安全数据采集代理顽健部署策略研究

陈黎丽[1,2]，王震[2,3]，郭云川[2]，华佳烽[1,2]，姚宇超[1]，李凤华[1,2,4]

（1. 西安电子科技大学综合业务网络国家重点实验室，陕西 西安 710071；
2. 中国科学院信息工程研究所第五研究室，北京 100093；
3. 杭州电子科技大学网络空间安全学院，浙江 杭州 310018；
4. 中国科学院大学网络空间安全学院，北京 100049）

摘 要：随着“网络黑产”事件频繁发生，攻击者以“趋利”的思想来策略地发动针对性的攻击。现有网络监测系统缺少针对“策略式攻击”精准有效的监测策略。因此，在敌对环境中，如何优化部署采集代理获取更好的监测效果成为一个极为重要的课题。针对该问题，提出了一种顽健采集代理部署策略。首先，引入攻防博弈思想，对采集代理和威胁事件及其之间的关系进行度量，构建度量攻防博弈模型——MADG 模型；然后，考虑传统精确求解算法无法求解该问题，利用目标函数的次模和非增的性质设计了顽健采集代理部署算法——RCD 算法进行近似求解；最后，对 RCD 算法进行了验证。实验结果表明，所提模型和方法是可行有效的，且具有可扩展性。

关键词：采集代理；安全数据；攻防博弈；顽健性；优化部署

中图分类号：TP302

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2019121

Robust deployment strategy for security data collection agent

CHEN Lili[1,2], WANG Zhen[2,3], GUO Yunchuan[2], HUA Jiafeng[1,2], YAO Yuchao[1], LI Fenghua[1,2,4]

1. State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710071, China
2. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
3. School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
4. School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: With the frequent occurrence of “network black production” incidents, attackers strategically launch target attacks with the idea of “profit-seeking”. Existing network monitoring systems lack accurate and effective monitoring strategies for “strategic attacks”. Therefore, in an adversarial environment, how to optimize the deployment of collection agents for better monitoring results becomes an extremely important issue. Based on this, a robust deployment strategy of collection agents was proposed for the above mentioned problem. Firstly, the idea of attack-defense game was introduced to measure the collection agents, threat events and their relations, then the MADG model was built. Secondly, considering that the traditional accurate solution algorithm cannot solve the problem, the robust acquisition agent deployment algorithm called RCD algorithm was designed to approximate the problem by using the sub-module and non-growths of the objective function. Finally, the RCD algorithm was verified. The experimental results show that the above model and method is feasible, effective and expandable.

Key words: collection agent, security data, defender-attacker game theory, robust, deployment strategy

---

收稿日期：2018−12−21；**修回日期**：2019−03−24

通信作者：李凤华，lfh@iie.ac.cn

基金项目：国家重点研发计划基金资助项目（No.2016YFB0800700, No.2016YFB0800702）；国家自然科学基金资助项目（No.61672515）；中国科学院大学生创新实践训练计划基金资助项目

Foundation Items: The National Key Research and Development Project (No.2016YFB0800700, No.2016YFB0800702), The National Natural Science Foundation of China (No.61672515), Innovative Practice Project of College Students in Chinese Academy of Sciences

1 引言

近年来，“网络黑产”事件频繁发生，攻击者以“趋利”的思想策略地发动针对性的攻击。例如，2018 年 12 月，攻击者通过加密电脑上的 doc、jpg 等常用文件的勒索病毒，利用微信支付二维码勒索赎金，但支付宝用户并未受到影响。目前，大部分的 Internet 服务提供商与大型企业网络部署了网络监测系统[1-3]，系统管理员通过网络监测器（如流量分析程序、网络入侵防御系统和防火墙等）对网络性能数据进行实时收集，从而监测网络的性能和安全状况。虽然现有的网络监测系统可以收集到一些安全数据（如网络流量、CPU 占用率等），但是针对上述“策略式攻击”相关的安全数据缺少精准有效的监测策略。同时，由于受到资源成本的限制，只能部署有限数量的采集代理进行安全数据的监测采集。因此，在敌对环境中，如何优化部署采集代理获取更好的监测效果成为一个极为重要的课题。

在敌对环境中，优化部署采集代理问题面临着以下几个挑战。首先，采集代理针对不同类型设备采集的内容存在差异，例如，载有防火墙的设备可采集到 UFW、Snort 等相关日志信息，载有数据库服务器的设备可采集到 MySQL 相关日志信息。上述日志信息存在数据异构性，为数据解析威胁事件带来困难。因此，需要对异构性的安全数据进行统一度量。其次，上述的安全数据与威胁之间存在关联关系，不同的异构数据组合能监测到不同的威胁。因此，为了更精准地监测“策略式攻击”，需要考虑如何部署有限数量的采集代理，使采集到的异构数据组合获得尽可能好的监测效果。最后，采集代理一旦部署，无法在短时间内移除，攻击者通过漏洞扫描、网络渗透、社会工程学等手段获取网络系统的信息（如漏洞信息、防火墙的位置等），并且策略地选择对其“最有利”（“最有利”指的是使网络攻击影响最大）的方式实施攻击。因此，需要考虑如何优化部署采集代理，在应对敌对情况时监测效果具有顽健性。

针对监测点部署问题，现已有不少学者开展了相关研究。在网络测量领域中，网络测量部署模型主要以优化为主体思想，并在此基础上提出了一系列的启发式近似算法[4-7]，为本文的研究奠定了基础。网络测量系统主要测量链路流量或端到端带宽、时延、分组丢失率等性能参数[8]。然而，本文研究的采集代理部署在不同类型的设备上，安全数据类型不同，包括网络流量数据、网络日志数据和设备状态数据，这些数据具有很强的异构性。在环境监测领域中，现有工作主要解决如何放置有限数量的传感器来检测污染物的问题[9-11]。其中，文献[10]对敌对情况下如何部署采集器进行了研究，将对抗设置为对手在知道传感器位置的情况下选择在哪里进行污染，该设置与本文敌对情景设置相近。然而，环境监测中的采集项主要包括污染物类型、污染物质量和污染时间等相关参数[11]。而本文主要从网络威胁层面考虑，例如，威胁事件发生的概率、威胁事件对目标系统的影响等。最后，在网络安全监测领域中，已有不少采集代理的部署方法。例如，基于访问控制策略的部署方法[2]和基于线性规划的部署方法[12-14]。其中，文献[14]对异构性强的日志信息进行度量，利用日志与威胁之间的关联关系构建优化目标，并对该方法的伸缩性进行讨论，该度量方法为统一量化异构数据提供了启发。然而，本文研究的是在敌对情况中进行采集代理的部署，不仅要考虑日志数据、流量数据和设备状态数据这 3 类异构数据，而且还要考虑敌对环境中部署方法的顽健性，以上 3 个方面的相关研究无法直接解决本文的问题，因此，本文提出了一种适用于敌对情况下的采集代理优化部署算法——采集代理顽健部署（RCD, robust collection deployment）算法。本文的主要贡献如下。

1) 将敌对环境下采集代理优化部署问题归结为一个攻防博弈问题，并且考虑到采集代理获取的安全数据的异构性，构建度量攻防博弈（MADG，metric attack and defense game）模型。

2) 在 MADG 模型的基础上，利用系统、威胁和采集代理以及三者之间的关联关系构建威胁−采集树，利用威胁事件发生的可能性和威胁事件对系统的影响构建攻击效用函数。

3) 利用上述目标函数的次模性，提出了一种基于贪婪的采集代理顽健部署算法——RCD 算法，该算法能找到一组性能至少与最优集合一样的部署位置集合，但时间成本会稍微增加。

