

资源受限网络中高效 XML 交换的性能评估

谢晓燕¹, 王浩¹, 陈彦萍^{1,2}

(1. 西安邮电大学 计算机学院, 陕西 西安 710121;

2. 西安交通大学 计算机系, 陕西 西安 710049)

摘要:为了研究高效 XML 交换 (EXI) 在资源受限网络中的使用情况, 文中基于传感器观测服务的真实数据, 对四种交换模式下的数据大小进行了对比: XML、高效 XML、Gzip 压缩的 XML 以及开启压缩模式的高效 XML。通过在 ns-2 中建模, 对资源受限网络环境下各交换格式的平均成功率、平均延时、平均能量消耗等性能进行了分析。实验结果表明, 无论是否开启压缩模式, 高效 XML 比 XML 具有更小的体积、更高的成功率、更短的延时和更少的能量消耗。高效 XML 交换更加适合于资源受限网络。

关键词:资源受限网络; XML 交换; 性能分析

中图分类号: TP312

文献标识码: A

文章编号: 1673-629X(2013)06-0054-05

doi: 10.3969/j.issn.1673-629X.2013.06.014

Performance Evaluation of Efficient XML Interchange in Resource-restricted Network

XIE Xiao-yan¹, WANG Hao¹, CHEN Yan-ping^{1,2}

(1. College of Computer, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;

2. Department of Computer, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In order to research the use of the Efficient XML Interchange (EXI) in a resource-restricted network, compare the size of four types interchanging data used in a factual sensor observation service, XML, efficient XML, Gzipped XML and efficient XML with compression. And then comparatively analyze the main performance in resource-restricted network in the average success rate, the average latency and the average energy consumption, by using those interchanging data with ns-2 simulator. The results show that EXI has smaller volume, higher successful rate, shorter latency and less energy consumption than XML. EXI with compression also has these merits compared with Gzipped XML. Thus, the EXI is more suitable for resource-restricted network.

Key words: resource-restricted network; Efficient XML Interchange; performance analysis

0 引言

随着物联网技术的出现, 传感器、RFID 等小型设备的无线互联和资源共享需求越来越迫切。Web 服务为异构资源的协同提供了有效手段, 但应用于这些资源受限设备的无线互联网络时面临诸多问题^[1]。特别是以 XML 为基础的数据交换存在大量冗余信息, 增加了设备的资源消耗, 降低了网络传输的效率。所以提高 XML 交换效率对于在资源受限网络中使用 Web 服务具有重要的作用。

针对 XML 交换格式效率的优化方法主要是对其进行编码和压缩。根据是否支持压缩后的 XML 文档查询, 可以将 XML 压缩技术分为两类^[2]: 不可查询的 XML 压缩技术 (如 Gzip^[3] 和 XMill^[4] 等); 可查询的 XML 压缩技术 (XGrind^[5] 和 Xpress^[6] 等)。除了这些对原始 XML 直接处理的方法, EXI (Efficient XML Interchange)^[7] 是一种新的 XML 表述形式, 它致力于同时优化数据交换的性能和计算资源的利用效率; EXI 是由 W3C (The World Wide Web Consortium) 提出的一种 XML 信息交换格式, 它采用语法驱动的方

收稿日期: 2012-09-21

修回日期: 2012-12-27

网络出版时间: 2013-03-05

基金项目: 国家自然科学基金资助项目 (60803162); 陕西省自然科学基金基础研究计划资助项目 (2012JQ8029); 陕西省教育科学研究计划资助项目 (12JK0938)

作者简介: 谢晓燕 (1972-), 女, 副教授, 硕士生导师, CCF 会员, 主要研究方向为网络及服务计算; 王浩 (1988-), 男, 硕士研究生, 主要研究方向为 Web 服务技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130305.0819.054.html>

式通过简单的编码算法和较小的数据类型集来达到高效的编码,能够有效地减少信息的冗余度,同时可以和现有的 XML 技术保持一定的兼容性;EXI 在显著减少网络带宽需求的同时仍然能够高效地使用电力能源、计算资源以及内存等,较适合部署在资源受限的设备上。

