面向可重构基础网络的部分扇出多播交换阻塞率模型

张博，汪斌强，朱圣平
（国家数字交换系统工程技术研究中心，河南 郑州 450002）

摘 要：传统网络采用调度前复制和扇出拷贝方式的多播交换模型不具备大规模可扩展能力。面向可重构基础网络，提出了一种部分扇出多播交换模型，该模型采用 2×2 布尔单元和布尔群组集线器建立基本交换结构，采用部分扇出拷贝方式和四状态分割编码实现自路由路径选择，进而推导了该模型在单多播混合业务源输入下的单播阻塞率、多播阻塞率和多播扇出率迭代过程。仿真实验表明：该模型在 Bernoulli 均匀业务源条件下，单播归一化负载强度为 0.2 时，多播阻塞率在 10^{-10}～10^{-2} 之间，多播时延总小于百纳秒量级上限，能够为到达业务提供时延上限保障。

关键词：可重构基础网络；部分扇出；布尔单元；阻塞率
中图分类号：TP 393 **文献标识码：**A **文章编号：**1000-436X(2012)05-0049-09

Partial fanout multicast switching model of reconfigurable foundation network

ZHANG Bo, WANG Bin-qiang, ZHU Sheng-ping
(National Digital Switching System Engineering Technological R&D Center, Zhengzhou 450002, China)

Abstract: The multicast switching model using copying before scheduling and fanout copying modes had worse large-scale extensibility capability in traditional technology system. A partial fanout multicast switching model with basic switching fabric comprised by 2×2 Boolean cells and Boolean group concentrators was proposed facing to reconfigurable foundation network. Self-routing path selected was realized by partial fanout copying mode and quaternary symbols cut-through coding. Unicast blocking probability, multicast blocking probability and multicast fanout ratio iterative course was deduced based on unicast and multicast mixed business source. The simulation result shows that the value of multicast blocking probability is from 10^{-10} to 10^{-2} based on Bernoulli uniform business source when normalized unicast load 0.2, the model can meet time QoS requirement because the value of multicast time delay is less than hundred nano-second level limit.

Key words: reconfigurable foundation network; partial fanout; Boolean cell; blocking probability

1 引言

目前，互联网开始进入下一代网络新时期，据 2010 年 Arbor 公司与美国密西根大学的一份持续 2 年的联合研究报告[1]，多媒体音视频的应用服务成为了当今互联网的发展主流。中国互联网络信息中心(CNNIC)2011 年 7 月中国互联网络发展状况统计报告[2]表明：网络视频继续保持平稳增长，2011 年

收稿日期：2011-11-22；修回日期：2012-03-28
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2012CB315901, 2012CB315905）；国家自然科学基金资助项目（61179028）
Foundation Items: The National Basic Research Program of China (973 Program) (2012CB315901, 2012CB315905); The National Natural Science Foundation of China (61179028)

6月中国网络视频用户达3.01亿，作为与线下电视最为相近的互联网服务，网络视频服务是使用最多的服务之一。视频流有典型的特性，其数据内容趋向于单点发送，多点接收，即多播。如今流行的基于视频应用的IPTV、视频会议、交互式仿真、多方游戏等应用的流量都具有多播特性。

多播交换结构研究开始于1984年，Huang和Knauer首次对ATM网络中多播数据的交换问题予以关注，并设计一种新型交换结构来支持多播，由此提出了第一个多播交换结构——Starlite[3]。1988年，Deering[4]提出了将多播功能结合到数据网IP层的多播结构。1992年，多播实验网——Mbone建立[5]。目前，较完整的多播协议体系已经形成，IP多播的研究工作逐渐集中于流量控制和拥塞控制研究，无线多播研究和大规模高效多播研究。

对于大规模高效多播研究，基于banyan的多播交换结构大多采用多播复制的方法，Tony Lee[6]提出了一种内部无阻塞的复制网络设计方案，但结构过于复杂。文献[7]对该结构进行改进，但输入分组复制要求的总和超出输出端口的数量时，满溢的现象便会发生。Yeh等人[8]提出了一种基于Knockout理论多播交换结构，每个输入端口采用完全连接的方式连接到所有的输出接口模块。但该方案的大规模扩展性能存在瓶颈。基于Clos结构的多播交换，文献[9]引入了一种路径分配向量，但整个路径决策时间复杂度为$O(K)$，无法大规模扩展。2010年的文献[10]证明了对Clos网络进行多播路径匹配是NP完全问题。基于Crossbar多播交换结构是当前的主流类型，Pan等人[11]提出了一种将信元载荷与目的地址信息分开存储的方案，但交换结构的加速比高。文献[12]针对IPTV提出了一种扇出无关的多播交换结构，但该结构要求到达过程满足强大数定律，且多播输出阻塞率高。基于Crossbar缓存单多播联合调度算法[13]采用了分级和层次化的调度机制，但该调度机制仍然存在$O(N(2^N-1))$队列复杂度瓶颈。

以上分析中多播结构和调度机制固有的瓶颈使其不具备大规模扩展能力，可重构信息通信基础网络理论体系[14]对多播问题的原因进行了分析。从多播研究的整体看，终端基本是以软件方式实现各种应用层，通常不存在瓶颈，而路由交换节点的实现是多播性能的关键。路由交换节点的多播实现方式可以分为多次单播软件调度方式和硬件电路线速扇出拷贝方式。目前路由交换节点上多播多数采用多次单播的软多播[15]，即在路由前对多播分组先进行复制而后各自在源端口队列排队，但软多播实时性差，不能保证服务质量。

在可重构基础网络中，有效弥合多播业务特性和网络服务能力的一个途径是将多播业务特征需求与网络承载服务二者抽象成一种特定的“业务—服务”映射模型，多播服务根据多播业务需求，构建节点和链路资源独享的可重构多播服务承载网，这就要求交换层面必须采用完全分布式的控制机制，才能将节点交换资源有效分割给多播业务。因此，可重构基础网络多播交换机制采用硬件电路线速扇出的多播拷贝方式和自路由寻路方式。其分布式自路由寻路方式有效地解决了缓存调度的瓶颈问题；其线速扇出与网络负载相适应，避免了输出竞争等待；其等级数的交换路径，时延抖动小，避免了时延抖动大造成的多播性能下降问题。

基于此，本文提出了部分扇出多播交换阻塞率模型，该模型具有完全分布式自路由，无需端口匹配，无内部缓存，无缓存时延的特点。当构建纯多播服务承载网或单多播混合服务承载网时，可采用关闭部分2×2布尔单元和级间比特置换的重构操作实现其交换资源的逻辑隔离，由于其分布式的特点，多播服务承载网之间无相关性，大大简化了多个可重构服务承载网构建的控制复杂度，增强了多播交换结构的可扩展性。

