

Analysis of Time-Varying Collusion Attacks in Fingerprinting Systems: Capacity and Throughput

Byung-Ho Cha and C.-C. Jay Kuo

Ming Hsieh Department of Electrical Engineering and Signal and Image Processing Institute

University of Southern California, Los Angeles, CA 90089-2564

E-mails: byungcha@usc.edu and cckuo@sipi.usc.edu

Abstract—We analyze time-varying collusion attacks for a fingerprinting system using concepts of capacity and throughput in this work. Continuous media provide a limited resource for data embedding, which is analogous to the capacity of a wireless communication channel. Here, we characterize the capacity of a host media using the just-noticeable-distortion (JND) of the human visual system (HVS). Furthermore, the collusion attack can be interpreted as a channel shared by multiple users. Based on this analogy, the colluder detection performance can be understood from the viewpoint of throughput. Specifically, we show how to determine instantaneous throughput using the fingerprint-to-interference-plus-noise ratio (FINR), and extend it to the total throughput and the averaged throughput over a time interval. Our analysis provides a good framework to the understanding of collusion attacks and ways to enhance the traitor tracing performance of a fingerprinting system.

I. INTRODUCTION

The fingerprinting technology provides a solution to traitor tracing in a video multi-cast environment, where different imperceptible watermarks are embedded in a media file and distributed to different buyers. Thus, each distributed copy carries the fingerprint of a specific buyer. One effective attack to fingerprinting is the collusion attack, where multiple attackers (or colluders) perform a linear combination of their copies to result in another copy with an objective to confuse the detector so that their individual fingerprints cannot be detected properly.

The time-varying collusion provides an effective means to the attack of traditional multimedia fingerprinting systems [1], [2]. Recently, this attack was examined and a solution called the multi-carrier code-division-multiple-access (MC-CDMA) fingerprinting system was proposed in [3], [4], where the time-varying collusion weights were estimated using a channel estimation technique. As a sequel to [4], we analyze time-varying collusion attacks from the viewpoint of capacity and throughput. The analysis shed light on collusion attacks, which is useful to the design of a better fingerprinting system.

Continuous media provide a limited resource for data embedding, which is analogous to the capacity of a wireless communication channel. Here, we characterize the capacity of a host media using the just-noticeable-distortion (JND) of the human visual system (HVS). Furthermore, the collusion attack can be interpreted as a channel shared by multiple users. Based on this analogy, the colluder detection performance can be understood from the throughput viewpoint. Furthermore, we show how to determine instantaneous throughput using the fingerprint-to-interference-plus-noise ratio (FINR), and extend it to the total throughput and the averaged throughput over a time interval. Our analysis provides a good framework to the understanding of collusion attacks and ways to enhance the traitor tracing performance of a fingerprinting system.

The rest of this paper is organized as follows. The MC-CDMA-based fingerprinting system is reviewed in Sec. II. Next, we analyze a fingerprinting system from the perspective of capacity and throughput

in Sec. III. It is demonstrated by computer simulation in Sec. IV that the proposed scheme can provide a good fingerprinting system for movie content protection. Finally, concluding remarks and future research work are given in Sec. V.

II. MC-CDMA-BASED FINGERPRINTING SYSTEM

In this section, we review the MC-CDMA-based fingerprinting system proposed in [3]. As shown in Fig. 1, it consists of three modules: 1) the fingerprint generation and embedding module, 2) the time-varying collusion attack module, and 3) the fingerprint detection module. They are detailed below.

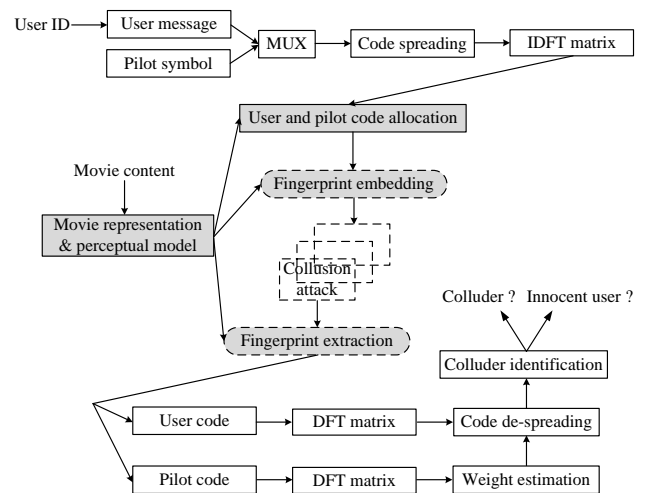


Fig. 1. The block-diagram of the overall system with three modules: 1) fingerprint generation and embedding, 2) time-varying collusion attack, and 3) fingerprint detection.

