BGP 异常事件影响风险区域快速识别方法

刘自勉, 邱 菡, 王 瑞, 朱俊虎, 王清贤

中国人民解放军网络空间部队信息工程大学网络空间安全学院 郑州 中国 450001

摘要 域间路由网络是互联网的关键基础设施。由于域间路由网络的自适应机制, 中断、攻击等 BGP 异常事件往往会引起级联效应, 对网络正常运行带来巨大危害。面向异常响应的及时性需求, 基于实时监测数据的 BGP 异常检测溯源方式存在滞后性, 难以及时阻断异常传播。在事件初期识别可能受影响的风险区域可以在异常扩散前提供告警信息, 支持提前进行有针对性的防护, 降低异常的影响。然而现有识别影响风险区域的方法通常基于级联失效模型模拟异常传播过程, 难以权衡识别准确率和识别速度。为此, 提出一种 BGP 异常事件影响风险区域快速识别方法 RRAI 以适应大规模域间路由网络的需求。通过分析域间路由网络级联失效过程中节点和边的失效机制及相互作用原理, 定义节点“风险度”用于识别易受当前异常区域影响的节点; 针对初始异常节点位置分布的两类情况: 集中式和分散式分布, 分别提出针对单区域异常和多区域异常的风险区域识别算法, 以初始异常区域为中心进行迭代式扩展, 层层筛选风险较高的节点加入风险区域。考虑到现有方法的高复杂度, 在小规模网络上与现有方法进行对比, 实验结果表明, RRAI 能在准确率提升的同时显著缩短运行时间。基于全球网络上的真实异常事件的实验结果表明, RRAI 能够在 10 分钟以内有效预测所有受损程度高的节点, 实现在大规模域间路由网络中快速识别风险区域。

关键词 域间路由安全; BGP 异常事件; 级联失效; 风险区域识别

中图法分类号 TN520.3040 DOI 号 10.19363/J.cnki.cn10-1380/tn.2025.03.07

A Rapid Identification Method for Risk Areas Affected by BGP Anomalies

LIU Zimian, QIU Han, WANG Rui, ZHU Junhu, WANG Qingxian

Institute of Cyberspace Security, PLA Strategic Support Force Information Engineering University, Zhengzhou 450001, China

Abstract Inter-domain routing networks are the key infrastructure of the Internet. Due to the adaptive mechanism of inter-domain routing networks, BGP anomalies such as disruptions and attacks often cause cascading failures which bring great harm to network. To meet the demand for timely anomaly response, the BGP anomaly detection traceability method based on real-time monitoring data has a lag, making it difficult to block the propagation of anomalies in a timely manner. Identifying potentially impacted risk areas at the early stage of the anomalies can provide alert information before anomalies spread, which can help support targeted protection in advance to reduce the impact of anomalies. However, existing methods for identifying impact risk areas are often based on cascading failure models that simulate anomaly propagation processes, making it difficult to balance identification accuracy and speed in large-scale inter-domain routing networks. For this reason, a rapid method called RRAI is proposed to identify risk areas affected by BGP anomalies that can meet the realistic needs of large-scale inter-domain routing networks. Analyzing the failure mechanisms and interaction principles of nodes and edges in the cascading failure process of inter-domain routing networks, a metric called “risk degree” was defined for identifying nodes that are vulnerable to the current anomaly area. For two types of initial anomaly node location distributions: centralized and decentralized distributions, we propose risk area identification algorithms for single-area anomalies and multi-area anomalies, respectively. The initial abnormal area was used as the center for iterative expansion, and nodes with higher risk were added to the risk area by layer screening. Comparison with existing methods is based on small-scale networks due to the high complexity of existing methods. Results show that RRAI can reduce the runtime while improving accuracy significantly. Experimental results based on real anomalous events on global networks show that RRAI can effectively predict all highly impaired nodes in less than 10 minutes, thus enabling rapid identification of risk areas in large-scale inter-domain routing networks.

Key words inter-domain routing security; BGP anomalies; cascading failure; risk area identification

通讯作者: 邱菡, 博士, 教授, Email: qiuhan410@aliyun.com。

本论文得到河南省自然科学基金项目(No. 242300421415)资助。

收稿日期: 2023-04-15; 修改日期: 2023-07-07; 定稿日期: 2025-01-10

1 引言

域间路由网络是互联网的关键基础设施。在域间路由网络中, 自治域(Autonomous System, AS)通过边界网关协议(Border Gateway Protocol, BGP)进行通信[1]。BGP 协议在设计之初未考虑安全性因素, 导致域间路由网络易受到攻击或故障的影响[2]。现有的攻击仿真实验[3]以及对现实安全事件的观测报告均表明, 域间路由网络中存在级联失效现象[4]。级联失效是指网络中少量节点或边的故障会在网络中不断传播, 进而引发大规模的网络瘫痪。多年来, 各类大规模 BGP 异常事件, 如 Slammer 蠕虫、日本地震事件、马来西亚电信路由泄露等, 均造成互联网大规模中断, 严重威胁全球网络安全[5]。

虽然目前针对 BGP 异常事件的检测溯源方法[6-7]及路由可视化技术[8-9]不断提升, 如 BGPMon[10]、Thousandeyes[11]这类互联网监测平台能够在几分钟内发现异常并定位, 但是这些方法只能识别已经受到影响的区域, 对于异常响应处置来说具有一定的滞后性。同时, 网络管理员也试图通过切断与异常节点相连接的对等节点以便尽快恢复网络[12]。但是由于级联失效现象的存在, 许多未与其直接相连的节点也受到不同程度的影响, 导致全球网络异常可持续数小时。因此, 预测异常影响传播范围, 在异常初始阶段快速识别可能受异常影响的面临网络中断风险的区域(简称风险区域), 可以有针对性地进行异常响应, 及时阻断异常传播, 维护网络稳定运行。

预测网络中的级联影响是复杂网络相关领域研究的一个重要内容。针对不同的应用场景, 研究者依据网络特性构建不同的模型刻画影响传播过程。例如针对社交网络中的信息级联传播, 研究者将信息在用户节点之间的传播过程刻画为概率图模型, 并利用基于级联模型、深度学习模型等方法学习用户被传播的概率进而预测信息传播的路径和范围[13-14]; 针对网络物理电力系统, 研究者构建综合刻画信息网络数据包传输以及电网动态潮流过程的模型, 利用仿真实验对级联故障影响进行分析[15]。而在域间路由网络领域, 绝大部分研究者采用基于负载容量模型的级联失效模型刻画 BGP 异常影响传播过程[16-19]。将域间路由网络抽象成 AS 节点构成的图, 通过摘除初始异常区域的 AS 节点和边并运行级联失效模型就可预测受 BGP 异常事件影响的区域[4]。

然而针对域间路由网络的现有方法在大规模网络中的运行效率低下。域间路由网络节点数量多且增长速度快, 根据对 CAIDA 公布的全球 AS 拓扑数据[20]的分析可知, 截止 2022 年 12 月, AS 节点数量已达 74713 个, 网络规模在近 5 年内扩大了 35%, 还在持续不断增长。现有针对级联失效模型的研究在提升与真实网络的逼真度的同时也增加了计算复杂度。以最新提出的基于最优有效路径的域间路由系统级联失效模型(简写为 CFM-VIRS 模型)为例, 计算两个点之间的最优有效路径的复杂度为 $O(V^2)$ [19], 其中 V 是网络节点数量, 则求全网所有节点对之间最短路径的计算复杂度为 $O(V^4)$ 。每当有新的节点和边失效都要重新计算最优有效路径, 假设失效 N 轮次, CFM-VIRS 的计算复杂度为 $O(NV^4)$ 。随着网络规模增大, 运行时间成指数级增加。因此, 对于规模庞大的域间路由网络来说, 利用现有级联失效模型无法在定位异常点后及时识别可能受影响的风险区域。虽然通过提取骨干网络减小网络规模可以提升级联失效模型运行速度, 但是诸如删除度为 1 的节点[21]和提取 k-core[22]等仅基于拓扑特征对网络进行缩减的方法没有考虑 AS 节点之间的商业关系, 难以适用于域间路由网络。并且无法保证初始异常节点还存在于缩减后的网络中, 难以进一步运用级联失效模型预测风险区域。