2 逐级部分扇出交换结构

定义1[16] **并播单元**：0-1并播单元输入定义在字母表$\{0-bound,1-bound,idle,bicast\}$，即定义在$\{10,11,00,bicast\}$域中，规定$10<00=bicast<00$，$bicast$为一个输入端的输入信号同时输出到Output-0和Output-1，在判决选路时，使用表1的规则。当$bicast$与空为输入时，则输出端口同时输出该信号。

表1 基于{10,bicast,11}域的输入控制

单元状态		Input-1控制信号			
		10	00	bicast	11
Input-0控制信号	10	Any	Bar	Bar	Bar
	00	Cross	Any	bicast	Bar
	bicast	Cross	bicast	Any	Cross
	11	Cross	Cross	Cross	Any

定义2[16] **布尔单元**：0-1并播单元可以看成2×2的布尔单元，其交换规则为：输入为含有2个元u和v的偏序集，输出相当于对输入偏序集中的

元求上确界和下确界，由格理论知 $u \wedge v$ 为下确界，$u \vee v$ 为上确界。

其中，$u \vee v$ $= u$ or v，$u \wedge v$ $= u$ and v，如图 1 所示，满足：当 $u \leq v$ 时，则 $u \wedge v = u, u \vee v = v$，当 $u \geq v$ 时，则 $u \wedge v = v, u \vee v = u$。

图 1　布尔单元

$0,1,I,B$ 分别代表 $0-bound, 1-bound, idle, bicast$ 输入，则有表达式 $0 \wedge 1 = 0, 0 \vee 1 = 1$，$I \wedge B = 0, I \vee B = 1$。

定义 3　布尔群组集线器：采用布尔单元组成的 banyan 类网络或扩展 banyan 类网络为布尔群组集线器，对于 $2G - \text{to} - G$ 的布尔群组集线器满足：该集线器将 $2G$ 个输入信号中最大的 G 个信号传输到具有最大输出地址（输出端口从上往下的地址编序为二进制表示的 $0 \to 2G-1$）的 G 个输出端口，并将其余信号传输到最小输出地址的 G 个输出端口。

$2G - \text{to} - G$ 布尔群组集线器构建方法可参见文献[17]，以双调循环排序器为例，其基本排序原理如图 2 所示，采用点线式表示双调循环排序器网络，设 $a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p \in P$，$P$ 为线性序集，根据布尔多项式比较方式可以得到排序序列 $A_1 \leq A_2 \leq \text{L} \leq A_{16}$，其中，$A_1 = a \wedge b \wedge \text{L} \wedge p$，$A_{16} = a \vee b \vee \text{L} \vee p$。如果 P 是 $\{0,1\}$，则 $a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p$ 中任意输入为 1 通过该网络都可以到达输出端的 8 个高地址，其余的到达 8 个低地址，假设该输入中有 6 个输入变量为 1，即 $a=1, f=1, g=1, k=1, m=1, o=1$，经该集线器，得到 $A_{11} = A_{12} = A_{13} = A_{14} = A_{15} = A_{16} = 1$.

对于一个 $N \times N$ 的单播交换结构，其输入信元目的地址可能组合为 N^N。而对于 $N \times N$ 多播交换结构，由于每一个输入信元的目的地址可能是部分或全部输出端口的组合，则整个交换结构其输入信元目的地址的组合多达 $(2^N - 1)^N$。当 N 大时，大规模多播交换结构的调度配对具有很高的复杂性，因此在大规模（如万端口交换结构）的情况下，要求交换结构在数据传输控制过程中不能采用集中式方式，而要分布式配对，以减小复杂度，因此采用自路由的方式来实现多播数据传输。由于一个多播信元有 $2^N - 1$ 种目的地址的组合，其信元分组自路由的编码开销的最少二进制位为 $O(\text{lb}\ (2^N - 1)) \approx O(\text{lb}\ 2^N) = O(N)$，当交换结构的规模变大时，如 $N = 1\,024$，则每个多播信元自路由目的地址最小的编码代价为 1 024bit=128byte，巨大的编码开销使得带宽的利用率变低，信元传输的时延增大，故在信元传输过程中必须采用分级控制。

参考多路径自路由模型[1]建立基于布尔群组集线器的 banyan 类交换网络，设 $N = 2^n$, $N = KG$，$K = 2^k, G = 2^g$，先构造一个 $K \times K$ banyan 类网络，这里采用分治 banyan 类（divide-and-conquer）网络。然后将网络中的 2×2 交换单元用 $2G - \text{to} - G$ 布尔群组集线器代替。取 $K = 2^4$，$G = 2^3$，可得图 3 所示的交换结构。该交换结构逐级进行自路由选路，先将分组交换到某个输出群组，而后进行时分输出。该交换结构的多播分组在每个布尔群组集线器中部分扇出。

