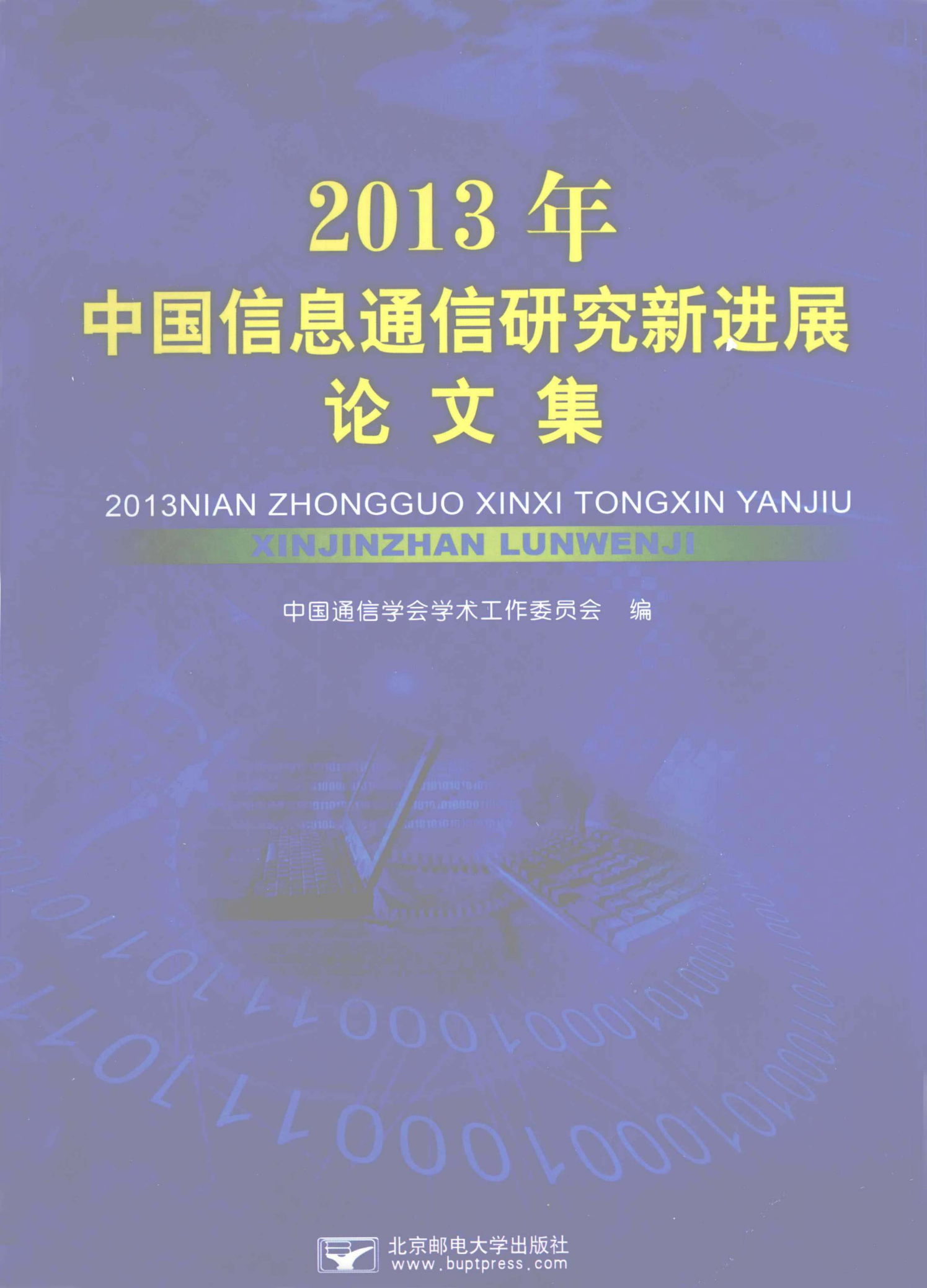
2013年
中国信息通信研究新进展
论文集

2013NIAN ZHONGGUO XINXI TONGXIN YANJIU
XINJINZHAN LUNWENJI

中国通信学会学术工作委员会　编



北京邮电大学出版社
www.buptpress.com

2013年中国信息通信研究新进展论文集

中国通信学会学术工作委员会　编

北京邮电大学出版社
www. buptpress. com

内 容 简 介

本论文集收录论文 73 篇，内容涉及计算机科学与技术、网络理论与技术、网络与信息安全、信号与信息处理等五大类，对通信技术、通信安全、物联网技术和信息安全等学科热点问题的最新研究进展和发展趋势开展深入的学术交流和探讨。

本论文集可供通信、计算机、信息技术、信息安全等领域的科技工作者和高等院校相关专业的师生参考。

图书在版编目(CIP)数据

2013 年中国信息通信研究新进展论文集/中国通信学会学术工作委员会编. --北京：北京邮电大学出版社，2014.4

ISBN 978-7-5635-3795-2

Ⅰ. ①2… Ⅱ. ①中… Ⅲ. ①信息技术—文集②通信技术—文集 Ⅳ. ①G 202-53②TN91-53

中国版本图书馆 CIP 数据核字(2013)第 296827 号

书　　名：2013 年中国信息通信研究新进展论文集

作　　者：中国通信学会学术工作委员会

责任编辑：艾莉莎

出版发行：北京邮电大学出版社

社　　址：北京市海淀区西土城路 10 号（邮编：100876）

发 行 部：电话: 010-62282185　传真: 010-62283578

E-mail：publish@bupt.edu.cn

经　　销：各地新华书店

印　　刷：北京联兴华印刷厂

开　　本：889 mm×1 194 mm　1/16

印　　张：22.75

字　　数：748 千字

印　　数：1—500 册

版　　次：2014 年 4 月第 1 版　2014 年 4 月第 1 次印刷

ISBN 978-7-5635-3795-2　　定　价：88.00 元

前　言

信息通信技术正越来越多地融入人们的日常生活。盒装的核心电子设备（计算机、智能手机、平板电脑等）越来越多地嵌入到日常物品，而这些物品往往是与互联网连接，使“物联网”成为“互联网的一切”。这个新的革命将使“智能”的地球（如智能城市、智能保健、智能电网、智能家居、智能交通和智能购物等）成为现实，提供更高质量、更具可持续性的社会前景。这也为企业发展提供了新契机。

本论文集以“信息通信技术新技术”为主题，收到了来自电信运营商和设备制造商的科技工作者，大学及科研院所的专家教授、科研人员、研究生，高科技企业的科技工作者，政府工作人员等的大量投稿。论文范围涉及到通信领域的各个方面，学术工作委员会组织专家对收到的论文进行了严格评审。

我们衷心感谢所有论文作者付出的辛勤劳动，感谢论文评审者对论文集的贡献，感谢通信学会领导和学术工作委员会对论文集出版的关心与支持，感谢北京邮电大学出版社对论文集的出版给予的大力支持。最后，我们向所有的领导和专家表示衷心的感谢！

