

基于 Web 的实时信息系统研究与应用

蒋大卫¹, 周 俊², 孔令和¹, 伍民友¹

(1 上海交通大学 计算机科学与工程系, 上海 200240;

2 耶鲁大学 计算机科学与工程系, 纽黑文 06520)

摘 要: 首先对实时信息以及实时信息系统给出定义. 在此基础上, 进一步提出实时信息价值评估方式以及实时信息系统的性能评估方式. 然后结合两个个现有的实时信息系统, 实时路况系统以及微博客系统, 探讨实时信息系统在设计以及实现过程中面对的挑战, 并且针对这些挑战, 对现有的实时信息系统设计提出改进思想. 根据这些改进思想, 设计了 INFOCOM LIVE 实时会务系统. 这个系统已经在国际顶级网络通信学术会议 IEEE INFOCOM 2011 的会场进行了实地运营.

关键词: 实时信息; 实时信息系统; 实时路况系统; 微博; INFOCOM LIVE

中图分类号: TP393.02

文献标识码: A

文章编号: 1000-7180(2012)06-0153-05

Real-Time Information System Theory and Its Application Based on Web

JIANG Da-wei¹, ZHOU Jun², KONG Ling-he¹, WU Min-you¹

(1 Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China;

2 Department of Computer Science and Engineering, Yale University, New Haven 06520, America)

Abstract: This paper starts from the concept of real-time information and the concept of real-time information system. Subsequently, we discuss how to evaluate the value of real-time information and the performance of real-time information system. Then, given two real-time information systems used in practice example; real-time traffic system and micro-blog system, we discuss the challenges in the design and the deployment of the real-time information system. Based on these discussions, we propose our methods to improve the performance of real-time information system. Then we implement these methods in the design of INFOCOM LIVE. This system has been used in the conference venue of IEEE INFOCOM 2011 (The 30th IEEE International Conference on Computer Communications), which is the top-level academic conference of communication and network research in the world.

Key words: real-time information; real-time information system; real-time traffic system; micro-blog; INFOCOM LIVE

1 引言

在互联网发展的过程中, 人们获得信息的速度越来越快, 信息对于某一些信息而言, 信息的实时性越高, 信息的价值也会增加, 不同种类的实时信息, 对信息的实时性的要求也不相同. 例如实时路况信息, 需要信息的实时性以分钟为单位^[1]. 而商场限时

优惠信息也属于实时信息, 其实时性则以小时, 甚至天为单位. 用来传递实时信息的系统就称为实时信息系统^[2].

通过传递高实时性的实时信息, 实时信息系统为人们的生活带来了巨大的便利. 实时信息系统对系统效率有着很高的要求, 这对实时信息系统的设计和实现提出了挑战. 实时信息的信息量巨大, 这也导致了系统需要对信息进行筛选, 对信息的准确性

进行分析.

文献[1-2],对实时信息以及实时信息系统的理论模型进行了研究.这些研究为实时信息系统的设计提供了理论依据,也为如何解决现在实时信息系统中的问题,提供了思路.在这个基础上,本文对实时信息的价值模型进行了研究,并以此为依据,找出实时信息系统设计以及实施中的关键点.

2 实时信息理论研究

能够在信息时效期望内从信源达到信宿的信息,称之为实时信息.实时信息是指必要在限定时间范围内完成采集、传输和处理的信息.

在真实的实时信息系统中,系统的信息采集能力,信息传输能力、信息处理能力、以及接受信息的信宿的信息处理能力都是有限的.因此对每一条实时信息的价值进行评估具有重要的意义.在实时信息系统中,依据信息的价值,合理地安排实时信息传输处理的优先级,将没有价值的信息从实时信息系统中剔除,将会很好地提高实时信息系统的工作效率.

不同场景下的实时信息,具有不同的时效期望,也就是信息的寿命.需要针对不同的需求,计算它的时效期望值,这牵涉到信源、信宿、传播媒介、信息量等因素.

