PaperTime检测报告简明打印版

相似度：67.26%
编　号：JPEA3MZM8JSSGEDT
标　题：基于机器学习算法的恶意代码检测技术研究
作　者：朱鹏博
长　度：13629字符
时　间：2017-12-17 14:58:21
比对库：本地库（学术期刊、学位论文、会议论文）；PaperTime云论文库；互联网

本地库相似资源（学术期刊、学位论文、会议论文）
1. 相似度：3.03%　篇名：《中国互联网产业报告》
来源：《财经界》　年份：2004　作者：玛丽·米克
2. 相似度：1.51%　篇名：《网络群体性事件的国内外研究现状》
来源：《群文天地:下半月》　年份：2012　作者：余爽
3. 相似度：0.53%　篇名：《统计信息网络安全研究》
来源：《统计与咨询》　年份：2013　作者：吕娜
4. 相似度：0.49%　篇名：《恶意代码检测技术研究》
来源：《扬州大学硕士论文》　年份：2012　作者：汪英爽
5. 相似度：0.34%　篇名：《计算机模块教学法初探》
来源：《职业教育研究》　年份：2004　作者：孙萍
6. 相似度：0.34%　篇名：《基于LZW算法的未知恶意代码检测方法》
来源：《北京工业大学学报》　年份：2012　作者：赖英旭
7. 相似度：0.18%　篇名：《机器学习应用于恶意代码检测的研究》
来源：《科技通报》　年份：2013　作者：史晓红
8. 相似度：0.14%　篇名：《采用静态分析检测android应用信息泄露的研究》
来源：《复旦大学硕士论文》　年份：2013　作者：彭智俊
9. 相似度：0.12%　篇名：《中职学校职业指导与创业教育的研究》
来源：《学园》　年份：2014　作者：李凯
10. 相似度：0.11%　篇名：《一种验证指针程序的方法》
来源：《微型机与应用》　年份：2011　作者：张志天
11. 相似度：0.10%　篇名：《基于机器学习的android恶意软件检测模型研究》
来源：《青岛理工大学硕士论文》　年份：2013　作者：蔡泽廷
12. 相似度：0.09%　篇名：《基于干净数据的恶意软件检测技术研究》
来源：《西南交通大学硕士论文》　年份：2011　作者：李鹏飞
13. 相似度：0.09%　篇名：《学习资源语义特征自动提取研究》
来源：《中国电化教育》　年份：2013　作者：杨现民
14. 相似度：0.08%　篇名：《基于数据流分析的网络协议逆向解析技术》
来源：《计算机应用》　年份：2013　作者：戴理
15. 相似度：0.08%　篇名：《基于人工神经网络的人脸识别方法》
来源：《吉林大学；曹欢欢硕士论文》　年份：2016　作者：曹欢欢
16. 相似度：0.07%　篇名：《网络恶意代码隐藏技术的分析及检测》
来源：《中国海洋大学硕士论文》　年份：2011　作者：王志乐
17. 相似度：0.07%　篇名：《基于CS和BS结构的工资管理系统的设计》
来源：《科学与财富》　年份：2011　作者：李华
18. 相似度：0.06%　篇名：《一种基于主动学习的数据库恶意行为检测方法》
来源：《网络安全技术与应用》　年份：2012　作者：车晶
19. 相似度：0.06%　篇名：《恶意代码的分析技术》
来源：《科技创新导报》　年份：2012　作者：袁慎芳
20. 相似度：0.06%　篇名：《机器学习中的特征选择算法研究》
来源：《中国海洋大学硕士论文》　年份：2009　作者：姜百宁

PaperTime云论文库(知网，万方，维普，百度文库等镜像）

1. 相似度：29.83% 标题：《基于操作码序列的静态恶意代码检测方法的研究 - 道客巴巴》
来源：http://www.doc88.com/p-0012453527162.html
2. 相似度：5.95% 标题：《一种基于操作码序列的快速病毒检测方法 - 道客巴巴》
来源：http://www.doc88.com/p-2502898166575.html
3. 相似度：5.60% 标题：《恶意代码检测中若干关键技术研究 - 道客巴巴》
来源：http://www.doc88.com/p-3728936276040.html
4. 相似度：3.52% 标题：《基于操作码序列的静态恶意代码检测方法的研究入 - 道客巴巴》
来源：http://www.doc88.com/p-6771221006690.html
5. 相似度：1.62% 标题：《结合语义的机器学习方法在软件安全中应用研究_百度文库》
来源：http://wenku.baidu.com/view/458cbfdd43323968001c9263.html
6. 相似度：1.14% 标题：《2016年中国互联网安全报告_图文_百度文库》
来源：http://wenku.baidu.com/view/596c1fd9bb0d4a7302768e9951e79b89680268e0.html
7. 相似度：0.49% 标题：《基于沙箱技术的恶意代码行为自动化检测方法_CNKI学问》
来源：http://xuewen.cnki.net/CMFD-1015909448.nh.html
8. 相似度：0.41% 标题：《基于行为分析的恶意代码检测与评估研究_CNKI学问》
来源：http://xuewen.cnki.net/CMFD-1014141151.nh.html
9. 相似度：0.38% 标题：《基于程序语义的静态恶意代码检测系统的研究与实现 - 道客巴巴》
来源：http://www.doc88.com/p-9951473603713.html
10. 相似度：0.34% 标题：《僵尸工具类恶意代码的检测研究_图文_百度文库》
来源：http://wenku.baidu.com/view/5be0ee4b2b160b4e767fcfc3.html
11. 相似度：0.34% 标题：《恶意代码行为自动化分析的研究与实现_图文_百度文库》
来源：http://wenku.baidu.com/view/5171bf11a216147917112857.html
12. 相似度：0.31% 标题：《移动设备恶意代码分析与检测技术研究》
来源：http://d.wanfangdata.com.cn/Periodical/xxydn201714012
13. 相似度：0.17% 标题：《(计算机应用技术专业论文)基于二进制多态变形的恶意代码反检...》
来源：http://www.doc88.com/p-3107130240505.html
14. 相似度：0.17% 标题：《基于可信基的恶意代码诊断技术研究_CNKI学问》
来源：http://xuewen.cnki.net/CMFD-2008032594.nh.html
15. 相似度：0.15% 标题：《内核Rootkit进程隐藏与检测技术研究-中国仪器仪表学会...》
来源：http://wap.cnki.net/huiyi-ZGYF201008001034.html
16. 相似度：0.13% 标题：《9、恶意代码与病毒 - 道客巴巴》
来源：http://www.doc88.com/p-297946058743.html
17. 相似度：0.12% 标题：《恶意代码检测系统的研究与实现》
来源：http://d.wanfangdata.com.cn/Thesis/Y1802445
18. 相似度：0.10% 标题：《浅谈恶意代码分析技术发展趋势_CNKI学问》
来源：http://xuewen.cnki.net/CJFD-ZXLJ201316015.html
19. 相似度：0.09% 标题：《基于动态污点分析的漏洞攻击检测技术研究与实现_图文_百度文库》
来源：http://wenku.baidu.com/view/08c2571402768e9950e73833.html
20. 相似度：0.09% 标题：《基于Hadoop海量数据处理关键技术研究》
来源：http://d.wanfangdata.com.cn/Thesis/D770182
21. 相似度：0.07% 标题：《基于语义的恶意行为分析方法_百度文库》
来源：http://wenku.baidu.com/view/bcb1e00633687e21ae45a93a.html
22. 相似度：0.07% 标题：《基于亲缘性分析的恶意代码检测技术研究与实现--《电子科技大学》...》
来源：http://cdmd.cnki.com.cn/Article/CDMD-10614-1015706364.htm
23. 相似度：0.07% 标题：《恶意代码检测及其行为分析 - 道客巴巴》
来源：http://www.doc88.com/p-7867394178177.html
24. 相似度：0.07% 标题：《基于虚拟化的恶意代码检测技术研究与实现 - 道客巴巴》
来源：http://www.doc88.com/p-9002855116973.html