1 EXI 技术简介

EXI 将 XML 文档的内容转化为 EXI 流 (EXI Stream), EXI 流包括两个部分: EXI 头 (EXI Header) 和 EXI 体 (EXI Body)^[7]。EXI 头中主要包括了一些 EXI 编码的版本信息和相关操作,用来编解码 EXI 流时采取相关级别的编解码操作。EXI 体是 EXI 中用来承载数据信息的部分,主要是由一系列的 EXI 事件 (EXI Events) 和数据混合而成的。EXI 有一套内置的事件表述类型,如文档开始使用 SD 表示,文档结束使用 ED 表示等。

EXI 的事件包括两个部分: 事件的结构编码 (Structure Coding) 和事件的内容编码 (Content Coding), EXI 的事件对应于 XML 文档的结构。

EXI 中包含一套内建的语法 (Build-In Grammar), 可以直接用来对 XML 文档进行编码处理,同时它还包含模式信息约束的语法 (Schema-informed Grammar), 可以根据给定的 Schema 模式文件来编码 XML 文档。采用该模式编码的文件体积会更小,但是由于需要预先解析 Schema 文件,所以时间消耗相对较多。

2 相关研究

目前公开发表过的相关文献有许多,文献[2]使用了大量的 XML 数据集评估了 9 种 XML 压缩工具的性能。文献[9]评估了 EXI 和 XML 以及 Gzip 压缩的 XML 这三种形式的数据文件的压缩效率和处理效率,该文献研究表明 EXI 的编码和解码速度快于 XML。然而,在这些工作中都没有考虑网络环境的影响,特别是在资源受限网络这样一种特殊的环境中,能量的消耗是一个非常重要的因素。文献[10]虽然考虑了网络环境的影响,也使用 NS2 评估了资源受限环境下的网络性能,但是其仿真时的主要场景是建立在 Web 服务发现的过程中的,涉及的数据量比较小,不能有效地反映出真实的网络使用情况。

和前面的工作不同,文中比较了四种 XML 交换形式的数据文件大小: 原始 XML、EXI、使用 Gzip 压缩的 XML 以及开启压缩模式的 EXI^[11]。并且在 NS2 建模仿真了这些交换格式,比较了它们在资源受限网络环境下的网络性能。

3 性能评估

3.1 XML 数据集

文中使用由 52°North 开发的 SOS 系统原始数据, SOS 是 SWE (Sensor Web Enablement)^[12] 的一个标准构件,主要负责返回用户请求的历史或实时观测数据,它建立在 SWE 的标准信息模型之上,采用 O&M 来编码观测数据,屏蔽了传感器观测数据的差异性;使用 SensorML 描述传感器信息,屏蔽了传感器设备的差异性^[13]。标准的 SOS 必须实现三个核心接口: GetCapabilities、DescribeSensor 以及 GetObservation。在众多的查询操作中,文中选择了典型的 8 种,包括了这三种核心操作还有一些增强操作,每一种查询请求对应于一种 XML 文档,使用数字 1 到 8 来编号。

编号为 1 的 XML 文件用来请求传感器的描述信息,编号为 2 的文件用于检索一个服务的元数据,编号为 3 的文件是请求传感器的 feature of interest^[7] 信息。其他文件都是用来根据不同要求查询特定服务的观测数据。这 8 种查询操作的 XML 文件大小参考图 1 中的 XML 部分所示。由于 SOS 系统没有插入过多的数据,所以返回的数据量并不是很大,这样做是为了比较真实的模拟传感器之间的数据交换。

