

# 基于 SLA 规范的服务性能管理

朱浩洋, 崔鹏帅, 曹介南

(国防科学技术大学 计算机学院, 湖南 长沙 410073)

**摘要:**服务等级协议被广泛应用于描述服务质量所需达到的要求。根据服务等级协议构建了服务质量模型。在这个质量模型中包含两种度量:一种是原子度量,一种是复合度量。根据这些度量参数开发了监测 SOA 系统的质量管理系统。这个管理系统由四个主要部分组成:监测器、QoS 计算器、SLA 规范分析器和修复方案。在不同的环境中测试了该管理系统,结果显示系统能够完成监测 QoS 并且分析 SLA 违例。由于系统监测接口是与技术无关的,所以能够拓展并且应用到其他种类的 SOA 系统中。

**关键词:**SLA 规范;QoS;度量;质量模型;SOA 架构;管理系统

中图分类号:TP302.1

文献标识码:A

文章编号:1673-629X(2014)04-0012-05

doi:10.3969/j.issn.1673-629X.2014.04.003

## Management of Service Performance Based on SLA

ZHU Hao-yang, CUI Peng-shuai, CAO Jie-nan

(College of Computer, National University of Defense Technology,  
Changsha 410073, China)

**Abstract:**Service Level Agreement (SLA) is widely used in defining requirements which the quality of a service must fulfill. Build the service quality model based on the SLA. There are two kinds of metrics in the quality model, one is atomic and the other one is composite. According to metrics in the model develop a quality management system to monitor services in a SOA system. This management system consists of four main parts, including monitor, QoS calculator, SLA analyzer and repair scheme. The system has been tested in many different situations. The result shows that the system is competent to monitor QoS and analyze SLA violations. As the system uses a technology independent on monitoring interface so that it can be extended and adopted in different kinds of SOA systems.

**Key words:**SLA; QoS; metrics; quality model; SOA; management system

## 0 引言

SOA 因为能利用现有的服务资源被视为一种有效的低成本开发手段, SOA 系统与传统的系统不同, 其特性为:松耦合、不断变化、黑盒、可拓展和不均一性等。使得在管理 SOA 系统时往往存在不断增加的管理成本和不断下降的管理效率等问题, 与此同时还伴随着一些难以预测和难以解决的系统内部服务及其相关问题。

考虑用户在 SOA 系统中使用服务的一般情况。用户首先搜索服务注册中心 (UDDI), 选择出合适的服务, 然后引用这个服务。但是如果服务在大于 SLA 规范中规定的最大响应时间时仍没有响应, 用户怀疑服务出现了问题, 却没有办法了解或者修复这个问题, 于是用户放弃这个服务。而在服务器端, 服务的管理

者并不能一定感知到服务出现了问题。这样的结果是服务被弃用。如果在 SOA 系统中加入服务管理系统, 在上述情况下, 被引用的服务被一直监测。如果监测器发现了服务的响应时间超过了 SLA 规范规定的最大响应时间, 服务管理系统按照 SLA 规范中的规定通知管理者和提供者并且采取相关措施。这两种截然不同的结果表明了在 SOA 系统中加入服务管理系统的重要性。

文中展示的服务管理系统是基于 SLA 规范的服务管理系统, 它能够在运行时监测服务的性能参数, 并且当探测到违反 SLA 规范的情况时采取相应的备用措施。现在该系统主要是针对网络服务, 但这个服务管理系统的接口是与技术无关的, 因此它能够应用在不同类型的 SOA 系统中。

收稿日期:2013-06-17

修回日期:2013-09-25

网络出版时间:2014-01-28

基金项目:国家自然科学基金资助项目(61170285)