互联网相似资源(博客，百科，论坛，新闻等）

1. 相似度：2.81% 标题：《盘点国内外2017年上半年网络安全事件_搜狐科技_搜狐网》
来源：http://www.sohu.com/a/161339624_604699
2. 相似度：2.01% 标题：《腾讯安全发布《2017年上半年互联网安全报告》-中新网》
来源：http://www.chinanews.com/it/2017/08-04/8295730.shtml

3. 相似度：1.24% 标题：《腾讯安全发布《2017年上半年互联网安全报告》-ITBEAR科技资讯》
来源：http://www.itbear.com.cn/html/2017-08/236988.html
4. 相似度：0.86% 标题：《腾讯安全发布《2017 年上半年互联网安全报告》-ZAKER新闻》
来源：http://www.myzaker.com/article/5983ed141bc8e0a00700003d
5. 相似度：0.78% 标题：《细数2017年国内信息安全大事件,提升个人信息安全防护意识》
来源：http://baijiahao.baidu.com/s?id=1585548828823156019&wfr=spider&for=pc
6. 相似度：0.66% 标题：《不可忽视的信息安全,国务院某App的H5遭遇流量劫持- 简书》
来源：http://www.jianshu.com/p/5a06d9934531
7. 相似度：0.60% 标题：《恶意代码依赖 - CSDN博客》
来源：http://blog.csdn.net/woshidaniu/article/details/2176383
8. 相似度：0.45% 标题：《不可忽视的信息安全，国务院某App的H5遭遇流量劫持 ...》
来源：https://www.trustauth.cn/news/security-news/15435.html
9. 相似度：0.45% 标题：《2016年中国互联网安全报告_图文 三亿文库》
来源：http://m.3y.uu456.com/mbp_3hlwf8hu6l2wkqq4mj6h371qz5d0ci00kps_1.html
10. 相似度：0.34% 标题：《12306官网用户数据又遭泄露:漏洞还没补完?-12306,网站,漏洞,火车...》
来源：http://news.mydrivers.com/1/528/528947.htm
11. 相似度：0.34% 标题：《静态检测技术_互动百科》
来源：
http://www.baike.com/wiki/%E9%9D%99%E6%80%81%E6%A3%80%E6%B5%8B%E6%8A%80%E6%9C
12. 相似度：0.24% 标题：《基于程序语义的静态恶意代码检测系统的研究和实现.pdf文档全文免费... 》
来源：
http://k.sogou.com/t/uID=bZPF1xoNdLaSQVWe/v=5/type=1/sp=1/ct=171217145908/keyword=%E9
aec0-4de2-8c95-4e4bc3faf168/sec=O56T6gei1DrfDdWtgkcruA../dp=1/pr=b/t?pg=webz&clk=4&url=
aec0-4de2-8c95-4e4bc3faf168&dp=1&userGroupId=21&key=%E9%9A%8F%E5%90%8E%E4%BB%96
13. 相似度：0.22% 标题：《360发布报告:中国成为全球APT攻击的第一目标国》
来源：http://field.10jqka.com.cn/20170213/c596405184.shtml
14. 相似度：0.22% 标题：《...计算机病毒具有寄生性、潜伏性、传染性、隐蔽性、破坏性、可...》
来源：http://bbs.pinggu.org/jg/bylw_dianzishangwubiyelunwen_93031_2.html
15. 相似度：0.22% 标题：《RSA2017:360谭晓生阐述处置高级威胁的行业趋势_财经频道_...》
来源：http://news.10jqka.com.cn/20170216/c596471723.shtml
16. 相似度：0.19% 标题：《恶意软件攻击防范与应急指导手册》
来源：http://www.360doc.com/content/14/0811/12/17799864_401008338.shtml
17. 相似度：0.17% 标题：《给文件加壳,加花,去壳,分别是什么意思?_百度知道》
来源：https://zhidao.baidu.com/question/569132743.html
18. 相似度：0.15% 标题：《《恶意代码分析实战》试读:第0章恶意代码分析技术入门》
来源：https://book.douban.com/reading/31367382/
19. 相似度：0.15% 标题：《如果反病毒软件不工作 有可能中了Rootkit Rootkit检测和清除还是有办...》
来源：https://yq.aliyun.com/articles/204653
20. 相似度：0.13% 标题：《普通恶意代码技术分析与检测 - 维维豆奶的日志 - 网易博客》
来源：http://blog.163.com/wajika@126/blog/static/7759339920084169405258 2/
21. 相似度：0.13% 标题：《普通恶意代码技术分析与检测_黑客吧_百度贴吧》
来源：http://tieba.baidu.com/p/554150774
22. 相似度：0.13% 标题：《基于动态二进制插桩的恶意代码行为分析方法的研究.pdf文档全文免费... 》
来源：
http://k.sogou.com/t/uID=ZMSusBlKBhgYdgbQ/v=5/type=1/sp=1/ct=171217145957/keyword=%E5%
f5f81d56ea3e/sec=O56T6gei1DqFxl7rAxe-
lA../dp=1/pr=b/t?pg=webz&clk=8&url=http%3A%2F%2Fmax.book118.com%2Fhtml%2F2015%2F092
f5f81d56ea3e&dp=1&userGroupId=17&key=%E5%8A%A8%E6%80%81%E5%88%86%E6%9E%90%E6
23. 相似度：0.13% 标题：《20159313网络攻击与防范第九周学习总结 - 20159313 ...》
来源：http://www.cnblogs.com/hougaopan/p/5444764.html
24. 相似度：0.12% 标题：《木马今年收入超百亿 黑客产业链正在形成 技术攻略 - 专..._专业玩家》
来源：http://www.zhuanyewanjia.com/news/09112614094236
25. 相似度：0.12% 标题：《计算机病毒模块主要由哪几个功能模块构成?_百度知道》
来源：https://zhidao.baidu.com/question/1988762335294865107.html
26. 相似度：0.12% 标题：《随着互联网的发展和普及,网络购物在中国也变得越来越普遍了,甚至...》

