应对灾难风险的多虚拟机快速协同撤离机制研究

鲍宁海* 李国平 冉 琴 岳渤涵
(重庆邮电大学通信与信息工程学院 重庆 400065)

摘 要：大规模灾难事件可能对通信网基础设施造成严重的威胁和破坏。针对大规模灾难风险下虚拟网的生存性问题，该文提出一种多虚拟机快速协同撤离(MRCE)机制。该机制采用后复制迁移技术实现虚拟机的在线迁移，通过基础迁移带宽的分配和升级，对属于同一虚拟网的多个风险虚拟机进行快速协同撤离，以减少单个虚拟网的撤离完成时长，降低损毁风险。仿真结果表明，该机制能在不同考察周期内获得较好的虚拟网撤离完成率和平均撤离完成时长。

关键词：虚拟网；生存性；灾难风险；多虚拟机迁移；协同撤离
中图分类号：TN915 **文献标识码**：A **文章编号**：1009-5896(2021)10-2886-08
DOI: 10.11999/JEIT200961

Research on Multi-virtual-machine Rapid Cooperative Evacuation Mechanism against Disaster Risks

BAO Ninghai LI Guoping RAN Qin YUE Bohan
(*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing* 400065, *China*)

Abstract: Large-scale disaster events might critically threaten and damage the infrastructure of telecom networks. In view of the virtual network survivability issue under large-scale disaster risks, a Multi-virtual-machine Rapid Cooperative Evacuation (MRCE) mechanism is proposed. The proposed mechanism employs the post-copy technique for live migration of virtual machines. By means of the basic migration bandwidth assignment and upgradation, a rapid cooperative evacuation is conducted on multiple risk virtual machines belonging to a single virtual network, so as to shorten the evacuation time of that virtual network and reduce its damage risk. Simulation results show that the proposed mechanism can achieve better performance than its counterparts in terms of virtual network evacuation completion ratio and average evacuation completion time in different observation periods.

Key words: Virtual Network (VN); Survivability; Disaster risk; Multi-virtual-machine migration; Cooperative evacuation

1 引言

近年来，网络虚拟化技术的不断发展和应用打破了传统互联网的僵化结构，为各类新兴业务的迅速推广提供了极大的便利。通过对计算、存储和带宽资源的抽象、汇聚与分配，底层的物理网络资源可以在不同的互联网业务提供商之间灵活共享，从而使众多基于虚拟网的应用得以广泛部署[1]。

虚拟网(Virtual Network, VN)通常是一定数量的虚拟节点和虚拟链路的集合。其中，每个虚拟节点以虚拟机的形式映射于一个物理节点中，实现计算和存储资源的汇聚，而每条虚拟链路映射于两个虚拟节点间的物理通路上，提供虚拟机之间的连通[2]。由于虚拟网各组件间的密切关联性，任何虚拟节点或虚拟链路的损毁都可能造成大量的业务中断和数据丢失，因此，虚拟网的生存性将成为影响未来网络运行和业务发展的重要问题。

本文针对大规模灾难事件(如地震、海啸等)对多个物理网络节点及相关链路造成的严重威胁，重点研究同一虚拟网中多个风险虚拟机的快速协同撤

收稿日期：2020-11-10；改回日期：2021-04-18；网络出版：2021-08-18
*通信作者： 鲍宁海 baonh@cqupt.edu.cn
基金项目：国家自然科学基金(61671092)，重庆市基础科学与前沿技术研究基金(cstc2016jcyjA0083)，重庆市高校创新团队基金(KJTD201312)
Foundation Items: The National Natural Science Foundation of China (61671092), The Fundamental Science and Frontier Technology Research Project of Chongqing (cstc2016jcyjA0083), The Chongqing City College Innovation Team (KJTD201312)

离机制，以减少单个虚拟网的撤离完成时长和损毁风险。

虚拟网生存性策略可分为网络组件抗毁和在线业务抗毁两类。前者可以采用虚拟网备份映射或重构来实现[3,4]，后者主要通过虚拟机在线迁移来完成[5]。而虚拟机的在线迁移技术又主要分为预复制迁移和后复制迁移两类，可广泛应用于在线系统维护、负载均衡、资源优化、能耗管理等领域[6]。

预复制迁移是先将原虚拟机的磁盘数据和主要内存页复制到目的虚拟机，再以迭代方式将原虚拟机产生的内存脏页(发生变化的内存信息)传递到目的虚拟机直到剩余脏页量或迭代次数满足要求，最后通过宕机切换将原虚拟机CPU状态和剩余脏页复制到目的虚拟机并启动目的虚拟机。后复制迁移则是首先通过宕机切换将原虚拟机CPU状态和少量必要内存页复制到目的虚拟机并启动目的虚拟机，然后再将剩余的内存页和磁盘数据主动复制并推送至目的虚拟机。

在虚拟机迁移技术的研究中，宕机时长和总迁移时长通常被用作衡量虚拟机在线迁移性能和效率的重要参考指标[7]。为减小虚拟机的预复制迁移时长，文献[8]通过预测迭代周期内的脏页数量，对迁移带宽进行动态预留和调整。文献[9]根据内存脏码率对迁移带宽和预复制迭代次数进行优化，提高网络带宽利用率，减小总迁移时长。文献[10]通过虚拟机迁移次序和迁移带宽的调度与优化，最小化预复制迁移时长。文献[11]采用速率感知的带宽共享传输满足预复制迁移的宕机时长和总迁移时长约束。文献[12]提出一种增强型距离自适应预复制迁移带宽分配算法，以降低迁移阻塞率。