作者简介:朱浩洋(1988-),男,硕士研究生,研究方向为计算机网络服务管理;曹介南,副教授,研究方向为计算机网络。

网络出版地址:<http://www.cnki.net/kcms/detail/61.1450.TP.20140128.1152.052.html>

## 1 相关工作

服务管理是 SOA 领域非常重要的议题。它包含的内容非常广泛,从安装、配置到收集性能信息行相关调整确保服务的执行。为了有效地管理负责的 SOA 系统,出现了很多研究方法。

Zhou、Chia 和 Lee 提出了自动在 Axis2 的运行链上安装网络服务监测器的方法<sup>[1]</sup>。Moser 等展示了一种面向方面的架构 VieDAME,用来监测 BPEL 过程并且最终实现服务的自适应<sup>[2]</sup>。Benharref 等提出了一个多观察者模式的架构,在客户端和服务端都布置监测器来获取网络服务的性能参数<sup>[3]</sup>。Comuzzi 等采用了一个基于事件的监测方式用来评价业务、软件和基础结构层的 SLA 规范<sup>[4]</sup>。Bertolino 等使用主动测试的方法来评价网络服务的 SLA 规范。它使用 WSDL 和 SLA 规范来进行主动测试<sup>[5]</sup>。Menasce 等提出了一个关于性能参数的服务管理的架构。在这个架构中有两个关键组件,性能参数中间人和服务提供者组件<sup>[6]</sup>。尽管在性能参数中间者、服务提供者组件和客户端之间有 7 个相互交互过程,但是在他们的技术方案中使用简单的布尔表达式来判断和计算性能参数属性而不是通过综合地使用布尔值来判断性能参数是否满足。所以方案在复杂的环境中频繁计算性能参数有局限性。

## 2 基于 SLA 规范的服务质量管理模型

在这个模型中根据 SLA 规范语言,针对网络服务设计了服务的质量模型与相关度量。

### 2.1 SLA 规范语言

有很多 SLA 规范语言被提出来,尽管它们的表达方式各有差异,但是它们都有一些相同的组成部分。相同的组成成分包括目标、参与方、有约束力的时期、适用范围、度量、服务等级目标(SLO)和惩罚。

目标是对 SLA 规范高等级的描述。参与方是签订 SLA 规范的参与者比如服务消费者、服务提供者或者第三方。有约束力的时期定义了 SLA 规范有效的时间间隔。度量是服务可以被测量的度量。度量分为两种:原子度量是直接测量得到的度量,比如当前响应时间、当前服务引用数;复合度量是通过原子度量计算得到的,比如吞吐率是在一定时间内服务的引用数。复合度量通常可以理解为原子度量的函数比如 maximum、average。只需要监测原子度量,而复合度量可以通过原子度量得到。服务等级目标(SLO)描述的是服务参与方对一些特定行为的承诺。它是一个基于定义的度量的逻辑表达式。比如一个 SLO 声明服务提供者保证服务每一天系统的可用性达到 99%。最后一个组成成分是对违反 SLO 的惩罚。这些成分中最重

要的是度量和 SLO。

文中提出了服务的质量模型,在这个质量模型中定义了服务质量的特征与相关的度量。通过测量这些度量可以判断服务是否达到了预先设定的目标。

### 2.2 质量模型与相关度量

参考 ISO9126 标准,结合 SLA 规范中的相应条款,设计了质量模型和相关度量。ISO9126 所包含的服务质量特征中维护性、移植性和可靠性这些特征是不可被监测的。因为这些特征是软件设计时候的特征,它们在软件运行的时候不发生改变。因此得出的结论是只有小部分属性是可以监测的,通常这些属性是与可用性、时间和正确性相关的属性。

确定好需要监测的属性,需要把它们和 SLA 规范中的度量对应起来。监测和管理的是网络服务,那么响应时间和可用性就是非常好的度量。所以构造如表 1 所示的质量模型。

