

· 实践与应用 ·

军用软件测试过程质量因素与度量指标*

杨玲萍¹ 张 丽²

(1 中国电子科技集团公司第二十八研究所 南京 210007)

(2 海军装备采购论证中心 北京 100071)

摘 要: 软件测试过程质量直接决定了测试质量,从而在一定程度上决定了软件质量。其间量化软件测试过程质量评价至关重要。描述了军用软件测试过程活动,分析了测试过程质量影响因素,构建了测试过程质量度量指标。实例应用证明,该评价方式可对量化测试过程质量提供有效支持。

关键词: 软件测试;测试过程质量;度量指标;目标-问题-度量(GQM)范式

中图分类号: TP311.52 **文献标识码:** A **文章编号:** 1674-909X(2014)03-0076-06

Influence Factor and Measurement Index for Testing Process Quality of Military Software

Yang Lingping¹ Zhang Li²

(1 The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

(2 The Purchase Argument Center of Naval Equipment, Beijing 100071, China)

Abstract: The testing process quality of the control software directly determines the testing quality, thus determining the software quality to some extent. The evaluation quantization of the software testing process quality is important. The military software testing process is described, influence factors of testing process quality are analyzed and the measurement index of the testing process quality is constructed. Examples prove that the evaluation method can effectively support the quantization of the testing process quality.

Key words: software testing; testing process quality; measurement index; goal, question, measurement (GQM) pattern

0 引 言

软件测试是保证软件质量不可或缺的重要手段。对于功能众多、软件之间以及软件内部信息关系复杂的军用软件来说,软件质量好坏直接影响作战效能和任务成败,需投入时间和资源予以测试保证。如何花费较少的成本和时间发现较多的软件缺陷,控制软件测试过程质量显得尤为关键,使得测试过程受控和管理。本文通过分析军用软件测试过程各阶段的活动要素,采用目标-问题-度量(GQM)范

式^[1],从影响测试的有效性、充分性和测试效率3方面的因素进行分析,总结出可以表征军用软件测试3方面度量指标,最后以实例化应用说明度量指标的有效性。

1 软件测试过程与活动

软件测试过程是软件工程的组成部分,开发过程质量决定了软件质量,而测试过程质量直接影响测试结果的准确性和有效性。软件测试过程与软件开发过程一样,遵循软件工程和管理学原理^[2]。软

* 基金项目:国防科工局“十二五”基础科研资助项目。

收稿日期:2013-12-10

件测试过程由不同阶段一系列活动组成,只有稳定的测试过程才是可度量和可监控预测的。军用软件测试过程有明确要求^[3],其过程分为测试需求分析、测试策划、测试设计与实现、测试执行和测试总结(包括评价过程和总结)5个阶段。它是一个完整的全面测试过程,每个阶段的输出是下一阶段的输入,确保测试过程规范化和稳定化。全面测试需对需求文档、设计文档、代码和用户文档等进行全面测试。软件测试过程与活动如图1所示。

1) 测试需求分析

关键活动:确定需要的测试类型及其测试要求;确定每个测试项的测试充分性要求。按确定的文档编写软件测试需求规格说明,并通过测试需求评审。测试类型包括功能测试、性能测试、接口测试、人机交互界面测试、强度测试、余量测试、容量测试、安全性测试、安装性测试、边界测试、恢复性测试、数据处理测试、可靠性测试、互操作测试、敏感性测试、代码审查、代码走查、逻辑测试、静态分析和文档审查等,测试要求包括状态、接口、数据结构和设计约束等。

活动输入:研制总要求、技术方案或其他等效文件,被测软件需求规格说明和软件设计文档。

活动输出:软件测试需求规格说明。

2) 测试策划

关键活动:确定测试策略,明确用于测试的资源

和技术要求;明确软件评价准则和方法、测试活动进度和人员安排;按确定的文档编写软件测试计划,并通过测试计划评审。

活动输入:研制总要求、技术方案或其他等效文件,被测软件需求规格说明和软件设计文档。

活动输出:软件测试计划。根据需要可将测试计划与测试需求规格说明合并为测评大纲或测试计划。

3) 测试设计和实现

关键活动:分解软件测试项,说明每个测试项的测试用例设计方法以及测试数据的选择依据;获取并验证测试数据,获取测试资源;开发必要的测试支持软件,如模拟器软件;建立并校核测试环境;按确定的文档编写软件测试说明。在转入测试执行阶段前,应进行测试就绪评审,以确定能否开始测试。