为有效降低在线迁移过程中的传输负载，文献[13]针对后复制迁移，利用远程失效页面过滤技术提高迁移内容的有效性。文献[14]提出一种分散集中式后复制迁移策略，利用多个中间节点作为缓存代理，解决目标主机接收速率过慢的问题。文献[15]提出一种反向增量检查机制，保障后复制迁移中断后虚拟机的快速恢复，避免宕机时长和总迁移时长的恶化。

针对关联性多虚拟机迁移问题，文献[16]通过并行预复制迁移策略优化迁移带宽分配，获得低于串行预复制迁移的平均宕机时长、总迁移时长，以及迁移阻塞率。文献[17]根据关联虚拟机业务流量的相关性确定迁移次序，减小预复制迁移的平均宕机时长。文献[18]通过多虚拟机迁移带宽和预复制迭代次数的联合优化，对虚拟机宕机时长和总迁移时长进行平衡和折中。

显然，在现有研究工作中，预复制迁移技术受到了更多的关注[8–12,16–18]。在预复制迁移过程中，由于原虚拟机始终保有CPU状态，可在迁移中断时迅速恢复，能为在线迁移提供更高的可靠性。然而，原虚拟机产生的大量脏页将不可避免地增加迁移负载，因此，基于预复制迁移的研究工作大多集中在脏页迭代周期规划和网络带宽资源优化上，以便在宕机时长与总迁移时长上获得较好的折中[8–12]。而后复制迁移技术由于避免了大量脏页的产生和传递，迁移负载显著减小，从而获得更短的总迁移时长，因此，基于后复制迁移的研究工作主要集中在对内存页的压缩、标识，以及网络资源的有效利用上，以进一步提高在线迁移的效率或可靠性[13–15]。

从以上分析可以看出，相对于预复制迁移，后复制迁移更适合虚拟机的快速紧急迁移。然而，在大规模灾难风险下，虚拟网快速撤离的关键不是单个虚拟机的迁移完成时间，而是该虚拟网中所有风险虚拟机最终完成迁移的时间。现有关于多虚拟机迁移的研究大多采用预复制迁移技术，主要目标还是通过迭代调度和资源优化解决宕机时长和总迁移时长的平衡问题[16–18]，并未就大规模灾难风险场景下的虚拟网生存性问题进行针对性研究。因此，本文基于后复制迁移技术，提出一种应对灾难风险的多虚拟机快速协同撤离机制。

2　问题描述及网络模型

2.1　问题描述

大规模灾难事件可能对通信网基础设施造成严重的破坏，并对映射于灾难风险区内的大量虚拟机及其相关业务构成极大的威胁。如何快速撤离这些风险虚拟机已成为虚拟网抗毁的一个重要问题。特别是，当虚拟网的多个甚至全部虚拟机均处于灾难风险区时，虚拟网的生存性将取决于所有风险虚拟机是否能够实现快速协同撤离，而目前对该问题的研究尚待深入。

针对以上情况，本文假设一次大规模灾难事件导致多个具有地理关联性的物理网络节点(及其邻接链路)陷入灾难风险区。当任何一个虚拟网有一个或多个虚拟机映射于风险物理节点上时，需要对该虚拟网进行快速撤离。虚拟网撤离问题主要包括虚拟网重构和虚拟机在线迁移两个子问题，二者均属于NP-hard问题[2,6]。由于大量虚拟网业务的运行依赖于各虚拟机的协同工作，虚拟网及其业务的生存性将取决于最后一个风险虚拟机的迁移结束时间。因此，本文以减少单个虚拟网的撤离完成时长，降低损毁风险为目标，重点研究同一虚拟网中多个风险虚拟机的快速协同撤离机制。

2.2 网络模型

假设底层物理网络为广域网$G_{\mathrm{s}}(\boldsymbol{N}_{\mathrm{s}},\boldsymbol{L}_{\mathrm{s}})$，其中，$\boldsymbol{N}_{\mathrm{s}}$代表物理节点集合，$\boldsymbol{L}_{\mathrm{s}}$代表物理链路集合。每个物理节点$n_{\mathrm{s}}(\in\boldsymbol{N}_{\mathrm{s}})$均具备一定的计算能力和存储能力，每条物理链路$l_{\mathrm{s}}(\in\boldsymbol{L}_{\mathrm{s}})$均具备一定的带宽容量。在物理网络$G_{\mathrm{s}}$中，受灾难风险威胁的部分标记为$G'_{\mathrm{s}}(\boldsymbol{N}'_{\mathrm{s}},\boldsymbol{L}'_{\mathrm{s}})$，其中，$\boldsymbol{N}'_{\mathrm{s}}$代表风险物理节点集合，$\boldsymbol{L}'_{\mathrm{s}}$代表风险物理链路集合，未受灾难风险威胁的部分标记为$\overline{G'_{\mathrm{s}}}(\overline{\boldsymbol{N}'_{\mathrm{s}}},\overline{\boldsymbol{L}'_{\mathrm{s}}})=G_{\mathrm{s}}-G'_{\mathrm{s}}$。