3.2 传输文件大小比较

文中比较了四种不同传输类型的文档大小: 原始 XML、EXI、使用 Gzip 压缩的 XML 以及开启压缩模式的 EXI。其中 EXI 是使用标准 EXI 模式构造的文档,其包含的信息和原始 XML 文档相同。EXI 可以使用一种压缩的模式来编码^[12],在这种模式下 XML 文件中的事件和数值能够按照相似的类型来重新排列,从而增加了数据块之间的相似度,再使用标准的 Deflate^[14] 算法来压缩处理过的文档,这样就能大幅度地减小文件的大小。使用 Gzip 压缩的 XML 主要用来和开启压缩模式的 EXI 做比较,所以把这种模式也放到了实验结果中。发出的请求文件的大小如图 1 所示,每一个请求所对应的响应数据文件大小如图 2 所示。图中的横坐标和表 1 中的类型标识相对应,图中 XML 表示原始的 XML 数据集,EXI 表示采用保留所有信息的标准模式编码,EXI-Compression 表示开启压缩模式的 EXI,XML+Gzip 表示使用 Gzip 压缩的 XML 文件。

从图 1 和图 2 中,可以看出 XML 文件的大小通常较大,几乎接近了两倍的 EXI 大小。这可能是由于 EXI 采用的是事件编码的方式,所以能够有效地减小文件所占用的体积。直接使用 Gzip 压缩的 XML,得到的文件大小通常小于 EXI,但是也有许多例外,比如图 1 和图 2 中横坐标为 1 时,使用 EXI 编码的文件大小小于直接使用 Gzip 压缩的 XML 文件大小。然而,从图 1 和图 2 看来,开启压缩模式的 EXI 总是具有最小

的文件体积。

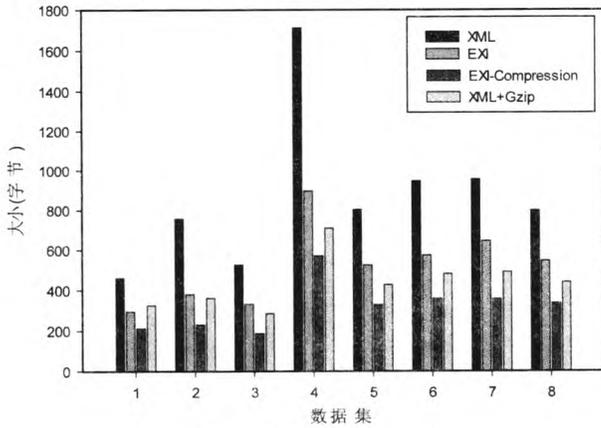


图 1 请求文件的大小

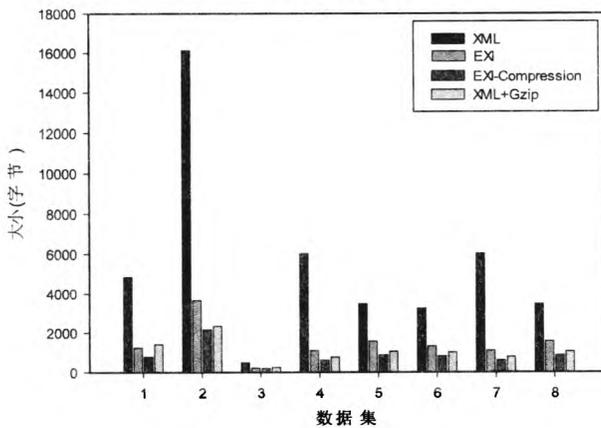


图 2 响应文件的大小

3.3 NS2 仿真网络性能

为了评估 EXI 的性能,在 NS2 中建模,将 EXI 和 XML 文件通过 UDP 的形式发送。假设所有的节点事先已经知道了在仿真过程中使用了 XML 和 EXI 格式的数据,所有的数据都通过 XML 或者 EXI 的形式来封装(XML+Gzip 的文件采用二进制的形式来传输)。文中使用前面第 3 部分和第 4 部分中的数据集进行仿真。除了原始的 XML 文件外,其他形式的文件都是由其转换过来的,包含了完全相同的信息。

