信息系统灾难恢复计划研究

王 琨[①] 尹忠海[①] 周利华[①] 蔡 震[②]
[①](西安电子科技大学计算机网络与信息安全教育部重点实验室 西安 710071)
[②](国家信息安全工程技术研究中心 北京 100093)

摘 要：灾难发生时，需要有量化的、精确的方法从多个子灾难恢复计划中选择最优集合实现灾难恢复。该文采用最优化理论提出一个灾难恢复计划数学模型。模型表示系统中不同应用、设施、资源、子灾难恢复计划、预算等以及它们之间的关系，对资源进行分类，解决不同子灾难恢复计划之间的冲突。模型具有较少的主观参数，以实现客观评价灾难恢复计划，易于对灾难恢复计划进行精确、量化的分析。论文给出了模型的实施步骤与模型分析，实验验证了模型的正确性。

关键词：信息安全；持续服务；风险评估；灾难恢复计划；最优化理论
中图分类号：TP393.08 **文献标识码**：A **文章编号**：1009-5896(2007)04-0776-05

Study on Information System Disaster Recovery Planning

Wang Kun[①] Yin Zhong-hai[①] Zhou Li-hua[①] Cai Zhen[②]
[①](*Key Laboratory of Computer Network and Information Security of the Ministry of Education, Xidian University, Xi'an* 710071, *China*)
[②](*National Information Security Engineering and Technology Research Center, Beijing* 100093, *China*)

Abstract: When disaster happens, it is necessary use quantitive and accurate method to select the best set from many sub disaster recovery plans to control the disaster recovery activities. This paper presents a practicable mathematical model using optimization theory. The model contains some parameters that stand for different applications, facilities, resources, sub plans, budget, etc. The model classifies resources and solves the conflicts among various sub disaster recovery plans. With less subjective parameter, the model can evaluate DRP more impersonally. The model possesses less parameter and can be easily implemented to analyze DRP accurately and quantitatively. The implementation and the analysis of the model are presented. Test also verifies the correctness of this model.

Key words: Information security; Continuous service; Risk assessment; Disaster Recovery Planning (DRP); Optimization theory

1 引言

越来越多应用日益严重依赖信息技术，一旦信息系统失效，也会对应用产成致命影响，例如：数据损失、员工生产率降低、客户关系损害、利润与收入的减少等。911 恐怖主义袭击再次警告人们信息安全的重要性，此后，越来越多IT经理人认识到灾难恢复的重要性[1]。本文中“灾难”已经超出了自然灾害的范畴，它指任何可能对信息系统产生危害导致系统失效一段时间的安全威胁。根据Dynamic Market公司对IT经理人的调查，大公司最重视的五大安全威胁分别是硬件失效(61%)、软件失效与病毒(both 59%)、火灾(56%)、黑客(36%)、员工意外失误(31%)[2]。

如何更加有效保护系统，提供持续服务已经成为一个研究热点。为了使系统在任何时候都持续可用，需要建设鲁棒的灾难恢复系统(Disaster Recovery System， DRS)。由于灾难恢复的代价通常非常昂贵，应该以合理的费用保护系统。由于 DRS 本身固有的复杂性，还需要开发辅助工具帮助决策者进行灾难恢复控制。

对于安全性要求非常高的组织(例如政府、军队、银行)尤其迫切需要建设DRS来保护它们的信息系统。作为中国“十五”重点科技攻关计划的某电子政务试点示范工程(E-government Experimental and Demonstration Project，EEDP)要求核心政府部门必须能够抵御多种人为和自然灾害的威胁。因此，本文提出了鲁棒的灾难恢复系统模型(Robust Disaster Recovery System Model， RDRSM)[3]，以指导EEDP中DRS的建设。DRS非常复杂，它涉及众多技术领域和人员。为了合理组织系统资源，保护目标应用，把DRS划分为几个环节，通过不同环节相互配合实现灾难恢复。灾难恢复计划(Disaster Recovery Planning， DRP)就