为解决上述问题, 考虑到风险区域识别的根本目的是筛选最可能受异常事件影响的高风险节点, 从而为及时响应提供决策支持, 因此本文将风险区域识别问题转化为节点相对于初始异常区域的风险评估问题。通过定义评价节点风险的指标并设计筛选节点的算法实现对风险区域的识别。提出了综合考虑不同初始异常情况的快速风险区域识别方法(Rapid Risk Area Identification, RRAI)。本文主要工作如下:

(1) 基于域间路由网络级联失效机理定义节点风险评估指标“风险度”, 用于筛选容易受异常事件影响的节点。

(2) 通过分析初始异常节点不同的位置分布情况及其影响传播原理, 分别提出单区域异常和多区域异常的风险区域识别算法, 以初始异常区域为中心, 通过层层筛选高风险节点迭代式扩展区域, 实现风险区域的识别。

(3) 从理论上分析 RRAI 的计算复杂度, 说明其在大规模网络中应用的可扩展性。通过与现有方法对比的实验, 并基于全球网络上的真实事件数据验证 RRAI 在准确率和速度方面的有效性。结果表明, RRAI 能够在大规模域间路由网络中实现风险区域的快速识别。

2 快速风险区域识别方法 RRAI

识别 BGP 异常事件影响的风险区域, 就要筛选出哪些节点容易受当前异常区域的影响, 即风险值高的节点。出于识别速度的考虑, 本文通过筛选风险高的节点构成子图作为风险区域, 从而将风险区域识别问题转化为节点相对于初始异常区域的风险评估问题。同时, 异常的传播在现有级联失效模型中均刻画为摘除初始异常节点后的多轮次失效过程[16-19]。因此, 本文采取以初始异常节点为核心迭代式向外扩展的方式, 通过层层筛选风险值高的节点实现风险区域识别。

由于导致BGP异常的原因不同, 其初始异常AS节点的位置分布可能各不相同。根据在域间路由网络中提取初始异常节点形成的子图是否连通, 本文将其位置分布分为两类: 集中式分布和分散式分布, 如图 1 所示。集中式分布是指初始异常节点可以构成连通图, 可对应于某一 AS 发生路由泄露的情况。这类异常以该初始异常区域为中心向外程辐射状传播。而无法构成连通图的分散式分布, 可对应于如俄乌战争中由于分布在不同位置的战争摧毁了当地的基础设施导致的 AS 失效[23]。其影响从各个区域分别向外扩散, 可能存在受影响区域重合的情况。

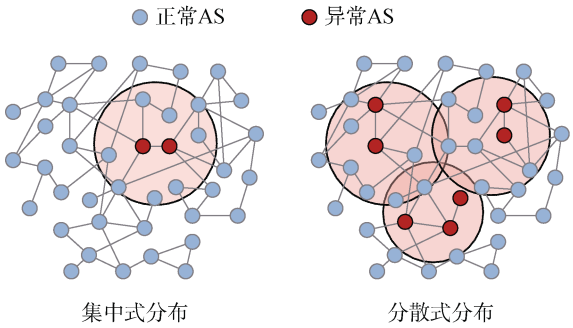


图 1 初始异常 AS 节点位置分布

Figure 1 Location distribution of initial abnormal AS nodes

基于上述分析, 提出一种快速风险区域识别方法 RRAI, 其核心要素有两个: 一是评价节点风险的指标——风险度; 二是针对不同初始异常情况的节点筛选算法——单区域和多区域异常的风险区域识别算法。

鉴于本节涉及的名词术语较多, 首先梳理关键名词术语的符号及其含义映射表, 然后详细描述 RRAI 方法的细节: 首先刻画节点的风险指标用于筛选易受影响的节点, 然后分别针对两种不同的初始异常节点分布情况设计对应的风险区域识别算法, 最后对方法的计算复杂度进行分析。

2.1 术语符号

本小节对方法中涉及的关键名词术语的符号及其基本含义进行梳理, 如表 1 所示。部分术语的具体内涵将在后文中进行详细介绍。

表 1 关键术语符号表

Table 1 Glossary of key symbols

符号	含义
CL_n^A	节点 n 与异常区域 A 的接近度
NAR_n^A	节点 n 的邻居位于异常区域 A 的比例
I_n	节点 n 的异常传播影响力
R_n^A	节点 n 相对于异常区域 A 的风险度
α	风险度计算中的节点异常原因比例控制参数
OA	初始异常区域
RA	风险区域
T	最终要识别的目标风险区域的节点数量
μ	迭代式识别过程中每轮次的节点筛选比例
RN	风险区域节点集合
NN	异常区域的邻居节点集合
TN	迭代式识别过程中每轮次的临时候选节点集合
CN	多区域异常时的扩展中心区域节点集合
CA	多区域异常时的扩展中心区域集合
V_A	区域 A 的节点数量
E_A	区域 A 的边数量
$\overline{Nei}$	节点平均邻居数量
k	网络中边与节点的数量比值
F	迭代扩展轮次

2.2 节点风险度

在信息安全领域, 风险是指网络中资源遭到破坏所造成的可能的损失[24]。本文将其对应于域间路由网络遭受攻击或故障时引发的 BGP 异常事件对 AS 节点造成的可能的影响。因此, 本小节通过分析域间路由网络级联失效机理, 综合考虑 BGP 选路模式、AS 节点的拓扑特征、以及 AS 间商业关系等多种要素刻画节点风险指标, 筛选出受当前异常区域影响较大的节点, 用于后续迭代式识别风险区域。

BGP 异常事件从以下两个方面对 AS 节点造成影响: 一方面是对节点本身的破坏, 使其无法承担流量传输任务; 另一方面是对与节点相连的边的破坏, 节点虽然功能正常但是无法与其他节点连通。依据对域间路由网络级联失效机制的分析可知, 网络中节点和边的失效条件不同[25]: 节点失效是由于路由器短时间接收并处理大量 UPDATE 报文导致 CPU、内存等资源耗尽; 边失效则是由于流量重分配

导致链路阻塞而无法传输数据包。因此, 判断一个节点是否容易受 BGP 异常事件影响而异常, 既要考虑其可能接收到的 UPDATE 报文是否会导致节点本身失效, 又要考虑与其相连的边是否容易阻塞, 从而使节点孤立。

考虑节点自身失效的情况, 依据 BGP 协议, 当某一节点发现 BGP 会话断开, 会向其邻居节点发送 UPDATE 报文进行路由更新[1]。由于已经异常的区域中会有大量 BGP 会话中断, UPDATE 报文便会从异常区域的各个节点发出并逐步向周围扩散。因此, 距离异常区域越近, 节点短时间内接收到的 UPDATE 报文数量就越多, 节点越容易失效。由此定义节点与异常区域的接近度指标如下。

定义 1. 接近度(CL).

用节点到异常区域中各点路径长度的平均值的倒数表示节点与异常区域的接近度。节点 n 与异常区域 A 的接近度 CL_n^A 为

$$CL_n^A = \frac{|A|}{\sum_{a\in A} l_n^a} \tag{1}$$

其中 $|A|$ 为异常区域 A 中节点的数量, a 是异常区域 A 中的节点, l_n^a 为节点 n 与 a 之间的路径长度。

考虑节点周围边失效的情况, 依据 BGP 协议, 当经过某一邻居的路径不可用时, 节点会选择其本地路由表中经过其他邻居的路径[1]。因而节点拥有越多位于异常区域的邻居, 其可选的备用路径越少, 越容易孤立。并且, 对于与异常节点相邻的节点来说, 原来流经异常节点的边上的流量会重分配至其未异常的邻居的边上, 造成这些边阻塞, 从而进一步促使节点孤立。由此定义节点的邻居异常率指标如下。

定义 2. 邻居异常率(NAR)

用节点邻居位于异常区域内的比例表示节点的邻居异常率。在当前异常区域为 A 时, 节点 n 的邻居异常率 NAR_n^A 为

$$NAR_n^A = \frac{N_n^A}{N_n} \tag{2}$$

其中 N_n 是节点 n 的邻居数量, N_n^A 是节点 n 位于异常区域 A 的邻居数量。

基于迭代式识别风险区域的思路, 本轮筛选出的节点将作为异常节点用于下一轮的节点筛选。因此还要考虑节点对其他节点的影响。本文将其定义为节点的异常传播影响力 I 。为提升计算速度, 可提前利用已有研究如 IKN-CF[18]和 NIE-GAT[25]等基于级联影响力评估节点的重要性的方法进行计算。本文采用适用于大规模网络的 NIE-GAT 方法, 该方法考虑了 AS 节点间的商业关系以及基于商业关系的选路策略刻画节点的异常传播影响力。