活动输入:软件测试计划和软件测试大纲。

活动输出:软件测试说明(含测试用例集)和准备的测试数据。

4) 测试执行

关键活动:执行测试用例,获取测试结果;分析并判定测试结果,同时根据不同判定结果采取相应措施;对测试过程的正常或异常终止情况进行核对,并根据核对结果,对未达到测试终止条件的测试用例,决定是停止测试,还是修改或补充测试用例集进一步测试,直至全部用例执行和软件问题回归测试

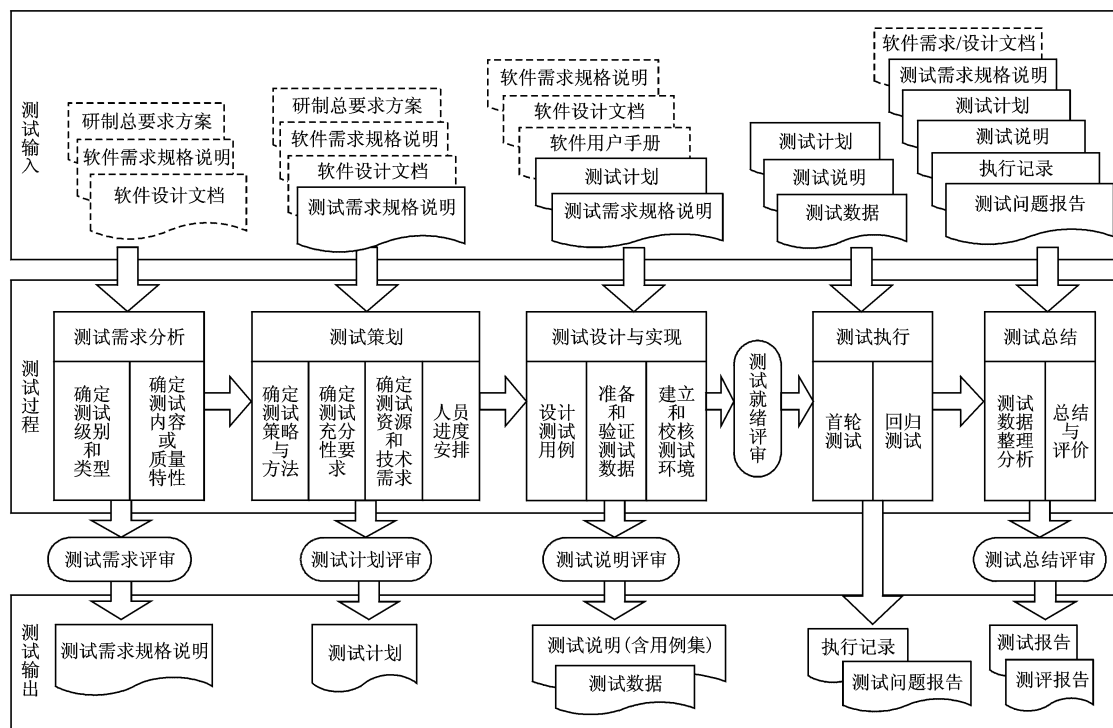


图1 软件测试过程与活动

完成。

活动输入:软件测试计划、软件测试大纲、测试说明和测试用例。

活动输出:测试执行记录和测试问题报告。

5) 测试总结

关键活动:分析和评价测试工作,确定无法解决的软件测试事件并说明不能解决的理由;分析和评价被测软件,总结测试反映的被测软件与软件需求(或软件设计)之间的差异,对测试问题和结果进行分类和总结;按确定文档编写测试报告。

活动输入:被测软件文档、测试计划、测试说明、测试执行记录和测试问题报告。

活动输出:软件测试报告。

2 测试过程质量因素分析与度量指标

一个良好成熟度的软件测试过程不仅是定性描述,而且是可量化和度量的过程,其度量指标选取是关键因素。本着有效性和易实施性原则,应能准确刻画过程属性,用以了解、控制和改进测试过程,且易于采集。本文分析了军用软件测试过程质量的内涵和影响因素,为量化测试过程的有效性、充分性和测试效率 3 方面的质量度量,以 GQM 范式为驱动,构建测试过程各阶段度量指标。