2005-12-26 收到, 2006-06-30 改回
国家“十五”重点科技攻关计划(2002AA1Z67101)资助课题

是DRS中非常关键的一个环节，它包含许多子灾难恢复计划(Sub Disaster Recovery Plans， SDRP)，通过有序的计划，DRP帮助决策者控制DRS的活动，使系统从灾难中恢复过来。

设计优秀的DRP不仅非常重要，而且非常困难，然而目前人们更多在研究有关DRS的软硬件工具，没有投入足够研究在DRP上。Chow[4]研究了香港的金融、制造、贸易和宾馆行业，发现对于这 4 种行业，在DRS中最为重要的五大因素分别是管理层支持、充足的资金保证、恰当的备份站点、离线存储备份和员工培训，这有助于集中注意力于DRS中的重要因素。还有一些研究侧重于某些特定应用领域，例如Smith研究防御信息系统网络中传输系统的DRP，目的在于加强卫星与地面通信系统的灾难恢复能力[5]。Hayes研究飓风中的灾难恢复问题[6]。还有一些相对通用的有关DRP的研究，例如Business Continuity Planning (BCP)[7]是一种过程驱动的通用模型，它把灾难恢复划分为 8 个环节，通过分析和更新商业持续计划实现灾难恢复。Disaster Management System (DMS)[8]集成传感器技术、建模和仿真工具、遥感和计算平台，为决策者提供预警、灾难发生中和灾难恢复中的信息，提高应对灾难的能力。它主要研究如何降低灾难中的人员与经济损失，并不是专门为保护信息系统而设计的。Fallara[9]和Hawkins[10]研究DRP中风险管理的重要性，以及灾难对商业应用的影响。

DRS 非常复杂，它包含许多重要因素。系统中应用经常具有不同的权值，显然具有较高权值的应用应该首先被恢复；由于一些应用需要依靠另外一些应用才能正常工作，在进行恢复工作时，还需要考虑到这种先后顺序；信息系统在灾难发生时经常会同时出现大量错误，不同应用可能需要占用同一个资源才能完成灾难恢复，因此还需要解决这种由于资源占用而产生的冲突；当系统崩溃时，可能有多种途径都可以恢复系统，每种途径需要消耗不同的资源，花费不同的时间。由于灾难恢复系统本身固有的复杂性，在灾难发生后短时间内仅通过直觉和经验从众多可选的 SDRP 中选择出最优集合，提供最好的恢复计划几乎是不可能的。这些研究主要在概念层次和宏观角度研究 DRP，并且作出了重要的贡献，但是也有其局限性。这些研究没有进一步分析业务应用系统，提供解决有关应用权值和不同应用之间关系问题的方法。并且没有考虑到 DRS 中不同类型资源具有不同的计费方式，这会严重影响最优化灾难恢复的代价；有些研究考虑到在灾难恢复中由于资源占用会产生冲突，但是没有进一步的解决方案。总之，这些对 DRP 的研究不能够精确评估 DRP，本文目的在于解决这些问题。

在深入研究 DRP 和信息系统需求的基础上，本文提出量化的数学模型用于建设和评估 DRP。这是一个为保护信息系统而设计的模型。模型能够区分和标示不同应用的重要性，分类不同类型的资源，解决不同 SDRP 之间的冲突，这些都是该模型优越于其它模型的特点。模型能够精确、量化地评估 DRP，帮助决策者从众多 SDRP 中选择最优的 SDRP 集合控制系统的灾难恢复工作。

2　灾难恢复计划数学模型

在制订DRP之前，需要建立长期的风险评估机制对系统进行风险评估，确定某个安全威胁发生的可能性，以及由此而引起的经济、信誉和商业伙伴损失等[11]。必须明确标识所有安全威胁和需要保护的应用与资源，通过分析不同安全威胁对不同应用与资源的破坏程度，应用与资源对破坏的时间、经济敏感度，区分应用与资源的优先级。

