

面向混合网络的半覆盖组播路由协议*

林彤^{1,2,3+}, 钱华林¹, 葛敬国¹, 牛广锋¹

¹(中国科学院 计算机网络信息中心,北京 100080)

²(中国科学院 计算技术研究所,北京 100080)

³(中国科学院 研究生院,北京 100049)

Half Overlay Multicast Routing for Hybrid Networks

LIN Tong^{1,2,3+}, QIAN Hua-Lin¹, GE Jing-Guo¹, NIU Guang-Feng¹

¹(Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100080, China)

²(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China)

³(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: Phn: +86-10-58812328, Fax: +86-10-58812306, E-mail: lintong@cnic.cn

Lin T, Qian HL, Ge JG, Niu GF. Half overlay multicast routing for hybrid networks. Journal of Software, 2007,18(10):2572-2583. <http://www.jos.org.cn/1000-9825/18/2572.htm>

Abstract: Although multicast can be implemented in every layer of network, the existing multicast protocols rarely meet the demands both for flexibility and efficiency at one time. In general, hardware multicast and IP multicast are more efficient than overlay multicast, but the situation of flexibility is quite the contrary. This problem is much pressing in hybrid networks, in which the two demands have the same priority. This paper proposes a novel protocol named half overlay multicast routing that integrates IP layer in-group regional broadcast, address-and-port-translation based overlay multicast and extensible support for hardware multicast into one single model, where the three mechanisms are chosen dynamically for data delivery. By this approach, HOMR obtains the flexibility of overlay multicast. The simulation results show that the overhead of HOMR is low and the performance of HOMR is comparable with pure IP multicasting.

Key words: multicast; routing protocol; hybrid network; half overlay; self-organizing

摘要: 组播已在多个协议层上分别实现,但现有协议仍未兼顾对灵活性与效率的共同需求.总的来说,硬件组播和 IP 组播较为高效,而覆盖组播则更为灵活.在混合网络中,底层技术的差异使这一问题变得更为突出.为此提出了一种动态自组织的半覆盖组播路由协议 HOMR(half overlay multicast routing).该协议把基于组内区域广播的 IP 组播、基于地址端口转换的覆盖组播以及对各类硬件组播的调用融入单一模型,在各层组播间按需转换.HOMR 具有与全覆盖组播相似的灵活性,模拟实验表明,HOMR 在保持较低协议开销的同时,获得了与完全 IP 组播相近的转发效率.

关键词: 组播;路由协议;混合网络;半覆盖;自组织

中图法分类号: TP393 文献标识码: A

* Supported by the Special Foundation of President of the Chinese Academy of Sciences under Grant No.KGCX2-YW-106 (中国科学院院长特别基金)

Received 2006-07-18; Accepted 2006-11-13

近年来,组播技术在3个主要方向上得到持续的发展:(1) 为了克服传统IP组播可扩展性较差、难以提供传输控制、需要全网部署等不足,人们提出了大量在应用层或传输层实现、完全依靠单播路由的覆盖组播*方案^[1,2];(2) 随着无线通信和自组织网络的发展,出现了多种Ad-hoc组播协议^[3],为适应链路窄带和物理拓扑的频繁变化,其中部分协议引入了按需触发(on-demand)、网状连接(mesh-based)等行之有效的路由策略;(3) 更多子网技术提供了高效的硬件组播支持,其中,以太网、ATM网等的硬件组播已非常成熟,WDM光组播^[4]、WLAN的MAC层组播^[5]等也都有相当多的提案出现。

然而,新兴的组播技术仍存在各自的局限性.覆盖组播对IP组播的改进是以性能为代价的,表现在组播逻辑拓扑与网络物理拓扑脱节、难以抑制链路重复报文、无法调用硬件组播支持等方面.Ad-hoc组播仅根据MANET的特点进行设计和优化,既不适用于包含基站、访问点等固定基础设施的主从式无线网,也不会被部署于有线网络.链路级的硬件组播则无法独立跨越异构子网的边界。

显然,各类组播技术间存在多方面的互补性,但现有研究还未能将它们各自的优势有效地结合起来.为此,本文提出了一种动态自组织的半覆盖组播路由协议 HOMR(half overlay multicast routing),它在单一的模型和协议框架内,把按接口适配的硬件组播、基于组内区域广播的IP层组播和通过地址端口转换实现的覆盖组播有机地融合在一起,其基本思想是优先选用底层组播机制,仅当下层机制不可用时,才借助上层投递方式进行“粘接”,从而尽可能地在性能与灵活性之间取得平衡。

我们首先将HOMR应用于混合网络环境.在本文中,混合网络是指具有多种底层通信技术,且在所有节点上都支持TCP/IP协议栈的接入网架构,一个典型的混合网络如图1所示.混合网络的特点包括:1) 在同一空间内,重叠覆盖有多个异构的有线或无线子网,节点拥有多种网络接口,可以并发访问与之相连的数个网;2) 端设备既可以连接固定的基础设施,形成主从式拓扑,也可以在无中心管理的情况下,相互连接形成对等的自组织网络,网络拓扑随时可能发生变化;3) 存在多个相互独立的管理区域,各个管理区域使用的单播选路算法可能互不相同;4) 接入子网既可能支持多路访问(如以太网、WLAN等),也可能只允许点到点连接(如帧中继、ADSL等),不同子网技术对硬件组播支持的力度和方式存在较大差异。

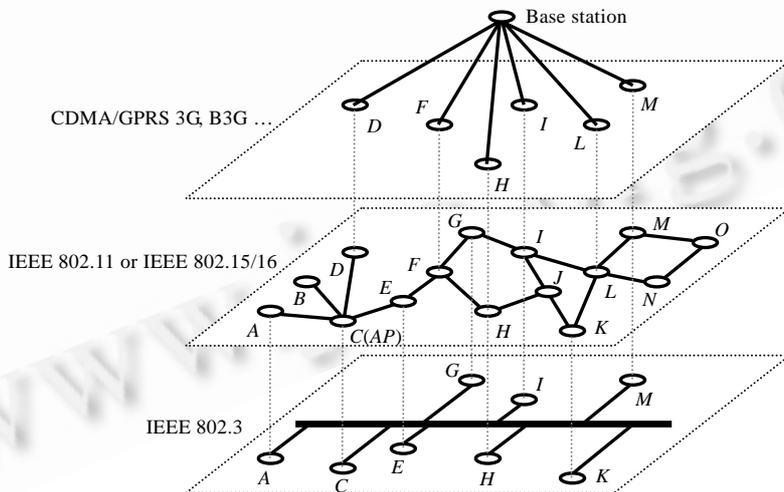


Fig.1 A case for hybrid access network

图1 一个混合接入网络的实例

* 与覆盖组播(overlay multicast,简称OM)相关的另一常用名词是应用层组播(application layer multicast,简称ALM),对于两者的关系存在不同观点:有学者认为它们是完全等同的概念^[1],也有学者把引入了转发代理的应用层组播称为覆盖组播^[2].本文则把应用层组播归入覆盖组播的范畴。