综上, CL 和 NAR 分别是刻画节点失效和节点周围边失效导致节点异常可能性的指标。因而本文用加权和的方式综合这两个指标刻画节点受到影响后异常的可能性。节点的风险, 即异常区域对节点造成的可能的影响, 可以理解为节点异常后造成的损失。因此, 本文通过将节点异常的可能性与其异常传播影响力相乘刻画节点的风险, 称之为“风险度”, 计算方式如下。

定义 3.风险度(R)

节点 n 相对异常区域 A 的风险度为

$$R_n^A = (\alpha CL_n^A + (1-\alpha) NAR_n^A) \times I_n \tag{3}$$

其中 CL_n^A 和 NAR_n^A 分别是节点 n 与异常区域 A 的接近度以及其邻居异常率; I_n 是节点 n 的异常传播影响力; α 是节点异常原因比例控制参数, 将在第 3 节详细说明参数的选择。

2.3 单区域异常的风险区域识别算法

对于初始异常节点集中式分布的情况, 异常节点构成单个连通子图, 简称其为单区域异常。由于单区域异常的影响从该区域向周围扩散, 本文将识别风险区域的过程转换成扩展异常区域的过程, 用迭代的方式模拟多轮次的级联失效。考虑到识别风险区域的目的是支持异常响应, 还需考虑识别速度和部署响应措施的资源限制, 因此本文通过限定要关注的目标风险区域节点数量作为扩展的终止条件。在实际域间路由网络中, 并不是所有检测出的异常事件都会发展成大规模的异常事件。为了服务于异常响应, 在识别出风险区域后需向高风险 AS 的管理员进行告警使其部署相应的防御措施。因此, 通过限定防御资源限制内可关注的 AS 节点数量作为扩展的终止条件, 可以避免过多小规模事件消耗大量计算资源和防御资源。

单区域异常的风险区域识别算法如算法 1 所示。首先将初始异常节点构成的子图作为初始异常区域。然后以该区域为中心迭代式扩展至设定的目标区域节点数量。在迭代过程中, 将当前异常区域内节点的一阶邻居, 即在网络中与节点直接相连的节点, 作为候选节点。通过计算候选节点相对于当前异常区域的风险度, 筛选一定比例的风险度高的节点纳入新的异常区域当中, 直到风险区域节点数量等于目标区域节点数量。筛选比例 μ 的设置方法将在第 3 节中详细说明。

算法 1. 单区域异常的风险区域识别算法

输入: 全球 AS 网络 $G=(V,E)$; 初始异常区域 $OA=(V_{OA},E_{OA})$; 目标风险区域节点数量 T; 节点筛选比例 μ

输出: 风险区域 RA

1. 初始化风险区域 $RA \leftarrow OA$, 风险区域节点集 $RN \leftarrow V_{OA}$;
2. WHILE $|RN| < T$:
3. 在 G 中查找 RA 中所有节点的一阶邻居节点, 作为邻居节点集合 NN;
4. 临时候选节点集合 $TN \leftarrow NN - RN$;
5. FOR n IN TN:
6. 计算 n 相对于 RA 的风险度 R_n^{RA};
7. IF $\mu|TN| < T - |RN|$:
8. 将 TN 中 R_n^{RA} 值最大的前 $\mu|TN|$ 个节点加入 RN;
9. ELSE:
10. 将 TN 中前 $T-|RN|$ 个节点加入 RN;
11. 在 G 中提取包含 RN 中所有节点的子图 g;
12. $RA \leftarrow g$;
13. RETURN RA;

以图 2 所示的网络为例对算法 1 进行说明。图 2 展示了两个轮次的异常区域扩展过程。第一轮扩展过程中, 初始异常节点为 1 和 2, 构成初始异常区域。首先将节点 1 和 2 的一阶邻居节点构成候选节点集合。然后计算所有候选节点相对于初始异常区域的风险度。在筛选比例为 50%的情况下, 节点 3、4 和 5 的风险高于其他节点, 则将其加入风险区域, 作为下一轮扩展的初始异常区域。同理, 在第二轮扩展过程中, 将节点 1～5 的一阶邻居节点作为候选节点, 其中节点 6～12 是相对于当前异常区域风险度更高的前 50%的节点。由此, 节点 1～12 构成的子图将作为第三轮扩展的初始异常区域。

2.4 多区域异常的风险区域识别算法

针对初始异常节点分散式分布的情况, 异常节点构成多个连通分量, 简称其为多区域异常。基于单区域异常的识别思路, 分别以初始异常区域的各个连通分量为中心进行迭代式扩展形成风险区域。在对多个互不连通的异常区域同时扩展的过程中, 可能会出现扩展后的新异常区域之间相互连通的情况。针对这一现象, 本文在筛选出相对于每个异常区域的风险节点后, 判断是否有重复节点。如果有重复节点, 说明这些区域在下一轮扩展时连通, 则将这些连通的区域融合为一个区域应用于下一轮扩展。如此可以减少后续扩展中对同时属于多个异常区域的候选节点重复计算风险度, 提升识别速度。多区域异常的风险区域识别算法如算法 2 所示。

算法 2. 多区域异常的风险区域识别算法

输入: 全球 AS 网络 $G=(V,E)$; 初始异常区域 $OA=(V_{OA},E_{OA})$; 目标风险区域节点数量 T; 节点筛选比例 μ

输出: 风险区域 RA

1. 计算 OA 在全网中的连通分量, 每一个连通分量的节点构成扩展中心区域节点集 CN, 所有 CN 构成扩展中心区域集合 CA;
2. 初始化风险区域节点集 $RN \leftarrow V_{OA}$;
3. WHILE $|RN| < T$:
4. FOR CN IN CA:
5. 在 G 中查找 CN 中所有节点的一阶邻居节点, 作为 CN 的邻居节点集合 NN_{CN};
6. 候选节点集合 $TN_{CN} \leftarrow NN_{CN} - RN$;
7. FOR n IN TN_{CN}:
8. 计算 n 相对于的 CN 风险度 R_n^{CN};
9. IF $\mu\sum|TN_{CN}| < T - |RN|$:
10. FOR CN IN CA:
11. 取 TN_{CN} 中 R_n^{CN} 值最大的前 $\mu|TN_{CN}|$ 个节点构成集合 RN_{CN} 加入 RN;
12. 将 RN_{CN} 中的节点加入 CN 形成新的扩展中心区域节点集 CN';
13. ELSE:
14. FOR CN IN CA:
15. 将 TN_{CN} 中的前 $(T-|RN|)|TN_{CN}|/\sum|TN_{CN}|$ 个节点构成集合 RN_{CN} 加入 RN;
16. 将 RN_{CN} 中的节点加入 CN 形成新的扩展中心区域节点集 CN';
17. 更新扩展中心区域集合 CA, 将每个 CN 替换成 CN';
18. WHILE CN' 之间有相同节点:
19. 将有相同节点的多个 CN' 合并成一个新的节点集合 CN'', 更新 CA;
20. 在 G 中提取包含 RN 中所有节点的子图 g;
21. $RA \leftarrow g$;
22. RETURN RA;

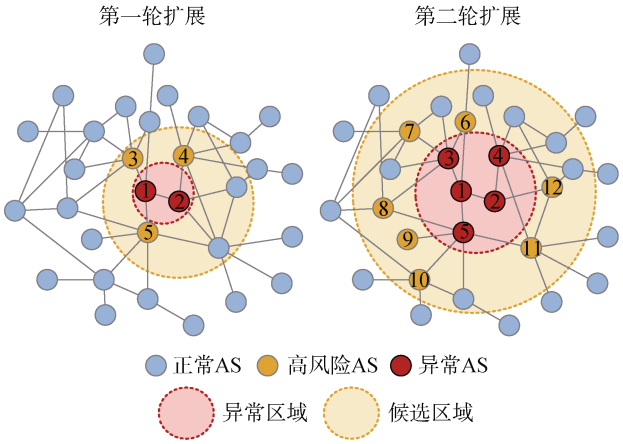


图 2　单区域异常的风险区域识别算法示例

Figure 2　Example of risk area identification algorithm for single area failure

2.5　算法复杂度分析

首先分析风险度 R 的计算复杂度。由于节点异常传播影响力 I 在事先计算完成, 因此只考虑 CL 和 NAR 这两个指标的计算复杂度。对每个节点来说, CL 需要计算节点与异常区域中各个节点之间的路径, 其复杂度为 $O(V_{An}E_{An})$, 其中 V_{An} 和 E_{An} 分别为由当前异常区域 A 和节点 n 组成的子图的节点和边的数量; NAR 的计算需要判断节点的每个邻居是否在异常区域内, 其复杂度为 $O(Nei_nV_A)$, 其中 Nei_n 和 V_A 分别为节点 n 的邻居数量和异常区域 A 的节点数量。因此, R 的计算复杂度为 $O(V_{An}E_{An}+Nei_nV_A)$。对于不同的节点, V_{An} 与 V_A 都相差 1, 假设 E_{An} 比 E_A 平均多 $\overline{E}$, 那么 $\overline{E}\in[1,V_A]$。则 R 的平均计算复杂度近似为 $O(V_A(E_A+\overline{E}+\overline{Nei}))$, 其中 $\overline{Nei}$ 为节点平均邻居数。

