电子与信息学报 2010 32 (5): 1136-1140 ISSN: 1009-5896 CN: 11-4494/TN

论文

基于高阶统计的网页隐秘信息检测研究

黄华军①, 谭骏珊①, 孙星明②

①中南林业科技大学计算机与信息工程学院 长沙 410004; ②湖南大学计算机与通信学院 长沙 410082

收稿日期 2009-4-13 修回日期 2009-12-29 网络版发布日期 2010-4-26 接受日期

摘要
隐秘信息能隐藏在网页标记字母中，虽在浏览器浏览时无法发现其存在，但却不可避免地改变了标记的内在特征标记偏移量。基于此，该文提出一种新的网页隐秘信息检测算法。根据标记偏移量在隐藏信息前和隐藏信息后的变换规律，确立高阶统计特征来检测网页标记中是否有隐秘信息。实验随机下载了30个不同类型网站的主页测试，实验结果验证了统计特征的正确性。检测的漏检率随嵌入信息的增大而减小，当50%的标记字母被用来隐藏信息后，检测的漏检率为0%。

关键词 信息隐藏 隐写术 隐写分析 网页 高阶统计 偏移

分类号 TP391

On Steganalysis of Information in Tags of a Webpage Based on Higher-order Statistics

Huang Hua-jun①, Tan Jun-shan①, Sun Xing-ming②

①School of Computer and Information Engineering, Central and South University of Forestry and Technology, Changsha 410004, China; ②School of Computer and Communication, Hunan University, Changsha 410082, China

Abstract
Secret message can be embedded into letters in tags of a webpage in ways that are imperceptible to human eye viewed with a browser. These messages, however, alter the inherent characteristic of the offset of a tag. This paper presents a new higher-order statistical steganalytic algorithm for detection of secret messages embedded in a webpage. The offset is used to build the higher-order statistical models to detect whether secret messages hide in tags. 30 homepages are randomly downloaded from different websites to test, and the results show the reliability and accuracy of statistical characteristics. The probability of missing secret messages decrease as the secret message increase, and it is zero, as 50% letters of tags are used to carry secret message.

Key words Information hiding Steganography Steganalysis Webpage Higher-order statistics Offset

DOI: 10.3724/SP.J.1146.2009.00530

通讯作者 黄华军 hhj0906@163.com

作者个人主页