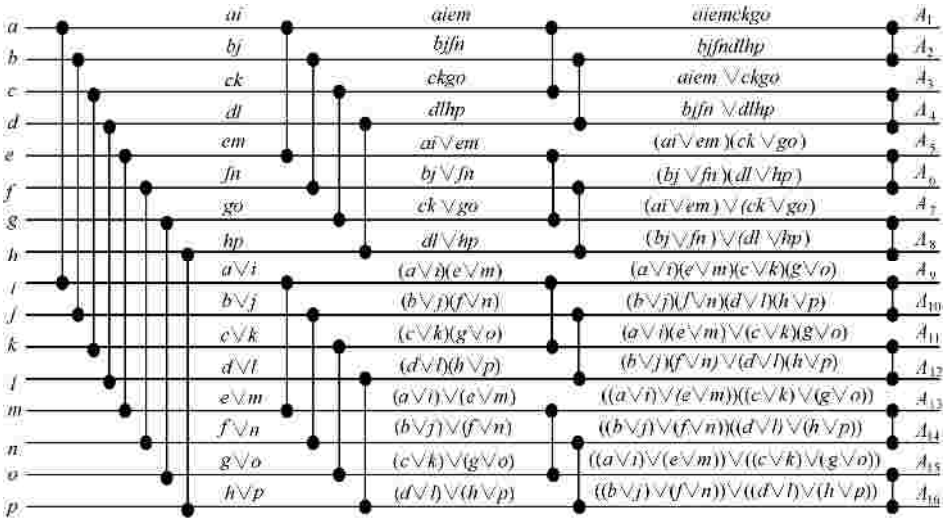


图 2　双调循环排序器点线式结构

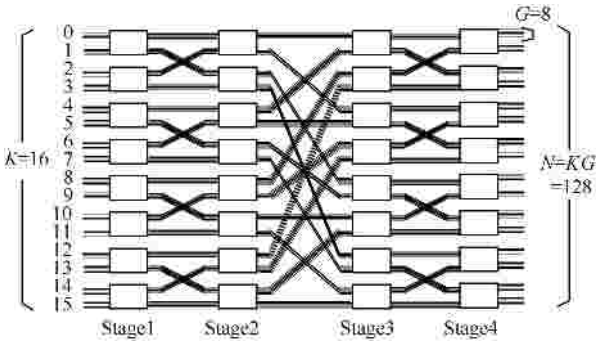


图 3 128×128 布尔群组集线器 banyan 类交换网络

3 部分扇出多播交换阻塞率模型

3.1 部分扇出自路由机制

如图 4 所示，交换结构的自路由分组格式分为带内控制信息和有效载荷 2 部分，带内控制信息进行选路控制。自路由方式有 2 种，一种为每经过一级，该级的有效位被处理后决定输出，所以对于第 $j,1 \le j \le m$ 级，$1g_j$ 为该级的有效选路信息，由此可实现该控制分组的正确寻路。另一种为每经过一级，其有效位在下一级依然保留被再利用，控制信号在每一级均为 $1g_j g_{j+1} \llcorner g_m$。在集线器互联的 banyan 类网络中选择前者，在布尔群组集线器中选择后者。

图 4 自路由分组格式

以 8-to-3 布尔群组集线器为例，由图 5(a)和图 5(b)对比不难看出，由于输出信息中仍然含有 B 控制信息，因此多播交换网络采用部分扇出的方式。当有单播加入时，即将图 5(a)中的 2 个 I 控制位，变为 0,1 控制位，则图 5(b)中输出信息中含有 2 个 B 的控制信息，多于图 5(a)中输出的 B 控制信息数，即有一个 B 没有与 I 进入过同一个布尔单元，所以其多播性能会下降。假设输入信息中无 I 信息，那么该多播信息就无法扇出。

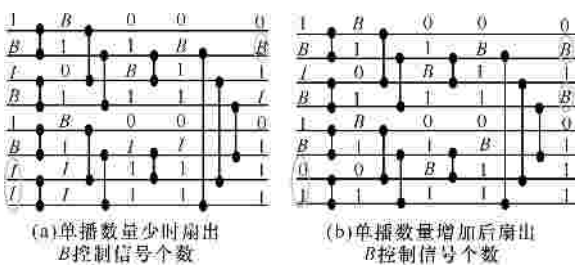


图 5 8-to-3 布尔集线器点线表示

结论：布尔群组集线器，具有部分扇出单播抢占优先的多播能力。

对于 $2^n \times 2^n$ 布尔多播交换结构，根据信息熵理论，其输出地址的任意子集带内控制信息都不能低于 2^n bit，但可以采用 2^n-1 的 $0,1,I,B$ 四状态带内控制信息来满足二进制比特的带内控制。带内控制信息必须包括每一级的控制信息，这里采用四状态的分割编码(cut-through coding)法[17]。其带内控制信息为

$$\llcorner Q_{110}Q_{010}Q_{100}Q_{000}Q_{11}Q_{01}Q_{10}Q_{00}Q_1Q_0Q$$

自右向左进行带内控制，前边的位决定后边控制位的分割。对于第一级输入，当 $Q=B$，控制信息 $\llcorner Q_{010}Q_{000}Q_{01}Q_{00}Q_0$ 从 0 端口输出，控制信息 $\llcorner Q_{110}Q_{100}Q_{11}Q_{10}Q_1$ 从 1 端口输出，同时有效载荷扇出。当 $Q=0$，控制信息 $\llcorner Q_{010}Q_{000}Q_{01}Q_{00}Q_0$ 从 0 端口输出，且 1 端口控制信息无效。当 $Q=1$，控制信息 $\llcorner Q_{110}Q_{100}Q_{11}Q_{10}Q_1$ 从 1 端口输出，且 0 端口控制信息无效。对于第二级输入，当 $Q_0=B$ 时，控制信息 $\llcorner Q_{000}Q_{00}$ 从 0 端口输出，控制信息 $\llcorner Q_{010}Q_{01}$ 从 1 端口输出，同时有效载荷扇出。每一级依次类推，可将分组传输到最终的目的地址。以 $2^3 \times 2^3$ banyan 网络为例。由图 6(a)可见其 2 个带内控制信号共含有 4 个 B，在无单播与其竞争的情况下，扇出到 6 个目的端口。当某一 $2^n \times 2^n$ 布尔多播交换网络输入只有一个带内控制信号为(2^n-1)bit B 时，可扇出到 2^n-1 个目的地址。由图 6(b)可见，当给网路加入 2 条单播数据分组，单播数据分组也会与该交换网络中的多播产生竞争，而使多播扇出很少的分组。

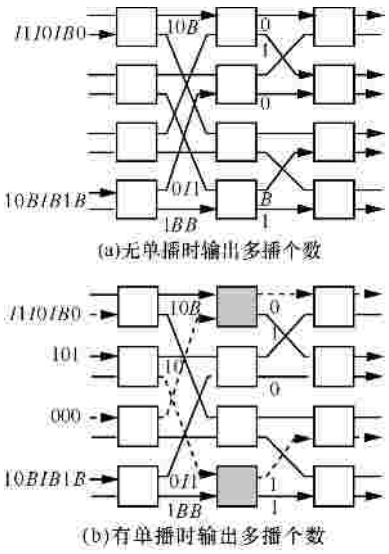


图 6 banyan 类网络多播自路由

结论：布尔多播 banyan 类交换结构，具有部分扇出单播抢占优先的多播能力。

3.2　多播交换阻塞率模型

单多播混合输入的交换网络，先确定单播与多播争用的布尔群组集线器中多播扇出的情况，扇出个数不同其阻塞率不同。由于布尔群组集线器具有部分扇出特性，因此随着级数的增加，其分组流的个数不断增大，因此其阻塞率不断增大。

单播自路由信息为 $d_{j(g)}\llcorner d_{j(2)}d_{j(1)}1$，其中，$j(1),j(2),\llcorner,j(g)$ 为导向转换信息[2]。多播分组头信息为 $\llcorner Q_{110}Q_{010}Q_{100}Q_{000}Q_{11}Q_{01}Q_{10}Q_{00}Q_1Q_0Q1$，第一位 1 为有效位，序列为从右向左依次排序，多播分组头信息可转换为多个单播分组头信息，且满足导向转换。

第 $i(1\le i\le m)$ 级布尔群组集线器内，单播控制信息不变为 $d_{j(i)}1$，多播控制信息为 $Q_{q_i}1$，其中，q_i 为二进制数，长度为 $i-1$，每次当 B 与 I 在同一个 2×2 布尔单元时，其后边的控制信息采用 2.2 节方法进行码分割。