接着分析单区域异常的风险区域识别算法的复杂度。风险区域的识别可以理解为是多次扩展异常区域的过程, 对于每一次扩展, 都需要对候选节点计算风险度 R 并依据 R 排序。假设候选节点数量为 V_C, 则针对异常区域 A 计算风险度的复杂度为 $O(V_CV_A(E_A+\overline{E}+\overline{Nei}))$。利用快速排序算法对节点排序的平均复杂度为 $V_C\log_2V_C$。因此, 扩展一次需要 $O(V_CV_A(E_A+\overline{E}+\overline{Nei})+V_C\log_2V_C)$, V_C、V_A、E_A 和 $\overline{E}$ 均随扩展轮次增加而增加。在最后一次扩展时, 最坏的情况下, V_A 比目标风险区域节点数量 T 少一个, V_C 为全网除了异常区域 A 以外的其他节点数。此时扩展一次的计算复杂度为 $O((V-V_A)V_A(E_A+\overline{E}+\overline{Nei})+(V-V_A)\log_2(V-V_A))$。域间路由网络是稀疏网络, 边一般为节点的几倍。因此假设 E_A 为 V_A 的 k 倍, $\overline{E}$ 取其最大值 V_A, 扩展轮次为 F, 则单区域异常的风险区域识别算法的最坏计算复杂度近似为 $O(F(V-T)((k+1)T^2+\overline{NeiT})+\log_2(V-T)))$。

多区域异常的识别算法比单区域异常的算法在每一轮扩展过程中需要判断多个候选节点集是否有交集, 其计算复杂度为所有候选节点之和, 最坏情况近似于 $O(V)$。因此, 多区域异常的风险区域识别算法的最坏计算复杂度近似为 $O(F(V-T)((k+1)T^2+\overline{NeiT})+\log_2(V-T))+FV)$。

对于全球域间路由网络来说, BGP 异常事件引发的级联失效现象从局部逐步扩散至更大的范围往往需要一定的时间[4]。同时, 基于对实际观测的真实大规模异常事件的数据[26]分析可知, 在异常发生后的前 30 分钟, 实际受影响较大的节点总数不超过 500 个。这表明异常初期受影响的节点数远远小于域间路由网络中的节点总数。由此, 设定的目标区域节点数量 T 可以远小于 V。在此情况下, 单区域异常和多区域异常算法的最坏计算复杂度分别可以近似为 $O(FV((k+1)T^2+\overline{NeiT}+\log_2V))$ 和 $O(FV((k+1)T^2+\overline{NeiT}+\log_2V+1))$。由于 RRAI 综合考虑单区域异常和多区域异常两种情况, 而多区域异常的识别算法复杂度高于单区域异常, 因此 RRAI 的最坏复杂度为 $O(FV((k+1)T^2+\overline{NeiT}+\log_2V+1))$。

综上, 即便在最坏的情况下, 本文所提方法 RRAI 也低于复杂度为 $O(NV^4)$ 的基于 CFM-VIRS 模型的方法, 并且随着域间路由网络规模的增长具有更好的适应性。

3　仿真实验与分析

由于现有方法基于级联失效模型, 计算复杂度高, 因此为了与现有方法进行对比验证 RRAI 的有效性, 本节在小规模网络中进行仿真实验。首先说明仿真实验环境配置和评价指标; 然后基于理论分析和仿真实验确定 RRAI 中关键参数的设置; 接着通过与现有经典的节点评估指标和影响区域预测方法进行对比, 验证 RRAI 在准确率和效率两方面的有效性。

3.1　仿真实验环境配置

当前针对域间路由网络级联失效现象研究的方法均依托于级联失效仿真模型。为了尽可能逼近真实网络的情况, 选择既考虑节点和边不同的失效条件, 又基于商业关系计算流量传输路径的 CFM-VIRS

模型模拟 BGP 异常传播过程。同时, 为了尽最大程度使得仿真实验的网络拓扑结构和商业关系符合当前域间路由网络的特点, 基于 CAIDA 公开的 2022 年 12 月全球域间路由网络拓扑数据[20]进行实验。该数据包含了 AS 拓扑连接关系及其商业关系信息, 可用于构建具有商业关系的全球 AS 网络拓扑。由于 CFM-VIRS 模型计算复杂度高, 从全网拓扑中提取规模分别为 500、1000、2000、3000 的子网络作为仿真实验网络。在每个网络上, 分别随机选择 10 组集中分布和分散分布的初始异常节点进行实验。

实验在 Ubuntu 20.04 服务器上运行, 配置为 Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz 处理器、256GB 内存、13T 硬盘。以 Python 3.8 为编程语言, 基于 NetworkX 实现级联失效模型。

3.2 评价指标

在仿真实验中, 将运行 CFM-VIRS 模型得到的受影响节点组成的子图作为基准区域, 通过对比不同方法识别的风险区域中节点属于基准区域的比例评估识别准确率。定义评价风险区域识别方法 m 的准确率指标如下,

$$Accuracy_m = \frac{\left|RN_m \cap BN\right|}{\left|RN_m\right|} \tag{4}$$

其中 RN_m 是方法 m 识别出的风险区域中的节点集合, BN 是基准区域中的节点集合。

3.3 关键参数设置

由第 2 节可知, RRAI 包含三个关键参数: 目标区域节点数量 T、扩展过程中的节点筛选比例 μ, 以及风险度计算时的控制参数 α。

本节将 CFM-VIRS 模型得到的基准区域节点数量作为目标区域节点数量 T。对于 μ, 当 T 值固定时, 迭代轮次随 μ 值的减小而增大。为了避免迭代轮次过少导致风险节点的丢失, 需要尽可能使得更多节点具有被筛选的可能性来保证识别的准确率。考虑到当网络直径为 d 时, 最少迭代 $d/2$ 次可遍历整个网络的节点。因此要使尽可能多的节点都有被筛选的可能, 需要满足迭代次数大于等于 $d/2$。用 V_C^i 表示第 i 次迭代的候选节点数量, 则筛选比例 μ 需要满足如下要求,

$$\mu \leqslant \frac{T}{\sum\limits_{i\in[1,\frac{d}{2}]} V_C^i} \tag{5}$$

用 V 表示网络节点数量, 则 V_C^i 小于 V, 可得到如下关系,

$$\frac{T}{\frac{d}{2}\times V} < \frac{T}{\sum\limits_{i\in[1,\frac{d}{2}]} V_C^i} \tag{6}$$

为满足公式(5), 本文取 μ 值如下,

$$\mu = \frac{2T}{dV} \tag{7}$$

α 参数刻画节点自身失效以及其周围边失效这两个因素导致节点异常的影响比例。为了获得使 RRAI 识别率最高的 α, 本文设置 α 取 0 到 1, 间隔为 0.1, 统计 RRAI 在不同规模网络中的平均识别准确率, 结果如图 3 所示。从图中可以看出针对初始异常节点集中式分布的情况, α 取 0.3 时准确率最高; 针对分散式分布的情况, α 取 0.4 时准确率最高。两种情况下最优的 α 均小于 0.5, 这说明相比于节点自身失效, 其周围的边失效导致节点异常的作用更大。另外, 针对分散式分布的准确率低于集中式分布的情况, 可能是由于多区域异常的风险区域识别算法未考虑不同区域失效对某些节点的叠加影响。

