Is Server Consolidation Beneficial to MMORPG? A Case Study of World of Warcraft

Yeng-Ting Lee and Kuan-Ta Chen
Institute of Information Science, Academia Sinica
habam.ytlee@iis.sinica.edu.tw, ktchen@iis.sinica.edu.tw

***Abstract*—MMORPG is shown to be a killer application of Internet, with a global subscriber number increased to 17 millions in 2010. However, MMORPG servers tend to be overly provisioned because 1) such games do not have standard architectures thus dedicated hardware is assumed; 2) MMORPGs normally adopt a "sharded design" to resolve the scalability challenges of content production and workload distribution; and 3) a game is commonly deployed in geographically distributed data centers to protect gamers from excessive network latencies. Therefore, an operator needs to deploy dedicated hardware for each game in each data center, even though hardware utilization is low.**

In this paper, we propose a *zone-based server consolidation strategy for MMORPGs*, which exploits the unique *spatial locality property* of players' interactions, to cut down the games' considerable hardware requirement and energy use. We evaluate the effectiveness of our strategy based on a nine-month trace from a popular MMORPG World of Warcraft. The evaluation results show that, with a per-hour dynamic zone reallocation policy, the server number required can be reduced by 52% and the total energy consumption can be reduced by 62%, while the user-experienced latency remains undegraded.

I. INTRODUCTION

Massively multiplayer online role-playing game (MMORPG) is a genre of computer role-playing games in which a very large number of players interact with one another within a virtual game world. MMORPGs are played throughout the world. In [27], it is reported that over 55% of Internet users are now also online gamers. There are total 354 MMORPGs collected on mmorpg.com [13], and over 16 million MMORPG active subscribers worldwide [24]. In 2008, Western consumer spending on subscribing MMOGs grew to $\$1.4$ billion [18], while in 2009, MMORPG market in China grew to $\$2.93$ billion [12].

Today, to provide a quality gaming experience to players, MMORPG operators tend to overly provision game servers mainly because of the following two reasons:

1) *Nonstandard architecture*: Unlike web applications, there are no standard or even *de facto* architectures for online game systems, thus different games may have totally different hardware and software (e.g., operating systems, middleware, database systems) requirements. This fact encourages or even forces a game operator to invest dedicated hardware for each game, even if spare resources from other games are available (as a game's subscriber number may decrease quickly over time for most game titles [23]).
2) *Sharded design*: Even a game can attract millions of online players, it is resource-prohibitive for a design team to produce sufficient game content, e.g., scenes, creatures, and missions, for such a huge number of players to explore and conquer at the same time. Therefore, most games adopt a "sharded" architecture [25] by providing multiple identical game worlds (Such a world is called a "realm" in World of Warcraft). The game content is duplicated to each realm and all the players' game data (e.g., the avatars' race, level, experience points earned, missions solved) are bound to a realm. This design resolves the scalability issues of content production and workload distribution; however, it worsens the server over-provisioning phenomenon, since as long as one player remains in a realm, the machines serving this realm cannot be retired.

The situation becomes worse when multiple games are hosted and multiple data centers are involved. Because of nonstandard architectures, an operator needs to invest $m \times n$ servers for m games, assuming n servers are required for a game on average. In addition, to protect gamers from experiencing excessive network latencies, a game may be deployed in multiple geographically distributed data centers. For example, World of Warcraft hosts over 700 realms worldwide, which are served by servers deployed in 10 data centers in Washington, California, Texas, Massachusetts, France, Germany, Sweden, South Korea, China, and Taiwan [11]. This further inflates the required hardware investment for MMORPGs by a factor of r assuming each game is deployed in r data centers on average.

With the advent of the virtualization technology [22], the consolidation [10] of MMORPG servers is now made possible despite that various system architectures may be involved. With a proper server consolidation strategy, an operator can safely reduce hardware investment and energy consumption (by putting idle servers into sleep mode whenever appropriate), while maintaining user-perceived service quality. We consider server consolidation is particularly suitable for MMORPGs because of the following reasons:

1) Like other online entertainment services, the daily and weekly workload variation is large, as many people do not (or cannot) play online games during work hours or on weekdays. This provides opportunities for power saving by consolidating workload to fewer servers.
2) In MMORPGs, most of the interactions between players possess a *spatial locality property*; that is, players tend to interact with others (by chatting, trading, fighting, or

teaming up to fight others, etc) in their vicinity in the game's virtual world. Thus, we can spatially partition a game world into multiple disjoint "zones," where each zone spans a certain continuous region. *We treat a zone as a unit for workload dispatching based on virtualization*, i.e., serving each zone with a separate VM (virtual machine); such design will not incur much inter-VM communication since user interactions mostly occur among players in the same zone. This property makes the workload of an MMORPG naturally *partitionable* and therefore suitable for server consolidation.