来源：https://zhidao.baidu.com/question/809326937902477572
27. 相似度：0.12% 标题：《特洛伊木马的检测2_muniao8488_新浪博客》
来源：http://blog.sina.com.cn/s/blog_56a7e4440100006u.html
28. 相似度：0.10% 标题：《人事考勤工资管理系统的设计与实现毕业设计doc下载_爱问共享资料》
来源：http://ishare.iask.sina.com.cn/f/32M2rA5xrJn.html
29. 相似度：0.10% 标题：《基于SVM的Android应用程序安全检测 - ...》
来源：http://blog.csdn.net/gaiyindexingqiu/article/details/62422058
30. 相似度：0.09% 标题：《恶意软件攻击防范与应急指导手册_安全白皮书_IT专家网》
来源：http://security.ctocio.com.cn/whitepapers/173/8286673.shtml
31. 相似度：0.08% 标题：《面向网页JavaScript恶意代码的智能检测方法》
来源：http://www.xjishu.com/zhuanli/55/201210092707.html
32. 相似度：0.08% 标题：《[转载]代码虚拟化- CSDN博客》
来源：http://blog.csdn.net/heikefangxian23/article/details/50247329
33. 相似度：0.07% 标题：《人工智能反欺诈三部曲之:设备指纹-猛犸反欺诈-51CTO博客》
来源：http://blog.51cto.com/12755572/2049360
34. 相似度：0.07% 标题：《特征离散化系列（一）方法综述 - CSDN博客》
来源：http://blog.csdn.net/CalCuLuSearch/article/details/52751218
35. 相似度：0.07% 标题：《面向体感游戏的人体运动生成方法 - 挑战杯》
来源：http://www.tiaozhanbei.net/project/197/
36. 相似度：0.07% 标题：《分析Android银行木马GM Bot的变异过程 - omnispace的博客 - CSDN博...》
来源：http://blog.csdn.net/omnispace/article/details/77964677
37. 相似度：0.06% 标题：《你不知道的Android SDK安全测试- CSDN博客》
来源：http://blog.csdn.net/liufangaliya/article/details/52180609

全文简明报告

{ 55% : 随着互联网的普及,恶意代码的危害也变得越来越来越难以控制。 }{ 63% : 各种病毒、木马、蠕虫等恶意代码在网络间广泛传播, }{ 57% : 已经给个人、企业甚至政府带来了难以估量的损失。 }{ 77% : 据国内知名互联网安全厂商奇虎360发布的《2016年中国互联网安全报告》显示, }{95% : 2016年全年,360互联网安全中心共截获PC端新增恶意程序样本1.9亿个。 }{100% : 敲诈者病毒在国内发生两次大规模传播,全国至少有497多万台用户电脑遭到了敲诈者病毒攻击。 }{100% : 通过对受害者调研,42.6%的受害者不知道感染病毒的原因。 }{100% : 预计在2017年敲诈者会增长10倍,且利用挂马攻击也将再次爆发。 }{100% : 2016年360互联网安全中心共截获Android平台新增恶意程序样本1403.3万个,其中资费消耗类程序为74.2%。 }{100% : 同PC端相似,手机端勒索软件也开始爆发,360全年截获新增手机勒索软件17万,170万台手机遭到攻击。 }{100% : 在截获盗取个人信息的手机恶意程序样本中,67.4%的样本会窃取短信信息,34.8%的样本会窃取手机银行信息,10.0%的样本会窃取手机联系人信息,3.7%的样本会窃取手机通话记录,2.0%的样本会窃取社交软件(例如微信、QQ等)聊天记录,1.8%的样本会窃取手机录音信息,0.1%的样本会窃取手机照片信息。 }{81% : 根据中国互联网安全报告显示,以下是2017年上半年5大典型安全威胁事件: }

{97% : (1)2017年上半年“WannaCry”、“暗云Ⅲ”、“Petya”等多种类型的病毒木马连续集中的爆发,为社会和行业敲响了网络安全的警钟。 }{100% : 《报告》显示,仅2017年上半年,腾讯安全反病毒实验室在电脑端总计已拦截病毒已超过10亿次,平均每月拦截木马病毒近1.7亿次,相较于2016下半年病毒拦截总量增长30%。 }{100% : 其中,受“WannaCry”刺激,勒索类病毒仅第二季度就新增了13.39%,但“WannaCry”在非感染型敲诈类病毒占比中仅排第三,而带有感染传播方式的PolyRansom勒索病毒传播力更巨大, }{100% : 其占所有勒索类病毒的78.84%。 }

{81% : (2)国务院某App的H5遭遇流量劫持。 }{100% : 5月中旬,某国字号的App遭遇流量劫持的传闻在业界流传。 }{99% : 有消息称,该App某H5页面被植入色情内容广告,经排查“基本确定为用户当地运营商http劫持导致H5页面被插入广告……”。 }

{80% : (3)12306官方网站再现安全漏洞。 }{100% : 4月21,有媒体记者发现在12306官方网站订票时发现,当退出个人账号,网站页面竟自动转登他人账号,且与账号相关联的身份证号、联系方式等个人信息均可见,随后记者在该页面点击常用联系人选项时页面再次刷新并显示他人账号及账号涵盖的所有信息。 }{100% : 而记者尝试在网站账户页面的个人信息栏等其他选项进行操作,点击进入后均得到不同的个人

身份信息。 }

{ 75% : (4)上亿优酷信息数据在暗网售卖。 }{93% : 4月17日,外国媒体hackread报道,100759591条优酷账户信息数据库在暗网售卖,该数据库售卖价格定为比特币 0.2559,,人民币约 2065.56 元。 }

{81% : (5) “土耳其犯罪家庭”的网络犯罪团伙掌握3亿苹果帐户。 }{100% : 3月底,国外媒体报道,自称为“土耳其犯罪家庭”的网络犯罪团伙,通过电子邮件告知苹果公司他们掌握了超过3亿苹果帐户,并能远程清除所有装置的内容。 }{100% : 他们宣表示只想苹果支付75000美元的比特币赎金,或者价值10万美元的iTunes礼品卡。 }{100% : 若苹果公司在4月7日拒绝遵守他们的要求,他们将大量清除iCloud帐户。 }

{ 77% : 在所有的网络安全事件中,尤以恶意代码的危害性最大,其带来的经济损失占很大比例。 }从八十年代初期开始出现第一个病毒开始,{ 65% : 到如今恶意代码的不断发展壮大, }入侵与防护的战役从未停止。{ 73% : 纵观恶意代码发展历史,在利益的驱使下,无论其发展速度还是其破坏性都在不断增强, }并且为了对抗杀毒软件的检测,各种抗查杀技术的应用,{ 77% : 使得恶意代码变得越来越复杂,从简单的AppleII病毒发展到复杂的内核病毒。 }{84% : 恶意代码的传播机制也发生了很大的变化, }从以前的被动传播到如今的主动传播。{ 61% : 由于现在出现了很多恶意代码的编制工具,并且极易获得,使得恶意代码制造成本越来越低,所以恶意代码的发布变得越来越频繁。 }{ 60% : 恶意代码的发展,给个人、企业以及政府带来的损失时难以估量的, } { 60% : 据统计80%以上的用户曾有意识或者无意识的遭受过恶意代码的侵袭。 }现如今,提高互联网安全已经成为国家的一项重要基本战略。

