

基于遗传算法的服务功能链贪心备份方法研究

杨雪, 陈卓

(重庆理工大学计算机科学与工程学院, 重庆 400054)

摘要: 在网络功能虚拟化(NFV)环境中, 虚拟网络功能及其基础设施的故障可能会严重损害服务功能链的可用性, 并影响整体网络性能, 成为NFV架构中的核心挑战之一。为了应对这一难题, 提出一种基于遗传算法的服务功能链贪心备份方法。首先, 提出一种基于生成对抗网络(GAN)的边缘节点故障预测模型, 以预测节点的健康状况; 其次, 优先选择能够显著提升系统可用性的VNF作为备份对象, 从而减少所需的备份VNF数量; 最后, 利用遗传算法的迭代优化, 为需要备份的VNF部署最优的节点和链路, 提出一种确保服务功能链高可用性的备份方法。仿真结果表明, 相比KSP方法, 该方案将资源消耗降低32%, 服务链请求接受率提升11%, 平均备份比例稳定维持在0.5。该方法通过“预测—筛选—优化”的闭环机制, 显著提升了动态网络环境下服务链的鲁棒性与资源利用效率。

关键词: 故障预测; 遗传算法; 虚拟网络功能; 可用性; 备份

DOI: 10.11907/rjdk.251100

中图分类号: TP393

文献标识码: A

文章编号: 1672-7800(2026)004-0140-08

扫描二维码阅读全文:



Research on Greedy Backup Method of Service Function Chain Based on Genetic Algorithm

YANG Xue, CHEN Zhuo

(School of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: In network function virtualization (NFV) environments, failures of virtual network functions (VNFs) and their infrastructure can severely compromise the availability of service function chains (SFCs) and degrade overall network performance, posing a critical challenge in NFV architectures. To address this issue, this paper proposes a genetic algorithm-based greedy backup method for SFCs. First, a Generative Adversarial Network (GAN)-enabled edge node fault prediction model is developed to assess node health status, thereby avoiding the use of high-risk nodes for backup. Second, a greedy strategy prioritizes VNF instances that maximize system availability improvements while minimizing resource consumption, ensuring efficient backup selection. Third, leveraging iterative optimization via genetic algorithms, the method optimally deploys backup VNFs across nodes and links to guarantee SFC high availability. Simulation results demonstrate that, compared to the K-Shortest Path (KSP) method, the proposed approach reduces resource consumption by 32%, increases SFC request acceptance rates by 11%, and maintains an average backup ratio of 0.5. By integrating a closed-loop "prediction-sifting-optimization" mechanism, this method significantly enhances SFC robustness and resource utilization efficiency in dynamic network environments.

Key Words: failure prediction; genetic algorithm; virtual network functions; availability; backup

0 引言

在网络功能虚拟化环境(Network Function Virtualization, NFV)中, 每一项业务请求都通过一系列虚拟网络功能(Virtual Network Functions, VNF)来处理。这些VNF各自承担不同的网络服务任务, 并按照预定的序列相互连

接, 从而构成了一条服务功能链(Service Function Chain, SFC, 简称服务链)。一条服务链的流量需按照预定义的顺序经过一系列指定的VNF^[1-2]。SFC将VNF组合成链式结构, 可以根据具体需求灵活地定制不同的服务链, 同时, 支持按需增添或移除VNF, 使得网络中的VNF可以根据实际需要进行动态调整和扩展。此外, SFC通过定义和管理网络服务链的顺序和流量处理规则, 确保流量能够按

收稿日期: 2025-03-11

作者简介: 杨雪(2000-), 女, CCF学生会会员, 重庆理工大学计算机科学与工程学院硕士研究生, 研究方向为网络虚拟化; 陈卓(1980-), 男, 博士, 重庆理工大学计算机科学与工程学院副教授, 研究方向为边缘计算、物联网及区块链技术等。本文通讯作者: 杨雪。

照指定的顺序和路径通过网络中的各个 VNF 进行处理和转发,实现复杂的网络服务需求,如安全服务链、加密服务链等^[3]。

然而,网络服务的可用性可能会受到网络功能的影响,这可能导致服务中断甚至整个网络的瘫痪。这些故障通常由承载网物理主机(Physical Machine, PM)的硬件故障、虚拟机(Virtual Machine, VM)故障或 VNF 的软件缺陷引起,从而影响 SFC 的正常运行,降低其可用性^[4-5]。这种情况会导致一定程度的数据丢失和资源浪费,是 NFV 使能网络中一个亟待解决的关键问题。

1 相关研究

为解决此问题,已有研究提出了 VNF 冗余备份机制。文献[6]通过将网络功能放置与链化问题建模为最小成本流问题,提出了一种基于分数解的优化策略,能够从复杂的配置方案中筛选出低成本路径,为静态或半静态网络环境下的资源分配提供数学上的全局最优解。但静态策略在高度动态的 5G 网络中可能面临计算效率不足的问题,难以适应实时变化的网络状态。文献[7]则聚焦于异构边缘环境中的多目标协同优化,通过博弈论方法平衡服务功能链部署与资源管理的矛盾,以常数近似比例实现总系统延迟的最小化,解决了边缘节点异构性与动态需求适配的难题。该方法极大地降低了理论上的延迟上限,但所产生的计算复杂度可能随网络规模呈指数级增长,限制了其在大规模边缘节点部署中的实际应用。针对 VNF 需求波动及服务链失败率的不确定性,文献[8]和[9]引入了在线学习算法,通过动态探索与利用的权衡机制,实时调整备份策略,显著增强了系统对突发流量和故障的鲁棒性。文献[10]提出了一种主备 VNF 映射的在线业务交换方法,综合考虑了排队延迟、通信延迟与交换延迟的协同影响。通过多维度延迟建模优化备份路径选择,提升了复杂网络环境下的服务连续性保障能力。

