

# 信息安全学报

Journal of Cyber Security

第 11 卷      第 2 期      2026 年 3 月

## 目 次

基于模型遗忘的深度神经网络鲁棒性水印方法 .....	1
任纪星, 许 葳, 汪 润, 李勃衡, 张钰洋, 王丽娜	
基于视觉 Transformer 的鲁棒伪造语音检测算法 .....	21
张 桐, 邓俊龙, 任延珍, 王丽娜	
面向语言模型的文本后门防御综述 .....	32
吴宗儒, 程彭洲, 张倬胜, 刘功申	
多源安全日志威胁量化分析 .....	62
冯文英, 顾钊铨, 赵昂霄, 罗 翠, 袁华平, 胡 宁	
面向 DeepFake 伪造模型溯源的逃避攻击 .....	80
吴梦洁, 于佳艺, 汪 润, 叶 茜, 简琛皓, 方黎明, 王丽娜	
基于说话行为相关面部关键点的鲁棒伪造人脸检测方案 .....	95
黄逸煊, 彭 荔, 任延珍, 王丽娜	
大模型对齐攻击综述 .....	110
宫润森, 王 凯, 张昱霖, 张伟哲, 乔延臣, 张玉清	
基于进程代数的 SP 网络结构密码形式化设计研究 .....	131
张 磊, 许弘可, 肖超恩, 王建新, 郑玉峭	
HTTPFuzzer: 强化学习引导的 Web 服务器程序灰盒模糊测试方法 .....	145
陈 乾, 洪 征, 江 川, 张国敏, 秦素娟, 古津榜, 崔 帅	
基于节点影响力和权重聚合签名的改进 PBFT 共识算法 .....	162
刘力汇, 邓小鸿, 刘 勇, 石亦燃, 张 丽	
基于秘密共享的可验证分层洗牌协议设计及其应用方案 .....	178
张艳硕, 满子琪, 周幸好, 杨亚涛, 谢绒娜	
Safety Classification Fine-tuning: 一种提高大模型输出内容安全性的微调方法 .....	191
于 淼, 孙 磊, 胡翠云, 臧韦菲, 郭 松, 胡 鹏	
面向大规模区块链网络的高效编辑方案 .....	209
高源芑, 冯 哲, 刘雪峰, 雷 静, 裴庆祺	
基于时频特征的多源融合信息泄漏检测方法 .....	221
冯 祺, 周永彬, 明经典, 张 倩	
抵御推断攻击的在线社交网络用户位置隐私保护综述 .....	237
马 卓, 曹玖新, 王 群, 胥 帅, 夏玲玲	
基于域适应的电磁泄漏还原图像中文文本识别 .....	258
吕志强, 于 超, 李海洋, 张 宁	
轻量级虚拟化技术安全研究综述 .....	273
孔 同, 王利明, 徐 震, 马多贺	
基于时序特征和结构特征的社交网络谣言检测方法 .....	289
卫玲蔚, 胡 斗, 鲍祎楠, 周 薇, 杨近朱, 虎嵩林	
跨社交网络用户身份链接回顾与展望 .....	300
薛 晖, 孙 波, 司成祥, 张 伟, 房 婧	
一种基于 RLWE 的三方口令认证密钥交换协议 .....	313
王梓梁, 顾小卓, 任培欣	

# Journal of Cyber Security

Volume 11 Issue 2 March, 2026

## Contents

A Robust Watermarking Scheme for Deep Neural Networks based on Machine Unlearning.....	1
<i>REN Jixing, XU Wei, WANG Run, LI Boheng, ZHANG Yuyang, WANG Lina</i>	
Robust Fake Audio Detection Algorithm based on Vision Transformer.....	21
<i>ZHANG Tong, DENG Junlong, REN Yanzhen, WANG Lina</i>	
A Survey on Textual Backdoor Defense for Language Models.....	32
<i>WU Zongru, CHENG Pengzhou, ZHANG Zhuosheng, LIU Gongshen</i>	
Quantitative Threat Analysis of Multi-source Security Logs.....	62
<i>FENG Wenying, GU Zhaoquan, ZHAO Angxiao, LUO Cui, YUAN Huaping, HU Ning</i>	
Evading Attacks for DeepFake Fake Model Traceability.....	80
<i>WU Mengjie, YU Jiayi, WANG Run, YE Xi, LIN Chenhao, FANG Liming, WANG Lina</i>	
A Robust Forged Face Detection Scheme Based on Speech-Related Facial Landmarks.....	95
<i>HUANG Yihuan, PENG Li, REN Yanzhen, WANG Lina</i>	
A Survey of Adversarial Techniques Against Large Model Alignment.....	110
<i>GONG Runsen, WANG Kai, ZHANG Yulin, ZHANG Weizhe, QIAO Yanchen, ZHANG Yuqing</i>	
Research on Formal Design of SP Network Cipher based on Process Algebra.....	131
<i>ZHANG Lei, XU Hongke, XIAO Chaoen, WANG Jianxin, ZHENG Yuzheng</i>	
HTTPFuzzer: Reinforcement Learning Guided Greybox Fuzzing for Web Server Programs.....	145
<i>CHEN Qian, HONG Zheng, JIANG Chuan, ZHANG Guomin, QIN Sujuan, GU Jinbang, CUI Shuai</i>	
Improved PBFT Consensus Algorithm Based on Node Influence and Weighted Aggregation Signature.....	162
<i>LIU Lihui, DENG Xiaohong, LIU Yong, SHI Yiran, ZHANG Li</i>	
Design of Verifiable Layered Shuffling Protocol based on Secret Sharing and Its Application Scheme.....	178
<i>ZHANG Yanshuo, MAN Ziqi, ZHOU Xingyu, YANG Yatao, XIE Rongna</i>	
Safety Classification Fine-tuning: A fine-tuning method to improve the output content safety of LLMs.....	191
<i>YU Miao, SUN Lei, HU Cuiyun, ZANG Weifei, GUO Song, HU peng</i>	
An Efficient Editing Scheme for Large-Scale Blockchain Networks.....	209
<i>GAO Yuanpeng, FENG Zhe, LIU Xuefeng, LEI Jing, PEI Qingqi</i>	
Time-Frequency Characteristics Based Multi-Channel Fusion Leakage Detection.....	221
<i>FENG Qi, ZHOU Yongbin, MING Jingdian, ZHANG Qian</i>	
A Survey for User Location Privacy Protection against Inference Attacks in Online Social Networks.....	237
<i>MA Zhuo, CAO Jiuxin, WANG Qun, XU Shuai, XIA Lingling</i>	
Chinese Text Recognition in Electromagnetic Emission Reconstructed Images Based on Domain Adaptive.....	258
<i>LV Zhiqiang, YU Chao, LI Haiyang, ZHANG Ning</i>	
Survey on Lightweight Virtualization Technology Security.....	273
<i>KONG Tong, WANG Liming, XU Zhen, MA Duohe</i>	
Jointly Exploiting Temporal and Structural Features for Rumor Detection on Social Media.....	289
<i>WEI Lingwei, HU Dou, BAO Yinan, ZHOU Wei, YANG Jinzhu, HU Songlin</i>	
Advance in user identity linkage across online social networks.....	300
<i>XUE Hui, SUN Bo, SI Chengxiang, ZHANG Wei, FANG Jing</i>	
A RLWE-based Three-party Password Authenticated Key Exchange Scheme.....	313
<i>WANG Ziliang, GU Xiaozhuo, REN Peixin</i>	