软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(10):2385−2400 [doi: 10.3724/SP.J.1001.2011.03905]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于池的 PMIPv6 移动接入网关容错方案*

张瀚文[1+], 许智君[1,2], 张玉军[1], 李忠诚[1], 周继华[3]

[1](中国科学院 计算技术研究所 网络技术研究中心,北京 100190)

[2](中国科学院 研究生院,北京 100190)

[3](重庆金美通信有限责任公司,重庆 400030)

Fault-Tolerant Approach Based on Pool for MAG in PMIPv6

ZHANG Han-Wen[1+], XU Zhi-Jun[1,2], ZHANG Yu-Jun[1], LI Zhong-Cheng[1] ZHOU Ji-Hua[3]

[1](Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

[2](Graduate University, The Chinese Academy of Sciences, Beijing 100190, China)

[3](Chongqing Jinmei Communications Co. Ltd, Chongqing 400030, China)

+ Corresponding author: E-mail: hwzhang@ict.ac.cn

Zhang HW, Xu ZJ, Zhang YJ, Li ZC, Zhou JH. Fault-Tolerant approach based on pool for MAG in PMIPv6. *Journal of Software*, 2011,22(10):2385−2400. http://www.jos.org.cn/1000-9825/3905.htm

Abstract: Proxy Mobile IPv6 (PMIPv6) is the network-based localized mobility management protocol proposed by IETF. The local mobility anchor (LMA) and the mobile access gateway (MAG) are the key entities realizing the system function of PMIPv6. To solve the reliability problem of PMIPv6 MAG, this paper proposes an MAG fault-tolerant method based on pool (MAGFT). MAGFT introduces MAG pool to realize MAG fault-tolerant. Several MAG pools are constructed in an PMIPv6 domain, and each MAG belongs to at least one of the pools. When an MAG fails, a given MAG in the same pool will take over. The results of theoretical analysis and simulation show that the fault-tolerant latency of MAGFT can be restricted within 35ms~340ms. When the fault-tolerant latency is lower than 120ms, MAGFT can avoid the influence of MAG failure on MNs' up-layer TCP applications. In the worst condition, MAGFT can resume an MN's TCP throughput within 1.1s, 1.6s, and 2.8s respectively when the MN resides in WLAN, 3G or satellite network. For the upper applications based on UPD, MAGFT can resume MNs' packet rate within 2s after the occurrence of an MAG failure. At the same time, MAGFT introduces low signaling cost, which can be neglected when compared with the PMIPv6 signaling. MN's access delay introduced by MAGFT is no more than 10ms.

Key words: proxy mobile IPv6; mobility access gateway; fault-tolerant; reliability

摘 要: 代理移动 IPv6(PMIPv6)是 IETF 提出的基于网络的区域移动管理协议,依赖于区域移动锚点(local mobility anchor,简称 LMA)和移动接入网关(mobile access gateway,简称 MAG)两类移动管理实体实现系统功能.针对 PMIPv6 网络中的 MAG 可靠性问题,提出一种基于池的移动接入网关容错方案(MAG fault- tolerant method based

* 基金项目: 国家自然科学基金(60803139, 61100177); 国家重点基础研究发展计划(973)(2007CB310702); 国家科技支撑计划(2008BAH37B02)

收稿时间: 2009-10-15; 定稿时间: 2010-06-28

on pool,简称 MAGFT).方案引入 MAG 池解决 PMIPv6 系统中 MAG 服务不可替代问题,实现对移动节点(mobile node,简称 MN)透明的 MAG 容错.针对 PMIPv6 系统所部署下层网络的不同,MAGFT 分别采用无重叠区部署和有重叠区部署两种模式在 PMIPv6 域内构建多个 MAG 池,使得域内各 MAG 至少归属于一个池.当某 MAG 失效时,它所在池内的某一有效 MAG 将快速接管其服务.理论分析和仿真实验结果表明,MAGFT 可将容错时间控制在 35ms~340ms.当容错时间在 120ms 以下时,MAGFT 可完全避免 MAG 失效对 MN 的 TCP 应用造成的影响;最差情况下,对分别处于 WLAN、3G 和卫星网络中的 MN 而言,MAGFT 也可在 MAG 失效发生后的 1.1s,1.6s 或 2.8s 内恢复其 TCP 应用吞吐量.对于 UDP 应用,MAGFT 可在 MAG 失效发生后 2s 内将 MN 的收包率恢复至其稳定值.同时,方案引入的容错开销小,当系统处于较饱和的稳定服务状态时,容错信令开销相比系统基本信令是可忽略的.MAGFT 的引入对 MN 接入延时略有增加,但增值控制在 10ms 以下.

关键词: 代理移动 IPv6;移动接入网关;容错;可靠性

中图法分类号: TP393 **文献标识码**: A

代理移动 IPv6(proxy mobile IPv6,简称 PMIPv6)协议[1]是 IETF 提出的基于网络的区域移动管理协议,它可与 MIPv6[2],HIP[3],MOBIKE[4]等任意广域移动管理协议相结合,高效地实现异构融合网络环境下的终端移动性支持.

与其他区域移动管理协议[5−7]相比,PMIPv6 具有无需终端支持、切换性能高、信令开销小、终端位置私密性保护、链路技术无关性支持及多穴支持等优势.目前,PMIPv6 已正式被 3GPP SAE,WiMAX 等多种无线通信技术标准所采用[8,9],成为新一代宽带无线网络建设中实现区域移动管理的首选技术方案.

PMIPv6 旨在实现无需移动节点(mobile node,简称 MN)参与的、基于网络的 IP 移动管理,为此,PMIPv6 引入了两类移动实体:区域移动锚点(localized mobility anchor,简称 LMA)和移动接入网关(mobility access gateway,简称 MAG),维持 MN 在 PMIPv6 域内的不同接入链路间移动过程中的可寻址性和数据连续性.LMA 和 MAG 作为 PMIPv6 系统中的重要基础设施,同时也是 PMIPv6 可靠性的瓶颈所在.移动管理实体本身不可能做到完全可靠,绝对可靠的、不发生故障的网络设备是不存在的.网络攻击日趋多样并难以防范,开放的移动网络环境更是增加了系统的故障率和不安全因素.一旦移动管理实体由于恶意或非恶意的、人为或非人为的因素发生故障,将不能继续为其所服务的移动节点维护路由信息,使得相关移动节点失去了可寻址性.因此,有必要研究 PMIPv6 移动管理实体的容错技术,使得网络中任意实体失效后,系统仍能为相关移动节点提供既定的路由服务.

本文针对 PMIPv6 MAG 容错方案,使得任意 MAG 失效后,系统仍能提供既定服务.LMA 容错作为我们另一部分研究工作,未在本文中涉及.

本文第 1 节是问题描述及相关工作介绍.第 2 节提出基于池的 MAG 容错方案.第 3 节和第 4 节分别通过理论分析和系统仿真,从 MAG 容错时间、TCP 及 UDP 应用下的 MAG 容错性能、容错开销等方面,对方案进行定量评价.最后总结全文并给出下一步工作方向.

1 问题描述及相关工作

如图 1 所示,在 PMIPv6 系统中,LMA 是 PMIPv6 域内所有 MN 的路由锚点,主要承担两方面的功能:

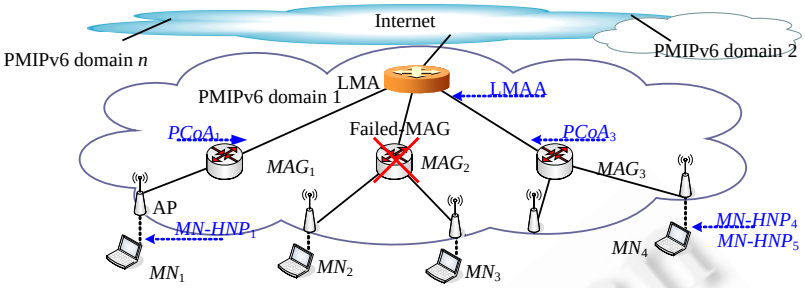
- 维护 MN 在 PMIPv6 域内的路由信息,记录其当前附着的 MAG;
- 通过与 MN 当前附着的 MAG 间的双向隧道为 MN 转发数据包.