3) While many web application providers host only a single service, e.g., Facebook[1] and Flickr[2], it is not uncommon for an online game operator hosting dozens of games at the same time. For example, SNDA, one of the most biggest game companies in China, hosts more than 20 MMORPGs[3]. Since the subscribers of a game may fluctuate greatly over its course of operation [24][4], it is important for the games of an operator to reuse the same infrastructure considering the huge amount of hardware investment.

In this paper, we propose to use a *zone-based server consolidation strategy for MMORPGs* to cut down the considerable hardware investment and energy use. In order to empirically evaluate the effectiveness of our proposal, we collect a player-usage[5] trace from a real-life World of Warcraft server during a nine-month period. We first apply a multi-scale analysis of the variability of avatar number over day, week, and month. Having shown that the avatar number is highly predictable at the hour and day scales, we apply the zone-based server consolidation strategy which reallocates zones among a set of server clusters regularly and evaluate its impact in terms of the number of servers required and energy. The evaluation results show that, with a per-hour dynamic zone reallocation policy, the server number requirement can be reduced by 52% and the total energy consumption can be reduced by 62%, while the user-experienced latency remains not degraded.

To sum up, our contributions in this paper is two-fold:

1) We elaborate why server consolidation is appropriate for MMORPGs and propose a zone-based strategy to facilitate server consolidation based on the virtualization technology.
2) We show that the zone-based server consolidation is beneficial in terms of both the number of servers required and energy consumption based on real-life traces from a popular MMORPG World of Warcraft.

The remainder of this paper is organized as follows. Section II describes related works. In Section III, we present how we collected the World of Warcraft trace and the summary of the trace. We analyze the variability and predictability of the avatar count in Section IV. In Section V, we evaluate the effectiveness of the proposed zone-based server consolidation strategies based on the collected trace. Finally, Section VI draws our conclusion.

[1] http://www.facebook.com

[2] http://www.flickr.com

[3] Shanda, http://ir1.snda.com/, March 2010

[4] It is common that a new game attracts millions of curious players. However, many of them will just "taste" the game and quickly switch to another one. The million userbase may soon decline to ten thousands or much less a few months later.

[5] In some MMORPGs, a player is allowed to have more than one avatars to play with; however, at the same time a gamer normally "plays" an avatar only. For the sake of brevity, we shall use "avatar" or "player" interchangeably throughout this paper.

II. Related Work

The energy consumption of data centers is becoming an increasingly important economic consideration. In [17], the authors collected empirical electricity costs from large companies and data centers and performed an analysis study. Their estimation showed that data centers account for an estimated 61 million MWh, which corresponds to 1.5% of US electricity consumption and costs 4.5 billion US dollars per year [4]. At worst, by 2011, data center energy use could double; at best, by replacing current hardware with state-of-the-art green-aware components, we may be able to reduce electricity usage to half the current level in 2011 [4]. These figures reveal the emergent demand in the reduction of energy consumption.

In [19], the authors proposed to use performance modeling and stepwise refinement to predict the benefit from server consolidation, which enables analysts understand the dynamics, dependencies, and interaction among system components. However, their analysis focuses on qualitative rather than quantitative results. While qualitative analysis may help us derive insights or detect latent problems, it cannot estimate the actual cost saving that may be required for decision making.

In [14], the authors investigated the resource allocation model for MMOGs running on a public cloud. They evaluated various resource renting models based on a real-life trace from RuneScape. While [14] is also related to online games, it focuses on the real-time interaction performance of gaming, rather than the cost of hardware investment and energy consumption as we do in this work. Also, it assumes that MMOGs run on public clouds, while our proposed strategy is general and can be used in both private- and public-cloud scenarios.

III. Data Description

In this section, we begin with an introduction of World of Warcraft, and describe how we collect players' game hours in an automated fashion. We conclude this section with a summary of the collected traces.

A. World of Warcraft

World of Warcraft is the fourth game set developed by Blizzard Entertainment Incorporation, and it is currently the most popular MMORPG in the world (March 2010). According to MMOGChart [24], the game's 11.5 million subscribers accounted for 62% of the MMOG market in Dec 2008 [1]. Because of its popularity, it has become a field for academia researchers to study psychology [26], social interaction [3, 15], and game play behavior [2, 6, 8, 9, 16].

