Audio Watermarking using DWT-DTMT-MLNCML: Implementation and Analysis

Meghadri Ghosh - 2023A3PS0314P
Pramit Pal - 2023AAPS0765P
Pranav Chandra N. V. - 2023AAPS0013P

November 15, 2025

1 Approach

We combined three transforms: MLNCML encrypts the watermark $W \in \{0,1\}^{32\times 32}$ using coupled logistic maps to produce $W' = W \oplus H_b$ where H_b is generated with $\epsilon = 0.3, \mu = 3.99$. The audio is segmented into $L_{\text{seg}} = \lceil (N \cdot M)/L_1 \rceil = 256$ blocks. Each segment undergoes 3-level Haar DWT:

$$\text{DWT}^3(\text{segment}) \rightarrow (A_3, D_3, D_2, D_1) \tag{1}$$

where A_3 contains low-frequency approximation coefficients. We divide A_3 into $L_1 = 4$ sub-blocks and apply DTMT with $K = \min(64, L_2/2)$ Tchebichef polynomial orders:

$$M = T \cdot \text{pa3}, \quad T \in \mathbb{R}^{K\times L_2} \tag{2}$$

Moments split into even/odd indices: $M_1 = M[0 :: 2], M_2 = M[1 :: 2]$. We embed bit b by modifying norms:

$$(\sigma_1', \sigma_2') = \begin{cases} (\bar{\sigma}+\delta, \bar{\sigma}-\delta) & b = 1 \\ (\bar{\sigma}-\delta, \bar{\sigma}+\delta) & b = 0 \end{cases}, \quad \bar{\sigma} = \frac{\|M_1\| + \|M_2\|}{2}, \quad \delta = 0.05 \tag{3}$$

Reconstruction uses inverse DTMT ($\text{pa3}' = T^T M'$) and inverse DWT. Extraction compares $\|M_1\|$ vs $\|M_2\|$ to recover bits, then decrypts: $\hat{W} = \hat{W}' \oplus H_b$. See Figure 1.

2 Results

Table 1 shows robustness against four attacks: LPF (4 kHz) should succeed since we embed in low frequencies, HPF (300 Hz) attacks our weak point, cropping (20%) tests localized damage, and Gaussian noise (20 dB) tests statistical perturbations.

Table 1: Attack robustness metrics: SNR (audio quality), BER (bit errors), NC (pattern correlation)

Attack	**SNR (dB)**	**BER**	**NC**
LPF 4000 Hz	—	—	—
HPF 300 Hz	—	—	—
CROP 20%	—	—	—
Gaussian 20 dB	—	—	—

3 Key Learnings

Frequency band selection. Embedding in detail bands (D_1, D_2, D_3) created tinny audio—high frequencies are perceptually obvious. Low-frequency A_3 works better but remains vulnerable to HPF.

Reconstruction isn't perfect. Haar DWT padding and DTMT truncation $(K = \min(64, L_2/2))$ introduce errors. Hard segmentation creates discontinuities like cutting audio without cross-fading.

Metrics vs perception. 40 dB SNR seems imperceptible but ears detect frequency distortions. Psychoacoustic masking is essential for true imperceptibility.

Numerical precision matters. MLNCML's 2^{100} keyspace provides security, but floating-point rounding breaks decryption. Chaotic systems demand exact parameters.

Moments provide robustness. Norm ratios survive attacks because perturbations affect M_1 and M_2 similarly. Direct coefficient modification fails.

Parameter tuning. $\delta = 0.05$ balances robustness and artifacts through trial and error. Future work needs adaptive embedding, error correction, windowing, and better wavelets.

Appendix: System Architecture

Figure 1 shows how the three transforms work together. The watermark flows through encryption, gets split across audio segments, and embeds into frequency-domain moments. Extraction reverses the process.

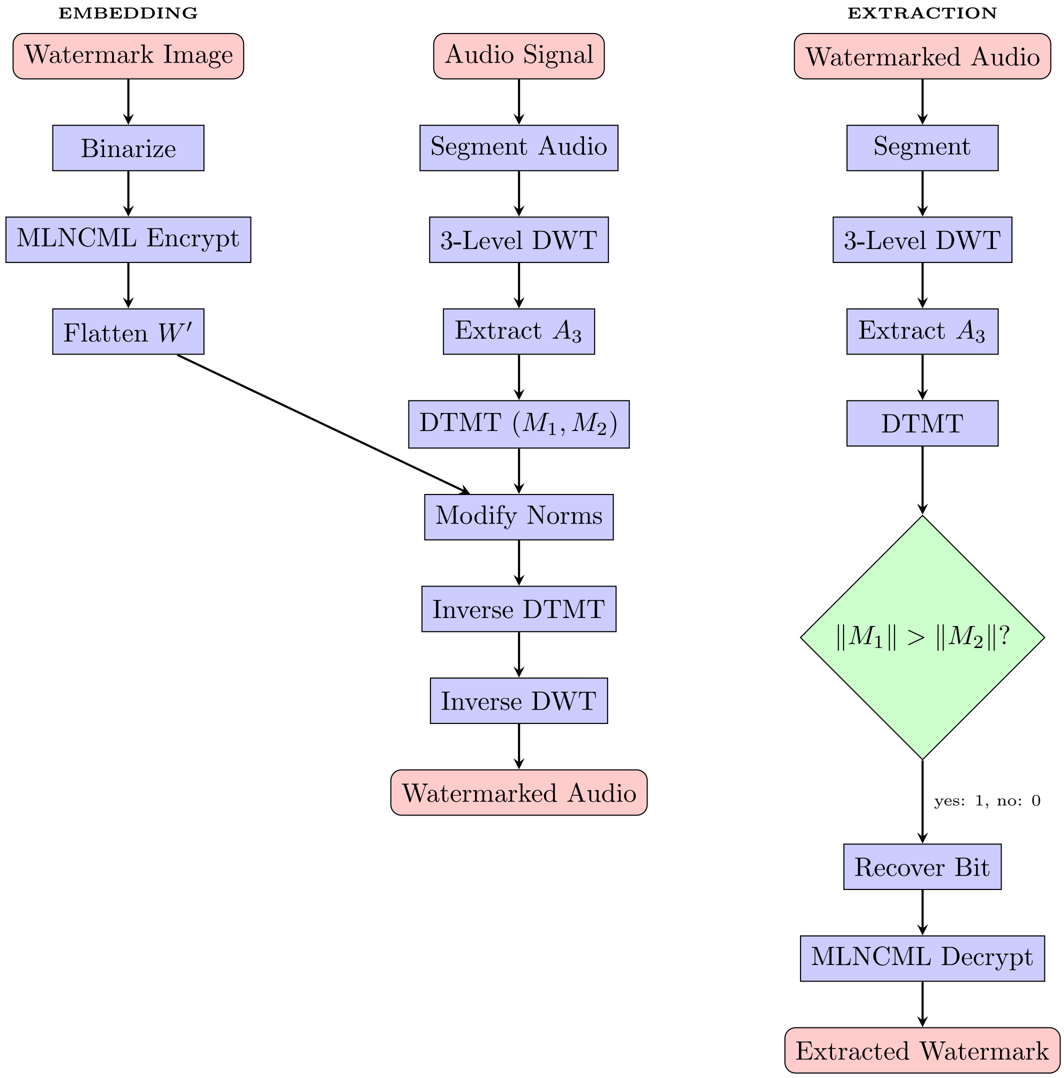

Figure 1: Watermarking system flowchart showing embedding (left/center) and extraction (right) pipelines