由于目前没有任何一种有线或无线技术能够在实施中取得压倒性的优势,因此,混合网络很可能会作为主流的网络形态长期存在,其复杂的结构对组播路由提出了苛刻的要求,例如,不稳定的物理拓扑要求逻辑拓扑更快地反映物理网络的现状,稀缺的带宽资源要求对信道的利用更充分,网络的异构特性要求协议的适用范围更广等等,这使得现有技术的局限性显得尤为突出.HOMR 的设计充分考虑了混合网络的特点.例如,使用组播区域代替组播树、使用地址端口转换代替组播隧道、使用适配协议访问硬件组播等等.

本文第 1 节介绍已有的相关工作,第 2 节详细说明 HOMR 协议的核心策略和主要算法,第 3 节给出对 HOMR 的仿真和分析结果,第 4 节对全文做出总结,并指出进一步的工作.

1 相关工作

在联合使用多种组播技术方面已有一些成果出现^[6-9].子网组播^[6]在子网内使用硬件组播,在子网间使用IP单播,弃用了难以部署的IP组播.HMTP(host multicast tree protocol)^[7]是一种能够连接IP组播孤岛的覆盖组播协议,它在每个孤岛内选举一个代理,然后构造一棵连接所有代理的覆盖共享树,通过UDP(user datagram protocol)隧道在代理间转发报文.孤岛组播^[8]与HMTP的基本想法相同,但在构造覆盖拓扑的方式上存在区别.文献[9]讨论了在Mobile IP模型下,节点移动对组播路由的影响,认为节点移动会对应用层组播产生严重影响,但对IP组播影响不大.

上述方案都只面向拓扑相对稳定的广域网环境,并不适用于混合网络:(1) 它们所使用的IP组播协议和覆盖组播协议是相互独立的,仅仅扩展了协议间的交互机制,这就需要在部分节点上部署至少两种组播路由协议.对于来源多样、存储和计算能力有限的个人设备来说,这种做法并不现实;(2) 它们所假定的子网和IP组播域的边界都是固定的,没有考虑组播域的合并和分片等情况;(3) 在移动性方面,Mobile IP与自组织网络也截然不同:Mobile IP组播^[10]所考虑的是在有固定网络和代理机制支持下,部分主机在域间的移动;而自组织网络是无管理的,整个拓扑随时都会发生改变,没有固定的参考方位可供利用.

在解决移动性问题上,Ad-hoc组播更接近混合网络的需要.然而,现有Ad-hoc组播协议都可明确归为IP组播^[3]或覆盖组播^[11,12],还没有看到两种方式混合路由的案例出现,因此,它们仍存在IP组播和覆盖组播各自的不足.尽管如此,由于Ad-hoc组播路由协议在自组织网络中有很好的表现,本文在仿真部分将以它们为参照,来对比验证HOMR协议的各项性能指标.

综上所述,目前旨在融合各类组播优势的研究仍处于较为初步的阶段,表现为:1) 不能在单一协议中同时具备覆盖转发和非覆盖转发的能力;2) 没有针对移动自组织网络的此方面论述出现;3) 没有充分考虑在子网技术多样化的情况下,对硬件组播的扩展问题.

2 HOMR-HN

HOMR 协议是以半覆盖拓扑(half overlay topology,简称HOT)模型为基础的.本节首先介绍HOT模型的定义,然后再详细介绍HOMR协议对HOT的实现.

2.1 HOT模型

在HOT模型中,整个网络表示为 $N=(V,E)$,其中, V 为节点集合, E 为物理链路集合.设 G 为组播组集合,对任意 $g \in G$,用 V_g 代表组 g 的成员集合, $V_g \subseteq V$, E_g 则为连接 V_g 中各节点的链路集合, $E_g \subseteq E$.

V 中的节点都必须支持IP协议,每个节点有一个全局唯一的标识(ID).一般可以从分配给节点各个网络接口的IP地址中选择一个作为节点ID.

设集合 $V_p \subseteq V_g$,如果对 $(\forall v_i)(\forall v_j)(v_i \in V_p \wedge v_j \in V_p)$,都能做到仅由链路协议即可完成 v_i 与 v_j 之间的组播数据投递,则 (V_p, E_p) 构成一个物理组播域(physical multicast region,简称PMR),其中, E_p 为 V_p 中节点间链路集合.

设集合 $V_n \subseteq V_g$,若对 $(\forall v_s)(\forall v_d)(v_s \in V_n \wedge v_d \in V_n)$,都存在路径 $(v_s, \eta_0, \dots, \eta_k, v_d)$,使得对路径中 $\forall \eta_i(i=0, \dots, k)$,都满足 $\eta_i \in V_n$;且要在 V_n 中完成组播,需要且至多仅需网络层机制在中间节点上参与转发至少一次,则 (V_n, E_n) 构成一个网络组播域(network multicast region,简称NMR),其中, E_n 为 V_n 中节点间链路集, $E_n \subseteq E_g$.

对节点集 V' , 若存在路径 $p(v_i, \dots, v_j, \eta_0, \dots, \eta_k, v_m, \dots, v_n)$, 使得 $v_j \in V', v_m \in V'$, 但对其中 $\forall \eta_h (h=0, \dots, k)$, 有 $\eta_h \notin V'$, 则 $p'(v_j, \eta_0, \dots, \eta_k, v_m)$ 构成一条相对 V' 的覆盖链路.

设集合 $V_o \subseteq V_g$, 若 $(\exists v_s)(\exists v_d)(v_s \in V_o \wedge v_d \in V_o)$, 使 v_s 与 v_d 之间的任意路径均包含至少一条相对 V_g 的覆盖链路, 则 (V_o, E_o) 构成一个覆盖组播域 (overlay multicast region, 简称 OMR), 其中 E_o 为 V_o 中节点间链路集合 (包括构成覆盖链路的物理链路). 一般地, $OMR_g = (V_o, E_o) = (V_g, E_g)$, 即 OMR_g 等价于 g 的 HOT 组播拓扑.

每个 PMR 和 NMR 都有一个指派节点 (designated node, 简称 DN), 分别缩写为 PDN 和 NDN.

