) in massive MIMO to analyze the attack strategies. Optimal attack strategies are investigated for two different cases: when the attacker knows the locations of the BS and users and when she does not have this information. Considering a fixed power allocation strategy at the BS (for downlink information signals), convex problems are formulated for the optimal PC attack. These problems are solved by the interior-point and Lagrangian minimization methods. We obtain a closed-form solution for the case of perfect information, i.e., known topology at the attacker. This solution represents a lower bound on the downlink sum-rate of massive MIMO systems under an optimal PC attack and a fixed BS transmission power. Then, we study the scenario where the BS optimizes its own power allocation scheme in the presence of PC attacks. For this case, a game-theoretic problem formulation is considered in which the BS and attacker are the players of the game. In particular, we obtain a *convex-concave* game, and propose an iterative algorithm that converges to the Nash equilibrium (NE) of the game. Our analysis provides an upper bound on the downlink sum-rate of massive MIMO systems under an optimal PC attack. Further, we study an attack model where the attacker generates jamming signals in both the pilot and downlink data transmission phases (hybrid attack). For this attack, the attacker is required to have a full-duplex radio. Stochastic optimization techniques are used to find the optimal power allocation at the attacker so as to minimize the downlink sum-rate of the system. In particular, the attacker estimates the channels between the users and itself while jamming the uplink pilot transmissions. These estimates are then used to strengthen the DoS attack during the downlink data transmission phase. Numerical results show that the downlink sum-rate significantly decreases under such an attack. Particularly, when the attacker is close to the BS, the downlink sum-rate of all users is reduced by more than 50%. Another important result of our paper is that an attacker without perfect information about the user locations is almost devastating as one with perfect information. This fact emphasizes the vulnerability of massive MIMO systems to PC attacks.

The rest of the paper is organized as follows. Section II describes the system model. In Section III, we compute the downlink transmission rates with/without the pilot contamination attack. Our PC attack under a fixed and optimal BS transmission power is analyzed in Section IV. We investigate the hybrid attack model in Section V. We provide numerical results in Section VI, and conclude the paper in Section VII.

Throughout the paper, we adopt the following notation. $\mathbb{E}[\cdot]$ indicates the expectation of a random variable. Row vectors and matrices are denoted by bold lower-case and upper-case letters, respectively. $(\cdot)^*$ and $(\cdot)^T$ represent the complex conjugate transpose and transpose of a vector or matrix, respectively. Frobenius norm and the absolute value of a real or complex number are denoted by $\|\cdot\|$ and $|\cdot|$, respectively. $\mathbf{A} \in \mathbb{C}^{M\times N}$ means that $\mathbf{A}$ is an $M \times N$ complex matrix, and $\mathbf{I}_M$ is an $M \times M$ identity matrix. $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex circularly symmetric Gaussian random variable of mean μ and variance σ^2. $[x]^+$ is defined as $\max(x, 0)$. For simplicity, $\log_2(\cdot)$ is referred to as $\log(\cdot)$.

II. System Model

We consider a single-cell massive MIMO system in which the BS (Alice) uses a large array of M antenna elements to transmit/receive independent data streams to/from K single-antenna users (Bobs), $M \gg K$. Because of the large M, the channel coherence time is not long enough to estimate the CSI of all M downlink channels per user [3]. Therefore, TDD is used instead of FDD (in the latter case, the downlink and uplink channels are estimated separately). In TDD, Alice estimates the CSI for uplink channels after receiving pilot sequences transmitted by Bobs. If these pilot symbols are not perfectly orthogonal to each other, interference among them causes erroneous channel estimates at the BS. Assuming channel reciprocity, these estimates are used for downlink data transmissions. There is no standardization for massive MIMO systems regarding the orthogonality of the pilot sequences. However, the authors in [4] suggested assigning an orthogonal time-frequency pilot sequence to each Bob. Orthogonal space-time block codes can also be utilized, as in 802.11ac systems, to increase the number of orthogonal pilot sequences. Fig. 1 shows an example of eight pilot sequences. Pilot sequences $\mathbf{p}_1$ and $\mathbf{p}_2$ are orthogonal space-time coded sequences. They are sent in the same time interval (t_1) over the same frequency (f_1) by two different Bobs. On the other hand, the orthogonality of $\mathbf{p}_1$ and $\mathbf{p}_5$ is guaranteed by transmitting them in different time intervals t_1 and t_2, e.g., $\mathbf{p}_1 = 0$ during t_2. Similarly, $\mathbf{p}_1$ and $\mathbf{p}_3$ are transmitted in different frequencies f_1 and f_2, respectively. The received signal at Alice during the pilot transmission phase is given by:

$$\mathbf{Y}_A = \sum_{i=1}^{K} \sqrt{P_k}\mathbf{h}_k^T\mathbf{p}_k + \mathbf{W} \tag{1}$$

where $\mathbf{h}_k^T \in \mathbb{C}^{M\times 1}$ represents the uplink channel from Bob$_k$ (kth Bob) to Alice. The mth entry, $m \in \{1, \cdots, M\}$, of this vector is given by $h_k^{(m)} = \sqrt{\theta_k}g_k^{(m)}$, where θ_k and $g_k^{(m)} \sim \mathcal{CN}(0, 1)$ represent the path-loss component (large-scale fading) and small-scale effects of the channel (Rayleigh fading), respectively. Note that θ_k is the same for all antennas, so $\mathbf{h}_k$ can be written as $\mathbf{h}_k = \sqrt{\theta_k}\mathbf{g}_k$, where $\mathbf{g}_k$ is a vector

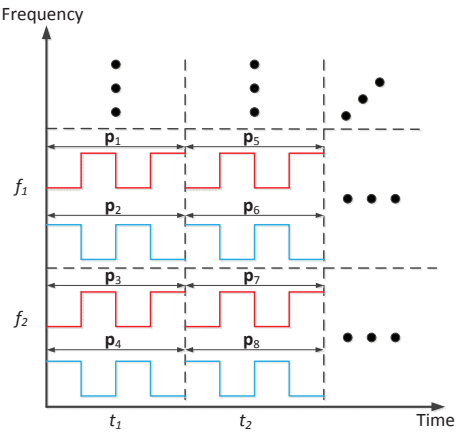


Fig. 1. Orthogonality of pilot sequences in space-time-frequency domain.