我们可以定义实时信息的价值模型是一个关于时间的函数.假设实时信息的价值为 v ,在某一时刻 t ,实时信息的价值为 $v=f(t)$.对于不同种类的实时信息,在某一时刻 t ,实时信息一定有一个确定的价值,以此为依据,实时信息系统可以通过提高传输的信息的总价值,来提升系统的性能.

实时信息系统由四个子系统组成:信息采集系统、信息处理系统、信息传输系统、信息发布系统.信息采集可以分为自动采集、人工采集、混合采集.信息处理系统负责对接收到的信息进行加工、分析和重构,使其成为真正可发布的实时信息.信息处理的方式有集中处理、分布式处理、云计算等.

信息传输系统负责在采集端、处理端和发布端传递实时数据.

信息发布系统负责将信息用各种各样的信息发布给信宿,这些发布形式包括 Web 页面,计算机程序,手机程序等等.

对于一个实时信息系统,在一段确定的时间内,假定信源所产生的信息的价值总和为 V_a .实时信息系统从信源采集到实时信息,并且在信息的时效期

望内将信息传递给信宿,最后到达信宿并被信宿接收的实时信息的价值总和为 V_c ,若 V_c/V_a 越高,则实时信息系统的性能越高.对于一个实时信息系统,在一段确定的时间内,信源的信息总量 $I=\{i_1, i_2, i_3, \dots, i_n\}$,相对应的信息的价值为 $V=\{v_1, v_2, v_3, \dots, v_n\}$,信息的时效期望为 $T=\{t_1, t_2, t_3, \dots, t_n\}$.采集信息需要的时间为 $TC=\{tc_1, tc_2, tc_3, \dots, tc_n\}$.传输信息需要的时间为 $TT=\{tt_1, tt_2, tt_3, \dots, tt_n\}$.

依据信息的价值以及时效期望,可以将实时信息系统分为三类:单态实时信息系统多态实时信息系统,混合实时信息系统.

在单态实时信息系统中,每一条信息具有相同的价值模型以及时效期望.这一类实时信息系统的代表是实时路况系统,实时股票系统.

在多态实时信息系统中,每一条信息具有不同的价值模型以及时效期望.这一类实时信息系统的代表是微博客系统.在微博客系统中,依据实时信息内容以及信源的不同,实时信息的价值模型和时效期望都不相同.

在混合实时信息系统中,一部分的信息具有相同的价值模型以及时效期望,另一部分的信息具有不同的价值模型以及时效期望.这一类实时信息系统的代表是酒店管理系统等.

针对以上三类实时信息系统,提出影响实时信息系统性能的几个要点.

(1) 对于每一条信息,必须满足系统采集信息,传输信息的时间之和小于信息的时效期望,信息才具有价值.即 $tc_i + tt_i \leq t_i (0 \leq i < n)$.这就要求实时信息系统设计的过程中,尽可能地降低信息采集的时间,尤其是在人工采集的过程中,信息采集时间的优化尤其重要.在信息的处理,传输过程中,也必须通过优化信息的数据结构,实时信息系统的体系结构的方式来减少信息的传递时间.在信息的发布过程中,必须选择实时性强的发布平台来将信息发布给用户.

(2) 在多态实时信息系统中,针对不同的实时信息,应该依据信息的价值高低,来安排信息在采集,处理,传输,发布过程中的优先级.

(3) 信宿的信息接收容量是有限的.在大多数的实时信息系统中,信息系统的信宿就是系统的人.用户不可能接收所有的信息.在用户的信息接收容量中,实时信息系统必须对传输的信息进行处理和筛选,从而使用户接收到的信息价值总和最大化.