由于时间仓促，水平有限，不足之处在所难免，欢迎批评指正。

中国通信学会学术工作委员会

目　录

面向 IPv6 网络的攻击图方法研究与实现 …… 张　腾，郭燕慧/1
一种针对流水线任务的云计算模型基于 MapReduce 的改进 …… 郑宇瀚，郭燕慧/9
基于行为分析的恶意代码潜伏性评估方法 …… 张　芮，胡　影，郑康锋/16
App Store 应用信息自动化采集系统设计与实现 …… 徐晓东，郭燕慧/20
基于层次分析法的移动互联网安全风险评估研究 …… 蔡建强，张　淼/26
一种基于黑盒评估的 iOS 平台应用安全评测方法研究 …… 严　炜，郭燕慧/34
加密云存储中文件共享关键技术分析 …… 张再东，肖　达/44
基于 HOTP 的移动终端用户认证机制研究 …… 韩　超，郭燕慧/50
基于卫星云图的风矢场（云导风）度量模型与算法探讨 …… 门方龙，白　健，王　坤/56
有限域上幂函数 S 盒构造及性质研究 …… 周　旋，王秋艳，端木庆峰，瞿成勤/61
基于波特五力模型的 4G 时代电信业经营策略分析 …… 张永辉，胡万里/67
基于 Kinect 的人体三维建模与尺寸测量技术的研究 …… 杨一飞，陈启军/73
铁路灾难恢复系统建设方案的研究 …… 周泽岩，马超群，于　洋，姚洪磊/78
基于资源优化下 FPGA 加密认证系统设计与实现 …… 刘　欢，赵　霞，何亚波，俞沁璐/83
基于物联网的智能家居监控系统的设计与实现 …… 蒋宇哲，连世兴，刘星成/90
iOS 平台上任务调度模块的设计与实现 …… 赵　辰，黄　玮，范文庆/96
移动终端通讯录数据同步去重算法 …… 吴朋朋，黄　玮，杨璐皓/100
基于 Kad 算法的 P2P-VoIP 系统的设计与实现 …… 尹　乐，马　严，吴　军/104
Numerical Simulation Method of Atmospheric Optical Communication Based on the “Frozen Turbulence” Hypothesis …… YU Song, JIANG Long, CHEN Zhixiao/108
一种基于逻辑子域的大规模网络攻击图生成方法 …… 赵　青，胡　影，戴方芳/113
CPG’s Parameters and Topology Co-Evolution for Walking Control of Biped Robots …… XIAO Hui, CHEN Qijun, LIU Chengju/120
基于拓扑图划分的仿真节点映射策略研究 …… 刘百川，郭燕慧/127
基于 RFID 的电子溯源卡券防伪系统设计与实现 …… 郭　俊，王艳艳，张　峻，陈锦华，赵璐琪/132
中国电信智能管道技术实现和演进 …… 贾聿庸，袁　博，范　亮/136
向 SDN 架构发展的电信宽带网络 …… 袁　博，范　亮，贾聿庸/142
Local Path Planning Method for Autonomous Driving in Structured Road …… WANG Zhuping, LI Hui, SHI Xingxin/148
Control System Designed for Home Service Robot based on FSM …… YE Yilong, CHEN Qijun/153
Design of a Real Time Mobile Home Monitoring System …… LI Xin, CHEN Qijun/159
FlexRay 网络的可扩展性解决方案 …… 赵金辉，韩　岗，朱耀铠/164
直放站性能优化专项研究 …… 魏俊松/169
Mobility Management in Internet of Things …… QIAN Linglong, SU Cui, LV Yuan, CHAI Rong/174
4G 技术应用研究综述 …… 刘　夏/184
基于 ZigBee 无线传感网络电压信息传输系统设计 …… 白文乐，吴晓旭，姜武希，穆　硕，李中仁/189
水电工程中的应急通信概述 …… 郑　迪，刘小飞/193
CCSK 软扩频系统抗干扰性能的分析与仿真 …… 李红领，霍景河，吴金亮，杨健康/197

野战通信枢纽避雷保护计算与雷电防护分析 …… 陈志元，孙玉铭/201
基于 NACA0012 翼型结构网格的 CFD 并行模拟 …… 廉　波，王正华，易晓山/206
基于 SyncML 协议的数据库同步安全传输技术 …… 罗　伟，蔡开裕/211
基于业务风险流的电子支付风险评估方法及其应用 …… 王　婷，郭燕慧/217
研究性教学模式在卓越工程师培养中的探索 …… 冀　杰，唐　超，彭　和/224
Small Cell 典型信令开销分析 …… 覃华忠，李　馨，刘　震/229
TD-LTE 室内深度覆盖分析 …… 秦　伟，邓　单，付杰尉/232
基于动态调整的 IT 支撑系统服务执行引擎研究 …… 王海明，王　欣/236
Small Cell 室内系统容量分流分析 …… 刘　震，吴汉光，李　馨/241
Design of a Novel Compact Microstrip Antenna with Enhanced Gain …… YANG Jun, ZHANG Wei/245
大数据时代电信运营商文件系统新思考 …… 祁昊颖/249
一种 SaaS 模式下服务单元个数估算方法 …… 孙志丹，谢　钧，李志刚/252
基于 Scyther 的密钥建立协议设计 …… 陆思奇，王　磊，王　伟，程庆丰/256
一种面向隐私保护的 MIPv6 网络切换认证方案 …… 林　翔，尚　涛，刘建伟/261
校园网话务分析和模型研究 …… 王建锋/269
关于通信机房冷水系统管路优化的创新 …… 张伟中，颜小彬，施　斌/273
Research on Fragile Watermarking based on Image Authentication …… ZHANG Qing，WANG Ying/277
数字媒体技术人才培养模式探索 …… 何　薇/281
基于任务驱动的“3D 建模技术基础”课程教学改革 …… 范士喜/284
“游戏脚本及编程”课程教学总结及方法探讨 …… 程明智/287
Teaching Innovation with BlueJ in Java Programming for Multimedia Undergraduates …… ZHANG Yongbin, MU Dazhong/291
“网络数据库技术”课程教学探讨 …… 舒　后/296
非完全信息博弈游戏的实践研究 …… 梅　险，孙德强，杜世锦，王陈云，曾凡夫，臧朝蕊/299
“以项目为驱动，以学生为中心，以教师为主导”的数字媒体工作室教学模式探讨 …… 王　渊，郭　斌，汪　浩，丛艳华，刘清华，李　果/303
数媒人才培养创立特色鼓励个性化发展的实践研究 …… 孙述和，牛翠霞，王竹海，张黎红/306
地方高校数字媒体技术专业建设策略探析——以湖北民族学院数字媒体技术专业为例 …… 杨顺钰，李　军，陈自根，国桂环/310
再谈产学合作模式数字媒体专业人才培养 …… 矫桂娥，石　潇/314
坚持艺术与技术结合，培养影视技术复合型人才 …… 丁友东，王毅敏，杨卫英，黄东晋/318
如何有效提高主变差动保护校验效率——浅谈降压变压器稳态比率差动保护校验算法 …… 李振计，张宸铭/323
电网 GIS 建设与规范化管理 …… 夏传鲲，梅林常/327
4G 技术在智能电网中的应用 …… 梅林常，王献军/331
开关偷合事故的原因分析及防范措施 …… 李爱叶/335
一起主变铁芯多点接地故障处理分析 …… 靳建坤，马慧娟/338
浅谈如何减少 10 kV 配电网计划停电时间 …… 孙素平/341
提高事故处理过程中调度员的工作效率 …… 孙素平/344
浅析送电线路建设通道费用的控制 …… 仝章文/347
倡导离柜交费　提升服务品质 …… 王国红/351
“学习型”电力调度管理经验述谈 …… 宋晓磊，陈亚辉，杨　涛/354

Research and Implementation of Attack Graph Method Based on IPv6 Network

ZHANG Teng, GUO Yanhui

(Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract: With the development of IPv6 network, security problems have become increasingly prominent. Because of the difference of IPv4 network and IPv6 network, the safety evaluation methods for IPv4 network can not be the original effectively used in IPv6 network and the special security evaluation for IPv6 network is displayed in a way that is more rare. Through the research of vulnerable characteristics on IPv6 network, this page constructs attack pattern knowledge database on IPv6 network. Compared with the previous research of attack graph on IPv4 network, the attribute attack graph for IPv6 network is specifically put forword and also how to construct the model of IPv6 network attack graph and the core algorithm is studied. By the manner of using depth first search for mining vulnerability information, finding the path can reach the final goal. In this paper, the test was carried out on the generation of IPv6 network attack graph baesd on the campus network. Using the document analysis method can permeate the weakness of the IPv6 network information uploaded to the repository, and through using the Graphviz API and the knowledge database, using Myeclipse developed the attack graph generation mode, the test finally generated the attack graph on IPv6 network. During the experiment of the attack graph, the theory about IPv6 network attack graph has been verified, the blank of using attack mode and weakness knowledge database for automatically generating attack graph on IPv6 network has been solved. This page has provided the research direction about IPv6 network attack risk assessment method based on attack graph.