4) 通过构建具体网络案例模型和扩展模型，从求解时间和求解质量方面进行了一系列的实验，并与精确求解算法和启发式算法进行了对比，表明 RCD 算法的顽健性和可扩展性。

2　相关工作

首先，从 3 个角度对网络监测点部署的相关研究进行了梳理。其次，为了直观地描述敌对环境，借鉴了威胁建模中威胁树的思想，并对威胁建模的相关工作进行了梳理。

2.1　监测点部署

近年来，很多学者从不同角度对网络监测点部署问题开展了研究。首先，在网络测量领域中，主要研究工作集中在网络测量部署模型及其优化算法上。Aqil[3]把网络测量部署问题映射为经典优化问题，利用经典优化问题的难解性和近似算法进行求解。Breitbart 等[4]使用同样的映射，采用整数规划来解决优化问题。Hochbaum[5]则设计了一个启发式算法求解近似解。此外，Chaudet 等[7]对网络测量部署模型和优化算法进行了归纳总结。Suh 等[6]针对主动监测时部署信标分配问题和被动监测时部署 tap 设备的分配问题进行研究，提出了一个问题组合和高效通用混合整数规划（MIP, mixed integer programming）公式。上述工作侧重于优化问题的描述和求解，没有考虑采集项中是否存在异构性，且在敌对环境方面没有给出相关讨论。

其次，在环境监测领域中，针对放置有限数量的传感器来检测污染物的问题，Leskovec 等[9]最早将部署采集器监测水污染的问题归结为“爆发检测”问题，并提出了基于贪婪的优化部署算法——CELF（cost-effective lazy forward selection）算法。Comboul 等[11]考虑了水分配网中不确定参数对部署方案的影响，从确定系统、不确定系统和灵敏传感器 3 个方面对优化部署问题进行优化，利用目标函数次模性提出贪婪算法求解最优化问题。此外，Kraused 等[10]针对敌对环境下水分配网络中采集器优化部署的问题进行了研究，其中对抗设置为对手在知道传感器位置的情况下选择污染位置，并对此提出了一种具有顽健性的优化部署算法——Saturation 算法。以上研究虽然与本文的研究领域不同，但是解决的问题与本文的问题都是有限个数节点的优化部署问题。

在网络安全监测领域中，为了尽早发现复杂网络中病毒的传播，Yu 等[12]提出了一种最小化最坏的感染规模的启发式算法——MMI（minimizing the maximum infection）算法，该算法具有较快的收敛速度。Zhou 等[13]研究了如何有效地部署传感器位置，以便在网络中早期发现动态有害级联问题，并将级联的动态属性因素考虑在内，在 CELF 算法[9]的基础上，提出了 UBG（upper bound based greedy）算法，求出收益函数的上界，以便删除不必要的检测时间估计，同时为了提升计算速度，提出了 2 个加速算法。Thakore 等[14]先对异构的日志信息进行了度量，并通过日志信息与威胁事件的关系建立了映射关系，然后又在度量的基础上提出了一个启发式优化部署算法——GOMD（greedy algorithm to compute the optimal monitor deployment）算法。此外，Talele 等[2]针对采集已知攻击问题的局限性，提出了一种计算网络监控位置的方法，该方法利用主机间可用访问控制策略的通用性来计算大规模系统的采集代理的部署位置。上述这些工作均忽略了攻击者对采集代理部署策略的影响，因此现存的优化部署方案顽健性不足，无法很好地应对敌对环境。

2.2　威胁建模

威胁建模是针对攻击者如何执行潜在攻击或对系统构成安全威胁进行简化、抽象描述的过程。对威胁进行建模可以更直观地描述威胁和评估威胁影响。在威胁建模的工具中，威胁树是较常用的一种。Marback 等[15]提出了一种基于威胁模型的安全测试方法，该方法可以从威胁树中自动生成安全测试序列，并将其转换为可执行测试。Pardue 等[16]提出了一种基于威胁树和蒙特卡罗模拟的风险模型和风险评估技术，其目标是对直觉或风险估计进行合理而简洁的量化。Morikawa 等[17]提出了威胁树模板来帮助非专家分析人员构建威胁树，每个模板都是一个冗余的威胁树，装载了代表许多可能攻击场景的分支，以及针对此类攻击的相应漏洞和对策的典型示例。目前，上述工作均从攻击的角度来描述威胁和威胁实现的条件，忽略了威胁和安全数据之间的内在联系。

3　模型

在描述模型之前，为了不出现异议，将采集器、采集代理、网络监测器统一用“采集代理”表示，涉及采集到的“安全数据”用“采集项”表示。在对敌对环境中优化部署采集代理问题进行建模之前，对该问题进行以下 2 个基本假设。

假设 1　采集代理方面。每个采集代理的采集能力是相同的，部署在不同类型的设备上，不同类型的设备输出不同类型的数据，因此能够采集到的

采集项具有差异。

假设 2 威胁方面。攻击者是贪婪的，攻击者通过扫描、渗透、社会工程学等手段获取部署位置，从而选择对自己“最有利”的方式进行攻击，且不会进行无效攻击。

3.1 问题描述

本文所要优化的问题是如何优化部署采集代理获取更好的监测效果，并且可以保证部署的成本尽可能最少。本文将其建模为一个基于度量的零和博弈模型——MADG 模型：管理员作为防守方（defender），其对应的防守策略是在通信网络上选取 k 个设备进行采集代理的部署；攻击者作为攻击方（attacker），其攻击策略是选取所有攻击方式中的一种。由于是零和博弈，因此有

$$\text{Utility}_{\text{attacker}} = -\text{Utility}_{\text{defender}}$$

攻击方根据获取的网络相关信息选择攻击效用值最大的一种攻击方式，而防守方选择使攻击方的攻击策略收益最小化的策略集合，即攻击方的目的是最大化攻击效用，防守方的目的是最小化攻击方的攻击效用。因此，该问题的目标函数可表示为

$$\min_{\text{defender}}\max_{\text{attacker}}\text{Utility}_{\text{attacker}} \tag{1}$$

3.2 部署目标相关度量定义

为了对式(1)中的目标函数进行精确的分析，本节借鉴文献[14]中给出的度量框架的一部分，对本文目标函数中相关元素进行了定义和度量。本文的符号及其含义如表 1 所示。

表 1 符号含义

符号	含义
V	系统中所有的设备节点集合
v	系统中任意一个设备节点
Ψ	威胁事件集合
ψ	威胁事件
S	内嵌式采集代理集合
S_d	采集代理的部署集合
s_i	第 i 个采集代理
C	所有特征信标集合
ς	特征信标
τ	最小特征信标集合
$\varphi(s)$	采集代理和特征信标对应关系
$\gamma(\psi)$	检测威胁事件的证据
$\text{Conf}(\psi, S_d)$	置信度
Risk_ψ	威胁事件 ψ 的风险
P_ψ	威胁事件 ψ 发生概率
I_ψ	威胁事件 ψ 影响值

定义 1 威胁事件。已经造成攻击影响的事件和可能会对网络造成攻击影响的事件，即系统中存在的攻击或入侵行为，或者系统中可能出现的攻击或入侵行为。

本文使用 Ψ 表示能够在系统中检测到的和可能发生的所有威胁事件的集合。

威胁事件并非是采集代理直接采集的采集项条目，而是通过对一个或多个采集信息相关性的分析确定能够检测出的威胁。

定义 2 内嵌式采集代理。实现网络监测的采集组件和采集器的统称。采集组件是安装在操作系统（如 Windows 系统、Linux 系统等）中的软件，采集器是部署在终端（如手机终端、卫星终端等）上的传感器，两者都可以对部署的设备上的各种类型的数据进行收集，本文用 $S = \{s_1, s_2, \cdots, s_m\}$ 表示内嵌式采集代理 s_i ($1 \leqslant i \leqslant m$)的集合。

