**海量存储技术结课总结**

**姓名：张衡**

**学号：2017202110049**

**云缓存：云计算的按需闪存管理**

**CloudCache: On-demand Flash Cache Management for Cloud Computing**

**Proceedings of the 14th USENIX Conference on File and Storage Technologies (FAST ’16).**

摘要

基于主机端的闪存有希望解决云计算中的虚拟机（virtual machine）存储这个可扩展的问题。本文提出了提出了云缓存：一种基于按需闪存管理的云计算方法满足虚拟机缓存需求的同时，减少高速缓存的磨损。首先，本文提出了一个新的高速缓存需求模型Reuse Working Set (RWS)来帮助高速缓存按需分配。该模型能保留时间局部性好的数据，同时使用一个RWS大小的量（RWSS）为工作中的缓存需求量建模。通过在线预测RWSS，只允许RWS访问高速缓存，云缓存能够准确地满足工作时的缓存需求而且也把缓存的消耗降到最低。其次，本文提出了一种动态缓存迁移方法来平衡主机中实时数据缓存迁移产生的缓存负载与虚拟机的负载。这包括对脏数据的按需迁移和RWS的背景迁移，进而优化虚拟机的性能。该方法同样支持对缓存数据传递速率的限制来减少对联合主机虚拟机的影响。最后，本文使用了实际数据进行综合性的实验，实验证明了云缓存的有效性。

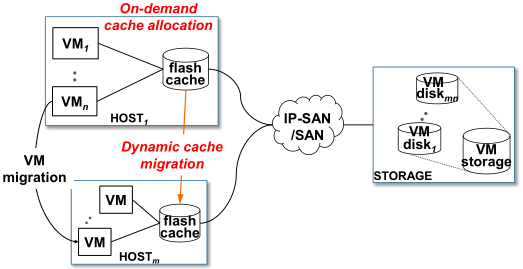
1．背景介绍

网络结构化的存储系统是云计算当中的重要部分，它有诸多好处，例如提高有效存储的利用率、提供可靠的虚拟机存储和实时的虚拟机迁移等。因此近些年网络结构化的存储系统受到了越来越多的关注。受关注的原因有两个，其一，随着各个云计算系统中融合的程度不断提高，共享的虚拟机存储服务器的可扩展性受到了影响；另外，因为访问一个局部闪存实际上要比通过网络访问一个远距离存储快，所以基于快速内存存储的闪存成为解决IO扩展性问题的一个选择。

然而由于容量、价格的限制，主机上闪存设备的数量相比于虚拟机中数据集的数量要少的多。因此为了充分利用闪存的潜力，需要根据实际需求在虚拟机之间竞争获取共享缓存。此外，因为读写会加重闪存的损耗，所以在对高速缓存管理时要尽量避免对无用数据的操作来减少对设备的损耗。

我们提出了云存储（CloudCache）上按需的闪存管理方法来解决上面的问题。方法主要分为如图所示的两个部分：

1. 按需的高速缓存分配机制：按照虚拟机对缓存容量的需求分配共享告诉缓存。
2. 动态缓存迁移：平衡虚拟机的缓存负载和缓存数据。



整体的结构可以这样描述：虚拟主机共享了一个网络存储来保存虚拟机的磁盘内容，然后通过SAN或者IP来访问。每个主机都有一个当地虚拟机共享的闪存，然后所有虚拟机通过这个高速缓存来访问自己位于远端磁盘里的内容。云缓存提供了上述按需的高速缓存分配机制和动态缓存迁移机制。尽管本文的重点关注块级别的虚拟机存储和高速缓存，但是我们的方法对于基于网络文件系统的虚拟机存储同样有用。

2．方法设计

按需分配高速缓存存在两个关键问题。其一，如何对一个缓存的工作负荷量建模？云计算中的工作负荷两包括由于时间局部性的程度不同而对高速缓存命中率造成不同影响的IOs。一个好的缓存需求模型能够通过提高性能的同时减少高速缓存的消耗。其二，如何使用缓存需求模型作为指导来分配缓存同时决定数据进入缓存？我们需要随时地精确预测高速缓存的工作负载为缓存的分配和保证有用的数据进入缓存提供依据。下面将详细提出解决这两个问题的办法。