2.2 混合网络的HOT表示

2.2.1 PMR 的形成

根据HOT定义, 设 $PMR_g = (V_p, E_p)$, V_p 内节点属于同一个支持链路层组播的子网. 在进行数据分发时, 只要把报文封装到以组 g 的二层组地址为目的地址的数据帧中, 就可以通过硬件组播送达 PMR 内的所有成员. PMR 是按接口定义的, 当一个节点有多个网络接口时, 它可以同时属于多个 PMR.

在图 1 中, 同组节点 A, C, E, H 各有一个以太网接口, 且属于同一碰撞域, 由于共享式以太网支持硬件组播, 所以, 这 4 个节点及其间的 802.3 链路构成了一个 PMR, 如图 2 所示. 对于交换式以太网, 如果交换机支持硬件组播^[13], 那么, 即使节点不属于同一碰撞域, 也能并入同一 PMR.

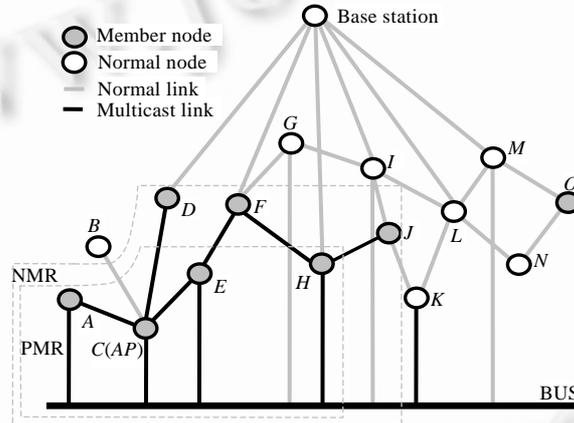


Fig.2 A HOT presentation of hybrid network

图 2 按半覆盖模型表示的混合网络示例

在无线子网中构造PMR会受到信号传输距离的限制. 在图 3(a)中, 3 个 WLAN 节点两两相邻, 能够形成一个 PMR; 但在图 3(b)中, A 到 C 的分组需通过 B 转发, 如果转发不能仅在链路层完成, 那么 A 和 C 就不属于同一 PMR. 设 $RXThresh_{min}$ 为子网传播模型所限定的、在各个传播方向上传输距离阈值中最小的那个, 对 $(\forall v_i)(\forall v_j) (v_i \in V_g \wedge v_j \in V_g)$, 设 $dist_{max}(i, j)$ 为 v_i 与 v_j 所能移动的最大间距. 为使问题简化, 在无线子网中构造 PMR, 需满足如下附加约束: $\max\{dist_{max}(i, j) | 1 \leq i \leq |V_g|, 1 \leq j \leq |V_g|\} < RXThresh_{min}$, 其中, $|V_g|$ 为组 g 的成员个数.

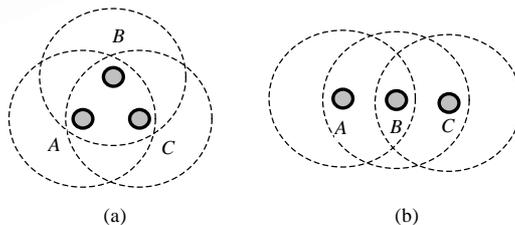


Fig.3 The influence of radio range on HOT construction

图 3 无线信号传输范围对 HOT 结构的影响

2.2.2 NMR 的形成

NMR_g 代表了相对于组 g 的、网络层组播可达的连续区域.NMR以节点为最小构造单位,一个节点在任何时刻都只能属于一个NMR.要构成NMR,除了需要组播协议的连续部署以外,还要求相邻节点必须是同组成员,这与“IP组播孤岛”的概念是不一样的.从网络层的角度看,PMR内的任意两个节点都互为同组邻居,因此,PMR能够成为NMR的组成部分(如图2所示),反之不然.构造PMR要求“链路组播相邻关系”的可传递性,而构造NMR无此要求.无法构造PMR,但能构造NMR的典型情况有:1) 转发路径包含点到点链路;2) 转发路径穿越了异构子网,例如图1中的路径(A,C,D);3) 传输距离受限的无线网络(如图3(b)所示).

当两个同组成员在物理拓扑中互为邻居时,它们至少可以建立NMR邻居关系.如果符合构造PMR的条件,就可以进一步建立PMR邻居关系,判断条件都是事先可以获得的.

2.2.3 OMR 的形成

一个典型的OMR是由多个NMR或PMR通过覆盖链路相连而成的.在覆盖链路上的分组转发,需借助于单播路由协议.根据定义,当全网单播可达时, OMR_g 必定包含了组 g 的全部成员和成员间的全部物理链路.在图2中,除了节点 B 和链路(B,C)外,整个区域构成了一个OMR,其中,节点 O 的邻居中没有同组成员,它与左侧NMR之间有多条覆盖链路相连.类似节点 O 的孤立点可以看作是最小规模的PMR及NMR.

2.2.4 DN 的作用

每个PMR和NMR都会选举一个DN.一种极端情况是PMR与NMR完全重叠,此时,在同一域内会分别选举一个PDN和一个NDN.PDN的唯一作用是作为PMR的标识节点,PDN的ID即为PMR的ID.NDN除了标识NMR的作用外,还作为NMR的代表节点与其他NDN交换信息,并进行NMR内外的组播数据转发.NMN间的覆盖链路实际上是NDN间的覆盖链路.

2.3 HELLO报文

HELLO机制是HOMR协议建立和维护HOT的最主要手段,每个组成员都会周期性地向自己的邻居节点发送HELLO报文.一个特例是,通过覆盖链路相连的NDN间也有多跳的HELLO报文交换,两个NDN互为覆盖网邻居(OMR peer).HELLO报文的作用有:1) 邻居发现和邻居关系的保持;2) 节点参数以及群组状态的通告和更新;3) 确保邻居间的双向通信;4) 触发对DN的选举.

HELLO报文包含的信息有:1) 群组ID;2) 节点ID;3) 节点类型(PDN,NDN或普通节点);4) 节点当前接口的PMR优先级;5) 节点的NMR优先级;6) 报文流水号;7) 若非PDN节点,包含节点当前接口所属PMR的PDN信息,其中包括PDN的ID、接口IP地址和PMR优先级;8) 若非NDN节点,包含节点所属NMR的NDN信息,其中包括NDN的ID、NMR优先级以及到NDN的路径矢量;9) 若是NDN节点,包含节点的OMR peer信息,包括各Peer的ID、接口IP地址和类型属性.