of all $g_k^{(m)}$, $m \in \{1, \cdots, M\}$. $\mathbf{p}_k \in \mathbb{C}^{1\times L}$ is the transmitted pilot sequence by Bob$_k$, where L is the number of symbols in the pilot sequence. As these pilot sequences are orthogonal to each other, $\mathbf{p}_k\mathbf{p}_l^* = 0 \ \forall \ k$ and $l \in \mathcal{K}$, where $k \neq l$ and $\mathcal{K} = \{1, \cdots, K\}$. P_k is the pilot transmission power at Bob$_k$, while $\mathbf{p}_k$ is a unit vector (i.e., $\mathbf{p}_k\mathbf{p}_k^* = 1$). $\mathbf{W}$ is the additive white Gaussian noise (AWGN) matrix, whose entries are zero-mean, unit-variance normal random variables. Without loss of generality, consider the estimation of $\mathbf{h}_i$ at Alice. Let $\hat{\mathbf{h}}_i$ represent this estimate. Under a priori knowledge of $\mathbf{p}_i$, Alice post-multiplies the received signal by $\mathbf{p}_i^*$ and divides it by $\sqrt{P_i}$ and L to obtain:

$$\begin{aligned}\hat{\mathbf{h}}_i^T &= \frac{\mathbf{Y}_A\mathbf{p}_i^*}{\sqrt{P_i}L} = \sum_{k=1}^{K}\frac{\sqrt{P_k}\mathbf{h}_k^T\mathbf{p}_k\mathbf{p}_i^*}{\sqrt{P_i}L} + \frac{\mathbf{W}\mathbf{p}_i^*}{\sqrt{P_i}L} \\ &= \mathbf{h}_i^T + \tilde{\mathbf{w}}_i^T \end{aligned} \tag{2}$$

where $\tilde{\mathbf{w}}_i^T \triangleq \frac{\mathbf{W}\mathbf{p}_i^*}{\sqrt{P_i}L} \sim \mathcal{CN}(0, \frac{1}{P_iL}\mathbf{I}_M)$

A. Attack Model

The attacker aims to contaminate pilot transmissions by imposing his signal. We assume that the attacker knows the pilot sequences used by Bobs (generally, pilots are publicly known sequences). Because the total number of orthogonal pilots is limited, after eavesdropping on the channels for a while, the attacker can learn which pilot is assigned to which Bob . If $\mathbf{x}_J \in \mathbb{C}^{1\times L}$ is the signal generated by the attacker, the received signal at Alice will be modified as follows:

$$\mathbf{Y}_A = \sum_{k=1}^{K}\sqrt{P_k}\mathbf{h}_k^T\mathbf{p}_k + \mathbf{h}_J^T\mathbf{x}_J + \mathbf{W} \tag{3}$$

where $\mathbf{h}_J^T \in \mathbb{C}^{M\times 1}$ represents the channel vector from the attacker to Alice. In the literature, $\mathbf{x}_J$ is often designed such that only a single user is targeted by the attacker [5], [9] (this user is selected randomly without any optimization). More specifically, $\mathbf{x}_J$ is often set to $\sqrt{P_J}\mathbf{p}_k$, where P_J is the average jamming power. In our model, we extend this strategy by setting $\mathbf{x}_J$ to:

$$\mathbf{x}_J = \sqrt{P_J}\sum_{k=1}^{K}\sqrt{\alpha_k}\mathbf{p}_k \tag{4}$$

where α_k is the ratio between the power allocated for $\mathbf{p}_k$ and the average jamming power. Note that $\sum_{k=1}^{K}\alpha_k \leq 1$. In our model, the objective of the attacker is to minimize the downlink sum-rate. Let R_k be the downlink transmission rate at Bob$_k$. The attacker's goal can be formulated as follows:

$$\underset{\{\alpha_k \ \forall k\in\mathcal{K}\}}{\text{minimize}} \sum_{k\in\mathcal{K}} R_k \tag{5}$$

subject to $\alpha_k \geq 0 \ \forall k \in \mathcal{K}$ and $\sum_{k=1}^{K}\alpha_k \leq 1$

III. Downlink Transmission Rates

In this section, we review and analyze the downlink sum-rate with/without the PC attack.

A. No PC Attack

In massive MIMO systems, the BS often applies MRT precoder [1]–[3], [10]. In conventional MIMO systems, MRT results in inter-user interference. However, as the number of antennas at the BS tends to infinity, the channels between BS and individual users become orthogonal to each other, and they individually reduce to single-input single-output (SISO) channels. In this case, MRT is the optimal precoder. Let s_k be the information signal intended to Bob$_k$ $\forall k \in \mathcal{K}$, and $\mathbf{v}_k^T \in \mathbb{C}^{M\times 1}$ be its normalized precoder, i.e., $\mathbf{v}_k\mathbf{v}_k^* = 1$. The received signal at Bob$_k$ in the downlink data transmission phase is given by:

$$y_k = \sum_{i=1}^{K}\sqrt{P_i^{(d)}}\mathbf{h}_k\mathbf{v}_i^T s_i + w_k^{(d)} \tag{6}$$

where $P_k^{(d)}$ and $w_k^{(d)}$ are the allocated power to s_k at Alice and the AWGN with zero-mean and unit-variance at Bob$_k$, respectively. Employing MRT precoding, $\mathbf{v}_k^T$ is given by $\mathbf{v}_k^T = (\hat{\mathbf{h}}_k^*/\|\hat{\mathbf{h}}_k\|)$. The achievable downlink rate at Bob$_k$ becomes:

$$R_k = \log\left(1 + \frac{P_k^{(d)}|\mathbf{h}_k\mathbf{v}_k^T|^2}{\sum_{l\in\{\mathcal{K}\setminus k\}}P_l^{(d)}|\mathbf{h}_k\mathbf{v}_l^T|^2 + 1}\right), \ k \in \mathcal{K}. \tag{7}$$

Note that the precoding vectors are computed based on channel estimates. Next, we study the asymptotic behavior of R_k as $M \to \infty$, with the objective of simplifying its expression. Such asymptotic analysis is needed later on for comparison with the case under a PC attack.

