

面向用户的普适计算系统可用性度量模型^{*})

王海鹏 周兴社 张涛 向冬

(西北工业大学计算机学院 西安 710072)

摘要 普适计算模式追求随时随地的计算境界,对相关可用性管理提出了很高的要求。现有研究大多基于定性的角度研究普适计算系统的可用性问题,本文尝试给出定量的可用性度量模型和分析。普适计算系统可用性具有典型的面向用户特点,为刻画用户行为需要引入具有非指数分布性质的模型状态,并且需要考虑用户态与系统态关联分析时的复杂情况。传统的连续时间马尔可夫链和半马尔可夫过程都不能很好地对以上情况给与分析。鉴于此,本文提出了一种基于马尔可夫重生过程(MRGP)的普适计算系统可用性度量方法。首先建立具有双扩展特点的用户模型,基于此构建了普适计算系统可用性度量的MRGP模型,对可用性进行了随机过程分析,给出了对应瞬态和稳态行为的度量方法,结合普适计算模式的特点定义了普适计算系统请求可用性度量。最后进行了数值分析,评价了用户和系统参数等因素对普适计算系统可用性度量的影响。

关键词 可用性度量,MRGP,用户模型,普适计算

User Oriented Measurement Model for the Availability of Pervasive Computing Systems

WANG Hai-Peng ZHOU Xing-She ZHANG Tao XIANG Dong

(School of Computer Science, Northwestern Polytechnical University, Xi'an 710072)

Abstract With the vision of anytime anywhere computing, the availability management is playing an important role in pervasive computing. Most research focus on the qualitative analysis. This paper adopts a quantitative approach to attack this topic. Due to the CTMC and SMP can not be applied to the modeling of the availability of pervasive computing system, which has the features of user oriented, non-exponential distributions and associative analysis between user states and system states, this paper presents an MRGP(Markov regenerative process)-based measurement for the availability of pervasive computing system. First, a bidirectional extended user model is developed, which provides a base point for describing the availability in pervasive computing. Second, an MRGP model for availability is established, and a stochastic process-based analysis is conducted for the availability of pervasive computing system. The transient and steady availability analysis are provided for this model, and a novel availability metric, called request availability of pervasive computing system, is suggested. Finally, numerical analysis has been carried out and the results argue that the proposed approach and availability metric can give reasonable evaluation for pervasive computing system.

Keywords Availability, MRGP, User model, Pervasive computing

1 引言

普适计算模式的出现和快速发展是近年来计算和通信领域产生的最为深刻的变化之一。普适计算追求随时随地的计算境界,资源是承载计算的物质基础,因此这对资源的可用性管理提出了很高的要求。现有研究大多基于定性的角度研究普适计算系统的可用性问题,随着研究的进一步深入,如何对普适计算系统的可用性给出定量的分析和度量成为现阶段发展的一个重要方向^[1~3]。

随时随地的可用性刻画了普适计算系统的可用性特点,这种特性要求普适计算系统在任何时间和任何地点向用户提供所需服务的能力。普适计算环境是一种更加复杂的分布式系统,其资源的分散性、移动性和异构性尤为突出,在此基础上要提供随时随地的可用性成为一个具有相当挑战性的课题。现有的实际情况是,普适计算环境包含的资源类型和规模不断扩大,但系统实际能够提供的可用性并未得到相应的提高,相反却随系统规模增长而下降^[4,5]。鉴于此,非常需要对普适计算系统可用性组成的主要因素和变化原因进行分析,以指导系统的设计、管理和维护,以达到普适计算随时随地可用性的要求。

普适计算模式以用户为计算主体,各种计算资源围绕用户而展开计算。因此,普适计算系统更加需要考虑用户在可用性建模中的作用,要考虑从用户角度出发感知到的可用性。其次,普适计算在资源管理模式上也有不同之处,即虽然用户任务执行依赖于不同资源,但由于资源的分散、自治和移动等特点,只有在任务最终执行时刻才能有效地确定资源可用性情况。基于此,在普适计算环境下,用户需要依赖于资源管理服务间接获得最终资源的可用性,资源管理服务一般包括资源发现、资源分配和资源调度等。一般来讲,用户、资源管理服务和资源的以上关系被描述为一种三角依赖关系,它是传统用户任务直接关联关系的扩展。因此普适计算环境大多将资源管理服务作为其基础设施而构建,其结果就是资源管理服务或系统成为资源可用性的提供者和管理者,并且实现资源与用户的解耦合机制。

