Enhancing Physical Layer Security of Random Caching in Large-Scale Multi-Antenna Heterogeneous Wireless Networks

Wanli Wen, Chenxi Liu, Yaru Fu, Tony Q. S. Quek, Fu-Chun Zheng, and Shi Jin

Abstract—In this paper, we propose a novel secure random caching scheme for large-scale multi-antenna heterogeneous wireless networks, where the base stations (BSs) deliver randomly cached confidential contents to the legitimate users in the presence of passive eavesdroppers as well as active jammers. In order to safeguard the content delivery, we consider that the BSs transmit the artificial noise together with the useful signals. By using tools from stochastic geometry, we first analyze the average reliable transmission probability (RTP) and the average confidential transmission probability (CTP), which take both the impact of the eavesdroppers and the impact of the jammers into consideration. We further provide tight upper and lower bounds on the average RTP. These analytical results enable us to obtain rich insights into the behaviors of the average RTP and the average CTP with respect to key system parameters. Moreover, we optimize the caching distribution of the files to maximize the average RTP of the system, while satisfying the constraints on the caching size and the average CTP. Through numerical results, we show that our proposed secure random caching scheme can effectively boost the secrecy performance of the system compared to the existing solutions.

Index Terms—Physical layer security, random caching, multi-antenna, heterogeneous wireless networks, stochastic geometry, optimization.

I. INTRODUCTION

With the rapid development of wireless technology, the global mobile data traffic is expected to reach 77 exabytes per month by 2022, which is a sevenfold increase over 2017 [1]. However, the majority of such traffic is asynchronously but repeatedly requested by many users at different times and thus a tremendous amount of mobile data traffic has actually been redundantly generated over the mobile networks [2]. Against this backdrop, caching popular contents at the base stations (BSs) has been proposed as a promising approach for shifting the enormous mobile data traffic from remote clouds to the edges in mobile networks [3, 4], thereby significantly improving the users’ quality of service (QoS). Motivated by this, cache-enabled wireless networks has recently been attracting great research attention.

However, due to the equipment size and cost issues, the storage resources at BSs are usually limited. Therefore, which contents should be cached at BSs becomes a vital design problem, referred to as the caching scheme design problem, in cache-enabled wireless networks. In [5]–[15], the authors proposed various caching schemes to improve the content transmission reliability (i.e., the probability that the requested content can be successfully transmitted from the BSs to the users, which can reflect the users’ QoS) in large-scale cache-enabled wireless networks. Specifically, in [5], the most popular caching scheme was studied, in which each BS only stores the most popular files. However, such a caching scheme may not yield the optimal network performance, since it cannot provide any content diversity [11]–[13]. To tackle this issue, random caching scheme was considered in [6]–[15]. More specifically, in [6], the uniform caching was studied, in which, each BS randomly stores a file according to the uniform distribution. In [7]–[13], the optimal random caching was examined to maximize the cache hit probability [7], [8], the successful offloading probability [9], and the successful transmission probability [10]–[14], or minimize the average caching failure probability [15].

Although significant efforts have been devoted to improve the reliability of content transmission [5]–[15], the content transmission secrecy is another important design aspect in the cache-enabled wireless networks, yet receives much less attention. Due to the broadcast nature of wireless communications, the transmissions of the contents cached at BSs are vulnerable to potential eavesdroppers (e.g., non-paying subscribers) and jammers [16]–[18]. By attempting to intercept the users’ content transmissions, the eavesdroppers can degrade the confidentiality of content transmissions. By transmitting harmful interfering signal in the wireless networks, the jammers can reduce the reliability of content transmissions. Currently, the most of mobile data services still rely on the traditional cryptographic encryption to guarantee the confidentiality of content transmission. Nevertheless, such encryption creates an insurmountable obstacle to caching content at BSs, since the encrypted content are uniquely defined for each user’s request and cannot be reused to serve other users’ requests [19]. Against this background, physical layer security has been regarded as a promising solution to support caching and enhance the confidentiality of content transmission simultaneously in large-scale cache-enabled single-tier [20], [21] or multi-tier heterogeneous wireless networks (HetNets) [22], due to the fact that it only relies on wiretap channel coding instead of...
source encryption. In particular, the authors in [20] and [22] proposed a partition-based hybrid caching scheme, where the cache of each BS was partitioned into two halves for storing both the most popular contents and fractions of other contents, and analyzed different physical-layer security performance such as the transmission capacity and energy efficiency [20] as well as the secure content delivery probability [22]. In [21], the authors leveraged the content diversity provided by the random caching to optimize the cache hit probability subject to different content confidentiality level constraints.

The authors in [20–22] only considered deploying single-antenna at each BS, and therefore, their proposed caching schemes may not achieve desirable security performance in multi-antenna wireless networks. In addition, unlike [20], [22], to deliberately confuse the eavesdroppers, an artificial-noise-aided transmission strategy was further considered in [21]. The key idea of this strategy is to use part of the transmit power at the BS to inject artificial noise into the null space of the legitimate user’s channel, thereby significantly reducing the link quality of the eavesdropper without interfering with the legitimate user [23]. However, since the BSs only have single-antenna, in [21], each BS is only able to transmit either artificial noise or content, resulting in that some users will not be able to download their requested content in a timely manner. Furthermore, it is worth pointing out that the authors in [20–22] only focused on combating the eavesdropping attack in cache-enabled wireless networks, leading to that their proposed caching schemes may not efficiently resist jamming attack.

In this paper, we would like to address the above issues. We consider a large-scale multi-antenna HetNet employing random caching and artificial-noise-aided transmission strategy, where the BSs deployed at multiple tiers transmit the randomly cached confidential content together with the artificial noise to the legitimate users in the presence of the jammers as well as eavesdroppers. We define and evaluate the content average reliable transmission probability (RTP) and the content average confidential transmission probability (CTP), which characterizes the reliability and the confidentiality of content transmissions, respectively. Our key goal is to determine a secure random caching scheme that maximizes the average RTP under the average CTP and storage resource constraints. Our main contributions are summarized as follows:

- We derive closed-form expressions for the average RTP and CTP in the considered HetNets. Based on which, we further derive tight upper and lower bounds on the average RTP. These analytical results allow us to reveal how the jamming and eavesdropping can have a significant impact on large-scale cache-enabled HetNets.
- We propose a new secure random caching scheme in which the caching distribution of the contents is judiciously determined to maximize the average RTP, while satisfying the constraints on the average CTP and the storage resource.
- Through numerical results, we show that the proposed secure caching scheme outperforms the existing baselines and significantly improves the average RTP, while being resist to the jamming and eavesdropping attack.

Beyond the above contributions, this paper provides valuable insights into the design of secure large-scale cache-enable HetNets. Specifically, we show that, when the constraint on the average CTP is stringent, caching less popular contents at BSs may be helpful for maintaining the level of average RTP.

In addition, we show that increasing the cache sizes at BSs does not always improve the average RTP when the secrecy constraint on the average CTP is taken into account.

The rest of the paper is organized as follows. Section II describes the system model and formulates a secrecy performance optimization problem. Section III analyzes the secrecy performance and derives the closed-form expressions for the average RTP and CTP. Based on these analytical results, Section IV develops two algorithms to solve the formulated optimization problem. Numerical results and the related discussions are provided in Section V. Finally, Section VI draws the conclusions. Unless otherwise specified, the notations used throughout the paper are summarized in Table I.

II. SYSTEM MODEL

A. Network Model

We consider the secure downlink transmissions in a large-scale cache-enabled HetNet, in which the BSs from $K$ co-channel deployed network tiers transmit to the users in the presence of randomly distributed jammers and eavesdroppers.

For ease of illustration, we denote the set of $K$ tiers by $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$. We assume that each BS in tier $k$ is active (as in [8], [10], [12]) and has $M_k$ antennas, while the users, the jammers, and the eavesdroppers are equipped with a single antenna each.

The eavesdroppers secretly intercept the downlink transmissions, while the jammers constantly broadcast the jamming signals into the channel to interrupt the downlink transmissions. The jammers and the eavesdroppers are not colluding, and therefore the jamming signals also interfere with the eavesdroppers. The BS locations in tier $k$ and the jammers are modeled by independent homogeneous Poisson point processes (PPPs) $\Phi_k$ and $\Phi_J$ with densities $\lambda_k$ and $\lambda_J$, respectively. In addition, the locations of the users and eavesdroppers are modeled by some other independent stationary point processes with certain densities.

We assume a interference-limited communication environment, in which the wireless channels undergo quasi-static Rayleigh fading along with a large-scale path loss. Specifically, we denote $h_{jk,x} \in \mathbb{C}^{M_k \times 1}$ as the small-scale fading vector (coefficient) between BS $b_k \in \Phi_k$ (jammer $j \in \Phi_J$) and receiver $x \in \{u, e\}$, where $\mathbb{C}$ denotes the complex number domain, $u$ and $e$ represent the user and the eavesdropper, respectively. As such, all the entries of $h_{bk,x}$ and $g_{jk,x}$ are identical and independent distributed (i.i.d.) circularly symmetric complex Gaussian random variables with zero mean and unit variance. For the path loss model, we use the standard power loss propagation model, i.e., the power of

1 We understand that the scenario where the users, the jammers, and the eavesdroppers are all equipped with multiple antennas is more general, interesting, as well as challenging. For analytical tractability, we leave it for future work.