所有的节点都均匀地分布在仿真的场景中,每一个节点和其邻居节点都在通信范围内。把发送请求数据的节点叫做客户端,而响应该请求的节点称为服务器。客户端可以发送表 1 中的八种请求类型的任意一种,但是一次只能发送一种类型的请求。而服务器只能响应这八种类型中的一种类型的请求,并且在仿真过程中不能改变该类型。服务器通过“使能-暂停-使能”的模式不断循环地提供服务,在仿真前随机选择部署在节点上。服务器提供服务的时间长短是随机的(采用均匀分布),并且其服务暂停的时间长短也是随机的(采用均匀分布)。客户端同样通过“使能-暂停-使能”的模式来发送请求,其模式和服务器类似。在仿真开始阶段,假设所有的节点都共享一个表,用来保

存服务器的相关信息,例如节点 ID,服务时间以及服务对应关系表等。当客户端想要发送一个服务请求时,必须先查找该表,找到一个此时正在提供该服务的服务器才能将请求发送出去,如果没有找到符合条件的服务器节点,则必须停止本次的请求发送而开始一个新的请求(如果该客户端处于使能状态)。这样做的目的是为了确保客户端总是能够发送请求给正在提供该服务的服务器节点。仿真的参数在表 1 中列出。

表 1 数据集信息

参数名	参数值	参数名	参数值
节点个数	61	背景流量类型	CBR
节点间距离(米)	50	背景流量大小	70 字节
仿真时间(秒)	10000	Channel	WirelessChannel
最小/大服务使能时间(秒)	50,60	Propagation	TwoWayGround
最小/大服务暂停时间(秒)	20,30	Phy	WirelessPhy
最小/大查询使能时间(秒)	50,60	Mac	802_11 DropTail/
最小/大查询暂停时间(秒)	10,20	Queue	PriQueue LL
服务的种类	8	Link layer	LL
客户端节点个数	19	Antenna	OmniAntenna
服务端节点个数	25	IFQ length	50
查询超时时间(秒)	10	Routing	AODV
UDP 数据包大小(字节)	81	Tx range	70

为了仿真网络传输的负载,仿真中加入了 CBR (Constants Bit Rate) 形式的背景流量,CBR 在仿真开始前随机地部署在网络节点间。并开启了 CBR 的随机模式来增加背景流量的随机性。为了分析性能,仿真结果通过处理后总结为以下几个方面:

a) 平均成功率:所有发出的查询请求中得到正确的响应数据的次数所占的百分比。当客户端发送了一个查询请求后,如果能接收到从服务器端传送来的完整响应数据,则认为这个查询是成功的。该项结果整理在图 3 中。

b) 平均请求延时:所有成功查询过程中经历的延时的平均值(以秒为单位)。一次成功查询的延时是指从发出查询开始,到收到完整的响应数据的时间。在计算过程中,只考虑成功传输的情况,因此查询超时或者返回的数据不正确都不会影响到该数值。这项结果如图 4 所示。

c) 平均节点能量消费:整个仿真过程中所有客户端和服务器节点的平均能量消耗。结果如图 5 所示。

图 3 ~ 图 5 可能不太符合通常的情况,比如图 3 中平均成功率应该随着背景流量发送速率的增大而减小。这可能是由于开启了 CBR 的随机间隔所导致的。仿真时的背景流量产生的间隔是随机的,这就使得背景流量发送的速率快不一定其发送的频率就高,从而较大的背景流量发送速率不一定对应较小的查询成功率。图 4 和图 5 中的情况和图 3 类似。