下面分析该结构在不同负载下的阻塞率。假定到达同一输出群组内的分组相互独立，设随机变量 $X_{I_i}/X_{O_i}(1\le i\le m)$，代表在某一时隙，到达第 i 级 $2G-\mathrm{to}-G$ 布尔群组集线器复用交换单元某输入/输出群组的分组数目，$0\le X_{I_i}\le G,0\le X_{O_i}\le G$。

X_{I_1} 为第一级交换单元某输入群组的分组数目，假设单播和多播业务都服从贝努利均匀分布，设单播业务源到达强度的参数 $l=p_1$，则单播业务源在单个时隙里以概率 p_1 到达，设多播业务源到达强度的参数 $l=p_2$，则多播业务源在单个时隙里以概率 p_2 到达，满足条件 $p_1+p_2\le1$，则第一级某一输入群组到达单播分组个数 $X_{I_{1s}}$ 为

$$P\left(X_{I_{1s}}=x_{1s}\right)=\binom{G}{x_{1s}}p_1^{x_{1s}}\left(1-p_1\right)^{G-x_{1s}}$$

第一级某一输入群组到达多播分组个数 $X_{I_{1m}}$ 为

$$P\left(X_{I_{1m}}=x_{1m}\right)=\binom{G}{x_{1m}}p^{x_{1m}}\left(1-p\right)^{G-x_{1m}}$$

其中，$p=p_1+p_2$，$x_1=x_s+x_m$，x_s 表示单播个数，x_m 表示多播个数。对于布尔群组集线器到达分组并不是最后的输出分组个数，为更多的扇出多播分组，假设每个多播分组带内控制信号由 2^k-1 个 B 构成(全 B 构成)，单个布尔集线器中单播的优势导致其多播以一定概率扇出，因为集线器中只有 1bit 有效位，所以每一个多播分组只能扇出 1 个或 2 个分组。对于第一级集线器有输入 X_{I_1}，Y_{I_1} 独立同分布，假设 A_{I_1} 表示第一级某 $2G-\mathrm{to}-G$ 布尔群组集线器输入分组数总和，为简化表示，采用 $D_I(d_i)$ 表示 $D_I=d_i$ 的概率 $P(D_I=d_i)$。

$$A_{I_1}(a_1)=\sum_{x_{1s}+x_{1m}+y_{1s}+y_{1m}=a_1}X_{I_{1s}}(x_{1s})X_{I_{1m}}(x_{1m})Y_{I_{1s}}(y_{1s})Y_{I_{1m}}(y_{1m})$$

$\forall a_1\le2G$，$\exists I\wedge B=0,I\vee B=1$，使得 $a_1+x_{1m_2}+y_{1m_2}\le2G$，其中，$x_{1m}=x_{1m_1}+x_{1m_2},y_{1m}=y_{1m_1}+y_{1m_2}$，$x_{1m_1},y_{1m_1}$ 为无扇出的多播分组，x_{1m_2},y_{1m_2} 为扇出 2 个分组的多播分组。实际输出的分组的总数 a'_1 为

$$A_{O_1}(a'_1)=\begin{cases}\sum\limits_{x_{1s}+2x_{1m}+y_{1s}+2y_{1m}=a'_1}X_{I_{1s}}(x_{1s})X_{I_{1m}}(x_{1m})\cdot\\ Y_{I_{1s}}(y_{1s})Y_{I_{1m}}(y_{1m}),\quad a'_1<2G\\ \sum\limits_{x_{1s}+y_{1s}=0}^{2G}\left(X_{I_{1s}}(x_{1s})Y_{I_{1s}}(y_{1s})\sum\limits_{2x_{1m}+2y_{1m}=a'_1-x_{1s}+y_{1s}}^{2G}\right.\\ \left.X_{I_{1m}}(x_{1m})Y_{I_{1m}}(y_{1m})\right),\quad a'_1=2G\end{cases}$$

当 $a'_1<2G$ 时多播全被扇出，当 $a'_1=2G$ 时，存在部分多播无扇出的情况。当有 $z'_1<G$ 的分组到达某输出群组，则至少有 $z'_1\le a'_1<2G$ 分组到达该集线器的输入，无路径错误选择，当有 $z'_1=G$ 的分组到达某输出群组，则有 $a'_1=G,a'_1=G+1,\llcorner,a'_1=2G$ 个分组竞争该输出群组，则会发生某些分组路径错误选择，导致阻塞。则第一级布尔集线器某输出群组单多播混合分组个数 z'_1 分布为

$$X_{O_1}(z'_1)=\begin{cases}\sum\limits_{a'_1=z'_1}^{2G}A_{I_1}(a'_1)2^{-a'_1}\binom{a'_1}{z'_1},\quad z'_1<G\\ \sum\limits_{a'_1=z'_1}^{2G}A_{I_1}(a'_1)2^{-a'_1}\sum\limits_{k=G}^{a'_1}\binom{a'_1}{k},\quad z'_1=G\end{cases}$$

则第一级布尔群组集线器输出单播分组概率分布 $X_{O_{1s}}(x'_{1s})$ 为

$$X_{O_{1s}}(x'_{1s})=\begin{cases}\sum\limits_{a'_1=x'_{1s}}^{2G-1}l_{1s}A_{I_1}(a'_1)2^{-a'_1}\binom{a'_1}{x'_{1s}}+\\ \sum\limits_{a'_1=2G}l'_{1s}A_{I_1}(a'_1)2^{-a'_1}\binom{a'_1}{x'_{1s}},x'_{1s}<G\\ \sum\limits_{a'_1=x'_{1s}}^{2G-1}l_{1s}A_{I_1}(a'_1)2^{-a'_1}\sum\limits_{k=G}^{a'_1}\binom{a'_1}{k}+\\ \sum\limits_{a'_1=2G}l'_{1s}A_{I_1}(a'_1)2^{-a'_1}\sum\limits_{k=G}^{a'_1}\binom{a'_1}{k},x'_{1s}=G\end{cases}$$

其中：

$$l_{1s}=\frac{X_{I_{1s}}(x'_{1s})}{X_{I_{1s}}(x'_{1s})+X_{I_{1m}}\left(\frac{1}{2}(a'_1-x'_{1s})\right)}$$

$$l'_{1s}=\frac{X_{I_{1s}}(x'_{1s})}{X_{I_{1s}}(x'_{1s})+X_{I_{1m}}\left(\frac{1}{2}(G-x'_{1s})\right)}$$