普适计算环境的以上特点对可用性研究的意义在于:普适计算系统可用性是一类广义的用户与系统的关联可用性^[7],需要对用户与系统因素加以综合考虑,不但要分析系统因素对可用性的影响,而且要考虑从用户角度出发的可用性度量。资源管理系统成为影响可用性度量的重要因素。从可用性意义上讲,资源管理系统演化为可用性管理系统。用户

^{*} 本课题得到国家自然科学基金(60573161),国家“八六三”高技术研究发展计划项目基金(2003AA1Z2100)和教育部博士点基金的资助。
王海鹏 博士生,主要研究领域为普适计算、分布计算、资源管理。周兴社 博士生导师,教授,研究方向为分布计算和嵌入式计算。张涛 博士生,研究领域为软件工程、普适计算。向冬 博士生,研究领域为计算机安全、普适计算。

与系统的交互,首先依赖于资源管理系统对动态时效资源的有效管理,其次才依赖于任务承载资源的可用性。由于资源管理系统一般作为基础设施而构建,其稳定性和可用性远高于一般资源;而且面对资源失效,资源管理系统能够执行资源替代,提供了一种可用性补偿或反馈机制,最终提供给用户一个高可用的系统。

虽然普适计算系统可用性研究仍处发展之中,但相关领域已经取得了部分研究成果,提供了一定的参考意义。例如 Web 系统、网络计算系统、面向系统、面向应用的可用性研究等^[1~6]。上述研究大多是从系统或应用的角度出发,以描述系统自身或面向不同应用的可用性为主。对于普适计算而言,更加关心从用户角度出发的可用性描述。例如,系统可用性度量对于只使用部分功能的用户而言并不具多大意义,而应用可用性度量也难以对跨越多应用的用户会话给出准确的描述。

基于以上分析,本文研究了面向用户的普适计算系统可用性度量。首先建立了描述用户行为的双扩展用户模型,在此基础上提出了一个基于马尔可夫重生过程(MRGP)的普适计算系统可用性度量模型,对可用性进行了随机过程分析,给出了对应瞬态和稳态行为的度量方法,结合普适计算模式的特点定义了普适计算系统请求可用性度量。最后进行了数值分析,评价了用户和系统参数等因素对普适计算系统可用性度量的影响。

2 双向扩展用户模型

首先给出对普适资源概念的一个定义。广义上讲,普适计算环境中的资源管理子系统和各种软硬件资源(设备和服务)都可纳入普适资源概念的范畴。但其中,资源管理子系统一般作为系统的基础设施而构建,具有对用户可见、相对稳定和具有高可用性的特点;而各类的软硬件资源,则具有对用户透明、变化频繁和可用性较低的特点。为了讨论方便,本文将后者专称为普适资源,以强调资源的用户透明性、动态性和分散异质等特点。

普适资源可用性具有如下特点:

- 面向用户。普适计算可用性具有面向用户的特点,即根据用户不同任务和不同需求,需要提供相关的资源可用性,因此不同于系统可用性或应用可用性的概念。

- 非完备性。用户不要求系统所有资源在使用时都处于可用状态。这一方面由面向用户的特点决定,另一方面受系统的动态性、移动性等限制,因此研究面向用户的可用性更为恰当。

- 时效性。在用户执行任务过程中,系统无需提供全程的资源可用性,只需在某些时段上处于可用状态。这是由用户状态空间的特点所决定。例如用户处于资源请求态时,对资源可用性有需求;而当用户处于思考态时,则对资源可用性不作要求。

用户模型主要对用户与系统之间的交互行为进行建模,因此对于关联用户因素的可用性问题,用户模型是可用性建模的一个关键。根据普适计算模式的交互特点,将用户行为分为如下三类,

- 第一类是普适资源请求行为。描述用户向所需资源发送请求和执行交互的行为,例如各种计算资源结点请求或信息资源请求等;对应该行为用户所处的状态称为资源请求状态,用表 P^r 示。该状态从用户发送请求开始,到请求处理执