TABLE I
DATA DESCRIPTION

Summary of Collected Trace	
Faction	Alliance
Realm	TW-Light's Hope
Start Date	2006-01-01
End Date	2006-09-30
Length	273 days
Sampling rate	144 samples per day

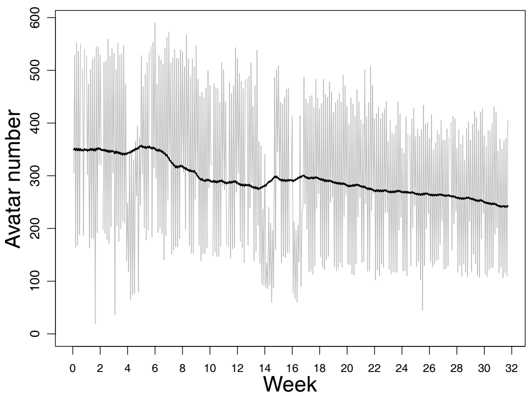

Fig. 1. Hourly online avatar number and its moving average in our trace.

B. Trace collection

We used the who command, which is publicly available to every player in the game, to collect our traces. The command asks the game server to reply with a list of players who are currently online. Thus, anyone can obtain the gameplay history of all the users on a server by issuing the who command with a regular interval. To do so, we create a character on a World of Warcraft server (the Light's Hope realm at Taiwan) and keep it online all the time. Our character is controlled by a program and automatically collects a list of online users every 10 minutes. If a player logins and logouts within 10 minutes, we may not be able to observe his/her re-login activity in consecutive snapshots. However, we do not consider this problem significant because most WoW session times are much longer than 10 minutes [20].

For scalability consideration, World of Warcraft servers restrict the number of users returned by a query to a maximum of 50 accounts. Thus, we have to narrow down our query ranges by dividing all the users into different races, professions, and levels. For example, we need to first ask the server to list all the users of the "Fighter" class with the first query, and then ask the server to list all the users with the "Wizard" class with the second query, and so on. This technique allows us to systematically list the entire set of online players despite the restriction of the query function.

C. Data Summary

We performed the data collection from Jan. 2006 to Sep. 2006. During the monitored 273 days, $8,761$ avatars and $166,758$ game sessions associated with the avatars were observed. Because the sampling interval was 10 minutes, we recorded $273 \times 144 = 39,312$ samples, where each sample represents the online avatar number at the sampling time. The data summary is presented in Table I.

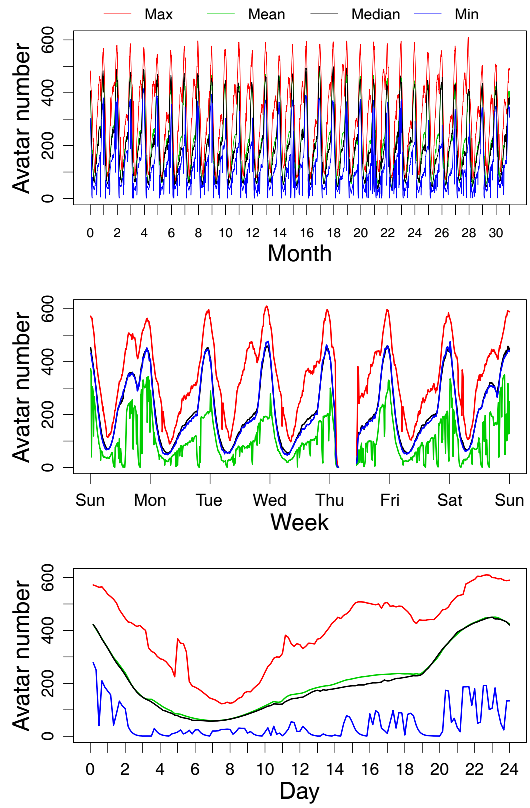

Fig. 2. The averaged avatar number process shown in different time scales. The top two graphs are plotted based on daily averages and the bottom one is plotted based on the hourly averages.

To give an overall feeling of how the trace looks like, we plot the avatar number over the trace period in Fig. 1. The grey line indicates the actual avatar number we observed at ten-minute intervals, while the black thick line is the moving average of the samples. The graph reveals the potential of server consolidation for the game. First, the grey line constantly fluctuates (e.g., between 200 and 600) in each day, which implies the power saving at daily off-peak times may be significant. Second, the average avatar number gradually decreases from 350 to 250 over the nine-month course, which implies potential benefits from merging multiple realms onto a physical server over a long-term course.

IV. PATTERNS IN THE AVATAR COUNT TRACE

In this section, we analyze the variability and predictability of the avatar number samples in our trace. The variability analysis is aimed to understand how much we could benefit by consolidating MMORPG servers, while the predictability analysis helps us determine whether zone allocations can be (partially) planned based on historical records.

