通信学报
Journal on Communications
主管单位：中国科学技术协会
主办单位：中国通信学会

徐小琳1,2,3,4,云晓春1,2,3,4,周勇林4 ,康学斌5.基于特征聚类的海量恶意代码在线自动分析模型[J].通信学报,2013,(8):146~153

基于特征聚类的海量恶意代码在线自动分析模型

Online analytical model of massive malwarebased on feature clusting

投稿时间： 2013-05-07

DOI：10.3969/j.issn.1000-436x.2013.08.019

中文关键词：恶意代码 在线自动分析 快速聚类 特征提取

英文关键词:malware on-line analytical fast clustering feature extraction

基金项目:国家高技术研究发展计划(“863”计划)基金资助项目(2013AA014700)；国家科技支撑计划基金资助项目(2012BAH46B02)；中国科学院战略性科技先导专项基金资助项目(XDA06030200)

作者	单位
徐小琳1,2,3,4，云晓春1,2,3,4，周勇林4 ，康学斌5	1. 中国科学院 计算技术研究所，北京100190；2. 中国科学院大学，北京100049；3. 中国科学院 信息工程研究所，北京100093；4. 国家计算机网络应急技术处理协调中心，北京100029；5. 安天实验室，黑龙江 哈尔滨150040

中文摘要:

针对传统海量恶意代码分析方法中自动特征提取能力不足以及家族判定时效性差等问题，通过动静态方法对大量样本行为构成和代码片段分布规律的研究，提出了基于特征聚类的海量恶意代码在线自动分析模型，包括基于API行为和代码片段的特征空间构建方法、自动特征提取算法和基于LSH的近邻聚类算法。实验结果表明该模型具有大规模样本自动特征提取、支持在线数据聚类、家族判定准确率高等优势，依据该模型设计的原型系统实用性较强。

英文摘要:

In order to improve the effectiveness and efficiency of mass malicious code analysis, an online analytical model was proposed including feature space construction, automatic feature extraction and fast clustering. Our research focused on the law of malware behavior and code string distribution by dynamic and static techniques. In this model, a sample was described with its API and key code fragment. This model proposed a fast clustering approach to identify group samples that exhibit similar feature when applied this model to real-world malware collections. The result demonstrates that the proposed model is able to extract feature automatically, support streaming data clustering on large-scale, and achieve better precision.

查看全文 查看/发表评论 下载PDF阅读器

关闭