2 In practice, the density of the jammers can be estimated by using the jammer detection policies [24].
We consider that each BS in tier \( k \) is equipped with a cache unit of size \( C_k \) (in files), \( C_k \leq \mathcal{N} \). As in \([7]–[13]\), file \( n \) is stored at each BS in tier \( k \in \mathcal{K} \) with a certain probability \( T_{n,k} \in [0,1] \), referred to as the caching probability of file \( n \) in tier \( k \). We denote \( \mathbf{T} \triangleq (T_{n,k})_{n \in \mathcal{N}} \in [0,1]^{\mathcal{N} \times 1} \), where \( \mathbf{T} \triangleq (T_{n,k})_{k \in \mathcal{K}} \in [0,1]^{\mathcal{K} \times 1} \), as the caching distribution of the \( \mathcal{N} \) files in the \( \mathcal{K} \)-tier HetNet. Then, the relationship between the caching probability of file \( n \) in tier \( k \) and the cache sizes of the BSs in tier \( k \) can be expressed as \([11]–[13]\).

\[
0 \leq T_{n,k} \leq 1, \quad n \in \mathcal{N}, \quad k \in \mathcal{K},
\]

\[
\sum_{n \in \mathcal{N}} T_{n,k} = C_k, \quad k \in \mathcal{K}.
\]

Let \( \Phi_{n,k} \subseteq \Phi_k \) and \( \Phi_{n,k} \subseteq \Phi_k, \ n \in \mathcal{N} \) denote the point processes of the BSs in tier \( k \) which store and do not store file \( n \), respectively. Then, we have \( \Phi_{n,k} \cup \Phi_{n,k} = \Phi_k. \) Due to the random caching and independent thinning \([26]\), we know that \( \Phi_{n,k} \) and \( \Phi_{n,k} \) are two thinned and independent homogeneous PPPs with densities \( \lambda_k T_{n,k} \) and \( \lambda_k (1 - T_{n,k}) \), respectively.

We now describe the user association rule adopted in this paper. Consider a user requesting file \( n \) at the beginning of a slot. If file \( n \) is not stored in the HetNet, it will not be served. Otherwise, it will be associated with a BS which not only stores file \( n \) but also provides the maximum long-term average received information signal power (ARISP) among all BSs in the \( \mathcal{K} \)-tier HetNet. This BS is referred to as the users’ serving BS and such association mechanism is called the content-based user association. Note that, under this association, a user may not be associated with the BS which provides the maximum ARISP if it has not stored file \( n \). As a result, the user usually receives a weak signal compared with the interference, and thus may not successfully receive the requested file and benefit from content diversity offered by random caching. To overcome this drawback, we assume that the channel state information (CSI) of each user is available at its serving BS and the serving BS adopts the maximal ratio transmission to deliver the requested files.

### C. Artificial-Noise-Aided Transmission

In order to deliberately confuse eavesdroppers in the HetNet while guaranteeing the reliable links to the users, we consider that each BS employs the artificial-noise-aided transmission strategy. Let \( s_{bk} \) denote the information-bearing signal with \( \mathbb{E}[|s_{bk}|^2] = 1 \) and \( \mathbf{v}_{bk} \in \mathbb{C}^{(M_b - 1) 	imes 1} \) denote the artificial noise vector at BS \( b \). Note that in this work, we only study serving cached files at BSs to get first-order insights into the design of cache-enabled \( \mathcal{K} \)-tier HetNets and characterize the benefits of caching, as in \([7]–[9]\). BSs may serve those uncached files through other service mechanisms, the investigation of which is beyond the scope of this paper.

### 3. Artificial-Noise-Aided Transmission

In order to deliberately confuse eavesdroppers in the HetNet while guaranteeing the reliable links to the users, we consider that each BS employs the artificial-noise-aided transmission strategy. Let \( s_{bk} \) denote the information-bearing signal with \( \mathbb{E}[|s_{bk}|^2] = 1 \) and \( \mathbf{v}_{bk} \in \mathbb{C}^{(M_b - 1) 	imes 1} \) denote the artificial noise vector at BS \( b \).

### 4. Artifical-Noise-Aided Transmission

In order to deliberately confuse eavesdroppers in the HetNet while guaranteeing the reliable links to the users, we consider that each BS employs the artificial-noise-aided transmission strategy. Let \( s_{bk} \) denote the information-bearing signal with \( \mathbb{E}[|s_{bk}|^2] = 1 \) and \( \mathbf{v}_{bk} \in \mathbb{C}^{(M_b - 1) 	imes 1} \) denote the artificial noise vector at BS \( b \). Note that in this work, we only study serving cached files at BSs to get first-order insights into the design of cache-enabled \( \mathcal{K} \)-tier HetNets and characterize the benefits of caching, as in \([7]–[9]\). BSs may serve those uncached files through other service mechanisms, the investigation of which is beyond the scope of this paper.

Note that, the analytical framework developed in this paper can be extended to other beamforming techniques.

---

**TABLE I. List of Notations.**

| Notation | Definition |
|----------|------------|
| \( \mathcal{K} \) | Set of network tiers. |
| \( K \) | Number of tiers. |
| \( \Phi_k \) | Point process of the BSs in tier \( k \). |
| \( \lambda_k \) | Spatial density of the BSs in tier \( k \). |
| \( M_k \) | Number of antennas at each BS in tier \( k \). |
| \( P_k \) | Transmit power of each BS in tier \( k \). |
| \( \Phi_j \) | Point process of the jammers. |
| \( \lambda_j \) | Spatial density of the jammers. |
| \( C \) | The complex number domain. |
| \( u \) | Notation representing for user. |
| \( e \) | Notation representing for eavesdropper. |
| \( h_{nk,x} \) | Small-scale fading vector between BS \( b_k \in \Phi_k \) and receiver \( x \in \{u,e\} \). |
| \( g_{j,x} \) | Small-scale fading coefficient between jammer \( j \in \Phi_j \) and receiver \( x \in \{u,e\} \). |
| \( \alpha \) | Path-loss exponent. |
| \( \mathcal{N} \) | Set of files in the network. |
| \( N \) | Number of files in the network. |
| \( a_n \) | Popularity of file \( n \). |
| \( \mathbf{a} \) | Popularity distribution of files in \( \mathcal{N} \). |
| \( C_b \) | Cache size of each BS in tier \( b \). |
| \( T_{n,k} \) | Caching probability of file \( n \) in tier \( k \). |
| \( \mathbf{T} \) | Caching distribution of the \( \mathcal{N} \) files. |
| \( \Phi_{n,k} \) | Point processes of the BSs in tier \( k \) which store file \( n \). |
| \( \Phi_{n,k} \) | Point processes of the BSs in tier \( k \) which do not store file \( n \). |
| \( s_{bk} \) | The information-bearing signal at BS \( b_k \). |
| \( \ell_{b,k} \) | The served BS of \( b_k \). |
| \( \mathbf{v}_{bk} \) | Artificial noise vector at BS \( b_k \). |
| \( \mathbf{w}_{bk} \) | Beamforming vector at BS \( b_k \). |
| \( \Phi_{b,k} \) | The matrix for transmitting the artificial noise at BS \( b_k \). |
| \( \phi_k \) | Fraction of power allocated to the signal \( s_{bk} \). |
| \( \epsilon_{a} \) | A typical user. |
| \( \epsilon_{e} \) | A typical eavesdropper. |
| \( r_{b,k} \) | Distance between BS \( b_k \) and \( \ell_{b,0} \). |
| \( r_{b,k} \) | Distance between BS \( b_k \) and \( \ell_{b,0} \). |
| \( d_{j,c} \) | Distance between jammer \( j \) and \( \ell_{c,0} \). |
| \( d_{j,c} \) | Distance between jammer \( j \) and \( \ell_{c,0} \). |
| \( b_{0,b,k} \) | Index of the serving BS of \( \ell_{c,0} \) in tier \( k \in \mathcal{K} \). |
| \( b_{0,b,k} \) | Index of an arbitrary BS storing file \( n \) in tier \( k \in \mathcal{K} \). |
| \( R_c \) | Target transmission rate of the wiretap code. |
| \( R_s \) | Secrecy rate of the wiretap code. |
| \( R_d \) | Redundancy rate against eavesdropping. |
| \( \epsilon \) | A given level of confidentiality. |
| \( p_{\text{RTP}}(T_n) \) | RTP of file \( n \). |
| \( p_{\text{RTP}}(T) \) | Average RTP. |
| \( p_{\text{CTP}}(T_n) \) | CTP of file \( n \). |
| \( p_{\text{CTP}}(T) \) | Average RTP. |

the transmitted signal with distance \( r \) is attenuated by a factor \( r^\alpha \), where \( \alpha > 2 \) denotes the path-loss exponent.
noise vector at BS \( b_k \). As per the rules of the artificial-noise-aided transmission strategy, all the entries of \( \nu_{b_k} \) are i.i.d. circularly symmetric complex Gaussian random variables with zero mean and variance \( \frac{\alpha}{\sqrt{M}} \). Then, the transmitted signal from BS \( b_k \in \Phi_k \) can be expressed as

\[
\tilde{s}_{b_k} = \sqrt{\phi_k P_k} w_{b_k} s_{b_k} + \sqrt{(1-\phi_k)P_k} W_{b_k} \nu_{b_k},
\]

where \( P_k \) denotes the transmit power of BS \( b_k \in \Phi_k \), \( \phi_k \in (0,1) \) denotes the fraction of power allocated to the information-bearing signal \( s_{b_k} \), \( w_{b_k} = \frac{\|b_{b_k,x,k}\|}{\|b_{b_k,x,k}\|} \) is the beamforming vector at BS \( b_k \) with \( \ell_k(b_k) \) representing the served user of BS \( b_k \) and \( \|h\| \) denoting the 2-norm of a vector \( h \), and \( W_{b_k} \in \mathbb{C}^{M_k \times (M-1)} \) is the matrix for transmitting the artificial noise at BS \( b_k \). We choose \( W_{b_k} \) as the projection matrix into the null space of \( h_{b_k,\ell_k(b_k)} \). As such, the artificial noise will not interfere with user \( \ell_k(b_k) \).