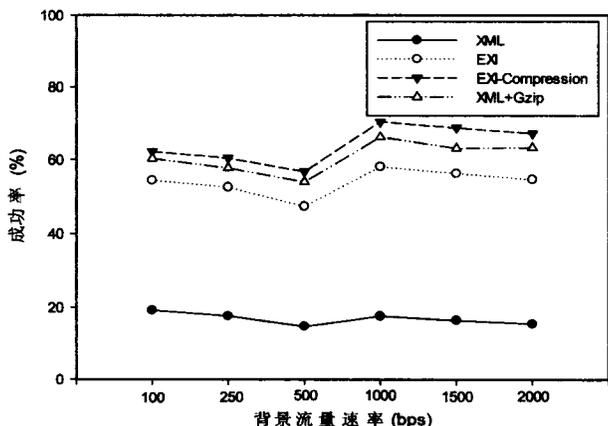


图 3 不同背景流量下的平均成功率

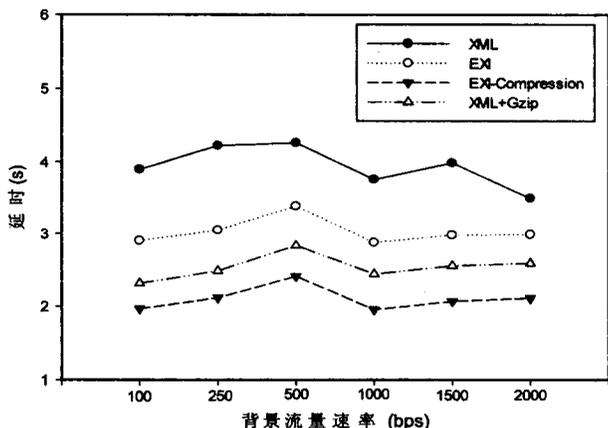


图 4 不同背景流量下的平均延时

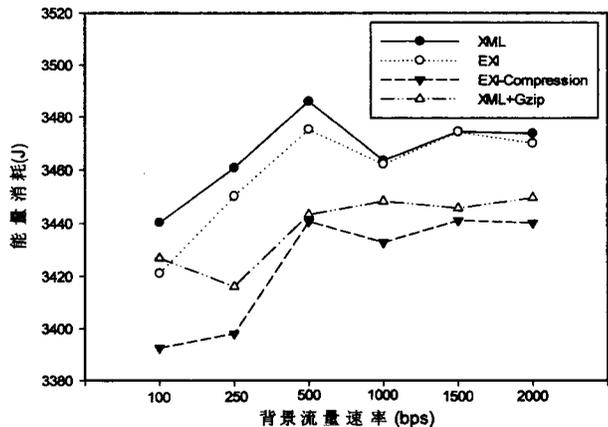


图 5 不同背景流量下的平均节点能量消耗

以上结果中的仿真数据经过了一定的处理,剔除了仿真中前 200 秒的数据,这是由于这个阶段的服务列表不稳定。在仿真开始时,服务列表里面的内容为空,只有当服务器节点初始化后才向表内添加相关信息,所以在服务器完全初始化前服务列表是不稳定的。如果把这部分数据也统计的话,对结果的正确性会有一定的影响。比如在该阶段中可能含有大量查询不成功的信息。

从仿真结果看出,使用传统的 XML 传输和其他三种形式的传输方式相比,平均成功率最低,平均延时最长,平均节点能量消费最大。这是由于传统 XML 文件可能含有大量冗余信息,导致其文件体积较大,而资源

受限网络中传输的数据包一般较小,所以较大的数据包必须分割成许多小的数据包来传输。然而,这些网络通常具有相对较高的丢包率,因此整个 XML 文件可能更难正确地传输,这就使得其具有较低的平均成功率。与此同时,较多的数据包传输将消耗更多的能量并且需要更多的时间来接收全部的数据,所以,传统 XML 传输具有更长的平均延时和更大的平均节点能量消耗。从这些结果看来,将传统 XML 传输机制部署在资源受限网络中似乎效率不高。

和 XML 相比 EXI 对数据的传输性能具有显著的提升。图 3 表明,EXI 的平均成功率是 XML 的两倍多,同时它和另外两种使用压缩传输模式的成功率比较接近。图 4 表明,EXI 的平均延时小于 XML 但是大于使用 Gzip 压缩的 XML 和开启压缩模式的 EXI。这是由于后两者都采用了压缩技术来处理数据文件,所以它们只需要传输较少的数据封包。从第 3.2 节可知,开启压缩模式的 EXI 比使用 Gzip 压缩的 XML 拥有更好的压缩效果,因此相比于使用 Gzip 压缩的 XML,EXI 具有更高的平均成功率和更小的平均延时。