A. Fingerprint Generation and Embedding

Let Φ be the user set and $|\Phi| = L$ the user number. The user message, m_l , of length M contains user identification (ID) u_l of length U and error correction codes of length $M - U$. Furthermore, $s_l(i)$ denotes the spreading code. Then, the user code, $w_l(i)$, is obtained via

$$w_l(i) = \text{IDFT} \{m_l s_l(i)\} \quad (1)$$

where IDFT is the inverse discrete Fourier transform. The number of users is decided by units of spreading codes. If we use $N \times N$ Hadamard-Walsh (HW) codes or $N \times N$ carrier interferometry (CI) codes, they can support N users [3].

In code embedding, let $x_l(i)$, $i = 0, \dots, T-1$, be selected discrete cosine transform (DCT) coefficients for user l according to the human

visual system (HVS). We divide this set into B segments, each of which has N samples, as

$$x_l(i) = x_l(b \cdot N + n), \quad \begin{cases} b = 0 \cdots B-1 \\ n = 0 \cdots N-1 \end{cases} \quad (2)$$

The additive code embedding method with shifted spreading [5] is given mathematically by

$$y_l(i - \Delta_l) = x_l(i - \Delta_l) + a_l(i - \Delta_l) \quad (3)$$

where

$$a_l(i - \Delta_l) = \alpha(i)w_l(i), \quad (4)$$

and where Δ_l is a shift amount under condition $P \ll N$ and $\alpha(i)$ is the embedded code strength, which is adaptively decided by the human perceptual model given in Sec. III-A. If we assume $|w_l(i)| = 1$, $\alpha(i)$ equals to P_F . By shifted spreading, we are able to increase the user number from N to $P \times N$.

B. Time-Varying Collusion Attack

We divide users into two groups: malicious users (or colluders) and innocent users, and use Ω to denote the set of colluders. Clearly, Ω is a subset of Φ . Without loss of generality, we assume that there are L users and K colluders in the system. That is, $|\Phi| = L$ and $|\Omega| = K$. A time-varying collusion attack can be expressed as

$$\hat{y}(i) = \sum_{k \in \Omega} h_k(i)y_k(i) + e(i) \quad (5)$$

where $y_k(i)$ is the host signal, $h_k(i)$ the time-varying weight for colluder k , $e(i)$ additive noise and $\hat{y}(i)$ the colluded signal on the i th sample. The weights should satisfy the following condition:

$$\sum_{k \in \Omega} h_k(i) = 1 \quad (6)$$

where $h_k(i) \neq 0$ for all i . Furthermore, colluders can change their colluder weights arbitrarily without the knowledge of embedding and detection algorithms. We express the value of $h_k(i)$ as

$$h_k(i) = \tilde{h}_k(r; q), \quad \begin{cases} r = 0, \dots, R(q) - 1, \\ q = 1, \dots, Q \end{cases} \quad (7)$$

where $R(q)$ represents the number of samples in segment q , which varies in each segment, and Q represents the number of segments in one media.

C. Fingerprint Extraction and Colluder Identification

We extract fingerprint codes from the host media via

$$\hat{y}(i) - x(i) = \sum_{l \in \Phi} h_l(i)\hat{w}_l(i) + (\alpha(i))^{-1}e(i). \quad (8)$$

After the extraction, the discrete Fourier transform (DFT) is taken before user detection:

$$\sum_{l \in \Phi} m_l \lambda_l(i) \hat{s}_l(i) = \text{DFT} \left\{ \sum_{l \in \Phi} h_l(i) \hat{w}_l(i) + \tilde{e}(i) \right\} \quad (9)$$

where $\tilde{e}(i) = (\alpha(i))^{-1}e(i)$. Then, user detection, which includes synchronization and group/subgroup ID identification with threshold τ , is performed by advanced detectors. When L users are supported, L user detection steps must be performed. Let \hat{u}_l be the user ID decoded from detected message \hat{m}_l . Colluders can be distinguished from innocent users by a colluder identification process based on the following principle:

- $\Pr[\hat{u}_k \neq u_k] \rightarrow 0$ for colluder set Ω ; and
- $\Pr[\hat{u}_j = u_j] \rightarrow 0$ for innocent user set Γ .