A. Variability Analysis

We begin by plotting the averaged maximum, mean, median, and minimum avatar number in different time scales, as shown in Fig. 2. The top graph shows that except variations within each day, there are no evident systematic variations at larger time scales. We zoom in the graph and stop at the week scale, as shown in the middle graph of Fig. 2, clearly the daily

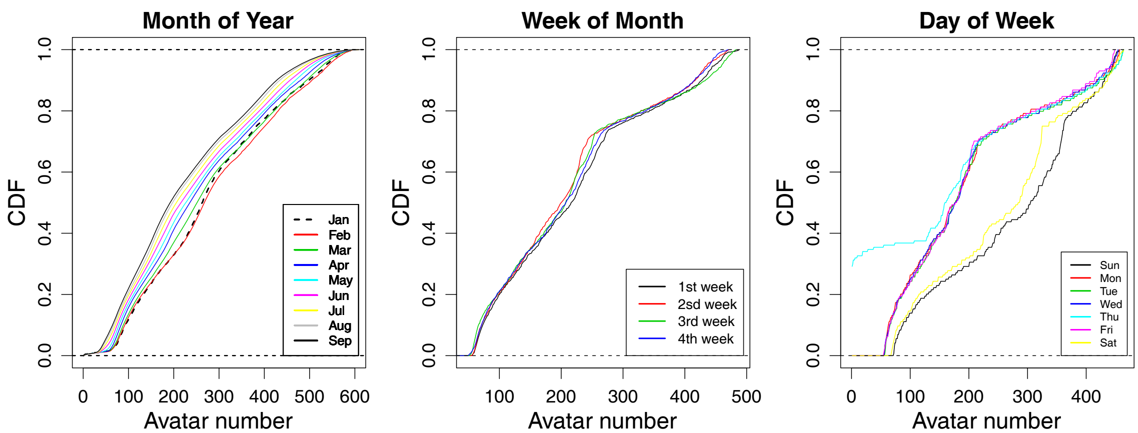

Fig. 3. The cumulative distribution functions of averaged avatar numbers in different time scales. The leftmost graph is based on monthly averaged numbers; the middle graph is based on weekly averaged numbers; the rightmost graph is based on daily averaged numbers.

variations dominate the variability component of the avatar number process. The blackout period in the Thursday morning is because the operator schedules a weekly maintenance down time during this period. We further zoom in and plot the daily variations of the avatar number in the bottom graph. The figure shows strong fluctuations of the avatar number over a range of 0 to 600 within the 24 hours a day. On average, the coldest hour is around 7am and the hottest hour is around 11pm, which implies the addictiveness of the game as many players often stay all the nights for gaming.

We further look at the distributions of avatar numbers over consecutive periods in different time scales, as shown in Fig. 3. From the leftmost graph, the avatar numbers in different months are similar with a slightly decreasing tread over the months. The only exception is January, which we suspect a consequence of the Chinese New Year vacation. During such vacations, Chinese people gather with their family, play mahjong, and travel; thus the avatar number in January was slightly fewer than February in our trace.

In the middle graph of Fig. 3, we can see that, as expected, the avatar number distributions in different weeks are similar. The rightmost graph, which shows the avatar number distributions on each day of the week, indicates that the overall game play time is significantly different on weekdays and on weekends, where Sunday attracts more gaming time than Saturday. The distribution curve of Thursday is due to the weekly maintenance, where the deviation of the curve from those of other weekdays implies that gamers come back immediately after the maintenance hours without "wasting" off-game time, which again reveals the addictiveness of the game.

B. *Predictability Analysis*

The strong regularity over consecutive weeks and days and strong variability within a day can be further examined by auto-correlation function (ACF) plots at corresponding time scales, as shown in Fig. 4. In this figure, the top graph exhibits the strong weekly regularity, while the middle graph exhibits the strongly daily regularity in the avatar number process. The ACF of the avatar number over hours in the bottom

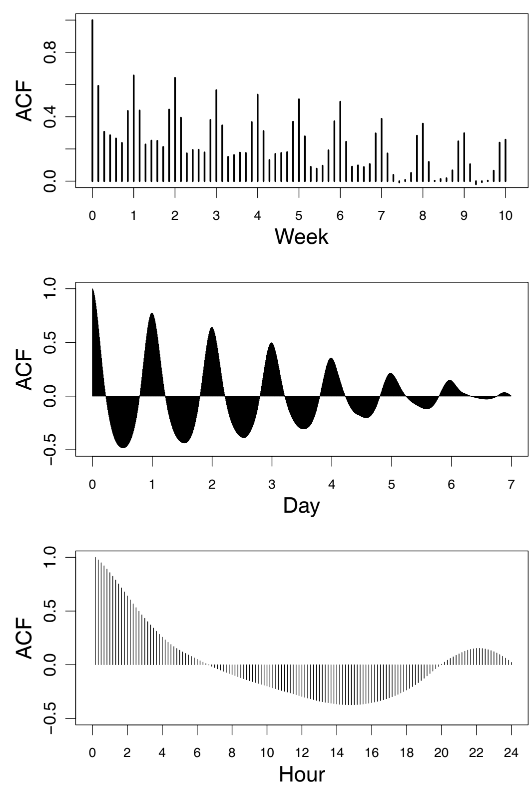

Fig. 4. The auto-correlation functions of the avatar number process in different times scales. The top graph is based on daily averaged numbers; the middle graph is based on hourly averaged numbers; and the bottom graph is based on the avatar number observed at each 10 minutes.

graph shows no regularity existing within a day; however, the large coefficients with a lag shorter than 2 hours indicates the potential for high predictability of avatar number within the future few hours.