表 1 质量模型

网络服务			
	度量	描述	分类
响应时间	当前响应时间	用户发送请求得到响应的时间/ms	原子
	最小响应时间	响应时间的最小值/ms	复合
	最大响应时间	响应时间的最大值/ms	复合
	平均响应时间	响应时间的平均值/ms	复合
可用性	当前可用性	当前一个网络服务的可用性	原子
	累积可用调用	用百分数的形式表示一个网络服务从第一次运行到现在有多少次是可用的	复合
	平均恢复时间	一个网络服务从不可用到可用所花费的平均时间/s	复合

当监测发现质量模型中的度量不能满足 SLA 规范中 SLO 对度量的规定时,就产 SLA 违例。针对产生的 SLA 违例进行修复或者产生告警信息。

## 3 监测系统架构

系统的架构可以按照系统内组件的功能关系分为监测器、QoS 计算器、SLA 分析器和修复方案四大核心功能组件<sup>[7]</sup>,如图 1 所示。

### 3.1 监测器

监测器是由一些测量工具组成的。测量工具只测量服务的原子度量。复合度量可以通过原子度量计算得到。这些组件是监测器完成测量工作的具体工作组件。监测器根据不同的服务规定的监测时间间隔,周期地调用测量工具完成对监测数据的更新。更新得到的数据通过监测器存入数据库,数据库预先设置触发器,当有更新操作完成时触发器触发,通知 QoS 计算器计算服务新的性能参数。

监测器是由 QoS 计算器管理的。每一个监测器对应一个特定的 SOA 系统。一个 QoS 计算器可以管理

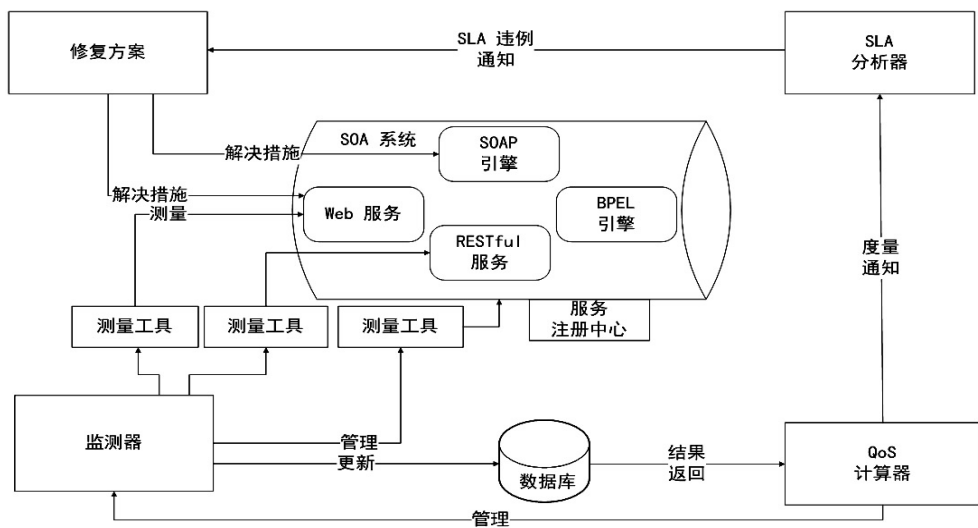


图 1 监测系统架构

一组监测器, QoS 计算器装载预先定义好的质量模型, 把需要监测的度量发送给监测器。监测器使用测量工具完成监测。

由于测量方式是主动监测, 采用的是请求-响应模式, 这种监测方式会干扰到被监测的服务, 所以每一个测量工具都必须把自己对被测量服务的干扰降低到最小。为了减小对被监测服务的干扰, 可以采用的方法是一个服务需要监测的原子度量使用同一个测量工具并且按照相同的监测间隔时间进行监测。比如想要测量一个服务的原子度量: 当前可用性和当前响应时间, 调用同一个测量工具在一次测量中把这两种信息获取到。