MAG 作为 MN 接入链路的第 1 跳路由器,是 PMIPv6 系统实现基于网络的移动管理服务的基础,其主要功能包括:

- 负责对 MN 在接入链路的附着及离开进行检测;
- 作为 MN 的代理向区域内的移动锚点 LMA 进行移动注册,更新 MN 位置信息;
- 通过向 MN 宣告其家乡网络前缀(MN’s home network prefix,简称 MN-HNP),为 MN 模拟家乡链路,使得

MN 在 PMIPv6 域内移动时感知不到 3 层移动;

- 通过与 LMA 建立的双向隧道为 MN 转发数据,建立网络连接.



LMAA: LMA address; Proxy-CoA: Proxy care-of-address; MN-HNP: MN's home network prefix

Fig.1 PMIPv6 architecture

图 1 PMIPv6 系统结构

PMIPv6 协议不支持 MAG 容错.一旦某 MAG 失效,将导致通过该 MAG 接入的所有 MN 失去网络连接.PMIPv6 系统所部署的不同下层接入网络自身对 MAG 容错的支持能力有所不同.比如,PMIPv6 部署于 WLAN 网络则完全无法实现 MAG 容错,部署于 3GPP SAE 网络则具有一定的 MAG 容错能力.在 WLAN 网络中,PMIPv6 MAG 部署于接入链路的缺省路由器,覆盖范围无重叠区,相互间无法替代服务.一旦某 MAG 失效,附着于该 MAG 的所有 MN 将失去网络连接,除非这些 MN 移动至其他 MAG 的覆盖范围.在 3GPP SAE 网络中,PMIPv6 MAG 实体功能实现于 SAE 分组核心网(EPC)中的服务网关(S-GW).当 MAG 失效时,通过 SAE 本身的管理机制,移动管理网元(MME)和当前通过失效 MAG 为 MN 提供数据面连接的 eNodeB(evolved node-B),经过一段时间将检测到该 MAG 的失效,MME 为失去数据面连接的 MN 重新选择可用 MAG(S-GW),eNodeB 与新选择的 MAG 为 MN 建立新的数据面连接.然而,这种容错处理对 MN 不透明,容错延迟大.更重要的是,这种基于接入网络技术本身的 MAG 容错机制依赖于特定的接入网络构架,不具有通用性,无法适用于其他类型的接入网络,这违背了 PMIPv6 具有链路技术无关性支持的协议设计目标.

为实现 PMIPv6 的大规模部署,需要对 PMIPv6 协议本身进行扩展,实现协议自身的 MAG 容错能力,而无需借助任何下层接入网络的管理架构.PMIPv6 于 2008 年 8 月被 IETF netlmm 工作组正式发布,目前针对 PMIPv6 的研究还主要集中在切换性能、多穴支持、双栈支持、PMIPv6-MIPv6 交互等方面[10−13].尽管可靠性技术已被确定为 IETF netlmm 工作组的重点工作内容之一,但该方面的工作尚处于起步阶段,目前仅有针对实体间可达性检测的方案提出[14].

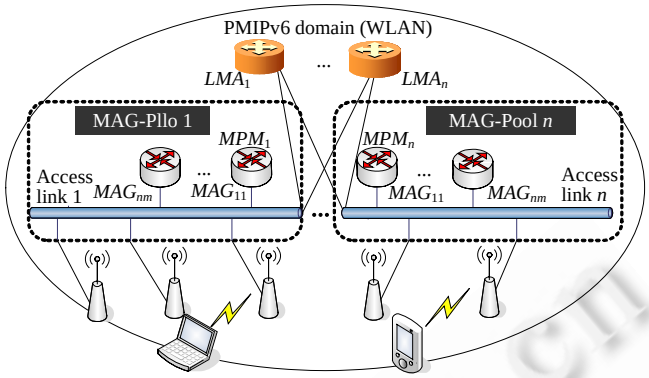
文献[14]提出了一种 PMIPv6 心跳机制,目的是实现 MAG 和 LMA 之间相互的可达性检测,而不再像以往那样依赖于应用层面和路由层面的数据传输中断来进行检测.该方案也可用于同种移动管理实体间(LMA 相互间、MAG 相互间)的可达性检测.然而,有效性检测仅仅是实现容错的第 1 步,针对 PMIPv6 移动管理实体的完备容错方案还有待进一步加以研究.

2 基于池的 MAG 容错方案

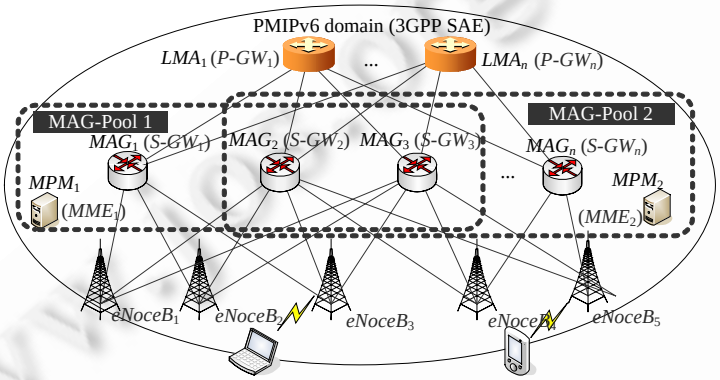
本文提出一种基于池的 MAG 容错方案(MAG fault-tolerant method based on pool,简称 MAGFT),其基本思想是:引入 MAG 池(MAG-pool),即能相互替代服务的 MAG 集合,解决 MAG 间服务不可替代问题.针对 PMIPv6 系统所部署的下层网络的不同,可分别通过无重叠区部署和有重叠区部署两种模式在 PMIPv6 域中构建多个 MAG 池,使得各 MAG 至少归属于一个池.当某 MAG 失效时,它所在的 MAG 池内的某一有效 MAG 将快速接管其服务,实现对 MN 透明的 MAG 容错.

MAGFT 方案系统结构如图 2 所示,在 PMIPv6 域中构建多个 MAG 池,每个 MAG 池中引入一个功能实体 MPM(MAG pool manager)作为集中管理者和面向 MN 的统一服务接口,负责 MAG 池的构建与维护,实现用户接

入、MAG 容错及负载均衡的统一决策.



(a) Non-Overlap mode
(a) 无重叠区部署模式



(b) Overlap mode
(b) 有重叠区部署模式

Fig.2 MAG-Pool deployment
图 2 MAG 池的构建

2.1 MAG池的构建与维护

根据 PMIPv6 系统所部署的下层网络的不同,MAGFT 设计了两种 MAG 池构建模式:

1) 无重叠区部署模式:典型地针对于 WLAN 一类网络.在这种部署模式下,系统原有的各 MAG 间无法相互替代服务,通过为各 MAG 部署覆盖范围相同的多个 MAG,构建 MAG 池.以 WLAN 为例(如图 2(a)所示),MAG 部署于各接入链路的缺省路由器,只有同一链路内的 MAG 可相互替代服务,需为 PMIPv6 域内的每条链路部署多 MAG,同一链路的多 MAG 构成一个 MAG 池;

2) 有重叠区部署模式:典型地针对于 3GPP SAE 一类网络.在这种部署模式下,系统原有的各 MAG 覆盖范围存在交集,具有相互替代服务的能力.因此,根据各 MAG 的覆盖范围,逻辑上将所有 MAG 组织成多个 MAG 池.这种构建方式下,一个 MAG 可能同时属于多个不同的 MAG 池.通过 MAG 池优先级设置,使得某 MAG 失效后,其所属的多个 MAG 池中仅有优先级最高的一个实现其服务接管.以 3GPP SAE 网络为例(如图 2(b)所示),MAG 功能实体实现于 SAE 分组核心网(EPC)中的服务网关(S-GW),与相同 eNodeB 存在连接的 MAG 都可相互替代服务,将这些 MAG 组织成一个 MAG 池.

功能实体 MPM(MAG pool manager)是 MAG 池的集中管理者和面向 MN 的统一服务接口.作为池的管理者,MPM 负责池内的全局状态管理;根据全局状态为各 MAG 配置备份节点(backup-MAG,简称 b-MAG);负责各

MAG 的失效检测,并在检测到失效后触发其 b-MAG 接管服务.作为 MAG 池面向 MN 的统一服务接口,MPM 在 MN 接入时负责接收其附着请求,根据全局状态为 MN 动态选择最优服务节点,将附着请求重定向至被选 MAG;被选 MAG 接收附着请求后,作为该 MN 的服务节点,触发正常的 PMIPv6 注册流程.这样做的目的首先是为遵循 PMIPv6 协议基于网络实现系统服务的设计原则,实现方案对 MN 的透明性.同时,通过统一服务接口的集中决策,实现了全局负载在 MAG 间的动态均衡,提升了系统性能.