Consider the inter-user interference term $P_l^{(d)}|\mathbf{h}_k\mathbf{v}_l^T|^2$ in (7). Scaling this term by M and taking the limit as $M \to \infty$, we end up with [10, Lemma 1]:

$$\lim_{M\to\infty}\frac{P_l^{(d)}|\mathbf{h}_k\mathbf{v}_l^T|^2}{M} = \lim_{M\to\infty}\frac{P_l^{(d)}|\frac{\mathbf{h}_k\hat{\mathbf{h}}_l^*}{M}|^2}{\|\hat{\mathbf{h}}_l\|^2/M} = 0 \tag{8}$$

$\forall k$ and $l \in \mathcal{K}$, where $k \neq l$. The reason is that entries of small-scale channel components of Bob_k and Bob_l are independent random variables of zero-mean and unit-variance. Hence, $\lim_{M\to\infty} \mathbf{g}_l \mathbf{g}_k^*/M = 0$. Similarly, $\lim_{M\to\infty} \mathbf{g}_l \tilde{\mathbf{w}}_k^*/M = 0$. This is a result of the channel orthogonality in massive MIMO systems. On the other hand, for the term in the numerator in (7), we have

$$\lim_{M\to\infty} \frac{P_k^{(d)} |\mathbf{h}_k \mathbf{v}_k^T|^2}{M} = \lim_{M\to\infty} \frac{P_k^{(d)} |\frac{\mathbf{h}_k \hat{\mathbf{h}}_k^*}{M}|^2}{\|\hat{\mathbf{h}}_k\|^2/M} = \frac{P_k^{(d)} \theta_k^2}{\theta_k + \frac{1}{P_k L}} > 0. \tag{9}$$

The last step follows from the Continuous Mapping Theorem and the fact that given a vector $\mathbf{x} \in \mathbb{C}^{1\times M}$ with a distribution $\mathcal{CN}(\mathbf{0}, c\mathbf{I})$, $\lim_{M\to\infty} \mathbf{x}\mathbf{x}^*/M = c$ [10, Lemma 1]. Hence, the downlink rate at Bob_k asymptotically behaves as:

$$R_k \sim \log\left(1 + \frac{P_k^{(d)} \theta_k^2}{(\theta_k + \frac{1}{P_k L})\frac{1}{M}}\right). \tag{10}$$

In our paper, we consider a finite but sufficiently large M, with $M \gg K$, so the channels are near-orthogonal. As a result, the inter-user interference can be neglected as in (8). Moreover, for a sufficiently large M, $|\mathbf{h}_k \mathbf{v}_k^T|^2/M$ approaches the result in (9) ([3], [4], [10]). In Section VI, we will numerically verify these results. As explained before, θ_k is the large-scale channel components at Bob_k. Equation (10) indicates that the SINR does not depend on the small-scale fading components, as they are averaged out by the large antenna array (channel hardening). The term $(1/M)$ in the equation comes from the AWGN $w_k^{(d)}$ at Bob_k. For example, as $M \to \infty$, the noise term vanishes and the SINR tends to infinity. Another noise term arises due to the channel estimation errors. For example, as the length of the pilots, L, increases, the second term in the denominator becomes smaller. This leads to an increase in the downlink rate. The same effect is also observed when the power allocated for pilots increases.

In this paper, we consider two different transmit power allocation strategies at Alice: "fixed" and "optimal". Both strategies are subject to an average power constraint P_A. Under the fixed power allocation, $P_k^{(d)}$ $\forall k \in \mathcal{K}$ is known to the attacker. For example, based on a fairness criterion, these values may be determined before the pilot transmission phase (e.g., when Bobs are registered with the network), and Alice may convey this information to Bobs through a feedback channel. If the attacker eavesdrops on this channel, she can obtain the power allocation values. In an instance of this setup, Alice may simply allocate powers uniformly to the information signals, i.e., $P_1^{(d)} = \cdots = P_K^{(d)} = P_A/K$. On the other hand, under the "optimal" power allocation strategy, Alice relies on the well-known water-filling technique to assign powers, using $(\theta_k + (P_k L)^{-1})/(M\theta_k^2)$ as the water levels [11].

B. Presence of PC Attack

Under the attack model in (4), the following channel estimation is performed at Alice for each Bob_k:

$$\hat{\mathbf{h}}_k = \mathbf{h}_k + \sqrt{\alpha_k u_k}\mathbf{h}_J + \tilde{\mathbf{w}}_k \tag{11}$$

where u_k is the ratio between the average power at the attacker and the pilot transmission power at Bob_k, i.e., $u_k = P_J/P_k$. In the rest of the paper, we assume that u_k is known to the attacker. Previously, we assumed that the attacker learns the pilot sequences by eavesdropping on the uplink transmissions. The attacker can similarly learn the pilot transmission power. Note that Alice is not aware of the presence of the attacker, so she treats $\hat{\mathbf{h}}_k$ as the correct channel estimate. Employing MRT precoding based on this estimation, the precoder vector of s_k is given by:

$$\mathbf{v}_k^T = \frac{(\mathbf{h}_k + \sqrt{\alpha_k u_k}\mathbf{h}_J + \tilde{\mathbf{w}}_k)^*}{\|\mathbf{h}_k + \sqrt{\alpha_k u_k}\mathbf{h}_J + \tilde{\mathbf{w}}_k\|}. \tag{12}$$

Using this precoder vector in (7) leads to a non-convex problem in (5). To obtain a tractable problem for the underlying attack model, we analyze the asymptotic behavior of R_k as $M \to \infty$. Following the same steps as in the case of no attacker, the following expression is obtained:

$$R_k = \log\left(1 + \frac{P_k^{(d)} M\theta_k^2}{\theta_k + \alpha_k u_k \theta_J + \frac{1}{P_k L}}\right). \tag{13}$$

Expectedly, as M increases, the massive MIMO system becomes more resilient to PC attacks. However, the vulnerability of the system against such an attack can be observed from (13), which shows that the SINR decreases with an increase in the jamming power $\alpha_k u_k$.

As in the previous section, a fixed or "optimal" power allocation strategy can be employed to calculate each $P_k^{(d)}$. Fixed power allocation is performed exactly as before, whereas "optimal" power allocation corresponds to the following strategy. Let $\phi_k \triangleq \theta_k + \alpha_k u_k \theta_J + \frac{1}{P_k L}$. Then, Alice tries to maximize $R_{\text{sum}} = \sum_{k=1}^K R_k$ to obtain the "optimal" power allocation vector:

$$\left[P_1^{(d)} \cdots P_K^{(d)}\right] = \underset{x_k \forall k \in \mathcal{K}}{\text{argmax}} \sum_{k=1}^K \log\left(1 + \frac{x_k M\theta_k^2}{\phi_k}\right) \tag{14}$$

subject to $\sum_{k=1}^K P_k^{(d)} \leq P_A$ and $P_k^{(d)} \geq 0$, $\forall k \in \mathcal{K}$. Because Alice is unaware of the attack, she will not necessarily solve the above problem. However, our goal is to observe the effect of PC attack, even if Alice employs the least favorable power allocation scheme from the perspective of the attacker. This way, we can establish an upper-bound on the downlink sum-rate under a PC attack.

IV. Analysis of Optimal PC Attack

A. Fixed Power Allocation at Alice

In this section, we study the optimal PC attack strategy. Our analysis provides a lower bound on the downlink sum-rate under a PC attack for a given power allocation at Alice.