每个组在创建时都会获得一个全局唯一的ID,HOMR并不限定群组ID的产生方式,如果不与其他IP组播协议相冲突,可以直接使用IP组播地址.为了选举PDN和NDN,每个节点都有一个NMR优先级,节点的每个接口有一个PMR优先级.对于单接口节点,节点ID与接口IP地址是相同的,但对多接口节点,两者可能不同.为避免收到陈旧报文,每个HELLO报文都有一个流水号,信宿根据流水号判断报文是否过期.

2.4 优先级的计算

HOMR采用优先级机制来选举DN.选举DN应尽量同时满足两个目标:1) DN须保持相对稳定;2) DN应具有较强的网络计算能力.为此,HOMR引入了多个参数用于优先级计算.

A) 固定参数 $param_{fix}$

固定参数反映了节点的计算能力和网络通信能力,该参数在协议初始化时即可算得,在运行期间不改变.

$$param_{fix} = \alpha_1 \times W + \beta_1 \times H + \gamma_1 \times M,$$

其中, W 为节点各接口带宽之和, H 为节点CPU主频总数, M 为内存总量, $\alpha_1, \beta_1, \gamma_1$ 为加权系数.

B) 动态参数 $param_{dyn}$

设 δ_{join} 为当前时刻与节点加入群组时刻之差, δ_{DN} 为当前时刻与节点成为DN时刻之差,有

$$param_{dyn} = \alpha_2 \times \delta_{join} + \beta_2 \times \delta_{DN},$$

其中, $\delta_{join} \geq 0, \delta_{DN} \geq 0, \alpha_2, \beta_2$ 为加权系数.

C) 优先级计算

HOMR 通过设置计时器,定期触发动态参数和优先级的计算,对 PMR 和 NMR 优先级的计算方法相同,公式为

$$prior = |\alpha_3 \times param_{fix} + \beta_3 \times param_{dyn}|,$$

其中, α_3, β_3 为加权系数.

2.5 PDN的选举、PMR的建立和保持

在开始时,每个加入组播的节点都建立一个以自己为 PDN 的 PMR,并在子网内广播 HELLO 报文.如果子网满足建立 PMR 的条件,那么,该报文能被子网内的其他所有节点收到.设节点 A 收到同组邻居 B 发来的 HELLO 报文,且报文声明 B 是一个 PDN,那么,A 会以自己 PDN 的优先级与 B 的优先级比较,选择优先级高的作为自己的 PDN,如果优先级相同,则选择节点 ID 大的作为 PDN.子网内所有同组成员都以相同的标准进行选举,经过一轮 HELLO 报文的交换过程后,各个成员都能获得 PMR 邻居的完整列表,并选举同一个节点作为 PDN.

节点为自己的每个 PMR 邻居设置一个超时值,在收到 HELLO 报文时,该值被更新.如果某邻居的超时值过期,则节点认为该邻居已经失效或已退出组,于是从自己的 PMR 邻居列表中删除它.如果 PDN 过期,将触发 PDN 的重新选举,如果子网内没有同组邻居,那么节点会选举自己为 PDN,并继续广播 HELLO 报文.那些组外节点会由此获得可加入的群组信息.当 PMR 内最后一个节点退出时,PMR 将自行消失.

2.6 NMR管理

NMR 模块是 HOMR 协议的核心组件,它在联系 PMR 所代表的硬件组播和 OMR 所包含的覆盖组播过程中,发挥“桥梁”的作用.

2.6.1 NMR 的归并

与 PMR 类似,节点建立 NMR 并初始化为 NDN.当节点在邻居中发现同组成员时,就会触发 NMR 的归并和 NDN 的选举.图 4 展现了引发 NMR 归并的两种典型情况:节点加入和无线节点移动.

在图 4(a)中,成员节点 1 逐渐移近,非成员节点 2 加入组播,这两个节点状态的变化最终使整个网络形成单一的 NMR(如图 4(b)所示).一般地,当节点 A 收到同组邻居 B 的 HELLO 报文,且 A 和 B 的 NDN 不一致时,A 就会选择两个 NDN 中优先级高的作为自己的 NDN,若优先级相等,则选择节点 ID 小的作为 NDN.之所以令 PDN 和 NDN 的选择条件不同,是为了使 DN 尽量分散.当 A 的 NDN 改变时,它会立即将新 NDN 的信息通知自己的其他邻居,通过消息在原 NMR 内的扩散,最终使归并在一起的各个 NMR 都选举相同的 NDN.

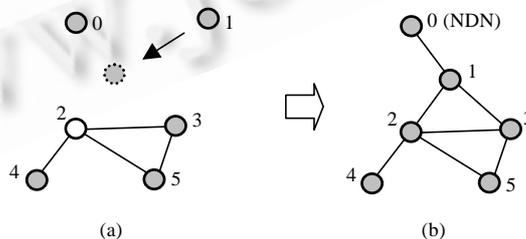


Fig.4 A merge of NMRs triggered by the change of network topology

图 4 由网络拓扑改变引发的 NMR 归并

2.6.2 路径矢量(path vector,简称 PV)算法

与 PDN 不同,NDN 通常只与 NMR 中的少数节点相邻,与 NDN 不相邻的节点只能从中间节点处间接获得 NDN 的信息,因此,非 NDN 节点仅仅知道 NDN 的 ID 和优先级是不够的.首先,当节点发现自己的某个邻居失效

时,可能意味着与 NDN 的联系也同时中断,是否真的中断需要额外的信息才能判定;第二,当邻居提供的 NDN 信息不一致时,如何采信也需要判断,例如在图 4(b)中,节点 2 能够收到 4 个邻居发来的关于 NDN 的信息,显然,它应该完全相信节点 1,而忽略来自节点 4 的情报.

HOMR 使用路径矢量算法来维护 NMR 的连通性,其基本思想是,相邻节点交换各自保存的、到 NDN 的严格路径,路径中包含从邻居到 NDN 沿途的所有节点 ID,格式为(NDN 邻居 ID,...,本节点邻居 ID).在收到来自邻居的关于 NDN 的路径后,节点从中选择一条有效的最短路径予以保存.

算法 1. 路径矢量的过滤更新算法.

输入:当前节点*i*的路径矢量表 T_i ,邻居节点*j*的路径矢量表 T_j ,节点*i*的邻居ID表 N_i ;

输出:更新后的 T_i .