以上方法大多基于网络处于理想状态的假设,未充分考虑实际环境中因硬件老化、链路拥塞或突发负载激增导致的节点故障风险。这种理想化设定限制了备份策略的实用性,尤其在动态复杂的 5G 网络场景中,节点失效可能引发服务链中断,难以满足高可用性需求。针对这一缺陷,本文提出基于对抗生成网络(Generative Adversarial Network, GAN)的边缘节点故障预测模型,更为精准的预测节点的健康状况,在此基础上将 VNF 备份到不易故障的节点,延长其寿命。同时,提出基于遗传算法的服务功能链贪心备份方法(Greedy Genetic Algorithm-based Backup Method for Service Function Chains, GGA-VNF)。该方法在降低资源消耗的同时强化服务的鲁棒性,核心策略是针对服务链中潜在可用性提升更大的部分,实施精细化的实例重建,维持服务的连续性和可用性。

2 边缘节点故障预测模型

在边缘计算网络中,导致服务器故障的主要原因大致分为硬件故障、软件问题、网络异常及外部因素 4 种^[11]。其中,硬件故障主要归结为物理节点磁盘故障。基于此,本文提出了一种基于 GAN 的边缘节点故障预测模型,如图 1 所示。

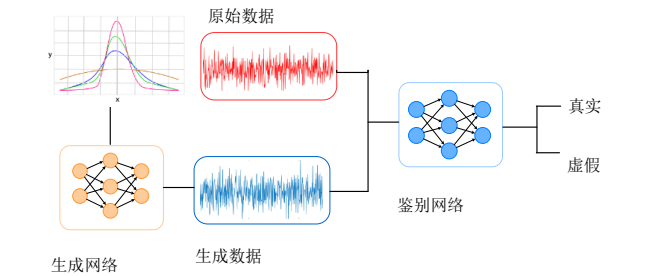


Fig. 1 Overview of the GAN model
图 1 GAN 模型概述图

2.1 故障预测

硬盘的自我监控、分析和报告技术(Self-Monitoring, Analysis, and Reporting Technology, SMART)特征数据是维护磁盘健康和预测潜在故障的关键。通过持续监控硬盘的多个关键指标来评估其状态^[12-13],包含读取错误率、寻道错误率、磁盘旋转启动/停止次数以及温度等。每当硬盘运行时,这些指标被定期记录下来,随着时间的推移,形成了相应的时间序列。其中,每个时间点的数据反映了磁盘在该时刻的状态。这组序列不仅提供了对当前磁盘健康状况的洞察,还能帮助检测异常行为。但利用磁盘 SMRAT 数据进行磁盘故障预测时,面临严重的数据不平衡和不丰富的问题,即正常数据远多于故障数据,如表 1 所示。

Table 1 ST4000DM000 disk SMRAT data classification
表 1 ST4000DM000 磁盘 SMRAT 数据分类

时间	故障磁盘数量	正常磁盘数量
2023-Q1	170	1 633 792
2023-Q2	167	1 607 661
2023-Q3	130	1 611 488
2023-Q4	93	1 295 836

由表 1 可知,正常运行的磁盘数量远超过故障磁盘的数量,正常磁盘数量可达 16 万,但故障磁盘数量却仅有 200 不到。数据不平衡和不丰富的问题很大程度上影响了机器学习模型的训练效果和预测准确度。

利用生成对抗网络 GAN 解决数据不平衡和不丰富的问题。GAN 包括 2 个主要部件:生成器和判别器^[14]。生成器将随机噪声(通常是从高斯分布中采样得到)转换为逼真的数据样本;判别器通过输出 0、1 来判别输入样本是来自真实数据集还是生成数据集。通过二者相互竞争的动态博弈过程来学习数据的分布,从而生成可靠且新的样本

数据。

从最初的 SMART 数据中选取与磁盘故障相关率较高的数据,即 SMART 5、9、187、188、193、194、197、198、241、242 作为特征^[15-16]。数据进行预处理,包括移除异常值和补充缺失信息,并将这些处理后的数据放入 GAN 实现数据生成。