MPM 可实现于不同网元:如在 WLAN 中,MPM 可实现于池内的任意 MAG;在 3GPP SAE 中,则可实现于核心网中的 MME(mobility management entity,移动管理网元).MPM 的可靠性问题通过 backup-MPM 解决,backup-MPM 同步备份 MPM 所维护的所有信息,一旦 MPM 失效,它将替代 MPM 对池进行管理.

MPM 维护 MAG 列表(MAG list,简称 MAGL),管理 MAG 池的全局信息,包括各 MAG 的优先级、当前状态(“有效”或“失效”)、备份节点等.为实现失效 MAG 的快速服务接管,MPM 根据全局状态为各 MAG 选择唯一的备份节点(b-MAG),实现 MAG 移动管理信息备份;当 MAG 失效时,由 b-MAG 基于备份信息快速接管其服务.MAG_j 为 MAG_i 的 b-MAG,称 MAG_i 为 MAG_j 所支持的节点(supported-MAG,简称 s-MAG).

2.2 MAG失效检测

MPM 负责池内所有 MAG 的有效性检测,一旦检测到某 MAG 失效,将触发 MAG 服务迁移,实现失效 MAG 的服务接管.同时,为失效 MAG 的 s-MAG 选择新的 b-MAG.MAG 失效检测实现如下(如图 3 所示):

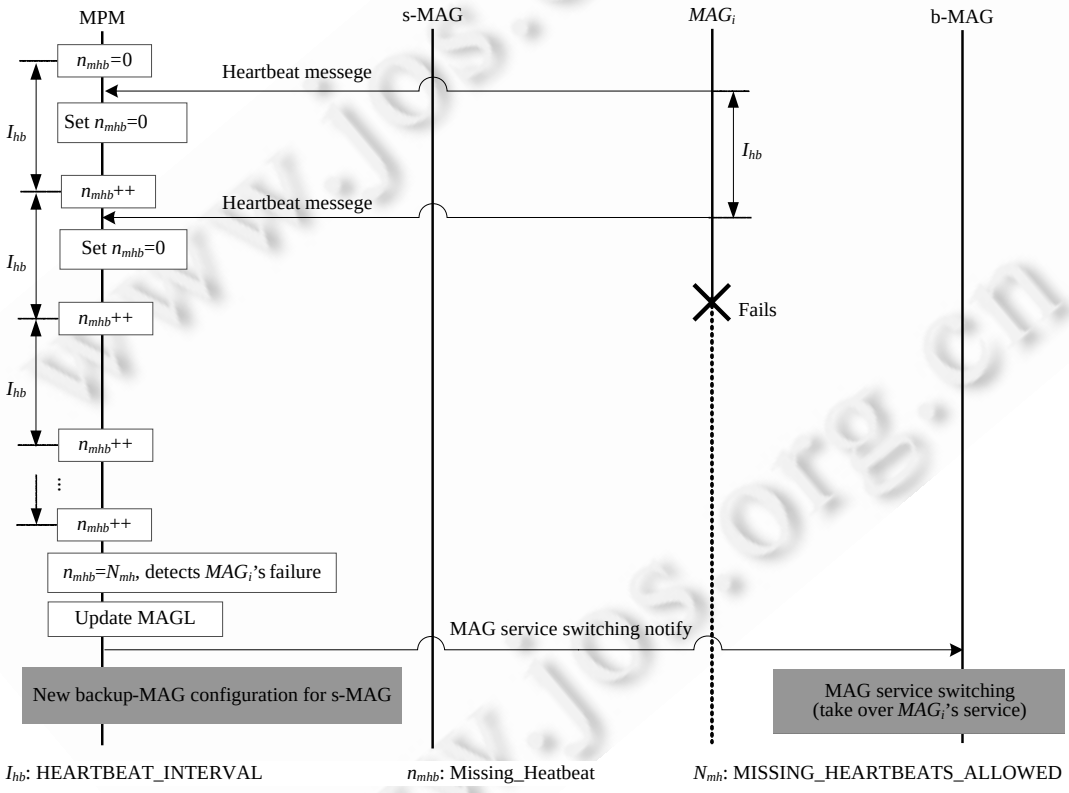


Fig.3 MAG failure detection

图 3 MAG 失效检测

1. 各 MAG 以间隔 I_{hb} 周期性地向 MPM 发送心跳消息(heart message,简称 HM),这里仅以 MAG_i 为例.

- MPM 接收 MAG_i 的 HM 后,将为其维护的丢失心跳消息计数器(missing_heartbeat)值 $n_{mhb}(i)$置 0.

2. MPM 为池内各 MAG 维护一个心跳计时器,这里仍以 MAG_i 为例,即 $TO(i)$.

- $TO(i)$初始化为 I_{hb};
- 每当定时器值 $TO(i)$到期,MPM 将 $n_{mhb}(i)$增 1,判断 $n_{mhb}(i)$是否已到达预设的上限值(missing_heatbeat_allowed)N_{mh}:若 $n_{mhb}(i)<N_{mh}$,则重置 $TO(i)=I_{hb}$;若 $n_{mhb}(i)=N_{mh}$,则判断 MAG_i 失效.

3. MPM 判断 MAG_i 失效后的处理.

- 在 MAGL 中将 MAG_i 的状态更新为“失效”;
- 向 MAG_i 的 b-MAG 发送 MAG 服务迁移通告(MAG service switching notify,简称 MSSN),触发失效 MAG 的服务迁移过程;
- 为 MAG_i 的 s-MAG 选择新的 b-MAG,触发备份节点配置过程.

2.3 MAG服务迁移

MAG 服务迁移机制支持失效实体全部服务迁移和过载实体部分服务迁移.当 MPM 检测到某 MAG 失效时,触发其 b-MAG 作为接管实体(takeover-MAG,简称 t-MAG),发起服务迁移过程.b-MAG 基于已同步备份的失效 MAG 的移动管理信息,接管附着于失效 MAG 的所有 MN(后续称为迁移 MN).为实现 MAG 池内的负载均衡,MPM 可能发起 MAG 间的部分服务迁移.当出现“过载”MAG 时,MPM 可在“空闲”MAG 中为其选择一个或多个 t-MAG,接管“过载”MAG 的部分负载.MAG 服务迁移过程具体实现如下(如图 4 所示):

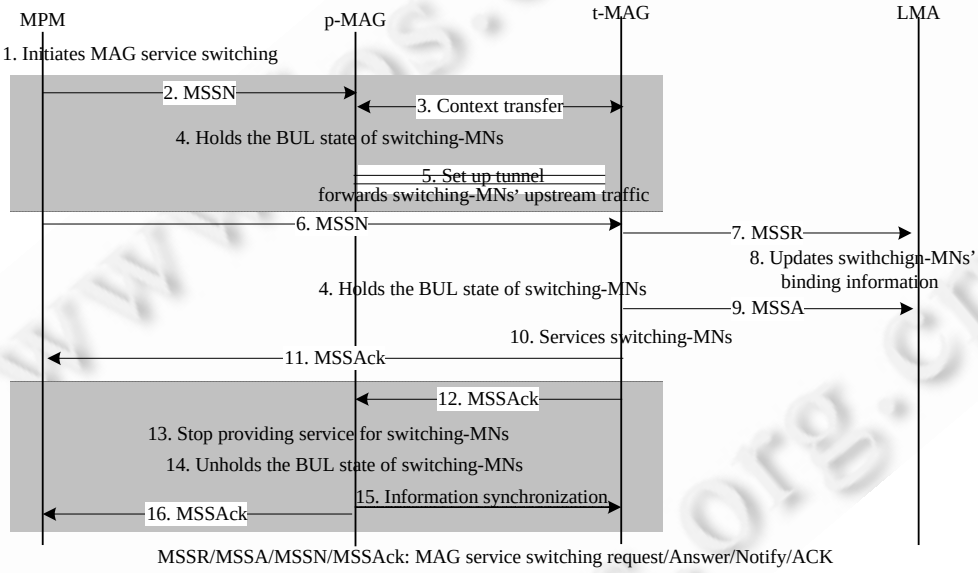


Fig.4 MAG service switching

图 4 MAG 服务迁移

1. MPM 触发 MAG 服务迁移过程.

- 触发事件:MPM 检测到 p-MAG 失效;或 MPM 要求“过载”p-MAG 将部分负载迁移至 t-MAG.