- (1) 初始化: $P' \leftarrow$ 空路径;删除 T_i 中包含*j*的所有路径;
- (2) **for** T_j 中的每条路径 P **do**
- (3) **if** P 中包含 i **then continue;**
- (4) **if** P 中包含 N_i 中的元素 **then**
- (5) $P \leftarrow$ 原 P 中从起始ID到首次出现属于 N_i 的ID(含)为止的部分;
- (6) **if** T_i 中包含 P **then continue;**
- (7) **if** P' 非空 **and** $length(P') \leq length(P)$ **then continue;**
- (8) 从 T_i 中删除 P' ;把 P 存入 T_i ; $P' \leftarrow P$.

仍以图 4(b)为例,各个节点处路径矢量表的一种稳定状态为

0: {} 1: {} 2: {(1)} 3: {(1)} 4: {(1, 2)} 5: {(1, 2), (1, 3)}.

节点 4 到 NDN 的唯一路径经过节点 2,节点 3 的路径则经过了节点 2 的邻居节点 1,在这两种情况下,节点 2 都不会接受它们所提供的关于 NDN 的信息.

2.6.3 NMR 的分片

NMR 分片是归并的逆过程,当发生链路失效或成员退出等情况时,就可能产生 NMR 的分片.断裂点附近的成员会最先发现情况的异常,设相邻成员 A, B 都不是 NDN,节点 B 失效, A 发现后先从自己的 PV 表中删除所有包含 B 的路径,然后判断自己是否与 NDN 相邻以及 PV 表是否为空.如果不与 NDN 相邻且 PV 表已空,则说明 NMR 已分片,而且 A 与 NDN 处于不同分片中,于是 A 声明自己为新的 NDN,并把 B 失效的消息通知自己分片内的原 NMR 成员,其他 NMR 成员在收到后采取相同的处理办法,从而触发在分片内的 NMR 归并和 NDN 选举.如果 A 与 NDN 相邻或 PV 表非空,则说明 NMR 没有分片, A 仅把 B 从自己的成员列表中删除.当 NDN 失效时,NDN 的所有邻居都会声明自己为 NDN,从而触发新的选举过程.

2.7 OMR 管理

考虑到混合网络规模较小、拓扑不稳定的特点,HOMR 没有构造层次化的覆盖拓扑,整个覆盖网络是以各个 NMR 的 NDN 为顶点、覆盖链路为边的完全图.为保持覆盖网络的连通性,设置了若干策略:1) OMR peer 交换各自所知的 NDN 列表;2) 非 NDN 节点同步维护从自己 NDN 处获得的全局 NDN 列表;3) 若非 NDN 节点收到误发的多跳 HELLO 报文,则通过应答报文回复自己所属区域 NDN 的信息;4) NDN 改选后,原 NDN 如未退出或失效,须向新 NDN 交付自己所知的 NDN 列表;5) 因链路断裂产生的孤立节点,在链路恢复后应立即通过多跳 HELLO 报文尝试恢复与其他成员的联系;6) 如果有 MSMP 服务器(参见第 2.10 节),NDN 须与服务器定期交换全局 NDN 列表;7) 节点应通过超时机制主动退出长期无数据交换的组.

出于性能考虑,HOMR 在传输覆盖组播报文时,没有采取常见的 UDP 或 IP 隧道方式,而是引入了地址端口转换.每个 NDN 都设置一个 MAT(multicast address and port translation)表,格式如下:

节点 ID	组 ID	目的端口	转换地址	转换端口
-------	------	------	------	------

当某 NDN 节点 A 有组播报文要转发给另一 NDN 节点 B 时,它首先根据 B 的节点 ID、报文的地址(组

ID)和目的端口来查询自己的 MAT 表.如果找到相应表项,则用表项中的转换地址(即 B 的单播 IP 地址)和转换端口替换报文原有目的地址和目的端口,然后借助单播路由把报文投递给 B .如果 A 没有找到对应表项,则需缓存数据报文,并向 B 发送一个转换请求. B 收到请求后,为(组 ID,目的端口)分配一个本地空闲端口并存入 MAT 表,然后把映射结果回复给 A . A 在保存映射信息的同时,转换并发送缓存的组播报文. B 收到报文后,再根据报文的端口反查 MAT 表,进而恢复报文的原始组播地址和端口.

2.8 节点的加入和退出

HOMR 采用显式加入机制.节点在加入前应了解可加入的群组列表,在向已有成员发送请求并得到确认后,方可加入.从确认应答中,节点将获知更多的组内情况信息.加入后的节点首先建立与组 ID 对应的 HOT 数据结构,然后开始发送 HELLO 报文;如果邻居中有同组成员,将与之建立 PMR 或 NMR 邻居关系;否则,它将初始化新的 PMR 和 NMR,并声明自己为 DN,而后与组内其他 NDN 建立 OMR 邻居关系.

正常退出的成员会向 HOT 邻居发送退出消息,其他同组成员可就此尽快调整 HOT 拓扑.但更经常发生的是成员的异常退出,异常退出的节点不会把自己退出的消息通知给 HOT 邻居,邻居节点将依靠超时机制来确定节点的异常退出并作相应处理.

2.9 数据报文转发

成员节点根据自己了解的 HOT 拓扑,从硬件组播、IP 组播和覆盖组播中选择可行的最有效方式来转发组播数据.在 PMR 内,数据帧的转发过程对 HOMR 是透明的,HOMR 仅需把封装后的组播帧交付链路协议即可.HOMR 的 IP 层拓扑为 Mesh 结构,这里采用逆向路径转发(reverse path forwarding,简称 RPF)算法实现 NMR 内 IP 层的广播.如前所述,PMR 是 NMR 的子集,当需要从硬件组播转换到 IP 组播时,PMR 边界节点就会从组播帧中提取 IP 组播报文并在 IP 层转发.NMR 间的组播报文交换由 NDN 完成,如信源在某个 NMR 内,这个 NMR 的 NDN 就会将 IP 组播报文转换成单播报文,通过覆盖链路投递给其他 NDN;当 NDN 收到覆盖组播报文时,将恢复报文的原始状态,并在所属 NMR 内转发.一次典型的转发过程如图 5 所示.

在图 5 中,PMR 内各节点是全互连的,因此,PMR 内的链路并未标出.节点 A 的两个网络接口都支持 PMR 构建,因此,PMR1 和 PMR2 产生了交集.这里说明两点:1) 根据定义,非成员节点不了解组内的半覆盖拓扑,因此,如果信源 S 不加入组,则必须将数据交由一个组内节点代为转发;2) 由于 NMR 内为 Mesh 结构,所以不能完全避免重复报文的出现.为抑制重复报文,HOMR 采用了源序列号机制,节点只可能从不同链路上收到重复报文,而不会将同一报文反复转发.如果序列号在穿越覆盖网络时被丢弃,则由 NDN 添加新序列号后再在所属 NMR 内转发.