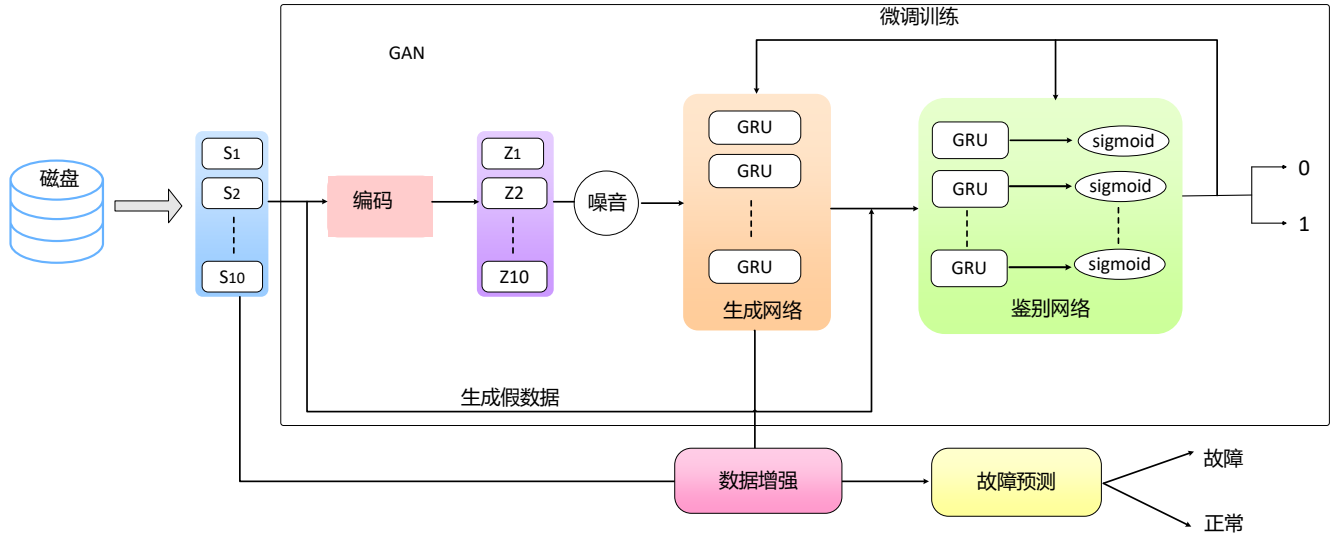


Fig. 2 Edge node fault prediction model based on GAN

图2 基于GAN的边缘节点故障预测模型

假正率(False Positives Ratio, FPR):实际为负例(真实标签为负)的样本中,被错误地预测为正例的比例。具体如公式(1)所示:

$$FPR = \frac{FP}{TN + FP} \quad (1)$$

真正率(True Positives Ratio, TPR):真正例占有所有实际正例样本的比例。具体如公式(2)所示:

$$TPR = \frac{TP}{TP + F_n} \quad (2)$$

准确率(Accuracy):所有预测正确的样本占总样本数的比例。具体如公式(3)所示:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

3 备份问题描述与网络模型

3.1 问题描述

在网络中,系统时刻接收来自众多用户的多样化请求。当用户提出请求时,系统会首先检查该请求中所指定的网络功能是否已在相应位置部署并就绪。此外,系统还需验证这些已部署的功能是否满足请求中明确指出的可用性指标。若此时用户请求的可用性不满足需求,则会对该请求进行备份,并逐步提高其可用性以满足需求。如果上述条件均得以满足,则系统会正式接受并处理用户的请求。

在备份过程中,VNF 备份问题的核心在于确定请求中

2.2 性能指标

故障预测模型结构,如图2所示。评估故障预测模型的性能涉及多项指标,这些指标基于混淆矩阵的4个基本部分:真正例(True Positives, TP)、真反例(True Negatives, TN)、假正例(False Positives, FP)和假反例(False Negatives, FN)。

的哪些 VNF 需要备份,以及将这些需要备份的 VNF 部署在哪个节点上。

3.2 网络模型

在多接入边缘计算(Multi-access Edge Computing, MEC)系统中,每个接入点(Access Point, AP)依附于一个边缘服务器。在边缘服务器上映射不同的网络功能,将特定顺序的 VNFs 链到一个 SFC 上,每个用户可以根据特定的可用性要求,请求一个服务功能链。系统宏观运作周期被细分为一系列离散的时间框架,记为 $T=\{1, 2, \dots, |T|\}$,每个数字代表一个独立的时间时隙。本文针对此场景将网络系统抽象为以下几个模型:

(1)物理网络:用一个赋权无向图 $G=(V, E)$ 表示,其中, V 表示物理节点集合,即每个 AP 粘附的边缘服务器, $V=\{1, \dots, v, \dots, |V|\}$ 。对于每个物理节点 $v \in V$,都关联着一组资源属性,使用二元组 (C_v, S_v) 分别表示计算和存储容量。 E 表示 v 连接 AP 的物理链路的集合,对于每条物理链路 $(v, v') \in E$ 都有一定的带宽容量 $B_{(v, v')}$ 。

(2)SFC 请求:每个时隙接收的 SFC 请求表示为 $Q(t)=\{1, \dots, q, \dots, |Q(t)|\}$ 。对于每个 SFC 请求 $q \in Q(t)$ 都表示一个特定的 SFC 请求,包括 F_q, a_q 两个关键组件。 F_q 表示 q 特定的 VNF 集合, $F_q = \{1, \dots, f, \dots, |F_q|\}$,是系统所有 VNF 类型集合的子集, q 中每个 VNF 被分配到不同的物理节点上,使用三元组 $(c_{(q, f)}, s_{(q, f)}, b_{(q, f)})$ 分别表示每个 VNF 所需的计算、存储容量、带宽容量。 a_q 表示对 SFC 请求 q 的可用

性需求。

3.3 SFC 请求可用性分析

每个 VNF 的可用性分为 2 个部分:软件可用性和其承载节点的硬件可用性^[15]。因此,VNF 的可用性可定义为 $R = R_f \cdot R_v$ 。其中, R_f 表示 NFV 的可用性, R_v 表示其承载节点的可用性。SFC 请求是由特定顺序的 VNF 链接组成,其可用性由 VNF 决定。给定一个 SFC 请求 q , 有 $|F_q|$ 个 VNF, 则 SFC 请求的可用性定义如公式(4)所示:

$$R_q = 1 - \prod_{i=1}^{|F_q|} (1 - R_i) \quad (4)$$

为了提升服务可用性并确保 VNF 故障时服务不中断,系统引入了冗余备份机制,如图 3 所示。在主 VNF 实例执行的同时进行备份,提前为备份实例分配资源并采用并联方式连接。当主 VNF 实例失效时,立即启用备份 VNF 实例恢复服务。在冗余 VNF 备份机制下,SFC 请求的可用性定义如公式(5)所示:

$$R_q = \prod_{f=1}^{|F_q|} [1 - (1 - R_p)(1 - R_b)] \quad (5)$$

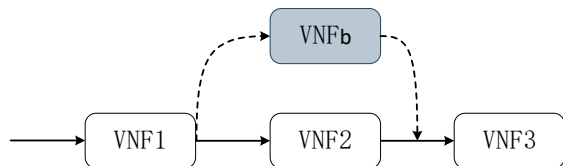


Fig. 3 Example of service function chain backup protection

图3 服务功能链备份保护示例

4 系统模型

4.1 变量与目标函数

引入二进制变量 $y_{(q,f)}$ 表示 SFC 请求 q 的 VNF f 是否需要备份,如公式(6)所示:

$$y_{(q,f)}(t) = \begin{cases} 0 \\ 1 \end{cases} \quad (6)$$

其中, $y_{(q,f)}(t) = 0$ 表示 SFC 请求 q 的 VNF f 不需备份; $y_{(q,f)}(t) = 1$ 表示 SFC 请求 q 的 VNF f 需要备份。

引入二进制变量 $z_{(v,v')}$ 表示 SFC 请求 q 的 VNF f 的流量走向,如公式(7)所示:

$$z_{(v,v')}(t) = \begin{cases} 0 \\ 1 \end{cases} \quad (7)$$

其中, $z_{(v,v')}(t) = 0$ 表示 SFC 请求 q 的 VNF f 的流量未通过链路 (v, v') ; $z_{(v,v')}(t) = 1$ 表示 SFC 请求 q 的 VNF f 的流量通过链路 (v, v') 。

为了优化系统性能并实现资源的最优配置,本研究确定了一个核心优化目标:在每个时间时隙 t 中,根据各项约束条件,力求最小化 VNF 备份资源的消耗,具体如公式(8)——(10)所示:

$$C(t) = \min(C_{(v,q)}(t) + C_{(e,q)}(t)) \quad (8)$$

$$C_{(v,q)}(t) = \sum_{q \in Q(t)} \sum_{f \in F_q} (c_{(q,f)} + s_{(q,f)}) y_{(q,f)}(t) \quad (9)$$

$$C_{((v,v'),q)}(t) = \sum_{q \in Q(t)} \sum_{f \in F_q} \sum_{v,v' \in V} b_{(q,f)} \cdot z_{(v,v')} \quad (10)$$

其中, $C_{(v,q)}(t)$ 表示 SFC 请求 q 在备份过程中所消耗的节点资源; $C_{((v,v'),q)}(t)$ 表示 SFC 请求 q 在备份过程中所消耗的链路带宽。

4.2 约束条件

为确保仅当 VNF 被选择作为备份,VNF 才需备份引入约束式,如公式(11)所示:

$$\sum_{v \in V} x_{b(q,f,v)}(t) = y_{(q,f)}(t), f \in F_q \quad (11)$$

为说明请求 q 的备份 VNF 需映射到不同的物理节点上,且只存在一个引入约束式,如公式(12)所示:

$$\sum_{f \in F_q} x_{b(q,f,v)}(t) \leq 1, v \in V \quad (12)$$

为保证请求 q 消耗的主 VNF 及备份 VNF 资源不超过节点及链路总量,引入约束式,如公式(13)——(15)所示:

$$\sum_{q \in Q(t)} \sum_{f \in F_q} [x_{b(q,f,v)}(t) + x_{p(q,f,v)}(t)] \cdot c_{(q,f)} \leq C_v, \forall v \in V \quad (13)$$

$$\sum_{q \in Q(t)} \sum_{f \in F_q} [x_{b(q,f,v)}(t) + x_{p(q,f,v)}(t)] \cdot s_{(q,f)} \leq S_v, \forall v \in V \quad (14)$$

$$\sum_{q \in Q(t)} \sum_{f \in F_q} [x_{b(q,f,v)}(t) + x_{p(q,f,v)}(t)] \cdot b_{(q,f)} \leq B_{(v,v')}, \forall v, v' \in V \quad (15)$$

为确保请求 q 的可用性,满足用户的可用性需求,引入约束式,如公式(16)——(18)所示:

$$1 - (1 - R_p)(1 - R_b) \geq a_q \quad (16)$$

$$R_p = \prod_{f=1}^{|F_q|} x_{p(q,f,v)}(t) \cdot R_f \cdot R_v \quad (17)$$

$$R_b = \prod_{f=1}^{|F_q|} x_{b(q,f,v)}(t) \cdot R_f \cdot R_v \quad (18)$$

5 基于遗传算法的 SFC 可用性保障备份方法

5.1 遗传算法

遗传算法 (Genetic Algorithm, GA) 是一种启发式搜索算法,它通过模拟自然界中生物进化的过程来寻找优化问题的解^[17]。遗传算法在一个由多个候选解组成的集合(称为种群)中进行搜索,且每个个体都需要通过某种方式编码成一个特定的形式。根据个体的适应度,决定哪些个体应该被选中来生成下一代。它通过组合 2 个或多个体的部分特征来创建新的个体,或以很小的概率随机改变某些个体中的某些特征,以避免算法过早地收敛到局部最优解,并增加种群的多样性。

5.2 备份 VNF 选择

过多冗余 VNF 的加入很大程度上增加了 SFC 请求的长度,且每个额外的 VNF 实例都需要占用网络和计算资源。因此,系统在运行时会选择对潜在可用性提升较大的关键 VNF 进行备份,在提高 SFC 可用性的同时有效降低了资源消耗。