We incorporate (13) into problem (5), considering fixed power allocation for the information signals at Alice:

$$\mathbf{P1}: \underset{\{\alpha_k\ \forall k\in\mathcal{K}\}}{\text{minimize}} \sum_{k=1}^{K} \log\left(1+\frac{P_k^{(d)} M\theta_k^2}{\theta_k+\alpha_k u_k \theta_J+\frac{1}{P_k L}}\right)$$
$$s.t. \qquad \alpha_k \geq 0\ \forall k\in\mathcal{K}, \quad \sum_{k=1}^{K}\alpha_k \leq 1.$$

For a given $k\in\mathcal{K}$, we assume that $\theta_k = A z_k^{-\gamma}$, where A is a constant that depends on the transmit and receive antennas, operating frequency etc., while γ and z_k are the path-loss exponent and the distance between Alice and Bob$_k$, respectively. Similarly, z_J is the distance between Alice and the attacker. For simplicity, the antennas at Bobs and the attacker are assumed to be identical, so the same A is considered for all of them. As a result, the objective function of **P1** is converted to the following one:

$$R_{\text{sum}} = \sum_{k=1}^{K} \log\left(1+\frac{P_k^{(d)} M A z_k^{-2\gamma}}{\alpha_k u_k z_J^{-\gamma}+z_k^{-\gamma}+\frac{1}{AP_kL}}\right) \tag{15}$$

Next, we discuss two different scenarios based on the information available to the attacker.

1) Perfect Information: Here, we assume that the attacker has perfect knowledge of the distances between Alice and individual Bobs as well as her own distance to Alice. Indeed, this is an idealized scenario (from the attacker's point of view), and is merely studied to provide a benchmark for comparison with the case of uncertainty in distances. **P1** is a convex programming problem, and we obtain the optimal solution as follows.

Theorem 1: **P1** has the following closed-form solution:

$$\alpha_k = \left[\frac{\sqrt{A_k(A_k+4/\lambda)}-A_k-2B_k}{2}\right]^+ \quad \forall k\in\mathcal{K} \tag{16}$$

where

$$A_k \triangleq \frac{P_k^{(d)} M A z_J^{\gamma}}{u_k z_k^{2\gamma}} \text{ and } B_k \triangleq \frac{z_J^{\gamma}}{u_k z_k^{\gamma}}+\frac{z_J^{\gamma}}{u_k A P_k L}.$$

λ is the *Karush-Kuhn-Tucker* (KKT) multiplier and is chosen such that $\sum_{k=1}^{K}\alpha_k = 1$. It can be easily computed by the *bisection* method as $\sum_{k=1}^{K}\alpha_k$ is a decreasing function of it.

Proof: See Appendix A. ∎

2) Uncertainty in Distances: Suppose that the attacker does not have perfect knowledge about various distances. Let Z_k and Z_J be random variables (rvs) that correspond to the Alice-Bob$_k$ and Alice-attacker distances, respectively. In this case, the expected value of R_{sum} is given by:

$$\begin{aligned}\mathbb{E}[\ R_{\text{sum}}\ ] &= \mathbb{E}\left[\sum_{k=1}^{K}\log\left(1+\frac{P_k^{(d)} M A Z_k^{-2\gamma}}{\alpha_k u_k Z_J^{-\gamma}+Z_k^{-\gamma}+\frac{1}{AP_kL}}\right)\right] \\ &= \sum_{k=1}^{K}\mathbb{E}\left[\log\left(1+\frac{P_k^{(d)} M A Z^{-2\gamma}}{\alpha_k u_k Z_J^{-\gamma}+Z^{-\gamma}+\frac{1}{AP_kL}}\right)\right]\end{aligned} \tag{17}$$

where Z is a generic rv that has the same distribution as Z_k for all k. In (17), the expectation is taken over Z and Z_J. The last equality follows from the assumption that the distributions of the distances between individual Bobs and Alice are identical. We further assume that Bobs and the attacker are randomly and uniformly located in a circular area around Alice. Hence, the CDF of Z is given by $\Pr[Z\leq x] = x^2/D_{\max}^2$ for $x\geq 0$, where $D_{\max}$ is the maximum possible distance between Alice and any Bob (e.g., the maximum communication range). Accordingly, the PDF of Z is given by $f_Z(x) = 2x/D_{\max}^2$, for $x\geq 0$.

Let $\Phi_k \triangleq \alpha_k u_k Z_J^{-\gamma}+Z^{-\gamma}+\frac{1}{AP_kL}$. Under fixed downlink power allocation, the optimal PC attack can be formulated by the following stochastic programming problem:

$$\mathbf{P2}: \underset{\{\alpha_k\ \forall k\in\mathcal{K}\}}{\text{minimize}} \sum_{k=1}^{K}\mathbb{E}\left[\log\left(1+\frac{P_k^{(d)} M A Z^{-2\gamma}}{\Phi_k}\right)\right]$$
$$s.t. \qquad \alpha_k \geq 0\ \forall k\in\mathcal{K}, \quad \sum_{k=1}^{K}\alpha_k \leq 1.$$

The objective function in **P2** can be rewritten as:

$$\sum_{k=1}^{K}\int_0^{D_{\max}}\int_0^{D_{\max}} \frac{2x}{D_{\max}^2}\frac{2y}{D_{\max}^2}\log(\Psi(x,y))\,dx\,dy \tag{18}$$

where

$$\Psi(x,y) \triangleq 1+\frac{P_k^{(d)} M A x^{-2\gamma}}{\alpha_k u_k y^{-\gamma}+x^{-\gamma}+\frac{1}{AP_kL}},\ \text{for } x,y\in[0,D_{\max}]$$

This is a convex programming problem, as the objective function and inequality constraints are all convex functions. The integral in (18) can be approximated by Simpson's Rule for double integrals, and can be solved efficiently by applying the interior point method. Note that **P2** need only be solved offline, so the time complexity of this solution method is not a concern. We also note that although we only study a uniform distribution for the locations of Bobs and the attacker, any arbitrary distribution can be considered. The integral operation preserves the convexity, so the same steps can be followed to solve **P2**. Our numerical results (not shown for brevity) indicate that for typical values of P_A, A, K, and $D_{\max}$, the attacker should target all Bobs by equally allocating its average power to various pilot sequences under uniform power allocation when $u_k = u_l\ \forall k,l\in\mathcal{K}$. That is, $\alpha_k = P_J/K$ $\forall k\in\mathcal{K}$. This is due to the symmetry of Bobs for this special case, as will be discussed in Section VI.