**D. Received Signal-to-Interference Ratios**

In this paper, w.l.o.g., according to Slivnyak’s Theorem \[26\], we focus on a typical receiver \( \ell_{x,0}, x \in \{a, e\} \) located at the origin. Let \( r_{b_k,\ell_{x,0}} \) denote the distance between BS \( b_k \in \Phi_k \) (jammer \( j \in \Phi_j \) and \( \ell_{x,0} \). Based on \([3]\), the received signal at \( \ell_{x,0} \) is given by \([4]\), as shown at the top of this page, where \( b_{n,k,0} \in \Phi_n,k,0 \) denotes the index of the serving BS of the typical user \( \ell_{x,0} \) in tier \( k_{x,0} \in K \), and \( h_{b_{n,k,0}} \) denotes the index of an arbitrary BS storing file \( n \) in tier \( k_{x,0} \in K \), and \( (\cdot)^T \) denotes the conjugate transpose operation. Then, the \( \text{SIR} \) at \( \ell_{x,0} \) can be expressed as \([5]\), as shown at the top of this page, where \( I_{b_{n,k,0},\ell_{x,0}}^{M_k} \), \( I_{n,k} \), \( I_{M_k} \), and \( I_{\ell_{x,0}} \) denote the power of artificial noise from BS \( b_{n,k,0} \), the interference and artificial noise from the BSs with storing file \( n \) in tier \( k_{x,0} \), the interference and artificial noise from the BSs without storing file \( n \) in tier \( k_{x,0} \), and the interference from jammers, respectively, given by \([6]\)–\([9]\), as shown at the top of the next page. Here, \( \xi_k \) is defined as

\[
\xi_k \triangleq \begin{cases} 
0, & \text{if } M_k = 1, \\
\phi_k^{-1} - 1, & \text{if } M_k = 2, 3, \ldots.
\end{cases}
\]

Note that, we have \( |g_{j,\ell_{x,0}}|^2 \sim \exp(1), |h_{b_{n,k,0}} w_{b_{n,k}}|^2 \sim \exp(1) \), \( |h_{b_{n,k,0}} w_{b_{n,k}}|^2 \sim \Gamma(M_k - 1, 1), \)

\( ^3 \)A typical user (eavesdropper) is a user (an eavesdropper) that is randomly selected from all the users (eavesdroppers) in the network.

\( ^4 \)The serving BS \( b_{n,k,0} \) of the typical user \( \ell_{x,0} \) is determined according to the content-based user association policy in Section II-B.

\( ^5 \)Note that \( X \overset{d}{=} Y \) means that the random variable \( X \) follows the distribution \( Y \).
In addition, the RTP (CTP) of file \( n \) can be thought of equivalently as i) the probability that a randomly chosen user (eavesdropper) succeeds in (fails to) decode file \( n \), or ii) the average fraction of users (eavesdroppers) who at any time succeed in (fail to) decode file \( n \).

Since each file \( n \in \mathcal{N} \) is requested with probability \( a_n \), according to the total probability theorem, the average RTP and CTP of a file, denoted as \( p_u(T) \) and \( p_e(T) \), respectively, are given by

\[
p_u(T) = \sum_{n \in \mathcal{N}} a_n p_{u,n}(T_n), \tag{15}
\]

\[
p_e(T) = \sum_{n \in \mathcal{N}} a_n p_{e,n}(T_n), \tag{16}
\]

where \( T \) is the design parameter related to random caching.

From (15) and (16), we see that the caching distribution \( T \) significantly affects both the RTP and the CTP. The key goal of this paper is to optimize \( T \) to maximize the average RTP of the system, subject to the caching size constraints in (1) and (2) as well as the constraint on the average CTP of the system, given by

\[
p_e(T) \geq \epsilon, \tag{17}
\]

where \( \epsilon \) denotes a given level of confidentiality. Mathematically, the optimization problem can be formulated as

**Problem 1 (Secrecy Performance Optimization):**

\[
T^* = \arg \max_T \ p_u(T) \quad \text{s.t.} \ (1), \ (2), \ (17),
\]

where \( T^* \) denotes an optimal solution.

## III. SECRECY PERFORMANCE ANALYSIS

In order to solve Problem 1 in this section, we first derive the closed-form expression of the RTP of file \( n \), i.e., \( p_{u,n}(T) \). Then, we derive the closed-form expression of the CTP of file \( n \), i.e., \( p_{e,n}(T) \). Note that, by substituting \( p_{u,n}(T) \) and \( p_{e,n}(T) \) into (15) and (16), respectively, we can directly obtain \( p_u(T) \) and \( p_e(T) \).

### A. Analysis of RTP

In this subsection, we analyze \( p_{u,n}(T) \), using tools from stochastic geometry. To calculate \( p_{u,n}(T) \), based on (11), we first need to analyze the distribution of the SIR, \( \text{SIR}_{t_n,k_n,\ell_n} \). Under random caching, we note that the interferers for the typical user at each tier can be classified into three categories, i.e., i) the interfering BSs storing the requested file of \( \ell_n \), ii) the interfering BSs without the requested file of \( \ell_n \) in each tier, iii) all the jammers in the network.

Under artificial-noise-aided transmission strategy, the artificial noise is embedded into the information signal transmitted from each BS in the network. Taking the impacts of these three categories of interferers and the artificial noise on the SIR of the typical user into account, we can derive the distribution of SIR at each tier, and then \( p_{u,n}(T) \), as summarized in the following theorem.

**Theorem 1 (RTP of file \( n \)):** The RTP of file \( n \) is given by

\[
p_{u,n}(T_n) = \sum_{j=1}^{K} \lambda_j T_n,j (\phi_j F_j)^\delta \left\| Q_{M,j}^{-1}(T_n) \right\|_1, \tag{18}
\]

where \( \left\| \cdot \right\|_1 \) is the \( L_1 \) induced matrix norm and \( Q_M(T_n) \) is a lower triangular Toeplitz matrix, i.e.,

\[
Q_M(T_n) = \begin{pmatrix}
q_0(T_n) & 0 & 0 \\
q_1(T_n) & q_0(T_n) & 0 \\
\vdots & \vdots & \ddots
\end{pmatrix} \tag{19}
\]

Here, \( q_m(T_n) = f_m(T_n, \theta_a) \), \( \theta_a \triangleq 2R_e - 1 \), \( m = 0, 1, \ldots, M_j - 1 \), where \( f_m(T_n, \theta) \) is given by (20), as shown at the top of the next page, with \( U_{m,M}(\xi, \theta), V_{m,M}(\xi, \theta) \) and \( W_m(\theta) \) given by (21)–(23), respectively, as shown at the top of the next page. Here, \( \eta_m(i, \xi, \theta) \triangleq \frac{(m+1)(m+2)}{m(\mu+2)} \Gamma(m+\delta+\ell_i, \theta) \), \( \kappa_m(i, \xi, \theta) \triangleq \frac{(-1)^m m! (\mu+2) \Gamma(1-\delta)}{i! \Gamma(1+\delta)} \sum_{\nu=0}^{\infty} \frac{\Gamma(\nu+1)}{\Gamma(\nu+1-\delta)} 2F_1(\alpha, \beta; \gamma) \) denote the Gauss hypergeometric function and Gamma function, respectively, and \( x(m) \triangleq x(x+1) \cdots (x+m-1) \) denote the falling and rising factorials, respectively.

**Proof:** See Appendix A.

We note that Theorem 1 provides a closed-form expression for \( p_{u,n}(T_n) \). In Fig. 1 we plot \( p_{u,n}(T_n) \) versus \( R_u \) for different values of \( \ell_j \). We see that the “analytical” curves, generated from Theorem 1 accurately match the points obtained from

\[\text{Note that when } m = 0, \text{ we define } x(0) = x^{(0)} = 1.\]
the derived expression of \( P^k \) as follows.

\[
\begin{align*}
  f_m(T_n, \theta) &\triangleq \sum_{k=1}^{K} \lambda_k (\phi_k P_k)^\delta (T_{n,k} U_{m,k}(\xi_k, \theta) + (1 - T_{n,k}) V_{m,k}(\xi_k, \theta)) + W_m(\theta), \\
  U_{m,k}(\xi, \theta) &\triangleq \begin{cases} 
    \eta_m(M - 1, 1, \theta), & \text{if } \xi = 1, \\
    \frac{1}{(1 - \xi)^m} \left( \eta_m(0, 1, \theta) - \sum_{i=0}^{M-2} (1 - \xi)^i \eta_m(i, \xi, \theta) \right), & \text{otherwise},
  \end{cases} \\
  V_{m,k}(\xi, \theta) &\triangleq \begin{cases} 
    \kappa_m(M + 1, 1, \theta), & \text{if } \xi = 1, \\
    \frac{1}{(1 - \xi)^m} \left( \kappa_m(2, 1, \theta) - \sum_{i=0}^{M-2} (1 - \xi)^i \kappa_m(i + 2, \xi, \theta) \right), & \text{otherwise},
  \end{cases} \\
  W_m(\theta) &\triangleq \lambda_3 P_3^\delta \kappa_m(2, 1, \theta).
\end{align*}
\]

**Proposition 1** (Upper and Lower Bounds of RTP of file \( n \)):
The RTP of file \( n \) satisfies \( p_{u,n}(T_n) \leq p_{u,n}(T_n) \leq p_{u,n}(T_n) \), where

\[
\begin{align*}
  p_{u,n}(T_n) &= 1 - \frac{1}{M^n} \sum_{k=1}^{K} \sum_{m=0}^{M-1} \left( \frac{M \lambda_k}{m} \right) \frac{(\phi_k P_k)^\delta}{f_0(T_n, m S_{M_k} \theta_m)}, \\
  p_{u,n}(T_n) &= \sum_{k=1}^{K} \sum_{m=0}^{M-1} \frac{\lambda_k}{S_{M_k} \theta_m} \left( 1 - \frac{1}{M^n} \right) f_m(T_n, \theta).
\end{align*}
\]

Here, \( S_{M_k} = (M_k!)^{-M_k^{-1}} \) and \( f_m(T_n, \theta) \) is given by (20).