第一级集线器输出多播分组概率 $X_{O_{1m}}(x'_{1m})$ 为

$$X_{O_{1m}}(x'_{1m})=\begin{cases}\sum\limits_{a'_1=x'_{1m}}^{2G-1} l_{1m}A_{I_1}(a'_1)2^{-a'_1}\binom{a'_1}{x'_{1m}}+\\ \sum\limits_{a'_1=2G} l'_{1m}A_{I_1}(a'_1)2^{-a'_1}\binom{a'_1}{x'_{1m}}, x'_{1m}<G\\ \sum\limits_{a'_1=x'_{1m}}^{2G-1} l_{1m}A_{I_1}(a'_1)2^{-a'_1}\sum\limits_{k=G}^{a'_1}\binom{a'_1}{k}+\\ \sum\limits_{a'_1=2G} l'_{1m}A_{I_1}(a'_1)2^{-a'_1}\sum\limits_{k=G}^{a'_1}\binom{a'_1}{k}, x'_{1m}=G\end{cases}$$

其中，$l_{1m}=1-l_{1s}$ $l'_{1m}=1-l'_{1s}$，已知第 i 级的输出为第 $i+1$ 级输入，即 $X_{O_{is}}$,$X_{O_{im}}$,X_{O_i} 与 $X_{I_{(i+1)s}}$,$X_{I_{(i+1)m}}$，$X_{I_{i+1}}$ 符合相同分布，实际输入分组总数 a_{i+1} 服从以下分布：

$$A_{I_1}(a_{i+1})=\sum_{x_{(i+1)s}+x_{(i+1)m}+y_{(i+1)s}+y_{(i+1)m}=a_{1+1}} X_{I_{(i+1)s}}(x_{(i+1)s})\cdot X_{I_{(i+1)m}}(x_{(i+1)m})Y_{I_{(i+1)s}}(y_{(i+1)s})Y_{I_{(i+1)m}}(y_{(i+1)m})$$

实际输出的分组总数 a'_{i+1} 服从以下分布：

$$A_{O_1}(a'_{i+1})=\begin{cases}\sum\limits_{x_{(i+1)s}+2x_{(i+1)m}+y_{(i+1)s}+2y_{(i+1)m}=a'_{1+1}} X_{I_{(i+1)s}}(x_{(i+1)s})X_{I_{(i+1)m}}(x_{(i+1)m})\cdot\\ Y_{I_{(i+1)s}}(y_{(i+1)s})Y_{I_{(i+1)m}}(y_{(i+1)m}), a'_{i+1}<2G\\ \sum\limits_{x_{(i+1)s}+y_{(i+1)s}=0}^{2G}\left(X_{I_{(i+1)s}}(x_{(i+1)s})Y_{I_{(i+1)s}}(y_{(i+1)s})\cdot\right.\\ \left.\sum\limits_{2x_{(i+1)m}+2y_{(i+1)m}=a'_{i+1}-x_{(i+1)s}+y_{(i+1)s}}^{2G} X_{I_{(i+1)m}}(x_{(i+1)m})Y_{I_{(i+1)m}}(y_{(i+1)m})\right),\\ a'_{i+1}=2G\end{cases}$$

第 $i+1$ 级布尔集线器某输出群组单多播混合分组个数 z'_{i+1} 分布为

$$X_{O_{i+1}}(z'_{i+1})=\begin{cases}\sum\limits_{a'_1=z'_{i+1}}^{2G} A_{I_1}(a'_{i+1})2^{-a'_{i+1}}\binom{a'_{i+1}}{z'_{i+1}}, & z'_{i+1}<G\\ \sum\limits_{a'_{i+1}=z'_{i+1}}^{2G} A_{I_1}(a'_{i+1})2^{-a'_{i+1}}\sum\limits_{k=G}^{a'_{i+1}}\binom{a'_{i+1}}{k}, & z'_{i+1}=G\end{cases}$$

从而推导出第 $i+1$ 级布尔群组集线器输出单播分组个数 $X_{O_{1s}}$ 分布和多播分组个数 $X_{O_{1m}}$ 分布。迭代可得第 k 级的布尔集线器某输出群组单多播混合分组个数 X_{O_k} 分布，单播分组个数 $X_{O_{ks}}$ 分布，多播分组个数 $X_{O_{km}}$。

单播阻塞率指平均输出单播分组数与平均输入单播分组数的比，得式(1)。

$$P_r=1-\frac{E(X_{O_{ks}})}{Gp_1}=1-\frac{\sum\limits_{x'_{ks}=0}^{G}x'_{ks}X_{O_{ks}}(x'_{ks})}{Gp_1}\tag{1}$$

多播阻塞率指平均输出多播分组数与最大输出多播分组数的比，得式(2)。

$$P_r=1-\frac{E(X_{O_{km}})}{(k+1)Gp_2\bmod(G-Gp_1)}=1-\frac{\sum\limits_{x'_{km}=0}^{G}x'_{km}X_{O_{km}}(x'_{km})}{(k+1)Gp_2\bmod(G-Gp_1)}\tag{2}$$

其中，$(k+1)Gp_2\bmod(G-Gp_1)$ 含义为：全 B 的带内控制信息，理想最大可扇出 $(k+1)Gp_2$ 个分组，但受到输出端口数 G 和单播输入分组数 Gp_1 的影响，最大输出 $(k+1)Gp_2\bmod(G-Gp_1)$ 个多播分组。

多播扇出率指输出多播分组个数与输入多播分组个数的比，得式(3)。

$$g_m=\frac{\sum\limits_{x'_{km}=0}^{G}x'_{km}X_{O_{km}}(x'_{km})}{Gp_2}\tag{3}$$

4 仿真实验

本节采用 MATLAB 软件对布尔群组集线器自路由交换网络在均匀分布业务源下的单播阻塞率、多播阻塞率和扇出率进行仿真。对于 Bernoulli 均匀分布业务源，群组业务到达服从二项分布，业务强度采用归一化负载强度表示，对于任意输入群组 i，满足 $l_{ij}=l/K$，用 l_{ij} 表示到达输入群组 i 去往输出群组 j 的速率。

1) 单播阻塞率仿真

设定多播负载强度 $p_2 = 0.2$，单播负载强度 $0 \le p_1 \le 1 - p_2$。对于 $N \times N$ 的交换网络，$N = KG$，$K = 2^k$，$G = 2^g$，在 $k = 6$ 的条件下比较 $G = 8,16,32$ 时的单播阻塞率，如图 7(a)所示，G 越大其阻塞率越小。在 $G = 16$ 的条件下比较 $k = 4,6,8$ 时的阻塞率，如图 7(b)所示，k 越大其阻塞率越大。该 2 种条件下，当归一化负载强度 > 0.6 时，其阻塞率大于等于 10^{-2}，会造成单播分组的大量错传。