3) Discussion: Let $\mathbf{z}\triangleq[z_1,\cdots,z_K]$ be the vector of distances from Alice to various Bobs (known to the attacker). Let $\boldsymbol{\alpha}^*(\mathbf{z},z_J) = [\alpha_1^*(\mathbf{z},z_J),\cdots,\alpha_K^*(\mathbf{z},z_J)]$ and $\boldsymbol{\alpha}^* = [\alpha_1^*,\cdots,\alpha_K^*]$ be the optimal solutions to **P1** and **P2**, respectively. In this case, the objective function of **P2** becomes $\mathbb{E}_{\mathbf{Z},Z_J}[R_{\text{sum}}(\boldsymbol{\alpha}^*)]$, and $\mathbb{E}_{\mathbf{Z},Z_J}[R_{\text{sum}}(\boldsymbol{\alpha}^*(\mathbf{Z},Z_J))]$ becomes the expectation of the optimal solution of **P1** under perfect information, where $\mathbf{Z}$ is the vector of i.i.d. distances $Z_1,\cdots,Z_K$. The expectations are taken over the random

distances, as previously explained. The *expected value of perfect information* (EVPI) is defined as follows:

$$\text{EVPI} \triangleq \mathbb{E}_{\mathbf{Z},Z_J}[R_{\text{sum}}(\boldsymbol{\alpha}^*)] - \mathbb{E}_{\mathbf{Z},Z_J}[R_{\text{sum}}(\boldsymbol{\alpha}^*(\mathbf{Z},Z_J))]. \quad (19)$$

Note that EVPI is always greater than or equal to zero, as the case with perfect information outperforms the one with uncertainty. If EVPI is small, the attacker does not gain much by knowing the exact distances. It can perform attacks almost as powerful as when perfect information is available. On the other hand, if EVPI is high, the attacker may try to acquire distance information by estimating Bobs' locations relative to its own. For example, a group of colluding adversaries can employ localization techniques (e.g., RSSI and time-of-arrival) to estimate Alice-to-Bobs distances [12], [13]. This requires more complex and costly systems at the attacker. In Section VI, we study the behavior of EVPI.

B. Optimal Power Allocation

In this section, we derive the optimal PC attack strategy when Alice adopts optimal (the least favorable from the perspective of the attacker) power allocation strategy for downlink data transmissions. Note that Alice is assumed to be unaware of the attack. Therefore, she cannot customize her power allocation strategy to combat such an attacker. However, while the attacker tries to minimize the downlink sum-rate, Alice tries to maximize this rate, without knowing about the attack. This is a *min-max* problem, and its solution is found as follows. As seen from (15), R_{sum} is a function $\mathbf{P}^{(d)} = \left[P_1^{(d)} \cdots P_K^{(d)}\right]$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_K]$. Thus, the problem can be formulated as a *convex-concave* game; for a fixed $\mathbf{P}^{(d)}$, $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ is a convex function of $\boldsymbol{\alpha}$, and for a fixed $\boldsymbol{\alpha}$, $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ is a concave function of $\mathbf{P}^{(d)}$. This means that the attacker needs to solve the following game:

$$\mathbf{P3}: \underset{\{\boldsymbol{\alpha}\}}{\text{minimize}} \left\{ \underset{\{\mathbf{P}^{(d)}\}}{\text{maximize}}\ R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha}) \right\}$$

$$s.t. \qquad \alpha_k \geq 0\ \forall k \in \mathcal{K}, \quad \sum_{k=1}^{K} \alpha_k \leq 1$$

$$P_k^{(d)} \geq 0\ \forall k \in \mathcal{K}, \quad \sum_{k=1}^{K} P_k^{(d)} \leq P_A$$

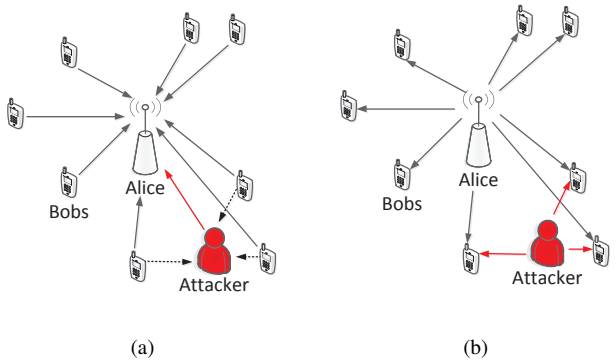
Let an optimal solution of this game, or a *saddle point*, be $(\mathbf{P}^{(d)*}, \boldsymbol{\alpha}^*)$. That is (for any possible power allocation $\mathbf{P}^{(d)}$),

$$R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha}^*) \leq R_{\text{sum}}(\mathbf{P}^{(d)*}, \boldsymbol{\alpha}^*) \leq R_{\text{sum}}(\mathbf{P}^{(d)*}, \boldsymbol{\alpha}).$$

This relationship shows that an upper-bound on $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ is obtained by solving **P3**. For instance, when $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$, $\mathbf{P}^{(d)*}$ maximizes $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha}^*)$. This optimal solution is obtained by a well-known water-filling technique. Specifically,

$$P_k^{(d)*} = \left[\eta - \frac{\alpha_k^* u_k z_J^{-\gamma} + z_k^{-\gamma} + \frac{1}{AP_kL}}{MAz_k^{-2\gamma}}\right]^+ \quad (20)$$

where η is a water-filling level chosen such that $\sum_{k=1}^{K} P_k^{(d)} = P_A$. η can be computed by bisection method as this summation is an increasing function of it. Similarly, when $\mathbf{P}^{(d)} = \mathbf{P}^{(d)*}$, $\boldsymbol{\alpha}^*$ minimizes $R_{\text{sum}}(\mathbf{P}^{(d)*}, \boldsymbol{\alpha})$. We propose to solve this game by using an iterative *Gauss-Seidel* method. To do that, we first solve $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ for some initial values of α_k, e.g., $\alpha_k = 0\ \forall k \in \mathcal{K}$ (initially, there is no PC attack). Then, the obtained $P_k^{(d)}$ values are used in $R_{\text{sum}}(\mathbf{P}^{(d)*}, \boldsymbol{\alpha})$, and this problem is solved with respect to $\alpha_k\ \forall k \in \mathcal{K}$ as explained in Theorem 1. After this step, the second iteration starts by solving $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha}^*)$ using the new values of α_k's. As the number of iterations increases, a better approximation for the saddle point is obtained. We evaluate the number of iterations required to reach the Nash equilibrium of this game, and observe that the algorithm almost always converges after 10 iterations. Due to space limitations, we omit the results here.



(a) (b)

Fig. 2. (a) Attacker contaminates the CSI estimation at Alice while overhearing the pilots from Bobs, (b) attacker generates the jamming signals to reduce the signal strength at Bobs during data transmission.

Theorem 2: Gauss-Seidel iterations converge when used to solve **P3**.