行完毕或失败中止结束。

- 第二类是资源管理或服务请求行为。描述用户请求资源管理系统提供资源管理服务的行为,例如,当 P^r 状态中发生资源失效时,请求替代资源的再发现或资源的重分配和调度等;对应该行为用户所处的状态称为资源管理请求状态或服务请求状态, P^m 用表示。

- 第三类是用户思考行为。描述不直接产生资源或资源管理请求的行为。例如,用户的思考时间或用户浏览网页的时间等;对应该行为用户所处的状态称为思考状态,用 T 表示。该状态从用户执行思考行为开始,到发起下一个资源或服务请求结束。

用户处于以上不同的状态,对普适计算系统有不同的可用性需求。在资源请求状态中,要求用户当前任务所依赖的普适资源要处于可用状态,但对资源管理系统不作要求;思考状态中,用户对普适计算系统的可用性不作要求;资源管理请求状态中,用户对普适资源和资源管理系统的可用性都有一定要求。以上因素都是普适计算系统可用性度量中需要加以考虑的问题。

图 1 给出了双扩展用户模型。图中结点表示用户状态,结点间连线表示状态迁移。用户状态包含普适资源请求状态 P_i^r , 思考状态 T_i , 服务请求状态 P_i^m , 分别对应上述三类用户行为和完成态 E 及失效态 F 。其中 E 和 F 是表示会话正常结束或非正常中止状态(均为吸收态)。状态标识 P_i^r 中 r_i 表示随用户请求的变化,所依赖的资源具有不同的状态;同理,状态标识 P_i^m 中 m_i 表示随用户请求的变化,所依赖的资源管理系统具有不同状态。在该模型中,当用户发出资源请求,如果所需资源可用,则进入 P_i^r 态;完成该请求后,进入 T_i 态;如果无法进入或 P_i^r 态发生资源失效,则进入 P_i^m 态;如果在 P_i^m 态能够完成资源的再发现和重分配等资源管理活动,则返回 P_i^r 态,否则进入 F 态。如果一次会话中所有的请求 P_i^r 都得到正确处理,则由 T_i 态进入到完成态 E ;如果存在任何一个请求失败,则由 P_i^m 态进入到失效态 F 。

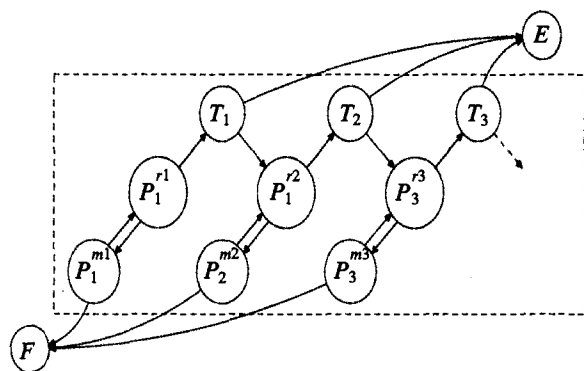


图 1 双向扩展用户模型

以上用户模型是对基本用户模型和单扩展用户模型^[2]的扩展,即以状态 P^r 为中心进行双向扩展,使用用户思考态 T 作到状态 E 的状态路径扩展,使用服务请求态 P^m 作到吸收态 F 的状态路径扩展;因此称之为双向扩展用户模型,简称双扩展用户模型。

从可用性的角度分析,基本状态模型不考虑用户的思考状态,因此在用资源请求态替代思考态时,会过高估计系统应该提供的可用性;单扩展用户模型不考虑服务请求态,因此在用资源请求态 P^r 替代服务请求态 P^m 时,会过低估计系统能

够提供的可用性。

双扩展用户模型的特点是：

- 更加范化的用户模型。包含了基本用户模型和单扩展用户模型的特例情况。根据普适计算系统中普适资源和资源管理系统分离的情况，去除了普适资源请求状态和服务请求状态的耦合假设和限制。

- 系统失效依赖转移。转移到依赖于作为普适计算环境基础设施的资源管理系统，而不是直接执行任务的普适资源；由于基础设施的可用性远高于具有分散特点的普适资源结点，因此降低了用户交互失败的概率，提高了用户会话完成的概率。