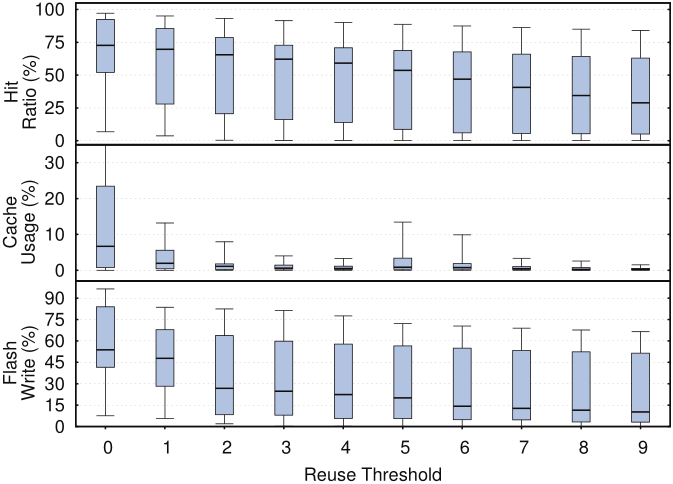
2.1 按需的高速缓存分配机制

Working Set（WS）是估计高速缓存需求量的经典模型。经典的WS模型中，在t时的WS(t,T)被定义为时间区[t −T,t]内以工作负载为参考的不同的块数据。这个定义使用了局部性的原则来估计保存在高速缓存中工作负载将访问的数据集。因此working set size（WSS）用来评估高速缓存工作负载的需求量。

尽管用WSS能很直接地估计虚拟机闪存的需求，但一个比较比较严重的缺点是这种方法没有对时间局部性的数据进行分级，因此结果就是分配得到的高速缓存中会存在时间局部性程度很低的数据。不仅浪费缓存空间，而且会加重缓存设备的使用负担，加速设备的损耗。

基于此，我们提出了一个称为Reuse Working Set

的模型。其中，RWSn(t,T)被定义为时间区[t −T,t]内以工作负载为参考的已经被重复使用n次的不同的块数据。相比于经典的WS模型，RWS模型保存的时间局部性的数据能够极大地提高工作负载的缓存命中率。当n=0时，RWS模型与WS相同。为了证明RWS的有效性，我们分析了不同n值情况下模型对于高速缓存命中率的影响。



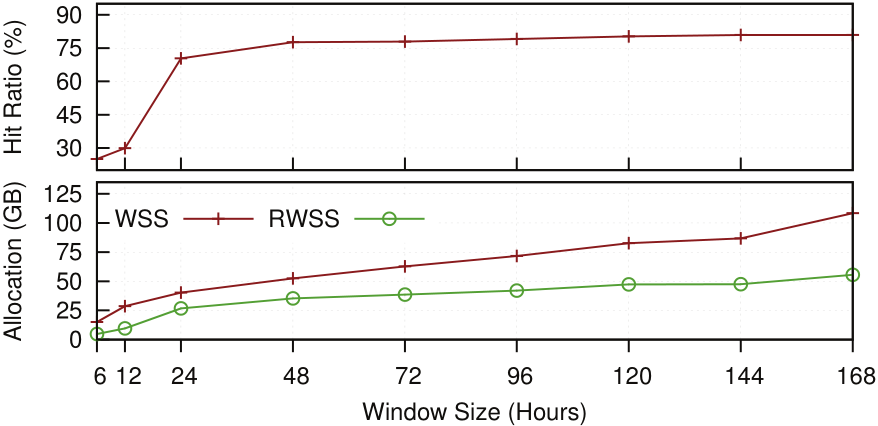
从图中可以看出，n值从0到1时，高速缓存命中率的中位数下降了8%，但是高速缓存的使用次数下降了82%，同时闪存的写的次数下降了19%。这种趋势随着n值不断增加而继续。