1.1.2 课题意义

很大一部分网络安全事件发生都是由恶意代码引起的,并且根据以往案例来看,往往都是恶意代码造成一定的损失之后,针对该恶意代码的分析及检测技术才会被提出。出现这种情况的原因无非有两个:首先是恶意代码越来越复杂,并且种类繁多,传播形式多种多样,使得用户很容易被感染;{ 62% : 其次是恶意代码检测技术不够成熟, }尽管很多学者为恶意代码检测做了很多研究,也提出了很多检测方法,但是理论到应用总是需要时间,而这些时间给了恶意代码去进一步变异的可能。此消彼长下,恶意代码的危害始终存在,而检测与反检测的斗争不会停止。

恶意代码能够在短时间内造成大范围传播的原因主要是信息共享技术和使用的普及。信息共享作为时代发展的必然产物,与人方便的同时,也给不法分子打开了新世界的大门,信息的快速流动加速了恶意代码的入侵。无论是Internet上的网页,还是光盘、U盘甚至接收到的Email都有可能携带恶意代码,防不胜防。可是时代的发展又离不开信息的共享技术,机遇与挑战并存,因噎废食是不明智也是不可取的。

近年来,恶意代码的使用与各种各样的经济甚至政治利益互相牵扯,其危害性和隐藏性日益增强,破坏的目标、目的以及要达成的后果更加具有针对性。{ 66% : 而恶意代码的编写者也从最初的技术炫耀逐渐变为经济或者政治利益的追逐者。 }{86% : 中国木马产业链一年的收入已逾上百亿元, }黑色产业链正在逐渐成型。{ 55% : 恶意代码的检测与反检测注定会是一场持久战。 }

{ 58% : 综上所述,恶意代码的危害无处不在, }{ 76% : 不仅给个人、企业带来了巨大的损失, }甚至可能给国家安全带来不可预期的危害。{ 66% : 对于个人,恶意代码的入侵会导致个人隐私的泄露,造成经济或者名誉损失;对于企业来说,企业数据一旦遭到恶意代码的入侵,会导致企业大量数据资产外泄,给企业带来无可挽回的损失,甚至因此产生一些灰色的产业链; }{ 73% : 对于国家,信心安全是国家安全的重点组成部分,网路安全已经是国策和民生的大问题,信息安全成为国家战略,随着互联网的发展, }信息安全的问题将会更加突出和重要。{ 71% : 因此研究更加有效的恶意代码检测技术是非常具有现实意义的。 }

1.2国内外研究现状

{88% : 自1981年第一个病毒 Apple II[2]出现以来,国内外许多计算机安全的学者 便投身于与恶意代码的对抗过程中。 }{97% : 恶意代码编写技术的发展也推动了检测技 术的发展,直到现在已经有很多恶意代码检测技术被广泛应用。 }{97% : Sung[3]等人提出了基于系统调用的静态恶意代码检测方法,主要针对恶意代码的变种。 }{100% : 该方法是将恶意代码反汇编并根据反汇编后的文本信息提取系统调用序列,并通过系统调用序列的相似度来判断。 }{98% : 基于系统调用序列方法也可以用在动态检测过程中,在虚拟环境中执行恶意程序时可以提取执行时的系统调用序列,并使用n-gram算法来提取特征,然后进行分类。 }{96% : 张波云[4]等人在虚 拟环境中动态获得可执行文件的系统调用序列,并使用 n-gram 算法提取特征,使用粗糙集理论对特征降维并使用支持向量机实施分类。 }{100% : 为了解决混淆技术带来的困惑,一些学者研究基于程序的语义分析方法。 }{96% : 语义分析是通过形式化抽象指令运行时的语义,通过符号执行[5]、模型检验[6]、逻辑推理证明等方法来分析程序的语义信息。 } Cousot.P 和 Cousot.{85% : R[7]提出了程序分析构造和逼近不动点语义理论,这为程序的语义分析提供了理论基础。 }{97% : M.Christodorescu[8]引入抽象模式库作为恶意行为自动机的符号,将恶意行为表