步骤 2~步骤 5 仅在 p-MAG 过载触发的服务迁移场景下执行:

2. MPM 向 p-MAG 发送 MAG 服务迁移通告(MAG service switching notify,简称 MSSN).

- MSSN 指示 p-MAG 和 t-MAG 的标识;指示切换原因为过载;可选的,MSSN 指示需要迁移的 MN 的百分比.

3. p-MAG 与 t-MAG 间上下文传递.

- p-MAG 通过上下文传递将迁移 MN 的移动管理信息发送至 t-MAG.

4. p-MAG 锁定待迁移 MN 的绑定更新列表(binding update list,简称 BUL).
- 在步骤 5~步骤 13 的迁移过程中,p-MAG 暂存会导致 BUL 中迁移 MN 信息改变的事件.

5. p-MAG 建立到 t-MAG 的单向隧道(可选步骤).
- p-MAG 通过该隧道,将迁移 MN 的上行数据转发至 t-MAG;t-MAG 缓存 p-MAG 隧道转发的迁移 MN 上行数据包.

6. MPM 向 t-MAG 发送 MSSN(p-MAG 过载触发的服务迁移场景下,与步骤 2 同时进行).
- MSSN 指示 p-MAG 标识、t-MAG 的标识;指示服务迁移触发原因(p-MAG 失效、p-MAG 过载迁移).

7. t-MAG 向待迁移 MN 的 LMA 发起 MAG 服务迁移请求(MAG service switching request,简称 MSSR).
- MSSR 指示 p-MAG 的标识和迁移 MN 的标识.

8. LMA 更新迁移 MN 的移动管理信息.
9. LMA 向 t-MAG 回复 MAG 服务迁移应答(MAG service switching answer,简称 MSSA).
10. t-MAG 向迁移 MN 提供移动管理服务.
- 替代 p-MAG 向迁移 MN 发送生命周期为 0 的路由器宣告;更新 BUL,为迁移 MN 创建路由;向迁移 MN 发送路由器宣告消息,宣告 MN-HNP;若执行了步骤 5,将缓存的迁移 MN 的上行数据包发往 LMA.

11. t-MAG 向 MPM 发送 MAG 服务迁移确认(MAG service switching ack,简称 MSSAck).

步骤 12~步骤 16 仅在 P-MAG 过载触发的服务迁移场景下执行:

12. t-MAG 向 p-MAG 发送 MSSAck.
13. p-MAG 停止为迁移 MN 提供移动管理服务.
- 更新 BUL,删除迁移 MN 的路由信息;停止向迁移 MN 发送路由器宣告消息;若执行了步骤 5,则拆除与 t-MAG 建立的单向隧道.

14. p-MAG 解锁 BUL 的状态.
15. 如果切换过程中,p-MAG 缓存了会导致 BUL 中迁移 MN 信息改变的事件,则切换完成后,p-MAG 将缓存的事件上报给 t-MAG,由 t-MAG 进行处理.
16. p-MAG 向 MPM 回复 MSSAck,指示服务迁移完成.

3 MAGFT 方案性能分析

容错效率和容错开销是评价 MAG 容错方案的重要指标.MAG 容错不仅要能恢复失效 MAG 当前服务 MN (failure-affected-MN,后文简称 fa-MN)的网络连接,同时应尽量缩短容错时间,降低 MAG 失效对 fa-MN 上层应用造成的影响.同时,容错方案应尽量减小引入的容错开销.

另一方面,容错机制的引入应尽量保证不降低移动管理系统本身的性能.由于 MAGFT 实现了 MAG 间的负载均衡,该方案可提升系统总体服务效率,但 MAG 池的引入也将影响 MN 的接入延时.

本节和下一节将分别通过理论分析和系统仿真,从 MAG 容错时间、TCP 和 UDP 应用下的 MAG 容错性能、容错信令开销、方案引入后的 MN 接入延时等方面,对 MAGFT 方案进行定量评价.

3.1 性能分析模型

图 5 为 MAGFT 方案的性能分析网络模型,表 1 为相关参数及其经验值[10,15,16].

Table 1 Pararameter description

表 1 参数说明

t_{mr}	The one-way link delay between the MN and the BS/AP, 10ms (WLAN), 50ms (3G), 150ms (satellite)
t_{rm}	The one-way link delay between the BS/AP and the entities in the MAG pool, 2ms~5ms
t_{mm}	The one-way link delay between the MN and the MAG, $t_{mm}=t_{mr}+t_{rm}$
t_{mp}	The one-way link delay between the MAG and the MPM, 2ms~5ms
t_{ml}	The one-way link delay between the MAG/MPM and the LAM, 10ms~30ms
t_{cl}	The one-way link delay between the CN and the LAM, 40ms~50ms

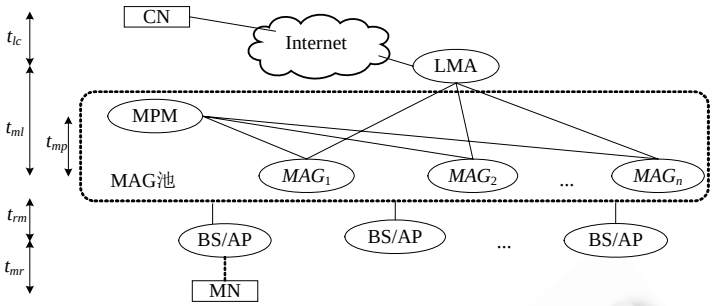


Fig.5 MAGFT network model for performance analysis

图 5 MAGFT 性能分析网络模型

3.2 MAG容错时间分析

MAG 容错时间 T_{ft}^{MAG} 包括失效检测时间 T_{fd}^{MAG} 和 MAG 服务迁移时间 T_{ss}^{MAG}.

MAG 失效检测时间分析如图 6 所示.MPM 以间隔时间 I_{hb} 接收 MAG_i 向其发送的心跳消息 HM.MAG_i 失效发生在 MPM 第 i 次接收到 MAG_i 发送的 HM 之后.从 t_{i+1} 开始,每次心跳间隔时间 I_{hb} 到期,MPM 都将为 MAG_i 维护的 $n_{mhb}(i)$值增 1,直到 $t_{i+N_{mh}}$. $n_{mhb}(i)$值达到 N_{mh},MPM 判断 MAG_i 失效.令θ表示 MAG_i 失效发生到 t_{i+1} 的时间,那么 t_{i+1} 服从[0,I_{hb}]的均匀分布.MA 失效检测时间为 $T_{fd}^{\mathrm{MAG}}=\theta+(N_{mh}-1)\times I_{hb}$.

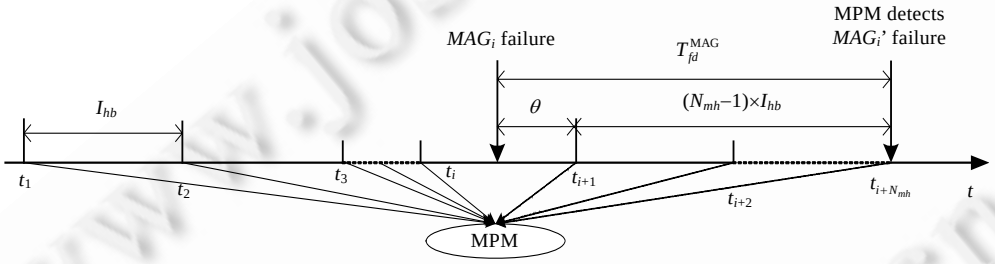


Fig.6 Analysis of MAG failure detection time

图 6 MAG 失效检测时间分析

MPM 检测到 MAG_i 失效,通告其 b-MAG 实现服务接管,当 LMA 接受 b-MAG 的服务迁移注册请求(MSSR)并为服务迁移 MN 更新绑定信息后,对迁移 MN 的服务恢复正常.忽略节点处理时间,服务迁移时间为 $T_{ss}^{\mathrm{MAG}}=t_{mp}+t_{ml}$.MAG 容错时间及其均值为

$$T_{ft}^{\mathrm{MAG}}=T_{\mathrm{fd}}^{\mathrm{MAG}}+T_{\mathrm{ss}}^{\mathrm{MAG}}=\theta+(N_{mh}-1)\times I_{hm}+t_{mp}+t_{ml};E[T_{ft}^{\mathrm{MAG}}]=(N_{mh}-1/2)\times I_{hm}+t_{mp}+t_{ml} \quad (1)$$

图 7 给出了当 t_{mp}=5ms,t_{ml}=20ms 时,MAG 容错时间均值 $E[T_{ft}^{\mathrm{MAG}}]$ 随心跳间隔周期 I_{hb}([20ms,70ms])和 N_{mh}(取值分别为 1~5)的变化情况. $E[T_{ft}^{\mathrm{MAG}}]$ 随 I_{hb} 呈线性增长,增长斜率由 N_{mh} 决定.N_{mh} 越大, $E[T_{ft}^{\mathrm{MAG}}]$ 随 I_{hb} 的增长率也越大.调节 I_{hb} 和 N_{mh} 两个容错参数,可以控制容错时间.N_{mh} 的建议取值为 2~5,具体取值应根据系统所部署网络的链路状况而定.在图 7 所示的容错参数取值范围内, $E[T_{ft}^{\mathrm{MAG}}]$ 可控制在 35ms~340ms.其中,N_{mh} 取 2 或 3,当 I_{hb} 取值在[20ms,50ms]时,可将 $E[T_{ft}^{\mathrm{MAG}}]$ 控制在 55ms~100ms 或 75ms~150ms.