(4) 实时信息系统必须实时去除因为超过时效期望而失去价值的信息,做到实时更新。

(5) 实时信息系统必须能够实时的排除错误采集的信息。

3 实时信息系统分析

3.1 实时路况系统

实时路况系统已经广泛的应用于我们的生活之中,在文献[3]中,提出了一个由 GPS 信息采集系统,数据分析系统,GIS 信息发布系统组成的实时路况信息系统模型。

实时路况系统是一个庞大的系统,在实际的应用中,信息采集系统,信息处理系统,信息发布系统甚至会由不同的运营商提供,例如 Google 的地图以及路况信息由 Mapabc 提供,而 Mapabc 的路况数据则由九州联宇公司^[4]提供。

实时路况系统的性能瓶颈存在于信息采集系统。实时路况信息的采集方式包括静态的采集方式以及动态的采集方式。静态的采集方式是指基于固定在路上或者路旁的监测设备来获得该点道路通行状况。动态的采集方式是指利用安装在车辆上的 GPS 信息设备来获得道路通行情况。其中动态的采集方式更具有实用性,在国外学术届已经提出了一系列的概念和部署方案,并且已经部署了几个大型的系 统,例如美国 MIT CarTel 项目^[5]、旧金山 Cabspotting 项目^[6]。在国内的实际的运用中,九州联宇公司^[4]通过一定数量带有信息回传功能的 GPS 车载系统的协议车辆(出租车、集团车辆等),来进行实时路况数据的采集。

利用车辆来采集实时路况信息,对信息采集系统的设计提出了一下几个挑战:

(1) 信息采集系统无法保证车辆回传数据的实时性,在实际的数据通信中,车辆必须通过无线网络与信息采集系统通信,传输的稳定性较差。

(2) 车辆在道路上的分布对信息采集系统的信息采集也有很大的影响,例如在某一个区域内,如果没有一辆可以回传 GPS 数据的车辆,那么信息采集系统就无法实时采集这个区域的路况数据^[3]。

3.2 微博客系统

微博客系统已经广泛的应用于我们的生活之中,比较著名的微博客系统有 twitter,新浪微博,腾讯微博等等。微博客系统的用户可以随时随地的发布信息,这些信息将会被实时发布给与此用户建立了关注关系的其他用户。

微博客系统的特点是每一条实时信息的价值模型以及时效期望都不相同。在微博客系统上面,每一条信息的内容决定了信息的时效期望,而每一条信息的发布者和内容,决定了信息的价值模型。

在信息发布系统中,微博客的性能瓶颈体现在两个方面:

(1) 微博客系统信息量大,用户接受信息的能力有限。如何为用户挑选有价值的信息,是实时信息系统的一个职责所在。

(2) 微博客系统中含有错误信息,影响巨大。由于微博客系统作为实时信息系统,具有信息量大,传播速度快的特点,导致微博客系统难以像传统的信息系统一样对信息建立复杂的审核机制。是否具备简单快速的审核机制来检测并且排除错误的信息,是评价微博客系统性能的一个重要指标。

4 INFOCOM LIVE 实时会务系统

本次 IEEE INFOCOM 2011 在上海召开。我们策划、设计与实施了 INFOCOM LIVE,这是一套服务于 IEEE INFOCOM 2011 会议现场的实时信息采集与展示的系统,其功能为采集 IEEE INFOCOM 2011 会场的各类实时信息,并通过会场核心区域放置的液晶大屏幕,可供国际 Internet 访问的实时网页进行显示。

4.1 INFOCOM LIVE 实时会务系统中的实时信息

INFOCOM LIVE 实时会务系统主要包括三个部分的实时信息:会场信息实时显示,官方会议信息发布,会议相关的自由信息发布。其中会场信息子系统是一个单态实时信息系统,官方会议信息子系统和自由信息子系统是多态实时信息系统。

(1) 会场信息

在 INFOCOM 2011 的三天正式会期内,共分为 80 个 Session。整个会场共有 8 个演讲厅,同时至多有 8 个 Session 并行进行。每一个 Session 长达 90 分钟,期间通常会安排 3~4 篇 Paper 的 Presentation。由于每篇 Paper 的展示时间是灵活的,所以在会场外的人难以把握 Session 的进展情况。另外,由于官方事先给出的 Schedule 可能会由于各种因素随时有变动。因此想要获得会场内的实时状况,就需要在会场内进行实时信息的采集,随后实时发布。