采集项可以分为 3 类，即日志数据、网络流量数据和设备状态数据。其中，日志数据可以分为以下 4 类：1) 主机上安装的 Windows 系统、Linux 系统等操作系统日志数据，2) 网络中部署的路由器、交换机等传输设备日志数据，3) 主机上记录的 SSH、MySQL、HTTP、Web 等具体服务运行日志数据，4) 安全防护系统防火墙、IDS 等安全设备日志数据。

定义 3 特征信标。通过采集代理生成的信息可以推断出系统中发生的威胁事件，在此用 ς 表示[14]。

特征信标并非采集代理采集到的实际数据，而是使用一些逻辑谓词对单个或多个采集代理获取的采集项进行连接和加工而生成的信息，是用来定义检测威胁的必要条件。特征信标的生成主要有 2 个步骤：1) 对于每一个威胁事件，将每个采集代理采集的采集项进行分组，根据它们提供的证据类型支持检测威胁事件；2) 根据采集项和威胁事件的内在联系，确定需要采集哪些信标、可以监测到哪些威胁事件。

文献[18-19]对逻辑规范的后续研究工作提供了一定的研究依据，但该问题不是本文的重点，在此不再赘述。此外，在现有的开源情报（如 OSINT）和入侵检测技术为关联关系映射提供了技术支持的同时，文献[20-23]针对不同类型网络中的威胁事件所对应的采集项进行了研究和调研，为该问题提供了一定的理论依据。

采用上述 2 个步骤，可以对采集到的异构数据

格式进行统一，避免了来自不同类型设备上的日志、流量、设备状态信息格式的差异，同时该方法可以在一定程度上简化安全领域专家对特征信标的枚举。尽管无法指定采集代理数据的确定格式，但是专家能够理解每条特征信标提供的语义信息。

定义 4　采集代理和特征信标对应关系。由一个映射函数表示$\varphi:s\to\Phi(C)$，即

$$\varphi(s)=\{\varsigma\in C\mid\text{由采集代理 } s \text{ 生成的特征信标}\ \varsigma\}$$

其中，$\Phi(C)$为 C 的幂集。现对该映射进行说明：映射是从采集代理到采集项一对多的关系，并表示由给定采集代理生成的一组采集项。考虑多维网络的异构性、设备差异性较大等因素，针对任何一个威胁事件，可能有多种检测方法，本文统一采用最小特征信标集合来监测分析威胁事件。

定义 5　最小特征信标集合。一组特征信标集合τ能够检测分析到一个威胁事件ψ，即所有最小特征信标集合中的元素足以监测到威胁事件ψ[14]。

定义 6　检测威胁事件的证据。一个给定映射$\gamma:\Psi\to\Phi(\Phi(C))$，其中，$\gamma(\psi)=(\tau\mid\tau$用于检测威胁事件$\psi)$[14]，$\gamma$是一个从威胁事件到特征信标的一对多的映射。在众多映射中，只要有一组特征信标监测到威胁事件即可。

定义 7　真实性。采集代理的真实性是指采集代理所产生的特征信标是正确的[14]。

采集代理的真实性不是二元的，因此本文应用模糊逻辑来评估采集代理的真实性，即采集代理可能不会从真实的信标转变为错误的信标，而可能会继续为少数威胁事件以外的其他威胁事件提供正确的信标。正如 Hosmer[24]所提出的，模糊逻辑比概率论更适合这种情况。因此，本文定义了一个函数σ_S，将采集代理映射到生成信标的真实性作为一个模糊逻辑度，表示由采集代理生成的信标中有多少是真实的，即

$$\sigma_S:S\to\{i\in\mathrm{R}\mid 0\leqslant i\leqslant 1\}$$

其中，R 为实数集合。

本文将采集代理的真实性映射到其产生的所有指标的真实性上。如果没有采集代理生成信标，默认情况下该采集代理的真实性为 0，即

$$\sigma_C(\varsigma,S_d)=\begin{cases}\max_{s\in S_d\mid\varsigma\in\varphi(s)}\sigma_S(s),\varepsilon(\psi,S_d)\\ 0,\ \text{其他}\end{cases}$$

为了监测一个威胁事件，至少需要采集来自该威胁事件对应最小信标集合中的一个特征信标。因此，对于一个最小的特征信标集合τ，本文结合所有最小特征信标集合τ的真实值来定义该最小特征信标集合τ的置信度。最后，给出监测到的威胁事件Ψ的置信度的定义为：该威胁事件Ψ对应的所有最小特征信标集合的置信度的 MTL 分离值即为Ψ的置信度。

定义 8　置信度。置信度定义为[14]

$$\mathrm{Conf}(\psi,S_d)=\max_{\delta\in\varphi(s)}\min_{\varsigma\in\delta}\sigma_C(\varsigma,S_d)\qquad(2)$$

本文考虑了攻击者在观察到采集代理的部署位置时，会选择破坏效果最大的威胁事件进行攻击，换言之，采集代理部署位置的集合，决定了攻击者的攻击策略。

为了直观地对敌对环境进行描述，本文借鉴威胁建模的思想通过威胁树对威胁事件、威胁特征信标和采集代理之间的内在联系，构建威胁−采集树模型，如图 1 所示，并给出了相关定义。

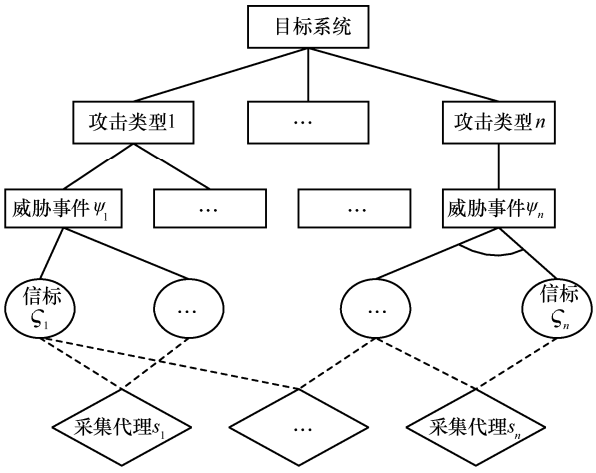

图 1　威胁−采集树模型

定义 9　威胁−采集树。威胁−采集树是描述威胁事件和采集代理之间通过采集项中信标确定映射关系的一个层次结构，包括目标系统、攻击类型、威胁事件、信标和采集代理 5 个层次。其中，第一层是目标系统，即管理员需要保护的系统；第二层是目标系统可能受到攻击的攻击类型；第三层是每种攻击类型可以通过一些威胁事件来实现；第四层是每个威胁事件所对应的威胁特征信标；第五层是能够采集到威胁信标的采集代理。由此，本文将攻击方的效用 $\mathrm{Utility}_{\mathrm{attacker}}$ 使用攻击方选取的威胁事件的风险函数 Risk 来表示。