- 系统成功依赖转移。转移到思考状态上，而不是资源请求状态（基本用户模型的情况）；同样提高了用户会话完成的概率（思考态对系统可用性不作要求），降低了用户交互失败的概率。

3 基于 MRGP 的普适计算系统可用性度量模型

普适计算系统可用性具有面向用户和系统关联的特点，对可用性的影响在于：用户行为不同于电子元器件需要非指数分布的模型状态描述^[1]，例如，用户思考态不能使用指数分布来准确描述。其二是要求能够支持用户状态内的系统状态变换，即用户状态驻留时，相关联的普适资源状态或资源管理状态能执行连续的变化。例如在用户处于普适资源请求态时，对应的普适资源可以在工作态和失效态间变换。这可以描述为一种马尔可夫重生性质，即重生点间状态的变化性质，其中用户态是重生点，在重生点间支持系统状态的变化行为^[7]。

基于以上分析，传统的连续时间马尔可夫链需要指数分布假设，而半马尔可夫过程假设所有的状态转换都是重生点，因此都不能很好地对普适计算系统可用性的情况给与恰当的分析。马尔可夫重生过程(MRGP)是包含上述随机过程的一类更加广泛的随机过程，其不但支持非指数分布的模型状态，而且可以基于重生理论建模用户和系统关联状态变化的复杂行为，因此本文提出基于 MRGP 的普适计算系统可用性度量方法。

3.1 基于 MRGP 的普适计算系统可用性度量模型

分析双扩展用户模型，其中 P^r 状态是其主要状态，以该状态为核心 T, P^m 状态作为附属状态出现。因此，完整的用户会话可以看做为一系列 P^r 状态之间的迁移，而以 P^r 状态为核心， P^r, T, P^m 间局部的状态迁移行为就组成了用户会话的基本单元，其相关可用性研究本文定义为普适计算系统请求可用性，也称为用户请求可用性。

根据双扩展用户模型并结合资源和资源管理子系统，定义随机过程 $\{Z(t), t \geq 0\} = \{(U(t), R(t), M(t)), t \geq 0\}$ 。其中随机变量 $U(t) \in \Omega_U = \{R, T, M, F\}$ 表示用户处于时间 t 的状态，其中 R, M 分别对应用户 P^r, P^m 状态；随机变量 $R(t) \in \Omega_R = \{U, D\}$ 表示普适资源在时间 t 的状态；随机变量 $M(t) \in \Omega_M = \{U, D\}$ 表示资源管理子系统在时间 t 的状态；其中 U, D 分别表示工作态(U_p)和失效态(Down)。

图 2 建立了普适计算系统可用性的状态模型。图中结点表示模型的状态，结点间连线表示状态迁移。每一个状态表示为一个三元组 $(U(t), R(t), M(t))$ ，模型的状态空间是 $\{(RU \cdot), (TU \cdot), (RD \cdot), (TD \cdot), (M \cdot U), (M \cdot D), (F \cdot D), (F \cdot U)\}$ 。图中用 \cdot 表示所对应的元组分量可取任意

值，例如 $(RU \cdot)$ 表示用户处于 P^r 状态，所依赖的普适资源处于工作态时，资源管理子系统可处于任何状态。资源管理子系统失效率为 λ_m ，修复率为 μ_m ；普适资源子系统失效率为 λ_r ，修复率为 μ_r 。当用户进入 P^r 状态后，随机活动一段时间进入思考态(从 $(RU \cdot)$ 到 $(TU \cdot)$)，该时间分布为 $F(t)$ ；在思考态驻留一段时间后返回到 P^r 态，该驻留时间分布为 $G(t)$ 。当用户处于 R 时普适资源发生失效(从 $(RU \cdot)$ 到 $(RD \cdot)$)，经过一段时间后普适资源子系统相关服务启动(从 $(RD \cdot)$ 到 $(M \cdot U)$)，时间分布为 $H(t)$ ；设普适资源的可用性为 A_R ，资源管理子系统在状态 $(M \cdot U)$ 随机驻留时间的均值为 $(M \cdot U) 1/\gamma$ ，则资源管理子系统发现新的可替代资源并且该资源可用的概率是 A_R ，该普适资源不可用的概率为 $1 - A_R$ ；当普适资源和资源管理子系统均失效后，模型迁移到 $(F \cdot \cdot)$ 状态，用户在该状态驻留一段时间后进行重试，返回到 $(M \cdot U)$ 态或 $(M \cdot D)$ ，驻留时间分布为 $R(t)$ ； $(M \cdot D)$ 态的驻留时间分布为 $T(t)$ 。