2.1 测试需求分析

软件测试需求分析质量好坏对被测对象的测试充分性产生影响。在进行测试需求分析时,需根据被测软件特征和重要性程度,在分析被测软件状态、接口、数据结构和设计约束基础上,选取测试类型,确定相应测试内容和要求。同时,基于每个测试类型进行测试项分解,分解后的测试项需全部覆盖软件设计文档规定的所有功能点,包括关键、重要和一般功能点。

在测试需求分析阶段,为保证测试需求分析充分性目标,采用 GQM 范式,针对测试需求覆盖软件需求充分性以及测试类型选取充分性等问题,需选取以下度量元:总功能点数、提取的关键/重要/一般需求功能点数、需求涉及的测试类型、计划执行的测试类型、关键/重要/一般功能测试项覆盖率和测试类型覆盖率。构成一组三层次结构的测试需求分析过程度量指标如表 1 所示。

2.2 测试策划

软件测试策划质量会影响测试过程的有效性和

表 1 测试需求分析过程度量指标

目 标	问 题	间接度量元	度量元
良好测试需求分析充分性	测试需求覆盖软件需求充分性	关键功能测试项覆盖率	总功能点数 提取的关键、重要和一般需求功能点数
		重要功能测试项覆盖率	
		一般功能测试项覆盖率	
	测试类型选取充分性	测试类型覆盖率	需求涉及的测试类型 计划执行的测试类型

注:1) 关键功能测试项覆盖率=提取的关键需求功能点数/总功能点数;

2) 重要功能测试项覆盖率=提取的重要需求功能点数/总功能点数;

3) 一般功能测试项覆盖率=提取的一般需求功能点数/总功能点数;

4) 测试类型覆盖率=计划执行的测试类型/需求涉及的测试类型。

测试效率。在测试策划时,需根据被测软件特征和测试类型,构建测试环境;同时,根据测试资源情况和计划执行的测试项,规划测试活动进度、需要的人员数和技能要求等。因此,在测试策划阶段,采用 GQM 范式,以良好的测试策划充分性和有效性为目标,针对测试环境平台构建有效性、测试团队角色分配有效性、测试时间估计有效性等问题,需选取以下度量元:总功能点数、测试环境支持的功能点数、测试环境对测试需求的支持率、测试团队各类角色人员数及比例、测试计划工作时间和有效工作时间等。构成一组三层次结构的测试策划过程度量指标如表 2 所示。

表 2 测试策划过程度量指标

目 标	问 题	间接度量元	度量元
良好测试策划有效性和充分性	测试环境平台构建有效性	测试环境对测试需求的支持率	总功能点数 测试环境支持的功能点数
			设计人员数 执行人员数 管理人员数 测试人员总数
	测试团队角色分配有效性	设计人员比例 执行人员比例 管理人员比例	
	测试时间估计有效性	有效工作时间	熟悉软件花费的时间 代码审查时间 测试设计花费的时间 测试执行时间 测试总结时间 测试计划工作时间

注:1) 测试环境对测试需求的支持率=测试环境支持的功能点数/总功能点数;

2) 设计人员比例=设计人员数/测试人员总数;

3) 执行人员比例=执行人员数/测试人员总数;

4) 管理人员比例=管理人员数/测试人员总数;

5) 有效工作时间=(熟悉软件花费的时间+代码审查时间+测试设计花费的时间+测试执行时间+测试总结时间)/测试计划工作时间。

2.3 测试设计和实现

软件测试设计质量会对被测对象的测试有效性、充分性和测试设计效率产生影响。在进行测试设计时,需对测试需求分析得到的测试项进行层次化分解,设计测试用例的实施步骤;同时,根据测试策划结果建立和校核测试环境,准备和验证所有测试用数据。而测试用例有效性依赖于测试环境支持情况,测试用例充分性取决于用例对软件所有功能点的覆盖情况,测试用例设计效率依赖于测试团队人员能力水平等。因此,在测试设计与实现阶段,采用 GQM 范式,以良好的测试设计充分性、有效性和测试设计效率为目标,针对测试用例设计的有效性、充分性和效率等问题,需选取以下度量元:总的设计用例数、合格用例数、平台支持的用例数、用例运行发现的问题数、用例合格率、关键/重要/一般功能点用例数及其用例覆盖率、动态测试设计时间和用例设计效率。构成一组三层次结构的测试设计和实现过程度量指标,如表 3 所示。