在可用性理论中,潜在提升值用于衡量通过提高某一组件的可用性所能实现的系统总体可用性的最大提升程度^[18-19]。在服务链中,为任何一个 VNF 提供备份可能都

会提升整条链路的可用性。SFC 中 VNF f 的潜在提升值定义,如公式(19)所示:

$$I_{(q,f)} = r_q^f - r_q \quad (19)$$

其中, r_q^f 表示对 SFC 请求 q 新增一个 VNF f 备份后的软件可用性; r_q 则表示增加之前 SFC 请求 q 的软件可用性。研究表明,在 SFC 中,选择能提供最大可用性增量的 VNF 最能有效提升整个链路的可靠性^[20]。因此,在所有无法满足可用性要求的 SFC 中,应选择 $I_{(q,f)}$ 更大的 VNF 进行备份。

5.3 SFC 可用性保障备份方法

为了尽可能降低系统中用于备份的资源消耗,利用本章第二节所提到的 VNF 选择策略,对 VNF 进行备份,但仍未解决 VNF 的部署问题。因此,本文提出了基于遗传算法的 SFC 可用性保障备份方法。根据网络结构和服务功能链的需求,利用该方法进行备份,直至满足服务功能链的全部要求。

在该方法中,所有备份 VNF 的部署位置被看作一个个体,而个体中的基因则表示每一个备份 VNF 的部署位置。在个体备份过程中,若只考虑资源消耗表示该个体对当前环境的适应度,则 SFC 的可用性提升较小。因此,将两者结合考虑,即个体的可用性不满足 SFC 可用性需求时,对适应度进行惩罚。首先,在已满足主 VNF 资源需求的 SFC 中,计算其可用性。如果 SFC 请求满足可用性需求,则接受该请求,无需部署备份;如果不满足,则根据上文所述选择策略选择备份 VNF。其次,将所有需要备份的 VNF 作为遗传算法的个体,并基于约束式(11)–(18)创建初始种群,计算各个体的适应度,并执行遗传算法运算。当遗传算法迭代结束后,系统会再次计算 SFC 的可用性。如果仍存在不满足可用性需求的 SFC,将对其再次添加备份 VNF 并重复执行上述操作,直至所有 SFC 均满足可用性需求。

算法 1 基于遗传算法的 SFC 贪婪优化备份方法

输入: $G = (V, E)$, $Q(t)$

输出: $x_{b(q,f,v)}(t)$, $C(t)$

1. $need_q(t) = 1(q \in Q(t), |Q(t)|)$; //表示 sfc q 是否需要备份

2. $while \sum(need_q(t)) > 0$

3. $For q = 1: |Q(t)|$

4. $if need_q(t) = 1 \text{ and } R_q \leq a_q$

5. $For f = 1: |F_q|$

6. 计算每个 VNF 的 $I_{(q,f)}$

7. 选择潜在可用性提高值最大的 VNF $y_{(q,f)}(t) = 1$

8. $end for$

9. $end if$

10. $end for$

11. 初始化种群: 给定种群规模 N , 根据 $y_{(q,f)}(t)$, 随机生成每个个体的基因组成 $x_{b(q,f,v)}(t)$

12. $for \text{gen} = 1: 20$ //迭代进行选出最优个体

13. 用目标函数 $C(t)$ 计算每个个体适应度(fitness)

14. 使用最佳适应度选择策略选择的个体作为父母种群

15. 从父代群体中随机选取配对个体, 执行单点交叉操作, 从而生成新的子代个体

16. 更新种群

17. 重新计算适应度得出最优个体 $x_{b(q,f,v)}(t)$

18. $For q = 1: |Q(t)|$

19. $if R_q \geq a_q$

20. $need_q(t) = 0$

21. $end while$

该算法的时间复杂度主要由 SFC 请求规模 $|Q(t)|$ 、VNF 数量 $|F_q|$ 及遗传算法参数共同决定。外层 while 循环在最坏情况下需遍历所有待处理的 SFC 请求, 其执行次数与未满足可用性需求的请求数量呈线性关系。在每次循环中, 每个 SFC 的 VNF 集合需要进行特征计算与关键节点选择, 时间复杂度为 $O(|Q(t)| * |F_q|)$ 。随后, 采用遗传算法进行优化。该过程包含种群初始化、20 次迭代的适应度评估(计算复杂度为 $O(N * |Q(t)|)$)以及选择交叉操作, 其中, N 为种群规模。综上, 算法整体复杂度可表示为 $O(K * N * |Q(t)|^2 * |F_q|)$, 其中, K 为外层循环次数。

6 实验结果及分析

本文分别对边缘节点故障预测模型以及可用性保障的 VNF 备份方法进行性能验证, 并对实验结果进行对比分析。

6.1 故障预测模型

本文基于 GAN 的边缘节点故障预测模型的预测效果进行分析, 构建了一个基于 pythonGAN 的神经网络, 输入的磁盘 SMART 数据来自 Blackblaze 公司真实生产环境中的公共数据集。输入层的维度为 10, 输出层的激活函数为 ReLu 函数, 隐藏层的激活函数为 sigmoid 函数, 学习率为 0.005, 最大迭代次数为 600。

将本文提出的基于 GAN 的边缘节点故障预测与 K 最近邻(KNN, K-nearest neighbor)、支持向量机(SVM, Support Vector Machine)进行比较, 具体如表 2 所示。

Table 2 Comparison of fault prediction by different algorithms

表 2 不同算法故障预测对比 (%)

	TPR	FPR
KNN	74.05	4.65
SVM	75.40	5.17
基于 GAN 的边缘节点故障预测	96.30	1.82

由表 2 可知, 相比 KNN 和 SVM, 本文算法在 TPR 上提高近 20%, 在 FPR 上降低近 3%。结果表明, GAN 模型通过生成对抗机制, 有效解决了小数据集场景下的数据稀疏性与模式学习不足的问题。KNN 依赖局部样本相似性进行故障分类, 但在小样本条件下, 稀疏的故障数据分布容易导致近邻搜索偏差。例如正常节点样本占据多数时, KNN