**Proof:** See Appendix [A].

The tightness of \( p_{u,n}(T_n) \) and \( p_{u,n}(T_n) \) are evaluated in Fig. 1. We can see that \( p_{u,n}(T_n) \) and \( p_{u,n}(T_n) \) have the similar trends as \( p_{u,n}(T_n) \), which show the correctness of our analysis in Proposition [A]. In addition, we see that, compared to \( p_{u,n}(T_n) \) and \( p_{u,n}(T_n) \), \( p_{u,n}(T_n) \) tightly matches \( p_{u,n}(T_n) \), and thus can serve as a good approximation for \( p_{u,n}(T_n) \). Based on Proposition [A], some properties of \( p_{u,n}(T_n) \) are established as follows.

**Property 1** (Effects of Jammer Transmit Power and Density): \( p_{u,n}(T_n) \) increases with \( P_3 \) and \( \lambda_3 \).

**Proof:** The proof is straightforward and we omit it for brevity.

Property [A] indicates that the jammers will degrade the transmission reliability of file \( n \), since the jamming signals degrade the link qualities of the typical user’s channels. Fig. 2 plots \( p_{u,n}(T_n) \) versus \( P_3 \) and \( \lambda_3 \). From this figure, we see that although Property [A] is obtained based on \( p_{u,n}(T_n) \), it holds for \( p_{u,n}(T_n) \) as well. To obtain more insights, we consider a special case where all the BSs have the same number of antennas \( M \) and power allocation ratio \( \phi \), i.e., \( M_k = M \) and \( \phi_k = \phi \), for all \( k \in K \). In this special case, we can further obtain the following properties.

**Property 2** (Effects of BS Transmit Power and Density When \( M_k = M \) and \( \phi_k = \phi \)): When \( M_k = M \) and \( \phi_k = \phi \), for all \( k \in K \), \( p_{u,n}(T_n) \) increases with \( P_k \) and \( \lambda_k \), if \( T_{n,k} \geq T_{n,k}^{\text{th}} \), and decrease with \( P_k \) and \( \lambda_k \) otherwise, where \( T_{n,k}^{\text{th}} \) is given by (26), as shown at the top of the next page, and \( \xi \) is given by

\[
\xi \triangleq \begin{cases} 
  0, & \text{if } M = 1, \\
  \frac{1}{M - 1}, & \text{if } M = 2, 3, \ldots.
\end{cases}
\]

**Proof:** Denote \( z_k \triangleq \lambda_k P_k^\delta \). When \( M_k = M \) and \( \phi_k = \phi \),
\[ T_{n,k}^{\text{th}} = \frac{\sum_{m=0}^{M-1} (1 - \frac{m}{M}) V_{m,M}(\xi, \theta_a) \sum_{i \neq j} \lambda_i P_i^k \phi_i^\delta + \sum_{m=0}^{M-1} (1 - \frac{m}{M}) W_m(\theta_a)}{\sum_{m=0}^{M-1} (1 - \frac{m}{M}) V_{m,M}(\xi, \theta_a) \sum_{i \neq j} \lambda_i P_i^k \phi_i^\delta}. \]  

(26)

Fig. 3. (a) \( p_{u,n}(T_n) \) versus \( P_2 \). (b) \( p_{e,n}(T_n) \) versus \( P_2 \). Here, \( K = 2, M_1 = M_2 = 4, \phi_1 = 0.9, \phi_2 = 0.5, P_1 = 20 \, \text{W}, P_2 = 1 \, \text{W}, \lambda_1 = \frac{1}{1500\pi}, \lambda_2 = \frac{1}{1500\pi}, \alpha = 3.5, R_n = 1.3 \, \text{bps/Hz}, \) and \( T_{n,1} = 0.9 \).

Finally, by letting \( \frac{\partial p_{u,n}(T_n)}{\partial P_k} \geq 0 \) and \( \frac{\partial p_{u,n}(T_n)}{\partial \lambda_k} \geq 0 \) or \( \frac{\partial p_{u,n}(T_n)}{\partial P_k} < 0 \) and \( \frac{\partial p_{u,n}(T_n)}{\partial \lambda_k} < 0 \), we obtain the desired result in (26), which completes the proof.

Property 2 indicates that if the caching probability of file \( n \) in tier \( k \) is relatively low, then increasing the BS transmit power or density in this tier will decrease the transmission reliability of file \( n \). This can be explained as follows. If file \( n \) is stored in tier \( k \) with a very low probability, all the BSs in this tier will be the interferer of the typical user. As such, the increase of the BS transmit power or density in tier \( k \) will increase the interference suffered by the typical user. Fig. 3 plots \( p_{u,n}(T_n) \) versus \( P_2 \) and \( \lambda_2 \). From the figure, we see that Property 2 also holds for \( p_{e,n}(T_n) \).

Property 3 (Effect of Caching Probability When \( M_k = M \) and \( \phi_k = \phi, k \in K \)): When \( M_k = M \) and \( \phi_k = \phi \), for all \( k \in K \), \( p_{u,n}(T_n) \) increases with \( T_{n,k} \).

Proof: See Appendix C.

B. Analysis of CTP

In this subsection, we analyze \( p_{e,n}(T_n) \), using tools from stochastic geometry. To derive \( p_{e,n}(T_n) \), we must derive the joint probability distribution of SIR_{b_{n,k},e,k} for all \( k_{e,0} \in K, b_{n,k}, e_{0} \in K_{b_{n,k},0} \), which, however, is extremely challenging due to the fact that the variables SIR_{b_{n,k},e,k} for all \( k_{e,0} \in K, b_{n,k}, e_{0} \in K_{b_{n,k},0} \) are correlated with each other. In order to make progress, motivated by [32], we focus on the high redundancy rate scenario, i.e., \( R_e > 1 \, \text{bps/Hz} \). Note that, it shows in [32] Lemma 1 that in this scenario, at most one BS in the entire network can provide channel capacity greater than \( R_e \), i.e., the typical eavesdropper can successfully decode message from at most one BS. Then, in this scenario, by carefully characterizing the impact of random caching and the artificial-noise-aided transmission strategy on the distribution of SIR_{b_{n,k},e,k} for each \( k_{e,0} \in K, b_{n,k}, e_{0} \in K_{b_{n,k},0} \),
we can derive the distribution of SIR\(b_{n,k,0}f_{x,0}\), for each \(k_{x,0}\in K\), \(b_{n,k,0}\in \Phi_{n,k}\), and then \(p_{e,n}(T_n)\), as summarized in the following theorem.

**Theorem 2 (CTP of file \(n\)):** When \(R_e > 1\), the CTP of file \(n\) is given by

\[
p_{e,n}(T_n) = 1 - \sum_{k=1}^{K} \lambda_k T_{n,k}^e(k + \xi_k)\psi_0\theta_k \theta_\psi \left[ 1 - \frac{W_0(\theta_k)}{W_0(\theta_k) + \sum_{j=1}^{K} \lambda_j T_{n,j}^e(k + \xi_j)\psi_0\theta_j \theta_\psi} \right] \]

(30)

where \(\theta_e \equiv 2\theta_e - 1\), \(V_{0,M}(\xi_e, \theta)\) and \(W_0(\theta)\) are given by (22) and (23), respectively.

**Proof:** See Appendix [E].

We note that Theorem 2 provides a closed-form expression for \(p_{e,n}(T_n)\) in the high redundancy rate scenario. In Fig. 5, we plot \(p_{e,n}(T_n)\) versus \(R_e\) for different values of \(\lambda_1\). We see that the “Analytical” curves, obtained from Theorem 2, closely match the “Monte Carlo” simulation points, indicating that Theorem 2 can also serve as a good approximation for \(p_{e,n}(T_n)\) in the low redundancy rate scenario.

Based on Theorem 2, we can obtain important insights into the behavior of \(p_{e,n}(T_n)\). We study the system parameters (e.g., \(P_j\), \(\lambda_j\), \(P_k\), \(T_n\), as follows.

**Property 5 (Effects of Jammer Transmit Power and Density):** \(p_{e,n}(T_n)\) increases with \(P_j\) and \(\lambda_j\).

**Proof:** The proof is straightforward and we omit it for brevity.

Property 5 indicates that, although the jammers in the HetNets can degrade the transmission reliability of file \(n\), they can improve the transmission confidentiality of this file, as shown in Fig. 5. This is because the jamming signals degrade the link qualities of the eavesdroppers’ channels.

**Property 6 (Effects of BS Transmit Power and Density):** \(p_{e,n}(T_n)\) decreases with \(P_k\) and \(\lambda_k\) if \(T_{n,k}^e\), and increases with \(P_k\) and \(\lambda_k\) otherwise, where

\[
T_{n,k}^e = \left[ V_{0,M}(\xi_e, \theta_e) \sum_{j=1}^{K} \lambda_j T_{n,j}^e(k + \xi_j)\psi_0\theta_j \theta_\psi \left[ 1 - \frac{W_0(\theta_j)}{W_0(\theta_j) + \sum_{j=1}^{K} \lambda_j T_{n,j}^e(k + \xi_j)\psi_0\theta_j \theta_\psi} \right] \right]^{-1}.
\]

**Proof:** The proof is similar to that in Property 5 and we omit it for brevity.

**Property 7 (Effect of Caching Probability):** \(p_{e,n}(T_n)\) decreases with \(T_{n,k}\).

**Proof:** From (30), we have

\[
\frac{\partial p_{e,n}(T_n)}{\partial T_{n,k}^e} = - \frac{\lambda_k (k + \xi_k)\psi_0\theta_k \theta_\psi \left[ 1 - \frac{W_0(\theta_k)}{W_0(\theta_k) + \sum_{j=1}^{K} \lambda_j T_{n,j}^e(k + \xi_j)\psi_0\theta_j \theta_\psi} \right]}{W_0(\theta_k) + \sum_{j=1}^{K} \lambda_j T_{n,j}^e(k + \xi_j)\psi_0\theta_j \theta_\psi}.
\]

Obviously, we see that \(\frac{\partial p_{e,n}(T_n)}{\partial T_{n,k}^e} < 0\) is a constant w.r.t. \(T_{n,k}\), which completes the proof of this property.