每一条会场信息都具有标准的格式,包括以下内容:演讲厅序号,Session 序号,论文序号,演讲开始时间,演讲者名字,会场人数。

(2) 官方会议信息

INFOCOM LIVE 实时会务系统同时也是一个会议主办方发布会议信息的平台. 在实际的运营过程中, 会议主办方利用 INFOCOM LIVE 发布了 INFOCOM 2011 开幕式相关信息, NSF Talk 相关信息, TCCC Meeting 相关信息等信息. 官方会议信息只能通过会议主办方来进行发布, 具有很高的价值和可信度. 我们为主办方提供了专用的发布接口, 不需要进行信息审核.

(3) 自由信息

INFOCOM LIVE 实时会务系统允许与会者发布自己的信息. 在实际的运营过程中, 自由信息系统为与会者提供了便利,

4.2 INFOCOM LIVE 实时会务系统的设计

(1) 系统架构

我们开发了一套基于 Web 和 WIFI 网络的数据采集、处理与发布的平台. 实时采集的信息通过带有 WIFI 功能的手机上传至中央服务器. 同时, 客户端持续接收服务器端所发送的最新信息. 系统维护人员可以通过后台接口对服务器端进行实时调整. 系统架构如图 1 所示.

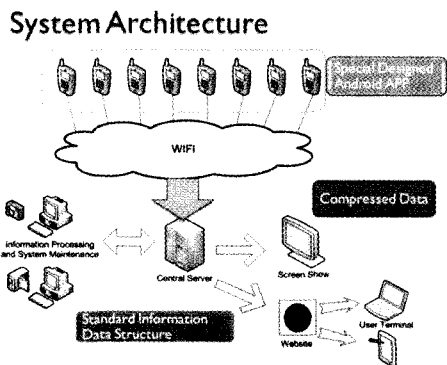


图 1 INFOCOM LIVE 系统架构

(2) 相关技术

系统开发部分包括作为信息采集终端的智能手机平台开发 (Android 2.2 智能手机, 型号为天星 A5000), 服务器端开发 (AJAX, MySQL, PHP), 信息发布端开发 (AJAX, Javascript, HTML, CSS). 参见图 2.

(3) 信息采集

在信息采集的过程中, 官方信息以及自由信息都是由官方和与会者发布的. INFOCOM LIVE 需要主动采集的信息就是会场信息. 而面对的主要问题是信息采集的实时性和准确性.

在设计过程中, 我们采用了智能手机作为信息



图 2 天星 A5000 智能手机

采集设备, 在系统的实际运行过程中, 为每一个会议室都配备了一台智能手机以及一个志愿者来进行信息采集.

并为志愿者设计了专门的智能手机软件. 由于会场信息的格式是固定的. 在软件中预先输入 Session 信息, 论文信息, 演讲者信息, 然后让志愿者只需要选择数个选项就可以快速的完成信息上传工作. 在软件中, 我们还对志愿者上传的信息进行了压缩, 同时在软件中对志愿者上传的信息进行了校验, 防止志愿者上传错误的信息. 信息采集软件界面如图 3 所示.

图 3 信息采集软件界面

(4) 信息处理

在信息处理的过程中, 主要面对两个挑战. 一是如何检验会场信息以及自由信息的准确性. 二是如何评估自由信息的价值高低, 从而在信息发布的时候, 将价值高的自由信息发布给用户. 对于会场信息的准确性, 在系统中预先记录了会议日程表, 志愿者上传会场信息以后我们会将志愿者上传的信息与会议日程表进行自动比对, 如果发现区别, 系统会要求志愿者对上传的信息进行确认. 从而判别是志愿者上传信息错误, 还是会议日程表被修改了. 对于自由信息的准确性以及价值评估, 我们则为系统设计专门的后台接口, 并配备一个专门的信息审核员对信息进行审核, 决定信息发布的优先级.