虚拟网标记为$G_{\mathrm{v}}(\boldsymbol{N}_{\mathrm{v}},\boldsymbol{L}_{\mathrm{v}})$，其中，$\boldsymbol{N}_{\mathrm{v}}$代表虚拟节点集合，$\boldsymbol{L}_{\mathrm{v}}$代表虚拟链路集合。每个虚拟节点$n_{\mathrm{v}}(\in\boldsymbol{N}_{\mathrm{v}})$均具有一定的计算和存储资源需求，每条虚拟链路$l_{\mathrm{v}}(\in\boldsymbol{L}_{\mathrm{v}})$均具有一定的带宽资源需求。在虚拟网$G_{\mathrm{v}}$中，受灾难风险威胁的部分标记为$G'_{\mathrm{v}}(\boldsymbol{N}'_{\mathrm{v}},\boldsymbol{L}'_{\mathrm{v}})$，其中，$\boldsymbol{N}'_{\mathrm{v}}$代表风险虚拟节点集合，$\boldsymbol{L}'_{\mathrm{v}}$代表风险物理链路集合，未受灾难风险威胁的部分标记为$\overline{G'_{\mathrm{v}}}(\overline{\boldsymbol{N}'_{\mathrm{v}}},\overline{\boldsymbol{L}'_{\mathrm{v}}})=G_{\mathrm{v}}-G'_{\mathrm{v}}$。

3 方案描述

为应对大规模灾难风险对虚拟网及其业务造成的严重威胁，本文提出一种多虚拟机快速协同撤离(Multi-virtual-machine Rapid Cooperative Evacuation, MRCE)机制。该机制将虚拟网的撤离过程分为两个阶段，即首先在风险区外对虚拟网进行重构，实现虚拟网的组件抗毁，再通过多虚拟机协同迁移，实现虚拟网的业务抗毁。

3.1 虚拟网重构

本文中的虚拟网重构主要是将G'_{v}中的虚拟节点及其相邻虚拟链路从G'_{s}上重映射到$\overline{G'_{\mathrm{s}}}$上，并且满足虚拟网映射的基本约束条件，如式(1)到式(5)所示。式(1)表示虚拟节点n_{v}的计算和存储资源需求$C_{n_{\mathrm{v}}}$不能大于所映射的物理节点n_{s}的可用资源容量$C_{n_{\mathrm{s}}}$。式(2)中，$\alpha_{n_{\mathrm{v}},n_{\mathrm{s}}}$表示虚拟节点$n_{\mathrm{v}}$与物理节点$n_{\mathrm{s}}$的映射关系。式(3)和式(4)分别表示同一个虚拟节点只能映射到一个物理节点上，且同一个虚拟网的不同虚拟节点必须分别映射到不同的物理节点上。

$$C_{n_{\mathrm{v}}}\le C_{n_{\mathrm{s}}} \tag{1}$$

$$\alpha_{n_{\mathrm{v}},n_{\mathrm{s}}}=\begin{cases}1, & n_{\mathrm{v}}\text{映射于}n_{\mathrm{s}}\\ 0, & \text{其他}\end{cases} \tag{2}$$

$$\sum_{n_{\mathrm{s}}\in G_{\mathrm{s}}}\alpha_{n_{\mathrm{v}},n_{\mathrm{s}}}=1,\forall n_{\mathrm{v}} \tag{3}$$

$$\sum_{n_{\mathrm{v}}\in G_{\mathrm{v}}}\alpha_{n_{\mathrm{v}},n_{\mathrm{s}}}\le 1,\forall n_{\mathrm{s}} \tag{4}$$

$$c_{n_{\mathrm{s}}}=\sum_{n_{\mathrm{v}}\in G_{\mathrm{v}}}h_{n_{\mathrm{v}},n_{\mathrm{s}}},n_{\mathrm{s}}\in\overline{G'_{\mathrm{s}}} \tag{5}$$

在广域网中，节点资源相对于带宽资源更易于升级与扩容，因此，假设每个物理节点中的计算资源和存储资源都是足够的。为在重构虚拟链路长度与虚拟机迁移通路长度上取得平衡，实现网络带宽资源的优化利用，首先建立物理节点代价函数，如式(5)所示。其中，$h_{n_{\mathrm{v}},n_{\mathrm{s}}}$表示虚拟节点$n_{\mathrm{v}}$到物理节点$n_{\mathrm{s}}$的最短距离(跳数)，$c_{n_{\mathrm{s}}}$表示虚拟网$G_{\mathrm{v}}$中所有虚拟节点到$\overline{G'_{\mathrm{s}}}$中物理节点$n_{\mathrm{s}}$的最短距离和。

通过式(5)，在风险区域外寻找代价最小的物理节点作为锚点，并将距离锚点$H(\le 2)$跳内的区域$(\subset\overline{G'_{\mathrm{s}}})$划为$\boldsymbol{N}'_{\mathrm{v}}$的节点重构区域。然后，在重构区域内根据最大资源优先原则均衡分配节点资源，即优先将具有最大节点资源需求的虚拟节点映射到具有最大可用节点资源的物理节点上。最后，利用最短路算法在风险区域外重构$\boldsymbol{L}'_{\mathrm{v}}$中的虚拟链路。

3.2 多虚拟机协同迁移

在网络带宽资源有限的情况下，虚拟机的迁移数据量是影响迁移完成时间的重要因素，因此，本文采用后复制迁移技术，避免大量脏页的产生和迭代传输，并假设迁移数据均为总量确定的必要数据，不涉及数据清理、过滤及压缩等问题。

本文中多虚拟机协同迁移的主要目标是减小单个虚拟网的撤离完成时长。在为风险虚拟机建立迁移通路时，过小的迁移带宽将导致该虚拟机的迁移完成时长超长，从而影响整个虚拟网的撤离完成时长。为解决这一问题，首先引入虚拟机基础迁移带宽约束条件，如式(6)所示。其中，$T_{\max}$表示虚拟机的最大迁移完成时长门限，$\tau_{\max}$表示虚拟机的最大宕机时长，D_i表示风险虚拟机m_i的待迁移数据量，B_i表示m_i的迁移带宽。