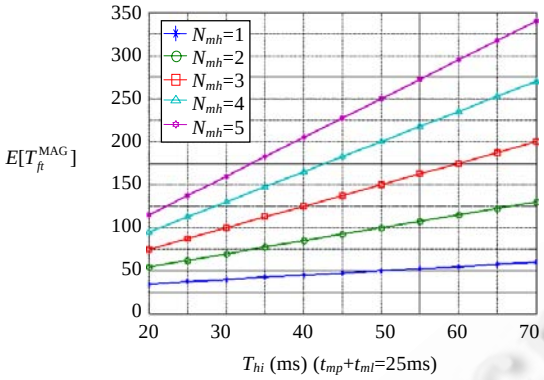


Fig.7 MAG fault-tolerant time

图 7 MAG 容错时间

3.3 TCP应用下的MAG容错性能分析

图 8 为 TCP 应用下 MAG 的容错性能分析模型.

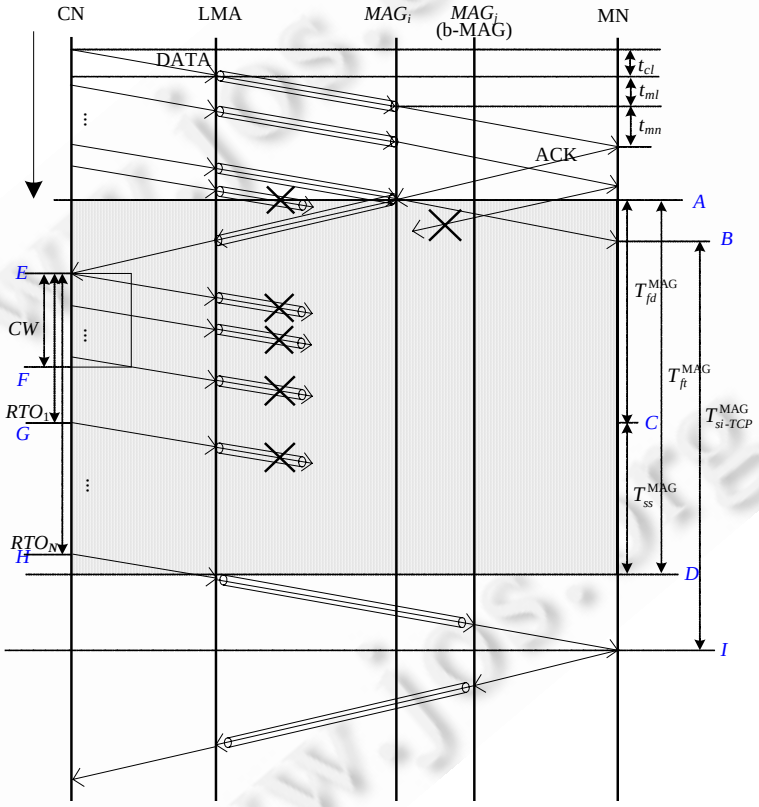


Fig.8 MAG fault-tolerant performance analysis model under TCP

图 8 TCP 应用下 MAG 容错性能分析模型

设 MAG_i 在时间点 A 失效,失效发生前,它通过与 LMA 间的双向隧道为 MN 转发数据包.MAG_i 在失效发生前,为 MN 转发的最后一个 CN→MN 方向数据包在时间点 B 到达 MN,最后一个 MN→CN 方向数据包在时间点

E 到达 CN.MAG_i 失效后,经过失效检测时间 T_{fd}^{MAG},MPM 在时间点 C 检测到其失效,触发 MAG 服务迁移;经过时间 T_{ss}^{MAG},在时间点 D,LMA 完成对 fa-MN 的移动注册信息更新.此后,LMA 将通过与 MAG_j 的双向隧道为 fa-MN 转发数据包.A 到 D 这段容错时间 T_{ft}^{MAG} 内,LMA 接收到的发往 fa-MN 的数据包将会丢失.

CN 在时间点 E 接收到最后一个 fa-MN 对在 MAG_i 失效发生前接收到的数据包的确认.CN 随后将向 fa-MN 发送拥塞窗口内(CW)的所有数据包,并等待 fa-MN 的确认.此时,可能有两种情况发生:

场景 1:CN 发送的拥塞窗口内的最后一个数据包(时间点 F 发送)到达 PMIPv6 域时失效,MAG 的服务迁移已经完成,即 $F+t_{cl}>D$.这种情况下,将有拥塞窗口内的数据包通过 LMA 和 MAG_j 到达 fa-MN,MAG_i 失效将不对 fa-MN 的上层 TCP 应用造成影响.

场景 2:CN 发送的拥塞窗口内的最后一个数据包到达 LMA 时,MAG 服务迁移尚未完成,即 $F+t_{cl}<D$.这种情况下,拥塞窗口内的所有数据包都将丢失,CN 不会接收到 fa-MN 为这些数据包回复的确认.当 CN 等待到时间点 G 时,它在时间点 E 发送的数据包的重传定时器到期(RTO_1).此时,CN 重传数据包,并将拥塞窗口置为 1.若该重传数据包到达 LMA 时,MAG 服务迁移仍未完成,该数据包仍将丢失.若第 N 次重传的数据包到达 LMA 时,服务迁移已完成,LMA 将通过 MAG_j 把数据包并转发至 fa-MN.此时,fa-MN 在 MAG_i 失效发生后首次成功接收到 CN 发送的数据包.fa-MN 接收到 MAG_i 在失效发生前为其发送的最后一个数据包(时间点 B)到 MAG 容错处理完成后 fa-MN 接收到第 1 个数据包(时间点 I)之间的时间间隔定义为 TCP 应用造成的服务中断时间 $T_{si\text{-}TCP}^{\mathrm{MAG}}$.

失效 MAG 服务迁移完成(时间点 D)后,CN 和 MN 通过 MAG_j 恢复正常通信.CN 在时间点 H 后开始 TCP 慢启动,直到拥塞窗口从 1 增大至 CW.

根据 TCP 慢启动机制,拥塞窗口增大到 CW 的恢复时间为 $\tau_s=[\log_2(1+CW)-1]\times RTT$($CW$ 取 20 时,$\tau_s=4RTT$).MAG_i 失效造成的 TCP 吞吐量下降时间为

$$T_{td\text{-}TCP}^{\mathrm{MAG}}=(H-B)+\tau_s+(t_{cl}+t_{ml}+t_{mm}),$$

其中,$(H-B)=(H-E)+(E-A)-(B-A)$,$(E-A)=t_{cl}+t_{ml}$,$(B-A)=t_{mm}$.吞吐量下降时间 $T_{td\text{-}TCP}^{\mathrm{MAG}}$ 与服务终端时间 $T_{si\text{-}TCP}^{\mathrm{MAG}}$ 相比主要增加了拥塞窗口恢复时间 τ_s.$(H-E)$由重传次数 N 确定.按照 TCP 重传机制,初始重传定时器值 $TO_1=\beta\times RTT$,RTT 为 TCP 连接的平均往返时延.第 N 次重传的定时器值 TO_N($N\geqslant 1$)为

$$TO_N=\begin{cases}\gamma^{N-1}TO_1, & \text{if } N\leqslant s\\ \gamma^{s-1}TO_1, & \text{if } N>s\end{cases} \tag{2}$$