Specific decision rules can be derived accordingly. Finally, to improve the performance of colluder detection, a maximum ratio combining (MRC) and a multiuser detector (MUD) technique based on pilot-aided colluder weight estimation (PACWE) can be used. We refer to [6], [7] for more detail.

III. ANALYSIS OF TIME-VARYING COLLUSION ATTACKS: CAPACITY AND THROUGHPUT

It is important to determine the proper power level of the fingerprint at different spatial/temporal locations. If the power of a fingerprint is higher, it is easier to detect. On the other hand, it will bring more noticeable degradation to the host media. Furthermore, in the context of collusion attack, we need to consider the quality of colluded media as well as the detectability of the fingerprint of any colluder. These issues will be discussed in this section.

A. Embedding Rate of Host Media and HVS Model

Following the discussion in Sec. II-A, we consider fingerprint embedding in DCT coefficients of each video frame. Let these DCT coefficients be

$$F(u, v), \quad 0 \leq u \leq \tilde{N} - 1, \quad 0 \leq v \leq \tilde{N} - 1$$

where a typical value of \tilde{N} is 8. The selected coefficient $F(u, v)$ based on the HVS model is mapped to the host sample $x(i)$ for fingerprint embedding.

For the HVS model, we consider the just noticeable difference (JND) introduced in [8]. There are two masking effects in JND; namely, the luminance mask and the contrast mask. The luminance mask is given by

$$z_b^L = z^F \left(\frac{F_b(0, 0)}{F(0, 0)} \right)^\beta \quad (10)$$

where b is the index of a DCT block, $F_b(0, 0)$ is the DC coefficient of the b th DCT block, $F(0, 0)$ is the averaged DC component of the entire frame, β is a constant (set to 0.649 empirically) and z^F is a frequency mask governed by the sensitivity of DCT coefficients. The contrast mask, z_b^C , is calculated from the luminance mask via

$$z_b^C = \max \left\{ z_b^L, |F_b(u, v)|^{\gamma(u, v)} (z_b^L)^{(1-\gamma(u, v))} \right\} \quad (11)$$

where $\gamma(u, v)$ is a value between 0 and 1 depending on the DCT basis function. It is set to 0.7 empirically. For details, we refer to [8]. The JND power P_J and the total number of selected samples B_J can be decided by z_b^C as introduced in [9]. The JND power, P_J , can be related to the embedding rate \mathbb{R}_E (in the unit of bits per sample). For a give host media, we can use the product of \mathbb{R}_E and B_J to determine the total number of embedding bits to be supported by the fingerprinting system. B_J can be explained by number of selected samples which are $P_J \neq 0$.

B. Power Relationship in Collusion Attacks

In the context of collusion attack, we use $P_J(n)$, $P_{F,k}(n)$ and $P_{IN,k}(n)$ to denote the power of corresponding to the maximum distortion allowed in multimedia by JND, the embedded fingerprint power and the interference-plus-noise power for an arbitrary colluder denoted by k at a certain selected DCT coefficient position denoted by n . They have to meet the following three conditions.

- 1) For fingerprint embedding with imperceptibility, we demand

$$P_{F,k}(n) \leq P_J(n). \quad (12)$$

- 2) For imperceptibility of a colluded media copy, we demand

$$P_{F,k}(n) + P_{IN,k}(n) \leq P_J(n). \quad (13)$$

3) For fingerprint detectability, we demand

$$P_{F,k}(n) > P_{IN,k}(n). \quad (14)$$

If condition 1) is violated, we can observe visual distortion from a fingerprinted media file. If condition 2) is violated, there exists visible distortion in the colluded media file. If condition 3) is violated, it will be difficult to detect colluder k at the selected DCT sample.