由于虚拟网的撤离完成时长取决于最后一个风险虚拟机的迁移结束时间，当式(6)中的B_i取最小值时，该虚拟网中各风险虚拟机的迁移完成时长相同，且均为$T_{\max}$。然而，在动态变化的网络资源环境下，如果固定采用$\min(B_i)$作为迁移带宽，严格执行多虚拟机的同步迁移，可能会牺牲网络带宽的资源利用率。因此，通过选择合适的基础迁移带宽，并根据网络资源状态对其进行调整和升级，提高风险虚拟网的撤离效率。为解决可能出现的迁移带宽升级受限的问题，采用多通路策略提高带宽升级的灵活性。同时，为使同一虚拟网中各风险虚拟机的迁移完成时长尽可能一致，还需对各虚拟机的迁移带宽进行协同调整。

$$B_i\ge\frac{D_i}{T_{\max}-\tau_{\max}} \tag{6}$$

$$B_i=\sum_{k(\le K)}b_i^k \tag{7}$$

$$T_i = \frac{D_i}{B_i} + \tau_{\max} \tag{8}$$

$$b_i^+ = \frac{D_i}{\min(T_i) - \tau_{\max}} - B_i \tag{9}$$

假设虚拟网G_{v}中有$|\boldsymbol{N}_{\mathrm{v}}'|$个风险虚拟机，为每个风险虚拟机$\mathrm{m}_i(\in \boldsymbol{N}_{\mathrm{v}}')$计算一条满足式(6)带宽约束的最短路作为第1迁移通路，设该通路的迁移带宽$b_i^k|_{k=1}$等于该通路的最大可用带宽$b_i^{k,\max}|_{k=1}$。通过式(7)和式(8)计算各风险虚拟机的预计迁移完成时长T_i。其中，$K(\leq 3)$为每个虚拟机的最大迁移通路数，b_i^k为第k条迁移通路带宽。以$\min(T_i)$为基准确定相应的基础迁移带宽，采用多通路$(1 \leq k \leq K)$策略为T_i较大的风险虚拟机寻找增量迁移通路并配置相应的增量带宽，使得各风险虚拟机的迁移完成时长尽可能一致。其中，增量带宽b_i^+的计算如式(9)所示。

对于风险虚拟网G_{v}，当$\boldsymbol{N}_{\mathrm{v}}'$中所有$\mathrm{m}_i$的迁移通路和带宽都分配完成后，启动后复制迁移。其中，每个m_i在宕机切换结束后，立即执行内存和磁盘数据的迁移，其迁移完成时间t_i^{c}如式(10)所示。其中，t_{c}为系统当前时间，D_i^{c}为m_i已迁移的数据量。若在$\min(t_i^{\mathrm{c}})$时间，$\boldsymbol{N}_{\mathrm{v}}'$中还有虚拟机$\mathrm{m}_i$未完成迁移，则其迁移完成时差$T_i^{\mathrm{d}}$如式(11)所示。为尽可能缩短虚拟网的撤离完成时长，当网络中出现可用带宽资源时，根据最大迁移完成时差优先原则，依次对相应虚拟机的迁移带宽进行升级。

$$t_i^{\mathrm{c}} = t_{\mathrm{c}} + \frac{D_i - D_i^{\mathrm{c}}}{B_i} \tag{10}$$

$$T_i^{\mathrm{d}} = t_i^{\mathrm{c}} - \min(t_i^{\mathrm{c}}) \tag{11}$$

3.3 启发式算法

根据以上方案描述，本文设计了相应的MRCE启发式算法，用于实现大规模灾难风险下虚拟网的快速撤离。其中，子算法-1完成虚拟网重构，子算法-2完成多虚拟机协同迁移的初始配置。

(1) MRCE算法

步骤1　将所有风险虚拟网放入集合$\boldsymbol{R}$，并按其所含风险虚拟机数量升序排列，初始化当前时间t_{c}=0，每个虚拟机已迁移数据量D_i^{c}=0，迁移完成时间t_i^{c}=∞；

步骤2　如果集合$\boldsymbol{R} \neq \varnothing$，调用子算法-1对$\boldsymbol{R}$中的每个风险虚拟网进行重构，调用子算法-2为重构虚拟网中的每个风险虚拟机分配迁移通路和带宽，将迁移带宽分配成功的虚拟网移入集合E中，跳转到步骤3，否则，跳转到步骤8；

步骤3　如果集合$\boldsymbol{E} \neq \varnothing$，对$\boldsymbol{E}$中虚拟网的各风险虚拟机$\mathrm{m}_i$执行后复制迁移，跳转到步骤4，否则，跳转到步骤2；

步骤4　根据式(10)更新所有虚拟机的迁移结束时间t_i^{c}，如果有t_i^{c}=t_{c}，相应虚拟机m_i完成迁移，释放其迁移通路及带宽，跳转到步骤5，否则，跳转到步骤3；

步骤5　如果该m_i为所属虚拟网内第1个完成迁移的虚拟机，根据式(11)计算该虚拟网内其他m_i的迁移完成时差T_i^{d}并放入集合$\boldsymbol{U}$，跳转到步骤7，否则，跳转到步骤6；

步骤6　如果该m_i为所属虚拟网内最后一个完成迁移的虚拟机，将该虚拟网从$\boldsymbol{E}$中删除，跳转到步骤7，否则，跳转到步骤7；

步骤7　更新$\boldsymbol{U}$中所有的T_i^{d}并删除$T_i^{\mathrm{d}} = 0$的记录，如果$\boldsymbol{U} \neq \varnothing$，将$T_i^{\mathrm{d}}$降序排列，依次将相应的虚拟机迁移带宽升级至上限，跳转到步骤2，否则，跳转到步骤2；

