虚拟化技术是云计算的基础，它对于数据中心资源的整合和高效配置起着重要的作用。内存作为一个有限的资源，如何在一个物理节点上的多个虚拟机之间动态分配来实现内存的高效利用一直是研究的难点。要实现内存的动态调配，首先要解决的问题是对虚拟机进行实时的内存预测。

在虚拟化环境下，虚拟机的内存访问对于虚拟机管理器来说是透明的，传统方法通过给虚拟机页表项置位,来使得虚拟机的每次页面访问都发生失效并陷入到虚拟机管理器，从而获得虚拟机的内存访问序列，之后根据内存访问序列构建失效率曲线(MRC)，预测工作集大小。然而这种方法会带来巨大的开销，因为虚拟机的每次内存访问都会发生缺页中断，这对于虚拟机的使用者来说是不可接受的。

传统的通过访存序列计算MRC的算法开销很大，模拟LRU过程的链表算法的时间复杂度为$O(NM)$(N为访存的次数，M为访存的不同地址的总数)，平衡树算法将时间复杂度降低到$O(Nlog\_M)$，采样的方法通过截获部分内存访问的办法减少N的大小，这些优化方法需要在精度和开销之间做出权衡，特别是需要准确分析引入采样方法对MRC的估算的影响。

最新的AET算法是基于平均淘汰时间的算法，重用时间相比重用距离能够在O(1)的时间获得，所以AET算法整体的时间复杂度降低到O(N)，更重要的是AET提出的采样算法能够在只截获部分访存情况下获得精确度很高的近似MRC，在离线情况下，使用采样的AET算法能够得到与传统LRU算法1\%误差的结果。本文将AET算法引入到虚拟化下内存工作集的在线预测上来，在全虚拟化环境下实现了内存的采样截获，对比了通过采样方式使用AET算法获得的MRC和不采样使用经典LRU算法计算的MRC，验证了AET算法在虚拟化环境下进行内存工作集预测上的可行性，通过动态采样率的方法将开销进一步降低，使得虚拟机内存预测系统能够在不影响虚拟机正常使用的情况下实现较为精准的工作集预测。

Virtualization technology is the basis of cloud computing, which plays an important role in the integration of resources and efficient allocation. How to dynamically allocate virtual machines' memory on one physical node to achieve full utilization has been a challenge. To achieve dynamic allocation of memory, the first problem to be solved is the real-time memory prediction.

In virtualization environment, virtual machines' memory access is transparent to the virtual machine manager(VMM). Our work intercepts virtual machine's memory access by marking page table's reserved bit. According to the memory access trace, we can build miss ratio curve (MRC) and make predictions of the working set size. However, this kind of method can bring huge cost, because every memory access of virtual machines will cause a page fault, which is unacceptable to the user of the virtual machine.

The traditional way to compute MRC using memory access trace has high overhead. The time complexity of LRU algorithm is O(NM)(N is the number of memory access, M is the number of different accessing address). Balancing tree algorithm reduces overhead to $O(Nlog\_M)$. Sampling method further cuts down the frequency of intercepting memory access meaning lower N. These optimization looks for balance between accuracy and overhead. Especially using sampling method is way to estimate MRC.

AET is the latest novel algorithm which reduces time complexity to O(N). More importantly, its sampling algorithm only captures a portion of memory accesses to obtain approximate MRC with high precision. This paper implements memory sampling technology in full virtualization environment. Compared the two MRCs computed by AET sampling algorithm and classical LRU algorithm, we verify the feasibility of using AET sampling algorithm in the working set prediction. Using dynamic sampling rate will further reduce the cost which makes it possible to obtain accurate working set prediction without any influence to the usage of virtual machines.