### 3.2 QoS 计算器

性能参数计算器是用来读取质量模型, 并且计算定义在质量模型中的性能参数。为了计算服务的性能参数, 这个部件包含了管理监测器、与数据库交互和向 SLA 分析器交互的接口。一个 QoS 计算器管理一组监测器。

监测器将得到的度量存入数据库。数据库中的触发器触发, 向 QoS 计算器提供更新后的数据。QoS 计算器根据数据库提供的更新的度量装载质量模型, 质量模型中包括了各个度量的描述。QoS 计算器根据质量模型计算这些度量并且把计算好的结果通知给 SLA 分析器。

### 3.3 SLA 分析器

SLA 分析器是用来探测低于定义在 SLA 规范中阈值的性能参数的值, 并且把违反 SLA 规范的情况通知系统管理者。SLA 规范在配置过程中被分解为一组判断条件的集合, 这些条件必须在 SLA 规范规定的特定时间范围内被同时满足。一个条件是由一个需要监测的度量、一个相关的操作符和一个进行比较的值组成的。比如当前响应时间 < 100 ms 是一个条件, 这个

条件在 SLA 规范规定的时间范围内必须为真。

对于性能参数值低于在 SLA 规范中规定的阈值, 定义为 SLA 违例。首先 SLA 分析器从 QoS 计算器中获得依据质量模型计算好的服务的度量。然后 SLA 分析器根据相应 SOA 系统中的 SLA 规范形成的条件逐个检验度量是否能满足要求。当发现不满足要求的情况就向系统管理者报告 SLA 违例。

### 3.4 修复方案

对于 SOA 系统中出现违反 SLA 规范的情况, 系统管理者选择一种最好的修复方案。因为在 SOA 系统中不仅是整个系统需要满足 SLA 规范中的规定, 系统中的每个服务也需要满足 SLA 规范中的规定。根据 SLA 违例采用相应的修复措施。

## 4 实验分析

针对构建的模型和系统, 设计了实验。实验的目的有两个。第一个是展示了如何构建质量模型; 第二个是在不同的实验环境中监测网络服务的性能参数, 绘制性能参数图。

在 SOA 系统中有两种网络服务: 网络服务 A、网络服务 B。这两种类型中各有三个功能相同的实例。

### 4.1 对单个 Web 服务监测实例演示

实验环境介绍: Windows 7 操作系统、AMD Athlon (tm) II X4 645 Processor 3.10 GHz、3 G 内存、Eclipse、Tomcat 6.0.18。实验先对  $a_1$ 、 $a_2$ 、 $a_3$ 、 $b_1$ 、 $b_2$ 、 $b_3$  进行性能测试<sup>[8-9]</sup>。测试方法为多线程多次调用。测试使用 1 000 个线程, 每个线程 20 次总共 20 000 次调用。实验结果为  $a_1$  的最小响应时间 1 ms、最大响应时间 28 182 ms、平均响应时间 2 081 ms, 如图 2 所示。

对于  $a_1$  的测试中尽管响应时间的范围跨度非常大, 但它们的值却相对集中。响应时间非常大的值

只占极少数。所以不能简单地使用平均值来评价服务的性能。

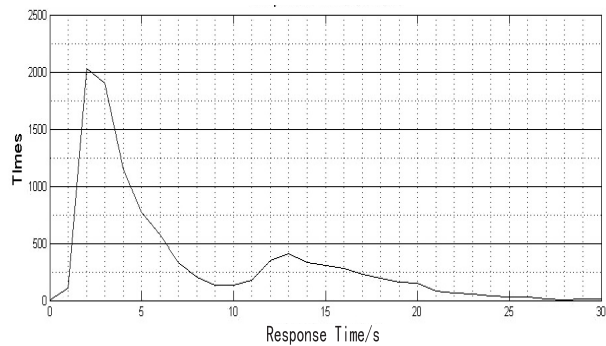


图 2 a<sub>1</sub> 响应时间分布图

响应时间在 30 ms 以内的线程总数有 9 592, 次几乎占 20 000 次服务调用总数的一半。80% 的响应时间在 9 503 ms 内。其中响应时间是 3 ms 的次数最多为 1 847 将近十分之一。根据这些实验数据制定 SLA 规范。把可用性定义为在 10 s 内服务响应。20 000 条线程完成测试的时间为 197 606 ms 约为 200 s。