示为带未解释符号的自动机,最后使用模型检验来实现检测。 }{96% : 随后他提出一种基于语义的检测方法,用迹语义来描述恶意代码的行为,采用抽象解释方法检测恶意行为[9]。 }{97% : D.Preda[10]也借鉴了抽象解释的思想,证明了关于混淆技术产生的恶意代码检测的正确性和完备性。 }{95% : Singh[11]通过分析反汇编文本的数据流信息,利用线性时态逻辑语义模型检测恶意行为。 }{97% : Kinder[12]分析了程序的控制流程图和函数之间的调用关系,用计算逻辑树描述恶意行为并公式化,最终使用模型检测方法检测。 }{98% : 李佳静[13]等人提出了一种基于语义的行为分析方法,对函数调用及函数调用序列之间的依赖关系进行了详细的描述,该方法能准确描述恶意行为并有很好的泛化能力。 }{100% : 用有穷自动机描述恶意行为,并引入数据流分析使用下推自动机描述程序的全局状态空间以提高分析精度,最终使用模型检测器实现检测。 }{98% : 王晓洁和王海峰[14]提出一种基于语义模型匹配的检测算法,通过语义描述恶意代码的行为,这样对经过代码混淆技术处理的恶意代码的检测有很好的效果。 }{81% : 孔德光[15]等人提出一种结合语义的多态蠕虫的签名提取算法,提高了检测的鲁棒性和准确性。 }{97% : G.Tahan[16]等人提出了一种新的自动签名提取算法,该算法 主要针对恶意的可执行文件,被应用到高速恶意代码过滤装置中。 }{97% : Y.Tang[17]等人提出了一种利用多序列对比技术的简化的正则表达式签名算法,这种方法能产生更加准确的基于漏洞的签名。 }{96% : Y.Chen[18]等人提出了在网络层没有任何主机分析的蠕虫执行的脆弱性驱动的签名,实验效果非常好。 }{100% : 现有的基于签名的恶意代码检测技术通过特殊的字符串特征来判断,其准确率非常高,但是其缺点是不能检测新出现的恶意代码,并且需要不断的更新特征库。 }{100% : 现在大部分研究用基于 n-gram 序列的字节序列代替二进制特征码序列,这会提高分类的准确率。 }{94% : Robert[19]等人提出了用操作码序列作为特征,然后使用文本分类的方法实现检测,并解决了数据不平衡问题[20]。 }{98% : Schultz[21]等人第一次提出了应用数据挖掘模型来检测恶意代码,他们提取三种特征并使用不同的分类方法:程序的头文件信息,字符串信息,字节的序列,应用基于签名、基于规则的学习器 Pipper、朴素贝叶斯等方法进行分类。 }{98% : 研究表明使用机器学习方法能提高准确率。 }{96% : 后来 Kolter[22]使用 n-gram 算法提取 字节序列作为特征,改进的决策树算法取得了很好的分类效果。 }{97% : 在参考文献[23]中,作者提出了使用 n-gram 算法提取特征后使用信息增益 的方法来选择一些分类效果好的特征,并使用 K 近邻,基于 TFIDF 的分类器、 朴素贝叶斯、支持向量机、决策树等分类方法,并取得了很好的实验效果。 }{93% : Kolterh 和 Maloof[22]研究了恶意代码的家族的分类,基于恶意代码的功能行 为,使用多分类方法将恶意代码分为蠕虫、木马、后门、病毒等,这更加细化 了分类的结果,有助于对每一种恶意代码的研究,发现它们的共性,这也为以 后的语义分析等方法奠定了基础。 }{97% : 文献[24]中作者提出了一个层次特征选择的方法,即使用 n-gram 算法提取特征后选择那些出现频率高于某个阈值的特征, 这种方法对于检测恶意代码的变种很有效。 }{86% : Raja[25]等人应用数据挖掘方法实现 恶意代码的检测,他通过反汇编技术提取了恶意代码的操作码序列,使用了一 种新的在文本分类领域的特征选择方法 CPD(Categorical Proportional Difference)。 }{98% : CPD 用来度量一个特征的区分能力,最终分类效果相对比较好。 }{95% : Dolev[26]提倡使用操作码来作为恶意代码的中间表示。 }{100% : 操作码是机器语言的一个操作的一部分,它包含着指令的行为和程序的控制。 }{94% : 近年来,操作码特征已经被用来检测蠕虫的变种和一些间谍软件[27]。 }{100% : 将操作码提取出来作为标签,然后产生签名来判别恶意代码的变种。 }{96% : 后来有些学者提取操作码并将其转化成操作码序列来检测未知的恶意代码[19],实验使用三种分类算法取得了很好的实验 效果。 }{ 75% : 在文献[28]中,作者提出了使用变长的指令序列作为特征,并使用 Bagging算法得到了很好的实验效果。 }{90% : 也有人使用了十六进制码作为特征[29]。 }{95% : 在恶意代 码检测技术中使用操作码序列作为特征的研究相对还是比较少的,但是研究结 果发现操作码序列是一种比较好的特征表示方法。 }{96% : 在文献[30]中,作者使用程序的控制流程图并用三种不同版本的黑客防御工具设计了一个分类算法,并取得了很好的实验效果。 }{90% : Ismail B[31]提出了将程序控 制流程图和函数调用拓扑用于将未知的恶意代码归类。 }{96% : 使用函数调用拓扑的缺 点是,攻击者能使用相似的函数调用或者改变函数调用的序列来逃避检测。 }{93% : Halvar[32]利用程序控制流程的拓扑图的同构来实现检测。 }{98% : 因为同一种族的恶意 代码的拓扑图基本相似,这种方法也是适合检测恶意代码的变种。 }{94% : Igor[33]提出了一个新的检测未知恶意代码族的方法。 }{96% : 该方法是基于操作码序列的出现的频 率,并挖掘了每一个操作码序列的相关性。 }{91% : 通过大量实验对比分析,该方法是 非常有效的。 }{96% : Perdisci[34]等人提出了从PE文件提取一些特征,如标准和非标准 部分的数目,可执行部分的数量以及 PE 头文件的熵信息,并使用不同的机器 学习模型实现分类。 }{94% : 后来他们开发了一个快速统计恶意代码的检测工具[35]。 }

{100% : 综上所述,现有的恶意代码检测技术有很多,每一种方法都有自身的优缺点。 }{100% : 这为后面的研究提供了基础的同时也带来了很多挑战。 }{100% : 本文提出了一种新的恶意代码检测方法,结合了特征码、行为及机器学习的方法,提出了基于操作码序列的静态恶意代码检测方法,能更好的检测恶意代码。 }

1.3研究内容

{ 74% : 本文改进了一种基于机器学习算法的恶意代码检测方法, }{ 65% : 并基于该方法实现了一个恶意代

码检测系统。 }方法主要是使用汇编操作码的抽象化技术将汇编码表示为中间码,并结合Eclat算法对中间码的频繁项集进行分析,预测出一种最优的中间码序列,{ 63% : 然后使用n-gram算法提取中间码特征序列生成概率矩阵, }作为代表恶意代码的特征,{ 58% : 最后使用机器学习算法进行建模。 }当抽象方式有多种的时候,这种方法能够针对n-gram算法提取出的中间码特征序列,预测出一种对分类效果最好中间码特征序列作为生成概率矩阵的依据。

在仿真实验中,本文提供了两种抽象方式,分别记为Abs1和Abs2,Abs1是根据作者的理解对汇编码提出的抽象方式,Abs2是文献[1]中提出了抽象方式,本文将会分别使用本文方法和文献方法进行实验并分析对比实验结果,给出结论。实验主要步骤有:首先,{ 60% : 为了逃避病毒检测系统的检测,一般的恶意代码作者都会对恶意程序进行加壳处理,所以本文的第一步就是对恶意代码进行查壳和脱壳处理;其次对恶意代码样本进行反汇编处理, }得到样本的汇编文本,并从中提取出汇编操作码序列;然后根据已配置的抽象方式对汇编操作码序列进行抽象化处理,得到各抽象方式对应的中间码序列;{ 64% : 接着使用n-gram算法获得中间码特征序列, }并对中间码特征序列的频繁项集进行分析,预测对分类效果最明显的中间码特征序列,并依据选择出的中间码特征序列生成概率矩阵;最后对比了随机森林算法、支持向量机以及K邻近三种机器学习算法使用概率矩阵作为输入的分类效果,结果显示随机森林算法效果最为显著和稳定。{ 59% : 另外根据本文提出的方法设计并实现了一个恶意代码检测系统。 }本文的工作有以下几点:

第一:收集实验样本,并对样本进行预处理操作,{91% : 然后提取基于文本的汇编操作码序列。 }

第二:{ 74% : 本文改进了一种基于机器学习算法的恶意代码检测方法。 }通过汇编操作码抽象化技术、Eclat算法频繁项集分析和n-gram算法提取特征序列获得概率矩阵,{ 59% : 并以此作为代表恶意代码的特征, }{ 60% : 最后使用机器学习算法构建分类模型。 }

第三:{ 62% : 对本文提出的方法进行实验仿真, }{ 71% : 并对实验结果进行对比分析,得出结论。 }

第四:结合实验仿真的结果分析,针对本文提出的改进方法,采用模块化编程技术实现了一个恶意代码检测系统。

1.4论文结构

本文总共分为四章:

{ 69% : 第一章为绪论,主要介绍课题背景及意义。 }{ 60% : 并且详细阐述了国内外对于恶意代码检测技术的研究现状和存在的问题,最后介绍了本文的研究内容和章节安排。 }

{ 69% : 第二章是相关理论与关键技术。 }{ 63% : 首先介绍了恶意代码的定义和分类,然后对现有恶意代码分析技术、检测技术和反检测技术做了相关介绍。 }