Key words: IPv6 network; Attack graph; Attack pattern knowledge database

面向 IPv6 网络的攻击图方法研究与实现

张　腾，郭燕慧

（北京邮电大学信息安全中心，北京 100876）

摘　要：随着 IPv6 网络的发展，安全问题日益突出。由于 IPv4 网络和 IPv6 网络的差异性，使得原本针对于 IPv4 网络的安全评估方式不能有效的用于 IPv6 网络，而专门针对 IPv6 网络的安全评估的展现方式却越发稀有，本文通过对 IPv6 网络脆弱性特征研究，构建了 IPv6 网络攻击模式知识库，经过对前人 IPv4 网络攻击图的方法研究对比，提出专门针对 IPv6 网络的属性攻击图，并研究了 IPv6 网络攻击图的构建模型，提出 IPv6 网络攻击图算法，采用深度优先搜索的方式对脆弱点信息进行挖掘，寻找路径已达到最终目标。本文最终以校园网的实际环境，对 IPv6 网络攻击图的生成进行了测试实验，使用文档解析的方式对渗透 IPv6 网络的弱点信息上传至知识库中，通过 Graphviz 的 API 和知识库数据的调用，使用 Myeclipse 开发了攻击图生成模式，最终加载数据生成了 IPv6 网络的攻击图。通过攻击图生成实验，使得 IPv6 网络攻击图理论研究得到了验证，弥补了国内有关 IPv6 网络攻击图通过攻击模式和弱点知识库自动化产生的空白，为以后基于 IPv6 网络攻击图的风险评估方法的研究提供了方向。

关键词：IPv6 网络；攻击图；攻击模式知识库

1 引　言

随着信息科技的发展，IPv4 网络已经满足不了全球的需要，加之 IPv4 本身设计的安全性缺陷，不可否认 IPv6 网络即将大面积覆盖甚至会替代 IPv4 网络。但是，由于某些老系统的无法更新化，IPv4 网络会与 IPv6 网络共存很长一段时间，进而出现了针对双栈系统的攻击方式以及专门利用 IPv6 网络协议进行攻击和威胁性分析的一些方案，同时，也由于 IPv6 网络和 IPv4 网络的区别，针对 IPv4 网络的脆弱性评估的方法已经不能再适用于 IPv6 网络，进而要求我们能够面向 IPv6 网络提出一种专门面向 IPv6 网络的攻击图。本文通过分析 IPv4 网络攻击图的优缺点，专门针对 IPv6 网络设计了基于属性攻击图的模型网络构建，弥补了在 IPV6 网络上人们进行网络脆弱性评估展示的不足，同时也克服了前人提出基于 IPv6 攻击树模型[8]的大规模建模的不足和绘制 IPv6 攻击网络的缺陷。

文章的第 1 部分介绍了攻击图的背景知识和相关介绍。第 2 部分详细阐述了 IPv6 网络攻击图的产生模型和相关算法。第 3 部分主要阐述 IPv6 攻击图与 IPv4 攻击图以及攻击树的区别。第 4 部分以具体实验的方式展示了基于攻击图产生模型的攻击图自动化生成的过程。第 5 部总结全文。

2 攻击图介绍

攻击图技术是一种基于模型的网络脆弱性评估方法，以图形的方式描述攻击者从攻击起点到达攻击目标的所有攻击路径的方法。它通过对目标网络建模，以攻击规则对攻击者建模，然后根据二者之间的相互作用关系产生攻击图，展示目标网络内各个脆弱性之间的关系、脆弱性与网络安全配置之间的关系。攻击图算法[4]通常分为状态攻击图和属性攻击图两类。状态攻击图显式地反映了攻击者所有的攻击轨迹，便于用户理解，但是由于状态攻击图的每个节点代表系统全局状态，因此创建状态攻击图的效率不高。与状态攻击图相比，属性攻击图更加简洁，便于分析攻击产生原因。属性攻击图是攻击图的一种，它的节点分为两类：一类节点表示原子攻击，另一类节点为属性节点，它表示这些原子攻击的每个前提条件或后果。原子攻击节点与属性节点间存在前提边和后果边。所有通过前提边与原子攻击节点相连的属性节点都满足时，该原子攻击才可被执行，从而使通过结果边与该原子攻击相连的属性都被满足。属性攻击图定义了网络中的安全属性在脆弱性的作用下的相互关系，同时也反映出网络中脆弱性之间的相互关系。属性攻击图可以展现出攻击者在攻击过程中依赖和改变的网络安全属性。

3 基于 IPv6 攻击图的实现

3.1 IPv6 攻击图生成模型

虽然 IPv6 网络的发展越来越迅速，但是 IPv6 网络下的脆弱性评估方案日益缺乏，本文在设计风险评估方案的过程中研发了面向 IPv6 网络的属性攻击图的方式弥补攻击方式展示的不足。以下是攻击图的方案步骤：（1）IPv6 网络整体环境的建模；（2）IPv6 网络攻击者和攻击模式建模；（3）构建 IPv6 网络属性攻击图算法；（4）进行 IPv6 网络攻击图自动化绘制；（5）传输攻击图至风险评估方案中；具体过程如图一所示。

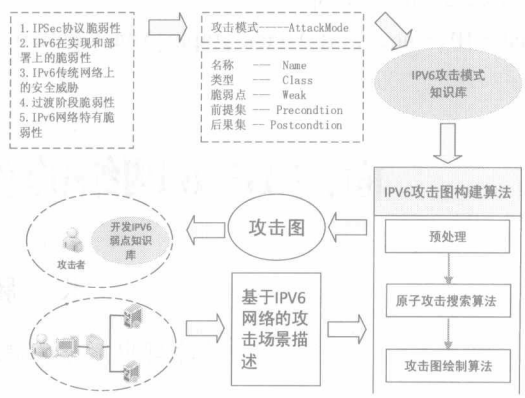


图 1　攻击图生成模型

3.2 攻击图算法与生成

3.2.1 攻击图构建阶段

IPv6 攻击图构建由三个步骤组成：（1）预处理，首先进行 IPv6 网络的整体环境建模以及攻击者建模，再从 IPv6 弱点知识库中以及攻击模式知识库中找出本次攻击的对于编号，通过模式化语言实现所有攻击模式模板化处理，对输入的攻击建模和环境建模的数据进行整理。（2）IPv6 网络原子攻击搜索算法，它

从攻击者初始攻击能力出发，计算攻击者进行多步骤多层次 IPv6 网络组合攻击过程中所能实施的所有原子攻击，通过逆向攻击搜索的方式，能够找寻到达到最终攻击目标的所有有效路径，并显示所能获得的最终攻击能力和效果。（3）攻击图绘制算法，它根据攻击者所能实施的所有原子攻击，使用 Graphviz 开发的 API，调用当前攻击产生的真实弱点数据，构造针对 IPv6 网络攻击图的程序化实现代码并进行数据赋值，绘制多目标攻击图，展示其详细的多步骤网络组合攻击过程及其攻击能力的扩张过程。