C. Capacity and Throughput

Capacity and throughput are two different concepts in wireless communications. Capacity is the property of a channel while throughput is affected by the channel as well as the interaction between users. In the context of fingerprinting, if there is no collusion, the analogy is that the messages of each user go through an independent channel. When collusion occurs, the analogy is that multiple users send their messages through a shared channel. When the system load is low, throughput is proportional to the colluder number. However, when the system load is high, throughput becomes lowered by interaction between colluders. The relationship between capacity and throughput in a single-user channel and a multi-user shared channel is shown in Fig. 2. As illustrated in Fig. 2 (a), throughput is saturated when it is reached by capacity in a single-user case, and as shown in Fig. 2 (b), throughput is lowered by collusion in a multi-user case.

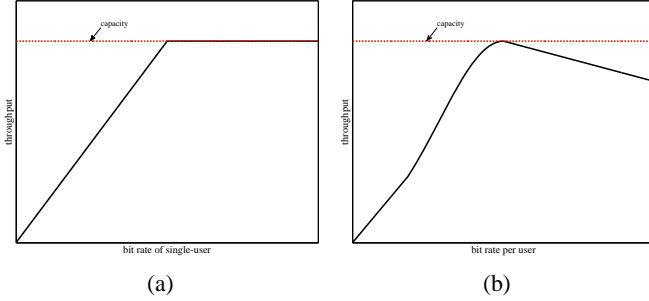


Fig. 2. The relationship between capacity and throughput in (a) a single-user channel and (b) a multi-user shared channel.

D. Throughput Analysis

We may define the following two concepts related to throughput:

- Instantaneous throughput $\mathbb{T}_{ins}(t)$
It is determined by the fingerprint power and the strength of collusion attacks at a given time interval centered around t .
- Total throughput \mathbb{T}_{tot}
It is the summation of all instantaneous throughput over the entire continuous host media.

Lower throughput means that we receive fewer messages of colluder k and, as a result, it will be more difficult to perform accurate detection. This relationship can be analyzed below.

Let us consider a time interval: $(t_1, t_2]$. On one hand, the capacity of the host media in this interval is equal to $\mathbb{R}_E \times (t_2 - t_1)$, which is governed by JND. On the other hand, the colluder throughput \mathbb{T}_{tot} in the same interval can be written as

$$\mathbb{T}_{tot}(t_1, t_2) = \int_{t_1}^{t_2} \mathbb{T}_{ins}(t) dt. \quad (15)$$

The averaged throughput is equal to the total throughput divided by the number of colluders, K . It can be written mathematically as

$$\mathbb{T}_{ave}(t_1, t_2) = \frac{\mathbb{T}_{tot}(t_1, t_2)}{K}. \quad (16)$$

The average detection probability of a colluder is related to the average throughput. The higher the average throughput, the higher the average detection probability. The average detection probability is closely related with the number of identified colluders [3]. The characterization between the average throughput and the average detection probability also depends on the detector design. This is however out of the scope of our current work and will be treated in our future work.

In the following, we only consider the effect of the fingerprint-to-interference-plus-noise ratio (FINR). The colluder detection performance of the MC-CDMA-based fingerprinting system is closely related with FINR. The inter-group interference (IGI) occurs among colluders from groups with different codewords [10]. If the fingerprint power of colluder k is denoted by $P_{F,k}$, the FINR of colluder k can be written as

$$\zeta_{FINR,k} = \frac{\sum_{n=0}^{N-1} |\lambda_k(n)|^2 p_{F,k}(n)}{\sum_{l \neq k} \sigma_{IGI_{l \rightarrow k}}^2 p_{F,l}(n) + N\sigma^2} \quad (17)$$

where σ^2 is the variance of the Gaussian noise, N is the spreading gain, $\lambda_k(n)$ is the frequency response of colluder weight and

$$\sigma_{IGI_{l \rightarrow k}}^2 = E\{\text{Re}^2\{IGI_{l \rightarrow k}\}\}.$$

When the spreading code length N is larger, $\sum_{n=0}^{N-1} |\lambda_k(n)|^2 p_{F,k}(n)$ is larger. If we can cancel out the term $\sigma_{IGI_{l \rightarrow k}}^2$ by MUD, the FINR is converted into the fingerprint-to-noise ratio (FNR).