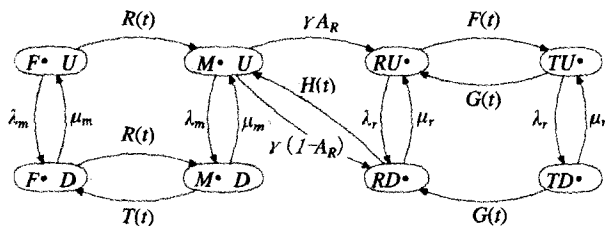


图 2 普适计算系统可用性度量 MRGP 模型

模型中 $F(t), G(t), H(t), R(t), T(t)$ 都是非指数的任意分布，并且模型中存在重生点间的状态变化，即存在局部行为，其中局部行为是指并发的指数分布迁移，即指数分布迁移连接的两状态具有相同的任意分布，包括 $(TU \cdot)$ 与 $(TD \cdot)$ 和 $(F \cdot U)$ 与 $(F \cdot D)$ 之间的迁移。除局部行为之外，每个状态改变都是重生点。分析上述模型的状态空间满足至多只有一个任意分布处于活动的条件，所以 $\{Z(t), t \geq 0\}$ 是马尔可夫重生过程(MRGP)^[7]。

普适计算请求可用性的 MRGP 模型由其 MRGP 随机过程的局部核 $E(t)$ 和全局核 $K(t)$ 所决定。其中局部核定义了重生点之间的行为，全局核定义了跨越重生点的行为。在局部核和全局核的基础上，瞬时可用性的求解基于式(1)，

$$V(t) = [V_{ij}(t)] = E(t) + K * V(t) \tag{1}$$

其中 $K * V(t)$ 是一个矩阵，其元素是 $\sum_r K_{rw} * V_w(t)$ ^[7]。稳态可用性的求解基于如下过程，

1. 计算 $\alpha = \int_0^{\infty} E(t) dt, \mu = \alpha \cdot e$ ，其中 e 是一个所有元素为 1 的列向量；
 2. 计算一步转移概率矩阵 $K(\infty)$ ，求得 $v = vK(\infty)$ ，其中 $v \cdot e = 1$ ；
 3. 稳态概率向量为 $P = (v \cdot \alpha) / (v \cdot \mu)$
- 限于篇幅本文不再给出局部核和全局核的推导。

3.2 普适计算系统请求可用性度量

基于以上度量模型，给出普适计算系统请求可用性的两个定义，狭义普适计算系统请求可用性 NUA 和广义普适计算系统请求可用性定义 GUA。这些定义都是基于用户角度出发的可用性度量。

定义狭义的普适计算系统请求可用性为，在不考虑用户 T 状态的情况下，用户能够成功发送请求的概率。定义如下：

$$NUA = \frac{P_{M,U} + P_{RU}}{P_{F,U} + P_{M,U} + P_{RU} + P_{F,D} + P_{M,D} + P_{RD}}$$

定义广义的普适计算系统请求可用性为,在 NUA 的基础上,综合考虑用户所有可能状态下,用户处于非失效状态 F 的概率。

$$GUA = \frac{P_{M,U} + P_{RU} + P_{TU} + P_{TD}}{P_{F,U} + P_{M,U} + P_{RU} + P_{TU} + P_{F,D} + P_{M,D} + P_{RD} + P_T}$$

以上定义中,NUA 和 GUA 的区别在于对用户、系统因素的折衷考虑。NUA 更加侧重系统因素,即忽略用户的思考状态;结果是简化了用户因素对可用性度量的影响,突出了系统因素在可用性度量中的作用,所以更加适合于偏重系统可用性评价的情况。例如,Web 系统中,评价 Web 服务器在线服务能力的情况,因为从 Web 服务器角度看到的是大量并发连续请求序列,而不考虑用户的思考行为,因此评价 NUA 就更加适合^[1]。GUA 更加侧重用户因素,即综合考虑用户的思考状态,能够对用户因素对可用性度量给出更加准确的分析。