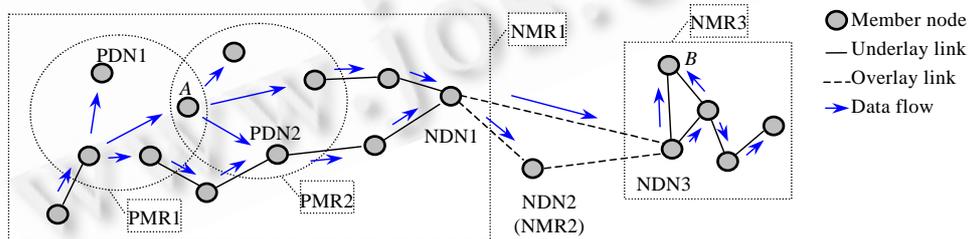


Fig.5 An illustration of data forwarding in a HOTA scenario

图 5 HOTA 场景示例及在此场景中的数据转发

2.10 对HOMR的辅助支持

HOMR 可在自身内部集成从 IP 组播到硬件组播的转换模块,但考虑到对未来技术的扩展性,将这种转换提取出来形成单独的子层将是更好的选择.为此,提出了一种组播适配协议(multicast adaptation protocols,简称 MAPs)框架,MAPs 是 HOMR 协议与链路层组播间的一层“垫片”,它们向 HOMR 提供一致的服务.每个 MAP 对

应一种硬件组播的技术实现,它与网络接口的硬件驱动一样按接口部署.在统一规范下,MAP 的实现应根据接口硬件提供.一个 MAP 的主要功能包括:1) 网络层组 ID 与链路组地址的双向映射;2) 对接口参数的配置和报告 3) 组播帧的封装和解封.

组播路由的最终目的是为网络应用服务,因此,有必要提供一个位于应用层的组播服务管理协议(multicast service management protocol,简称 MSMP).MSMP 的功能包括:1) 组播服务的注册、发布、索引和查询;2) 实现可控组播,包括认证、授权、完整性检验等在内.

对 MAPs 和 MSMP 的完整论述已经超出了本文的范围.在没有辅助支持时,HOMR 仍能完成 IP 层组播、覆盖转发以及简单的硬件组播转换.

2.11 讨论

与单层的组播协议相比,HOMR 涉及了多个协议层,这对协议复杂度和性能的影响值得关注.在复杂度方面我们认为,相对 IP 组播和覆盖组播而言,HOMR 在逻辑上既有扩充又有简化,其实现难度不会有明显增加,理由是:1) HOMR 的组播逻辑并非平均分布于各层,拓扑重构、转发方式决策等核心功能仍集中于网络层;2) MAP 模块的内部实现对路由协议透明,因此,HOMR 不会为了扩充硬件组播支持而不断扩大规模;3) 覆盖转发的引入使 HOMR 无须再构造组播树,这让协议内部逻辑和控制报文种类都得到简化;4) HOMR 仅通过 MAT 完成报文的覆盖转发,没有因构造覆盖组播拓扑而增加的复杂度.

在性能方面,以网络层为起点,HOMR 向硬件组播的扩展与 IP 组播类似,没有新增转换开销;向传输层的扩展有额外开销,但 NAT 的成功经验表明,基于地址端口转换的做法,其时间延迟和资源占用都在可接受的范围内.半覆盖拓扑有 3 种极限情况:1) 所有成员在同一 PMR 内的完全硬件组播;2) 不满足情况 1),但成员连片形成单一 NMR 的 IP 层组内广播;3) 组内任意两个成员均互不相邻的对等覆盖组播.显然,3 种情况的性能表现存在很大差异,因此,HOMR 的性能不仅取决于实现方式,更与成员密度及网络拓扑紧密相关.另外我们认为,以组播区域取代组播树还能使协议更快地适应网络拓扑和成员关系等的变化.

3 仿真及结果分析

为验证 HOMR 的性能表现,我们在 NS2 平台上进行了测试,并与 ODMRP(on-demand multicast routing protocol)^[14]和 MAODV(multicast operation of the ad-hoc on-demand distance vector)^[15]做了比较.ODMRP 和 MAODV 均为面向 MANET 的完全 IP 组播协议,其中,ODMRP 被认为是目前性能最好的 Ad-hoc 组播路由协议之一^[16,17];MAODV 则是 AODV(ad-hoc on-demand distance vector)^[18]的组播扩展.由于我们选用了不含组播功能的 AODV 作为支撑 HOMR 覆盖转发的单播路由协议,所以,与 MAODV 比较很有意义.

根据 IETF MANET 工作组的建议^[9],在性能测试和对比分析中采用以下评价标准:

- (1) 报文投递率(packet delivery ratio)=信宿实收非重复数据报文总数/信宿应收非重复数据报文总数;
- (2) 报文传送率(packet transmission ratio)=数据报文传送总次数/信宿实收非重复数据报文总数;
- (3) 控制开销率(control overhead ratio)=控制报文传送总比特数/信宿实收非重复数据报文总比特数;
- (4) 信道占用率(channel access efficiency)=全部报文传送总次数/信宿实收非重复数据报文总数.

为保证性能比较的公平性,在仿真中关闭了 HOMR 的硬件组播功能.另外,在计算 HOMR 的控制报文开销时,把 AODV 的所有开销也都一并计算在内.

仿真环境参数包括:共 50 个节点随机分布于 1km×1km 的平面内,采用 IEEE 802.11 接口,最大传输距离为 250m,信道带宽为 11Mb/s.数据为恒定比特率(CBR)分组,长度为 512B,发送频率为 2pkt/s.仿真持续 600s,节点在前 50s 内加入,不再退出.总共使用了 60 个不同的移动场景,各场景分别采样而后取平均值.

图 6 显示了在不同移动速度下协议性能的变化.(1) HOMR 与 MAODV 的变化趋势十分相似,这表现了单播路由策略对组播的影响.在 HOMR 协议中,当 PMR/NMR 邻居关系失效时,节点会首先尝试通过基于单播的覆盖链路转发报文,链路越不稳定,单播路由的使用频率越高.(2) HOMR 的 PDR 指标明显高于 MAODV,这源于 HOMR 以 Mesh 为主的转发结构和 MAODV 所完全基于的 Tree 结构.研究表明,在自组织网络中,Mesh 结构的转发成功率

高于Tree结构^[16,17]。(3) HOMR的 3 项耗费指标PTR(packet transmission ratio),COR(control overhead ratio)和 CAE(channel access efficiency)都高于MAODV,这表明在MAODV中,组播和单播得到了较高度度的整合,而 HOMR必须能与各种单播协议共存,这决定了它不会针对特定的协议进行优化。(4) 在多数情况下,ODMRP的 PDR较高,但相应的协议开销也明显高于HOMR和MAODV,这说明ODMRP是以额外控制来换取高投递率的.考虑到ODMRP本身具有单播路由能力,因此可以推知ODMRP的单播性能不及AODV.