定义 10　风险。风险是威胁事件发生的可能性与威胁事件的影响的乘积，即风险效用函数，具体

可形式化定义为

$$\text{Utility}_{\text{attacker}}=\text{Risk}=P_{\psi}I_{\psi} \tag{3}$$

其中，P_{ψ}为攻击方在防守方部署了采集代理时，威胁事件ψ可能发生的概率，I_{ψ} 为威胁事件ψ对系统的影响，即

$$I_{\psi}=w_cE_c+w_iE_i+w_aE_a$$

其中，E_c表示影响系统机密性，w_c表示影响系统机密性的权重，E_i表示影响系统完整性，w_i表示影响系统完整性的权重，E_a表示影响系统可性，w_a表示影响系统可用性的权重。“影响”表示一种威胁事件对系统的影响，本文从机密性、完整性和可用性 3 个方面对其进行衡量。

P_{ψ} 为攻击方在防守方部署了采集代理时，某一个威胁事件可能发生的概率。该概率是该威胁事件发生但未被最小特征信标集合监测到的概率与该威胁事件发生且被最小特征信标集合监测到的概率之和，即

$$\begin{aligned}P_{\psi}=&\prod\nolimits_{\tau_i\in\gamma(\psi)\wedge\text{Conf}_{\tau_i}\neq 0}(1-P_{\tau_i})+\\&\sum\nolimits_{\tau_i\in\gamma(\psi)}\prod\nolimits_{\tau_i\in\gamma(\psi)}P_{\tau_i}(1-\text{Conf}_{\tau_i})\cdot\\&\prod\nolimits_{\tau_j\in\gamma(\psi)\wedge i\neq j}(1-P_{\tau_j})\end{aligned} \tag{4}$$

根据数学归纳，可以将式(4)简化为

$$P_{\psi}=\prod\nolimits_{\tau_i\in\gamma(\psi)}(1-P_{\tau_i}\text{Conf}_{\tau_i}) \tag{5}$$

通过上述定义和度量，现将目标函数表示为

$$\text{Risk}_{\psi}(S_d)=\prod\nolimits_{\tau_i\in\gamma(\psi)}(1-P_{\tau_i}\text{Conf}_{\tau_i})I_{\psi} \tag{6}$$

3.3 部署目标属性

根据 3.2 节中给出的效用函数可以发现，该效用函数具有次模性，即在采集代理部署节点足够的情况下，再添加一个采集代理的部署点能获得的收益并不比部署该点前的收益大。后文中使用函数 R 来代替 Risk_{ψ}，目标函数 R 具有以下特性。

1) 次模性：若对所有 $S_{d_1}\subseteq S_{d_2}\subseteq V$ 且 $s\in V\backslash S_d$，则有

$$R(S_{d_1}\cup\{s\})-R(S_{d_1})\geqslant R(S_{d_2}\cup\{s\})-R(S_{d_2})$$

2) 单调性：若对所有的 $S_{d_1}\subseteq S_{d_2}\subseteq V$，则有

$$R(S_{d_1})\leqslant R(S_{d_2})$$

3) 规定：$R(V)=0$。

因此，攻击者收益问题可以形式化为

$$\max\nolimits_{S_d\subseteq V}R(S_d) \tag{7}$$

$$\text{s.t.}\ |S_d|\leqslant k \tag{8}$$

其中，R 具有标准单调次模性，k 是部署点个数的限制。式(7)是一个 NP-hard 问题[25]，该问题经常使用启发式算法求近似解。Nemhauser 等[26]提出一个基本结论：对于具有次模性的函数，贪婪算法实现了一个常数因子近似值，通过贪婪算法集合 S_d 至少获得一个常数分数$(1-\frac{1}{\text{e}})$，该常数分数是基于通过最优解获取的观察点值。例如，$R(S_d)\geqslant(1-\frac{1}{\text{e}})\min_{|S_d|\leqslant k}R(S_d)$，除非 P=NP[22]，否则没有多项式时间算法可以提供更好的近似保证。

因此，针对本文的问题选择使用贪婪算法，其算法思想为：从一个空集开始，迭代地添加一个元素 $s^*=\text{argmin}_{s\notin V\backslash S_d}R(S_d\cup\{s\})$，直到添加元素的个数满足 k 的要求，则贪婪算法选择完毕。

针对敌对环境，优化采集代理部署问题解决以下问题，即

$$\min\nolimits_{S_d\subseteq V}\max\nolimits_i R_i(S_d) \tag{9}$$

$$\text{s.t.}\ |S_d|\leqslant k \tag{10}$$

防守方的目标是选择一组设备点部署采集代理，能够应对攻击方时监测效果表现最好。因此要考虑攻击方对部署策略的影响。本文的敌对环境设置为攻击方知道防守方采集代理部署的位置 S_d，选择对防守方结果最坏的 R_i。应注意的是，尽管所有的 R_i 都具有次模性，但是 $G(S_d)=\max_i R_i(S_d)$不具备次模性。在这个设置中，简单贪婪算法可以执行。

本文的目标是找到在应对策略式攻击时表现良好的采集代理部署位置。因此，本文在连续博弈的矩阵中寻找一种平衡，每一个 S_d 作为一行，每个 R_i 作为一列。

定理 1 对于式(9)所示问题，除非 P=NP，否则不存在任何多项式时间逼近算法。

证明 设 n 是实际问题中的规模大小，$\beta(\cdot)>0$ 是任意一个关于 n 的正函数。如果存在一个多项式时间算法，能够保证找到一组个数为 k 的集合 S'_d，则有

$$\min\nolimits_i R_i(S'_d)\geqslant\beta(n)\max\nolimits_{|S_d|\leqslant k}\min\nolimits_i R_i(S_d)$$

则 P =NP。

因此，除非 P=NP，否则就不存在任何可以提供保证的算法。证毕。

4 顽健性采集代理部署算法

由于本文的优化部署目标函数是最小化攻击

方效用，该效用使用了度量值作为目标函数和约束条件，目标函数具有非线性和非凸性。因此，不可能使用凸优化技术（如内部点方法）来解决此目标函数。此外，还有一个额外的限制，即采集代理部署变量是二进制的，所以使用梯度下降方法的混合整数非线性程序解决方案也没有用处，因为度量函数在搜索空间上不是连续的。同时考虑到敌对环境的设置，本文借鉴了文献[10]对抗设置来针对目标函数的优化。

4.1　RCD 算法思路

RCD 算法是在贪婪算法的基础上进行的。首先，将式(9)问题进行转换，找到替代的方案。

$$\min_{z,S_d} z \tag{11}$$

$$\text{s.t.}\quad z \geqslant R_i(S_d) \tag{12}$$

$$1 \leqslant i \leqslant m \tag{13}$$

$$|S_d| \leqslant k \tag{14}$$

设集合 S_d 的大小最大为 k，对于所有 i 可以满足 $R_i(S_d) \leqslant z$，z 尽可能小。

然后，解决目标转换后成为式(11)问题：对于每个 z 的取值，可找到成本最低的集合 S_d，对于所有的 i 可以满足 $R_i(S_d) \leqslant z$。如果成本最低的集合最多具有 k 个元素，则 z 是可行的。

最后，用一个固定的 z 来描述近似解方程(10)。对于 $z>0$，有

$$R'_{i,z}(S_d) = \max\{R_i(S_d), z\} \tag{15}$$

最初的函数 R_i 在 z 的位置被截断，其中，$R'_{i,z}(S_d)$ 函数也是次模函数[27]，其平均值是

$$\overline{R}_z(S_d) = \frac{1}{m}\sum_i \max\{R'_{i,z}(S_d), z\} \tag{16}$$

次模函数在凸组合下是封闭的，因此，$\overline{R}_z$ 是单调的次模函数。

4.2　RCD 算法

本节详细介绍 RCD 算法过程，该算法能找到一组，至少和最优集合一样的结果，但求解时间会稍微增加。贪婪算法和 RCD 算法具体伪代码如算法 1 和算法 2 所示。