这个结果说明，使用RWSS去估计高速缓存的需求是更有效的：它降低了工作负载对高速缓存的使用次数即以轻微降低高速缓存命中率的代价减少了高速缓存的损耗。系统的使用者能够根据对于系统性能的要求来对n值进行设置，从而达到平衡高速缓存命中率和高速缓存损耗的问题。通常情况下，当n为1或者2时，能较好地在两者之间取得平衡。

模型建立完成后，我们需要对工作负载的RWSS进行测量，因此我们需要选择一个合适的时间窗口来观测工作负载。这其中需要讨论两个问题。其一，如何跟踪窗口？在经典WS模型中，窗口是根据访问次数而不是时间设置的。然后，因为虚拟机可能会闲置很长一段时间，且从不会被窗口填满，所以使用虚拟机的被访问次数来决定窗口的设置是很困难的。因此，我们使用了基于时间的窗口来观察RWSS的工作负载。

其二，如何决定窗口的大小？如果窗口太小，RWS模型将不能获取工作负载当前的局部性信息，进而得到的RWSS将会低估工作负载的缓存需求，如果窗口太大，它将包含旧的局部性数据从而影响缓存的空间和增加不必要的设备损耗。我们的解决办法是概括一段时间的工作负载，即同时统计使用不同的窗口大小情况下缓存的命中率。画出分配命中率与窗口大小的关系图像，选取图像中的knee point对应的量作为实际的窗口大小。