DRP 关注的是风险评估之后剩余风险环境下的业务运转，包括商业影响分析和恢复计划设计。商业影响分析研究某个应用或资源的中断对其它应用的影响。主要有两个度量：一个度量是某个应用或资源崩溃后，其它应用还能持续正常运行的最大时间；另一个度量是恢复崩溃应用需要占用的资源，系统中的不同资源各有其特点，其占用代价的计算方式也有所不同(例如一次性购买付费或多次租用)。应尽量要找到两者的最佳接合点。恢复计划设计根据系统需求和目标设计 DRP，它包括许多策略和计划，例如数据备份策略，组建救援维护小组，资源维护计划，权衡维护或替换损毁设备的代价等。在设计 DRP 时，比较重要的几个限制是预算、复杂性和对资源的使用。不同的 SDRP 有可能能够保护恢复相同的设施，但是可能具有不同的效率和代价。另一方面，每个 SDRP 都要使用一定资源，某些资源可以被其它资源所替代，这些不同资源在解决相同问题时可能具有不同效率，花费不同代价，并且某些资源可以用于保护恢复多种设施。由于信息系统在灾难发生时出现的故障具有大批量的特点，因此在 DRP 的设计中应制订排除大量并发网络和系统故障的计划，并且保证不同 SDRP 之间不会由于资源抢占而发生冲突。

2.1　灾难恢复计划模型

当业务应用崩溃时，需要花费一定资源修复崩溃的应用。通常许多 SDRP 构成的多个子集都可以解决同一问题，但是哪个子集是最优的？仅仅依靠直觉和经验是几乎不可能选择出最优方案的，理想的方法是建立数学模型，通过量化的计算选择出最优解。由于灾难恢复系统本身固有的复杂性，必须对问题进行适当抽象，使抽象后的数学模型既要忽略一些相关度不大的细节，使问题不会由于过于复杂而无法解决，又必须符合现实世界的要求，能够解决实际问题。另外，模型中应尽量避免使用过多受主管因素影响的参数，以提高模型的客观性和精确性。基于上述思想构建模型如下：

符号说明：B 为灾难恢复的预算限额。A 为系统中包含的业务应用的集合。s_a 为业务应用 a 的重要程度。F 为需要保护的设施的集合，可以是软硬件设备等，它对业务应用具有重要价值。通常信息系统包含多个业务应用 a，每个

业务应用需要依靠多个设施 f 才能够实现其功能。另一方面，一个设施 f 可能会对多个业务应用 a 提供服务。可见，a 与 f 之间通常是多对多关系。d_{fa} 为设施 f 的失效会影响业务应用 a 正常运行的可能性。w_f 为设施 f 相对所有业务应用的重要性，$w_f=\sum_{a\in A}s_a d_{fa}$。$R_o$ 为对于有些消耗性资源，它们只需要付一次费用，只能被使用一次(例如打印耗材)，这些资源构成集合 R_o。R_m 为还有一些资源，只需要付一次购买费用，之后可以反复被使用，这些资源构成集合 R_m。R_h 为某些资源需要在每次使用时交付租用费用(例如人力资源、租用设备等)，这些资源构成集合 R_h。R 为灾难恢复需要使用的所有资源的集合，$R=R_o\cup R_m\cup R_h$。P 为所有 SDRP 构成的集合。P_f 为可以保护恢复设施 f 的 SDRP 构成的集合。P_r 为需要使用资源 r 的 SDRP 构成的集合。P_{c_i} 为由相互之间会产生冲突的 SDRP 构成的集合的集合，其元素编号为 $1,\cdots,k$。i_{pf} 为当子灾难恢复计划 p 能够提供对设施 f 的保护恢复能力时，取值为 1，否则为 0。m_p 为对子灾难恢复计划 p 能力的度量，$m_p=\sum_{f\in F}i_{pf}w_f$。$n_p$ 为当需要子灾难恢复计划 p 时，取值为 1，否则为 0。q_r 为资源 r 的数量。q_{pr} 为子灾难恢复计划 p 需要的资源 r 的数量。c_r 为资源 r 的单价。如果 $r\in R_o\cup R_m$，则 c_r 指一次性所付的费用。如果 $r\in R_h$，则 c_r 指相应的租金或薪水。e_r 为当选中资源 r 时，取值为 1，否则为 0。