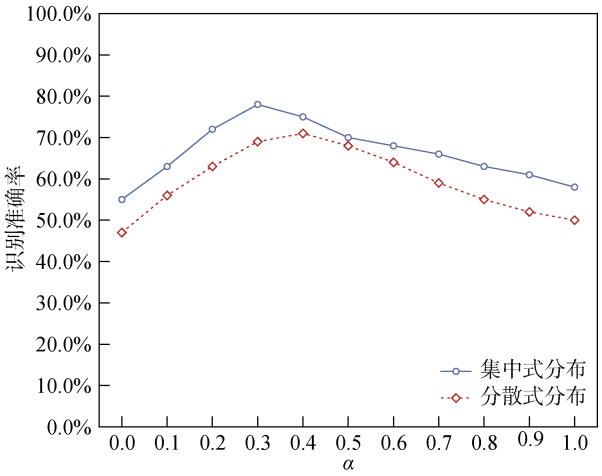


图 3 设置不同 α 的准确率变化

Figure 3 Changes in accuracy for different α

3.4 准确率对比与分析

首先对风险度指标的有效性进行验证。风险区域识别的关键是筛选风险度高的节点从而进行提前的针对性保护。以往的研究往往通过对节点重要性进行排序来决定优先保护哪些节点。常见的节点重要性评价指标包括复杂网络领域经典的度、介数, 以及域间路由网络 AS 排名指标客户锥(Customer-cone)[27]。因此本节通过对比在不同规模的网络中使用这些指标划定风险区域的平均识别准确率来验证利用风险度指标识别高风险节点的有效性。基于图 3 的结果, 在计算风险度时, α 在集中式分布和分散式分布的情况下分别取 0.3 和 0.4。实验结果如图 4 所示。

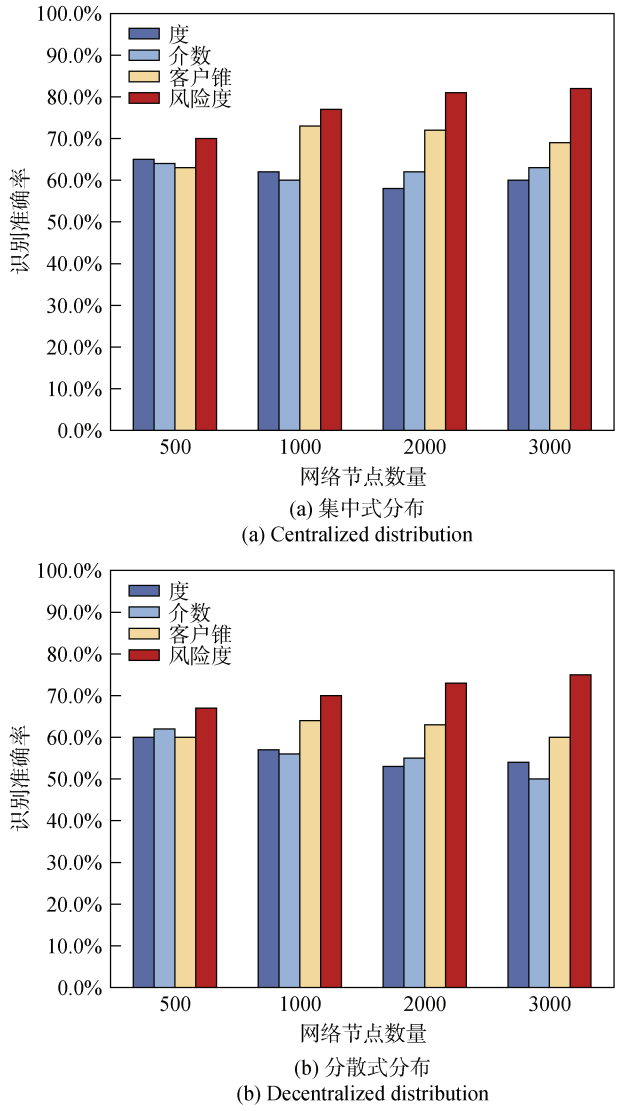


图 4　不同节点评估指标识别准确率对比

Figure 4　Accuracy comparison among different node metrics

总体来看, 本文提出的风险度指标的准确率在所有规模的网络中都明显高于其他指标。在不同规模网络中, 相比于其他指标中的最高值, 应用风险度进行筛选的准确率对初始异常节点集中式分布和分散式分布的情况分别平均提升 7.75%和 9%。这可能是因为其他指标度量的是节点在全网中的重要性, 忽略了节点与异常区域的关系。另外, 在规模较大的网络中, 度和介数的表现不如考虑域间路由网络特性的指标。这可能是因为域间路由网络流量传输与节点间的商业关系密切相关, 因此单纯的网络拓扑特征难以刻画域间路由网络节点的重要性特征。

接下来验证 RRAI 与现有方法相比的优势。现有方法基于级联失效模型识别风险区域, 低复杂度的模型虽然运行速度快但是逼真度低, 而逼真度较高的模型复杂度也高, 在大规模网络中往往通过缩减网络规模提升模型运行速度。因此, 本文选择低复杂度的 IRS-CFM 模型[16]和高复杂度的 CFM-VIRS 模型[19]进行对比。在运行 CFM-VIRS 时, 通过去掉既无 Customer 又无 Peer 的节点缩减网络, 后文简写为 CFM-VIRS-PN。另外, 为了进一步验证迭代式算法的有效性, 再添加非迭代式的 RRAI (Non-iterative RRAI, 简写为 RRAI-NI)作为对比。RRAI-NI 算法首先针对初始异常区域计算全网其他所有节点的风险度, 然后根据给定的目标区域节点数量 T, 提取风险度排名最高的 T 个节点组成的区域, 该区域即为 RRAI-NI 算法识别出的风险区域。在不同规模网络中的实验结果如图 5 所示。

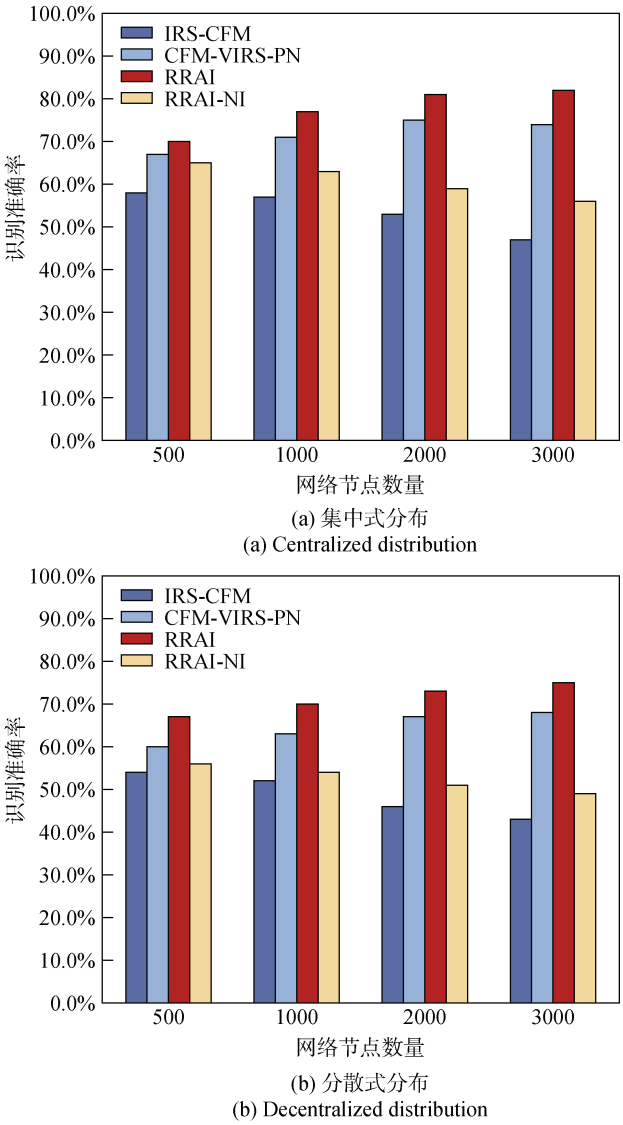


图 5　不同识别方法识别准确率对比

Figure 5　Accuracy comparison among different identification methods

由图 5 可知, RRAI 在不同规模的网络中的表现均优于其他方法。在不同规模网络中, 相比于其他方

法中的最高值, RRAI 的准确率对初始异常节点集中式和分散式分布的情况分别平均提升 5.75%和 7%。同时, RRAI 的准确率随网络规模增大稳步提升, 当网络节点数为 3000 时, 准确率可达 82.35%。同样基于风险度筛选, RRAI-NI 的准确率不高, 且随着网络规模增大准确率在下降, 说明迭代式算法的有效性。