如下图示：

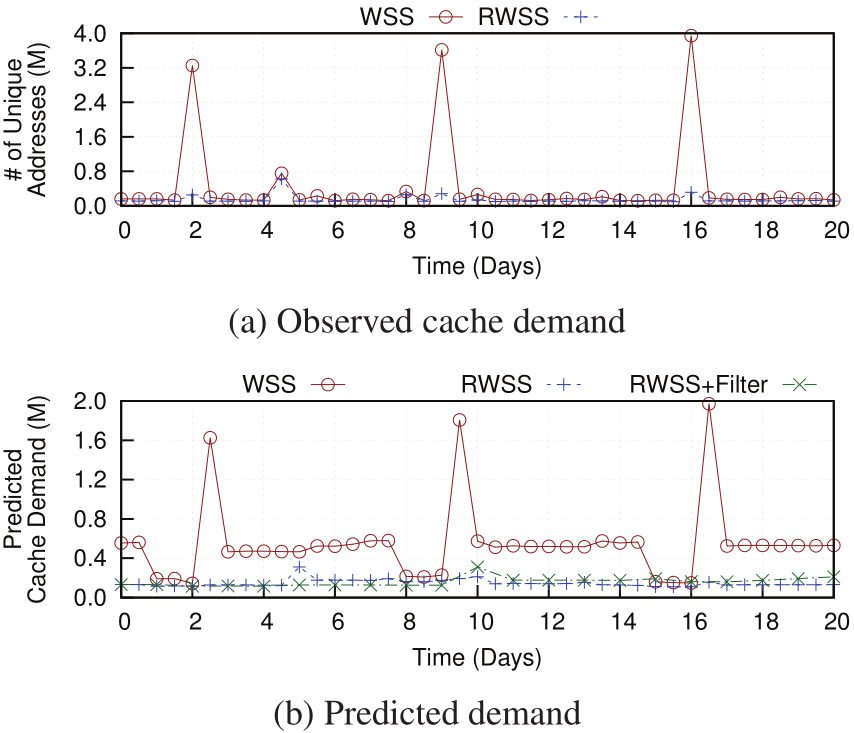


在该图选择命中率即将放平的拐点作为参考点。另外，可以周期性地在线做这样的性能分析来决定窗口的大小。

至此，高速缓存的需求模型建立完成。接下来需要在线对高速缓存的需求进行预测。

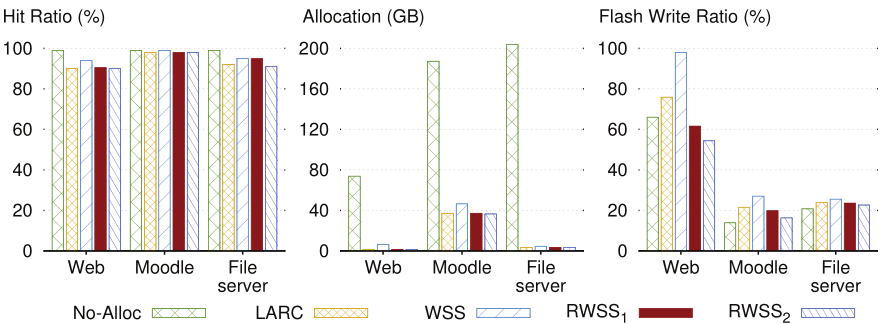
对高速缓存的需求预测存在两种方法：指数平滑法和双指数平滑法。他们分别一个平滑参数α，一个趋势参数β。参数的设置对于预测的精确度有着很大的影响。我们基于预测值与RWSS的观测值之间的误差来估计以上参数，并轮转使用两个方法。另外，为了提高这种预测方法的鲁棒性，我们设计了一个过滤方法把观测过程中出现的异常值舍弃，来抑制这些值对预测准确度的影响，称为RWSS-filter方法。

下图展示的结果是RWSS三周内预测一个网站服务器的缓存去求的结果：



可以看出，RWSS模型自动地过滤出备份的IO设备，预测得到的RWSS只有WSS平均值的26%，同时，使用过滤技术的RWSS模型筛选出了由于IO偶然的爆发导致的不能正常反应工作负载变化趋势的异常值。

上述实验证明了RWSS模型对于缓存需求预测的有效性，在使用了RWSS模型对缓存的需求进行估计和预测之后，需要同时对虚拟机进行缓存分配。我们要求对缓存的分配不会导致对于数据复制和冲洗的额外消耗。因此我们设计了一个不需要跨分区对缓存分配的替代措施（replacement-time enforcement）。同时，我们还需要一个只允许数据被使用过的块的访问策略（admission policy）。

对基于WSS和基于RWSS的缓存分配进行比较，实验结果如下：

实验结果证明了：我们提出的基于RWSS模型的高速缓存分配策略能够在稍许降低缓存命中率的情况下在实质上减少工作负载的的缓存使用次数，也相当于减少了对于缓存设备的损耗。在现实的使用场景中，由于工作负载额外的缓存分配在窗口刚开始变动时不需要立刻丢弃，因此我们的方法的性能开销会变得更少。此外，因为基于WSS的方法对于同样的工作负载会更多的缓存分配，云服务提供商不得不提供更大的缓存，这也会导致额外的现金支出。

2.2 动态缓存迁移

上一小节讨论的按需的缓存分配策略允许云缓存更准确地估计工作负载的缓存需求然后动态地把共享的的缓存容量分配给它们。但是当缓存容量不足时，该如何分配内存？同时，如何平衡缓存负载和数据之间的关系？本节将提出一种动态的缓存迁移方法来解决上述问题。

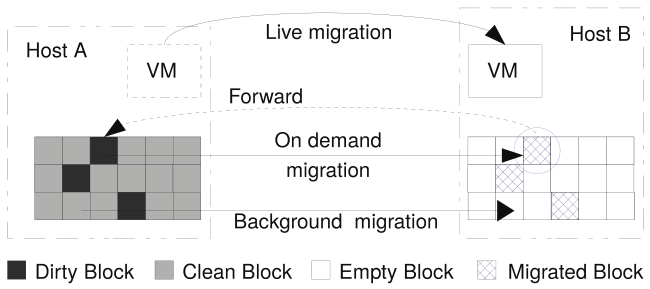
2.21 实时缓存迁移

实时虚拟机迁移允许工作负载能够明显地在主机之间传递。在云存储中，我们使用实时虚拟机迁移方法来平衡闪存的负载，使得一些虚拟机能够被迁移到其他拥有多余缓存的主机上。

需要解决两个问题：迁移源主机的迁移虚拟机脏数据必须跟目的主机的同步，才能保证他们能被虚拟机再次访问；迁移虚拟机需要唤醒目标主机的高速缓存，这可能会耗费很长时间，其性能会在高速缓存被唤醒前大幅度降低。为了解决上述两个问题，云缓存的动态缓存迁移方法混合使用了被动和主动性的迁移技术：