Property 7 indicates that caching a file at more BSs will compromise the transmission confidentiality of this file, since caching the file at more BSs incurs a higher risk of being eavesdropped. Fig. 6 verifies Property 7.

**Property 8 (Linearity of \(p_{e,n}(T_n)\)):** \(p_{e,n}(T_n)\) is a linear function of \(T_{n,k}\).

**Proof:** The proof is straightforward and we omit it for brevity.
the Euclidean projection of \(\Pi_{\mathcal{P}}(\mathbf{T})\) with a diminishing step size. If it is chosen improperly, Algorithm 1 may need a great number of iterations to converge. Recall that in Property 4 in the special case of \(M_k = M\) and \(\phi_k = \phi\), for all \(k \in \mathcal{K}\), \(P_{u,n}(\mathbf{T}_n)\) has a DC structure. Thus, this case Problem 2 becomes a DC optimization problem and thus can be solved using convex-concave procedure (CCP) [36]. Compared to GPM, CCP does not depend any step size and thus may lead to robust convergence performance. As such, in the following, we consider this special case and develop an iterative algorithm to obtain a stationary point of Problem 2 based on CCP. The core idea of CCP is to linearize the convex term of the DC objective function to obtain a concave objective function for a maximization problem, and then solve a sequence of convex optimization problems successively. Specifically, at iteration \(t + 1\), based on Property 4 we have the following approximation problem:

**Problem 3 (Approximation of Problem 2 at Iteration \(t + 1\))**

When \(M_k = M\) and \(\phi_k = \phi\), for all \(k \in \mathcal{K}\):

\[
\mathbf{T}^{(t+1)} \triangleq \arg \max_{\mathbf{T}} \sum_{n \in \mathcal{N}} a_n \left( P_{u,n}^{(1)}(\mathbf{T}_n) - P_{u,n}^{(2)}(\mathbf{T}_n; \mathbf{T}_n^{(t)}) \right) + 1
\]

s.t. (1), (2), (17)

where \(P_{u,n}^{(1)}(\mathbf{T}_n)\) is given by (29) and is the gradient of \(P_{u,n}^{(1)}(\mathbf{T}_n)\) at \(\mathbf{T} = \mathbf{T}_n^{(t)}\), given by (34), as shown at the top of the next page. Here, (33) is given by (24). As such, we can transform Problem 1 into the following problem:

**Problem 2 (Simplified Secrecy Performance Optimization):**

\[
\mathbf{T}^* \triangleq \arg \max_{\mathbf{T}} P_{u,n}^{(1)}(\mathbf{T})
\]

s.t. (1), (2), (17)

Problem 2 maximizes a non-concave objective function over a linear set, denoted by \(\mathcal{T} \triangleq \{ \mathbf{T} \in [0,1]^{|\mathcal{K}|} (1), (2), (17) \}\), and hence is a non-convex optimization problem in general. Since the objective function \(P_{u,n}^{(1)}(\mathbf{T})\) is continuously differentiable on \(\mathcal{T}\), we can obtain a stationary point of Problem 2 using the gradient projection method (GPM) with a diminishing step size, denoted by \(\zeta^{(t)}\) with \(t = 0, 1, 2, \ldots\) being the iteration index. The details for solving Problem 2 using GPM are summarized in Algorithm 1 where the diminishing step size satisfies \(\zeta^{(t)} > 0\), as \(t \to \infty\), \(\sum_{t=1}^{\infty} \zeta^{(t)} = \infty\) and \(\sum_{t=1}^{\infty} (\zeta^{(t)})^2 < \infty\) [33, pp. 227]. In addition, in Step 3 of Algorithm 1 \(\Pi_{\mathcal{T}}(\mathbf{T}) \triangleq \arg \min_{\mathbf{T} \in \mathcal{T}} \| \mathbf{T} - \mathbf{T} \|\) denotes the Euclidean projection of \(\mathbf{T}\) onto \(\mathcal{T}\), and \(\nabla_{\mathbf{T}^*} P_{u,n}^{(1)}(\mathbf{T}) \triangleq (\partial P_{u,n}^{(1)}(\mathbf{T}_n; \mathbf{T}_n^{(t)}) / \partial \mathbf{T}_n)_{|\mathbf{T}_n = \mathbf{T}_n^{(t)}}\) is the gradient of \(P_{u,n}^{(1)}(\mathbf{T})\) at \(\mathbf{T} = \mathbf{T}_n^{(t)}\), given by (34), as shown at the top of the next page. Here,

\[\lambda_k(M_k, \phi_k, \theta) \triangleq \lambda_k(\phi_k P_k)^{1/2} (U_{0,M_k}(\xi_k, \theta) - V_{0,M_k}(\xi_k, \theta))\]

and \(\psi(M_k, \phi_k, \theta) \triangleq \sum_{k \in \mathcal{K}} \lambda_k(\phi_k P_k)^{1/2} (U_{0,M_k}(\xi_k, \theta) + V_{0,M_k}(\xi_k, \theta))\), where \(U_{m,M}(\xi, \theta)\), \(V_{m,M}(\xi, \theta)\) and \(W_{m,M}(\xi, \theta)\) are given by Theorem 1. According to [34], we know that the sequence \(\{\mathbf{T}^{(t)}\}_{t=1}^{\infty}\) generated by Algorithm 1 converges to a stationary point of Problem 2. The computation cost of Algorithm 1 is dominated by the Euclidean projection at Step 3, which has the computational complexity \(O(|\mathcal{K}|^3)\) in the use of interior point algorithm [35].

Note that, the convergence rate of Algorithm 1 is sensitive to the choice of the step size. If it is chosen improperly, Algorithm 1 may need a great number of iterations to converge. Recall that in Problem 2, the objective function \(\mathbf{T} \in \{ \mathbf{T} \in [0,1]^{|\mathcal{K}|} (1), (2), (17) \}\), and hence is a non-convex optimization problem in general. Since the objective function \(P_{u,n}^{(1)}(\mathbf{T})\) is continuously differentiable on \(\mathcal{T}\), we can obtain a stationary point of Problem 2 using the gradient projection method (GPM) with a diminishing step size, denoted by \(\zeta^{(t)}\) with \(t = 0, 1, 2, \ldots\) being the iteration index. The details for solving Problem 2 using GPM are summarized in Algorithm 1 where the diminishing step size satisfies \(\zeta^{(t)} > 0\), as \(t \to \infty\), \(\sum_{t=1}^{\infty} \zeta^{(t)} = \infty\) and \(\sum_{t=1}^{\infty} (\zeta^{(t)})^2 < \infty\) [33, pp. 227]. In addition, in Step 3 of Algorithm 1 \(\Pi_{\mathcal{T}}(\mathbf{T}) \triangleq \arg \min_{\mathbf{T} \in \mathcal{T}} \| \mathbf{T} - \mathbf{T} \|\) denotes the Euclidean projection of \(\mathbf{T}\) onto \(\mathcal{T}\), and \(\nabla_{\mathbf{T}^*} P_{u,n}^{(1)}(\mathbf{T}) \triangleq (\partial P_{u,n}^{(1)}(\mathbf{T}_n; \mathbf{T}_n^{(t)}) / \partial \mathbf{T}_n)_{|\mathbf{T}_n = \mathbf{T}_n^{(t)}}\) is the gradient of \(P_{u,n}^{(1)}(\mathbf{T})\) at \(\mathbf{T} = \mathbf{T}_n^{(t)}\), given by (34), as shown at the top of the next page. Here,
\[
p_c(T) = \sum_{n \in N} a_n \sum_{j \in K} \lambda_j T_{n,j} \left( \phi_j P_j \right)^\delta \left\| Q_{M_j}^{-1}(T_n) \right\|_1,
\]
(31)
\[
p_e(T) = \sum_{n \in N} a_n \left( 1 - \sum_{k=1}^{K} \frac{\lambda_k T_{n,k}(\phi_k P_k)^\delta \left(1 + \xi_k \theta_k\right)^{-M_k}}{W_0(\theta_k) + \sum_{k=1}^{K} \lambda_k (\phi_k P_k)^\delta V_{0,M_k}(\xi_k, \theta_k)} \right).
\]
(32)
\[
\frac{\partial p_{1j}(T)}{\partial T_{n,k}} = -a_n \sum_{m=1}^{M_k} \frac{M_k}{m} f_{0j}(T_n, mS_{M_k} \theta_0) \left( \sum_{i \in K \setminus \{k\}} \lambda_i (\phi_i P_i)^\delta T_{n,i} A_i (M_k, \phi_k, mS_{M_k} \theta_0) + \Psi(M_k, \phi_k, mS_{M_k} \theta_0) \right).
\]
(34)
\[
\frac{\partial p_{2j}(T)}{\partial T_{n,k}} = -a_n \sum_{m=1}^{M_k} \frac{M_k}{m} f_{0j}(T_n, mS_{M_k} \theta_0) \left( \sum_{i \in \hat{K} \setminus \{k\}} \lambda_i (\phi_i P_i)^\delta T_{n,i} A_i (M, \phi, mS_{M_k} \theta_0) + \Psi(M, \phi, mS_{M_k} \theta_0) \right)
\]
(35)

Problem 2 using CCP are summarized in Algorithm 2. To initialize Algorithm 2 we can randomly choose a point and then project it onto the linear constraint set of Problem 2. According to [56], we declare that the sequence \(\{T(t)\}_{t=0}^\infty\) generated by Algorithm 2 converges to a stationary point of Problem 2. Similar as Algorithm 1, the computation cost of Algorithm 2 is dominated by the solution of Problem 3 at Step 3 as well, which has the computational complexity \(O((KN)^3)\) in the use of interior point algorithm [55].