- SLA 规范的内容为:
- (1) 有约束力的时期: 整个网络服务运行时期;
  - (2) 度量: 响应时间、可用性;
  - (3) SLO: 平均响应时间小于 3 s、可用性大于 80%。

用 XML 语言表示为<sup>[10-11]</sup>:

```
<? xml version="1.0" encoding="UTF-8"? >
<SLA name="sla.xml">
<Metric name="Responsetime" type="float" unit="millisecond"></Metric>
<Metric name="Availability" type="float" unit="1"></Metric>
<Period>
<strat>2013-01-04</strat>
<end>2013-03-09</end>
</Period>
<SLO name="slo">
<! -- Description of SLO -->
</SLO>
<Action>
<! -- violation Repairing -->
</Action>
</SLA>
```

4.2 基于 SLA 规范构建质量模型

根据表 1 中质量模型的内容,对于网络服务使用响应时间和可用性作为质量模型的度量<sup>[12-13]</sup>。由实验得到的 SLA 规范和建造的质量模型生成 SLA 分析器中判断条件。一个条件是由一个需要监测的度量、一个相关的操作符和一个进行比较的值组成的。对于性能参数值低于在 SLA 规范中规定的阈值,定义为

SLA 违例。

生成 SLO 如下:平均响应时间<3 000 ms、累积可用调用>80%。

用 XML 语言表示为:

```
<? xml version="1.0" encoding="UTF-8"? >
<SLO name="slo">
<Metric name="Responsetime" type="float" unit="millisecond"></Metric>
<!-- Average Response Time -->
<Metric name="AverageResponseTime" type="float" unit="millisecond">
<Function type="Divide" Output="ART">
<Operand>
<Metric name="ResponseTime" type="float" unit="millisecond"></Metric>
<Metric name="count" type="int" unit="1"></Metric>
</Operand>
</Function>
<Expression type="less">
<Threshold unit="millisecond">3 000</Threshold>
</Expression>
</Metric>
<Metric name="Availability" type="float" unit="1">
<Expression type="more">
<Threshold unit="1">0.8</Threshold>
</Expression>
</Metric>
</SLO>
```

4.3 对组合服务的监测实例演示

实验环境和单个网络服务的实验环境一样,组合服务 C 中包含服务 A、服务 B 与服务 N。监测子服务 A 与组合服务 C 的响应时间,实验结果如图 3 所示。

上方曲线显示的是 C 的响应时间,下方曲线是子服务 A 的响应时间。可以发现子服务的性能影响组合服务的性能,组合服务的性能与全部子服务的性能有关。所以对组合服务进行性能管理需要对其中的所有子服务进行管理。

4.4 本地和异地访问服务监测

通过本地和异地测试可以计算出服务在服务器的运行时间和在网络中的传输时间<sup>[14]</sup>。当在客户端调用服务,这个时候的响应时间是服务请求在网络中的传输时间加上服务在服务器上的处理时间。

实验使用 3 个 IP 地址的主机进行测试,一个是本机 localhost,一个是地址为 192. 169. 0. 6 的主机,一个是地址为 192. 169. 0. 43 的主机。当服务器的负载没有超过上限的时候,网络服务 A 在服务器端的执行时间小于 2 ms,而且基本在 1 ms 左右。用响应时间减去在服务器上的执行时间就是在网络上的传输时间。