按需迁移：当迁移的虚拟机访问了一个源主机的脏块数据，它的局部缓存将请求推送给源主机并从源主机获得数据。源主机脏块上的元数据将被随着迁移的虚拟机转移，所以目的主机上的局部缓存知道那些块在源主机上是脏块。

隐蔽迁移：除了被动性的迁移技术，源主机上的缓存同样能主动地转移虚拟机上缓存的数据。这个转移操作是隐蔽进行以降低对源主机上虚拟机的影响。这种隐蔽迁移允许目的主机快速地唤醒它的局部高速缓存器，除了能提高迁移的虚拟机的性能，他还允许源主机快速地减少其缓存负载进而提高其剩下的虚拟机的性能。

按需迁移允许迁移虚拟机快速地访问其脏块，但不适合转移许多块。隐蔽迁移能够高效地转移数据但不适合为当前等待服务的迁移虚拟机提供服务。因此，我们选择将这两者策略合并运用。下图描述了云缓存中的缓存迁移是如何进行的：

当一个虚拟机是一个从A主机到B主机的实时虚拟机，为了保持数据一致性同时避免冲洗数据的操作，高速缓存的元数据上的脏块被转移到了B主机上。一旦虚拟机的实时迁移完成，B主机上的虚拟机将被激活，其局部闪存可以很快的对一些服务请求进行回应。

通过使用转移的元数据，主机B上的高速缓存可以判断一个块是否是在脏块，以及该块的位置。如果主机A上扔存在一个脏块，一个请求将被传送去满足需求，同时，主机A也会传送隐蔽迁移中迁移虚拟机的RWS。当缓存块无论是使用按需迁移的方法还是隐蔽迁移的方法从A主机被转移到B主机，A主机将为其他虚拟机腾出它的高速缓存空间。

2.22 迁移速率限制

上述两种方法的混合使用能够优化迁移虚拟机的性能，但是这个过程对于其他虚拟机的影响同样值得我们注意。缓存迁移要求源主机上的读和对目的主机缓存的写操作，这些操作会降低其他联合虚拟主机高速缓存IOs，除了被虚拟机内存迁移所消耗的带宽，这些操作也会占用额外的带宽。

因此，云缓存将限制高速缓存迁移的转移速度。当转移速度确定后，即从源主机到目的主机的数据块的最大数量将被确定。这个速度可以基于虚拟机的优先级和迁移主机的RWSS值来确定。云缓存允许系统管理者调节转移速度来降低高速缓存迁移对其他联合虚拟主机造成的影响。

3．总结

闪存拥有解决存储设备发展瓶颈和提高云计算系统中虚拟机性能的优秀潜力。按照虚拟机的需求并有限制地分配缓存是有效利用高速缓存和优化虚拟机性能的关键。

本文提出了提出了云缓存：一种基于按需闪存管理的云计算方法满足虚拟机缓存需求的同时，减少高速缓存的损耗。首先，本文提出了一个新的高速缓存需求模型Reuse Working Set (RWS)来帮助高速缓存按需分配。该模型能保留时间局部性好的数据，同时使用一个RWS大小的量（RWSS）为工作中的缓存需求量建模。通过在线预测RWSS，只允许RWS访问高速缓存，云缓存能够准确地满足工作时的缓存需求而且也把缓存的消耗降到最低。其次，本文提出了一种动态缓存迁移方法来平衡主机中实时数据缓存迁移产生的缓存负载与虚拟机的负载。这包括对脏数据的按需迁移和RWS的背景迁移，进而优化虚拟机的性能。该方法同样支持对缓存数据传递速率的限制来减少对联合主机虚拟机的影响。