V. NUMERICAL RESULTS

In this section, we provide numerical results to verify the effectiveness of our proposed secure random caching scheme. Specifically, we first demonstrate the convergence of Algorithm 1 and Algorithm 2. Then, we compare the performance of the proposed secure random caching scheme with that of some existing baseline schemes. In the simulations, we consider a two-tier HetNet, i.e., \(K = 2\), consisting of a macrocell network as the 1st tier overlaid with a picocell network as the 2nd tier. Unless otherwise stated, the simulation settings are as follows: \(\epsilon_{err} = 10^{-6}\), \(t_{max} = 100\), \(P_1 = 20\) W, \(P_2 = 0.13\) W, \(P_3 = 1\) W, \(\lambda_1 = \frac{200}{250+7}\), \(\lambda_2 = \frac{250}{250+7}\), \(\lambda_3 = \frac{200}{200+7}\), \(\alpha = 4\), \(R_s = 1.3\) bps/Hz, \(R_c = 0.2\) bps/Hz, \(c = 0.7\), \(M_1 = M_2 = 10\), \(\phi_1 = \phi_2 = 0.9\), \(C_1 = 15\), \(C_2 = 10\), \(N = 20\) and \(a_n = \frac{\sum_{i \in N} a_i}{n}\), where \(\beta = 0.6\) is the Zipf exponent.

A. Convergence of Algorithm 1 and Algorithm 2

In Fig. 9 we present the convergence trajectories of Algorithm 1 and Algorithm 2 as functions of iteration index for five different initial points. We see that the convergence rate of Algorithm 1 is highly sensitive to the choices of step size, while Algorithm 2 has more robust convergence performance, due to the fact that it does not rely on any step size. We also see that, for different initial points, both Algorithm 1 and Algorithm 2 converge to the same RTP, demonstrating their effectiveness in solving Problem 2. Moreover, compared to Algorithm 2, the convergence trajectory of Algorithm 1 is not monotonically decreasing, due to the Euclidean projection onto the feasible set.

B. Performance Comparisons between Proposed Scheme and Baselines

In Fig. 10 – Fig. 13 we examine the effects of system parameters and demonstrate the superiority of our proposed secure random caching schemes over two baseline schemes, in terms of the average RTP. Specifically, Baseline 1 adopts the most popular caching scheme, where each BS in tier \(k\) selects the \(C_k\) most popular files to store, and Baseline 2 adopts the uniform caching scheme, where each BS in tier \(k\) randomly selects \(C_k\) files to store, according to the uniform distribution.

Fig. 10 shows the effect of the confidential level \(\epsilon\). Specifically, in Fig. 10(a) we plot the average RTP versus \(\epsilon\). From the figure, we can see that for a large \(\epsilon\), the two baseline schemes may be infeasible (e.g., \(\epsilon > 0.65\) for the most popular caching and \(\epsilon > 0.75\) for the uniform caching), but our proposed secure random caching scheme may still be feasible (e.g., \(\epsilon < 0.81\)), at the cost of RTP decrement (e.g., \(0.65 < \epsilon < 0.81\)). This can be explained as follows. The
gradually decreases with \( \epsilon \), which in turn leads to a reduction of the average RTP.

In Fig. 11(b), we examine the effects of the jammer transmit power. (b) Effect of the jammer density.

CTPs of the two baseline schemes (each with fixed caching distribution of files) are deterministic if given the system parameters. Therefore, the decrease of the confidential level will easily violate the confidential level constraint in (17).

On the contrary, our proposed scheme can wisely adjust the caching distribution of the files to increase the CTP such that the confidential level constraint is satisfied. The corresponding adjustment in caching distribution is illustrated in Fig. 11(b). To be specific, in the region of \( \epsilon < 0.65 \), a file with higher popularity has a higher caching probability, while in the region of \( \epsilon \geq 0.65 \), the caching probability of a file with higher popularity gradually decreases with \( \epsilon \), which in turn leads to a reduction of the average RTP.

In Fig. 11 we examine the effects of the transmit power and density of the jammers, i.e., \( P_j \) and \( \lambda_j \). We first see that the average RTP of each caching scheme decreases as \( P_j \) or \( \lambda_j \) increases, which verifies Property 11. We also see that, when \( P_j \) or \( \lambda_j \) is relatively high (e.g., \( P_j \geq 90 \text{ W} \) or \( \lambda_j \geq \frac{180}{2500} \text{ W} \)), the most popular caching can achieve almost the same average RTP as our secure random caching. This is because, when \( P_j \) or \( \lambda_j \) is large, the received SIR at the typical user is small. As such, our secure random caching scheme tends to cache the files with higher popularity in order to provide the largest possible average RTP. In addition, we see that, when \( P_j \) or \( \lambda_j \) is low (e.g., \( P_j \leq 1 \text{ W} \) or \( \lambda_j \leq \frac{16}{2500} \text{ W} \)), the most popular caching may be infeasible, but our secure random caching scheme can still be feasible, demonstrating that compared to the baselines, our proposed scheme can effectively resist jamming attack.

Next, we examine the effects of the number of BS antennas \( M_k \), the power allocation \( \phi_k \), the target transmission rate \( R_u \), and the secrecy rate \( R_s \) in Fig. 12(a). We see that, when \( M_1 \) is small, \( \phi_1 \) is low or \( R_u \) is large, the most popular caching scheme can achieve almost the same average RTP as our proposed secure random caching scheme, since caching the most popular files at each BS can compensate lower received SIR at the typical user. In addition, we also see that when \( \phi_1 \) is large, \( R_u \) is small or \( R_s \) is large, the two baseline schemes may be infeasible but our proposed schemes may still be feasible.
denoting that our schemes can well adapt to the changes of these parameters.

Finally, we examine the effects of the cache size $C_k$ in Fig. 13. From the figure, we can see that, when the cache size is large enough, the baselines may not be feasible, while our proposed scheme can still be feasible, but at the cost of the average RTP decrement. This observation indicates that increasing the cache sizes at the BSs does not always increase the average RTP, which is significantly different from the observations in the existing works without considering secrecy constraints, e.g., [10]–[13]. This can be explained as follows. If the cache size increases, for the most popular caching scheme, each BS will store more files, while for the uniform caching scheme, each BS will store more files, both of which will increase the risk of eavesdropping, i.e., the average CTP for the two baselines will decrease. If the average CTP is below the given confidential level, the baselines will be infeasible. On the other hand, our proposed secure caching scheme can wisely adjust the caching distribution of the files (i.e., reduce the caching probability of a file with higher popularity as shown in Fig. 10 (b)) to maintain the average CTP above the given confidential level. However, such adjustment leads to a reduction of the average RTP.

VI. CONCLUSIONS

In this paper, we examine the security issue in the cache-enabled large-scale multi-antenna HetNets, in which the multi-antenna BSs deployed at multiple independent network-tiers deliver their cached files to the requesting users in the presence of the eavesdroppers and the jammers. To confuse the eavesdroppers, the BSs transmits artificial noise as well as the useful signals, simultaneously. We first derive closed-form expressions for the average RTP and the average CTP, characterizing the impacts of the eavesdroppers and jammers on the secrecy performance of the system. We then derive closed-form upper and lower bounds for the RTP, which facilitates us to understand the effects of key system parameters. In addition, we propose a secure random caching scheme which optimizes the caching distribution of the files to maximize the average RTP of the system, while simultaneously meeting the requirements on the caching sizes at the BSs and the average CTP of the system. Numerical results show that the proposed secure random caching scheme significantly outperforms the existing baseline solutions.

APPENDIX A

PROOF OF THEOREM 1

Denote $I_n \triangleq \sum_{k \in \mathcal{K}} (I_{n,k}^u + I_{n,k}^l) + I_{n}^e$ and $\theta_n \triangleq 2^{R_n} - 1$. Substituting (11) into (13), we have (35), as shown at the top of the next page, where (a) is due to the fact that $h_{n,k,\ell_0,e_0,\ell_0}^{-1} w_{b_{n,k,\ell_0,e_0}}^2 \sim \Gamma(M_{k_0,0},1)$ and $A_{k_0,0}$ is the probability that the typical user $k_0,0$ is associated with tier $k_0,0$; (b) follows from $\mathbb{E}[I_n] \sim (1)^m \mathcal{L}_n(s)$ with $\frac{d^m}{ds^m} \mathcal{L}_n(s)$ denoting the $m$-order derivative of the Laplace transform of random variable $I_n$, i.e., $\mathcal{L}_n(s) \triangleq \mathbb{E}[I_n^e e^{-st}]$; (c) uses the probability density function (p.d.f.) of the distance $r_{b_{n,j},e_0,0}$, i.e., $f_{r_{b_{n,j},e_0,0}}(r_0)$, which is given by

\[
f_{r_{b_{n,j},e_0,0}}(r_0) = \frac{2\pi \lambda J_{n,j} A_{j}}{A_j} \exp(-r_0^2 \sum_{k=1}^{K} \pi \lambda_k T_{n,k} \left(\frac{\phi_k P_k}{\sigma^2_f R_j}\right)^\delta). \tag{37}
\]

Thus, to calculate $p_{u,n}(T_n)$, we only need to calculate $\frac{d^m}{ds^m} \mathcal{L}_n(s)$.