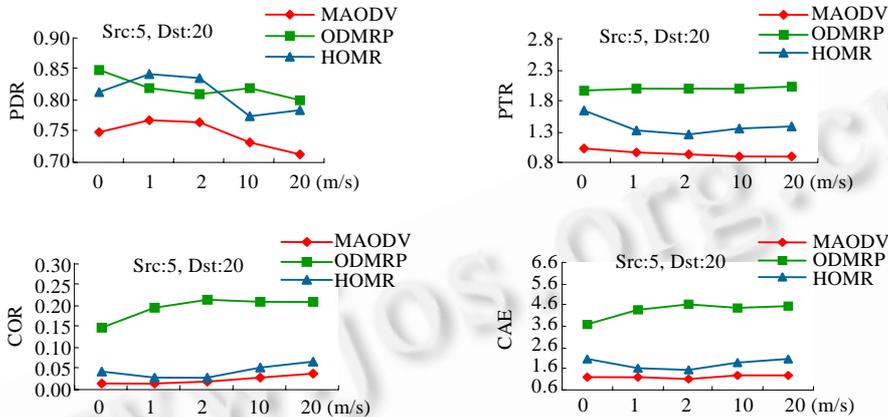


Fig.6 Performance metrics as a function of mobility speed

图 6 节点移动速度对协议性能的影响

图 7 显示了信源个数对性能的影响.当信源不断增加时,3 种协议的 PDR 都显著下降,ODMRP 的下降趋势最快,HOMR 其次,MAODV 最慢,这反映出链路承载能力已经成为决定性能表现的主要因素.由于信源生成报文的频率不变,当信源个数增加时,同时存在于网络中的报文个数就会同比例增加,信道逐渐变得拥挤,大量报文因竞争不到信道资源而被丢弃.此时,协议开销的增加会进一步加剧 PDR 的下降.从 COR 图可以看出,ODMRP 和 HOMR 的协议开销都会随着信源个数的增加而增长,ODMRP 开销的增长最为明显,MAODV 的开销则呈下降趋势,这个差别最终影响了 PDR 变化的速度.另外,当链路拥挤时,MAODV 的 Tree 结构显示出了相对于 Mesh 结构的优势,Tree 结构在转发时不产生重复报文,因此,PTR 和 CAE 不会随着信源的增加而上升,这虽然不能缓解链路的拥挤状况,但对 PDR 的负面影响却能相对降低.

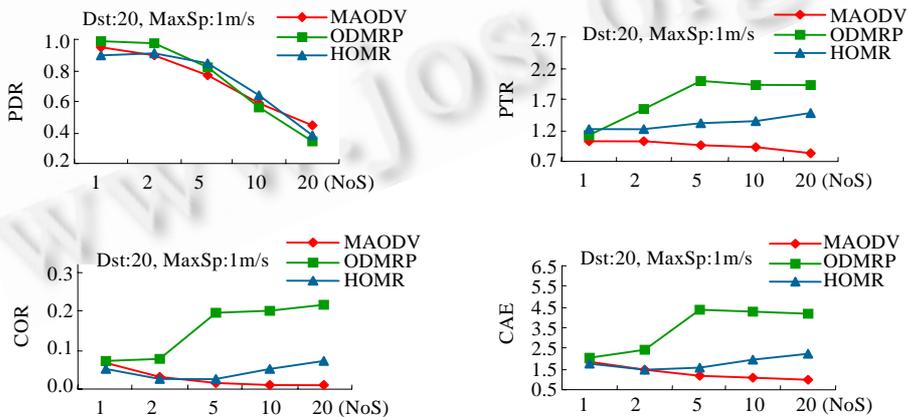


Fig.7 Performance metrics as a function of number of senders

图 7 信源个数对协议性能的影响

在仿真中,通过调整信宿个数验证了群组规模对协议性能的影响,如图 8 所示.当规模增加时,HOMR 的 PDR

指标好于 ODMRP 和 MAODV,这显现出了成员密集程度对 HOMR 行为的影响.在一个空间内,如果同组成员越多,则 HOMR 有更多的机会构造出规模较大的 NMR,以非覆盖方式完成转发操作的比例将会随之增加,这显然有助于提高转发效率和投递成功率.HOMR 的报文投递率之所以比 ODMRP 还要高,是因为 HOMR 在 NMR 内的转发算法与 Flooding 相似,而 Flooding 算法的 PDR 指标要好于 ODMRP.

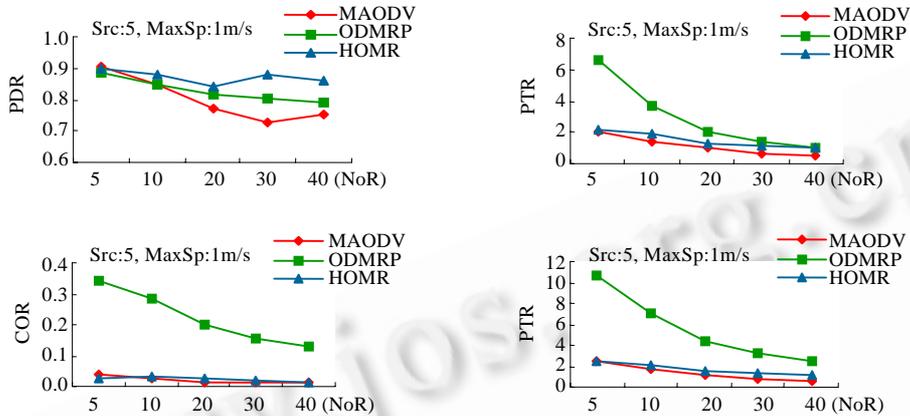


Fig.8 Performance metrics as a function of multicast group size

图 8 群组规模对协议性能的影响

总的来看,作为一种半覆盖协议,HOMR 与 ODMRP 和 MAODV 相比并没有显著的性能损失,特别是与 MAODV 相比,HOMR 以较少的额外开销,换取了在部署和负载分配等方面的灵活性以及在多数情况下更高的投递率.另外,HOMR 可以适应更为广泛的网络环境,这种通用性也是 MANET 组播协议所不具备的.