Proof: See Appendix B. ■

Note that the above analysis applies to the case of perfect information where distances are known to the attacker. It can be easily extended to the case where only the probability distribution of distances is known. The same steps in Section IV-A2 are applied to account for the uncertainty. In particular, the expectation of $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ over Z_J and Z_k's is considered in the objective function of **P3**. The resulting problem is still a convex-concave game that can be solved by the Gauss-Seidel method. We skip this analysis here due to space limitations.

V. Hybrid Full-duplex Attack

So far, we have considered jamming the pilot transmission phase. However, if the attacker is equipped with a full-duplex (FD) radio that allows it to transmit and receive signals simultaneously over the same frequency, a more sophisticated attack can be launched. Further, a stronger attack can also be launched with a multi-antenna (MIMO) FD-based attacker. In particular, consider an attacker with an average power constraint over the whole transmission phase (pilot and downlink data phases). Using an FD radio, the attacker can generate jamming signals during both phases. For instance, the attacker may contaminate the CSI estimation process at Alice, as in

Fig. 2(a), without knowing the channels between itself and Bobs. At the same time, the attacker can overhear the pilots (dashed lines in Fig. 2(a)) from Bobs using the FD radio, and exploit this knowledge to transmit jamming signals during the downlink transmission phase, as shown in Fig. 2(b). We call this attack a *hybrid* attack, as it combines the PC attack and conventional data-jamming attack. Notice that even though the hybrid attack performs at least as good as the PC attack, it requires an additional hardware capability (FD radio) at the attacker.

Even though the attacker needs one antenna to generate a jamming signal in data transmission phase, we study a more general scenario where she is equipped with $N+1$ antennas, where $N > 0$. Our goal is to find an optimal strategy for the attacker to minimize the downlink sum-rate, exploiting its multiple antennas. One of these antennas is reserved for the PC attack, while the others receive the pilot signals from Bobs. The attacker estimates $\mathbf{h}_{Jk} \in \mathbb{C}^{1\times N}$, the channel between Bob$_k$ and itself, during the pilot transmission phase. The self-interference signal at the receiving antennas of the attacker is canceled by employing FD radio design techniques in [14], [15]. For example, the self-interference channel is obtained by transmitting a pilot from the antenna that jams the pilot signal. Then, the self-interference signal is extracted from the received signals using this information. Let n_i be the ith jamming signal in the downlink transmission phase, $i \in \mathcal{N} = \{1, \cdots, N\}$. Let $h_{Jk}^{(i)} = g_{Jk}^{(i)}\sqrt{Az_{Jk}^{-\gamma}}$, $i \in \mathcal{N}$ and $k \in \mathcal{K}$, be the channel gain between the ith antenna of the attacker and Bob$_k$, where z_{Jk} denotes the distance between the attacker and Bob$_k$ and $g_{Jk}^{(i)}$ is the small-scale fading. $\beta_i\ \forall i \in \mathcal{N}$ denotes the ratio between the allocated power for n_i and P_J. By using the same PC attack model in Section II-A and MRT precoding at Alice, the received signal at Bob$_k$ during the downlink data transmission phase is given by:

$$y_k = \sum_{i=1}^{K}\sqrt{P_i^{(d)}}\mathbf{h}_k\frac{\hat{\mathbf{h}}_i^*}{\|\hat{\mathbf{h}}_i\|}s_i + \sum_{i=1}^{N}\sqrt{\beta_i P_J}h_{Jk}^{(i)}n_i + w_k^{(d)}.$$

Adding the jamming term to (10), the following downlink sum-rate is obtained:

$$R_{\text{sum}} = \sum_{k=1}^{K}\log\left(1 + \frac{C_k}{D_k\left(\sum_{i=1}^{N}\beta_i P_J|g_{Jk}^{(i)}|^2 Az_{Jk}^{-e}+1\right)}\right) \tag{21}$$

where

$$C_k \triangleq P_k^{(d)}MAz_k^{-2\gamma} \text{ and } D_k \triangleq \alpha_k u_k z_J^{-\gamma} + z_k^{-\gamma} + \frac{1}{AP_kL}.$$

Given the setup above, we formulate a two-stage stochastic optimization problem to find the optimal attacking strategy. This problem can be solved for various scenarios (e.g., perfect information, uncertainity in the distances and channels, etc.) by utilizing the techniques in Section IV and the ones presented in this section. The solutions of these problems are discussed in Section VI. For now, we explain our solution approach for one of these scenarios. Specifically, we assume that the distances, powers of information signals, and other constants in (21) are known to the attacker. In the first stage of the problem, the attacker finds the optimal values of $\alpha_k\ \forall k \in \mathcal{K}$ without knowing any $g_{Jk}^{(i)}\ \forall k \in \mathcal{K}$ and $\forall i \in \mathcal{N}$. In the second stage (after learning $g_{Jk}^{(i)}\ \forall k \in \mathcal{K}$ and $\forall i \in \mathcal{N}$ during the pilot transmission phase), the attacker optimally allocates the remaining power to the N jamming signals in the data transmission phase, i.e., $\beta_i\ \forall i \in \mathcal{N}$. Let ω represent a certain realization of the channel, $g_{Jk}^{(i)}$, and let Ω be the set of all realizations. (Note that $g_{Jk}^{(i)}$ and β_i are functions of these realizations.) Let t_p and t_d be the duration of pilot and data transmission phases, respectively. The two-stage stochastic problem can be formulated as follows:

$$\mathbf{P4}: \underset{\substack{\{\alpha_k\ \forall k\in\mathcal{K}\}\\ \{\beta_i(\omega)\ \forall i\in\mathcal{N},\ \forall\omega\in\Omega\}}}{\text{minimize}} \quad \mathbb{E}_\omega\left[\sum_{k=1}^{K}\log\left(1+\frac{C_k}{D_k\left(E_k+1\right)}\right)\right]$$
$$s.t. \quad \alpha_k \geq 0\ \forall k \in \mathcal{K}$$
$$\beta_i(\omega) \geq 0\ \forall i \in \mathcal{N},\ \forall \omega \in \Omega$$
$$\frac{F_k}{t_p+t_d} \leq 1\ \forall \omega \in \Omega$$

where $F_k \triangleq t_p\sum_{k=1}^{K}\alpha_k + t_d\sum_{i=1}^{N}\beta_i(\omega)$ and $E_k \triangleq \sum_{i=1}^{N}\beta_i(\omega)|g_{Jk}^{(i)}(\omega)|^2 P_J Az_{Jk}^{-e}$. Note that $g_{Jk}^{(i)}$ is a continuous random variable. **P4** can be approximately solved by creating T realizations, e.g., Ω has a cardinality of T. In particular, we replace the expectation in **P4** by the sum of these equiprobable T realizations. Therefore, we end up with K first-stage decision variables, namely $\alpha_k\ \forall k \in \mathcal{K}$, and NT second-stage decision variables, namely $\beta_i(\omega)\ \forall i \in \mathcal{N}$ and $\forall \omega \in \Omega$. The underlying problem is a convex programming problem, and can be solved by the interior point method. When T is large (for better approximation), the complexity of solving the problem increases. However, as the problem is solved offline, the time complexity is not a concern.