(5) 信息发布

采用了两种集中化的方式来实时发布信息:会场核心区域放置的液晶大屏幕,可供国际 Internet 访问的实时网页。

大屏幕以及实时网页的展示内容,包括会场的实时基本信息、官方授权发布的通知与消息,经过审核的关于会议的自由讨论信息。实时信息页面是基于 AJAX 网页技术而开发的动态 Web 页面。在展示页面设计上,我们将页面分为三个部分。在最显眼的左上角安排了最重要的官方会议信息。而在左下角安排了进过审核和编辑的自由信息。在右半边则安排了会场信息,在这个模块中将页面分成了 8 个部分,每个部分实时更新对应会议室中的会议进程。

5 结束语

本文对实时信息理论以及实时信息系统进行了研究。在此基础上,结合现有的实时路况系统,微博客系统,对实时信息系统的性能进行分析评估,找到了影响实时信息系统性能的关键要素。提出了改进实时信息系统性能的设计方案。

(上接第 152 页)

6 结束语

本验证方法充分利用了断言的优点,将其应用于 WISHBONE 到 AHB 的转换桥的验证中。首先抽象出总线的时序属性,然后通过 SVA 对其描述定义,同时插入覆盖率检查机制,对验证覆盖率进行分析。在总线验证中,断言以简洁的表达方式,获得较高的覆盖率,达到良好的验证效果。

参考文献

- [1] 董杨鑫,郑建宏. 基于断言的 SoC 设计验证方法[J]. 理论与方法, 2007,26(10):14-16
- [2] 查欣. 基于断言语言 SVA 的设计验证方法[J]. 现代电子技术, 2010(10):5-7
- [3] 李洋洋,吴武臣,王龙伟,等. 基于断言的验证方法在 UART 模块中的应用研究[J]. 微电子学与计算机, 2010,27(1):151-153
- [4] Cao Fan, Chen Lan, Yi Bo. Designing WISHBONE to AMBA Wrapper[C] // International Conference on ASIC. [S. L.]: IEEE, 2009:403-406

参考文献

- [1] 吕炳潮,伍民友. 实时信息系统——物联网的基础[J]. 中国计算机学会通讯, 2010(4):38-46.
- [2] 吕炳潮,杨扬,伍民友. 实时信息的理论研究及其应用[J]. 计算机工程与设计, 2010(31):4012-4014.
- [3] 葛晓峰,曹斌. GPS 与 GIS 结合的高速公路实时路况分析系统设计及实现[J]. 电子技术, 2007(7):155-157.
- [4] 九州联宇公司. [EBIOL]. [2011-11-24]. <http://www.tranbbs.com/topic/realtraffic/index.shtml>
- [5] Cabspotting [EBIOL]. [2011-11-24]. <http://cabspotting.org/>.
- [6] Ryan Kelly. Twitter Study Reveals Interesting Results About Usage—40% is “Pointless Babble” [EBIOL]. [2011-11-24]. <http://www.pearanalytics.com/blog/2009/twitter-study-reveals-interesting-results-40-percent-pointless-babble/>

作者简介:

蒋大卫 男,(1988-),硕士.研究方向为实时信息系统。

孔令和 男,(1983-),博士.研究方向为物联网、传感器网络。

伍民友 男,(1954-),教授,博士生导师.研究方向为无线与传感器网络,高性能计算。

[5] Janick Bergeron, Eduard Cerny, Alan Hunter, et al. SystemVerilog 验证方法学[M]. 北京:航空航天大学出版社, 2007

[6] ARM. AMBA specification. rev. 2.0[R] 1999:14-22

[7] Srikanth Vijayaraghavan, Meyyappan Ramanathan. A practical guide for systemverilog assertions[M]. Berlin: Springer, 2006.

作者简介:

张挺 男,(1987-),硕士研究生.研究方向为 SoC 设计与验证。



陈岚 女,(1968-),博士生导师.研究方向为超深亚微米芯片设计方法学及计算机系统机构。

冯燕 女,(1980-),助理研究员.研究方向为 SoC 设计与验证。