目标函数：

$$\text{Max}\quad \sum_{s\in S}n_p m_p \tag{1}$$

约束条件：

$$\sum_{p\in P_f}n_p\le v,\quad \forall f\in F,\quad v=1,2,3,\cdots \tag{2}$$

$$\sum_{p\in P_{c_i}}n_p\le 1,\ \forall 1\le i\le k \tag{3}$$

$$\sum_{p\in P_r}n_p q_{pr}-e_r q_r\le 0,\quad \forall r\in R_o \tag{4}$$

$$\underset{p\in P_r}{\text{Max}}(n_p q_{pr})-e_r q_r\le 0,\quad \forall r\in R_m\cup R_h \tag{5}$$

$$\sum_{r\in R_o\cup R_m}e_r c_r+\sum_{p\in P}n_p\left(\sum_{r\in R_h}e_r c_r\right)\le B \tag{6}$$

使式(1)具有最大值的 SDRP 构成的集合就是最优的 SDRP 集合。

2.2 模型分析

首先必须深入研究系统中 A, F, R 和 P 及其之间的关系，在此基础上，还必须确保 DRP 中没有遗漏任何重要资源，使 DRP 具有完整性。约束条件式(2)使得针对每个设施，最多只能选中 v 个 SDRP($v=1,2,3,\cdots$)。当 $v=1$ 时，限制条件最严格，得到的结果也最好，意味着针对每个设施，只选中 1 个 SDRP。当限制条件过于严格时，可能由于不同 SDRP 在保护不同设施时产生冲突，从而无法得出最优解。例如 SDRP p_1 能够保护设施 f_1 和 f_2，SDRP p_2 能够保护设施 f_2 和 f_3，如果需要使用 p_1 和 p_2 同时保护设施 f_1，f_2 和 f_3 时将会产生冲突。此时可以放松限制条件，逐渐增大 v 的值，直到能够找到最优解为止。约束条件式(3)保证所有选中的 SDRP 之间不会产生冲突，保证选中的 SDRP 需要的所有人力、物力、甚至时间等资源是都可行的，不同 SDRP 之间不会产生冲突。必须避免 SDRP 单独可行，不同 SDRP 在相同时刻却由于资源竞争冲突导致它们在一起时却不可行这种现象。约束条件式(4)保证所需要的一次付费、一次使用的资源能够满足要求，约束条件式(5)保证使选中 SDRP 需要的可以多次使用的资源也满足要求。约束条件式(6)使最优解中 SDRP 耗用资源的费用(一次性费用和累积租用费用)不会超过预算。

模型中，最初只有 s_a 和 d_{fa} 是主观参数，w_f 和 m_p 是通过两个主观参数计算得出的参数，所有其它参数具有客观特性。这能够减少人为主观因素对评估 DRP 而产生的影响。为了使模型尽可能的精确，必须在最初认真评估 s_a 和 d_{fa} 这两个主观参数。

相比其它 DRP 模型，本文中的模型结合实际需求对不同种类的资源进行详细分类，分别评估其重复使用情况和费用，能够更加精确地评估灾难恢复计划；模型充分考虑到了不同 SDRP 之间可能的冲突，并给出数学方法加以解决，使其优越于其它 DRP 数学模型；模型具有较少的参数，使其更易于实施；模型中除少数参数与人为的主观评价有关外，绝大多数都能够给出精确的评估，因此，模型能够对灾难恢复计划给出更加客观、精确的评估。