{ 74% : 第三章是基于机器学习算法的恶意代码检测方法。 }首先对汇编操作码的特点进行了分析,并介绍了有关学者对汇编码抽象化的一些研究;然后概要的介绍了本文提出方法的大体流程,接着对方法的各部分进行了详细说明,包括数据预处理、概率矩阵的生成过程和恶意代码的分类;最后根据本文方法进行了实验仿真,{ 56% : 并和传统方法的实验结果进行了对比分析, }得出结论。

{ 67% : 第四章是系统设计与实现。首先介绍整个系统的架构设计, }包括系统的功能分析及组成模块,并对各模块的功能进行了简单介绍;接着对各模块的实现方式进行了详细介绍;最后对本章进行总结。

第五章为总结与展望。{ 68% : 概括总结了本文的主要研究成果和不足, }{87% : 对未来的可研究方向进行了展望。 }

第二章 相关理论与关键技术

{ 79% : 本章首先介绍了恶意代码相关概念, }并对现有的恶意代码检测技术以及反检测技术进行了详细介绍;{ 68% : 然后对恶意代码的反检测技术做了相关介绍; }最后对恶意代码的分析技术做了总结。

2.1 恶意代码简介

2.1.1 恶意代码的定义

恶意代码也成为恶意软件,是对各种敌对和入侵软件的概括性术语。{ 55% : 包括各种形式的计算机病毒、蠕虫、特洛伊木马、勒索软件、间谍软件、广告软件以及其他的恶意软件。 }形式上多种多样,可以是可执行文件、脚本、插件等等。其违背使用者的意愿去执行一些操作,损害用户的利益以达到入侵者不可告人的目的。

2.1.2 恶意代码的分类

{ 65% : 根据不同的依据,恶意代码有很多种不同的分类方法,没有一种标准的分法,但是常见的种类有: }{ 75% : 计算机病毒、蠕虫、特洛伊木马、间谍软件、勒索软件等等。 }{90% : 下面对几种恶意代码做简要介

绍: }

{100% : (1)计算机病毒。病毒是早期产生的最主要的恶意代码之一,病毒是能够自我繁殖并寄生在其他程序中的代码,这个被寄存的程序被称为宿主程序,但是病毒不能单独运行,必须通过激活宿主程序并满足一定条件下,病毒就能干扰电脑正常工作,扰乱或破坏己有存储的信息,甚至引起整个系统不能正常工作。 }{99% : 一般而言计算机病毒通常由三个单元和一个标志构成:引导模块、感染模块、破坏表现 模块和感染指标。 }{ 79% : 1、引导模块是指将计算机病毒感染的宿主程序设法引导安装到 }

计算机操作系统中,{100% : 为以后的感染、破坏两个后期模块提供前期的有效准备,一般而言不同的计算机病毒有不同的引导操作,而且引导操作往往是隐蔽的,不易被用户察觉和发现的。 }{98% : 2、感染 模块包括两个部分,一个是用来激活感染功能的判断部分。 }该模块提供一个感染

{ 72% : 的标志,用来判断计算机是否被感染。 }{100% : 另一个是执行感染功能部分。这一部分主要的功能就是监控宿主满足条件的时机,并及时的将计算机病毒存入到系统特定的位置。 }{100% : 3、破坏表现模块与感染模块一样包括两个部分,一是具有触发破坏表现功能的判断部分。 }{100% : 二是具有破坏表现功能的实施部分。 }{93% : 计算机病毒一般具有寄生 性、传染性、隐藏性、破坏性、潜伏性等特征。 }

{96% : (2)特洛伊木马。木马分为客户端和服务端,客户端安装在攻击者的主机 上是控制端,服务端安装在受害者的机器上。 }{100% : 木马可以使攻击者远程控制受害者的主机,造成受害者信息丢失等问题。 }{98% : 木马有很好的隐蔽性,通过模仿正常的系统文件命名、与其他程序绑定、进程注入及拦截系统调用的方法伪装自己。 }{100% : 木马也有很好的自启动性和自恢复性。 }{100% : 常见木马有远程访问型木马、键盘记录型木马、密码发送型木马、FTP 型木马以及破坏型木马等。 }

{100% : (3)蠕虫。蠕虫是一种可以独立运行、自我复制及自动传播的恶意程序。 }{100% : 它通过网络、共享文件、电子邮件、移动存储设备以及有漏洞的主机等自我复制和传播。 }{100% : 蠕虫的传播速度非常快,根据它的危害性可以简单分为无害型、消耗型和破坏型。 }{99% : 无害型蠕虫感染主机后会产生很多垃圾文件减少系统的可用空间;消耗型蠕虫感染主机后,发送大量扫描数据包,消耗主机的 CPU 和内存资源,与此同时增加了网络的负载,降低网络的性能;破坏型蠕虫感染主机后会 删除和破坏程序和文件,有时会泄露一些重要信息。 }

{97% : (4)后门。它是一种运行在目标系统中,能够绕过安全控制机制获得对系 统的访问权,为攻击者提供通道的恶意代码。 }{100% : 后门可以使攻击者远程控制目标主机,危害无穷。 }{100% : 后门提供的通道有几种类型:本地权限提升、远程命令行访问、单命令远程执行、远程控制等。 }

{96% : (5)Rootkit。它是指帮助攻击者获取主机管理权限后,实现维持拥有管理权限的程序[36]。 } {100% : 通常攻击者通过后门获取管理权限,并使用 Rootkit 维持管理权限使的恶意代码能隐藏在目标系统中。 }{100% : Rootkit 分为用户模式和内核模式。 }{100% : 用户模式通过通道插入恶意代码、覆盖文件、API 钩子和 DLL 注入等方式达到目的。 }{100% : 而内核模式通过安装恶意的设备驱动程序、打补丁、修改内存中运行的内核以及虚拟伪造系统的方式实现。 }

{97% : (6)间谍软件。它是在未授权的情况下窃取用户的信息并通过网络发送给 攻击者的一种恶意代码。 }{99% : 这种恶意代码不仅仅能泄露目标主机的数据信息,还 可以提供恶意代码的植入接口使得被侵系统受到更加严重的破坏。 }

{99% : (7)广告软件。它是指在未经用户授权的情况下和别的程序捆绑在一起, 以便经常弹出一些用户不想接受的广告。 }{98% : 这种恶意程序目的是通过这种强制的 方式做商业宣传。 }{100% : 一些广告插件的安装会降低主机的性能。 }{96% : 广告软件主要的危 害是弹出一些色情或者恶意的广告,这会给用户带来很大的困扰。 }

{94% : (8)恶意网页脚本。它是指在网页中嵌入一些用脚本语言编写的有恶意行 为的代码。 }{98% : 当用户点击带恶意脚本的网站后,脚本通过修改目标系统的注册表、 下载病毒或者加载木马程序等方式对被侵系统实施破坏行为。 }