3.2.2 IPv6 网络整体环境的建模

IPv6 网络环境建模首先要依靠自动化网络检测工具，构建大致网络拓扑结构图。根据 IPv6 与 IPv4 网络共存与否分为两种拓扑构建方案：

1.当 IPv4 和 IPv6 网络共存时，首先可以利用 nmap 扫描器对各网段进行扫描，该扫描器会通过拓扑发现模块自动构建整个网络结构，然后利用绿盟扫描器或者 nessus 扫描器，扫描出整个网络段中所有的主机和网络单元的弱点信息。

2.对于单独存在的 IPv6 网络，可以通过 THC-IPv6 攻击工具箱的 alive6 工具，它可识别一个 LAN 上的其他 IPv6 主机节点。当一个攻击者向一个 IPv6 多播地址发送一条探测消息时，这个工具能够找到默认网关的路由器以及在该 LAN 上的其余的主机。此外，也可以利用邻居缓存来发现整个网络段上的其他主机，因为邻居缓存是 IPv6 中 ARP 缓存的 IPv6 对应物。它包含其他邻接 IPv6 节点的 IPv6 地址到二层 MAC 地址的映射。如果一名攻击者远程得到 LAN 上一台主机的访问权限，就可看到邻居缓存，并用之开始对所列主机进行攻击。例如在一台 Cisco 路由器上，可使用 show ipv6 neighbors 命令显示邻居缓存。在找到网络段上所有存活主机的 IP 地址后，可以针对每个 IP 地址使用 nessus 进行弱点信息的扫描发现，最终无论 IPv4 和 IPv6 网络是否共存，都能够挖掘出所有的弱点信息。

接着对于发现的 IPv6 网络环境的主机脆弱信息，网络服务信息，网络连接情况，用户权限等等进行模式化的语言翻译，有利于攻击图的算法绘制和攻击模式的对比分析。

定义 1 主机脆弱性定义为 5 元组

< hostsrc,ipv6error,cnvd,service,frequency> 其中 hostsrc 表示存在该脆弱性的 IPv6 地址；ipv6error 表示是哪一类 IPv6 的问题导致的主机存在了脆弱性，cvnd 为该类脆弱性 IPv6 弱点知识库中对应的编号；service 表示与该脆弱性相关的服务；frequency 为该脆弱性出现的频率。

定义 2 网络服务定义为 4 元组

<hostsrc,service,protocol,port>其中 hostsrc 表示服务 service 所在主机的 IPv6 地址；protocol 和 port 分别为服务 service 所用的 IPSec 协议和端口。

定义 3 网络连接定义为 4 元组

<hostsrc,hostend,potocol,port>。该 4 元组表示主机 hostsrc 使用协议 protocol 在端口 port 和主机 hostend 进行通信。

定义 4 用户的权限定义为 2 元组

<hostsrc,privilege> ,表示在主机 hostsrc 上用户具有权限 privilege。不失一般性，将权限定义为三类：admin 、user 和 guest 表示用户的管理员权限，user 表示普通用户权限，guest 表示基本的访问权限。

3.2.3 IPv6 攻击者建模

攻击者建模包括 IPv6 弱点知识库集合和 IPv6 攻击模式知识库集合。

IPv6 弱点知识库依据国家弱点知识库构建方法，以 IPv6 五大弱点类型构造完整的弱点知识库，图 2 是 IPv6 弱点信息的定义的情况：

字段名	字段意义	类型	举例	备注
flaw_id	漏洞信息 ID	int(16)	0001	
flaw_name	漏洞名称	varchar(32)	IPv6 协议漏洞	
flaw_grade	漏洞等级	varchar(16)	高	
flaw_inforamation	漏洞信息	text	www.anzhi.com	
flaw_number	漏洞编号	varchar(32)	Zdxx001	
submit_confirm_id	漏洞验证 id	int(16)	242156	
category_id	漏洞类型 id	int(16)	124725	
Category	分类	varchar(32)	IPSec 协议弱点	
find_count	相同漏洞上传数	int(16)	15	
flaw_influence	漏洞造成影响	text	跨过鉴定	

图 2 弱点信息定义

IPv6 攻击模式知识库[7]通过分析 IPv6 渗透测试情况进行定义，面向 IPv6 渗透测试攻击模式包括传统渗透测试攻击模式和针对 IPv6 特有攻击模式：

（1）传统渗透测试攻击模式，例如外网层面的 SQL 注入攻击，跨站脚本攻击，数据包重放攻击，跨站点请求伪造攻击，文件上传漏洞攻击，点击劫持攻击，Web 框架安全，DDos 攻击，钓鱼等等，内网层面的如嗅探攻击，ARP 欺骗，网络共享扫描攻击，RPC 系统级漏洞利用，中间人攻击等等。

（2）IPv6 特有攻击模式[2]，包含 RH0 扩展首部的攻击，分段攻击，Smurf6 多播放大攻击，无状态地址自动配置问题，邻居发现的问题，重复地址检测问题，重定向问题等等。

通过构建当前被测 IPv6 网络的系统拓扑图，以及依据 IPv6 网络的弱点知识库集合和 IPv6 网络下的多种不同层次渗透测试的方案模式集合库，最终构建攻击模式知识库。图 3 是攻击模式知识库分类情况：

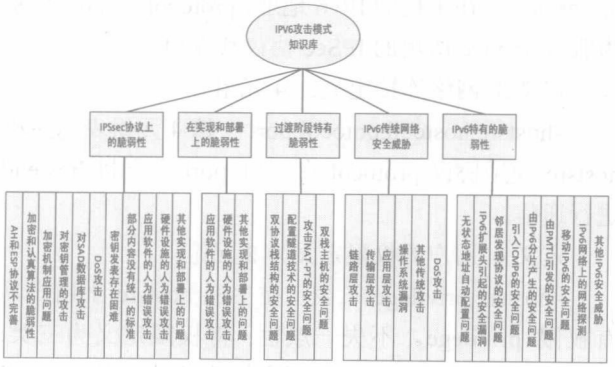


图 3　攻击模式分类

定义 5 IPv6 攻击模式定义为 5 元组

<name,precondition,result,ipv6error,cnvdset> 。其中，设定攻击模式为 name，其利用的条件为 precondition，通过攻击后产生的结果为 result，ipv6error 代表发现的 IPv6 弱点问题是属于攻击模式知识库中哪一种类型的， cnvdset 为 ipv6error 脆弱性下的哪一类 IPv6 漏洞信息的 cnvd 编号。

当攻击成功时，即是在 ipv6error 中某一个漏洞信息被成功利用后，要求利用的前提条件为 precondition，而 result 为攻击完成后最终产生的结果。我们用以下四个网络结构来定义一个完整的攻击模式 IPv6RunCodeAdmin：

Name: IPv6RunCodeAdminPrecondition:

<hostsrc,hostend,protocol,port>

<hostsrc,privilege>

<hostsrc,ipv6error,cnvd,service,frequency>

<hostsrc,service,protocol,port>

Result:

<hostend,admin>

作为有特征性的 IPv6 攻击模式，它主要是一种攻击的方法的自然语言化描述。可以得出，当其各个网络结构都存在而且完成满足时，那么可以设想一次攻击 IPv6RunCodeAdmin 就会发生了。当然，在发生后其产生的结果获得主机的权限或者是拿到主机网络的部分账号信息等等。当进行攻击时，可以设想出会同时发现多种漏洞信息，也可以设想出，当一个漏洞信息的产生，也可以由多种 IPv6 攻击方式进行得出的。在 IPv6 攻击模式中，可以做的是把我们普遍遇到的攻击问题和攻击的环境通过形式化的语言进行描述，从而简化攻击流程语言分析。