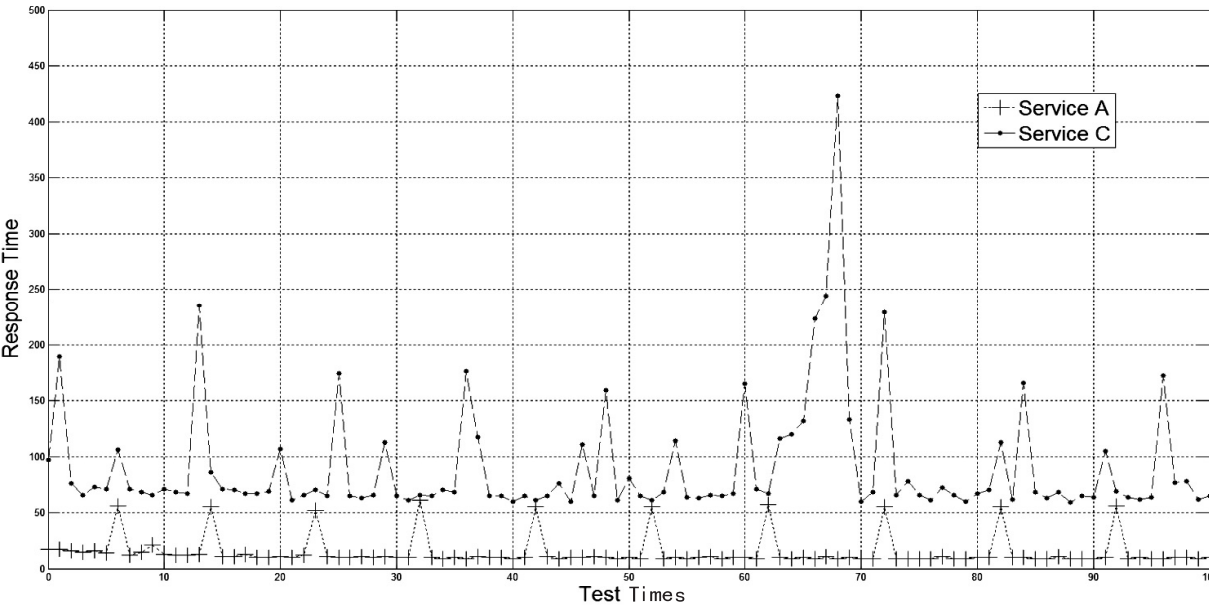


图 3 组合服务响应时间分布图

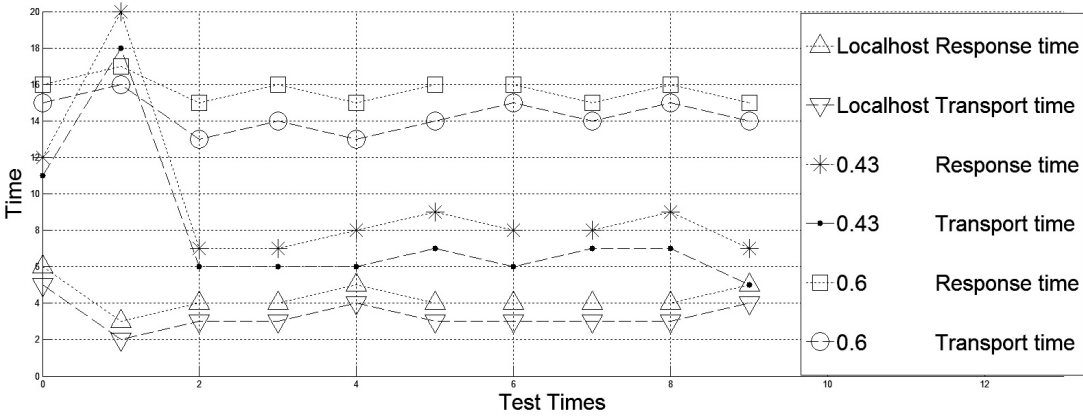


图 4 响应时间和网络传输时间分布图

图 4 为响应时间和网络传输时间分布图。

通过监测服务实例在服务器上的运行时间和网络中的传输时间可以对服务的性能进行调整。如果是网络传输时间过长可以采取增加网络带宽等措施,如果是服务器上运行时间过长可以选择一台负载低的并且部署由相同服务实例的服务器。这些性能参数的采集为性能管理提供了依据。