VI. Numerical Results and Discussion

We model the channel gain from each transmit antenna to each receive antenna as $h = g\sqrt{Ad^{-3.522}}$, where $g \sim \mathcal{CN}(0,1)$ and $A = 3.0682\times10^{-5}$. The path-loss is modeled using the COST-Hata Model with center frequency is 2 GHz [16]. The average transmit powers at Alice, Bob$_k$, and the attacker are 46, 20, and 30 dBm, respectively. The durations of the pilot and data transmission phases are set to be equal [4]. We consider a 20 MHz channel with noise floor of -101 dBm. Bobs and the attacker are uniformly and randomly distributed within a circle whose center is Alice and whose radius is $D_{\max}$ and $D_{\max,\text{J}}$, respectively. We set $D_{\max}$ to 750 meters. Our results are averaged over 10^5 different network realizations.

We set the number of users $K = 10$. In Fig. 3(a), we consider uniform power allocation for both the information signals at Alice and the jamming signals at the attacker. The figure depicts the downlink sum-rate vs. M. It shows that (10) and (13) are good approximations for the downlink rates in (7). Note that the approximation-based sum-rate is slightly higher than the exact values, as the inter-user interference does not

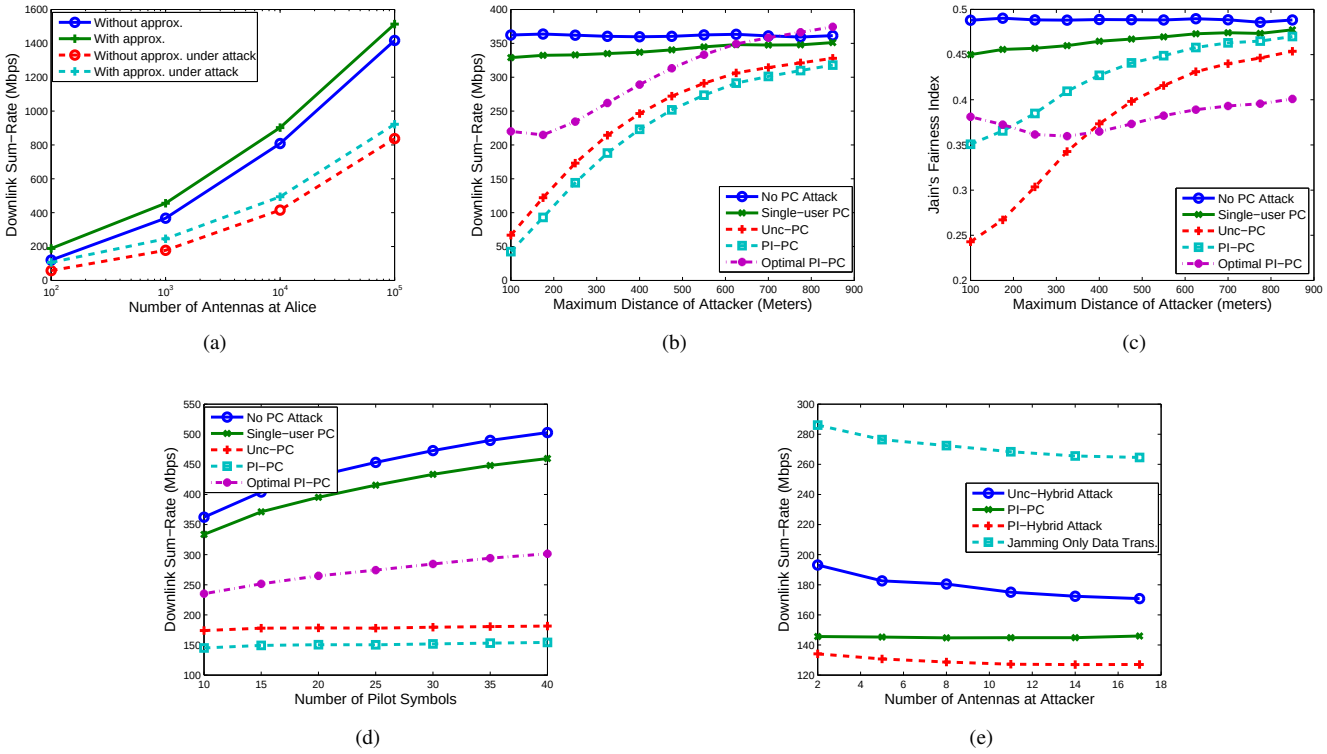


Fig. 3. (a) Downlink sum-rate vs. M under uniform power allocation at both Alice and the attacker, (b) downlink sum-rate vs. $D_{\max,\mathrm{J}}$, (c) Jain's fairness index vs. $D_{\max,\mathrm{J}}$, (d) downlink sum-rate vs. number of pilot symbols, (e) downlink sum-rate vs. number of antennas at the attacker.

perfectly vanish at a finite M. In our subsequent results, we set M to 1000.

We observe the effect of the maximum distance between Alice and the attacker ($D_{\max,\mathrm{J}}$) in Figs. 3(b) and 3(c). In the case of a single-user PC attack, only one randomly selected Bob is targeted by the attacker. This attack can also be interpreted as an unintentional interference from a user in an adjacent cell. It does not have a big impact on the sum-rate. PC with uncertainty (Unc-PC) and PC with perfect information (PI-PC) were explained in Section IV-A, and optimal PI-PC was studied in Section IV-B. Note that optimal PI-PC gives an upper-bound on the sum-rate of a massive MIMO system under an optimal PC attack. As the attacker moves farther from Alice, the sum-rate increases in all attack schemes. In Fig. 3(b), EVPI is around 20 Mbps. This says that when the attacker knows the distribution of Bobs, it can launch attacks that are almost as powerful as when the attacker has complete CSI. Note that Alice uniformly allocates downlink transmission powers in no PC attack scheme, whereas she employs optimal power allocation in the optimal PI-PC. Therefore, the downlink sum-rate without an attack is less than the one with the optimal PI-PC when $D_{\max,\mathrm{J}} > 700$ meters. In Fig. 3(c), we depict Jain's fairness index for different schemes. Jain's fairness index ranges from $1/K$ to 1 for the worst and best cases, respectively (if all users have the same downlink rate, the fairness index is 1). The figure shows that fairness among Bobs is significantly reduced when PC attacks take place. Unc-PC decreases the fairness more than PI-PC. The reason behind this phenomena is that when the attacker is close to Alice and knows the distances, Bobs with higher downlink rates are targeted. Therefore, Bobs are forced to have closer downlink rates, which increases the fairness index.