根据以上定义，可以定义出攻击图的设计方法如下：

定义 6 IPv6 攻击图 $IG = (H_p, H_d, H_e, E)$ 其中，H_p, H_d, H_e 为不会有交集的集合：H_p 为初始化的网络信息集合,对应 IPv6 网络整体网络情况；H_e 为利用节点集合，代表对探测到的 IPv6 网络脆弱点可以进行利用；H_e 为当经历网络攻击后所暴露出的漏洞信息和可以利用的条件；E 为攻击图将生成的图的有向边

$$E \subset ((H_p \cup H_d) \times H_e) \cup (H_e \cup H_d)。$$

3.2.4　基于深度优先反向搜索的 IPv6 网络原子攻击算法

在实际问题中，假设攻击所获得的弱点信息不会再成为第二次攻击的目标。经过分析，使用深度优先迭代算法[1]，在搜索攻击路径[9]时使用反向方式，并标明在利用攻击的每一个渗透的过程所经历的步骤数，最后产生 IPv6 攻击图。其算法核心思想为：首先要确定出要进行攻击的目标其中包含最终目标和到达最终目标所需的条件信息也可以作为攻击对象 ，然后参照 IPv6 攻击模式知识库，可以判断该攻击将会使用哪种攻击模式；最后查询通过 IPv6 渗透已经获得的攻击模式和网络中的脆弱性信息，把可以实现该网络条件的攻击模式实例化，并把在前一阶段渗透得出的结果作为新的渗透的前提条件信息，对攻击的目标进行多层先迭代攻击。采用反向的搜索是为了明确目标信息，对于最终攻击结果无关的漏洞信息不会进行处理，这样会使得所找到的漏洞信息更加具有针对性和高效性。

IPv6 攻击图的生成算法如下：

算法名：

GenerateIG = (*IG*, *I*, *ipv*6, *Step*, *MaxStep*, *ipv*6*error*)

Input：初始 IPv6 网络信息集合 I，攻击目标 $ipv6$，当前渗透攻击的步骤 $Step$，最大渗透步骤 $MaxStep$；

Output：IPv6 攻击图 IG 步骤

GenerateIG = (*IG*, *I*, *ipv*6, *Step*, *MaxStep*, *ipv*6*error*) {

1　*if*($ipv6I \in ipv6error$) {

2　$H_p \leftarrow H_p\{ipv6\}$;

return;}

3　*if*((*Step* + 1) > *MaxStep*)

```
        return;
4     else{
5       G_current ← {e | e ∈ G, ipv6 ∈ post(e)};
6       if(G_current = φ)
          return;
7       for each e G_current{
8       each G_current ← pre(e) ∩ Ī ∩ H̄_d;
9     if((G_current ≠ φ) ∧ (H_current ∩ H_ipv6 = φ){
10              for each n ∈ H_current
11  GenerateIG(IG, I, n, Step + 1, MaxStep);
      }
12          H_current ← pre(e) ∩ Ī ∩ H̄_d;
13          if(H_current ≠ φ) continue;
            else{
14              H_p ← H_p ∪ {I ∩ pre(e)};
15              H_d ← H_d ∪ {ipv6};
16              H_ipv6 ← H_ipv6 − {ipv6};
17              H_e ← H_e ∪ {e};
18              E ← E ∪ {< e, ipv6 >};
19          E ← E ∪ {< n, e > | n ∈ pre(e)};
                }
            }
          return;
      }
    }
```

算法说明如下：

首先判断如何能够达到最终攻击目标所采用的攻击模式是属于哪一种 $ipv6error$，而其渗透的最多步骤为 $MaxStep$。设定最多步骤作为最大迭代搜索的限制条件。集合 I 为经过 IPv6 网络扫描后得出的所有当前网络信息的集合。算法首先对能够到达最终目标的攻击方式类型进行判断，放弃最最终攻击结果不大的漏洞信息，在扫描的过程中为 IPv6 攻击图添加边和节点，最后以攻击图链表的方式生成 IPv6 攻击图。H_{ipv6} 设定为在攻击过程中所要完成而没有完成的攻击目标，H_{ipv6} 的作用不让攻击者进行攻击搜索的死循环，第 4 步是对攻击目标进行入队操作，而第 16 步则是攻击完成后目标信息的出队。$G_{current}$ 设定为目前进行的 IPv6 攻击所使用的攻击模式。$H_{current}$ 为当前进行渗透时发现所缺少的攻击条件，在第 11 步时通过递归方式调用，意味着渗透步骤加 1，然后在第 12 和 13 步判断经算法迭代后，这些进行攻击的条件是否都已经被满足了。

3.2.5 攻击图绘制

Graphviz 是一款开源的软件，由贝尔实验室设计。它采用图形构成算法，把设计在途中的信息进行均匀布局，并且能够把减少交叉的数目以及减少节点的边长。Graphviz 使用 dot 语言来编写其图形构成脚本，它通过图形布局器，对脚本的图形类别进行分类，通过类别和数量进行完美的自动化布局。经实验发现，目前可以使用 Myeclipse 可以加载 Graphviz 的 API 包，通过调用那些存储在 IPv6 检测数据库中刚完成的 IPv6 网络脆弱性情况，以及通过调用 IPv6 弱点知识库里的攻击模式相关数据，把这些数据加载到 Graphviz 的 API 包中，进而能够自动化的构建完整的攻击图[6]。

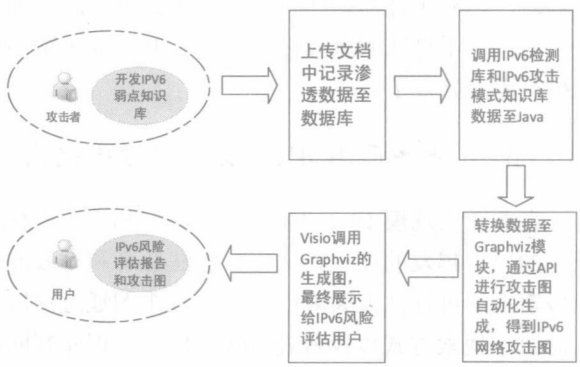


图 4　攻击图程序化实现

4　IPv6 攻击图与 IPv4 攻击图，IPv6 攻击图对比分析

由于 IPv6 网络的特殊性，以及针对 IPv6 进行渗透测试所挖掘的脆弱点的特殊性，挖掘与 IPv4 攻击图区别，同时基于前人提出的 IPv6 攻击树的方法，分析对比出 IPv6 网络攻击图的优点。

4.1 IPv6 攻击图与 IPv4 攻击图对比

1. 攻击者建模区别

在攻击模式上，IPv6 不仅包含了在 IPv4 上出现的传统弱点问题，还包含了 IPv6 特有的四大网络安全问题，IPSec 协议弱点，实现部署弱点，传统网络安全弱点，过渡阶段特有弱点等，为此，专门开发了 IPv6 弱点类型知识库，帮助建模。

2. 环境建模区别

由于 IPv4 的地址 32 位，IPv6 的地址 128 位[3]，因此在主机发现方面，IPv4 更易于 IPv6 建模，可以通过 nmap 实现大规模拓扑发现，而单独的 IPv6 网络只能通过邻居发现协议或者邻居缓存的方式搜索在线主机，进而再利用 nessus 进行各个主机脆弱点的信息挖掘。