5 结束语

文中讲述了根据 SLA 规范建立质量模型,在 SOA 系统的环境中监测 SLA 规范中声明的 SLO。提出了基于 Web Service 的性能特性,按照 SLA 规范生成服务的质量模型方法。

系统通过监测组合服务、本地异地服务的性能为性能瓶颈分析提供实验数据基础。这个服务性能管理系统的架构由于其高度的技术无关性,能够在其他类型服务上使用。系统为各种服务提供接口,能够采用现有的监测工具作为架构的一部分。

参考文献:

[1] Zhou C, Chia L T, Lee B S. QoS measurement issues with DAML-QoS ontology [C]//Proc of IEEE international conference on e-business engineering. [s. l.]:[s. n. ],2005.

[2] Moser O, Rosenberg F, Dustdar S. Non-intrusive monitoring and service adaptation for WS-BPEL [C]//Proceeding of the 17th international conference on World Wide Web (WWW 2008). [s. l.]:[s. n. ],2008.

[3] Benharref A, Dssouli R, Serhani M A, et al. Efficient traces collection mechanisms for passive testing of Web Services [J]. Information and software technology, 2009, 51: 362 - 374.

[4] Comuzzi M, Kotsokalis C, Spanoudakis G, et al. Establishing and monitoring SLAs in complex service based systems [C]//Proc of IEEE 7th international conference on Web services. [s. l.]:[s. n. ],2009.

[5] Bertolino A, Inverardi P, Pelliccione P, et al. Automatic synthesis of behavior protocols for composable web - services

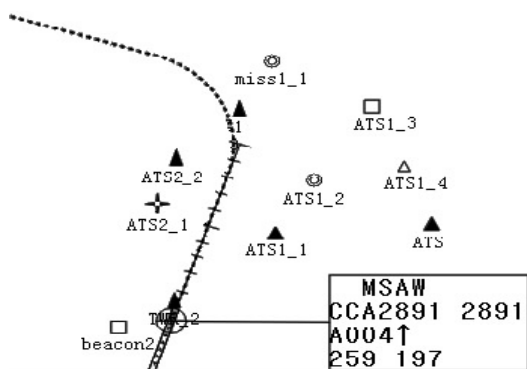


图 4 降落时切跑道

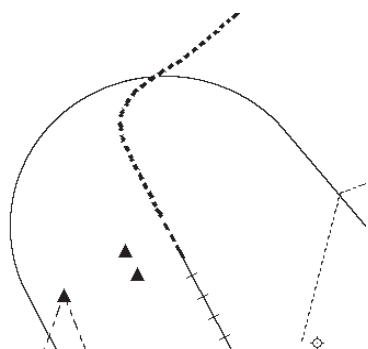


图 5 真实飞机降落过程

仿真,通过实时校正当前航向与期望航向的偏差,实现了准确的沿期望路径飞行的目的。经过实验,该模型能够高效准确地控制飞机横向运动,经过与真实飞机的航迹数据对比,比传统模型更接近真实飞行。下一步的工作将继续应用自动控制原理模拟飞机的纵向运动,提高飞机高度、速度方向的仿真度。

#### 参考文献:

- [1] 罗 瑜. 空管模拟培训逼真度问题的研究与实现[D]. 成都:四川大学,2005.