表 3 测试设计和实现过程度量指标

目 标	问 题	间接度量元	度量元
良好测试设计有效性、充分性和效率	测试用例设计有效性	用例合格率 用例缺陷发现率	合格用例数 总的设计用例数 平台支持用例数 用例运行发现问题数
	测试用例设计充分性	关键功能点用例覆盖率 重要功能点用例覆盖率 一般功能点用例覆盖率	关键、重要和一般功能点数 用例设计覆盖的关键、重要和一般功能点数
	测试用例设计效率	用例设计效率	总的设计用例数 动态测试设计时间

注:1) 用例合格率=合格用例数/总的设计用例数;
2) 用例缺陷发现率=用例运行发现问题数/平台支持用例数;
3) 关键功能点用例覆盖率=用例设计覆盖关键功能点数/关键功能点数;
4) 重要功能点用例覆盖率=用例设计覆盖重要功能点数/重要功能点数;
5) 一般功能点用例覆盖率=用例设计覆盖一般功能点数/一般功能点数;
6) 用例设计效率=总的设计用例数/动态测试设计时间。

2.4 测试执行

测试执行质量直接影响测试过程的有效性、充分性和效率。按照测试计划和测试说明的内容和要求,在执行测试时,一方面,实施代码走查/审查和设计文档审查等静态测试;另一方面,实施动态测试,在构建的测试环境中运行测试用例,判定测试用例是否通过,分析是否需进行补充测试。最后根据测

试发现的问题,对修改后被测软件进行回归测试,重复该过程,直至所有问题归零。测试执行的有效性依赖于测试用例缺陷的发现率、有效性和检错率等;测试执行的充分性取决于测试用例对需求的覆盖率、用例执行率、代码走查/审查力度和文档审查检错率等;测试执行的效率依赖于代码审查效率、测试人员工作效率和当前过程工作偏差率等。测试执行过程度量指标如表 4 所示。在测试执行阶段,采用

表 4 测试执行过程度量指标

目 标	问 题	间接度量元	度量元
良好测试执行有效性、充分性和效率	动态测试充分性	用例执行率 用例对需求的覆盖率 代码覆盖率	测试执行用例数 平台支持用例数 已执行用例覆盖功能点数 总功能点数 语句覆盖率 分支覆盖率 条件覆盖率 条件组合覆盖率
	代码走查/审查力度	代码审查率	有效代码行数 审查过的有效代码行数
	代码审查效率	代码审查效率	审查过的有效代码行数 代码审查时间
良好测试执行有效性、充分性和效率	当前过程有效性	当前过程的问题归零率 缺陷有效率 文档审查检错率 代码审查检错率 动态测试检错率	已归零问题数 开发方确认问题数 文档审查发现问题数 代码审查发现问题数 动态测试发现问题数 测试发现总问题数
	测试人员工作效率	有效工作时间比	熟悉软件花费的时间 测试执行时间 测试总结时间 测试中断时间 测试等待时间
	当前过程工作偏差情况	当前过程工作偏差率	当前过程计划工作量 当前过程实际投入工作量 测试计划工作量

注:1) 用例执行率=测试执行用例数/平台支持用例数;
2) 用例对需求覆盖率=已执行用例覆盖功能点数/总功能点数;
3) 代码审查率=审查过的有效代码行数/有效代码行数;
4) 代码审查效率=审查过的有效代码行数/代码审查时间;
5) 当前过程的问题归零率=已归零问题数/开发方确认问题数;
6) 缺陷的有效率=开发方确认问题数/测试发现总问题数;
7) 文档审查检错率=文档审查发现问题数/开发方确认问题数;
8) 代码审查检错率=代码审查发现问题数/开发方确认问题数;
9) 动态测试检错率=动态测试发现问题数/开发方确认问题数;
10) 有效工作时间比=(熟悉软件花费的时间+测试执行时间+测试总结时间)/(熟悉软件花费的时间+测试执行时间+测试中断时间+测试等待时间);
11) 当前过程工作偏差率=(当前过程计划工作量-当前过程实际投入工作量)/测试计划工作量;
12) 代码覆盖率强弱顺序^[4]:条件组合覆盖率→条件覆盖率→分支覆盖率→语句覆盖率。若 4 种覆盖率均达 100%,则为包含关系,即条件组合覆盖率达 100%,可判断其他 3 种覆盖率均达 100%。实际测试时,需具体分析被测对象后提出代码覆盖率指标要求,作为测试依据。