步骤8　算法结束。

(2) 子算法-1

步骤1　在$\overline{G_{\mathrm{s}}'}$内，根据式(5)为风险虚拟网$G_{\mathrm{v}}$寻找最小代价物理节点作为锚点，并根据参数$H$确定$\boldsymbol{N}_{\mathrm{v}}'$的节点重构区域；

步骤2　依次为$\boldsymbol{N}_{\mathrm{v}}'$中的虚拟节点按最大节点资源需求优先原则，在节点重构区域内选取具有最大可用资源的物理节点进行重映射；

步骤3　如果$\boldsymbol{N}_{\mathrm{v}}'$中的所有虚拟节点重映射成功，依次为$\boldsymbol{L}_{\mathrm{v}}'$中的虚拟链路在$\overline{G_{\mathrm{s}}'}$内寻找满足带宽资源需求的最短路，并进行重映射，跳转到步骤4，否则，跳转到步骤5；

步骤4　如果$\boldsymbol{L}_{\mathrm{v}}'$中的所有虚拟链路重映射成功，风险虚拟网$G_{\mathrm{v}}$完成重构，跳转到步骤6，否则，跳转到步骤5；

步骤5　风险虚拟网G_{v}重构失败，释放已重构的虚拟节点及虚拟链路资源，跳转到步骤6；

步骤6　算法结束。

(3) 子算法-2

步骤1　初始化$k = 1$，分别为$\boldsymbol{N}_{\mathrm{v}}'$中的每个风险虚拟机$\mathrm{m}_i$计算一条满足式(6)约束的最短迁移通路，如果成功，跳转到步骤2，否则，跳转到步骤6；

步骤2　对$\boldsymbol{N}_{\mathrm{v}}'$中每个$\mathrm{m}_i$，令$b_i^k = b_i^{k,\max}$，根据式(7)和式(8)计算其预计迁移完成时长$T_i$，将$T_i > \min(T_i)$的风险虚拟机$\mathrm{m}_i$放入集合$\boldsymbol{M}$中，并按$T_i$值降序排列；

步骤3　如果集合$\boldsymbol{M} \neq \varnothing$，令$k = k + 1$，跳转到步骤4，否则，跳转到步骤6；

步骤4　依次为$\boldsymbol{M}$中每个m_i计算一条最短通路作为增量迁移通路，根据式(9)计算增量带宽b_i^+，如果$b_i^{k,\max} \geq b_i^+$，令$b_i^k = b_i^+$，将该m_i从$\boldsymbol{M}$中删除，跳转到步骤5，否则，令$b_i^k = b_i^{k,\max}$，跳转到步骤5；

步骤5　如果$k < K$，跳转到步骤3，否则，跳转到步骤6；

步骤6　算法结束。

根据以上算法步骤分析，MRCE的时间复杂度为$O(|\boldsymbol{N}_\mathrm{v}|^2 \cdot |\boldsymbol{N}_\mathrm{s}| \cdot \log_2 |\boldsymbol{N}_\mathrm{s}| + |\boldsymbol{N}_\mathrm{v}| \cdot K \cdot |\boldsymbol{N}_\mathrm{s}| \cdot \log_2 |\boldsymbol{N}_\mathrm{s}|)$，其中，$\boldsymbol{N}_\mathrm{s}$与$\boldsymbol{N}_\mathrm{v}$分别表示物理网$G_\mathrm{s}$和虚拟网$G_\mathrm{v}$的最大节点数，$K$表示单个虚拟机的最大迁移通路数，$|\boldsymbol{N}_\mathrm{s}| \cdot \log_2 |\boldsymbol{N}_\mathrm{s}|$为最短路算法的时间复杂度。

4　仿真测试与分析

本文对提出的多虚拟机快速协同撤离算法(MRCE)进行仿真测试，并采用常规后复制迁移算法(Regular Post-Copy Migration, RPCM)和基于自适应带宽升级的后复制迁移算法(Adaptive bandwidth Upgraded Post-copy Migration, AUPM)进行风险虚拟网撤离的性能对比。3种算法均采用相同的虚拟网重构策略，其中，RPCM为每个虚拟机寻找最短迁移通路并分配最大可用迁移带宽，而AUPM在RPCM的基础上，根据网络带宽资源状态，对迁移带宽进行尽力而为的动态升级。

仿真采用的物理网络拓扑包含24个节点和43条链路，如图1所示。假设各物理节点均拥有充足的计算和存储资源，各物理链路的带宽资源均为240 Gbps。图中阴影部分代表3个独立的灾难风险模型，分别包含(3, 4, 5)，(9, 12, 13)和(16, 17, 22) 3组物理节点及其相邻物理链路。

随机产生90套初始业务模型，每套包含500个虚拟网络，且随机分布(映射)于上述物理网络中。其中，每个虚拟网络随机产生3～5个虚拟节点，任意两个虚拟节点间以0.5的概率建立虚拟链路，虚拟链路带宽在0.5～3 Gbps间随机产生。采用后复制迁移技术对风险虚拟机进行迁移，每个虚拟机的待迁移数据量在5～10 GB间随机产生，宕机时长在0.5～1.5 s间随机产生[9]。