4 结论及进一步工作

本文分析了混合网络的特点,提出了一种能够满足混合网络组播要求的半覆盖拓扑模型,并据此给出了 HOMR 协议的核心路由策略.HOMR 在单一协议中融合了覆盖组播和非覆盖组播的多项特性,包括:(1) 无须对自身的连续部署或加载,能使数据穿越不可控的转发网络;(2) 仅要求节点间单播可达,对单播路由方式无特殊要求,具有与 PIM 类似的协议无关性;(3) 节点仅管理已加入的群组状态,资源占用少;(4) 能够直接感知网络拓扑的变化,加快了反应速度;(5) 尽可能使用非覆盖转发,有效减少重复报文.此外,除了在覆盖转发中用到了地址端口转换外,HOMR 没有引入其他的跨层机制;通过对 PMR 和 MAPs 框架的定义,HOMR 保持了对未来硬件组播发展的可扩展性.仿真研究表明,HOMR 的协议开销较低,转发能力与完全 IP 组播路由协议水平相近.

尽管本文是结合混合网络环境对 HOMR 进行介绍的,但事实上,HOT 模型和 HOMR 的大部分路由策略都能被用于广域网组播.使 HOMR 完全适应广域网环境将是研究的重点,例如,广域网内 NDN 间的覆盖链路大多是长路径.所以,有必要引入传输控制机制来保证覆盖转发的可靠性.

References:

- [1] Sonia F, Minseok K. Characterizing overlay multicast networks. In: Frances T, ed. Proc. of the 11th IEEE Int'l Conf. on Network Protocols (ICNP). Atlanta: IEEE Computer Society Press, 2003. 61-70.
- [2] Li L, Cui JH, Mario G, Dario M. A comparative study of multicast protocols: Top, bottom, or in the middle? In: Kia M, Edward K, eds. Proc. of the IEEE INFOCOM 2005. Miami: IEEE Communications Society Press, 2005. 2809-2814.
- [3] Carlos de MC, Hrishikesh G, Dharma PA. Multicast over wireless mobile ad hoc networks: Present and future directions. IEEE Network, 2003,17(1):52-59.
- [4] Aijun D, Gee-Swee P. A survey of optical multicast over WDM networks. Computer Communications, 2003,26(2):193-200.

- [5] Hrishikesh G, Nagesh N, Kumar A, Dharma PA. Supporting MAC layer multicast in IEEE 802.11 based MANETs: Issues and solutions. In: Stiller B, ed. Proc. of the IEEE Int'l Conf. on Local Computer Networks (LCN). Tampa: IEEE Press, 2004. 172–179.
- [6] Juyoung P, Seok JK, Shin GK, Dae YK. Multicast delivery based on unicast and subnet multicast. IEEE Communications Letters, 2001,5(4):181–183.
- [7] Beichuan Z, Wenjie W, Sugih J, Daniel M, Lixia Z. Universal IP multicast delivery. Computer Networks, 2006,50(6):781–806.
- [8] Cheuk KWR, Chan SHG, Jack YBL. Island multicast: The combination of IP multicast with application-level multicast. In: Heather Y, Steve W, eds. Proc. of the IEEE Int'l Conf. on Communications (ICC). Paris: IEEE Press, 2004. 1441–1445.
- [9] Garyfalos A, Almeroth K, Finney J. A hybrid of network and application layer multicast for mobile IPv6 networks. In: Proc. of the Int'l Workshop on Large-Scale Group Communication (LSGC). Florence, 2003. <http://litiwww.epfl.ch/~pedone/workshop.html>
- [10] Wu Q, Wu JP, Xu K, Liu Y. A survey of the research on IP multicast in mobile Internet. Journal of Software, 2003,14(7):1324–1337 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1324.htm>
- [11] Min G, Srikanth VK, Michalis F. Overlay multicasting for ad hoc networks. In: Proc. of the 3rd Annual Mediterranean Ad Hoc Networking Workshop (MedHocNet). Bodrum, 2004. <http://www.ece.osu.edu/medhoc04/>
- [12] Chao G, Prasant M. Efficient overlay multicast for mobile ad hoc networks. In: Tachikawa K, ed. Proc. of the IEEE Wireless Communications and Networking (WCNC). New Orleans: IEEE Communications Society Press, 2003. 1118–1123.
- [13] Wang J, Wu ZM. Multicast protocol over switch Ethernet. Journal of Software, 2003,14(3):496–502 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/496.htm>
- [14] Lee SJ, William S, Mario G. On-Demand multicast routing protocol in multihop wireless mobile networks. Mobile Networks and Applications, 2002,7(6):441–453.
- [15] Elizabeth MR, Charles EP. Multicast operation of the ad-hoc on-demand distance vector routing protocol. In: Harel K, ed. Proc. of the 5th ACM/IEEE Annual Conf. on Mobile Computing and Networking (MOBICOM). Seattle: ACM Press, 1999. 207–218.
- [16] Lee SJ, William S, Julian H, Mario G, Rajive B. A performance comparison study of ad hoc wireless multicast protocols. In: Sidi M, ed. Proc. of the IEEE INFOCOM 2000. Tel-Aviv: IEEE Communications Society Press, 2000. 565–574.
- [17] Kumar V, Katia O, Gene T. Exploring mesh and tree based multicast routing protocols for MANETs. IEEE Trans. on Mobile Computing, 2006,5(1):28–42.
- [18] Charles EP, Elizabeth MR. Ad-Hoc on-demand distance vector routing. In: Kristine K, ed. Proc. of the 2nd IEEE Workshop on Mobile Computer Systems and Applications (WMCSA). New Orleans: IEEE Computer Society Press, 1999. 90–100.
- [19] Corson S, Macker J. Mobile ad hoc networking (MANET): Routing protocol performance issues and evaluation considerations. IETF RFC2501, 1999.

附中文参考文献:

- [10] 吴茜,吴建平,徐恪,刘莹.移动 Internet 中的 IP 组播研究综述.软件学报,2003,14(7):1324–1337. <http://www.jos.org.cn/1000-9825/14/1324.htm>
- [13] 王军,吴志美.交换式以太网上的多播协议.软件学报,2003,14(3):496–502. <http://www.jos.org.cn/1000-9825/14/496.htm>



林彤(1975—),男,福建闽侯人,博士,工程师,主要研究领域为计算机网络体系结构,协议设计与分析,自组织网络与组播技术.



葛敬国(1973—),男,博士,助理研究员,主要研究领域为互联网体系结构.



钱华林(1940—),男,研究员,博士生导师,主要研究领域为计算机网络体系结构,数据通信.



牛广锋(1976—),男,研究实验员,主要研究领域为计算机网络.