First, we calculate $\mathcal{L}_n(s)$. Note that we have $\mathcal{L}_n(s) = \prod_{k \in \mathcal{K}} \mathcal{L}_{n,k}^u(s) \mathcal{L}_{n,k}^l(s) \mathcal{L}_{n,k}^e(s)$. In the following, we calculate $\mathcal{L}_{n,k}^u(s)$, $\mathcal{L}_{n,k}^l(s)$, and $\mathcal{L}_{n,k}^e(s)$, respectively. Let $X_b \triangleq \phi_k P_k (h_{k,e_0,0} w_b^1 + \xi_k |h_{e_0,0} w_b^2|$ with the p.d.f. being given by [37] Lemma 1. We calculate $\mathcal{L}_{n,k}^e(s)$ as (38), as shown at the top of the next page, where (d) follows from the probability generating functional (PGFL) over PPP and $\mathcal{L}_{X_b}(s)$ is the Laplace transform of $X_b$; (e) follows from [38] eq. (49)) and $A(s)$ and $B(s)$ are, respectively, given by (49) and (40), as shown at the top of the next page, with $\varphi(i,x) \triangleq 2 F_1(\delta, i + \delta, i + \delta + 1; -x^{-1})$. Similarly, $\mathcal{L}_{n,k}^u(s)$ and $\mathcal{L}_{n,k}^l(s)$ are calculated as (41) and (42), respectively, as shown at the top of the next page. Based on (38), (41) and (42), we have

\[
\mathcal{L}_n(s) = \exp(\eta(s)), \tag{43}
\]

where $\eta(s)$ is given by (44), as shown at the top of the next page, with

\[
f_k(s) \triangleq \frac{1}{T_{n,k}} \left(\frac{\phi_k P_k}{\sigma^2_f R_k}\right)^\delta A(s) - \frac{r_0^2}{2} B(s) + g_k(s) \triangleq \frac{1}{2} \left(2 F_1(\delta, M_k; 1 - \delta; -sr_0^{-\alpha}) - 1\right) + \frac{1}{T_{n,k}} - 1 \right)^\delta \Gamma(M_k,0) \Gamma(1 - \delta).
\]

Next, we calculate $\frac{d^m}{ds^m} \mathcal{L}_n(s)$. Note that, directly computing the derivatives will lead to intractable expressions. To address this issue, according to [39] Lemma 1, we obtain the following recursive relations

\[
\frac{(-s)^m}{m!} \frac{d^m}{ds^m} \mathcal{L}_n(s) = \begin{cases} \mathcal{L}_n(s), & \text{if } m = 0, \\ \sum_{i=0}^{m-1} \frac{m-i}{m} z_{m-i}(-s)^i \frac{d^i}{ds^i} \mathcal{L}_n(s), & \text{if } m = 1, 2, \ldots, \end{cases} \tag{45}
\]

where

\[
z_m(s) = \frac{(-s)^m}{m!} \frac{d^m}{ds^m} \eta(s). \tag{46}
\]

From (45), we can see that in order to calculate $\frac{d^m}{ds^m} \mathcal{L}_n(s)$, we only need to calculate $z_m(s)$, which are related to the
\[
\mathbf{p}_{w,n}(T_n) = \mathbb{E}_{k_n} \mathbb{E}_{\theta_{k_n},\kappa_n,\kappa_0,\kappa_0} \mathbb{E}_{I_n} \left[ \Pr \left[ \left| \mathbf{b}_n^T \mathbf{r}_{k_n,\kappa_n,\kappa_0,\kappa_0}^\alpha \mathbf{w}_{b_n,\kappa_n,\kappa_0} \right| \geq \theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0} I_n \right] \right] \\
\overset{(a)}{=} K \sum_{k_n=0}^{K} A_{k_n} \mathbb{E}_{\theta_{k_n},\kappa_n,\kappa_0,\kappa_0} \mathbb{E}_{I_n} \left[ \sum_{m_{k_n}=0}^{M_n} \left( \frac{(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0})^{m_{k_n}}}{m_{k_n}!} \right) \right] e^{-\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0} I_n} \\
\overset{(b)}{=} K \sum_{j=1}^{K} A_j \mathbb{E}_{\theta_{j},\kappa_j,\kappa_0,\kappa_0} \left[ \sum_{m_j=0}^{M_j} \left( -s \right)^{m_j} \frac{d^{m_j}}{d s^{m_j}} \mathcal{L}_{I_n}(s) \right] \bigg|_{s=\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}} \\
\overset{(c)}{=} K \sum_{j=1}^{K} \int_{0}^{\infty} f_{r_{b_n,\kappa_n,\kappa_0,\kappa_0}}(r_0) \left[ \sum_{m_j=0}^{M_j} \left( -s \right)^{m_j} \frac{d^{m_j}}{d s^{m_j}} \mathcal{L}_{I_n}(s) \right] \bigg|_{s=\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}} \, dr_0. 
\]

(36)

\[
\mathcal{L}_{I_n}(s) = \mathbb{E}_{I_n} \left[ e^{-\mathbf{I}_n} \right] = \exp \left( -2\pi \lambda \hat{T}_{n,k} \int_{0}^{\infty} \left( \frac{d^{m_j}}{d s^{m_j}} \mathcal{L}_{I_n}(s) \right) \bigg|_{s=\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}} \right)
\]

(38)

\[
A(s) \equiv s \delta \left( \frac{\Gamma(1 + \delta) \Gamma(1 - \delta)}{(1 - \xi_k)^{M_k - 1}} - \sum_{i=0}^{M_k - 2} \xi_k^{i+\delta} \Gamma(i+\delta+1) \Gamma(1-\delta) \right), \tag{39}
\]

\[
B(s) \equiv 1 - \frac{\delta \varphi(1, \frac{P_k}{\varphi_{T_k}} \sqrt{\sigma_0^\alpha})}{(1 + \delta) \Gamma(1 - \xi_k)^{M_k - 1}} \sum_{i=0}^{M_k - 2} \frac{\delta \xi_k \varphi(i+1, \xi_k \frac{P_k}{\varphi_{T_k}} \sqrt{\sigma_0^\alpha})}{(i+1) \Gamma(1 - \xi_k)^{M_k - i-1}}. \tag{40}
\]

\[
\mathcal{L}_{I_n}(s) = \exp \left( -\pi \lambda_k (\hat{T}_{n,k} - \mathbf{g}_s) \right) \bigg|_{s=\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}} \left( \frac{\varphi_{T_k}}{\sqrt{\sigma_0^\alpha}} \right) s^\delta \Gamma(1 + \delta) \Gamma(1 - \delta) \right), \tag{41}
\]

\[
\mathcal{L}_{I_n}(s) = \exp \left( -\pi \lambda_k \left( \frac{P_k}{\varphi_{T_k}} \right) \frac{\delta}{\varphi_{T_k}} s^\delta \Gamma(1 + \delta) \Gamma(1 - \delta) \right). \tag{42}
\]

\[
\eta(s) = \begin{cases} 
-\pi \lambda_j \mathbf{T}_{n,j} \left( \sum_{k=1}^{K} \frac{\lambda_k \mathbf{T}_{n,k}}{\lambda_j \mathbf{T}_{n,j}} \left( \frac{\varphi_{T_k}}{\varphi_{T_j}} \right) \delta \mathbf{g}_s(s) + \frac{\lambda_j \mathbf{T}_{n,j}}{\lambda_j \mathbf{T}_{n,j}} \left( \frac{\varphi_{T_j}}{\varphi_{T_j}} \right) \delta \mathbf{g}_s(1 + \delta) \Gamma(1 - \delta) \right), & \text{if } \xi_k = 1, \\
-\pi \lambda_j \mathbf{T}_{n,j} \left( \sum_{k=1}^{K} \frac{\lambda_k \mathbf{T}_{n,k}}{\lambda_j \mathbf{T}_{n,j}} \delta \mathbf{g}_s(s) + \frac{\lambda_j \mathbf{T}_{n,j}}{\lambda_j \mathbf{T}_{n,j}} \left( \frac{\varphi_{T_j}}{\varphi_{T_j}} \right) \delta \mathbf{g}_s(1 + \delta) \Gamma(1 - \delta) \right), & \text{otherwise},
\end{cases} \tag{44}
\]

\[
\frac{d^m}{d s^m} \eta(s) = \begin{cases} 
-\pi \lambda_j \mathbf{T}_{n,j} \left( \sum_{k=1}^{K} \frac{\lambda_k \mathbf{T}_{n,k}}{\lambda_j \mathbf{T}_{n,j}} \delta \mathbf{g}_s(s) + \frac{\lambda_j \mathbf{T}_{n,j}}{\lambda_j \mathbf{T}_{n,j}} \left( \frac{\varphi_{T_j}}{\varphi_{T_j}} \right) \delta \mathbf{g}_s(1 + \delta) \Gamma(1 - \delta) \right), & \text{if } \xi_k = 1, \\
-\pi \lambda_j \mathbf{T}_{n,j} \left( \sum_{k=1}^{K} \frac{\lambda_k \mathbf{T}_{n,k}}{\lambda_j \mathbf{T}_{n,j}} \delta \mathbf{g}_s(s) + \frac{\lambda_j \mathbf{T}_{n,j}}{\lambda_j \mathbf{T}_{n,j}} \left( \frac{\varphi_{T_j}}{\varphi_{T_j}} \right) \delta \mathbf{g}_s(1 + \delta) \Gamma(1 - \delta) \right), & \text{otherwise}. \tag{47}
\end{cases}
\]

derivatives of \(\eta(s)\). As shown in [39], obtaining a closed-form solution for \(\frac{d^m}{d s^m} \eta(s)\) is generally much easier than for \(\frac{d^m}{d s^m} \mathcal{L}_{I_n}(s)\). Specifically, we have [47], as shown at the top of this page, where \(\frac{d^m}{d s^m} = \delta^{m_j} \mathbf{g}_s^{m_j} \frac{d^m}{d s^m} \mathbf{g}_s(s)\) and \(\frac{d^m}{d s^m} \mathbf{f}_k(s)\) are given by [48] and [49], respectively, as shown at the top of the next page. Here, \(\frac{d^m}{d s^m} \varphi(1, \frac{\varphi_{T_k}}{\sqrt{\sigma_0^\alpha}})\) and \(\frac{d^m}{d s^m} \varphi(i+1, \xi_k \frac{\varphi_{T_k}}{\sqrt{\sigma_0^\alpha}})\) can be calculated as [50] and [51], respectively, as shown at the top of the next page. Based on [39] Theorem 2, we have