可能将少数故障节点的异常特征误判为噪声,从而降低TPR并抬高FPR。SVM虽通过核函数映射特征空间从而构建分类超平面,但在小数据集时,由于特征维度与样本量的不平衡,难以捕捉故障模式的非线性边界,导致其分类性能显著下降。相比之下,GAN通过生成器与判别器的对抗训练,能够从有限的历史数据中学习节点状态并生成与真实故障模式高度相似的合成样本。这种数据增强能力不仅扩充了训练集的多样性,还缓解了传统方法因样本不足导致的过拟合问题。

如图4所示,在不同特征数量下,基于GAN的边缘节点故障预测算法始终优于KNN和SVM方法。在低特征数(如3个特征)场景中,尽管所有算法准确率均未超过70%,但GAN的预测性能仍显著领先。这一优势源于GAN模型对潜在特征关系的学习能力与数据分布重构机制的协同作用。当特征数从3增至10时,GAN通过引入更多上下文信息(如历史负载趋势、丢包率波动周期等),进一步强化了故障模式的动态建模能力,使得准确率跃升至97.36%;而传统方法因依赖手工特征工程与固定模型架构,难以充分提取高维特征中的隐藏规律。此外,GAN的生成式特性使其对噪声与异常值具有更强的鲁棒性,即使部分特征缺失或失真,仍能通过分布生成机制补偿信息损失,从而在复杂实际场景中维持稳定的预测性能。

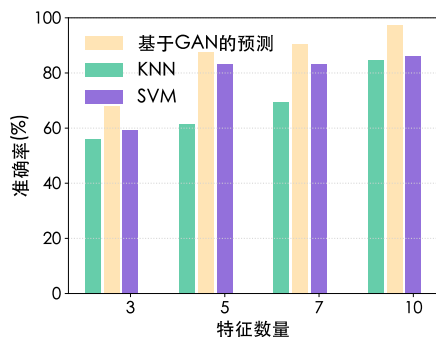


Fig. 4 Accuracy comparison under different number of features

图4 不同特征数量下准确率比较

6.2 实验场景与平台

本文实现了一个基于python的模拟器,使用Salam网络拓扑随机生成算法生成物理网络拓扑和SFC拓扑。设定2个物理网络Case1、Case2以验证算法的普适性。物理网络中分别包含15个边缘服务器节点和25个边缘服务器节点。每个节点的可用性水平在(0.97~0.99)随机生成。每个服务器节点拥有相同的资源数量,每条链路根据链接节点远近拥有带宽资源。网络中提供8种不同的VNF,每种VNF所需的资源随机分布在10~30内,可用性随机分布在0.99~0.999之间^[20]。每条SFC请求中包含2~5个VNF,每条请求的可用性从(0.96, 0.97, 0.98, 0.99)中随机选择,请求链越长则可用性越低。

6.3 仿真结果与分析

将本文提出的基于遗传算法的SFC贪婪优化备份方

法与K-最短路径(K-Shortest Path)算法进行比较,分析其优越性。经过多次试验验证,算法参数设置为初始种群大小100、交叉率0.5和变异率0.1时,能够有效平衡全局搜索能力与计算效率,其适应度函数值曲线,如图5所示。其中,初始种群大小100确保了候选解的多样性,避免陷入局部最优;交叉率0.5通过适度交换基因片段保留优质解特征,同时维持种群进化潜力;变异率0.1则在不破坏优势基因结构的前提下,引入可控随机扰动以增强算法对动态网络状态的适应性。

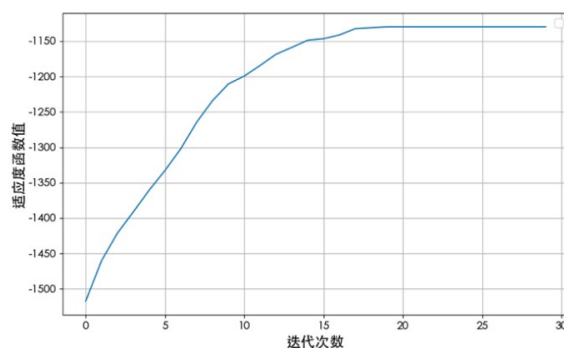


Fig. 5 Change curve of fitness function value of genetic algorithm

图5 遗传算法适应度函数值变化曲线

KSP算法按照主VNF的部署顺序选择备份VNF,并为需要备份的每个SFC寻找K条最短路径。通过评估每条路径的性能(包括可用性、资源消耗等),从而选择性能最佳的路径作为备份路径。

如图6所示,在Case1和Case2两种环境下,本文提出的GGA-VNF方法在网络请求量增加时,相较于KSP方法表现出了显著的优势。同时,相较于Case1环境,2种方法在Case2环境中的请求接受量均有所下降,KSP方法下降趋势更为明显。这种现象主要是由于网络规模扩大使得环境更为复杂,KSP方法在此种变化下执行备份VNF操作时难以选择出资源及可用性相匹配的路径,从而受到更大的影响。而GGA-VNF通过遗传算法的全局搜索能力,在解空间中探索备份路径与资源分配的多种组合,并利用贪心策略对候选解进行局部精细化调整。这种“全局探索-局部优化”的协同机制,使得GGA-VNF在请求量增加时仍能维持较高的请求接受率。实验数据显示,当请求量达到峰值时,GGA-VNF的接受率较KSP提升约11%,验证了其在复杂动态网络环境下保障服务可用性的优越性。