算法 1　贪婪算法 Greedy($\overline{R}_z$, z)

输入　$\overline{R}_z$，z

输出　采集代理 S 的部署位置集合 S_d

1) $S_d \leftarrow \varnothing$

2) while　$\overline{R}_z(S_d) > z$　do

3)　　for each　$s \in V_{\text{asset}} \setminus S_d$　do

4)　　$\mu_s \leftarrow \overline{R}_z(S_d \cup \{s\}) - \overline{R}_z(S_d)$

5)　　$S_d \leftarrow S_d \cup \{\text{argmin}\ \mu_s\}$

6)　　end for

7) end while

8) 输出 S_d

算法 2　RCD 算法 RCD($R_1, R_2, \cdots, R_i, k$)

输入　效用函数 $R_1, R_2, \cdots, R_i$，采集代理个数 k

输出　采集代理 S 的部署位置集合 $S_{d\text{best}}$

1) $z_{\min} \leftarrow \max_i R_i(V)$;
　$z_{\max} \leftarrow \max_i R_i(\varnothing)$;
　$S_{d\text{best}} \leftarrow \varnothing$;

2) while　$z_{\max} - z_{\min} > \frac{1}{m}$　do

3)　　$z \leftarrow \frac{z_{\max} + z_{\min}}{2}$;

4)　　$\forall S_d$ 定义 $\overline{R}_z(S_d) \leftarrow \frac{1}{m}\sum_i \max\{R'_{i,z}(S_d), z\}$;

5)　　$S_d \leftarrow \text{Greedy}(\overline{R}_z, z)$;

6)　　if　$|S_d| > k$　then

7)　　$z_{\min} \leftarrow z$;

8)　　else $z_{\max} \leftarrow z$;

9)　　$S_{d\text{best}} \leftarrow S_d$

10)　　end if

11) end while

12) 输出 $S_{d\text{best}}$

首先，计算出算法 2 中所能取到的最大值 $z_{\max}$ 和最小值 $z_{\min}$，其中，最大值 $z_{\max}$ 是当所有采集代理都没有部署时，攻击方效用值最大；最小值 $z_{\min}$ 是当所有设备节点上都部署采集代理，攻击方效用最小。其次，求出最大值 $z_{\max}$ 和最小值 $z_{\min}$ 的平均值 z，同时，针对任意一组采集代理集合 S_d 都可以计算出对应的收益 $\overline{R}_z(S_d)$。再次，调用算法 1，根据均值 z 与 $\overline{R}_z(S_d)$ 依次找出每一轮中增量绝对值最大的设备节点 ID 的组合，并且将其赋值给 $S_{d\text{best}}$；如果 $S_{d\text{best}}$ 个数不满足 k 的要求，则使用 z 当前的取值赋给 $z_{\max}$ 或 $z_{\min}$。最后，再次调用算法 1，依次循环来找到满足目标函数的部署集合。需要注意的是，每次调用算法 1 时，都是从空集开始的。

5　实验

5.1　算例

本节以一个小型网络为目标，如图 2 所示，将

MADG 模型和 RCD 算法进行实例化，该算例中网络设备节点共有 5 个，可以部署采集代理的设备，根据开放式 Web 应用程序安全项目（OWASP, Open Web Application Security Project）中 top10 的攻击类型，选取排名靠前的 4 类网络威胁事件。

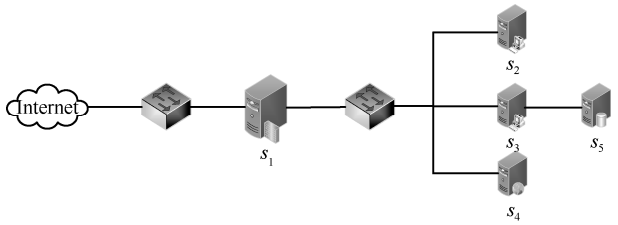

图 2 小型网络

采集代理与各指标之间的映射关系为

$\varphi(s_1)=\{\varsigma_1, \varsigma_2\}$

$\varphi(s_2)=\{\varsigma_2, \varsigma_3, \varsigma_4, \varsigma_5, \varsigma_6, \varsigma_7\}$

$\varphi(s_3)=\{\varsigma_3, \varsigma_8, \varsigma_9, \varsigma_{10}\}$

$\varphi(s_4)=\{\varsigma_{11}, \varsigma_{12}\}$

$\varphi(s_5)=\{\varsigma_{13}\}$

最小信标集合与威胁事件的对应关系为

$\gamma(\psi_1)=\{\{\varsigma_1\},\{\varsigma_2\}\}$；暴力破解

$\gamma(\psi_2)=\{\{\varsigma_3\},\{\varsigma_8\},\{\varsigma_{11}\}\}$；DDOS 攻击

$\gamma(\psi_3)=\{\{\varsigma_4\},\{\varsigma_5\},\{\varsigma_6\}\}$；XSS 攻击

$\gamma(\psi_4)=\{\{\varsigma_7, \varsigma_{10}\},\{\varsigma_{12}\},\{\varsigma_{13}\}\}$；SQL 注入

根据上述思路可以在采集项中提取对应信息，生成特征信标。本算例中生成的特征信标为

ς_1: SSH尝试失败次数>阈值

ς_2: SSH开始尝试次数>阈值

ς_3: Syn半连接个数

ς_4: XXS尝试通过资源上的URL字符串 /logfile/index.php?page =capture_data.php

ς_5: XXS尝试通过表格NET_STAT_INFO注入

ς_6: XXS尝试通过资源上的URL字符串 /logfile/index.php

ς_7: 包含MySQL版本的字符串

ς_8: 接收到网络数据分组的个数>正常值

ς_9: HTTP PHP文件POST请求

ς_{10}: MySQL注入HTTP获取尝试

ς_{11}: CPU利用率>正常值

ς_{12}: 表格NET_STAT_INFO尝试SQL注入

ς_{13}: MySQL注入类型询问

根据威胁事件与威胁事件特征信标之间的关系、威胁事件特征信标和采集代理之间的内在联系，构建威胁−采集树模型，如图 3 所示。每个威胁事件的最小特征信标集合发生的概率如表 2 所示。

表 2 最小特征信标集合发生的概率

最小特征信标集合	概率
$\{\varsigma_1\}$	0.3
$\{\varsigma_2\}$	0.8
$\{\varsigma_3\}$	0.5
$\{\varsigma_8\}$	0.7
$\{\varsigma_{11}\}$	0.8
$\{\varsigma_4\}$	0.5
$\{\varsigma_5\}$	1.0
$\{\varsigma_6\}$	0.3
$\{\varsigma_7, \varsigma_{10}\}$	0.9
$\{\varsigma_{12}\}$	0.3
$\{\varsigma_{13}\}$	0.8