3.5 效率对比与分析

首先分析仿真实验中对比的其他方法的计算复杂度如下:

对于 IRS-CFM, 本文采用张等[28]的方法计算介数作为节点负载, 计算复杂度为 $O(VE)$, 其中、V 和 E 分别是网络节点数和边数。依据其失效原理可知, 每轮节点失效后其负载按邻居容量大小的比例分配到各个邻居[16]。每轮需要遍历节点和边查找所有失效节点的邻居, 计算复杂度最坏是 $O(V+E)$。假设 N 是失效轮次, 则其计算复杂度为 $O(VE+N(V+E))$。

对于 CFM-VIRS-PN, 前文分析可知, CFM-VIRS 模型计算的复杂度为 $O(NV^4)$, 其中 N 是失效轮次。则 CFM-VIRS-PN 的复杂度为 $O(NV'^4)$, 其中 V' 是缩减后网络的节点数量。

对于 RRAI-NI, 需要对全网中除了初始异常节点以外的其他节点计算风险度并排序。依据第 2.5 节中的分析, 计算某个节点风险度 R 的平均计算复杂度近似为 $O(V_A(E_A+\overline{E}+\overline{Nei}))$。除了初始异常区域以外所有节点的数量为 $V-V_A$。因此, 计算这些节点的风险度的复杂度为 $O((V-V_A)V_A(E_A+\overline{E}+\overline{Nei}))$。利用快速排序算法对这些节点进行排序的计算复杂度为 $(V-V_A)\log_2(V-V_A)$。综上, RRAI-NI 的复杂度为 $O((V-V_A)V_A(E_A+V_A+\overline{Nei})+(V-V_A)\log_2(V-V_A))$。初始异常节点的数量远远小于全网节点总数, 则 $V-V_A\approx V$, 所以 RRAI-NI 的计算复杂度可近似为 $O(VV_A(E_A+V_A+\overline{Nei})+V\log_2 V)$, 其中 V 是全网节点总数, V_A 和 E_A 是初始异常节点和边的数量, $\overline{Nei}$ 是网络中节点平均邻居数量。

本小节将 3.4 节中用度(Degree)、介数(Betweenness)和客户锥(Customer-cone)这三个指标替换 RRAI 方法的风险度指标 R 后得到的识别方法分别简写为 RRAI_Degree、RRAI_Betweenness 和 RRAI_Customer-cone。由于这些指标评价的是节点在全网中的重要性, 与初始异常区域无关, 因此首先计算除初始异常节点以外的所有节点的指标值, 然后在迭代过程中取候选节点中指标值较大的节点即可。这些方法的计算复杂度可分为两部分, 一是计算节点指标值的复杂度, 二是迭代式扩展算法的复杂度。在前期工作中已经分析了在全网中计算度、介数和客户锥的计算复杂度分别为 $O(V+E)$、$O(VE)$ 和 $O(V^2+VE)$ [25], 可作为第一部分的计算复杂度。对于第二部分的复杂度, 在单区域失效的情况下, 每轮次扩展只需要对候选节点按指标值大小进行排序即可, 假设候选节点数量为 V_C, 则扩展一次的计算复杂度为 $O(V_C\log_2 V_C)$。最坏的情况下, V_C 为全网除了异常区域以外的其他节点数, 可近似为 V。假设扩展轮次为 F ,则最坏的情况下迭代式扩展的计算复杂度为 $O(FV\log_2 V)$。对于多区域失效的情况, 基于第 2.5 节的分析, 相比于单区域失效的算法还需要增加对候选节点是否有交集的判断, 计算复杂度最坏为 $O(FV\log_2 V+FV)$。综上, RRAI_Degree、RRAI_Betweenness 和 RRAI_Customer-cone 算法的计算复杂度分别近似为 $O(V+E+FV(\log_2 V+1))$、$O(VE+FV(\log_2 V+1))$ 和 $O(V^2+VE+FV(\log_2 V+1))$。

对不同识别方法的计算复杂度梳理总结如表 2 所示。可以看出, RRAI 的计算复杂度显著低于基于级联失效模型的方法。

表 2 不同识别方法计算复杂度

Table 2 Computational complexity of different identification methods

方法	计算复杂度
IRS-CFM	$O(VE+N(V+E))$
CFM-VIRS-PN	$O(NV'^4)$
RRAI	$O(FV((k+1)T^2+\overline{Nei}T+\log_2 V+1))$
RRAI-NI	$O(VV_A(E_A+V_A+\overline{Nei})+V\log_2 V)$
RRAI_Degree	$O(V+E+FV(\log_2 V+1))$
RRAI_Betweenness	$O(VE+FV(\log_2 V+1))$
RRAI_Customer-cone	$O(V^2+VE+FV(\log_2 V+1))$

在实际实验中, 不同方法的平均运行时间如表 3 所示。由表 3 可知, RRAI 运行速度相比于基于级联失效模型的方法有了明显提升, 且随着网络规模增大, 速度提升倍数越大。当网络规模为 3000 时, RRAI 将运行时间从 10 小时提升至 1 分钟以内。虽然 RRAI 不如 RRAI-NI、RRAI_Degree 和 RRAI_Customer-cone 速度快, 但是结合准确率来看, RRAI 显然是兼顾效率和准确率的最佳方法。

表 3 不同识别方法平均运行时间

Table 3 Average running time of different identification methods

方法	网络规模			
	500	1000	2000	3000
IRS-CFM	00:00:38	00:01:19	00:05:35	00:11:54
CFM-VIRS-PN	00:04:51	00:32:06	03:35:48	10:18:33
RRAI	00:00:08	00:00:15	00:00:27	00:00:38
RRAI-NI	00:00:04	00:00:09	00:00:14	00:00:25
RRAI_Degree	00:00:01	00:00:02	00:00:04	00:00:07
RRAI_Betweenness	00:00:03	00:00:08	00:01:26	00:04:32
RRAI_Customer-cone	00:00:01	00:00:03	00:00:12	00:00:29

4 真实事件验证分析

由于级联失效模型的运行速度限制, 仿真实验仅在几千个节点的网络中进行。为了进一步说明 RRAI 在更大规模的实际网络中的应用能力, 本节以 2020 年 8 月 30 日 CenturyLink 数据中心错误配置导致的互联网中断事件为例说明 RRAI 在真实的大规模网络中应用的有效性。该事件是由单个 AS 上的 BGP 路由配置错误引发的, 造成众多互联网服务瘫痪, 全球 Web 流量下降了 3.5%[12]。

由于 CFM-VIRS 难以运行在如此大规模的网络中, 本节采用 Thousandeyes 公布的观测分析数据[26]作为 AS 受影响情况的基准来验证 RRAI 的有效性。该数据包含了事件发生过程中不同时间段内受影响的节点和链路。从中提取事件发生时间段内所有受影响的节点以及受影响链路涉及的节点作为基准区域节点集合。从 CAIDA 公开的 AS 商业关系数据库[20]中提取 2020 年 8 月的数据绘制 RRAI 以及其他方法需要的网络拓扑, 其中包含 69564 个节点。由于该事件是初始异常节点集中式分布的情况, 取 α 的值为 0.3。初始异常区域设置为引发配置错误的 AS3356 节点, 并根据第 3 节中的公式(7)计算 μ。

在第 3 节的对比的方法中, IRS-CFM 方法和 RRAI_Betweenness 方法由于计算复杂度高无法在全网拓扑中运行。对于 CFM-VIRS-PN 方法, 即便多次去除既无 Customer 又无 Peer 的节点, 最后得到的网络节点数量收敛于 13476 个, 依然无法在本文的实验环境下运行。因此仅将 RRAI 与其他三种方法进行对比。

首先对比随着目标区域节点数量增加, 不同方法准确率的变化, 实验结果如图 6 所示。相比于其他方法, RRAI 具有最高的准确率, 其准确率随着目标区域范围的增加有所波动, 其最低准确率为 71.4%, 最高准确率能够达到 81.2%, 相比于其他方法中的最高值, 准确率平均提升 11.5%。