图7(a)(b)描述了不同网络服务请求数量下,备份数量的比例变化。

由图7可知,随着网络服务请求数量的增加,2种策略的备份数量比例均有所波动,备份数量比例计算,如公式(20)所示。无论在何种网络环境下,本文提出的GGA-VNF方法备份数量比例始终保持在0.5左右,显著低于KSP方法,有力的证明了GGA-VNF方法的普适性。GGA-VNF通过遗传算法的全局搜索能力,将备份问题建模为多

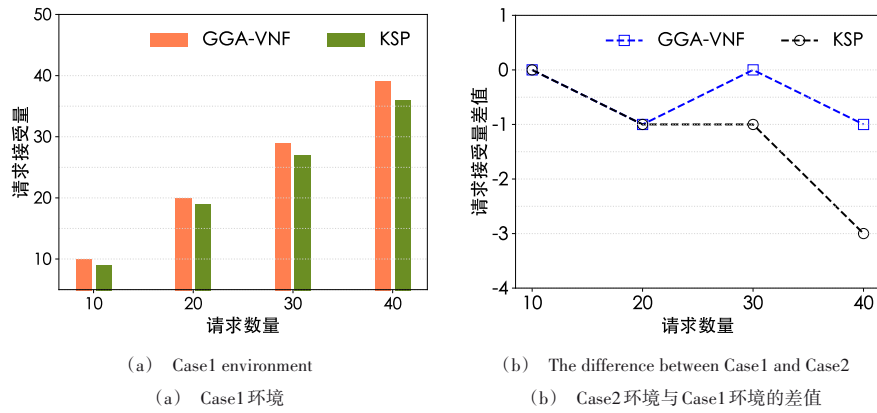


Fig. 6 Comparison of the number of requests accepted

图6 请求接受数量比较

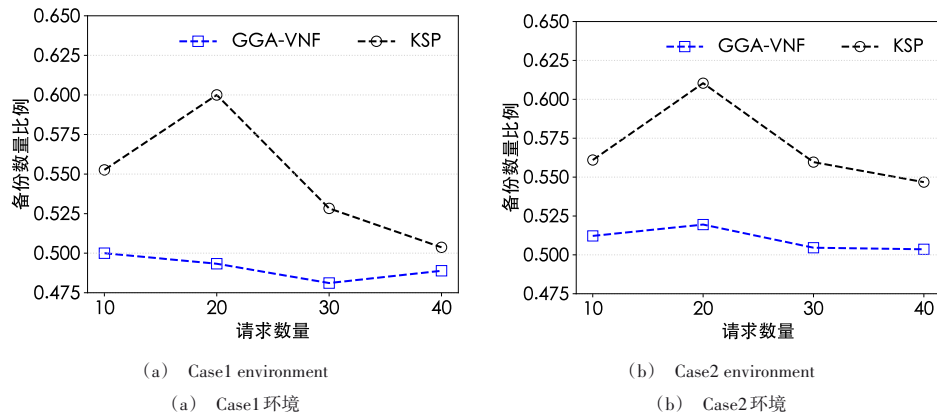


Fig. 7 Comparison of backup quantity ratio

图7 备份数量比例比较

约束优化任务,综合考量节点故障概率(来自GAN预测模型)、资源剩余量、服务链端到端延迟及负载均衡度,确保备份方案在降低冗余的同时最大化服务可用性。此外,GGA-VNF通过整合故障预测结果,主动规避潜在故障节点,从而减少因节点失效触发的重复备份需求。这种前瞻性策略与KSP的被动响应模式形成鲜明对比,使得备份实例的部署更具针对性。

$$\text{备份数量比例} = \frac{\text{备份VNF数量}}{\text{请求接受VNF数量}} \quad (20)$$

图8显示了2种方法在不同网络请求数量下的备份资源消耗。随着网络服务请求数量的增加,2种方法的备份资源消耗都在上升,这表明随着请求量的增加,系统需要消耗更多的备份资源来保障SFC的可用性,从而提高网络的请求接受率。Case2环境中系统节点增加导致备份数量

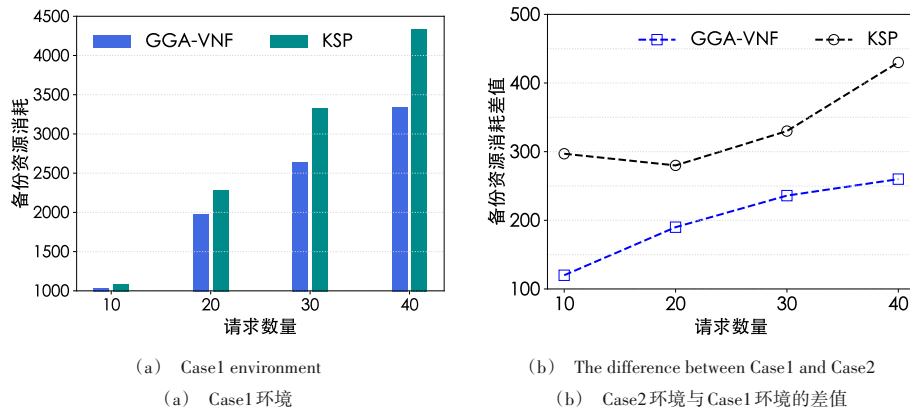


Fig. 8 Comparison of backup resource consumption

图8 备份资源消耗比较

增长,因此2种方法消耗的备份资源均多于Case1环境。但无论在何种请求量下,Case1与Case2环境中GGA-VNF的备份资源消耗始终低于KSP,其关键在于,GGA-VNF通过贪心策略动态量化每个VNF实例的“潜在可用性提升值”(指标定义为实例故障对服务链整体中断风险的影响权重与其资源消耗成本的比值),并优先选择比值最高的实例进行备份。这一机制能在解空间中高效筛选出对服务连续性贡献最大而资源消耗最小的备份组合。具体而言,遗传算法的交叉与变异操作生成多样化的备份候选方案,贪心策略则通过局部迭代快速剔除低效解(例如备份资源消耗高但可用性增益低的实例),并在每一代种群中保留资源-可靠性均衡的最优个体。实验数据显示,在相同请求量下,其资源消耗较KSP减少约32%。