在以上仿真环境中，分别对MRCE, RPCM和AUPM进行仿真测试，性能指标主要包括风险虚拟网的平均撤离完成时长、撤离完成时长标准差，以及撤离完成率。最终仿真结果对3个灾难风险模型和90套随机业务模型取平均。考虑到我国的地震监测系统已具备一定程度的预警能力(十几秒至几十秒的预警时间)，但网络组件的损毁时间仍旧难以确定。因此，以10 s为基数设立6个考察周期，分别在算法执行后的10 s, 20 s, 30 s, 40 s, 50 s, 60 s时间点对以上性能指标进行考察记录。

在MRCE中，虚拟机最大迁移完成时长门限T_max的选取对网络带宽资源利用率有较大的影响。如果T_max取值过大，则虚拟机的迁移效率降低，可能造成虚拟网的平均撤离完成时长过大；如果T_max取值过小，则虚拟机的基础迁移带宽需求较大，可能造成过高的迁移带宽阻塞率。因此，首先考察MRCE在不同T_max设置下的平均撤离完成时长和迁移带宽阻塞率，并用T10, T20, T30, T40, T50, T60分别代表T_max取值为10 s, 20 s, 30 s, 40 s, 50 s, 60 s时MRCE的相应指标。

如图2和图3所示，在10～60 s考察周期内，MRCE的平均撤离完成时长随着T_max取值的增加不断变大，而其迁移带宽阻塞率却随着T_max的增加而

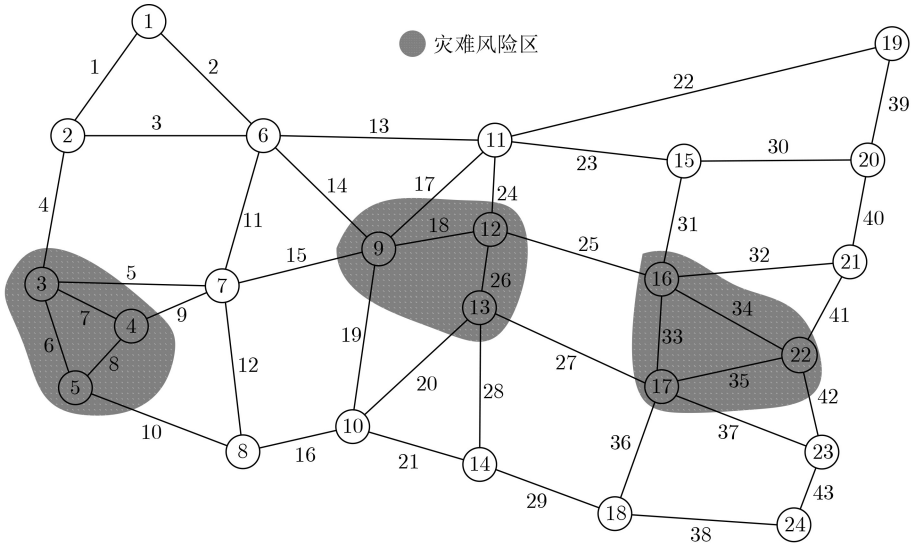

图 1 物理网络拓扑

降低。显然，由式(6)可知，增大$T_{\max}$的取值等同于放松对基础迁移带宽的约束，使得虚拟机更容易找到可用的迁移通路。然而，当网络中的带宽资源较为紧张，且碎片化较为严重时，过小的基础迁移带宽会使以$\min(T_i)$为基准的增量带宽策略变得低效，从而导致虚拟网的撤离完成时长变大。图3数据显示，在此仿真环境下，当$T_{\max}$≥40 s时，迁移带宽阻塞率较低，且数值变化明显趋缓，因此在后续的仿真测试中，采用40 s作为MRCE的虚拟机最大迁移完成时长门限。

风险虚拟网平均撤离完成时长如图4所示。在10 s的考察时间点，MRCE和AUPM的平均撤离完成时长均高于RPCM，其原因包括两个方面：其一是在撤离开始的初期，包含风险虚拟机数量最少的虚拟网将被优先撤离，特别是仅包含单个风险虚拟机的虚拟网，能以最小代价尽快脱离风险区。而对此类风险虚拟网来说，MRCE的协同撤离策略和AUPM的自适应升级策略的优势不够突出。其二是在每一个考察时间点，虚拟网的平均撤离完成时长只统计当前已完成撤离的虚拟网，而RPCM在10 s周期内完成撤离的虚拟网数量远小于MRCE和AUPM，且完成撤离的主要是仅包含单一风险虚拟机的虚拟网。从图4中可以发现，在20 s及以后的考察时间点，MRCE的平均撤离完成时长有稍许增加但非常平稳，AUPM的平均撤离完成时长逐渐增长但较缓慢，而RPCM的平均撤离完成时长增长较为迅速。显然，相对于RPCM, AUPM的自适应升级策略能够更好地提高网络带宽资源利用率，缩短风险虚拟机的迁移完成时长，从而在一定程度上缩短风险虚拟网的撤离完成时长。而MRCE的协同撤离策略通过压缩同一虚拟网中风险虚拟机间的迁移完成时差，能够进一步减小各风险虚拟网的撤离完成时长。

风险虚拟网的撤离完成时长标准差如图5所示。在10 s考察时间点，RPCM的撤离完成时长标准差与MRCE的接近且稍低于AUPM。其原因除了前面提到的RPCM在该考察周期内完成撤离的虚拟网数量少且主要包含单一风险虚拟机以外，另一个原因是在撤离初期，风险区外的网络资源相对较充足，带宽资源碎片化情况不严重，因此，RPCM在此期间完成虚拟网撤离所耗费的时长较短且差距不大。然而，随着包含多风险虚拟机的虚拟网开始大量撤离，以及风险区外网络带宽资源的减少，RPCM的撤离完成时长标准差迅速上升。虽然AUPM可以通过自适应带宽升级提高带宽资源利用率，但对缩短同一虚拟网内不同风险虚拟机的迁移完成时差并无确定性的作用，即迁移带宽升级的结果可能使时差减小，也可能使时差增大。随着风险区外可用带宽资源的逐渐短缺和碎片化，AUPM的自适应带宽升级效率逐渐降低，这使得不同风险虚拟网撤离完成时长的差距逐渐变大。相对于RPCM和AUPM，MRCE通过基础迁移带宽的选取和增量迁移带宽的