通过以上定义,普适计算系统请求可用性(NUA 和 GUA)与系统可用性关系的一个定性分析是:

- 当系统处于失效状态时,用户可能没有向系统发送请求,因此用户感知不到相应的系统不可用性,即最终请求可用性大于系统可用性;
- 当系统从失效状态迁移到可用状态时,由于用户仍然滞留在不可用状态(F 状态)而导致用户对系统可用性感知的延迟,即请求可用性小于系统可用性,这和用户的重试率有紧密关联。

4 实验结果及分析

实验所使用的模型参数如表 2 所示。其中,根据 ON-OFF 模型的实验数据^[8],定义双扩展用户模型状态的非指数分布为,用户状态 P^r, P^m 服从 Weibull 分布,其形状参数 k 取值范围定义为 $[0.77, 0.91]$,规模参数取值范围定义为 $\theta \in [e^{4.4}, e^{4.6}]$ 。状态 T 服从 Pareto 分布,设定 Pareto 分布的形状参数 α ,门槛参数 m 和会话门槛参数 n 的取值分别为, $\alpha \in [0.5, 0.9], m=120, n=6000$ 。

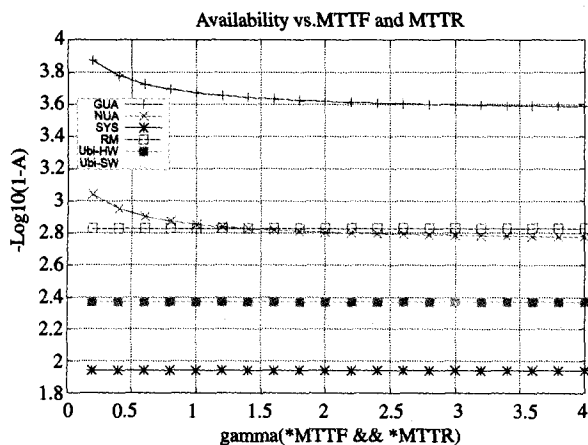


图 3 GUA/NUA vs. MTTF and MTTR

图 3 给出了当保持系统可用性不变时 GUA 和 NUA 的变化情况。图中纵轴为 $-\log_{10}(1-A)$ 表示可用性中 9 的数量,横轴表示系统 MTTF 和 MTTR 的倍乘因子 $\gamma(\text{gamma})$,

在 0.2 到 4 之间变化。为了便于比较,同时给出了资源管理子系统、普适硬件资源、普适软件资源和系统的稳态可用性。由于普适资源一般具有自治特点,可以假设资源管理子系统、普适硬件资源和普适软件资源之间具有随机独立性。系统可用性(SYS)定义为以上三部分全部处于可用状态的概率。由图可见,GUA 和 NUA 均随系统平均失效时间和平均修复时间的增加而减小,当 MTTF 和 MTTR 的值较小时下降幅度较大,而当 MTTF 和 MTTR 的值较大时变化幅度趋于平缓。其原因是 MTTR 对用户感知可用性的影响大于 MTTF 对用户感知可用性的影响。

表 2 模型参数

参数	缺省值	描述
$MTTF_m$	672hours	资源管理子系统 MTTF
$MTTF_{hw}$	168hours	普适硬件资源 MTTF
$MTTF_{sw}$	48hours	普适软件资源 MTTF
$MTTR_m$	3600seconds	资源管理子系统平均手工修复时间
$MTTR_{hw}$	2581seconds	普适硬件资源结点平均修复时间
$MTTR_{sw}$	1000seconds	普适软件资源平均重启时间
g	300seconds	思考时间
h	100seconds	重试时间

基于以上分析可得,

- 给定具有相同系统可用性的两个普适计算系统,其中具有较小平均失效时间和修复时间的普适计算系统能够提供较高的普适计算系统请求可用性;
- 系统平均修复时间对普适计算系统请求可用性的影响要大于平均失效时间;
- 系统的快速修复和恢复能力对于普适计算系统请求可用性影响比之系统提供长失效时间的能力更重要。

