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

Yuntao Wang¹,², Xiaowei Yi¹,², Xianfeng Zhao¹,², Ante Su¹

1. State Key Laboratory of Information Security, Institute of Information Engineering, CAS
2. School of Cyber Security, 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.

**Network Structure**

Rich High-Pass Filtering Module

\[
M_q = \begin{bmatrix}
Q_{1,1} & Q_{1,2} & Q_{1,450} \\
Q_{1,2} & Q_{1,3} & Q_{1,450} \\
Q_{200,1} & Q_{200,2} & Q_{200,450}
\end{bmatrix}
\]

\[
M^*_{min} = Q_{ij} - Q_{i+1,j+1} \\
M^*_{max} = Q_{ij} - Q_{i-1,j-1} \\
A^*_{min} = \lfloor Q_{ij} \rfloor - \lfloor Q_{i+1,j+1} \rfloor \\
A^*_{max} = \lfloor Q_{ij} \rfloor - \lfloor Q_{i-1,j-1} \rfloor \\
A^*_{min} = 2 \times Q_{ij} + Q_{i+1,j+1} + Q_{i-1,j-1} \\
A^*_{max} = 2 \times Q_{ij} + Q_{i+1,j+1} + Q_{i-1,j-1} \\
A^*_{min} = \lfloor Q_{ij} \rfloor - 2 \times \lfloor Q_{i+1,j+1} \rfloor + \lfloor Q_{i,j+1} \rfloor + \lfloor Q_{i-1,j-1} \rfloor \\
A^*_{max} = \lfloor Q_{ij} \rfloor - 2 \times \lfloor Q_{i+1,j+1} \rfloor + \lfloor Q_{i,j+1} \rfloor + \lfloor Q_{i-1,j-1} \rfloor
\]

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

<table>
<thead>
<tr>
<th>Bitrate</th>
<th>RER</th>
<th>RHFCN</th>
<th>WASDN</th>
<th>MDI2</th>
<th>ADOTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.1</td>
<td>87.18</td>
<td>83.71</td>
<td>58.48</td>
<td>56.84</td>
</tr>
<tr>
<td>320</td>
<td>0.3</td>
<td>92.77</td>
<td>88.05</td>
<td>68.11</td>
<td>65.13</td>
</tr>
<tr>
<td>640</td>
<td>0.5</td>
<td>95.18</td>
<td>93.34</td>
<td>80.35</td>
<td>74.95</td>
</tr>
</tbody>
</table>

**Steganalysis with Size Mismatch**

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

<table>
<thead>
<tr>
<th>Bitrate</th>
<th>RER</th>
<th>RHFCN</th>
<th>WASDN</th>
<th>MDI2</th>
<th>ADOTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.1</td>
<td>93.26</td>
<td>90.08</td>
<td>68.79</td>
<td>68.30</td>
</tr>
<tr>
<td>320</td>
<td>0.3</td>
<td>97.96</td>
<td>82.17</td>
<td>60.79</td>
<td>60.30</td>
</tr>
<tr>
<td>640</td>
<td>0.5</td>
<td>98.74</td>
<td>74.37</td>
<td>57.71</td>
<td>56.74</td>
</tr>
</tbody>
</table>

**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.