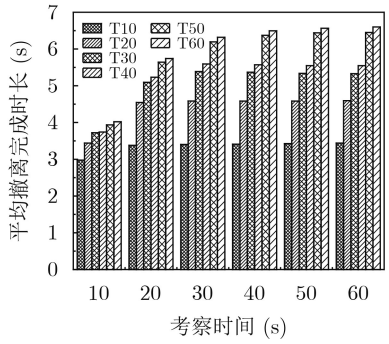

图 2 $T_{\max}$对平均撤离完成时长的影响

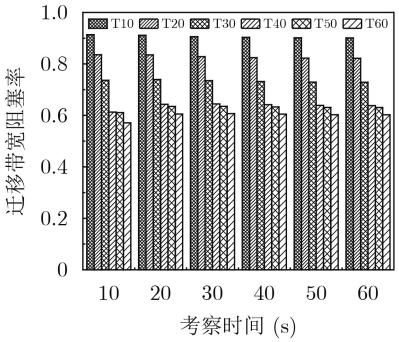

图 3 $T_{\max}$对迁移带宽阻塞率的影响

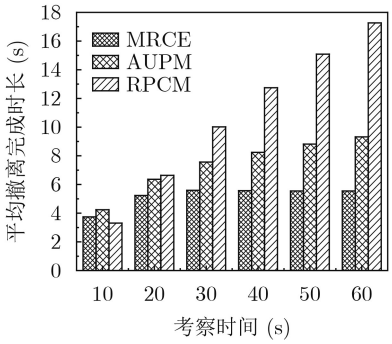

图 4 不同考察时间下的平均撤离完成时长

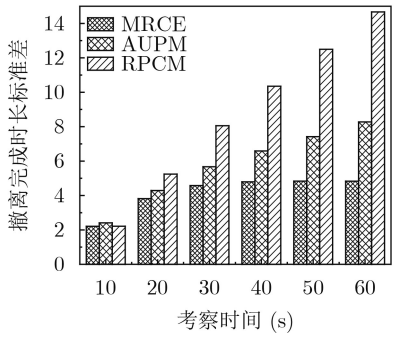

图 5 不同考察时间下的撤离完成时长标准差

配置，以及基于虚拟机迁移完成时差的迁移带宽动态升级，能够在不同的网络带宽资源状态下获得较为稳定的撤离完成时长标准差。

风险虚拟网的撤离完成率可以直观反映灾难风险下虚拟网的抗毁能力。如图6所示，相对于AUPM和RPCM，MRCE在各个阶段都能取得更好的撤离完成率。该结果表明，在灾难损毁时间不确定的情况下，MRCE能够更快更多地完成虚拟网的撤离。值得注意的是，随着考察时间的不断推移，3种算法的撤离完成率也在不断增长，但增长率有逐渐变缓的趋势。这是因为随着风险区外的虚拟网(已重构且完成撤离)不断增多，该区域的带宽资源消耗也在不断增加，可用于虚拟机迁移的带宽资源大量减少，从而限制了并行撤离的虚拟网数量。因此，在大量风险虚拟网的撤离过程中，撤离完成率越高，其增长率下降越明显。

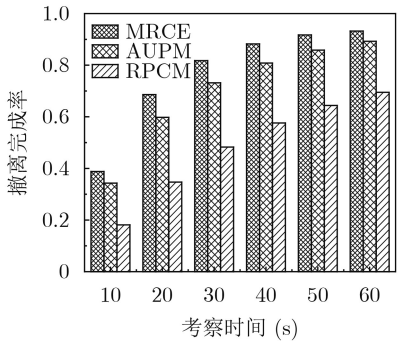

图 6 不同考察时间下的撤离完成率

5 结束语

为应对大规模灾难事件对虚拟网生存性造成的严重威胁，本文研究并提出一种多虚拟机快速协同撤离机制。该机制采用风险虚拟网的重构实现网络组件的抗毁，采用风险虚拟机的后复制迁移实现在线业务的抗毁。特别针对同一虚拟网的多个或全部虚拟机均受到灾难威胁的情况，通过虚拟机最大迁移完成时限设定，约束虚拟网撤离完成时长，通过虚拟机迁移完成时长预估和基础迁移带宽分配，协调各虚拟机的迁移速度，通过虚拟机迁移完成时差评估和迁移带宽动态升级，进一步压缩虚拟网撤离完成时长。仿真测试证明，与常规后复制迁移算法和基于自适应带宽升级的后复制迁移算法相比，多虚拟机快速协同撤离算法能够在不同考察周期内获得较好的虚拟网撤离完成率和平均撤离完成时长。本文提出的虚拟网抗毁机制主要针对自然灾难风险，在应对未知或不可测风险(如军事打击、意外事件)，以及更为严苛的时间约束等问题上，还有较大的研究空间。

参 考 文 献

[1] CHOWDHURY N M M K and BOUTABA R. A survey of network virtualization[J]. *Computer Networks*, 2010, 54(5): 862–876. doi: 10.1016/J.COMNET.2009.10.017.