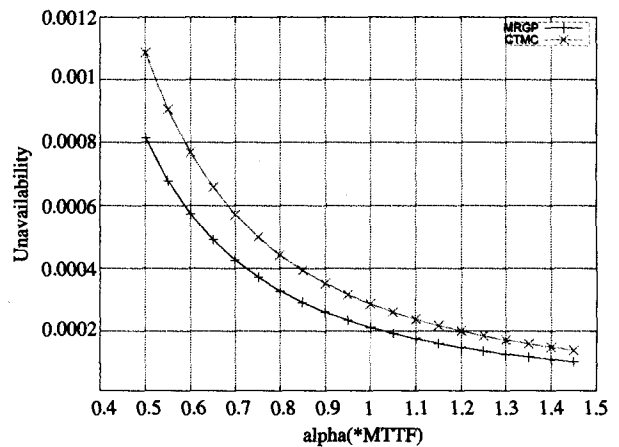


图 4 MRGP vs CTMC

图 4 给出了本文基于 MRGP 的分析结果和基于连续时间马尔可夫链(CTMC)分析结果的比较。其中 CTMC 模型的构建是通过将 MRGP 中的任意分布替代为具有相同均值的指数分布所得,即将本文的 MRGP 模型简化为对应的 CTMC 模型。图中纵轴为不可用性,横轴表示系统 MTTF 的倍乘因子 $\alpha(\text{alpha})$,在 0.5 到 1.45 之间变化。

如果定义 $\eta = |\bar{A}_{CTMC} - \bar{A}_{MRGP}| / \bar{A}_{MRGP}$,其中 $\bar{A}_{MRGP}, \bar{A}_{CTMC}$ 分别 MRGP 和 CTMC 模型下的普适计算系统请求不可用性的计算值,则由表 3 可见 CTMC 模型较之 MRGP 模型高估不可用性 33.25153%~35.45454%。

表 3 \bar{A}_{MRGP} vs. \bar{A}_{CTMC}

alpha(α)	\bar{A}_{MRGP}	\bar{A}_{CTMC}	η
0.50000	0.000815	0.001086	33.25153%
0.60000	0.000574	0.000768	33.79790%
0.70000	0.000426	0.000571	34.03755%
0.80000	0.000328	0.000442	34.75609%
0.90000	0.000261	0.000352	34.86590%
1.00000	0.000212	0.000287	35.37735%
1.10000	0.000127	0.000172	35.43307%
1.20000	0.000110	0.000149	35.45454%

结束语 本文对面向用户的普适计算系统可用性进行了定量分析。首先针对普适计算模式下用户行为的特点建立了双扩展用户模型,从用户请求层次分析了普适计算系统的可用性。基于此,建立了普适计算系统可用性的马尔可夫重生过程模型,模型取消了用户态内系统态不变的约束(即重生点间状态不变的约束),并且模型允许使用任意分布来刻画用户行为,因此该模型可以更加准确地描述用户请求可用性的变化。基于随机过程的方法对普适计算系统可用性进行了分析,给出了对应瞬态和稳态行为的度量方法,并以此为基础定义了普适计算系统的请求可用性度量。最后给出了实验分

析,对影响相关因素进行了讨论,结果表明该度量模型和方法能够较为恰当地评估普适计算系统可用性,为进一步的研究提供参考。

参考文献

- 1 Xie W, Sun H, Cao Y, Trivedi K S. Modeling of user perceived webserver availability. ICC2003, Anchorage, Alaska, May 2003, 1796~1800
- 2 Wang Dazhi Trivedi K S. Modeling user-perceived service availability. In: Intl. Service Availability Symposium (ISAS'05), 2005
- 3 Govindan R, Reddy A. An analysis of internet inter-domain topology and route stability. In: Proc. of the INFOCOM1997, 1997, 850~857
- 4 Kalyanakrishnan M, Iyer R K, Patel J U. Reliability of internet hosts: A case study from the end user's perspective. Computer Networks, 1999, 31:45~57
- 5 Kaaniche M, Kanoun K, Martinello M. A user-perceived availability evaluation of a Web based travel agency. In: Proc. of the 2003 Intl. Conf. on Dependable Systems and Networks (DSN'03), pages 709-718, San Francisco, California, June 2003. 709~718
- 6 李春江, 李东升, 肖依, 杨学军. 计算网络应用可用性的度量模型. 计算机研究与发展, 2003, 40(12): 1705~1709
- 7 Choi H, Kulkarni V G, Trivedi K S. Markov regenerative stochastic petri net. Performance Evaluation, 1994, 20: 335~357
- 8 Deng S. Empirical model of WWW document arrivals at access link. ICC1996, 1996. 1797~1802