图 5 展示了服务器节点和客户端节点的平均能量消耗,其大致的曲线表明 EXI 比 XML 拥有更低的平均能量消耗,开启压缩模式的 EXI 比使用 Gzip 压缩的 XML 平均节点能量消耗更低。通常,使用 Gzip 压缩的 XML 比 EXI 的平均节点能量消耗要低,但是也不是绝对如此,比如当背景流量的速率为 100bps 的时候,EXI 的平均节点能量消耗要低于 Gzip 压缩的 XML。这是由于使用 Gzip 压缩的 XML 文件并不总是减小文件的体积的,有时候反而使文件体积增大,或者由于文档的结构导致其压缩效果不佳,压缩后的文件大小比 EXI 编码的文件体积要大。相比之下,开启压缩模式的 EXI 性能则相对稳定,从图 5 看来相比其他传输形式它的平均能量消耗总是最低的。

尽管从以上图中来看 Gzip 压缩的 XML 和开启压缩模式的 EXI 比 XML 和 EXI 具有更高的效率,但是由于在数据处理阶段采用了压缩技术,其消耗的设备资源必定要多于后两者,所以通常将 EXI 和 XML 相比较,而把开启压缩模式的 EXI 和 Gzip 压缩的 XML 相比较。

通过前面的分析可以得出,EXI 比 XML 拥有更高的平均查询成功率,更短的平均延时,以及更少的平均节点能量消耗,因此,和 XML 相比它能更加高效地使用网络中的设备资源。

从数据的压缩传输来看,开启压缩模式的 EXI 比 Gzip 压缩的 XML 传输效率也要更高。所以,从文中实验上来看,EXI 比 XML 更加适合于资源受限的网络中。

4 结束语

文中通过在 NS2 上建模仿真,对 EXI 和 XML 在资源受限网络中的性能进行了评估。文章比较了 EXI 和 XML 以及开启压缩模式的 EXI 和 Gzip 压缩的 XML 之间的网络传输性能,对仿真结果从平均成功率、平均延时、平均节点能量消耗上给出了相应的分析。最后得出 EXI 比 XML 更加适合在资源受限网络中使用。下一步的工作主要是评估 EXI 在资源受限网络中使用的稳定性。

参考文献:

- [1] 李刚,孙红梅,李智,等.资源受限 Web 服务[J].计算机学报,2010,33(2):193-208.
- [2] Sakr S. XML compression techniques: A survey and comparison[J]. Journal of Computer and System Sciences,2009,75(5):303-322.
- [3] Deutsch P. GZIP file format specification version 4.3[S]. IETF RFC 1952,1996.
- [4] Liefke H,Suciu D. Xmill: An efficient compressor for XML data[C]//Proceeding of ACM SIGMOD International Conference on Management of Data. Dallas, Texas, USA: ACM Press, 2000:153-164.
- [5] Tolani P M, Haritsa J R. XGRIND: A Query-friendly XML Compressor[C]//Proceeding of 18th International Conference on Data Engineering. San Jose, California, USA: IEEE Press, 2002:225-234.
- [6] Min J K, Park M J, Chung C W. XPRESS: A Queriable Compression for XML Data[C]//Proceeding of the ACM SIGMOD International Conference on Management of Data. California, USA: ACM Press,2003:122-133.
- [7] Schneider J, Kamiya T. Efficient XML Interchange (EXI) Format 1.0. [S/OL]. 2011. <http://www.w3.org/TR/2011/REC-exi-20110310/>.
- [8] Na A, Priest M. Sensor Observation Service[S]. OGC 06-009r6,2007.
- [9] Boumez C. Efficient XML Interchange Evaluation[EB/OL]. 2009-04. <http://www.w3.org/TR/2009/WD-exi-evaluation-20090407/>.
- [10] Abangar H, Ghader M, Gluhak A, et al. Improving the Performance of Web Services in Wireless Sensor Networks[C]//Proceeding of Future Network and Mobile Summit. Florence, Italy: IEEE Press,2010:1-8.
- [11] Peintner D, Pericas-Geertsen S. Efficient XML Interchange (EXI) Primer[EB/OL]. 2009-12. <http://www.w3.org/TR/exi-primer/#compression>.
- [12] Bröring A, Echterhoff J, Jirka S, et al. New Generation Sensor Web Enablement[J]. Sensors,2011,11(3):2652-2699.
- [13] 王建国.一种新型的传感器 WEB 标准-传感器 WEB 整合框架[J].小型微型计算机系统,2008,29(9):1647-1651.
- [14] Deutsch P. DEFLATE Compressed Data Format Specification version 1.3[S]. IETF RFC 1951,1996.