GQM 范式,以良好的测试执行充分性、有效性和执行效率为目标,针对动态测试的充分性、代码走查/审查力度、当前过程有效性、当前过程工作偏差情况、测试人员工作效率等问题,需选取以下度量元:测试执行用例数、平台支持用例数、语句/分支/条件/条件组合覆盖率、用例执行率、用例对需求覆盖率、有效代码行数、代码审查时间、代码审查效率、总问题数、缺陷有效率、已归零问题数和有效工作时间比等。

2.5 测试总结

按照测试计划,在完成全部测试项及其测试用例后,应对测试工作和被测软件进行分析和评价,分类和总结测试结果及软件缺陷,报告软件测试结果。利用 GQM 范式,以良好的测试过程有效性、测试过程充分性和测试效率为目标,针对是否结束本次测试工作问题,选取的度量元应是测试需求分析、测试策划、测试设计和测试执行 4 个测试阶段度量指标的全部集合。

3 应用实例

本章以一个应用实例说明上述评价方式的可行性。实例基本信息如下:

被测软件类型:非嵌入式

被测软件规模:989 万行

软件运行平台:Windows2000 Server sp4

软件开发平台:VC/C++

测试级别:配置项级和系统级

使用测试工具:Klockwork

基于本文构建的测试过程度量指标,采用模糊数学的模糊综合评价原理^[5],运用文献[6]中软件测试过程模糊决策模型,评价实例在软件测试过程的测试有效性、测试充分性和测试效率 3 方面的测试质量。

在模糊决策模型中评价集选取如下:

测试过程充分性: $V=\{v_1,v_2,v_3,v_4\}=\{\text{不充分,基本充分,较充分,充分}\};$

测试过程有效性: $V=\{v_1,v_2,v_3,v_4\}=\{\text{无效,基本有效,较好,好}\};$

测试效率: $V=\{v_1,v_2,v_3,v_4\}=\{\text{低,中等,较高,高}\}。$

在该实例计划执行的测试工作全部完成后,以本文提出的度量指标为评价指标,采集实例测试过程基础数据,测试过程充分性评价指标及基础数据如表 5 所示。对实例的测试过程在测试过程的有效性、充分性和效率 3 方面分别进行评价指标权重分配、单因素评价和综合评价,计算得到评价结果,如表 6 所示。

表 5 测试过程充分性评价指标及基础数据

主因素(第一层指标)	权重	子因素(第二层指标)	权重	度量元(第三层指标)	基础数据
测试类型选取充分性	0.213	测试类型覆盖率	1.000	需求涉及的测试类型	13
				已执行测试类型	13
测试用例设计充分性	0.334	关键功能点用例覆盖率	0.479	关键功能点数	61 个
				用例设计覆盖到的关键功能点数	61 个
		重要功能点用例覆盖率	0.321	重要功能点数	306 个
				用例设计覆盖到的重要功能点数	306 个
		一般功能点用例覆盖率	0.200	一般功能点数	702 个
				用例设计覆盖到的一般功能点数	702 个
动态测试充分性	0.240	用例执行率	0.257	测试执行的用例数	3 624 个
				平台支持的用例数	3 624 个
		用例对需求覆盖率	0.397	一般功能点数	702 个
				总的功能点数	1 069 个
				关键功能点数	61 个
				重要功能点数	306 个
		代码覆盖率	0.346	已执行用例覆盖到的功能点数	1 069 个
				分支覆盖率	0
				语句覆盖率	0
				条件覆盖率	0
条件组合覆盖率	0				
代码走查/审查力度	0.213	代码审查率	1.000	有效代码行数	989 万行
				审查过的有效代码行数	350 万行

表6 测试过程质量模糊综合评价结果

测试过程 指标	最大隶属度法	模糊分布法	中心平均解模糊法 评判得分
充分性	充分	83.88%(充分)	93.924
有效性	有效性好	79.46%(有效性好)	92.714
效率	较高	45.22%(较高)	78.006