(上接第 51 页)

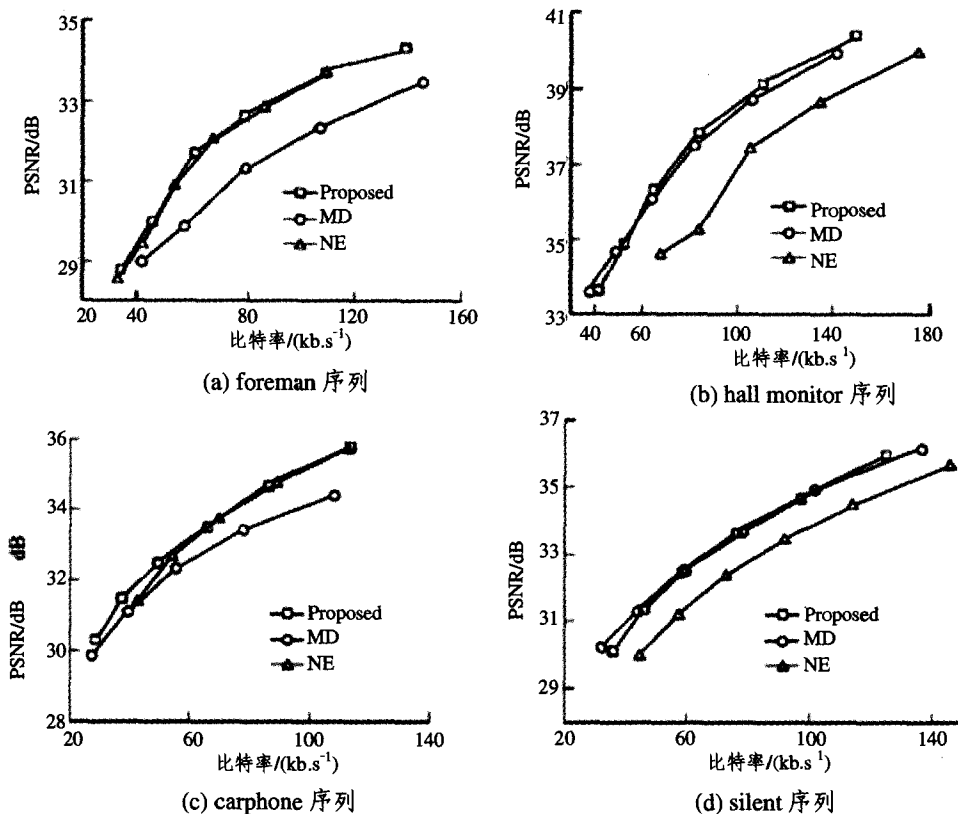


图 4 无线信道比特误码率在 $10^{-5} \sim 10^{-3}$ 之间的相关率失真算法性能比较

参考文献

- 1 Wiegand T. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H. 264 | ISO/IEC 14496-10 AVC), JVT-G050, March 2003
- 2 Wenger S. H. 264 Complexity Analysis according to VCEG-L36 section 2.1.4, document VCEG-M23, ITU-T Video Coding Experts Group (VCEG) Meeting, 2-4 April, 2001
- 3 Schäfer R, Wiegand T. the Emerging H. 264/AVC Standard. EBUT echnical Review, Jan. 2003
- 4 傅祖芸著. 信息论——基础理论与应用. 电子工业出版社, 2001. 9

- 5 张剑. H. 264 标准的量化 16 位算法. 计算机科学, 2004, 11
- 6 Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H. 264 ISO/IEC 14496-10 AVC), Draft 7 5th Meeting: Geneva, Switzerland, 9-17 October, 2002
- 7 Zhang jian. The Computation Model of Code Error Distortion Based on the Rate-Distortion. Theory, international conference on embedded software and systems(ICESS), 2005, 12
- 8 Zhang jian. Research on the Computation Model of Macro block Coding Based on the Rate-Distortion Optimization, International Symposium on Communications and Information Technologies 2005 (ISCIT 2005), 2005. 10