(上接第 53 页)

流的仓储管理策略制定,对我国物流业的发展具有重要意义。

参考文献:

- [1] 郭俊义. 计算机模拟理论方法及其应用[M]. 北京:宇航出版社,1988.
- [2] 李保刚. 空空导弹攻击过程视景仿真研究[J]. 计算机技术与发展,2012,22(3):113-116.
- [3] 刘瑞叶,任洪林,李志民. 计算机仿真技术基础[M]. 第 2 版. 北京:电子工业出版社,2011.
- [4] Inderfurth K, van der Laan E. Leadtime effects and policy improvement for stochastic inventory control with remanufacturing[J]. International Journal of Production Economics,2011,71(2):381-390.
- [5] 肖田元,范文慧. 系统仿真导论[M]. 第 2 版. 北京:清华大学出版社,2010.
- [6] 王维平. 离散事件系统建模与仿真[M]. 北京:科学出版社,2007.
- [7] 刘满凤,陶长琪,柳键,等. 运筹学教程[M]. 北京:清华大学出版社,2010.
- [8] Çetinkaya S, Parlar M. Optimal myopic policy for a stochastic inventory problem with fixed and proportional backorder costs[J]. European Journal of Operational Research, 1998, 110(1):20-41.
- [9] 胡运权. 运筹学基础及应用[M]. 哈尔滨:哈尔滨工业大学出版社,1998.
- [10] 邱丽绚. 随机库存系统的统计计算研究[J]. 淮海工学院学报(自然科学版),2005,14(3):8-10.
- [11] 杨肇夏. 计算机模拟及其应用[M]. 北京:中国铁道出版社,1999.
- [12] Zhong Yao, Ke Liu, Stephen C H, et al. Single period stochastic inventory problems with ordering or returns policies[J]. Computers & Industrial Engineering,2011,61(2):242-253.
- [13] Dorfman M J, Medanic J. Scheduling trains on a railway network using a discrete event model of railway traffic[J]. Transportation Research Part B: Methodological,2004,38(1):81-98.
- [14] Rey G D. Reading direction and signaling in a simple computer simulation[J]. Computers in Human Behavior, 2010, 25(6):1176-1182.
- [15] 张延林,程锡礼. 随机库存系统的存货策略计算机仿真[J]. 系统工程,2005,23(1):119-123.

资源受限网络中高效XML交换的性能评估

作者: [谢晓燕](#), [王浩](#), [陈彦萍](#), [XIE Xiao-yan](#), [WANG Hao](#), [CHEN Yan-ping](#)
作者单位: [谢晓燕, 王浩, XIE Xiao-yan, WANG Hao \(西安邮电大学计算机学院, 陕西西安, 710121\)](#), [陈彦萍, CHEN Yan-ping \(西安邮电大学计算机学院, 陕西西安710121; 西安交通大学计算机系, 陕西西安710049\)](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2013, 23(6)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjz201306014.aspx