We now apply simple estimation rules to test whether avatar number can predicted based on its historical data with high accuracy. The three rules we use are as follows:

- Prediction based on the last hour;
- Prediction based on the same hour in the last day;
- Prediction based on the same hour in the last week.

As shown in Figure 5, we find that the avatar number is highly

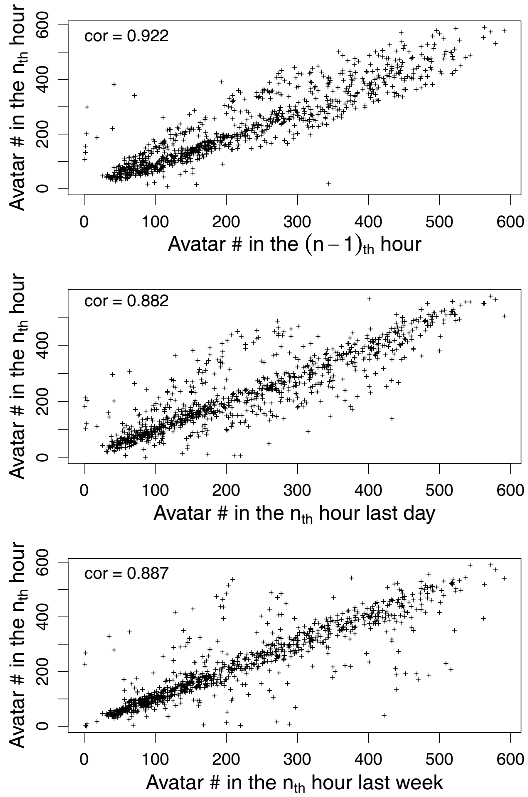

Fig. 5. Predictability analysis. The scatter plots of the avatar number in time t and $t+1$, where the unit of time is hour, day, and week, respectively.

predictable with the three simple prediction rules, where the prediction based on the last hour performs the best with a correlation coefficient as high as 0.922. We believe that the slightly lower predictability of "the same hour in the last day" is mainly due to the differences between weekdays and weekends, and that of "the same hour in the last week" is due to occasional game events that cause flash crowds to gather at certain points in the game.

In sum, our analysis of the variability and predictability of the avatar number process shows that the variations, especially at the day scale, are significant, which implies large potential benefits from server consolidation. On the other hand, the observation that the avatar number is highly predictable is also a good news, because it implies that the zone allocation computations need not to be performed very frequently and can even be pre-planned in advance.

V. Proposed Strategy and Its Effectiveness

In this section, we first elaborate the proposed zone-based server consolidation strategy and then present our trace-based simulation methodology for evaluating the effectiveness of the proposed strategy. Our evaluations are conducted in single- and multiple-game scenarios respectively, and the results show that the proposed strategy can indeed greatly save server investment and energy consumption in both scenarios.

A. Zone-based Server Consolidation

As we have stated in Section I, MMORPGs possess a *spatial locality property* in that players mostly interact with others in their vicinity areas in the virtual world. Because of this property, we can spatially partition a game world into multiple disjoint "zones." We consider *such a zone a perfect unit for workload dispatching based on virtualization*. Specifically, a game whose virtual world containing n zones are served by n processes, where each process is responsible for a zone respectively, and each process (zone) resides in a separate VM for the sake of flexible migration. This design incurs little inter-VM communications because of the aforementioned spatial locality property.

One advantage of the strategy is that, as far as we know, most MMORPGs incorporate the concept of zones in their system design. In some games, such as World of Warcraft, the concept of zones is *implicit*, as its virtual world is continuous and flat; however, a player can query the current zone his avatar resides by issuing the `who` command. In some other games, such as Lineage, players can be *explicitly* aware of the existence of zones, because whenever their avatars move across different zones, the screen will show a "Data Loading" progress indicator and they cannot continue to play until the loading of the target zone is finished. Therefore, the zone-based server consolidation fits to the *de facto* design of current MMORPGs and requires minimum efforts for game designers to adopt.