3. 算法优点

IPv6 攻击图算法中其网络属性加载了有关 IPv6 弱点类型的属性说明，同时由于 IPv6 网络 IP 地址的非连贯特性，不能像 IPv4 地址属性构建时采用掩码聚合的方式缩减地址信息，本攻击图引入了关于 IPv6 地址特性和攻击方式的属性特点来联合构建算法的初始化的，弥补无地址缩减方式的不足，同时对无关系统攻击目标的一类攻击方式，会事先在初始化属性前进行判断，忽略不计与最终目标无关的弱点信息，因此大大的优化了算法的选择性执行，比 IPv4 攻击图的算法执行更加高效了。

4.2 IPv6 攻击图与 IPv6 攻击树对比优点

在面向大规模 IPv6 网络时，构建 IPv6 攻击树会让树的模型越发的庞大，进而在算法效率和有效的路径搜索上的可行性大大降低，IPv6 攻击图通过深度反向的原子搜索方式以及构建 IPv6 模型知识库判断的方式，大大降低非有效路径的搜索以及以图的方式循环搜索不会漏掉任何一个可靠的攻击路径，这两点都是目前 IPv6 攻击树没有实现的。

5 实验演示

5.1 环境搭建

以校园 IPv6 网络作为实例，在实验室搭建攻击环境，图 5 为校园搭建环境拓扑情况：

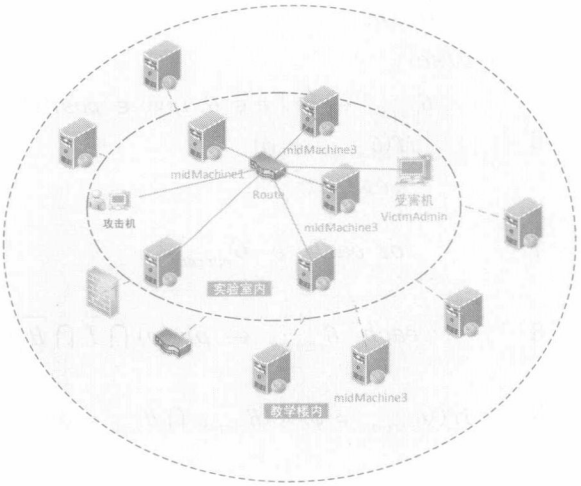


图 5 攻击图程序化实现

5.2 攻击图生成实验步骤：

步骤一：使用 J2EE 开发了 IPv6 弱点知识库模块，存储了大量了有关 IPv6 攻击模式和 IPv6 相关的弱点信息，并可以提供查询功能，数据都存储在后台数据库中，为生成攻击图[5]提供弱点信息。图 6 为开发完成的 IPv6 弱点知识库系统模块 Log：

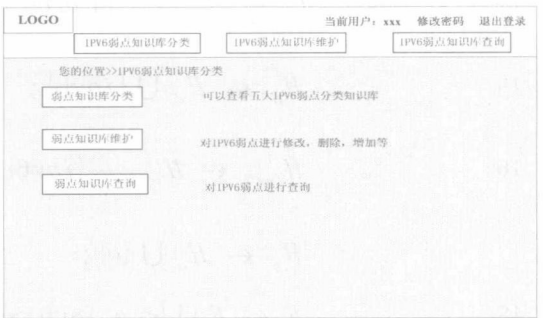


图 6 弱点知识库

图 7 为开发完成的 IPv6 弱点分类信息模块 Log：

图 7 弱点分类

步骤二：对系统进行渗透测试，查找关键问题点和 IPv6 网络主机信息，采用 THC-IPv6 黑客工具箱里

的 alive6 工具，进行邻居发现实验，因为是在实验室网络与教学楼网络相通，所以发现了 321 台主机，因此找寻到整个实验室网络中所有的主机的地址信息，图 8 为渗透发现的活跃 IPv6 网络主机截图。

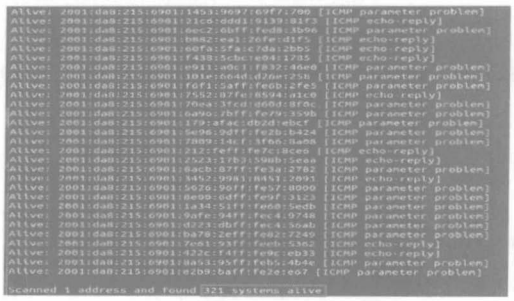


图 8　活跃主机

步骤三：面向 IPv6 的渗透测试过程，开发了 IPv6 渗透弱点文档解析模块，可以上传五大 IPv6 渗透测试过程文档，把本次检测的 IPv6 环境弱点信息和攻击者信息存储到 IPv6 弱点知识库和攻击模式知识库中。图 9 是开发的系统进行 IPv6 渗透测试结果报告的上传和解析模块。

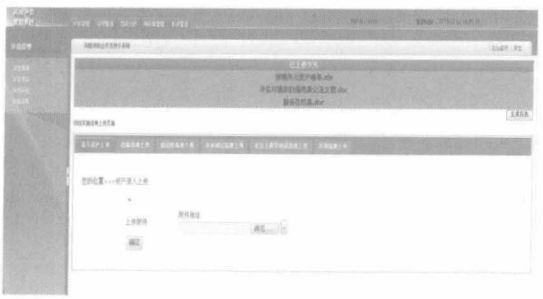


图 9　报告解析和上传

步骤四：从 IPv6 检测库和 IPv6 弱点知识库中调取本次渗透测试的信息，传输到以 Java 编写的攻击图生成模块中，通过 Graphviz 的 API 进行攻击图的自动化生成部分核心生成代码如下图 10 所示：

```
private void start()
{
  GraphViz gv = new GraphViz();
  gv.addln(gv.start_graph());
  gv.addln("size =\"10,10\";");
  gv.addln("Attacker [shape=box];");
  gv.addln("EnvironmentInit ->Attacker [style=dotted];");
  gv.addln("AttackInit ->Attacker [style=dotted];");
  gv.addln("WeakDataBase ->Attacker [style=dotted];");
  gv.addln("Attacker  -> VictmAdmin[weight=4 style=bold,label=\"VulExists RPC \n2001:db8:111:0:20c:29ff:feb8:7e30\"];");
  gv.addln("Attacker  -> midMachine1[label=\"VulExists neighbourDiscovery \n2001:db8:120:202:e3ff:fe11:4585\"];");
  gv.addln("Attacker  -> Route[label=\"VulExists RH0 \n2001:db8:111:0:202:e3ff:fe11:4585\"];");
  gv.addln("Attacker  -> midMachine2[label=\"VulExists IPSecError\n 2001:db8:111:0:250b:f64b:d943:8256\"];");
  gv.addln("Attacker  -> midMachine3[label=\"VulExists SqlInsert\n 2001:db8:111:0:250b:764b:d943:8226\"];");
  gv.addln("Attacker  string [label=\"make a\nstring\"]; ");
  gv.addln("midMachine1 -> VictmAdmin[style=dotted label=\"VulExists getData\"];");
  gv.addln("edge [color=red];");
  gv.addln("midMachine2 -> VictmAdmin[style=bold label=\"VulExists smurf6 \n2001:db8:111:0:20c:29ff:feb8:7e30\"];");
  gv.addln("midMachine3 -> VictmAdmin[style=bold label=\"VulExists WebShell \n2001:db8:111:0:20c:29ff:feb8:7e30\"];");
  gv.addln("Route-> VictmAdmin[style=bold label=\"VulExists sniffer \n2001:db8:111:0:20c:29ff:feb8:7e30\"];");
  gv.addln("}");
  System.out.println(gv.getDotSource());
  String type = "gif";
  File out = new File("d:/Program Files/graphviz-java-api\out." + type);   // Windows
  gv.writeGraphToFile( gv.getGraph( gv.getDotSource(), type ), out );
}
```