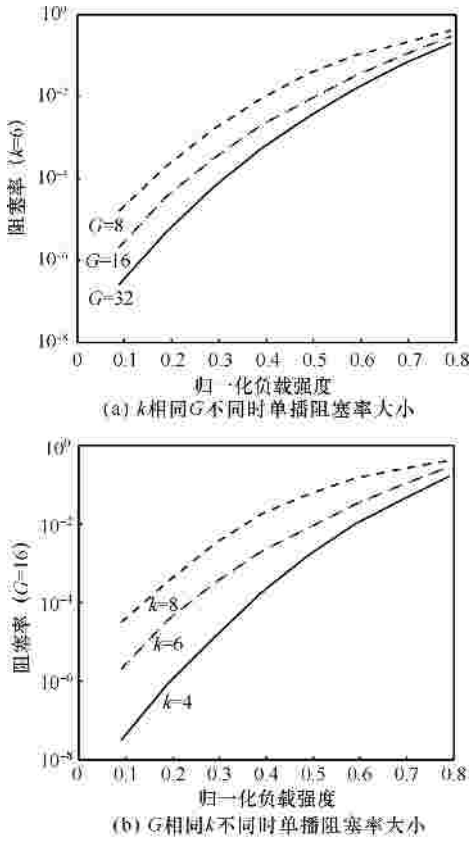


图 7　单播阻塞率

原因分析：k 决定交换网络的级数，在 k 不变的情况下，多播负载强度 $p_2 = 0.2$，不同的 G 时，由于级数相同，其多播分组扇出个数对单播的影响基本相同。但 G 越大，其 $x \le G$ 的概率越大，从而减少了 $x > G$ 时分组在布尔群组集线器中的内部阻塞概率。当 G 一定时，k 越大，交换结构的级数越大，由于多播的扇出特性，随着级数的增加，其多播扇出数就会增加，增加了分组的个数，加大了内部阻塞率。而且每一级由于布尔群组集线器中分组对输出群组的均匀选择，存在一定的路径错传概率，同样加大了单播阻塞率。

2) 多播阻塞率仿真

设定单播负载强度 $p_1 = 0.2$，多播负载强度满足 $0 \le p_2 \le 1 - p_1$。在 $k = 6$ 的条件下比较 $G = 8,16,32$ 时的多播阻塞率，如图 8(a)所示，G 越大其阻塞率越小。在 $G = 16$ 的条件下比较 $k = 4,6,8$ 时的多播阻塞率，如图 8(b)所示，k 越大其阻塞率越大，但 k 不同时，其阻塞率区别不明显。2 种条件下，多播阻塞率小于等于 10^{-2}。

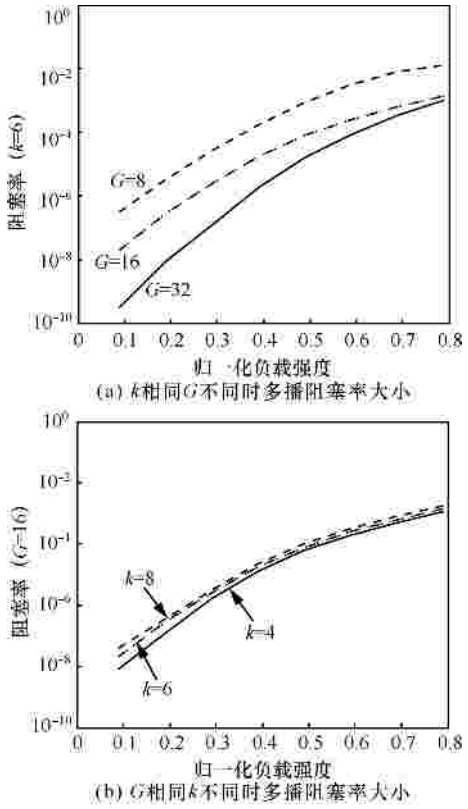


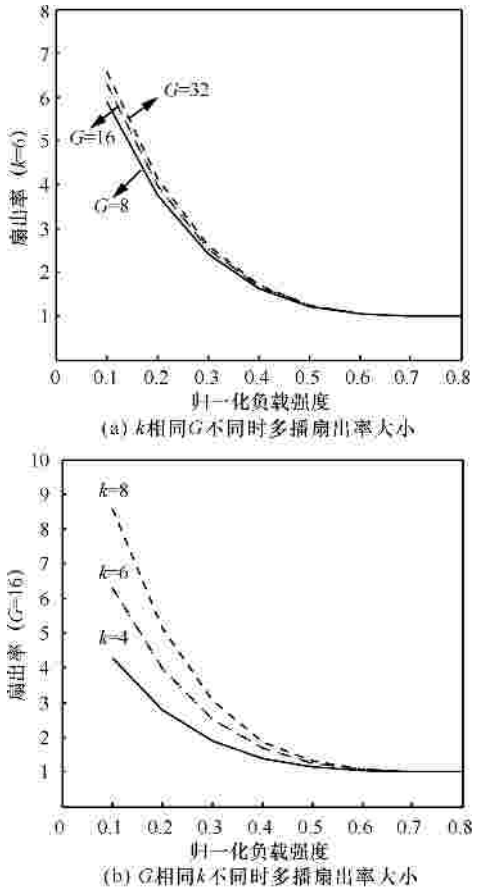
图 8　多播阻塞率

原因分析：k 决定交换网络的级数，在 k 不变的情况下，单播负载强度 $p_1 = 0.2$，不同的 G 时，由于级数相同，单播负载强度相同，其单播分组对多播的影响基本相同。但 G 越大，其 $x \le G$ 的概率越大，从而减少了 $x > G$ 时分组在布尔集线器中的内部阻塞概率。当 G 一定时，k 越大，交换结构的级数越大，但由于多播的扇出特性，当级数大于等于 4 时，负载 $p_2 \ge 0.2$ 时，其 4 次扇出的多播分组个数就接近 $G - 0.2G = 0.8G$，后几级的多播扇出会很小，而对于式(2)，需要做模 $(G - Gp_1)$ 运算，所以 $k = 4,6,8$ 时，其阻塞率区别不明显。

3) 多播扇出率仿真

设定单播负载强度 $p_1 = 0.2$，多播负载强度

$0 \le p_2 \le 1 - p_1$，在 $k = 6$ 的条件下，比较 $G = 8,16,32$ 时的多播扇出率，如图 9(a)所示，归一化负载强度为 0.1 时，多播扇出率接近 7，G 越大其扇出率越大，多播归一化负载强度接近 0.8 时，其扇出率都趋近于 1，但 G 不同时，其多播扇出率区别不明显。在 $G = 16$ 的条件下，比较 $k = 4,6,8$ 时的多播扇出率，如图 9(b)所示，k 越大其多播扇出率越大。



