RHFCN: FULLY CNN-BASED STEGANALYSIS OF MP3 WITH RICH HIGH-PASS FILTERING

Yuntao Wang[1,2], Xiaowei Yi[1,2], Xianfeng Zhao[1,2], Ante Su[1,2]

1. State Key Laboratory of Information Security, Institute of Information Engineering, CAS
2. School of Cyber Security, University of Chinese Academy of Sciences

中国科学院信息工程研究所 INSTITUTE OF INFORMATION ENGINEERING,CAS

中国科学院大学 University of Chinese Academy of Sciences

Github: https://github.com/Charleswyt/tf_audio_steganalysis

Introduction

1. **MP3** is the most commonly-seeing compressed audio format on the Internet.
2. Various MP3 **steganographic algorithms** with large capacity and good imperceptivity have been proposed.
3. The performance of existing MP3 **steganalytic algorithms** are needed further improving.

Rich High-Pass Filtering Module

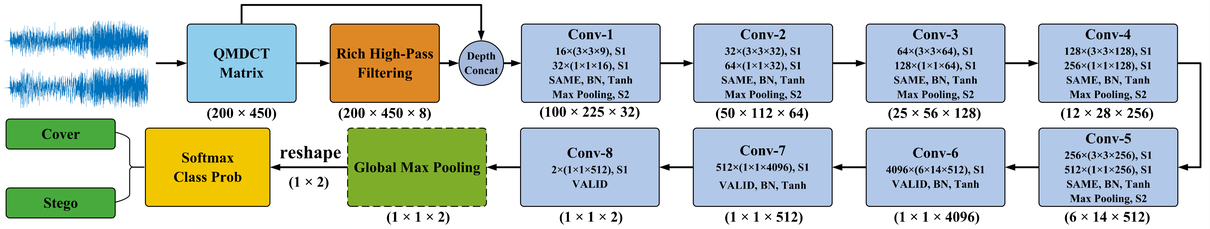
$$M_Q = \begin{bmatrix} Q_{1,1} & \cdots & Q_{1,j} & \cdots & Q_{1,450} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Q_{i,1} & \cdots & Q_{i,j} & \cdots & Q_{i,450} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Q_{200,1} & \cdots & Q_{200,j} & \cdots & Q_{200,450} \end{bmatrix}$$

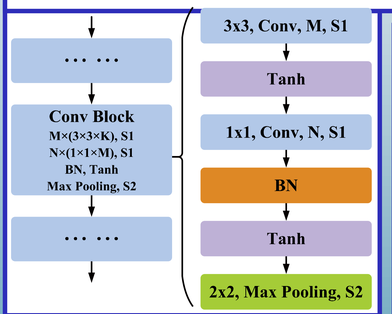
$$M^{\rightarrow}_{m,n} = Q_{i,j} - Q_{i,j+1}$$
$$M^{\downarrow}_{m,n} = Q_{i,j} - Q_{i+1,j}$$
$$A^{\rightarrow}_{m,n} = |Q_{i,j}| - |Q_{i,j+1}|$$
$$A^{\downarrow}_{m,n} = |Q_{i,j}| - |Q_{i+1,j}|$$
$$M^{\rightrightarrows}_{m,n} = Q_{i,j} - 2 \times Q_{i,j+1} + Q_{i,j+2}$$
$$M^{\downdownarrows}_{m,n} = Q_{i,j} - 2 \times Q_{i+1,j} + Q_{i+2,j}$$
$$A^{\rightrightarrows}_{m,n} = |Q_{i,j}| - 2 \times |Q_{i,j+1}| + |Q_{i,j+2}|$$
$$A^{\downdownarrows}_{m,n} = |Q_{i,j}| - 2 \times |Q_{i+1,j}| + |Q_{i+2,j}|$$

Table 1 Percentages of **modified QMDCT coefficients** via each HPF (128 kbps, $W = 4$)

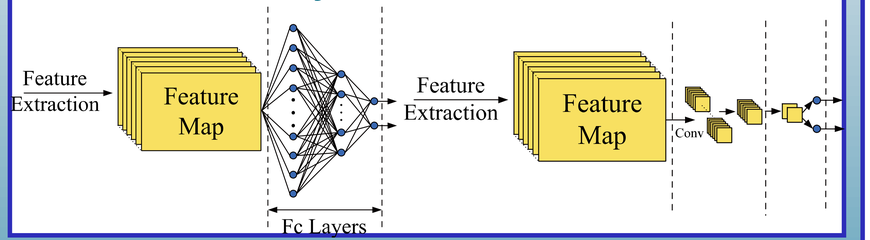
M_Q	$M^{\rightarrow}$	$M^{\downarrow}$	$M^{\rightrightarrows}$	$M^{\downdownarrows}$	$A^{\rightarrow}$	$A^{\downarrow}$	$A^{\rightrightarrows}$	$A^{\downdownarrows}$
1.45	2.00	2.82	2.87	4.14	2.17	2.81	2.88	4.14

Network Structure





Removal of Fc Layers



Network Variants

Table 2 Description and detection accuracy of each network variant (128 kbps, $W = 4$)

ID	Description of the network	Accuracy (%)
a	**RHFCN**	**80.44**
b	Remove rich HPF module	78.13
c	Quit Removing Fc layers	79.09
d	Remove rich HPF module and quit removing Fc layers	77.36

Conclusion

1. The **rich HPF module** "enlarges" the traces of the signal introduced by secret messages, so that the network is more sensitive to the existence of stego signal.
2. The design of **fully CNN structure** does not only improve the performance of the network due to the utilization of spatial and structural correlation of feature maps, but also contributes to the steganalysis of MP3 with **size mismatch**.

Steganalysis with Size Mismatch

Table 3 Detection accuracy (%) of MP3 steganalysis with size mismatch (128 kbps, $W = 4$)

Size	**200×450**	230×450	200×480	230×480
Accuracy	**80.44**	78.22	77.07	75.53

Experimental Results

Table 4 Detection accuracy (%) of **HCM** algorithm

Bitrate	*RER*	**RHFCN**	WASDN	MDI2	ADOTP
128	0.1	**87.18**	83.71	58.48	56.84
	0.3	**92.77**	88.05	68.11	65.13
	0.5	**95.18**	93.34	80.35	74.95
320	0.1	**98.84**	93.26	82.45	68.21
	0.3	**99.23**	94.99	88.44	80.44
	0.5	**99.51**	98.27	93.55	88.34

Table 5 Detection accuracy (%) of **EECS** algorithm

Bitrate	*W*	**RHFCN**	WASDN	MDI2	ADOTP
128	2	**93.26**	90.08	68.79	68.30
	3	**87.96**	82.17	60.79	60.30
	4	**80.44**	74.37	57.71	56.74
	5	**74.76**	64.55	54.72	54.34
	6	**68.50**	55.97	52.02	51.54
320	2	**98.46**	95.57	76.59	73.41
	3	**95.57**	90.17	66.86	61.66
	4	**88.63**	80.15	61.75	57.03
	5	**83.23**	72.54	58.86	54.82
	6	**78.71**	66.67	54.24	53.28