图 10　攻击图代码

最终由 myeclipse 自动化生成的攻击图展示，下图 11 为生成的攻击图：

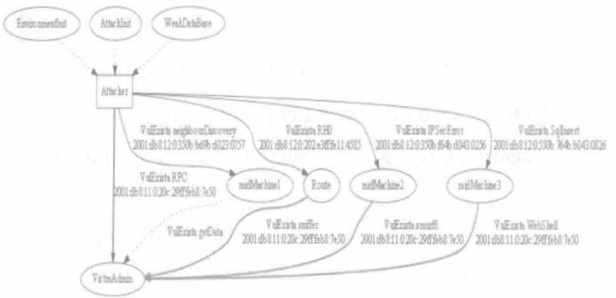


图 11　攻击图

攻击图说明：EnvironmentInit，AttachInit，WeakDataBase 分别是攻击者建模，IPv6 网络环境建模和 IPv6 弱点知识库信息导入。攻击者通过攻击模式知识库存储的攻击信息进行有效的原子攻击索引，通过基于深度优先的迭代算法去反向搜索攻击路径，得到能够进行完整攻击的攻击图。

攻击图实现红色实现代表能够达到最终目标的有效路线，黑色实线代表达到中间目标，黑色虚线表示达不到。midMachine1 代表邻居发现的方式，对最终目标不构成威胁，Route 代表路由北 RH0 攻击方式入侵，再以路由为节点进行嗅探最终目标，midMachine2 是由于 IPSec 协议缺陷被攻击，再利用 smurf6 攻击最终目标，midMachine2 是由于传统的 Sql 注入方式被攻击，再以此机为跳板，攻击加载在最终目标的 Web 服务器，上传 WebShell 最终获得权限。

6　结　论

本文通过对属性攻击图的研究，根据以往 IPv4 攻击图的比较分析，以 IPv6 网络脆弱性特征情况，设计了 IPv6 属性攻击图，其中包括核心搜索算法和 IPv6 攻击模式知识库的构建，并以实验的方式，使用 Graphviz 的代码自动化编辑出 IPv6 网络攻击图，解决了在目前 IPv6 网络研究中没有攻击图进行脆弱性展示的不足继。但是面对 IPv6 网络的风险评估工作还需要进行更深层次的研究，下一步的工作安排将放到利用 IPv6 网络攻击图如何有效的进行风险评估方法的计算当中去。

参 考 文 献

[1] Ammann P, Wijesekera D, Kaushik S. Scalable, graph-based network vulnerability analysis[C]//Proceedings of the 9th ACM Conference on Computer and Communications Security. ACM, 2002: 217-224.

[2] Zagar D, Grgic K. IPv6 security threats and possible solutions[C]//Automation Congress, 2006. WAC'06. World. IEEE, 2006: 1-7.

[3] Durdağı E, Buldu A. IPV4/IPV6 security and threat comparisons[J]. Procedia-Social and Behavioral Sciences, 2010, 2(2): 5285-5291.

[4] Jha S, Sheyner O, Wing J. Two formal analyses of attack graphs[C]//Computer Security Foundations Workshop, 2002. Proceedings. 15th IEEE. IEEE, 2002: 49-63.

[5] Ou X, Boyer W F, McQueen M A. A scalable approach to attack graph generation[C]//Proceedings of the 13th ACM conference on Computer and communications security. ACM, 2006: 336-345.

[6] Zhang T, Hu M Z, Li D, et al. An effective method to generate attack graph[C]//Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on. IEEE, 2005, 7: 3926-3931.

[7] 陈锋. 基于多目标攻击图的层次化网络安全风险评估方法研究[D]. 长沙: 国防科技大学, 2009.

[8] 苗延强. 面向 IPv6 的入侵检测系统实现技术研究[D]. 国防科学技术大学, 2009.

[9] 赵豹, 张怡, 孟源. 基于攻击模式的反向搜索攻击图生成算法[J]. 计算机工程与科学, 2009, 33(7): 18-24.

Cloud Computing Model for Pipelined Tasks Improvement of MapReduce

ZHENG Yuhan, GUO Yanhui

（Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China）

Abstract: With the development of cloud computing, more and more tasks rely on cloud computing platform. This paper analyzes shortcomings of Hadoop in pipelined tasks, which is a mainstream instrument of cloud computing. Based on the MapReduce model, Map-Reduce-Reload model is proposed to solve the problem. The effectiveness of the proposed Map-Reduce-Reload model in handling pipelined tasks is verified by the results of contrast experiment.

Key words: Cloud Computing; pipelined tasks; MapReduce; Map-Reduce-Reload.

一种针对流水线任务的云计算模型 基于 MapReduce 的改进

郑宇瀚，郭燕慧

（北京邮电大学信息安全中心，北京，中国，100876）

摘　要： 随着云计算发展，越来越多的任务依靠云计算来执行。本文针对当前云计算的主流工具 Hadoop，分析了其在执行流水线任务的缺点。将其 MapReduce 模型进行改进，提出了一种 Map-Reduce-Reload 模型，并通过实验将两模型进行了对比，结果表明基于 Map-Reduce-Reload 模型的云计算框架在处理流水线任务时的优越性。

关键词： 云计算；流水线任务；MapReduce；Map-Reduce-Reload

1 引　言

Hadoop[1]是基于 Google 提出的 MapReduce 并行编程模型开发的云计算框架平台，可有效实现对海量数据的并行处理。MapReduce 并行编程模型的最大优势是能够简单实现大规模并行计算，其屏蔽了底层实现细节，有效降低并行编程难度，提高编程效率，程序员可轻松地编写简单、高效的并行程序，用户可根据需要自定义 Map、Reduce 等函数，可使用 MapReduce 提供的接口和运行支持库高效管理任务调度、系统负载、容错等[2]。

随着大数据时代的到来，越来越多的数据分析和数据挖掘任务需要通过云计算技术来处理，其中尤其以基于 MapReduce 的 hadoop 云计算平台发展最为迅猛。需要用到云计算技术的任务的数量和类别越来越多，流水线任务便是其中一个。流水线任务是指可以分解为多个独立或者可迭代的任务的一种综合性任务。其应用场景比较多，例如常见的 PageRank 算法，单源最短路径问题。这些任务具有多步骤，不可变数据多、可重用数据等特点。

虽然 MapReduce 模型可以支持这些任务，但是这需要使用者在编程上花费很多的精力，即在这方面易用性较差。而且由于任务调度的原因，传统的 MapReduce 没有考虑到流水线任务中数据的可重用性，在执行流水线任务时效率降低。本文通过对这些缺陷的分析，对模型进行了改进，并提出了支持流水线任务的 Map-Reduce-Reload 模型，解决了易用性和效率的问题，最后通过实验测试证明了该模型的实际性能的优势。

本文第 1 部分是引言，第 2 部分介绍了流水线任务和 MapReduce 的相关技术特点。第 3 部分详细阐述了 MapReduce 在执行流水线任务时遇到的问题。第 4 部分提出了 Map-Reduce-Reload 模型框架，具体阐述了该模型的优点。第 5 部分以实验的方式对比了 2 个模型的实际性能。第 6 部分总结全文。

2 流水线任务与 MapReduce 模型

2.1 经典任务

在数据挖掘，信息检索等领域，有很多算法需要多个流水才能完成，或者要经过多次迭代直到约定值。例如两个很经典的算法，一个是 PageRank，另一个是 SSSP（Single Source Shortest Path）。