2.2 恶意代码检测与反检测技术

2.2.1 恶意代码检测技术

{ 71% : 目前用于商业的恶意代码检测软件中,一般采用的都是基于“特征码”的检测技术,基本思想是,当新的恶意代码被发现后,对其进行采集取样,分析代码构成,提取有用的特征码,然后将新的特征码加入已有数据库中,用户更新病毒库之后,就会使用新的特征库去匹配恶意代码,如果匹配成功则进行相应的处理。 }{ 65% : 但是特征码检测技术的缺点是只能对已知恶意代码进行有效的将测,对与未知或者稍加变动的恶意代码无能为力。 }{100% : 因此,在恶意代码检测领域提出了启发式检测算法来预防和检测新的恶意代码。 }

{ 63% : 根据对恶意代码分析原理的不同对现有恶意代码检测方法进行分类,主要分为基于特征码的检测技术、基于行为的检测技术、基于启发式的检测技术、基于语义的检测方法和基于机器学习算法的检测技术等。 }{100% : 下面将详细介绍几种检测技术。 }

(1)基于特征码的检测技术

{ 77% : 基于特征码的检测方法是使用最古老和最广泛的方法。 }{80% : 被Symantec等多有著名病毒检测厂商所使用,是目前已知的所有恶意代码检测方法中最简单、开销最小的方法,广泛用于文件类型的病毒检测中。 }检测软件的核心就是恶意代码特征库的完全性,{ 60% : 当需要扫描某个程序是否有恶意企图时,启动特征扫描提取特征,然后再与特征库进行匹配,如果匹成功,则判断该程序是恶意的。 }{ 66% : 此技术的关键在于如何选取最能代表恶意程序的特征值。 }采用该方法,检验结果准确,鲜有误报情况,{ 58% : 但该方法对于未知或者变形恶意代码无能为力。 }还有,这种方法使得特征库不断增加,这需要用户经常更新特征库,{ 55% : 随着时间的流逝,特征库会越来越庞大, }这会影响检测的速度和系统的性能。

(2)基于行为的检测技术

{ 67% : 基于行为的检测方法是利用恶意代码的特有行为来检测恶意代码的方法。 }{ 73% : 恶意代码的行为有相对的稳定性和已于检测的特点,比如特定的系统调用,恶意代码要完成自身逻辑功能,即完成对系统的入侵和破坏,就必须获取系统非法权限,调用系统的资源[37],这样通过分析恶意代码的行为就可以方便的分析检测出恶意代码。 }{85% : 当程序运行时,监控其行为,如果发现了异常行为,则立即报警。 }{84% : 一般用于检测恶意代码的行为特征如下: }

{ 69% : 1) 对特定文件执行写操作:有些恶意代码时依附而生,所以在其执行时,就要将自身代码附加在感染文件中,可以监控是否有异常写操作。 }

{ 79% : 2) 监控系统调用序列:某些系统调用序列可以体现某种程度的程序语义。 }{ 65% : 系统调用是用户态和内核态的唯一接口, }恶意代码想要获取高级权限实施破坏行为,就必然要经过系统调用接口。

{ 61% : 3) 修改内存总量:恶意代码为了完成特定的恶意意图, }{ 62% : 经常会常驻在内存中,并且不能被覆盖, }{ 65% : 那么将会减少系统内存的总量, }使得该段内存不受系统内核控制。

(3)基于启发式的检测技术

{ 73% : 启发式检测方法是对恶意代码特征提前设定一个阈值,在对文件进行扫描后,当提取的特征和恶意代码特征的相似度达到一定的值,这认定该文件是恶意代码。 }{86% : 例如一些恶意代码都会固定的对一些内核函数进行调用, }{ 66% : 通常这些调用的顺序是有一定的规律,因此利用对内核函数的名称和调用次数进行分析,可以构建一个恶意代码对内核函数的特征。 }{81% : 启发式方法属于主动防御技术,对未知的恶意代码检测具有明显的效果,因此,这种方法在现如今的商业开发被重点应用。 }启发式检测可分为静态启发和动态启发。

{ 76% : 静态启发方法其实是对传统的特征识别方法的一种扩展,通过分析程序对系统API的调用序列作为特征,有领域专家根据自身经验,研究总结出某些恶意代码的行为特征,当对行为进行监控时,此类特征一旦被发现,就立即报警并做相应的处理。 }{96% : 这种方法能够有效的检测出已知的恶意代码,并发现部分未知的恶意代码,但在发现恶意代码的时候,系统往往已经被感染。 }{ 60% : 另外,行为检测是对系统进行实时的监控,因此可能会持续占用大量内存、CPU等系统资源。 }在商业领域中,该方法主要用于辅助性检测。

{ 73% : 动态启发式技术主要的工作原理是在计算机系统中划分出一各独立的虚拟环境,当发现可疑程序时,并不立即停止,而是让其继续运行。 }{ 73% : “沙盒”技术就是动态启发式技术的一种,沙盒会对可疑程序的行为进行记录,直到恶意代码完全暴露后,它在执行回滚操作,使计算机恢复到执行可疑程序之前的状态。 }{ 67% : 近年来病毒检测厂商已经将沙盒技术应用与商业的查杀工具中并进行了推广。 }

(4)基于语义的检测技术

{83% : 基于语义的检测技术是现在研究的热点。 }因为 混淆技术只是通过插入垃圾指令、改变指令顺序及寄存器重新分配等方法来改变程序,但是程序的基本语义是等价的。通过分析恶意程序,抽象程序指令的行为并建立其行为模型,使得该模型既描述恶意程序的基本行为,又具有很强的泛化能力。这样因其有很强的泛化能力,使得检测恶意程序的变种更加方便快捷。{ 60% : 除此之外,也可以检测未知类型的恶意代码。 }现阶段基于语义的检测方法分为基于内存和函数调用的方法。{ 60% : M.Christodorescu[8][9]提出了一套抽象理论和语义框架,使用自动机描述程序的行为, }{ 56% : 通过抽象理论描述程序的行为建立抽象模式库,并将其作为自动机的符号表, }最终经恶意行为描述为自动机表述的模板,{ 64% : 最后通过模型检测方法检测样本是否含有恶意行为。 }{81% : 模型检测是通过 遍历系统所有状态空间, }看其中是否有一条符合的路径状态。{ 62% : 之后,他还提出迹语义这一概念,将迹语义作为程序的基本语义,并定义了等价的条件,通过 抽象解释的方法给出了近似的检测算法。 }抽象解释理论为解决不可判定和复杂 问题的逼近求解提供可很好的构造方

法。{ 55% : 基于函数调用的方法是将程序中使用的函数提取出来, }并结合程序的控制流程图,通过图的同构、线性时态逻辑、计算逻辑树、有穷状态机及下推自动机等方法描述恶意行为,最终通过模型检 测完成恶意代码的检测。