4 结束语

基于度量的测试过程管理是高质量测试的重要标志。本文描述了软件测试过程活动,分析了影响测试过程质量的因素,建立了软件测试过程度量所涉及的指标。实例应用表明,测试过程度量指标的建立与应用,能够为量化测试过程质量评价提供基础支持,使评价结果真实、准确和客观。

参考文献(References):

- [1] 杨玲萍,蔡东华,王建强. 软件测试效率度量指标体系研究[J]. 指挥信息系统与技术,2010,1(6):1-5.
Yang Lingping, Cai Donghua, Wang Jianqiang. Research on the metric index system of software testing

(上接第11页)

动的指挥处置环节或片段进行规范化设计,依据各种作战条令和政策法规进行关联互动,实现流程化体系联动;是一个具有高度政策性、技术性和渗透性的跨领域跨部门的系统工程,对提升空军作战行动指挥处置能力,具有重要的理论价值和现实意义。能够实现作战指挥处置流程、组织与信息系统融合一体,从根本上解决信息技术效力发挥的传统结构性矛盾,提高信息化作战指挥效能。为指挥信息系统建设提供正确的作战需求和发展思路,从而推动信息技术发展,实现作战需求与信息技术发展的互动协调。

参考文献(References):

- [1] 王伟海. 加快构建中国特色现代军事力量体系[J]. 前线,2013(5):48-50.
Wang Weihai. Speed the construction of modern military power system with Chinese characteristics[J]. Battlefield,2013(5):48-50. (in Chinese)
- [2] 陈志新,谈兴秋. 指挥信息系统流程控制技术[J]. 指挥信息系统与技术,2013,4(3):20-24.
Chen Zhixin, Tan Xingqiu. Workflow control technology for command information system[J]. Command Information System and Technology,2013,4(3):20-24. (in Chinese)
- [3] 陈建林,巴宏欣,朱孟平,等. 联合作战共用战场态势图的构建[J]. 指挥控制与仿真,2013,35(2):21-24.

effectiveness[J]. Command Information System and Technology,2010,1(6):1-5. (in Chinese)

- [2] Rex B. 软件测试过程管理:Managing the testing process[M]. 龚波,但静培,林生,等译. 2版. 北京:机械工业出版社,2003.
- [3] 中国人民解放军总装备部电子信息基础部. 军用软件测试实验室测评过程和技术能力的要求[S]. 北京:中国人民解放军总装备部,2005.
- [4] Pressman R S. 软件工程:实践者的研究方法:Software engineering;a practitioner's approach[M]. 梅宏,译. 4版. 北京:机械工业出版社,1999.
- [5] 杨纶标,高英仪,凌卫新. 模糊数学[M]. 广州:华南理工大学出版社,2001.
- [6] 杨玲萍. 军用软件测试过程模糊决策技术研究及应用[D]. 南京:东南大学软件学院,2006:30-53.

作者简介:

杨玲萍,女(1962—),研究员级高级工程师,研究方向为软件过程改进、软件测试方法与评估技术。
张丽,女(1975—),高级工程师,研究方向为系统总体与集成以及软件质量度量。

Chen Jianlin, Ba Hongxin, Zhu Mengping, et al. Constructing common battle-space situation picture for joint operation[J]. Command Control & Simulation, 2013,35(2):21-24. (in Chinese)

- [4] 江振华. 无线电台抗复杂电磁干扰的几点措施[J]. 四川兵工学报,2010,31(6):46-55.
Jiang Zhenhua. Measurements of anti-electromagnetic interference on radio[J]. Journal of Sichuan Ordnance,2010,31(6):46-55. (in Chinese)
- [5] 陈鸿. 战场环境建模与态势生成关键技术研究[D]. 长沙:国防科学技术大学信息与通信工程,2010.
- [6] 张得志,张洁. EX-XPDL 集成化企业建模和仿真的研究[J]. 机械设计与制造,2005(9):122-124.
Zhang Dezhi, Zhang Jie. Research on EX-XPDL integrated enterprise modeling and simulation[J]. Machinery Design & Manufacture,2005(9):122-124. (in Chinese)

作者简介:

朱伟清,男(1963—),处长,研究方向为空军指挥信息系统。
陈朝阳,男(1969—),副处长,研究方向为空军指挥信息系统。
高勇,男(1980—),参谋,研究方向为空军指挥信息系统。
袁峰,男(1982—),参谋,研究方向为空军指挥信息系统。