\[
\frac{d^m}{d s^m} \mathcal{L}_{I_n}(s) \bigg|_{s=\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}} = \left\| \mathbf{M}_j \right\|_1, \tag{52}
\]

where \(\mathbf{Z}_M\) is the following \(M \times M\) lower triangular Toeplitz matrix, i.e.,

\[
\mathbf{Z}_M = \begin{pmatrix} 
z_0(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) & 0 & 0 & 0 \\
z_1(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) & z_0(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
z_{M-1}(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) & \cdots & z_1(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) & z_0(\theta \mathbf{r}_{b_n,\kappa_n,\kappa_0,\kappa_0}) 
\end{pmatrix}. \tag{53}
\]
The lower bound on $p_{u,n}(T_n)$ can be directly obtained by using the Lemma 1 in [40], and thus we omit the details due to page limitation. In the following, we focus on the calculation of the upper bound on $p_{u,n}(T_n)$. Specifically, based on the derivation in [56], we have [54], as shown at the top of the next page, where (a) is due to the fact that $|h^T_{b_{k_o},k_{o}} w_{b_{k_o},k_{o}}|^2 \leq \Gamma(M_{b_{k_o}}+1)$ and $A_{b_{k_o}}$ is the probability that the typical user $k_{o}$ is associated with tier $b_{k_o}$; (b) follows from a lower bound on the incomplete gamma function, i.e., $\Gamma(1+a) > (1-e^{-2x})^a$ for $a > 1$; and (c) follows from the binomial expansion. Substituting [57] and [43] into [54] and using some algebraic manipulations, we obtain the upper bound on $p_{u,n}(T_n)$, as shown in (24). Thus, we complete the proof. 

**APPENDIX C**

**PROOF OF PROPERTY [5]**

When $M_k = M$ and $\phi_k = \phi$, for all $k \in K$, based on (25), we have (55), as shown at the top of the next page, where $V_{m,M}(\xi, \theta_a)$ and $W_m(\theta_a)$ are given by (22) and (23), respectively. First, it is easy to see that $\sum_{m=0}^{M-1} (1 - \frac{m}{M}) W_m(\theta_a) \geq 0$ due to $W_m(\theta_a) \geq 0$. Denote $Y \triangleq \sum_{m=0}^{M-1} (1 - \frac{m}{M}) V_{m,M}(\xi, \theta_a)$. Then, based on the following Lemma 1 we have $Y \geq 0$. Hence, we have $\frac{\partial^k \phi_k(T_n)}{\partial t_k^{M-n}} \geq 0$, which completes the proof.

**Lemma 1 (Sign of $Y$):** We have $Y \geq 0$.

**Proof:** According to (22), we consider the cases of $\xi = 1$ and $\xi \neq 1$, respectively. If $\xi = 1$, we have $Y = \sum_{m=0}^{M-1} (1 - \frac{m}{M}) \frac{\Gamma(M+\delta)}{\Gamma(1+\delta)} \Gamma(1-\delta) \theta^\delta$ and thus $Y \geq 0$ holds clearly. If $\xi \neq 1$, for ease of illustration, we rewrite $Y$ as $Y = Y_1 Y_2$, where $Y_1$ is given by (56), as shown at the top of the next page, and $Y_2 \triangleq \sum_{m=0}^{M-1} (1 - \frac{m}{M}) \frac{\Gamma(M+\delta)}{\Gamma(1+\delta)} \Gamma(1-\delta) \theta^\delta$. We first show $Y_1 \geq 0$. To be specific, we further rewrite $Y_1$ as $Y_1 = \Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi)$, where $y(\xi) \triangleq \Gamma(1+\delta) - \frac{\Gamma(M+\delta)}{\Gamma(1+\delta)} \Gamma(1-\delta) \theta^\delta + \frac{\Gamma(M+\delta)}{\Gamma(1+\delta)} \Gamma(1-\delta) \theta^\delta$. Clearly, if $\xi < 1$, we have $Y_1 \geq 0$ due to $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi) \geq 0$. We next consider $y(\xi)$, i.e., $y(\xi) \geq 0$ due to $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi) \geq 0$. Clearly, if $\xi > 1$ and $M$ is odd, i.e., $M = 1, 3, \ldots$, we have $Y_1 \geq 0$ due to $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi) \geq 0$. We next consider $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi)$, i.e., $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi) \geq 0$. Clearly, if $\xi > 1$ and $M$ is even, i.e., $M = 2, 4, \ldots$, we have $Y_1 \geq 0$ due to $\Gamma(1-\delta) \theta^\delta (1-\xi)^{M-1} \theta^\delta y(\xi) \geq 0$. Specifically, we have
\[ p_{u,n}(T_n) = 1 - \mathbb{E}_{k_n \sim \mathbb{E}} \mathbb{E}_{r_{n,k_n}, \ell_n, \xi_n} \mathbb{E}_{I_n} \left[ \Pr \left[ \left| h_{n,k_n, \ell_n, \xi_n} w_{b_n, r_{n,k_n}, \ell_n} \right| \leq \theta_n r_{b_n, r_{n,k_n}, \ell_n} I_n \right] \right] \]

\[ = 1 - \sum_{k_{n} = 1}^{K} A_{k_n} \mathbb{E}_{r_{n,k_n}, \ell_n, \xi_n} \mathbb{E}_{I_n} \left[ \gamma \left( \frac{M_{k_n}}{M} \right) \frac{\theta_n r_{b_n, r_{n,k_n}, \ell_n} I_n}{\gamma(\xi_n)} \right] \]

\[ \leq 1 - \sum_{k_{n} = 1}^{K} A_{k_n} \mathbb{E}_{r_{n,k_n}, \ell_n, \xi_n} \mathbb{E}_{I_n} \left[ \left( 1 - e^{-S M_k \theta_n r_{b_n, r_{n,k_n}, \ell_n} I_n} \right)^{M_k} \right] \]

\[ = 1 - \sum_{k_{n} = 1}^{K} \sum_{m_{n} = 1}^{M_k} M(m) \left( 1 - \right)^{m} A_{k_n} \mathbb{E}_{r_{n,k_n}, \ell_n, \xi_n, \theta_n} \left[ \int_{0}^{\infty} f_{r_{n,k_n}, \ell_n, \xi_n} (\eta) \mathcal{L}_{I_n} \left( m S M_k \theta_n r_{b_n, r_{n,k_n}, \ell_n} \right) d\eta \right] \]

\[ \dot{\mathbb{P}}_{u,n}(I_n) = \frac{\lambda_k (\phi_k)^{\delta}}{\left( f_m(T_n, \theta_k) \right)^{\delta}} \sum_{m=0}^{M-1} \left( 1 - \frac{m}{M} \right)^{\delta} \frac{V_m \delta M_k \theta_n r_{b_n, r_{n,k_n}, \ell_n} + W_m (m S M_k \theta_n)}{\delta M_k \theta_n r_{b_n, r_{n,k_n}, \ell_n} + W_m (m S M_k \theta_n)} \]

\[ \frac{\partial^2 \mathbb{P}_{u,n}^{ij}(I_n)}{\partial T_n^{2}} = -2 \left( \sum_{k \in \mathbb{K}} \lambda_k (\phi_k)^{\delta} \delta M_k \theta_n r_{b_n, r_{n,k_n}, \ell_n} + W_m (m S M_k \theta_n) \right) \left( \dot{U}_m - \dot{V}_m \right) \]

\[ Y_2 = \sum_{m=0}^{M-1} \left( 1 - \right)^{M-1} \delta > 0. \]
have $\mathcal{L}_I^n(s) = \mathcal{L}_I^{n*}(s)\mathcal{L}_I^n(s)\prod_{k \in K} \mathcal{L}_I^{n,k}(s)\mathcal{L}_I^{n,-k}(s)$. By following the similar steps as in the derivation of $\mathcal{L}_I^n(s)$ in Appendix A, we have $\mathcal{L}_I^{n,k}(s) = (1 + \xi_k s^{-\alpha})^{1-M_k}$, $\mathcal{L}_I^{n,k}(s) = \mathcal{L}_I^n(s)$ and 
\[
\prod_{k \in K} \mathcal{L}_I^{n,k}(s)\mathcal{L}_I^{n,-k}(s) = \frac{\exp\left(-\sum_{k=1}^{K} \lambda_k \phi_P \phi_T \Gamma\left(1-M_k\right)\Gamma\left(1-\delta\right)s\right)}{W_0(\delta) + \sum_{k=1}^{K} \lambda_k \phi_P \phi_T \Gamma\left(1-M_k\right) \Gamma(1, \delta) s},
\]
if $\xi_k = 1$, and 
\[
\frac{\exp\left(-\sum_{k=1}^{K} \lambda_k \phi_P \phi_T \Gamma\left(1-M_k\right)\Gamma(1, \delta) s\right)}{W_0(\delta) + \sum_{k=1}^{K} \lambda_k \phi_P \phi_T \Gamma\left(1-M_k\right) \Gamma(1, \delta) s},
\]
otherwise.

where $A(s)$ is given by (69). Then, we have 
\[
\mathcal{L}_I^n(\theta s^{\alpha}) = \frac{\phi_P \phi_T \delta^{\frac{\delta}{\theta}} }{W_0(\delta) + \sum_{k=1}^{K} \lambda_k \phi_P \phi_T \Gamma\left(1-M_k\right) \Gamma(1, \delta) s},
\]
where $W_0(\delta)$ and $V_{0,M}(\xi, \theta)$ are given by Theorem 1. Substituting (69) into (69), we complete the proof of Theorem 2.