The classic information-theoretic capacity region for colluders with white Gaussian noise $e(i)$ in Gaussian multiple-access channel (GMAC) can be written as

$$\sum_{k \in \Omega} \mathbb{R}_k < \frac{1}{2} \log(1 + \zeta_{FNR}), \quad (18)$$

which is in the unit of bits per message symbol [11]. Here, ζ_{FNR} is given by

$$\zeta_{FNR} = \frac{\sum_{k \in \Omega} \sum_{n=0}^{N-1} |\lambda_k(n)|^2 p_{F,k}(n)}{N\sigma^2} \quad (19)$$

Eq. (18) represents the maximum sum rate which can be achieved by the total power of colluders in Ω without interactions among colluders. It can be used to represent the rough upper bound of instantaneous throughput $\mathbb{T}_{ins}(t)$. For more details, we refer to [6], [12].

IV. EXPERIMENTAL RESULTS

We study the performance of the proposed MC-CDMA-based fingerprinting system applied to a real movie in this section. The length N of spreading codes is chosen to be 256. We apply shift spreading with $P = 2$ (*i.e.*, 1-bit shift) in all examples. Thus, the collusion attack is formulated as the frequency selective fading channel in a wireless communication. Colluder weights in the collusion attacks are randomly generated using a normal Gaussian distribution with zero mean and unit variance, then normalized by a total sum. K Colluders are selected from user set of size L with combinations C_K^L . Experimental results are obtained using 10^4 simulation runs.

We first evaluate the average throughput which is introduced in Sec. III-D with correlation detection and MRC with PIC-MUD. The correlation detection is widely used in the watermarking and fingerprinting field. In Fig. 3 (a), The average throughput is increased when the bit rate of colluders is increased at the fingerprint embedder. The average throughput can be improved by MRC with PIC-MUD.

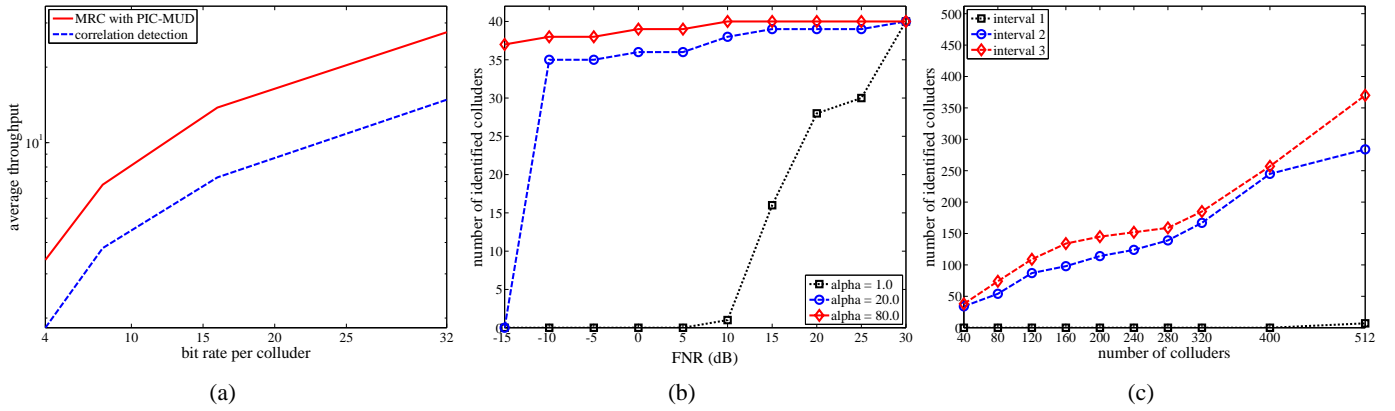


Fig. 3. Simulation results with 1-bit shift and $L = 512$ (a) the average throughput as a function of bit rate per colluder with conventional correlation detection and MRC with PIC-MUD ($K = 512$) (b) The number of identified colluders as a function of the fingerprint-to-noise ratio (FNR) depending on fingerprint power ($K = 40$), and (c) The number of identified colluders as a function of number of colluders depending on the host media intensity (FNR = 0 dB).