那么,$(H-E)$,即从第 1 次发送到第 N 次重传某数据包的总时间为

$$\begin{aligned}(H-E)=RTO_N=\sum_{i=1}^{N}TO_i&=\begin{cases}TO_1+\gamma TO_1+\ldots+\gamma^{N-1}TO_1, & \text{if } N\leqslant s\\ TO_1+\gamma TO_1+\ldots+\gamma^{s-1}TO_1+(N-s)\gamma^{s-1}TO_1, & \text{if } N>s\end{cases}\\&=\begin{cases}TO_1\times((\gamma^N-1)/(\gamma-1)), & \text{if } N\leqslant s\\ TO_1\times((\gamma^s-1)/(\gamma-1))+(N-s)\gamma^{s-1}TO_1, & \text{if } N>s\end{cases}\end{aligned} \tag{3}$$

重传次数 N 由容错时间 T_{ft}^{MAG} 确定:

$$RTO_{N-1}+(E-A)+t_{cl}<T_{ft}^{\mathrm{MAG}}\leqslant RTO_N+(E-A)+t_{cl}\Leftrightarrow RTO_{N-1}+2t_{cl}+t_{ml}<T_{ft}^{\mathrm{MAG}}\leqslant RTO_N+2t_{cl}+t_{ml} \tag{4}$$

MAG 失效造成的吞吐量下降时间为

$$T_{td\text{-}TCP}^{\mathrm{MAG}}=\begin{cases}0, & \text{if } T_{ft}^{\mathrm{MAG}}\leqslant 2t_{cl}+t_{ml}\\ RTO_1+2(t_{cl}+t_{ml})+\tau_s, & \text{if } 2t_{cl}+t_{ml}<T_{ft}^{\mathrm{MAG}}\leqslant RTO_1+2t_{cl}+t_{ml}\\ RTO_N+2(t_{cl}+t_{ml})+\tau_s, & \text{if } RTO_{N-1}+2t_{cl}+t_{ml}<T_{ft}^{\mathrm{MAG}}\leqslant RTO_N+2t_{cl}+t_{ml}\ (N\geqslant 2)\end{cases} \tag{5}$$

图 9 给出了 t_{mr} 分别取值 10ms,50ms,150ms(WLAN、3G、卫星通信网络中的典型值)时,MAG 失效造成的吞吐量下降时间 $T_{td\text{-}TCP}^{\mathrm{MAG}}$ 随 MAG 容错时间 T_{ft}^{MAG} 的变化情况.如图 9 所示,随着 T_{ft}^{MAG} 的增大,$T_{td\text{-}TCP}^{\mathrm{MAG}}$ 阶梯性上升,上升的幅度逐次递增,为相邻两次重传定时器到期时间的差值.因为每当 T_{ft}^{MAG} 跨越某值,使得 CN 的重传次数增大,服务中断时间就增大一个重传超时时间.t_{mr} 越大,同等次数的 $T_{td\text{-}TCP}^{\mathrm{MAG}}$ 上升幅度就越大.这是因为每次的超时时

间与 CN-MN 间的平均往返时延相关,t_{mr} 越大使得超时时间越长,T_{ft}^{MAG} 的上升幅度也就越大.针对具体网络环境,调节容错参数 I_{hb} 和 N_{hb},可以有效降低 LMA 失效对 fa-MN 上层 TCP 应用造成的影响.根据上节分析,容错参数在建议取值范围内可将 T_{ft}^{MAG} 控制在 35ms~340ms.如图 9 所示,理论上,当 T_{ft}^{MAG} 小于 120ms 时,即前文分析的场景 1,MAG 失效不对 fa-MN 上层 TCP 应用造成影响.最差情况是,对分别处于 WLAN、3G 和卫星网络中的 fa-MN 而言,MAGFT 也可将其 TCP 应用吞吐量下降时间控制在 1.1s,1.6s 和 2.8s 左右.

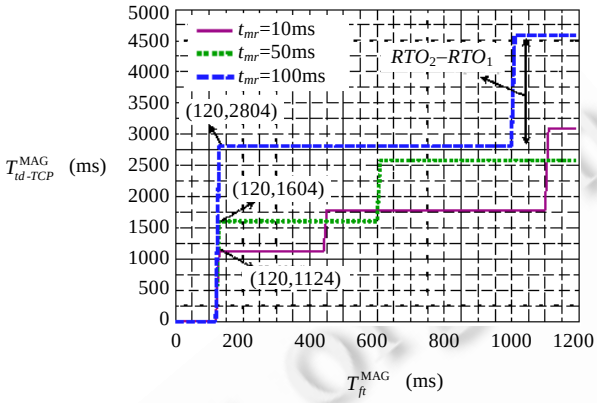


Fig.9 TCP throughput degradation time caused by MAG failure

图 9 MAG 失效造成的 TCP 吞吐量下降时间

3.4 UDP应用下的MAG容错性能分析

图 10 为 UDP 应用下的 MAG 容错性能分析模型.设 MAG_i 在时间点 A 失效,MAG_i 在失效发生前为 MN 转发的最后一个数据包在时间点 B 到达 MN.MAG_i 失效后,经过失效检测时间 T_{fd}^{MAG},MPM 在时间点 C 检测到 MAG_i 失效,在时间点 D 完成服务迁移.此后,LMA 将通过与 MAG_j 的双向隧道为 MN 转发数据包.CN 在时间点 E 到 F 之间向 MN 发送的数据包无法到达 MN.设 CN 的发包率为 λ_p,则 MAG_i 失效造成 MN 的 UDP 应用丢包为

$$N_{pl-UDP}^{MAG}=\lambda_p\times(F-E)=\lambda_p\times[(D-A)-(D-F)+(A-E)]=\lambda_p\times[T_{ft}^{MAG}-t_{cl}+(t_{cl}+t_{ml})]=\lambda_p\times(T_{ft}^{MAG}+t_{ml}) \quad (6)$$

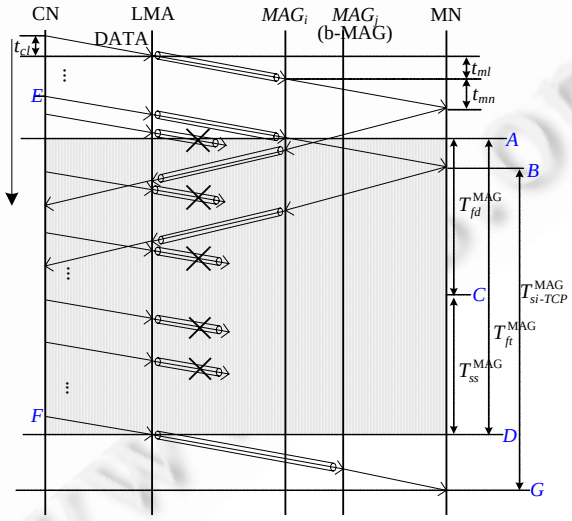


Fig.10 MAG fault-tolerant performance analysis model under UDP

图 10 UDP 应用下 MAG 容错性能分析模型

图 11 为 t_{ml}取值 20ms、CN 发包率λ_p分别为 20/s,50/s,100/s 时, $N_{pl\text{-}UDP}^{MAG}$随 MAG 容错时间 T_{fd}^{MAG} 的变化情况. $N_{pl\text{-}UDP}^{MAG}$随 T_{fd}^{MAG} 呈线性增长,增长斜率由 CN 的发包率决定.当 MAGFT 的容错时间控制在 55ms~150ms 时,对于发包率分别为 20/s,50/s,100/s 的 UDP 应用,可将 MAG 失效造成的丢包个数控制在 1~3,3~8,7~17 范围内.

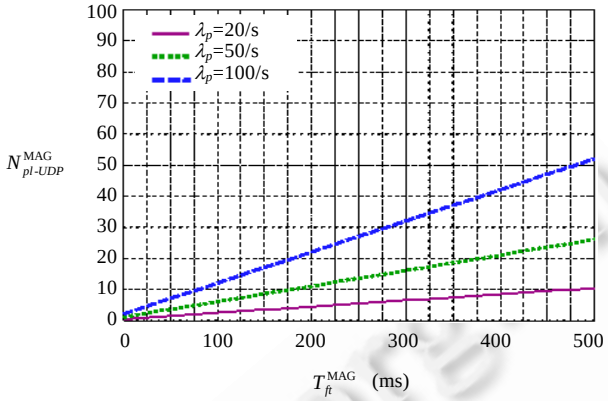


Fig.11 UDP packet loss caused by MAG failure

图 11 MAG 失效造成的 UDP 应用丢包

4 MAGFT 方案系统仿真

本文在 NS-2 中,按照协议标准实现了 PMIPv6 网络仿真平台,并在此基础上进一步实现了 MAGFT 方案.本节通过仿真分析,从 TCP 和 UDP 应用下的 MAG 容错性能、MAG 容错信令开销、MN 接入延时等方面,对 MAGFT 方案进行评价.