最后需要说明的是，需要对 DRP 进行尽可能充分的测试。由于完整测试 DRP 非常困难，因此必须首先集中测试 DRP 中的所有 SDRP，保证所有 SDRP 单独可行。在此基础上根据实际情况尽量制订多种联合测试方案，确保多个 SDRP 在一起时也是可行的，从而在某种程度上确保 DRP 的有效性。

2.3 模型实施

在研究被保护信息系统和灾难恢复系统的基础上，这个数学模型能够通过以下步骤实现对 DRP 的评估分析：

(1)评估系统中所有重要应用，根据每个应用的重要程度赋以相应权值，构建应用向量 $(s_{a_1},s_{a_2},s_{a_3},\cdots)$。向量中元素记录应用的权值，元素的下标代表系统中不同的应用。

(2)对所需资源进行标识，对资源进行分类，构建 3 个资源数量向量 $(R_{o_1},R_{o_2},R_{o_3},\cdots)$，$(R_{m_1},R_{m_2},R_{m_3},\cdots)$ 和 $(R_{h_1},R_{h_2},R_{h_3},\cdots)$，分别记录系统中 3 种不同类型的资源数量。构建资源代价向量 $(R_{o_1},R_{o_2},\cdots,R_{m_1},R_{m_2},\cdots,R_{h_1},R_{h_2},\cdots)$ 记录资源的单价。

(3)评估系统中需要被保护的所有设施，以及设施对应用的影响。通过这一步骤，构建设施矩阵：

$$\begin{bmatrix} d_{f_1a_1} & d_{f_1a_2} & d_{f_1a_3} & \cdots \\ d_{f_2a_1} & d_{f_2a_2} & d_{f_2a_3} & \cdots \\ d_{f_3a_1} & d_{f_3a_2} & d_{f_3a_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

矩阵的行下标代表不同设施，列下标代表不同应用，矩阵的元素记录某个设施对某一应用产生影响的可能性。通过这个设施矩阵，可以计算出设施权值向量$(w_{f_1}, w_{f_2}, w_{f_3}, \cdots)$，该向量记录每个设施对所有应用总的影响程度。

(4)深入研究所有 SDRP 并构建 SDRP 设施矩阵：

$$\begin{bmatrix} i_{p_1f_1} & i_{p_1f_2} & i_{p_1f_3} & \cdots \\ i_{p_2f_1} & i_{p_2f_2} & i_{p_2f_3} & \cdots \\ i_{p_3f_1} & i_{p_3f_2} & i_{p_3f_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

矩阵的行下标代表不同 SDRP，列下标代表不同设施，矩阵的元素值为 1 或 0，分别代表相应的 SDRP 是否能够保护相应的设施。类似可以构建 SDRP 资源矩阵，矩阵的元素取值代表 SDRP 需要占用的相应资源的数目：

$$\begin{bmatrix} q_{p_1r_1} & q_{p_1r_2} & q_{p_1r_3} & \cdots \\ q_{p_2r_1} & q_{p_2r_2} & q_{p_2r_3} & \cdots \\ q_{p_3r_1} & q_{p_3r_2} & q_{p_3r_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

(5)借助于 SDRP 资源矩阵，分析系统并构建 SDRP 冲突矩阵，矩阵的行下标和列下标都代表不同应用，矩阵的元素为 1 或 0，分别代表不同应用之间是否会产生冲突：