(a) k相同G不同时多播扇出率大小

(b) G相同k不同时多播扇出率大小

图 9 多播扇出率

原因分析：k 决定交换网络的级数，在 k 不变的情况下，单播负载强度 $p_1 = 0.2$，不同的 G 时，由于级数相同，单播负载强度相同，其单播分组对多播的影响基本相同。G 不同，但多播分组的输入分组个数和输出分组个数同等比例增加，所以扇出率变化不明显。当 G 变大时，$x \le G$ 的概率变大，从而减少了 $x > G$ 的概率，减少了分组在布尔集线器中的内部阻塞概率，使式(3)分子变大，所以扇出率变大。当 G 一定时，k 越大，交换结构的级数越大，由于多播的扇出特性，当级数变大时，多播就会有更多的扇出分组，$k = 8$ 比 $k = 6$ 时多了两级扇出，$k = 6$ 比 $k = 4$ 时多了两级扇出，所以随着 k 的变大，其扇出率明显变大。当多播负载非常小时，其多播可以完全扇出，其扇出率正比于 $k+1$，所以当归一化负载强度为 0.1 时，$k = 8,6,4$ 时，多播扇出率分别接近于 9,7,5。

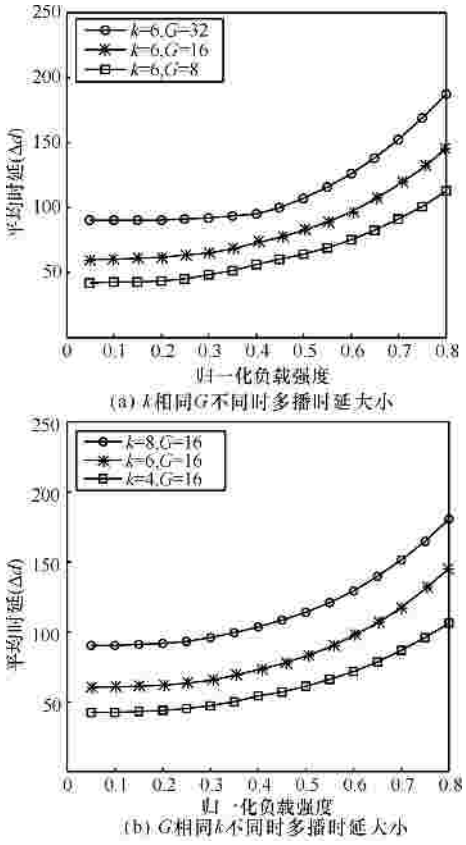
4) 多播时延仿真

时延分为分组平均时延 D_a 和分组最大时延 D_m。若交换系统在一定时间内交换的分组数为 n，d_i 代表第 $i, 1 \le i \le n$ 个分组的时延，那么 D_a 和 D_m 描述为

$$D_a = \frac{1}{n}\sum_{i=1}^{n} d_i, D_m = \max\left(d_1, d_2, \mathrm{L}, d_n\right)$$

分组的时延主要由分组经过 2×2 布尔单元个数决定，假设分组每经过一个 2×2 布尔单元的时延为 Δd，一般在 ns 级。

在 $k = 6$ 的条件下比较 $G = 8,16,32$ 时的时延，如图 10(a)所示，G 越小时延越小。在负载强度小于等于 0.5 时，其时延增长速度 T_v 的关系为 $T_{v(G=32)} < T_{v(G=16)} < T_{v(G=8)}$，在负载强度 > 0.5 时，其时延增长速度 T_v 的关系为 $T_{v(G=32)} > T_{v(G=16)} > T_{v(G=8)}$。在 $G = 16$ 的条件下比较 $k = 8,6,4$ 时的时延，如图 10(b)所示，k 越小其时延越小。且负载强度在 $(0, 0.8]$ 范围内时，k 不同，时延增长速度 T_v 变化不明显。



(a) k相同G不同时多播时延大小

(b) G相同k不同时多播时延大小

图 10 多播时延

原因分析：在阻塞率相同的情况下，时延主要受 2×2 布尔单元级数的影响，该级数由集线器中 2×2 布尔单元的级数和 banyan 类网络级数相乘得到。G 不变，G 越大，集线器中 2×2 布尔单元的级数越大，2×2 布尔单元的总级数越大，所以时延变大。对于图 10(a)，当负载强度小于等于 0.5 时，由于 $G=32$ 的阻塞率远远小于 $G=16, G=8$，因此 $T_{v(G=32)}$ 最小。当负载强度大于 0.5 时，由于 $G=32$ 的阻塞率与 $G=16, G=8$ 在相同的数量级，但 $G=32$ 时 2×2 排序器级数最大，所以 $T_{v(G=32)}$ 最大。G 不变，k 越大，2×2 交换单元的总级数越大，时延越大，而由图 8(b)可知，k 不同时，其阻塞率变化不明显，因此在图 10(b)中时延增长速度 T_v 无明显变化。不论何种 G,k 参数的交换结构，其时延总小于某个定值。例如：当 $G=32, k=6$ 时，时延小于等于 $D_{\max}$，约等于 $187.5\Delta t$，当 $G=16, k=8$ 时，时延小于等于 $D_{\max}$，约等于 $180\Delta t$。

不同的业务对带宽、丢失率、时延、抖动等 QoS 需求是不同的，如电子邮件对实时性要求低，但对丢失率要求高，丢失后必须重传；实时流媒体业务对实时性要求高，一定丢失率下不影响服务质量。该单多播混合输入部分扇出多播交换模型，当单播负载强度控制在一定范围内时，如 $p_1 \le 0.2$ 时，其多播阻塞率小于等于 10^{-2}，所以该交换结构可以满足某些网络场景的应用，其多播时延总小于某个定值，且该定值在百纳秒量级，所以该交换结构能够为到达业务提供时延上限保障。

5　结束语

本文的阻塞率模型是部分扇出多播模型，由于单播抢占或阻塞，在单个交换节点扇出率小的多播分组可在下一个交换节点中再次进行部分扇出。本文假设多播的带内控制信息为全 B，即对所有输出端口均有扇出，该假设忽略了多播对输出端口集的限制，下一步工作，基于可控多播输出端口集进行多播扇出模型的建立，该模型需要考虑每一级是否扇出的问题。

参考文献：

[1] LABOVITZ C, IEKEL-JOHNSON S, MCPHERSO D. Internet inter-domain traffic[A]. Proc of the ACM SIGCOMM[C]. New Delhi, India, 2010, 75-86.

[2] 中国互联网络信息中心. 中国互联网络发展状况统计报告[EB/OL]. http://www.cnnic.net.cn/dtygg/dtgg/201201/t20120116_23667.html, 2011.
China Internet Network Information Center. Statistical China Internet network development status[EB/OL].http://www.cnnic.net.cn/dtygg/dtgg/201201/t20120116_23667.html, 2011.