PageRank 是一个非常有名的网页重要性衡量因素，它是一个多次迭代的过程[3]，如图 1 所示，每次迭代，PageRank 由两个作业 MR1 和 MR2 完成，这样迭代多次，直到相邻的两次迭代中 PR 之差小于某一个阈值。

单源最短路径问题也是流水线任务的一种，包含多次迭代的过程，主要思想是：设 $G=(V,E)$是一个带权有向图，R 是 G 的邻接矩阵。整个算法始终把图中顶点集合 V 分成两组，第一组为已求出最短路径的顶点集合（用 S 表示，初始时 S 中只有一个源点，在每次迭代中求得一条最短路径，并将该路径的另一顶点加入到集合 S 中，直到全部顶点都加入到 S 中，算法就结束了），第二组为其余未确定最短路径的顶点集合（用 U 表示）。在每次迭代中，从 U 中选择一个当前路径最短的顶点，转存到 S 中，直到 U 为空。

初始值 R_0	
URL	RANK
www.a.com	1
www.b.com	1
www.c.com	1
www.d.com	1

	网页链接关系表 L	
	URL	OUT_LINK
	www.a.com	www.b.com
	www.a.com	www.c.com
	www.b.com	www.a.com
	www.b.com	www.c.com
	www.c.com	www.d.com
	www.d.com	www.b.com

(a)

$$MR_1\begin{cases} T_1 = R_i \bowtie_{url=url\_source} L \\ T_2 = \gamma_{url,rank,\frac{rank}{\text{COUNT}(url\_dest)} \to new\_rank}(T_1) \\ T_3 = T_2 \bowtie_{url=url\_source} L \end{cases}$$

$$MR_2\{\ R_{i+1} = \gamma_{url\_dest \to url,\text{SUM}(new\_rank) \to rank}(T_3)$$

(b)

图 1 PageRank 的两个作业

2.2 MapReduce 介绍

MapReduce 是一个云计算思想，也是一个编程模型，也是一个处理和生成超大数据集的算法模型的相关实现。用户首先创建一个 Map 函数处理一个基于 key/value pair 的数据集合，输出中间的基于 key/value pair 的数据集合；然后再创建一个 Reduce 函数用来合并所有的具有相同中间 key 值的中间 value 值[4]。现实世界中有很多满足上述处理模型的例子。

MapReduce 架构的程序能够在大量的普通配置的计算机上实现并行化处理。这个系统在运行时只关心：如何分割输入数据，在大量计算机组成的集群上的调度，集群中计算机的错误处理，管理集群中计算机之间必要的通信。采用 MapReduce 架构可以使那些没有并行计算和分布式处理系统开发经验的程序员有效利用分布式系统的丰富资源。

2.3 MapReduce 编程模型

MapReduce 编程模型的原理是：利用一个输入 key/value pair 集合来产生一个输出的 key/value pair 集合。MapReduce 库的用户用两个函数表达这个计算：Map 和 Reduce。其主要利用的分布式思想是数据分布式，将输入数据划分为等长的小数据块，称为输入分片，为每个分片构建一个 map 任务，该任务来运行用户自定义的 map 函数从而处理分片中的每条记录。最后再通过 reduce 任务，回收任务结果。由于基于数据分布式思想，所以其主要用对大数据的处理[5]。

用户自定义的 Map 函数接受一个输入的

key/value pair 值，然后产生一个中间 key/value pair 值的集合。MapReduce 库把所有具有相同中间 key 值的中间 value 值集合在一起后传递给 reduce 函数。

用户自定义的 Reduce 函数接受一个中间 key 的值和相关的一个 value 值的集合。Reduce 函数合并这些 value 值，形成一个较小的 value 值的集合。一般的，每次 Reduce 函数调用只产生 0 或 1 个输出 value 值。通常我们通过一个迭代器把中间 value 值提供给 Reduce 函数，这样我们就可以处理无法全部放入内存中的大量的 value 值的集合。

3 传统MapReduce模型执行流水线任务的缺点

传统的 MapReduce 框架把一个作业的执行过程分为两个阶段：map 和 reduce，在 map 阶段，每个 map task 读取一个 block，并调用 map()函数进行处理，然后将结果写到本地磁盘（注意，不是 HDFS）上；在 reduce 阶段，每个 reduce task 远程地从 map task 所在节点上读取数据，调用 reduce()函数进行数据处理，并将最终结果写到 HDFS。

以上过程如表 1 所示，从中可以看出，map 阶段和 reduce 阶段的结果均要写磁盘，这虽然可以提高可靠性，但降低系统性能。正是由于这个原因，传统的 MapReduce 不能显式地支持流水线编程，如果用户硬要在传统 MapReduce 上运行流水线作业，性能将非常低。

表 1　MapReduce 的 IO 操作

	Map 阶段	Reduce 阶段
读取来源	HDFS	本地存储
写入目标	本地存储	HDFS

除了 IO 缺点，还存在以下技术难点：

(1) 输入数据都动态数据和静态数据两部分组成。对于 PageRank，L 属于静态数据，而 R 属于动态数据；对于 SSSP，R 属于静态数据，S 和 U 属于动态数据。传输动态数据是不可避免的，而静态数据可以采用某种策略避免重复传输。怎样才能避免传输静态数据是一个需要解决的问题。就算没有静态数据，只有动态数据，假设 100W 行数据 3.22G，64M 的 split 有 52 个，每个 2W 行数据。由于是随机生成的，平均每行 500 个链接地址，每个连接地址都会生成一行临时结果<URL_ID AER_PR>，估算一下也有 150M（实际 140M），那么 3.22 数据，最后生成临时数据为 7G+。

(2) 每个流水线，是否需要继续 reduce 后再 map 再 reduce，这样效率会很低，如何提高效率。

(3) 每个流水线，如果所有 task 重复重新创建，代价将非常高。怎样重用 task 以提高效率是一个重要问题。hadoop 自身提供了 task JVM reuse 的功能。不过该功能仅限于同一个 job 内，而我们每个流水线都会重新运行一个 job，故自带功能不适用。

(4) 每次迭代，数据如何存储，如果总是写磁盘，代价将非常高。

(5) 何时迭代终止，怎样改变编程模型，允许用户指定合适终止迭代，以适合可变流水线任务。

4 改进的Map-Reduce-Reload模型框架

4.1 流水线任务建模

流水线任务，可以用有向图来表示。分为两类

(1) 不带有迭代的流水线任务，有向图中没有回路，是有向无环图 DAG(Directed Acyclic Graph)。

(2) 带有迭代的流水线任务，有向图中有回路。

对于第一类，任务可以表示为一个 DAG 图，$G=(T, E)$，其中 T 表示图中节点的集合，节点的权表示任务的处理时间，E 表示边的集合，边的权值表示数据依赖关系和任务间的通信时间。一个 DAG 的优先约束关系表现为一个节点不能在获得它的前序节点的所有信息之前执行,如果两个节点被分配到同一个处理器，则它们的通信成本视为零。

对于第二类，实际上可以转化为 DAG 图与环，在有向图中找出环路可以改进的深度优先遍历 DFS 算法来实现或者使用回路向量空间算法来实现[6]，此处省略。

由于 MapReduce 在对用户提交的任务请求进行划分后，生成的子任务有些存在顺序关系，有些不一定有顺序关系，因此改进的 MapReduce 模型设计目标是：先把子任务分类，分为循环类 C，顺序类 S，平行类 P。分别进入 Task Queue 模块，后让 Master 先执行存在顺序和循环关系的子任务，当子任务执行到与原来的子任务不存在偏序关系时，让 Master 进行多次 Map 过程。

4.2 MRR（Map-Reduce-Reload）模型

图 2 是 MRR 模型的主要技术架构，是基于 MapReduce 模型并针对流水线任务进行改进的云计