(上接第 16 页)

- [C]//Proc of the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on the foundations of software engineering. [s. l.]: ACM, 2009: 141-150.
- [6] Menasce D, Ruan H, Goma H. QoS management in service-oriented architectures[J]. Performance evaluation, 2007, 64(7): 646-663.
- [7] Ameller D, Franch X. Service level agreement monitor(SA-LMon)[C]//Proc of seventh international conference on composition-based software systems. [s. l.]: IEEE, 2008: 224-227.
- [8] 程 旭. 基于 SLA 的 SaaS 应用性能管理机制研究[D]. 济南:山东大学,2010.

- [2] 张 健. 空管自动化系统中关键算法分析[J]. 阴山学刊(自然科学版), 2012, 26(2): 37-40.
- [3] 郭运韬, 朱衍波, 黄智刚. 民用飞机航迹预测关键技术研究[J]. 中国民航大学学报, 2007, 25(1): 20-24.
- [4] 董天罡, 李新胜, 蔡 葵. 场面监视雷达仿真系统的地面运动数学模型[J]. 计算机应用, 2008, 28(z2): 378-379.
- [5] 李新胜, 兰时勇, 李 纲, 等. 利用 BeZier 和 B 样条曲线模拟飞机航迹的方法[J]. 西南交通大学学报, 2011, 46(6): 1040-1045.
- [6] 郭 勇, 张瑞峰, 吴 伟. 基于 Cardinals 三次样条曲线算法的飞行过程再现[J]. 计算机工程, 2005, 31(8): 173-174.
- [7] 杨友良. 自动控制原理[M]. 北京:电子工业出版社, 2011.
- [8] 刘 豹, 唐万生. 现代控制理论[M]. 第 3 版. 北京:机械工业出版社, 2011.
- [9] 陈楷民. 新一代塔台视景模拟机仿真核心的研究与实现[D]. 成都:四川大学, 2006.
- [10] Nikolos I K, Valaanis K P, Tsourveloudis N C, et al. Evolutionary algorithm based offline/online path planner for UAV navigation[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2003, 33(6): 898-912.
- [11] 李 楠, 刘 刚, 王 林, 等. 机场塔台管制仿真系统核心计算模块的建模及实现[J]. 交通运输工程与信息学报, 2010, 8(1): 39-43.
- [12] Feigh K M. An airspace simulator for air traffic management research[D]. Granfield: Granfield University, 2003.
- [13] Eurocontrol Experimental Centre. User manual for the base of aircraft data (BADA)[R]. Brussels: Eurocontrol, 2012.
- [14] Peters M, Konyak M A. The engineering analysis and design of the aircraft dynamics model for the FAA target generation facility[R]. Atlantic City: Federal Aviation Administration William J. Hughes Technical Center, 2012.

- [9] 张若英, 邱雪松, 孟洛明. SLA 的表示方法和应用[J]. 北京邮电大学学报, 2003, 26(z2): 12-17.
- [10] Martin D. XML 高级编程[M]. 北京:机械工业出版社, 2001.
- [11] Seely S, 杨 涛, 杨晓云, 等. SOAP: XML 跨平台 Web Service 开发技术[M]. 北京:机械工业出版社, 2002.
- [12] Gaur H, Zirn M. BPEL cookbook[M]. Birmingham: Packt Publishing Ltd, 2006.
- [13] Muthusamy V, Jacoben H A, Chau T, et al. SLA-driven business process management in SOA[C]//Proc of the 2007 conference of the center for advanced studies on collaborative research. Riverton: [s. n.], 2007.
- [14] 张 业. 网络性能数据采集与分析方法研究[D]. 西安:西安电子科技大学, 2010.

基于SLA规范的服务性能管理

作者：[朱浩洋](#)，[崔鹏帅](#)，[曹介南](#)，[ZHU Hao-yang](#)，[CUI Peng-shuai](#)，[CAO Jie-nan](#)

作者单位：[国防科学技术大学 计算机学院, 湖南 长沙, 410073](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

ISTIC

年，卷(期)：2014(4)

本文链接：[http://d.g.wanfangdata.com.cn/Periodical\\_wjfz201404003.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjfz201404003.aspx)