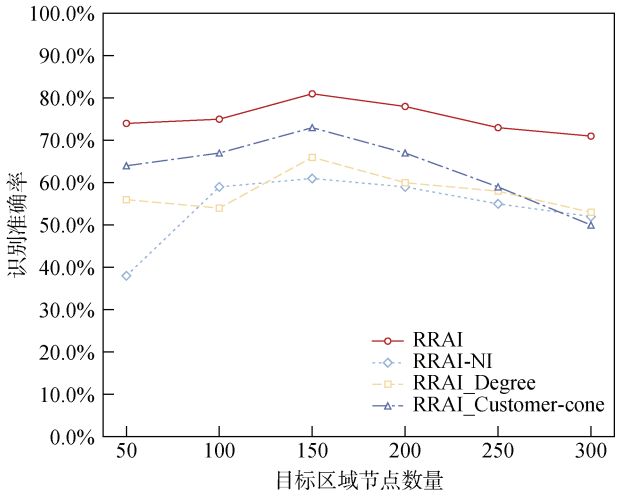


图 6 不同方法识别准确率随目标区域节点数量变化对比

Figure 6 Accuracy comparison among different methods varies with the number of nodes in the target area

为了进一步验证 RRAI 对支持异常响应的有效性, 本文对识别出的风险区域中节点的受影响程度进行分析。基于观测数据可知, 在所有受影响的节点中, 包含 32 个网络接口损坏的节点, 其余节点则是由于通过损坏接口与这些节点相互通信而受到间接的影响。因此, 识别出这些具有受损接口的节点对修复网络至关重要。本文通过设定不同的目标区域节点数, 对比不同方法识别出的风险区域包含具有受损接口的节点数, 实验结果如图 7 所示。相比于其他方法, RRAI 能够在设定最少的目标区域节点数量时包含所有受损接口的节点。在目标区域节点数量 T 为 50 时, 未包含在风险区域中的具有受损接口的节点仅有 4 个。此时 RRAI 对受损程度高的节点的识别准确率达到了 87.5%。通过计算所有节点在异常时间段内的平均受损接口数量, 发现未识别出的 4 个节点的平均受损接口数量在所有具有受损接口的节点中是最少的, 说明其在异常时间段的受损程度也相比其他节点更小。而当 T 为 100 时, RRAI 就能够识别出所有具有受损接口的节点。这说明即便限定的目标区域节点数量较小时, RRAI 仍然够有效筛选受损严重的节点, 可为异常事件的及时响应提供有力支持。

对比不同方法的运行时间如表 4 所示。综合图 6 的准确率来看, 虽然 RRAI_Degree 和 RRAI-NI 方法的速度最快, 但是其准确率均在 70%以下。而 RRAI 虽然不是最快的, 但是结合图 7 的结果, 当识别风险

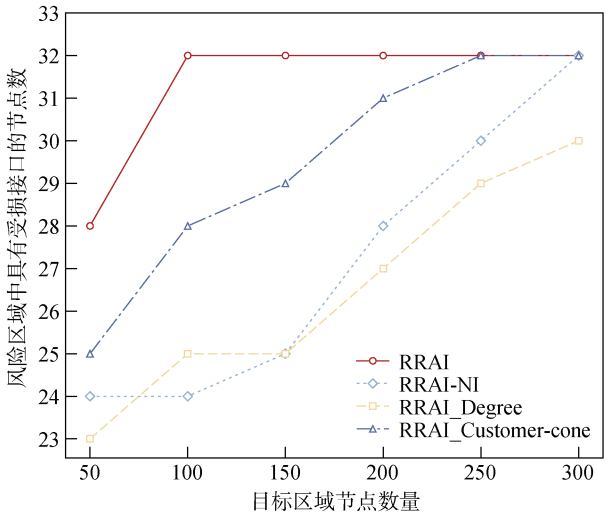


图 7　不同方法识别高受损节点数量随目标区域节点数量变化对比

Figure 7　Comparison of highly damaged node number among different methods varies with the number of nodes in the target area

节点数量小于 100 时, RRAI 方法能在 10 分钟内识别所有具有受损接口的节点, 其运行时间也是可接受的。并且, 级联失效导致的异常扩散也需要时间, 往往从初始异常到扩展为大规模异常需要 1 小时左右的时间[4]。因此在实际应用时可以先设置较小的目标区域范围, 对其进行及时控制, 然后随着观测的真实数据更新异常区域, 进而识别更大范围的风险区域。

5　讨论

基于第 3、4 节的实验可以验证 RRAI 与现有方法相比的优越性能, 以及其在真实域间路由网络中的有效性。为了进一步提升 RRAI 在实际网络中的可用性, 本节从前提条件、应用流程以及优化方式三个方面讨论 RRAI 在实际网络中的应用方法。

在实际现网中利用 RRAI 进行风险区域识别的前提条件有两个: 一是事先构建全网拓扑, 二是在 BGP 异常事件初期已经检测出网络异常并定位到初始异常区域。对于第一个条件, 本文基于 CAIDA 公开的最新 AS 商业关系数据[20]构建全网的网络拓扑, 该数据为每月更新。通过对 CAIDA 公开的 AS 拓扑关系数据进行统计分析发现, 2023 年来每个月的网络节点数量变化平均为 145.75。虽然多年来域间路由网络节点数量不断增长, 但是每个月平均增长的节点数量占全网规模的 0.19%, 相比于全网来说可忽略不计。因此本文假设异常事件发生时的全网拓扑与基于最新公布数据的全网拓扑一致。对于第二个条件, 目前已有大量关于 BGP 异常检测和定位的研究, 能够在异常初期快速检测和溯源, 相关方法已在引言中列举。对 BGP 的检测和定位并非本文研究的重点, 在此不再赘述。

表 4　不同识别方法在全球域间路由网络中的运行时间

Table 4　Running time of different identification methods in the global inter-domain routing network

方法	目标区域节点数量					
	50	100	150	200	250	300
RRAI	00:03:07	00:09:13	00:24:45	00:29:56	00:47:41	01:02:14
RRAI-NI	00:11:47	00:11:48	00:11:48	00:11:48	00:11:47	00:11:48
RRAI_Degree	00:00:09	00:00:15	00:00:31	00:00:42	00:00:54	00:01:02
RRAI_Customer-cone	00:38:53	00:38:58	00:39:12	00:39:22	00:39:36	00:39:48

应用 RRAI 识别风险区域的具体流程可分为三个步骤: 识别算法选择、关键参数设置和识别算法执行。首先判定初始异常区域是否连通, 如果该区域连通则选择单区域异常的风险区域识别算法进行识别, 否则选择多区域异常的风险区域识别算法。然后是对算法中关键参数的设置, 包括风险度计算时的控制参数 α、目标区域节点数量 T 以及扩展过程中的节点筛选比例 μ。基于第 3.3 节分析的结果, 对于单区域异常和多区域异常两种情况, α 分别取 0.3 和 0.4。通过第 4 节对真实事件的验证分析, 可以先将 T 设置为较小的数快速识别出风险区域, 进而对涉及到的 AS 的管理员进行告警, 以便尽快部署防御措施抑制异常传播。另外, 也可同时设置较大的 T 值尽可能多的识别具有风险的节点, 以便在异常传播后期进行全面的抑制。当设置好 T 值后, 可根据全网拓扑的直径大小和节点数量, 利用第 3.3 节的公式(7)计算节点筛选比例 μ。设置好上述参数后, 在事先构建好的网络拓扑上以初始异常区域为核心执行迭代式的风险区域识别算法, 就能识别该异常事件影响的风险区域。

由于 RRAI 以目标区域节点数量作为迭代的终止条件, 在提升识别速度的同时也会难以避免地造成识别准确率的下降。而基于本文给出的关键参数设置方法得到的识别结果可能并不是权衡速度与准确率的最优解。为了在实际应用中达到更好的识别

效果, 即在能够保证识别速度的情况下得到最优的准确率, 可以在网络拓扑已知的情况下, 基于当前的计算资源, 在不同的参数组合下, 随机选择一些初始异常区域进行仿真实验, 得到对应于该参数组合的平均识别速度。然后基于级联失效模型的仿真结果或以往的异常事件数据, 通过设定速度需求区间, 选择识别速度在该区间的参数组合进行识别, 从中选择具有最优识别准确率的参数组合。

6 结论

本文提出了针对 BGP 异常事件的风险区域快速识别方法 RRAI。考虑到识别风险区域就是要筛选出最可能受异常事件影响的高风险节点为网络管理员提供告警信息以进行提前防护, 本文将风险区域识别问题转化为节点相对于初始异常区域的风险评估问题。RRAI 包含两个核心要素: 节点风险评价指标和节点筛选算法。综合考虑 BGP 选路策略、AS 间商业关系以及 AS 节点的拓扑结构特征定义了节点风险评价指标“风险度”。针对初始异常节点位置的两种分布情况, 分别设计了单区域和多区域异常的风险区域识别算法, 以初始异常区域为中心迭代式筛选节点。在小规模网络上的仿真实验结果表明, RRAI 的识别准确率在不同规模网络中均优于现有方法, 且相比于基于级联失效模型的方法显著提升了识别速度。在基于全球网络上真实事件数据的实验中, RRAI 最高识别准确率可以达到 80%以上, 准确率平均提升 11.5%, 并且能够在可接受的时间内及时识别受损严重的节点为异常响应提供支持。

本文方法提供了一个全新的 BGP 异常影响风险区域识别思路, 后续可进一步考虑识别速度和准确率的权衡问题。可行的思路是以识别速度作为限制条件, 求解满足该速度并达到最优识别准确率的参数组合。另外, 目前提出的风险度指标主要是用于预测容易受影响的节点。后续可通过刻画节点抑制异常传播的能力定义其他的节点相对异常区域的重要性指标, 从而筛选出控制异常传播的关键节点。

参考文献

[1] Rekhter Y, Li T, Hares S. A Border Gateway Protocol 4 (BGP-4). RFC 4271. 2006.