Our zone-based server consolidation strategy works as follows. Assuming an operator owns s servers and hosts r realms, where each realm contains z zones, and each zone is served by a VM. We can treat the VM allocation problem as the *bin-packing* problem, and obtain its near-optimum solutions by using any of the existing heuristic algorithms [21]. For example, one of the best known heuristics is called the First-Fit Decreasing (FFD) algorithm, which can be illustrated as follows:

- Given:
  - $n = r \times z$ objects (zones) to be placed in bins (servers) of capacity L each.
  - zone p contains l_p avatars in the next time period T.
- Objective: determine the minimum number of servers needed to accommodate n zones.
- Algorithm:
  - sort zones so that $l_p \geq l_{p+1}, 1 \leq p < n$;
  - label s servers as $1, 2, \ldots, s$;
  - zones are considered for packing in the order $1, 2, 3, \ldots, n$;
  - designate zone p on server q, where server q is the server which can accommodate zone p with the smallest q.

The algorithm processes the zones in an arbitrary order. For each zone, it attempts to place the zone in the first server that can accommodate the zone. If no such a server is found, it boots a new server and puts the zone on the new server.

By employing the FFD or similar algorithms, we can find a near-optimum arrangement to designate the $r \times z$ zones on s servers while minimizing the required number of servers. The problems remain include 1) how to choose a time period T, and 2) how can we know the avatar number in each zone *in advance*. Choosing the time period T is a trade-off between

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Simulated period	2 months
Server capacity	7500 avatars
Realm number (r)	100
Server number (s)	100
Zone per realm (z)	83
Avatar number per realm	$\mathcal{N}(2640, 1500^2)$
Power consumption per server	300 W (idle), 750 W (peak)
VM allocation algorithm	First-Fit Decreasing

server utilization efficiency and overhead of zone migrations. We consider a period of one hour or shorter is appropriate since the variations in each zone within one hour would not be large. On the other hand, predicting the (approximate) maximum avatar number in each zone is plausible because the total avatar number is highly predictable, as we have shown in Fig. 5, and the avatar number in each zone changes in a fairly low frequency (e.g., on average 5.8 avatars are moving across different zones per minute in a realm).

B. Experiment Setup

In the following, we present the design of our simulations for evaluating the effectiveness of the proposed strategy. In our evaluations, we assume a scenario that an operator owns $s = 100$ servers hosting $r = 100$ realms of a game, where each realm contains $r = 83$ zones[6]. Supposing that a server is capable of serving $7,500$ avatars[7], which is also the maximum avatar number each realm supports. To simulate realistic avatar numbers in different realms, we model the avatar number in a realm as a normal distribution with mean $2,640$ and standard deviation $1,500$, which is derived from the data set on Warcraft Census[8] and Wow Database[9] assuming the maximum avatar number per realm is $7,500$. For each realm, after obtaining an avatar number (with the above normal distribution), we compute the avatar number in each zone based on the relative avatar number in each zone in our collected trace. The simulation parameters and details are summarized in Table II.

To evaluate the effectiveness of server consolidation, we run the simulations with three different policies:

- Fixed: This is the baseline policy. With this policy, no virtualization technology is used. We simply host each of the r realms on a physical server.
- Dynamic (day): With this policy, we adopt the zone-based server consolidation strategy with the zone reallocation period T set to 24 hours. That is, we reallocate the zones on the servers at 6am each day. We assume that the maximum avatar number in each zone can be predicted by using an oracle predictor.
- Dynamic (hour): This policy is similar to the "Dynamic (day)" policy except that T is set to one hour.

[6]This number is chosen because we have observed 83 zones in our World of Warcraft trace.

[7]Because there are 11.5 million subscribers, and 779 realms according to WoWWiki. Thus, on average each realm has $115,000,000/779 = 14,400$ players. We assume that the most players will not player the game for more than 12 hours a day, so we expect that at maximum half of the players would be online at any time. Therefore, we can expect that only half the world's players would be on at any point in time, $14,400/2 = 7,200$, thus we choose the approximation $7,500$ as the server capacity.

[8]http://www.warcraftrealms.com/census.php

[9]http://www.hotwow.com/

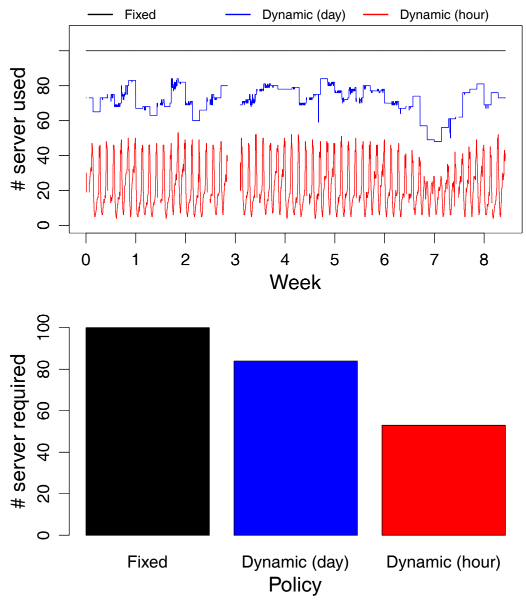

Fig. 6. The top graph shows the number of servers required over time in the single-game simulation; the bottom graph shows the maximum number of servers required over the simulation term.