4.1 仿真场景及参数设置

MAGFT 方案仿真场景设置如图 12 所示,在一个 PMIPv6 域内部署 3 个 LMA、3 个 MAG 池.每个 MAG 池中部署 3 个 MAG,其中一个 MAG 同时作为 MPM.通过两个路由器相互连接各 LMA 和 MAG.MN 在各 MAG 间随机移动,通过无线链路与当前附着 MAG 建立连接.在仿真中,分别使用了 802.11g、UMTS 和卫星网络 3 种无线接入方式,其链路延时分别为 10ms,50ms,150ms.分别仿真 MN-CN 间建立 TCP 应用、运行 UDP 应用两种场景下的 MAG 失效,分析 MAGFT 方案在两种应用下的容错性能.同时,统计方案引入的容错信令开销,分析 MN 的接入延时.

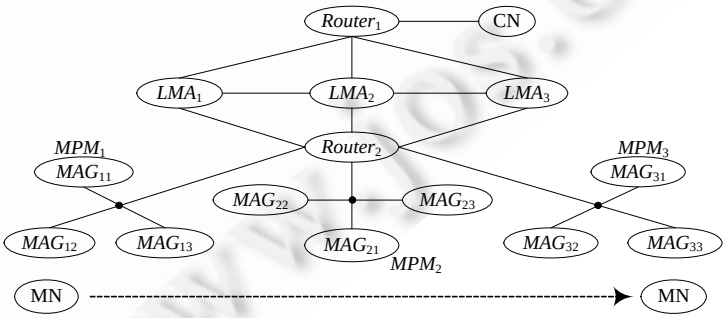


Fig.12 MAG simulation scenario

图 12 MAGFT 仿真场景

仿真参数设置见表 2.

Table 2 Simulation parameter description

表 2 仿真参数说明

Number of MNs	100
MN's mobility rate	0.01~10 (m/s)
CN-Router link	The bandwidth is 100Mb, the one-way link delay is 40ms
LMA-Router link	The bandwidth is 100Mb, the one-way link delay is 10ms
MAG-Router link	The bandwidth is 100Mb, the one-way link delay is 10ms
MN-MAG wireless link	802.11g, UMTS, and satellite, the one-way delay are 10ms, 50ms and 150ms, respectively
MAG-MPM link	802.3, the bandwidth is 100Mb, the one-way link delay is 5ms
Simulation duration	50s
The time that the MAG-failure takes place	The 25th second
The steady state value of the TCP congestion window (*CW*)	20
The initial value of the TCP's retransmission timeout	200ms
The UDP's packet transmission rate	25kbps

4.2 TCP应用下的MAG容错性能

图 13 给出了分别运行 PMIPv6 基本协议(base-PMIPv6)和引入 MAGFT 方案,MN-MAG 间无线链路分别为 802.11g、UMTS、卫星网络的 6 种场景下,50s 仿真时间内 fa-MN 的 TCP 吞吐量变化情况.其中:心跳间隔 I_{hb} 取值为 50ms;心跳消息允许丢失数 N_{mh} 取值为 3;MAG 失效发生在第 25s.横坐标是仿真时间,纵坐标是 MN 的收包率(packets/s).如图 13 所示:在第 5s 左右,各场景下的 TCP 应用吞吐量达到稳定状态;同种无线链路场景下,Base- PMIPv6 和 MAGFT 方案的吞吐量值相同.在第 25s,MAG 失效发生后,运行 Base-PMIPv6 的 3 种场景下的 fa-MN 吞吐量值都迅速降为 0,且不再恢复.MAGFT 方案的 3 种场景下,fa-MN 的吞吐量值在失效发生后有较明显的下降,但很快恢复至失效发生前的稳定值.

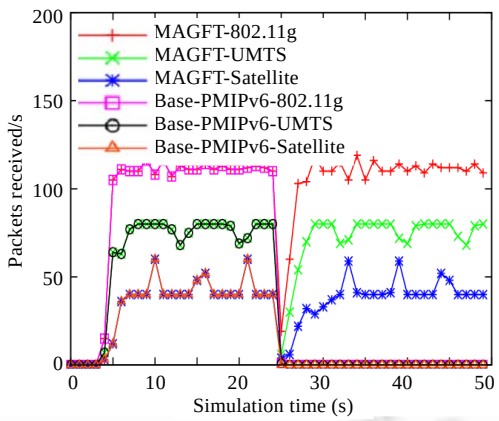


Fig.13 UDP throughput during MAG failure

图 13 MAG 失效时的 UDP 收包率

4.3 UDP应用下的MAG容错性能

图 14 给出了分别运行 Base-PMIPv6 和 MAGFT 方案两种场景下,50s 仿真时间内 fa-MN 上层 UDP 应用的收包率.其中,I_{hb} 取值为 50ms,N_{mh} 取值为 3.由于 MAG 失效对 fa-MN 上层 UDP 应用的影响与 fa-MN 所在无线链路无关,因此图中仅给出了 MN-MAG 间无线链路为 802.11g 的场景.在 MAG 失效发生之前,Base-PMIPv6 和 MAGFT 两种场景下,fa-MN 的收包率变化情况是一致的,都在第 5s 左右达到稳定状态,为 25packets/s.第 25s,LMA 失效发生后,Base-PMIPv6 场景下的 fa-MN 收包率立即降为 0,且不再恢复.MAGFT 场景下,fa-MN 丢包率在第 26s 下降到 20packets/s 左右,在第 27s 已恢复至失效发生前的稳定值,整个容错过程的丢包在 10 个以下.

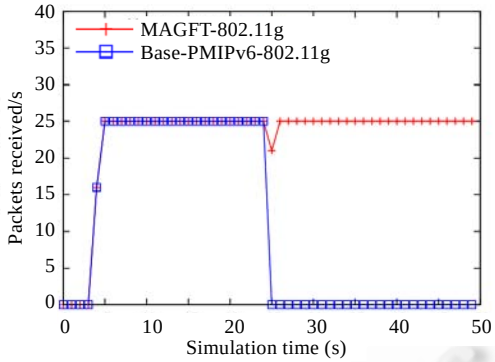


Fig.14 UDP throughput during MAG failure

图 14 MAG 失效时的 UDP 应用收包率

4.4 MAG容错信令开销

图 15 为 50s 仿真时间内的 LMA 容错信令开销和 PMIPv6 基本信令开销.左图为 I_{hb} 取值在 20ms~500ms 变化时的容错信令和 PMIPv6 基本信令开销,其中,系统所服务的 MN 数目为 100.方案引入的 MAG 容错信令开销是 I_{hb} 的反比例函数,随着 I_{hb} 的增大,容错信令开销急剧减小.图中,当 I_{hb} 取值 50ms 时,容错信令开销约为 PMIPv6 基本信令开销的一半.右图为容错信令和 PMIPv6 基本信令开销随系统所服务的 MN 数目变化的情况,其中,I_{hb} 取值为 50ms.如图 15 所示,容错信令和基本信令开销都随 MN 数目的增加呈线性增长.这时,因为 MAGFT 中各 MAG 需向 b-MAG 同步备份移动管理信息,MN 越多,信息备份信令越多.然而,容错信令始终小于基本信令,且增长率更小.图 15 所显示的是系统处于较空闲状态的情况,容错信令开销和 PMIPv6 基本信令开销具有一定的可比性.在实际网络部署中,9 个 MAG 的部署规模所能支持的 MN 远大于 300.随着 MN 数目的增多,当系统处于较饱和的稳定服务状态时,容错信令开销将远小于基本信令.