This result is constant with our previous result in [3]. Then, we investigate the relationship between the noise and the number of identified colluders in a fixed interference environment. The length of user message is set to $M = 32$. In Fig. 3 (b), we fix the number of colluders $K = 40$ and adjust the fingerprint-to-noise ratio (FNR) range from -15 dB to 30 dB. The embedding strength $\alpha(i)$ is set to 1.0, 20.0 and 80.0, which governs the fingerprint power. When the power of fingerprint codes increases, we see that the number of identified colluders increases in the lower FNR case (*i.e.*, high noise power). Finally, we analyze the effect of the host media intensity (*i.e.*, dynamic range), which is equal to 10 in interval 1, 128 in interval 2, and 256 in interval 3, with FNR equal to 0 dB. The length of message is $M = 32$. In Fig. 3 (c), we see that the number of identified colluders is much higher when the maximum intensity range of the host media is higher.



Fig. 4. Embedding imperceptibility of the *Toystory* movie based on JND: (a) one original frame and (b) the corresponding fingerprinted frame with PSNR = 62.99 dB.

We embed the fingerprint to movie *Toystory* (of resolution 720×480 and frame rate 30 frames per second) in the DCT domain to test the HVS model, where DCT is applied to every 8×8 block unit. Fig. 4 shows one frame of *Toystory* with the luminance component only. The total number of DCT coefficients are $N_{tot} = 345,600$, and we embed fingerprints to a subset of these samples, whose total length is $N_f = 14,506$. Thus, the averaged embedding rate 4% per sample. The quality of any fingerprinted frame is well preserved after embedding, and the peak signal-to-noise ratio (PSNR) is higher than 60 dB.

V. CONCLUSION AND FUTURE WORK

In this work, we related the JND of the HVS model to the capacity of a host media and the colluder detection performance

in an MC-CDMA-based fingerprinting system and also connected it to the throughput analysis in a wireless communication system. Furthermore, we showed how to determine instantaneous throughput using FINR, and extended it to the total throughput and the averaged throughput over a period of time. Finally, it was demonstrated by computer simulation that the proposed scheme can provide a systematic approach for movie content protection. In the future, we will examine the theoretical analysis for throughput and the problem of optimal detector design based on the average detection probability (or throughput).

REFERENCES

- [1] Z. J. Wang, M. Wu, H. V. Zhao, and K. J. R. Liu, "Anti-collusion forensics of multimedia fingerprinting using orthogonal modulation," *IEEE Transactions on Image Processing*, vol. 14, pp. 804–821, June 2005.
- [2] S. He. and M. Wu, "Joint coding and embedding techniques for multimedia fingerprinting," *IEEE Transactions on Information Forensics and Security*, vol. 1, pp. 231–247, June 2006.
- [3] B.-H. Cha and C.-C. Jay Kuo, "Advanced colluder detection techniques for OSIFT-based hiding codes," in *Proc. IEEE Int'l Sym. Circuits and Systems*, Seattle, Washington, May 2008, pp. 2961–2964.
- [4] B.-H. Cha and C.-C. Jay Kuo, "Robust MC-CDMA-based fingerprinting against time-varying collusion attacks," *IEEE Transactions on Information Forensics and Security*, 2008, submitted.
- [5] B.-H. Cha and C.-C. Jay Kuo, "Design of multiuser collusion-free hiding codes with delayed embedding," in *Proc. IEEE Int'l Conf. Intelligent Information Hiding and Multimedia Signal Processing*, Kaohsiung, Taiwan, November 2007, pp. 379–382.
- [6] S. Verdu, *Multiuser detection*, Cambridge University Press, Cambridge, UK, 1998.
- [7] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*, John Wiley & Sons, West Sussex, UK, 2004.
- [8] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE, Conf. Human Vision, Visual Processing, and Digital Display*, San Jose, CA, USA, February 1993, pp. 202–216.
- [9] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 525–539, May 1998.
- [10] B.-H. Cha and C.-C. Jay Kuo, "Design and analysis of high-capacity anti-collusion hiding codes," *Journal of Circuits, Systems, and Signal Processing*, March 2008.
- [11] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, pp. 641–657, March 1999.
- [12] D. N. C. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, Cambridge, UK, 2005.