We evaluate the three policies in terms of server investment (i.e., the number of servers used) and energy consumption, while ensuring that the response time is not degraded due to high server utilization. The energy consumption of each server is computed by [5]:

$$P_{idle} + (P_{peak} - P_{idle}) \times U + (PUE - 1) \times P_{peak},$$

where P_{peak} and P_{idle} are the server power usage at peak and idle states, U is the average server utilization, and PUE is the power usage effectiveness, with a typical value 2. While the model is obviously simplified, our goal is to provide an estimate how server consolidation helps save energy consumption rather than an accurate computation of energy actually used.

In our dynamic zone allocation algorithms, we ensure that the avatar number on any server will not achieve 80% of its capacity in order to maintain user-perceived latencies. The choice of 80% is based on an assumption that the server's CPU utilization is approximately proportional to the number of avatars the server currently serve and Fig. 5 of [7], which shows that if the CPU utilization is over 0.8, the system's mean response time will increase rapidly.

C. Single-Game Evaluation

In the first scenario, we compute the number of servers required and energy consumed using the three policies by assuming that a game operator only hosts one game at a time. The number of servers required over time is shown in the top graph of Fig. 6. From the graph, we can see that while the Fixed policy constantly uses 100 servers, the Dynamic (day) policy requires around 84 servers and the Dynamic

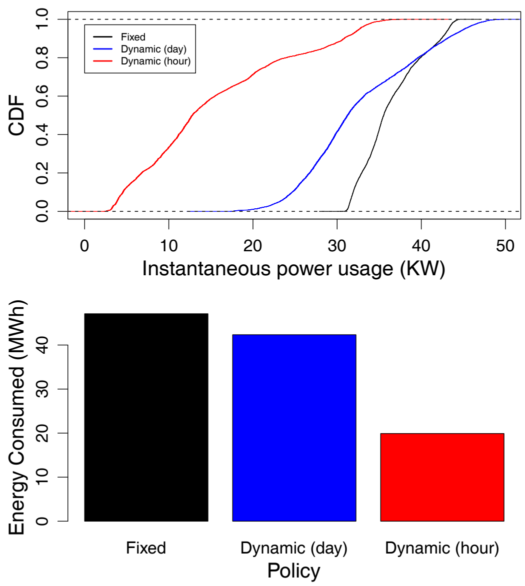

Fig. 7. Cumulative distribution functions of instantaneous power usage (top graph) and accumulative energy consumption (bottom graph) in the single-game simulation.

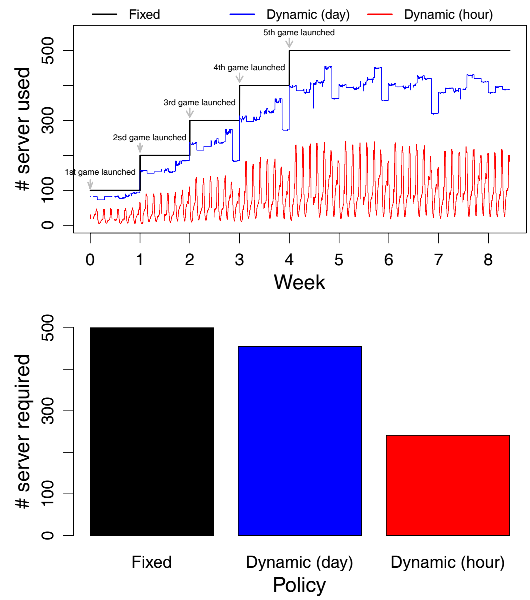

Fig. 9. The top graph shows the number of servers required over time in the multiple-game simulation; the bottom graph shows the maximum number of servers required over the simulation term.