7 结语

在网络功能虚拟化(NFV)环境中,服务功能链(SFC)的可用性与资源效率高度依赖于虚拟网络功能(VNF)的可靠部署。本文针对传统备份策略在动态网络场景下的局限性,提出了一种基于遗传算法的贪心备份方法(GGA-VNF)。该方法核心创新在于融合节点故障预测与多目标优化机制,通过GAN驱动故障预测模型主动规避高风险节点,结合遗传算法的全局优化能力,在动态网络环境中快速生成高鲁棒性备份方案。但当前模型主要聚焦于资源与可用性,却尚未集成延迟、能耗等关键指标,可能会影响系统在这些复杂场景下的综合性能。下一步工作将引入帕累托最优框架,联合建模延迟、能耗与资源效率等,探索多指标协同的均衡解。同时,本文结合VNF迁移(即在确保服务连续性和状态一致性的前提下,动态调整VNF的部署位置以适应网络环境变化和业务需求波动),考虑使用深度强化学习设计自适应的VNF迁移调整机制,旨在应对突发流量以及长期负载波动。

参考文献:

- [1] WANG J W, ZHANG X L, LI Q, et al. Research progress on network function virtualization technology [J]. *Journal of Computer Science*, 2019, 42(2): 185-206.
王进文, 张晓丽, 李琦, 等. 网络功能虚拟化技术研究进展[J]. *计算机学报*, 2019, 42(2): 185-206.
- [2] WANG P. Research on key technologies application of core network in fifth generation mobile communication [J]. *Electronic Communication and Computer Science*, 2023, 5(9): 125-127.
王鹏. 第五代移动通信中核心网关键技术应用研究[J]. *电子通信与计算机科学*, 2023, 5(9): 125-127.
- [3] GOU J G, LYU G F, SUN Z G, et al. Overview of network function virtualization technology [J]. *Computer Engineering and Science*, 2019, 41(2): 260-267.
苟建国, 吕高峰, 孙志刚, 等. 网络功能虚拟化技术综述[J]. *计算机工程与科学*, 2019, 41(2): 260-267.
- [4] DOE J, SMITH J. A survey on virtual network functions reliability in nf: challenges, solutions, and research directions[J]. *Journal of Network and Computer Applications*, 2022, 40(3): 221-235.
- [5] LIU Y, SHANG X, MAO Y, et al. Availability aware online virtual network function backup in edge environments[J]. *IEEE Transactions on Mobile Computing*, 2023, 23(5): 3909-3922.
- [6] ZHANG J, WU W, LUI J C S. On the theory of function placement and chaining for network function virtualization [C]//*Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2018: 91-100.
- [7] LIU Y, SHANG X, YANG Y. Joint SFC deployment and resource management in heterogeneous edge for latency minimization[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2021, 32(8): 2131-2143.
- [8] WANG C, HU Q, YU D, et al. Proactive deployment of chain-based VNF backup at the edge using online bandit learning[C]//*2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*, 2021: 740-750.
- [9] WANG C, HU Q, YU D, et al. Online learning for failure-aware edge backup of service function chains with the minimum latency [J]. *IEEE/ACM Transactions on Networking*, 2023, 31(6): 2730-2744.
- [10] LI J, GUO D, XIE J, et al. Availability-aware provision of service function chains in mobile edge computing [J]. *ACM Transactions on Sensor Networks*, 2023, 19(3): 1-28.
- [11] LU S, LUO B, PATEL T, et al. Making disk failure predictions SMARTer! [C]//*18th USENIX Conference on File and Storage Technologies*, 2020: 151-168.
- [12] WEN M S, LI J Q, HAI T X, et al. Failure prediction based VNF migration mechanism for multimedia services in power grid substation monitoring [C]//*2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2022: 1-6.
- [13] GUAN Y, LIU Y, ZHOU K, et al. A disk failure prediction model for multiple issues [J]. *Frontiers of Information Technology & Electronic Engineering*, 2023, 24(7): 964-979.
- [14] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2014: 2672-2680.
- [15] KLEIN A. What SMART hard disk errors actually tell Us [EB/OL]. <https://www.backblaze.com/blog/what-smart-stats-indicate-hard-drive-failures/>
- [16] EL-SHIMI A. Predicting storage failures [C]//*VAULT-Linux Storage and File Systems Conference*, 2017: 16-22.
- [17] GE J K, QIU Y H, WU C M, et al. A review of genetic algorithm research [J]. *Application Research of Computers*, 2008, 25(10): 2911-2916.
葛继科, 邱玉辉, 吴春明, 等. 遗传算法研究综述[J]. *计算机应用研究*, 2008, 25(10): 2911-2916.
- [18] HOY A, RAUSAND M. System reliability theory: models and statistical methods [M]. Hoboken: Wiley, 2004.
- [19] FAN J, JIANG M, QIAO C. Carrier-grade availability-aware mapping of service function chains with on-site backups [C]//*2017 IEEE/ACM 25th International Symposium on Quality of Service*, 2017: 1-10.
- [20] WANG M, CHENG B, ZHAO S, et al. Availability-aware service chain composition and mapping in NFV-enabled networks [C]//*2019 IEEE International Conference on Web Services*, 2019: 107-115.

(责任编辑:陈维捷)