[3] HUANG A, KNAUER S. Starlite: a wideband digital switch[A]. Proc of Globecom[C]. Atlanta, GA, 1984. 121-125.

[4] DEERING S, CHERITON D. Multicast routing in datagram internetworks and extended LANs[J]. ACM Transactions on Computer Systems, 1990, 8(2):85-110.

[5] ERIKSSON H. Mbone: the multicast backbone[J]. Communication of the ACM, 1994, 37(8):54-55.

[6] LEE T T. Non-blocking copy networks for multicast packet switching[J]. IEEE Journal on Selected Areas in Communications, 1988, 6(9): 1445-1467.

[7] GUO M H, CHANG R S. Multicast ATM switches based on input cells scheduling[A]. Proc of APCC[C]. Sydney, Australia, 1997. 1629-1632.

[8] YEH Y S, HLUCHYTJ M G, ACAMPORA A. The knockout switch: asimple, modular architecture for high-performance packet switching[J]. IEEE Journal on Selected Areas in Communications, 1987, 5(8): 1274-1283.

[9] YANG Y, MASSON G M. Broadcast ring sandwich networks[J]. IEEE Trans Computers, 1995, 44(10):1169-1180.

[10] JASTRZEBSKI A, KUBALE M. Rearrangeability in multicast clos networks is np-complete[A]. 2010 2nd International Conference on Information Technology (ICIT)[C]. 2010. 183-186.

[11] PAN D, YANG Y. FIFO-based multicast scheduling algorithm for virtual output queued packet switches[J]. IEEE Transactions on Computers, 2005, 54(10):1283-1297.

[12] 胡曦明. 有效支持 IPTV 业务的扇出无关交换结构[D]. 郑州: 信息工程大学, 2007. 29-49.
HU X M. Fanout-free Switch Fabric to Support IPTV Business[D]. Zhengzhou: Infromation Engeering University, 2007. 29-49.

[13] HU H C, LIN P, GUO Y F. Integrated uni- and multicast traffic scheduling in buffered crossbar switches[A]. 3rd Intentional Conference on Communications and Networking in China (ChinaCOM'08)[C]. Hangzhou, China, 2008. 66-72.

[14] 兰巨龙. 可重构信息通信基础网络体系研究[R]. 国家重点基础研究发展计划（973 计划）项目计划任务书, 2011.
LAN J L. Research on Foundition Network System of Reconfigurable Information Communication[R]. Planning Task Document of National Basic Research Program of China(973 Program), 2011.

[15] 刘莹, 徐恪. Internet 多播体系结构[M]. 北京：科学出版社, 2008.
LIU Y, XU K. Internet Multicast Architecture[M].Beijing: Science Press, 2008.

[16] LI S Y R. Unified algebraic theory of sorting, routing, multicasting, and concentration networks[J]. IEEE Transactions on Communications, 2010, 58(1):247-256.

[17] LI S Y R. Algebraic Switching Theory and Broadband Applications[M]. San Diego, CA, USA: Academic Press, 2001. 281-321.

（下转第 65 页）

dominating set based on serial maximum independent set[J]. Journal of Huazhong Science and Technology (Natuaral Science Edition), 2011,39(3):61-65.

[12] **郑杰，郭淑杰，屈玉贵等. 无线传感器网络能效时延平衡数据收集机制**[J].**中国科学技术大学学报**,2008, 38(12):1414-1421.
ZHENG J, GUO S J, QU Y G, *et al*. Energy efficiency and delay balancing data gathering for wireless sensor networks[J]. Journal of University of Science and Technology of China, 2008, 38(12):1414-1421.

[13] **郜帅，张宏科. 时延受限传感器网络移动 Sink 路径选择方法研究**[J]. **电子学报**,2011, 39(4):742-747.
GAO S, ZHANG H K. Optimal path selection for mobile sink in delay-guaranteed sensor networks[J]. Acta Electronica Sinica, 2011, 39(4):742-747.

[14] SAMIR K, BALAJI R, NEAL E. Young balancing minimum spanning trees and shortest path trees[J]. Algorithm, 1994, 120(4):305-321.

[15] LI Y S, THAI M T, WU F. On the construction of a strongly connected broadcast arborescence with bounded transmission delay[J]. IEEE Transactions on Mobile Computing, 2006, 5(10):1460-1470.

[16] **孙彦景，钱建生，顾相平等.联合约束无线传感器网络连通支配集算法**[J].**电子科技大学学报，**2009，38(2)：231-235.
SUN Y J, QIAN J S, GU X P, *et al*. Distributed connected dominating set algorithm with combined constraints in wireless sensor network[J]. Journal of University of Electronic Science and Technology of China, 2009，38(2)：231-235.

[17] LI Y, CHENG M X, DU D Z. Optimal topology control for balanced energy consumption in wireless networks[J]. Journal Parallel and Distributed Computing, 2005, 65(2): 124-131.

[18] WAN P J, KHALED M A, OPHIR F. Distributed construction of connected dominating sets in wireless ad hoc networks[J]. ACM/Kluwer Mobile Networks and Applications, 2004, 9(2): 141- 149.

作者简介：

孙彦景（1977-），男，山东滕州人，博士，中国矿业大学教授、博士生导师，主要研究方向为无线传感器网络和嵌入式实时系统。

钱建生（1964-），男，浙江桐乡人，博士，中国矿业大学教授、博士生导师，主要研究方向为矿山通信和无线传感器网络。

马姗姗（1978-），女，河南焦作人，中国矿业大学博士生、讲师，主要研究方向为无线传感器网络和信息处理。

任鹏（1985-），男，江苏徐州人，中国矿业大学博士生，主要研究方向为无线传感器网络和无线认知网络。

（上接第 57 页）

[18] **李挥，何伟，伊鹏. 排序集线器多级互连交换结构的多路径自路由模型**[J]. **电子学报**, 2008, 36(1):1-8.
LI H, HE W, YI P, *et al*. Modeling multi-path self-routing switching structure from multistage interconnection of sorting concentrators[J]. Acta Electronica Sinica, 2008, 36(1):1-8.

作者简介：

张博（1982-），男，河北张北人，国家数字交换系统工程技术研究中心博士生，主要研究方向为可重构路由理论与技术、可重构交换理论与技术。

汪斌强（1963-），男，安徽桐城人，国家数字交换系统工程技术研究中心教授、博士生导师，主要研究方向为宽带信息网络和高速路由器核心技术。

朱圣平（1968-），男，浙江东阳人，国家数字交换系统工程技术研究中心工程师，主要研究方向为可重构网络理论与技术。