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

(6)基于这个数学模型和以上步骤，设计开发软件工具，或借助于一些优秀的软件(例如 CPLEX)，设计开发辅助工具，辅助决策者进行灾难恢复决策。

3　模型测试

不同 DRS 差异巨大，不同 DRP 也具有不同特点，在对 DRP 的测试上也没有通用标准。为测试该模型，本文假定了一些场景。表 1 是测试场景中有关应用、设施、资源、SDRP 和冲突的测试参数。测试中，假定系统中有 10 和 20 个需要保护的关键应用，由于设施往往可以共享，因此设置设施数量为 15，20，30，40。假定 SDRP 的数量是相应应用数量的两倍多，因此设置 SDRP 的数量为 20，25，40，50。在测试中设置冲突的 SDRP 的数量分别为 0，1，2，3 个，用于测试模型对冲突的解决能力。在每个测试场景中随机生成模型中其它的参数。在每个测试场景中，我们均能够成功找到 SDRP 的最优集合。通过测试，可以确信模型能够正确评估 DRP。

表 1　测试场景

A	F	R_o	R_m	R_h	P	冲突
10	15	15	10	5	20	0
10	15	15	10	5	20	1
10	20	15	10	5	25	0
10	20	15	10	5	25	1
20	30	25	15	10	40	0
20	30	25	15	10	40	1
20	30	25	15	10	40	2
20	30	25	15	10	40	3
20	40	25	15	10	50	0
20	40	25	15	10	50	1
20	40	25	15	10	50	2
20	40	25	15	10	50	3

4　结束语

在研究信息系统和灾难恢复计划的基础上，论文使用最优化理论提出一个灾难恢复计划数学模型，并给出了模型分析和实施。模型对现实世界进行了必要的抽象和简化，具有较少的参数，能够解决 SDRP 之间的冲突，对资源的分类，实现对 DRP 的精确评估。实验证明了模型的正确性和可行性。

参 考 文 献

[1] Petroski H. Technology and architecture in an age of terrorism[J]. *Technology in Society*, 2004, 26(2-3): 161–167.

[2] Anon. Blackouts. Threat of terrorism spur disaster recovery planning[J]. *International Journal of Micrographics and Optical Technology*, 2003, 21(4-5): 2–2.

[3] Wang K, Su R D, and Li Z X, *et al*.. Robust disaster recovery system model[J]. *Wuhan University Journal of Natural Sciences*, 2006, 11(1): 170–174.

[4] Chow W S. Success factors for IS disaster recovery planning in Hong Kong[J]. *Information Management and Computer Security*, 2000, 8(2): 80–86.

[5] Smith D R, Cybrowski W J, and Zawislan F, *et al*.. Contingency/disaster recovery planning for transmission systems of the defense information system networks[J]. *IEEE Journal on Selected Areas in Communications*, 1994, 12(1): 13–22.

[6] Hayes P E and Hammons A. Disaster recovery project management[C]. Proceedings of IEEE 47th Petroleum and Chemical Industry Conference, San Antonio, TX, USA, 2000: 55–63.

[7] Lam W. Ensuring business continuity[J]. *IT Professional*, 2002, 4(3): 19–25.

[8] Uddin N and Engi D. Disaster management system for southwestern Indiana[J]. *Natural Hazards Review*, 2002,

3(1): 19–30.

[9] Fallara P. Disaster recovery planning[J]. *IEEE Potentials*, 2004, 22(5): 42–44.

[10] Hawkins S M, Yen D C, and Chou D C. Disaster recovery planning: A strategy for data security[J]. *Information Management and Computer Security*, 2000, 8(5): 222–229.

[11] Corley J and Lejerskar D. Homeland defense center network-capitalizing on simulation modeling and visualization for emergency preparedness response and mitigation[C]. Proceedings of the 2003 Winter Simulation Conference: Driving Innovation, Piscataway, NJ, USA, 2003: 1061–1067.

王 琨： 男，1973 年生，博士生，研究方向为网络与信息安全.

尹忠海： 男，1964 年生，博士生，副教授，研究方向为网络与信息安全.

周利华： 男，1942 年生，教授，博士生导师，研究方向为网络与信息安全，网络多媒体.

蔡 震： 男，1976 年生，高级工程师，研究方向为网络与信息安全.