[2] CAO Haotong, WU Shengchen, HU Yue, *et al*. A survey of embedding algorithm for virtual network embedding[J]. *China Communications*, 2019, 16(12): 1–33. doi: 10.23919/JCC.2019.12.001.

[3] HE Fujun, SATO T, and OKI E. Backup resource allocation model for virtual networks with probabilistic protection against multiple facility node failures[C]. The 15th International Conference on the Design of Reliable Communication Networks, Coimbra, Portugal, 2019: 37–42. doi: 10.1109/DRCN.2019.8713736.

[4] SHAHRIAR N, AHMED R, CHOWDHURY S R, *et al*. Generalized recovery from node failure in virtual network embedding[J]. *IEEE Transactions on Network and Service Management*, 2017, 14(2): 261–274. doi: 10.1109/TNSM.2017.2693404.

[5] GHALEB A M, KHALIFA T, AYOUBI S, *et al*. Surviving multiple failures in multicast virtual networks with virtual machines migration[J]. *IEEE Transactions on Network and Service Management*, 2016, 13(4): 899–912. doi: 10.1109/TNSM.2016.2616283.

[6] ZHANG Fei, LIU Guangming, FU Xiaoming, *et al*. A survey on virtual machine migration: Challenges, techniques, and open issues[J]. *IEEE Communications Surveys & Tutorials*, 2018, 20(2): 1206–1243. doi: 10.1109/COMST.2018.2794881.

[7] NOSHY M, IBRAHIM A, and ALI H A. Optimization of live virtual machine migration in cloud computing: A survey and future directions[J]. *Journal of Network and Computer Applications*, 2018, 110: 1–10. doi: 10.1016/J.JNCA.2018.03.002.

[8] 李湘, 陈宁江, 杨尚林, 等. 感知应用特征与网络带宽的虚拟机在线迁移优化策略[J]. 通信学报, 2017, 38(S2): 147–155. doi: 10.11959/J.ISSN.1000-436x.2017268.

LI Xiang, CHEN Ningjiang, YANG Shanglin, *et al*. Optimization strategy of virtual machine online migration with awareness of application characteristics and network bandwidth migration[J]. *Journal on Communications*, 2017, 38(S2): 147–155. doi: 10.11959/J.ISSN.1000-436x.2017268.

[9] MANDAL U, CHOWDHURY P, TORNATORE M, *et al*. Bandwidth provisioning for virtual machine migration in cloud: Strategy and application[J]. *IEEE Transactions on Cloud Computing*, 2018, 6(4): 967–976. doi: 10.1109/TCC.2016.2545673.

[10] WANG Huandong, LI Yong, ZHANG Ying, *et al*. Virtual machine migration planning in software-defined networks[J].

IEEE Transactions on Cloud Computing, 2019, 7(4): 1168–1182. doi: 10.1109/TCC.2017.2710193.

[11] ZHANG Jiao, REN Fengyuan, SHU Ran, *et al*. Guaranteeing delay of live virtual machine migration by determining and provisioning appropriate bandwidth[J]. *IEEE Transactions on Computers*, 2016, 65(9): 2910–2917. doi: 10.1109/TC.2015.2500560.

[12] AYOUB O, MUSUMECI F, TORNATORE M, *et al*. Efficient routing and bandwidth assignment for inter-data-center live virtual-machine migrations[J]. *Journal of Optical Communications and Networking*, 2017, 9(3): B12–B21. doi: 10.1364/JOCN.9.000B12.

[13] SU Kui, CHEN Wenzhi, LI Guoxi, *et al*. RPFF: A remote page-fault filter for post-copy live migration[C]. 2015 IEEE International Conference on Smart City/Socialcom/Sustaincom, Chengdu, China, 2015: 938–943. doi: 10.1109/SmartCity.2015.191.

[14] DESHPANDE U, CHAN D, CHAN S, *et al*. Scatter-gather live migration of virtual machines[J]. *IEEE Transactions on Cloud Computing*, 2018, 6(1): 196–208. doi: 10.1109/TCC.2015.2481424.

[15] FERNANDO D, TERNER J, GOPALAN K, *et al*. Live migration ate my VM: Recovering a virtual machine after failure of post-copy live migration[C]. IEEE INFOCOM 2019-IEEE Conference on Computer Communications, Paris, France, 2019: 343–351. doi: 10.1109/InfoCom.2019.8737452.

[16] SUN Gang, LIAO Dan, ZHAO Dongcheng, *et al*. Live migration for multiple correlated virtual machines in cloud-based data centers[J]. *IEEE Transactions on Services Computing*, 2018, 11(2): 279–291. doi: 10.1109/TSC.2015.2477825.

[17] NARANTUYA J, ZANG Hannie, and LIM H. Service-aware cloud-to-cloud migration of multiple virtual machines[J]. *IEEE Access*, 2018, 6: 76663–76672. doi: 10.1109/ACCESS.2018.2882651.

[18] CERRONI W and ESPOSITO F. Optimizing live migration of multiple virtual machines[J]. *IEEE Transactions on Cloud Computing*, 2018, 6(4): 1096–1109. doi: 10.1109/TCC.2016.2567381.

鲍宁海：男，1973年生，博士，教授，研究方向为网络生存性、网络虚拟、网络节能、移动边缘计算等.

李国平：男，1992年生，硕士，研究方向为网络生存性、网络虚拟.

冉　琴：女，1995年生，硕士生，研究方向为网络生存性、移动边缘计算.

岳渤涵：男，1995年生，硕士生，研究方向为网络生存性、移动边缘计算.

责任编辑：马秀强