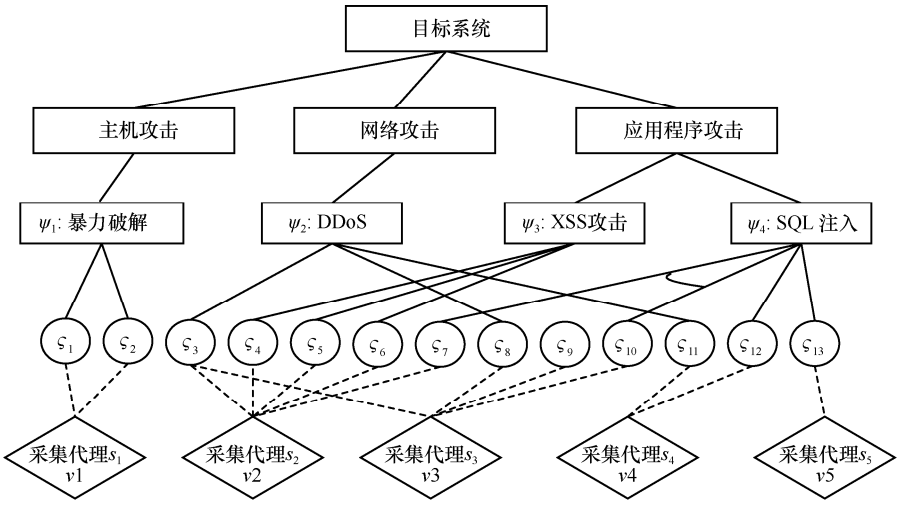

图 3 小算例的威胁−采集树模型

由于每个特征信标是由不同的采集代理采集的采集项生成的，因此每个特征信标的置信度与生成它的采集代理的置信度保持一致。根据图 2 的网络结构可知，s_1 为防火墙被攻击方攻击的概率比较大，那么它的置信度相对就会小一些，因此，通过 MTL 给出 s_1 的置信度为 0.3，依次类推，给出其他设备置信度，如表 3 所示。

表 3　　采集代理置信度

采集代理	置信度
s_1	0.3
s_2	0.3
s_3	0.5
s_4	0.5
s_5	0.9

通过对系统的机密性（confidentiality）、完整性（integrity）和可用性（availability）3 个方面的考虑，同时参照 OWASP 中 top10 列表中的信息，给出本节算例中每个威胁事件的影响值，如表 4 所示。

表 4　　威胁事件影响值

威胁事件	影响值
ψ_1	14
ψ_2	20
ψ_3	5
ψ_4	10

本节算例分别使用 EXACT 算法、GOMD 算法[14]、RCD 算法进行计算，相关数据如表 5 所示，3 种算法的具体求解数据见附录。

表 5　　3 种算法求解

求解算法	部署集合 S_d	增量	效用函数值
EXACT 算法	[1,3,4]	—	9.682 4
GOMD 算法[14]	[1,3,4]	0.076 148 413	9.682 4
RCD 算法	[1,3,4]	—	9.682 4

5.2　扩展

为测试本文提出的 RCD 算法的扩展性，在 BA 网络上使用 50 个设备节点、100 个威胁事件作为基准进行实验，并从中生成不同规模的网络和不同规模的威胁事件集合，分别对每种规模的网络进行 100 次实验，取平均值作为实验结果。

本节主要对 RCD 算法的时效性和顽健性进行验证，分别对其运行时间和求解质量进行实验，并与 EXACT 算法、MMI 算法[12]和 GOMD 算法[14]进行对比。以上所有实验均在 Windows 10 的 64 位系统上运行，该系统加载 Inter(R) i7 处理器、500 GB 硬盘和 8 GB 内存。

首先，本文在不同规模的网络（设备节点数量从 0 个节点增加到 50 个节点，步长为 5）上进行实验，采集代理的数量 k=3，各算法的运行时间如图 4 所示，求解质量如表 6 所示。

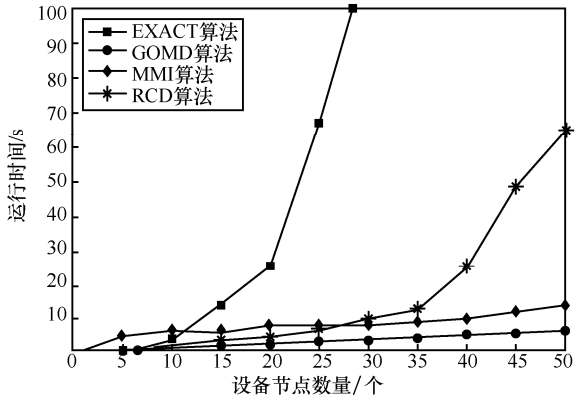

图 4　不同规模网络的运行时间对比

表 6　　不同规模网络的求解质量对比

设备节点数量/个	EXACT 算法	MMI[12] 算法	GOMD[14] 算法	RCD 算法
0	19	19	19	19
5	18	19	18	18
10	17	18	17	17
15	17.76	19	18	17.76
20	18	18	19	18
25	17	19	19	17
30	17	19	19	17
35	18	19	19	18
40	18	19	19	18
45	17	19	18	17
50	18.24	19	19	18.24

实验表明，EXACT 算法随着网络规模增长（20 个设备节点时）求解时间开始急剧增长，然而 MMI 算法[12]和 GOMD 算法[14]的求解时间增长相对缓慢；RCD 算法求解时间的增长介于 EXACT 算法和 GOMD 算法之间。其中，当设备节点为 35 个时，由于要多次调用 Greedy 算法，RCD 算法的运行时间出现急剧增长。

在求解质量方面，攻击者对目标系统的影响值越大，求解数值越大，求解的质量就越差。实验表明，MMI 算法的求解质量最差，GOMD 算法在小规模的网络下（如设备节点在 10 个以内）所得解

与 EXACT 算法保持一致，但随着网络规模的增加，GOMD 算法与 EXACT 算法之间存在很大的偏差，而 RCD 算法始终与 EXACT 算法保持一致。

其次，在网络规模固定（设备节点数量为 50）的情况下，通过改变采集代理的数量（k=0，3，5，10，15，20，25），对 RCD 算法和 3 种对比算法进行实验。各算法的运行时间如图 5 所示，求解质量如表 7 所示。

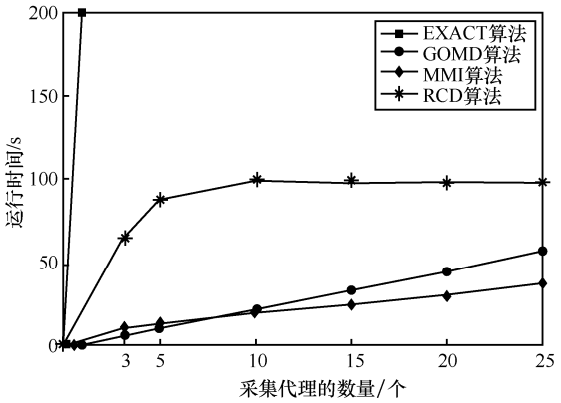

图 5 相同规模网络的运行时间对比

表 7 相同规模网络的求解质量对比

采集代理数量/个	EXACT 算法	MMI[12] 算法	GOMD[14] 算法	RCD 算法
0	19	19	19	19
3	18.24	19	19	18.24
5	—	19	19	18.24
10	—	19	19	18.24
15	—	19	19	18.24
20	—	19	19	18.24
25	—	19	19	18.24

随着采集代理数量的增加，产生的排列组合数量过多，EXACT 算法计算时间急剧上升。当采集代理数量为 3 时，运行时间为 730.057 s，当采集代理数量为 5 时，运行时间为 1 500 s；MMI 算法和 GOMD 算法可以在相对较短的时间内计算出有效解，计算时间的趋势呈线性增长；RCD 算法在开始阶段，计算时间随着采集代理数量呈线性增长，当采集代理数量增加到 3 个时，运行时间增长趋势变缓，当采集代理数量增加到 5 个时，计算时间趋于稳定。

实验表明，当网络设备数量总数相等时，采集代理数量增加到一定数量后，EXACT 算法求解质量会趋于恒定，且无法给出有效解。虽然 MMI 算法和 GOMD 算法可以给出有效解，但求解质量无法达到精确解值。然而，RCD 算法基本与 EXACT 算法保持一致，随着采集代理的数量增加，当 EXACT 算法无法给出有效解时，RCD 算法仍然可以继续给出有效解。