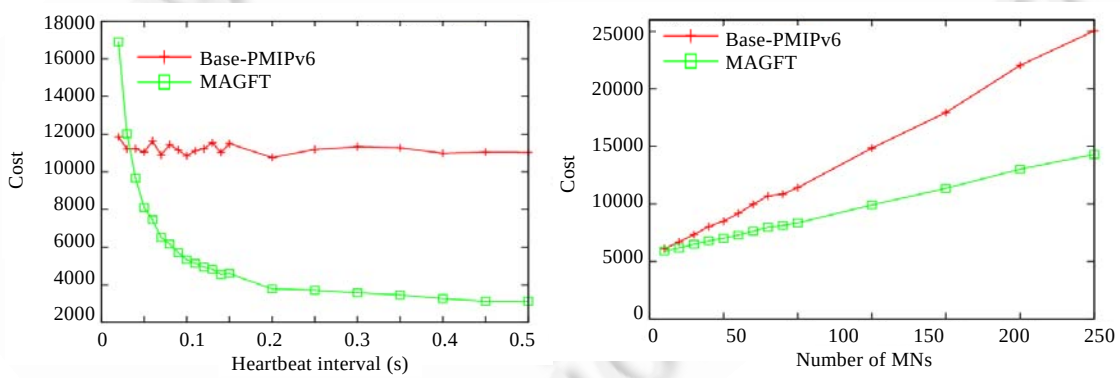


Fig.15 signaling cost (MAGFT vs. PMIPv6)

图 15 信令开销(MAGFT vs. PMIPv6)

4.5 引入MAGFT方案后的MN接入延时

图 16 为在不同 MN 数量情况下,引入 MAGFT 方案和运行 PMIPv6 基本协议两种场景下的 MN 接入平均延时.随着 MN 数量的增加,两者的平均接入延时都缓慢增加.与运行 PMIPv6 基本协议相比,在引入 MAGFT 方案后, MN 的接入延时略有增大.这主要是因为 MN 接入请求需经过 MPM 转发至 MAG,而后触发 PMIPv6 注册过程,增加了 MPM-MAG 间的传输延时.但增加的接入延时保持在 10ms 以下,且随着 MPM-MAG 间链路延时的

缩短,该增值会更小.

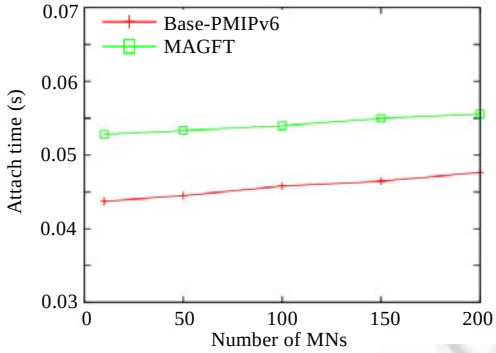


Fig.16 Access delay (MAGFT vs. PMIPv6)

图 16 接入延时(MAGFT vs.PMIPv6)

5 全文总结及下一步工作

本文提出一种基于池的 MAG 容错方案(MAGFT).方案引入 MAG 池,解决 PMIPv6 系统中的 MAG 服务不可替代问题.针对 PMIPv6 系统所部署的下层网络的不同,可分别采用覆盖范围无重叠区部署和有重叠区部署两种模式,在 PMIPv6 域内构建多个 MAG 池,使得域内各 MAG 至少归属于一个池.当某 MAG 失效时,它所在池内的某一有效 MAG 将快速接管其服务,实现对 MN 透明的 MAG 容错.理论分析和仿真实验结果表明: MAGFT 方案容错效率高,当某 MAG 失效后,系统能够快速恢复相关 MN 的网络连接,并有效减小对 MN 上层应用造成的影响.方案容错开销小,当系统处于较饱和的稳定服务状态时,容错开销相比系统基本信令开销可以忽略.MAGFT 的引入对 MN 接入延时略有影响,但增加的延时在 10ms 以下.

在 PMIPv6 系统中,移动管理服务是由 LMA 和 MAG 协同实现的,某 MAG-LMA 间链路失效也将导致系统部分服务失效.因此,MAG-LMA 链路容错将是下一步的研究重点.

References:

[1] Gundavelli S, Leung K, Devarapalli V, Chowdhury K, Patil B. Proxy mobile IPv6. IETF RFC 5213, 2008.

[2] Johnson D, Perkins C, Arkko J. Mobility support in IPv6. IETF RFC 3775, 2004.

[3] Moskowitz R, Nikander P. Host identity protocol (HIP) architecture. IETF RFC 4423, 2006.

[4] Eronen P. IKEv2 mobility and multihoming protocol (MOBIKE). IETF RFC 4555, 2006.

[5] Soliman H, Castelluccia C, El Malki K, Bellier L. Hierarchical mobile IPv6 mobility management (HMIPv6). IETF RFC 4140, 2005.

[6] Ramjee R, Varadhan K, Thuel SR, Wang SY, La Porta T. HAWAII: A domian-based approach for supporting mobility in wide-area wireless networks. IEEE/ACM Trans. on Networking, 2002,10(3):396−410. [doi: 10.1109/TNET.2002.1012370]

[7] Valko A. Cellular IP: A new approach to Internet host mobility. ACM SIGMOBILE Computer Communication Review, 1999,29(1): 50−65. [doi: 10.1145/505754.505758]

[8] Third Generation Partnership Project—Technical specification group services and system aspects—Architecture enhancements for non-3GPP accesses (release 8). 2008.

[9] WiMAX forum network architecture—Stage 2: Architecture tenets, reference model and reference points [part 2]. 2007.

[10] Kong KS, Lee WJ, Han YH, Shin MK. Handover latency analysis of a network-based localized mobility management protocol. In: Proc. of the IEEE Int'l Conf. on Communications (ICC 2008). Beijing: IEEE Press, 2008. 5838−5843. http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4534128&queryText%3D.+Handover+latency+analysis+of+a+network-based+localized+m

obility+management+protocol%26openedRefinements%3D*%26filter%3DAND%28NOT%284283010803%29%29%26searchField%3DSearch+All [doi: 10.1109/ICC.2008.1092]

[11] Li Y, Kum DW, Seo WK, Cho YZ. A multihoming support scheme with localized shim protocol in proxy mobile IPv6. In: Proc. of the IEEE Int'l Conf. on Communications (ICC 2009). Dresden: IEEE Press, 2009. 1–5. http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=5198634&queryText%3DA+multihoming+support+scheme+with+localized+shim+protocol+in+proxy+mobile+IPv6.+In%3A+Proc.+of+the+IEEE+Int%E2%80%99l+Conf.+on+Communications%26openedRefinements%3D*%26filter%3DAND%28NOT%284283010803%29%29%26searchField%3DSearch+All [doi: 10.1109/ICC.2009.5198634]

[12] Jeong SJ, Shin MK, Kim HJ. Implementation of route optimization mechanism supporting IPv4/IPv6 traversal in proxy mobile IPv6. In: Proc. of the IEEE 11th Int'l Conf. on Advanced Communication Technology (ICACT 2009). Phoenix Park: IEEE Press, 2009. 1242−1244. http://ieeexplore.ieee.org/search/searchresult.jsp?newsearch=true&queryText=Implementation+of+route+optimization+mechanism+supporting+IPv4%2FIPv6+traversal+in+proxy+mobile+IPv6&x=73&y=16

[13] Kong KS, Lee WJ, Han YH, Shin MK, You KR. Mobility management for all-IP mobile networks: Mobile IPv6 vs. proxy mobile IPv6. IEEE Wireless Communication, 2008,15(2):36−45. [doi: 10.1109/MWC.2008.4492976]

[14] Devarapalli V, Koodli R, Lim H, Kant N, Krishnan S, Laganier J. Heartbeat mechanism for proxy mobile IPv6. IETF RFC 5847, 2010.

[15] Akan OB, Akyildiz IF. ATL: An adaptive transport layer suite for next generation wireless Internet. IEEE Journal on Selected Area in Communications (JSAC), 2004,22(5):802−817. [doi: 10.1109/JSAC.2004.826919]

[16] Mohanty S, Akyildiz IF. Performance analysis of handoff techniques based on mobile IP, TCP-migrate, and SIP. IEEE Trans. on Mobile Computing, 2007,6(7):731−747. [doi: 10.1109/TMC.2007.1040]

张瀚文(1981－),女,四川成都人,博士,助理研究员,CCF 会员,主要研究领域为移动网络,可信网络.

许智君(1974－),男,讲师,主要研究领域为可信网络.

周继华(1979－),男,博士,高级工程师,主要研究领域为宽带无线通信,移动网络.

张玉军(1976－),男,博士,副研究员,CCF 高级会员,主要研究领域为移动网络,可信网络.

李忠诚(1962－),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为计算机网络.