[2] Wang N, Du X H, Wang W J, et al. A Survey of the Border Gateway Protocol Security[J]. *Chinese Journal of Computers*, 2017, 40(7): 1626-1648.
(王娜, 杜学绘, 王文娟, 等. 边界网关协议安全研究综述[J]. *计算机学报*, 2017, 40(7): 1626-1648.)

[3] McDaniel T, Smith J M, Schuchard M. The Maestro Attack: Orchestrating Malicious Flows with BGP[C]. *Security and Privacy in Communication Networks*, 2020: 97-117.

[4] Qiu H, Zhu H H, Li Y F, et al. FD-SP: A Method for Predicting Cascading Failures of Inter-Domain Routing System[C]. *2018 IEEE 4th International Conference on Computer and Communications*, 2018: 290-295.

[5] Muosa A H, Ali A H. Internet Routing Anomaly Detection Using LSTM Based Autoencoder[C]. *2022 International Conference on Computer Science and Software Engineering*, 2022: 319-324.

[6] Shapira T, Shavitt Y. BGP2Vec: Unveiling the Latent Characteristics of Autonomous Systems[J]. *IEEE Transactions on Network and Service Management*, 2022, 19(4): 4516-4530.

[7] Shapira T, Shavitt Y. SASA: Source-Aware Self-Attention for IP Hijack Detection[J]. *IEEE/ACM Transactions on Networking*, 2022, 30(1): 437-449.

[8] Ulmer A, Sessler D, Kohlhammer J. ProBGP: Progressive Visual Analytics of Live BGP Updates[J]. *Computer Graphics Forum*, 2021, 40(3): 37-48.

[9] Youn J, Kim K, Kang D, et al. Research on Cyber ISR Visualization Method Based on BGP Archive Data through Hacking Case Analysis of North Korean Cyber-Attack Groups[J]. *Electronics*, 2022, 11(24): 4142.

[10] BGPMon, https://bgpmon.net/blog/. Apr.15 2023.

[11] Thousandeyes. https://www.thousandeyes.com/. Apr. 2023.

[12] CenturyLink Level 3 Outage Analysis. https://www.thousandeyes.com/blog/centurylink-level-3-outage-analysis. Sept. 2020.

[13] Zhou F, Xu X, Trajcevski G, et al. A Survey of Information Cascade Analysis: Models, Predictions, and Recent Advances[J]. *ACM Computing Surveys*, 2021, 54(2): 1-36.

[14] Xu X, Zhou F, Zhang K P, et al. CasFlow: Exploring Hierarchical Structures and Propagation Uncertainty for Cascade Prediction[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(4): 3484-3499.

[15] Gao X L, Peng M F, Tse C K. Cascading Failure Analysis of Cyber–Physical Power Systems Considering Routing Strategy[J]. *IEEE Transactions on Circuits and Systems II*: *Express Briefs*, 2023, 70(1): 136-140.

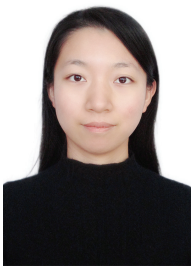
[16] Yang B, Zhang Y Q, Lu Y L. A New Methods for Cascading Failures Analysis in Inter-Domain Routing System[C]. *2015 Fifth International Conference on Instrumentation and Measurement*, *Computer*, *Communication and Control*, 2015: 382-385.

[17] Zhu H H, Qiu H, Wang Q X, et al. Double Damage Factor Based Inter-Domain Routing System Cascading Failure Model[J]. *Computer Engineering and Applications*, 2019, 55(2): 92-99.
(朱会虎, 邱菡, 王清贤, 等. 基于双毁伤因素的域间路由系统级联失效模型[J]. *计算机工程与应用*, 2019, 55(2): 92-99.)

[18] Zhao W D, Wang Y J, Xiong X L, et al. IKN-CF: An Approach to Identify Key Nodes in Inter-Domain Routing Systems Based on Cascading Failures[J]. *Entropy*, 2021, 23(11): 1456.

[19] Zhang J, Wang Y J, Zhang J Y, et al. Cascading Failure Model for Inter-Domain Routing System Based on Optimal Valid

Path[J]. *Netinfo Security*, 2021, 21(5): 90-99.
(张俊, 王永杰, 张敬业, 等. 基于最优有效路径的域间路由系统级联失效模型[J]. *信息网络安全*, 2021, 21(5): 90-99.)
[20] The CAIDA AS Relationships Dataset. http://www.caida.org/data/active/as-relationships/. Apr. 2023.
[21] Giakatos D P, Kostoglou S, Sermpezis P, et al. Benchmarking Graph Neural Networks for Internet Routing Data[C]. *The 1st International Workshop on Graph Neural Networking*, 2022: 1-6.
[22] Hoarau K, Tournoux P U, Razafindralambo T. Suitability of Graph Representation for BGP Anomaly Detection[C]. *2021 IEEE 46th Conference on Local Computer Networks*, 2021: 305-310.
[23] Luconi V, Vecchio A. Impact of the First Months of War on Routing and Latency in Ukraine[J]. *Computer Networks*, 2023, 224: 109596.
[24] Ganin A A, Quach P, Panwar M, et al. Multicriteria Decision Framework for Cybersecurity Risk Assessment and Management[J]. *Risk Analysis*, 2020, 40(1): 183-199.
[25] Liu Z M, Qiu H, Guo W, et al. NIE-GAT: Node Importance Evaluation Method for Inter-Domain Routing Network Based on Graph Attention Network[J]. *Journal of Computational Science*, 2022, 65: 101885.
[26] Internet Insights Snapshot 08/30/2020 09:30 UTC. https://pxakqbup.share.thousandeyes.com/. Sept. 2020.
[27] ASRank. https://asrank.caida.org/. Apr. 2023.
[28] Zhang G Q, Zhang G Q. An Algorithm for Internet AS Graph Betweenness Centrality Based on Backtrack[J]. *Journal of Computer Research and Development*, 2006, 43(10): 1790-1796.
(张国强, 张国清. 基于回溯机制的互联网 AS 拓扑的 Betweenness 算法[J]. *计算机研究与发展*, 2006, 43(10): 1790-1796.)



刘自勉 于 2018 年在信息工程大学网络工程专业获得学士学位。现在信息工程大学网络空间安全专业攻读博士学位, 研究领域为网络安全评估、域间路由安全。Email: mmian1314 @163.com

邱菡 于 2008 年在信息工程大学通信与信息系统专业获得博士学位。现任信息工程大学教授, 硕士生导师。研究领域为域间路由安全、网络安全模拟与评估。Email: qiuhan410@aliyun.com

王瑞 于 2014 年在解放军电子工程学院网络工程专业获得学士学位, 现在信息工程大学电子信息专业攻读硕士学位, 研究领域为网络威胁态势、域间路由安全。Email: 598251783 @qq.com

朱俊虎 于 2013 年在信息工程大学计算机软件与理论专业获得博士学位。现任信息工程大学教授, 博士生导师。研究领域为网络对抗、网络安全测试与评估。Email: zhujunhu74@163.com

王清贤 于 1982 年在北京大学计算机科学技术专业获得硕士学位。现任信息工程大学教授, 博士生导师。研究领域为网络安全。Email: wqx196008@163.com