通过实验表明，本文提出的 RCD 算法以时间成本为代价，保证了求解质量的精确度，因此该算法是有效可行的，且具有顽健性，还能够进行一定的扩展。

6 结束语

本文针对敌对情况下采集代理优化部署问题进行了研究。首先，将攻防双方建模为一个攻防博弈问题，提出一个度量攻防博弈模型——MADG 模型；其次，对模型中涉及的系统、威胁和采集代理的相关属性进行定义和度量，并通过威胁事件和采集项之间的复杂联系构建了威胁−采集树；再次，采用数学规划对攻击方的效用函数进行构建，即优化的目标函数，并设计了一个具有顽健性的采集代理部署算法——RCD 算法，使防守方在面对攻击方进行策略式攻击时，部署策略具有比较好的顽健性；最后，通过构建实际算例，分析了本文方法的有效性和可扩展性。

附录 3 种算法的具体求解数据

3 种算法的具体求解数据如表 8~表 10 所示。

表 8 EXACT 算法精确求解

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	威胁事件中最大效用	最小效用
123	1	14	0.691 6	9.682 4	9.75	9.682 4
	2	20	0.487 5	9.75		
	3	5	0.541 45	2.707 25		
	4	10	0.73	7.3		
124	1	14	0.691 6	9.682 4	10.2	
	2	20	0.51	10.2		
	3	5	0.541 45	2.707 25		
	4	10	0.85	8.5		

（表 8 续表）

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	威胁事件中最大效用	最小效用
125	1	14	0.691 6	9.682 4	17	9.682 4
	2	20	0.85	17		
	3	5	0.541 45	2.707 25		
	4	10	0.28	2.8		
134	1	14	0.691 6	9.682 4	9.682 4	
	2	20	0.292 5	5.85		
	3	5	1	5		
	4	10	0.85	8.5		
135	1	14	0.691 6	9.682 4	9.75	
	2	20	0.487 5	9.75		
	3	5	1	5		
	4	10	0.28	2.8		
145	1	14	0.691 6	9.682 4	12	
	2	20	0.6	12		
	3	5	1	5		
	4	10	0.238	2.38		
234	1	14	0.76	10.64	10.64	
	2	20	0.292 5	5.85		
	3	5	0.541 45	2.707 25		
	4	10	0.620 5	6.205		
235	1	14	0.76	10.64	10.64	
	2	20	0.45	9		
	3	5	0.541 45	2.707 25		
	4	10	0.204 4	2.044		
245	1	14	0.76	10.64	10.64	
	2	20	0.51	10.2		
	3	5	0.541 45	2.707 25		
	4	10	0.238	2.38		
345	1	14	1	14	14	
	2	20	0.292 5	5.85		
	3	5	1	5		
	4	10	0.238	2.38		

表 9　　GOMD 算法近似求解

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	本轮最大	上一轮	增量	增量最小
1	1	14	0.691 6	9.682 4	20	0	20	14
	2	20	1	20				
	3	5	1	5				
	4	10	1	10				
2	1	14	0.76	10.64	17	0	17	
	2	20	0.85	17				
	3	5	0.5414 5	2.707 25				
	4	10	1	10				
3	1	14	1	14	14	0	14	
	2	20	0.487 5	9.75				
	3	5	1	5				
	4	10	1	10				

（表 9 续表）

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	本轮最大	上一轮	增量	增量最小
4	1	14	1	14	14	0	14	14
	2	20	0.6	12				
	3	5	1	5				
	4	10	0.85	8.5				
5	1	14	1	14	20	0	20	
	2	20	1	20				
	3	5	1	5				
	4	10	0.28	2.8				
31	1	14	0.691 6	9.682 4	10	14	−4	−4
	2	20	0.487 5	9.75				
	3	5	1	5				
	4	10	1	10				
32	1	14	0.76	10.64	10.64	14	−3.36	
	2	20	0.487 5	9.75				
	3	5	0.541 45	2.707 25				
	4	10	0.73	7.3				
34	1	14	1	14	14	14	0	
	2	20	0.292 5	5.85				
	3	5	1	5				
	4	10	0.85	8.5				
35	1	14	1	14	14	14	0	
	2	20	0.487 5	9.75				
	3	5	1	5				
	4	10	0.28	2.8				
41	1	14	0.691 6	9.682 4	12	14	−2	
	2	20	0.6	12				
	3	5	1	5				
	4	10	0.85	8.5				
42	1	14	0.76	10.64	10.64	14	−3.36	
	2	20	0.45	9				
	3	5	0.541 45	2.707 25				
	4	10	0.85	8.5				
45	1	14	1	14	14	14	0	
	2	20	0.6	12				
	3	5	1	5				
	4	10	0.238	2.38				
312	1	14	0.691 6	9.682 4	9.75	10	−0.25	−0.317 6
	2	20	0.487 5	9.75				
	3	5	0.541 45	2.707 25				
	4	10	0.73	7.3				
314	1	14	0.691 6	9.682 4	9.682 4	10	−0.317 6	
	2	20	0.292 5	5.85				
	3	5	1	5				
	4	10	0.85	8.5				
315	1	14	0.691 6	9.682 4	9.75	10	−0.25	
	2	20	0.487 5	9.75				
	3	5	1	5				
	4	10	0.28	2.8				

表 10　　**RCD 算法近似求解**

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	$C=\frac{C_{max}+C_{min}}{2}$	$\max(f, c)$	FCA	上一轮	增量	增量最小
0	1	14	1	14	9.722 703 175	14	13.430 675 79	-	-	-
	2	20	1	20	9.722 703 175	20				
	3	5	1	5	9.722 703 175	9.722 703 18				
	4	10	1	10	9.722 703 175	10				
1	1	14	0.691 6	9.682 4	9.722 703 175	9.722 703 18	12.361 351 59	13.430 675 79	−1.069 324 206	
	2	20	1	20	9.722 703 175	20				
	3	5	1	5	9.722 703 175	9.722 703 18				
	4	10	1	10	9.722 703 175	10				
2	1	14	0.76	10.64	9.722 703 175	10.64	11.840 675 79	13.430 675 79	−1.59	
	2	20	0.85	17	9.722 703 175	17				
	3	5	0.541 45	2.707 25	9.722 703 175	9.722 703 18				
	4	10	1	10	9.722 703 175	10				
3	1	14	1	14	9.722 703 175	14	10.868 175 79	13.430 675 79	−2.562 5	−2.5625
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	1	5	9.722 703 175	9.722 703 18				
	4	10	1	10	9.722 703 175	10				
4	1	14	1	14	9.722 703 175	14	11.361 351 59	13.430 675 79	−2.069 324 206	
	2	20	0.6	12	9.722 703 175	12				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	0.85	8.5	9.722 703 175	9.722 70318				
5	1	14	1	14	9.722 703 175	14	13.361 351 59	13.430 675 79	−0.069 324 206	
	2	20	1	20	9.722 703 175	20				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	0.28	2.8	9.722 703 175	9.722 70318				
31	1	14	0.691 6	9.682 4	9.722 703 175	9.722 70318	9.798 851 588	10.868 175 79	−1.069 324 206	−1.069 324 206
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	1	10	9.722 703 175	10				
32	1	14	0.76	10.64	9.722 703 175	10.64	9.958 851 588	10.868 175 79	−0.909 324 206	
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	0.541 45	2.707 25	9.722 703 175	9.722 70318				
	4	10	0.73	7.3	9.722 703 175	9.722 70318				
34	1	14	1	14	9.722 703 175	14	10.792 027 38	10.868 175 79	−0.076 148 413	
	2	20	0.292 5	5.85	9.722 703 175	9.722 70318				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	0.85	8.5	9.722 703 175	9.722 70318				
35	1	14	1	14	9.722 703 175	14	10.798 851 59	10.868 175 79	−0.069 324 206	
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	0.28	2.8	9.722 703 175	9.722 70318				
312	1	14	0.691 6	9.682 4	9.722 703 175	9.722 70318	9.729 527 381	9.798 851 588	−0.069 324 206	−0.076 148 413
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	0.541 45	2.707 25	9.722 703 175	9.722 70318				
	4	10	0.73	7.3	9.722 703 175	9.722 70318				