In Fig 3(d), we set $D_{\max,\mathrm{J}}$ to 250 meters, and study the effect of the number of pilot symbols L. As L increases, the sum-rate increases as well in no PC, single-user PC, and optimal PI-PC attacks. The reason is that the error in MRT precoding vectors due to erroneous channel estimates decreases, and the signal strength at Bobs increases. On the other hand, the sum-rate does not increase under the Unc-PC and PI-PC attacks. Note that in these cases, a fixed power is allocated for the information signals at Alice, and she does not exploit the decrease in channel estimation errors.

In Fig. 3(e), we compare hybrid and PC attacks under a similar average jamming power constraint. We observe that as the number of antennas at the attacker increases, the sum-rate slightly decreases for the hybrid attack. Note that the hybrid attacks utilizes multiple antennas, whereas PC attacks use a single-antenna. Interestingly, even though the hybrid attacks outperform PC attacks with respect to the sum-rate, a larger number of antennas at the attacker does not lead to more powerful attacks. EVPI for the hybrid attacks is around 60 Mbps, which is much higher than the one for PC attacks. The reason is that the hybrid attack includes one more source of uncertainty due to the channels between Bobs and the attacker.

Another important result is that attacking only downlink data transmissions (no jamming during pilot transmission phase) does not have as a great of an impact on performance as the impact of the PC attack.

VII. Conclusion

We considered a single-cell massive MIMO system with several mobile users, and demonstrated vulnerabilities of uplink pilot transmissions against jamming attacks. Specifically, the attacker generates pilot sequences similar to those of users and contaminates the pilot transmissions to distort channel estimation at the BS. This PC attack reduces the downlink transmission rates, as the beamforming techniques utilized by the BS heavily depend on accurate CSI estimates. We formulated an optimization problem from the standpoint of the attacker to minimize the downlink sum-rate. Both cases when the attacker knows or does not know the distances between the BS and users were considered. Using (stochastic) optimization and game theory, we derived the optimal attacking strategies when the BS employs either fixed or optimal power allocation for downlink transmissions. Numerical results showed that the downlink sum-rate is reduced by more than 50% if the average distance between the attacker and the BS is less than the one of the users. We also observed that even if the attacker does not know the channels and the locations of the users, it can launch powerful attacks as if it has the perfect information. In this work, we assumed that the BS and users are not aware of the attacker. An interesting future work is to develop counter algorithms to prevent PC attacks.

Appendix A
Proof of Theorem 2

Let us define

$$A_k = \frac{P_k^{(d)} M A z_J^{\gamma}}{u_k z_k^{2\gamma}} \text{ and } B_k = \frac{z_J^{\gamma}}{u_k z_k^{\gamma}} + \frac{z_J^{\gamma}}{u_k A P_k L}$$

$\forall k \in \mathcal{K}$. Therefore, the objective of **P1** can be written by

$$R_{\text{sum}} = \sum_{k=1}^{K} \log\left(1 + \frac{A_k}{\alpha_k + B_k}\right) \tag{22}$$

Hence, the Lagrangian function of this problem is given by

$$L(\boldsymbol{\alpha}) = \sum_{k=1}^{K} \log(1 + \frac{A_k}{\alpha_k + B_k}) + \lambda(\sum_{k=1}^{K} \alpha_k - 1). \tag{23}$$

Its first derivative with respect to α_k becomes

$$\frac{\partial L(\boldsymbol{\alpha})}{\partial \alpha_k} = \frac{-A_k}{(\alpha_k + B_k)(\alpha_k + A_k + B_k)} + \lambda. \tag{24}$$

Let $\alpha_k^* \ \forall k \in \mathcal{K}$ be the optimal value that minimizes the objective function of **P1**. These values are also the roots of the polynomial functions where the equation (24) is equal to zero. Also, note that $\alpha_k^* \ \forall k \in \mathcal{K}$ is a nonnegative number. Thus,

$$\alpha_k^* = \left[\frac{\sqrt{A_k(A_k + 4/\lambda)} - A_k - 2B_k}{2}\right]^+ \tag{25}$$

where λ is chosen such that $\sum_{k=1}^{K} \alpha_k^* = 1$.

Appendix B
Proof of Theorem 3

The players of the game described in **P3** are Alice and the attacker. In this game, the utility function of Alice is $R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$, and her strategy is to choose the optimal power allocation for the downlink transmissions. Similarly, $-R_{\text{sum}}(\mathbf{P}^{(d)}, \boldsymbol{\alpha})$ is the attacker's utility, and her strategy is to find the optimal $\boldsymbol{\alpha}$ to maximize this utility. The strategy sets of both players are non-empty, compact, and convex subsets of real numbers (the constraints in **P3** are linear functions). Furthermore, their utility functions are continuous and diagonally strictly concave. As a result, the existence and uniqueness of NE is proved for this game, and Gauss-Seidel method converges to this point [17].

References

[1] E. Bj, E. G. Larsson, T. L. Marzetta *et al.*, "Massive MIMO: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[4] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[5] X. Zhou, B. Maham, and A. Hjorungnes, "Pilot contamination for active eavesdropping," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 903–907, 2012.

[6] J. K. Tugnait, "Self-contamination for detection of pilot contamination attack in multiple antenna systems," *IEEE Wireless Communications Letters*, vol. 4, no. 5, pp. 525–528, 2015.

[7] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 21–27, 2015.

[8] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission in the presence of an active eavesdropper," in *Proc. of the IEEE International Conference on Communications (ICC)*, 2015, pp. 1434–1440.

[9] Y. O. Basciftci, C. E. Koksal, and A. Ashikhmin. (Mar. 2015) "Physical layer security in massive MIMO". [Online]. Available: http://arxiv.org/abs/1505.00396

[10] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 192–201, 2013.

[11] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[12] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. F. Abdelzaher, "Range-free localization and its impact on large scale sensor networks," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 4, no. 4, pp. 877–906, 2005.

[13] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, 2005.

[14] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 375–386, 2013.

[15] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 359–372.

[16] V. Abhayawardhana, I. Wassell, D. Crosby, M. Sellars, and M. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," in *Proc. of the IEEE VTC'05*, pp. 73–77.

[17] R. Cominetti, F. Facchinei, and J. B. Lasserre, *Modern optimization modelling techniques*. Springer Science & Business Media, 2012.