(5)基于机器学习的检测技术

{80% : 基于机器学习和数据挖掘的检测方法。 }{ 62% : 随着检测技术的不断发展,机器学习和数据挖掘的方法已经被开始应用在恶意代码检测的领域。 }主要应用分类、关联规则挖掘、序列模式分析以及聚类等几种技术。主要思想是利用数据挖掘技术从现有的数据中挖掘一些有意义的模式,用机器学习技术归纳出已有样本的特征,然后根据特征的相似性等完成分类的任务。{ 57% : 其中,最主要的是选择好的特征和有效的分类器。 }检测步骤如下:首先,要分析样本确定提取哪种特征或者特征序列;其次,根据特征的特性选择合适的特征选择方法从所有提取的特征中选择一些分类效果好的特征;最后,{ 61% : 根据实际情况选择较好的模型实现分类。 }

2.2.2 恶意代码反检测技术

{ 76% : 恶意代码检测与反检测技术总是相互促进, }相辅相成。{ 65% : 检测技术的进步也带动了反检测技术的发展, }{ 60% : 当前出现了各种各样的恶意代码反检测技术, }现总结如下:

(1)加壳技术

恶意代码作者为了防止自己的程序被检测软件发现,利用一些软件技术给恶意代码加外壳,可以是利用算法将自己伪装成正常程序,或者利用特殊的算法将自己压缩或加密,使得检测软件很难检测。但是这些“壳”都有一个特点就是,{ 62% : 他们先于程序获得执行控制权, }{ 62% : 然后把伪装后的程序还原,再把执行权交还给原始代码: }是一类自修改代码。

(2)反虚拟执行技术

不可否认,虚拟执行的系统和真实系统或多或少存在差异[41]。比如,硬件上,调试器总是会设置硬件断点,而虚拟机总是在模拟硬件,这和真实的硬件是有差别的;执行环境,内核地址空间,对于虚拟机和真正的机器是不同的,还有调试器必须挂靠某些进程来插桩进程用于监控;应用程序,{ 56% : 虚拟机和调试器都有外部应用程序, }对进程可见,用于检查运行环境。一些指令在虚拟机环境中,执行时间总是远远长于真实环境,一个经常执行此类指令的程序能够指示它在虚拟机环境中运行。

(3)代码迷惑技术

恶意代码迷惑技术是指通过某种程序代码变换,改变自身在空间和时间上的结构,但是完成相同的逻辑功能。恶意代码在进行迷惑处理之后,使得逆向工程分析变得难以进行。迷惑技术本身是一种保护软件的手段,但是常常被用来对抗分析和检测。恶意代码的迷惑技术可以有效的对抗恶意代码的静态分析技术和动态反汇编技术。目前主要有基于加密的迷惑技术和基于代码变换的迷惑术。其中代码变换主要指在源程序中,利用等价指令替换、指令位置交换、添加新指令等手段改变程序形式,但逻辑功能保持不变。

2.3 恶意代码分析技术

恶意代码分析是确定恶意代码意图的过程,是实行恶意代码检测的必要前提。恶意代码分析的直接结果是用于实现恶意行为建模的元数据信息,{ 60% : 如指令流、API调用序列等, }{ 59% : 为后续恶意代码的检测工作提供必要的支持。 }

{ 72% : 恶意代码的分析技术一般可分为静态分析和动态分析。 }

(1)静态分析技术

{ 72% : 静态分析技术是指对被测软件的源程序或者二进制码进行扫描,从语法、语义的层面去理解程序的行为,以期望获取程序在运行过程中的信息, }而不需要运行程序。

要进行恶意代码的静态分析,{ 71% : 首先需要对恶意程序进行反汇编, }{ 55% : 常用的反汇编工具有:W32DASM、objdump、PEid、HIEW、IDA Pro等。 }

{ 56% : 静态分析技术由于不会运行程序, }因此不会对计算机系统造成任何伤害,{ 57% : 其分析效率相对动态分析而言较高, }同时由于静态分析技术从程序本身入手,因此可以获得程序的全部信息,分析结果较为全面。但是由于静态分析技术的前提条件是对程序进行正确的反汇编,现如今很多恶意代码编写者常常会对恶意代码进行加壳、加密或者压缩使得恶意程序很难被正确的反汇编。总之,{ 55% : 如果恶意代码无法被正确的反汇编, }那么静态分析将会失效。

(2)动态分析技术

动态分析技术是指在可控环境下实际运行程序,监控执行过程中的程序行为,记录程序执行的信息。{ 65% : 由

于动态分析需要先运行程序,}{ 61% : 所以为了防止恶意代码对当前环境的破坏,}系统在普或者恶意代码相关信息之后,会自动恢复到恶意代码执行前的最初状态,防止影响下一次的分析结果,但是动态分析技术能够获得恶意代码执行时的真实信息,可以有效地解决静态分析中譬如加壳、加密的干扰。

当前最流行的动态分析技术是动态污点分析技术[38],它的基本原理是将一切不信任的外部数据标记为污点,{ 57% : 然后跟踪标记为污点的数据的传播情况,}并记录相关的系统调用或者指令执行等相关信息,然后以此信息进行检测。动态污点分析能够记录恶意代码更细粒度的精确特征,{ 57% : 是当前非常热门的恶意代码检测技术。 }

动态分析技术也存在缺陷,比如开销大,一次只能分析一条路径,恶法应对恶意代码存在多路径的问题,同时由于恶法模拟出一个完全真实的计算机环境,{ 64% : 对某些环境敏感的恶意代码无法进行有效的检测,}因为恶意代码能够检测到虚拟机或者仿真机存在的情况,从而隐藏自身的真实行为,也无法知道某些恶意代码何时才会触发,动态分析技术也会受到行为层的混淆技术的干扰[39]如等价行为替换、模拟序列或者混淆序列等。于是有学者开始尝试静态分析与动态分析相结合[40]的方式进行恶意代码的检测,充分利用两种分析技术的优点。

两种分析技术各有优缺点,静态分析技术开销小,关注的是恶意程序本身的语法或者结构特征,动态分析技术开销大,{ 60% : 关注的是恶意代码的行为特征,}各有侧重点。静态分析技术分析全面,可获得恶意代码的全部信息,但获取特征的方式一般都是无导向的,因此可能包含大量无用信息,也易受代码迷惑技术的影响,{ 56% : 动态分析技术能够获得程序的真实行为信息,}但一次只能获得一种行为并且与当前的检测环境相关,信息不够全面。当前应用最为广泛的技术还是静态分析技术。总之,{ 65% : 无论是静态分析还是动态分析,}都需要借助恶意代码分析技术和监控技术获得恶意代码的基本属性和执行信息,{ 57% : 以便深入理解恶意代码的功能,}{ 76% : 进一步实现恶意代码的检测和抑制。 }

2.4 本章小结

{ 74% : 本章首先介绍了恶意代码的定义以及恶意代码的分类,}具体介绍了病毒、特洛伊木马、蠕虫等恶意代码的特征和危害。其次,{ 57% : 介绍了恶意代码的检测技术,详细阐述了基于特征码的检测技术、基于行为的检测技术、基于启发式的检测技术、基于语义的检测技术和基于机器学习的检测技术。 }最后,针对目前主流的恶意代码反检测手段以及分析技术做了详细说明。