（表 10 续表）

组合	威胁事件	影响 I	攻击可能发生的概率 P	攻击效用值 R	$C=\frac{C_{\max}+C_{\min}}{2}$	$\max(f, c)$	FCA	上一轮	增量	增量最小
314	1	14	0.691 6	9.682 4	9.722 703 175	9.722 70318	9.722 703 175	9.798 851 588	−0.076 148 413	−0.076 148 413
	2	20	0.292 5	5.85	9.722 703 175	9.722 70318				
	3	5	1	5	9.722 703 175	9.722 70318				
	4	10	0.85	8.5	9.722 703 175	9.722 70318				
315	1	14	0.691 6	9.682 4	9.722 703 175	9.722 70318	9.729 527 381	9.798 851 588	−0.069 324 206	
	2	20	0.487 5	9.75	9.722 703 175	9.75				
	3	5	1	5	9.722 703 175	9.722 703 18				
	4	10	0.28	2.8	9.722 703 175	9.722 703 18				

参考文献：

[1] 马莉波，李星，张亮. 有效扫描监测系统建模与部署[J]. 软件学报，2009, 20(4): 845-857.
MA L B, LI X, ZHANG L. On modeling and deploying an effective scan monitoring system[J]. Journal of Software. 2009, 20(4): 845-857.

[2] TALELE N, TEUTSCH J, ERBACHER R, et al. Monitor placement for large-scale systems[C]//The 19th ACM symposium on Access control models and technologies (SACMT'14). 2014: 29-40.

[3] AQIL A. Resource efficient frameworks for network and security problems[D]. California: University of California, Riverside, 2017.

[4] BREITBART Y, CHAN C Y, GAROFALAKIS M, et al. Efficiently monitoring bandwidth and latency in IP networks[C]// INFOCOM 2001: 1-10.

[5] HOCHBAUM D S. Approximation algorithm for NP-Hard problems[M]. Boston: PWS Publishing Company, 1997.

[6] SUH K, GUO Y, KUROSE J, et al. Locating network monitors: complexity, heuristics and coverage [C]//INFOCOM 2005. 2005: 351-361.

[7] CHAUDET C, FLEURY E, GUÉRIN LASSOUS I, et al. Optimal positioning of active and passive monitoring devices[C]//The CoNEXT. 2005: 71-82.

[8] 蔡志平，刘芳，赵文涛，等. 网络测量部署模型及其优化算法[J]. 软件学报, 2008, 19(2): 419-431
CAI Z P, LIU F, ZHAO W T, et al. Deploying models and optimization algorithms of network measurement[J]. Journal of Software, 2008, 19(2): 419-431.

[9] LESKOVEC J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C]//The 13th ACM SIGKDD International Conference on Knowledge Discovery and Datamining. 2007: 420-429.

[10] KRAUSE A, MCMAHAN B, GUESTRIN C, et al. Selecting observations against adversarial objectives[C]//International Conference on Neural Information Processing Systems. 2007: 777-784.

[11] COMBOUL M, GHANEM R. Value of information in the design of resilient water distribution sensor networks[J]. Journal of Water Resources Planning and Management, 2012, 139(4): 449-455.

[12] YU Y, XIAO G. On early detection of strong infections in complex networks[J]. Journal of Physics A Mathematical & Theoretical, 2014, 47(6): 881-892.

[13] ZHOU C, LU W X, ZHANG J Z, et al. Early detection of dynamic harmful cascades in large-scale networks[J]. Journal of Computational Science, 2018(28): 304-317.

[14] THAKORE U, GABRIEL A W, WILLIAM H S. A quantitative method-ology for security monitor deployment[C]//2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). 2016: 1-12.

[15] MARBACK A, DO H, HE K, et al. A threat model‐based approach to security testing[J]. Software: Practice and Experience, 2013, 43(2): 241-258.

[16] PARDUE H, LANDRY J, YASINSAC A. A risk assessment model for voting systems using threat trees and Monte Carlo simulation[C]//2009 First International Workshop on Requirements Engineering for e-Voting Systems (RE-VOTE). 2010: 55-60.

[17] MORIKAWA I, YAMAOKA Y. Threat tree templates to ease difficulties in threat modeling[C]//2011 14th International Conference on Network-Based Information Systems. 2011: 673-678.

[18] ZHOU D, YAN Z, FU Y, et al. A survey on network data collection[J]. Journal of Network and Computer Applications, 2018, 116(8): 9-23.

[19] LIU G, YAN Z, PEDRYCZ W. Data collection for attack detection and security measurement in mobile Ad Hoc networks: a survey[J]. Journal of Network and Computer Applications, 2018, 105(3): 105-122.

[20] LIN H, YAN Z, CHEN Y, et al. A survey on network security-related data collection technologies[J]. IEEE Access, 2018, 6(3): 18345-18365.

[21] HE L, YAN Z, ATIQUZZAMAN M. LTE/LTE-a network security data collection and analysis for security measurement: a survey[J]. IEEE Access, 2018, 6(1): 4220-4242.

[22] CUPPENS F, ORTALO R. LAMBDA: a language to model a database for detection of attacks [C]//International Workshop on Recent Advances in Intrusion Detection. 2000:197-216.

[23] TOTEL E, BERNARD V, LUDOVIC M. A language driven intrusion detection system for event and alert correlation[C]//Security and Pro-

tection in Information Processing Systems. 2004: 209-224.

[24] HOSMER H H. Security is fuzzy!: applying the fuzzy logic paradigm to the multipolicy paradigm[C]//Workshop on New Security Paradigms. 1993: 175-184.

[25] FEIGE U. A threshold of ln *n* for approximating set cover[J]. Journal of the ACM, 1998, 45(4): 634-652.

[26] NEMHAUSER G L, WOLSEY L A, FISHER M L. An analysis of approximations for maximizing submodular set functions—I[J]. Mathematical Programming, 1978, 14(1): 265-294.

[27] FUJITO T. Approximation algorithms for submodular set cover with applications[J]. IEICE Transactions on Information and Systems, 2000, 83(3): 480-487.

[作者简介]

陈黎丽（1985− ），女，甘肃天水人，西安电子科技大学博士生，主要研究方向为网络空间安全、攻防博弈。

王震（1984− ），男，山东聊城人，博士，中国科学院信息工程研究所在站博士后，杭州电子科技大学副研究员，主要研究方向为网络空间安全、博弈论。

郭云川（1976− ），男，四川营山人，博士，中国科学院信息工程研究所副研究员，主要研究方向为网络空间安全、访问控制。

华佳烽（1989− ），男，湖北浠水人，西安电子科技大学博士生，主要研究方向为信息安全、隐私保护。

姚宇超（1997− ），男，新疆库尔勒人，主要研究方向为网络与系统安全。

李凤华（1966− ），男，湖北浠水人，博士，中国科学院信息工程研究所研究员、博士生导师，主要研究方向为网络与系统安全、信息保护、隐私计算。