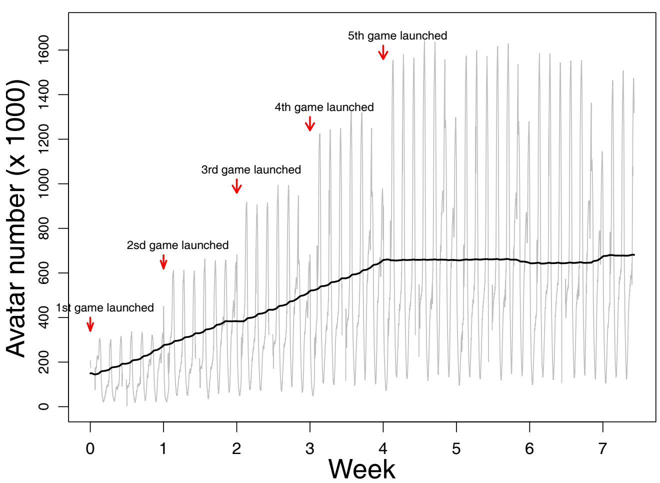

Fig. 8. The number of participating avatars over time in the multiple-game simulation.

(hour) policy requires much fewer servers (with an average around 25), though the number of servers required fluctuates significantly in each hour. The blackout period at the 3rd week was caused by occasional server downtime. We also plot the maximum number of servers required at any time for the three policies in the bottom graph of Fig. 6. The three policies require 100, 84, and 53 servers respectively, which indicates that server consolidation can greatly save the hardware investment of MMORPG operators, and the dynamic zone allocation at an one-hour frequency is needed to ensure high server utilization.

From the perspective of energy consumption, server consolidation provides further benefits because during off-peak hours, we can reduce the servers' power usage by putting idle servers into sleep mode. We plot the cumulative distribution function of the instantaneous power usage and the total energy consumed during the simulation period in Fig. 7. From the top graph, we can see that the Dynamic (hour) policy can reduce power use greatly by putting idle servers into sleep mode, so that the off-peak power use (1 KW) is only around $1/10$ of the peak use (10 KW). On the contrary, the power usage with the other two policies are more concentrated around their respective peak usage. The bottom graph plots the total energy consumed with different policies, which indicates the Dynamic (day) and Dynamic (hour) policies can save 11% and 57% energy from the zone-based server consolidation strategy.

D. Multiple-Game Evaluation

Our second scenario is a more realistic one, in which a game operator hosts up to five MMORPGs at the same time. We assume that the five games are released one by one with an one-month interval, as shown in Fig. 8. We then evaluate the effectiveness of zone-based server consolidation by similar procedures in the single-game scenario.

In Fig. 8, we can see that the number of servers required using the fixed policy increases by a fixed amount every time a game is launched over the simulation period. With the Dynamic (day) policy, the server number increase is slightly less, but still significant. On the other hand, the multiplexing effect plays its role with the Dynamic (hour) policy in that the number of servers required increases much more slowly than the other two policies. As a result, the overall server investment is 500, 455, and 241 servers respectively. The energy consumption evaluation shows similar behavior as well, with the results plotted in Fig. 10.

To quantify the effectiveness of the different policies, we summarize the server investment and energy consumption of the two server consolidation policies (normalized by that of Fixed policy) in Table III. We can see that, in both scenarios, the Dynamic (day) policy does not yield good results (around 10% and 15% reduction in hardware investment and energy use respectively), which implies that the long-term

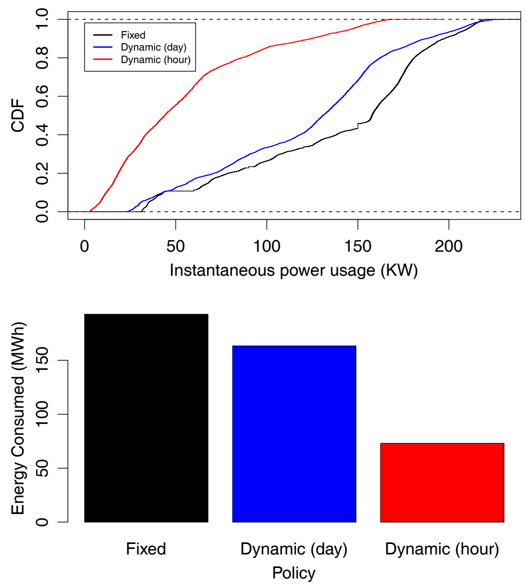

Fig. 10. Cumulative distribution functions of instantaneous power usage (top graph) and accumulative energy consumption (bottom graph) in the multiple-game simulation.

variation may not be the key to efficient server consolidation for MMORPGs. On the contrary, the Dynamic (hour) policy performs significantly better even though the only difference from the Dynamic (day) policy is on the zone reallocation frequency. In the five-game scenario, it reduces the server investment by 52% and energy consumption by 62%, which is a significant saving in any case. Note that the effect could be more prominent if an operator hosts more games at the same time, which is commonly seen especially for Asian game operators.

VI. Conclusion and Future Work

In this paper, we have shown that MMORPGs, because of the spatial locality property of the players' interactions, can be made much more resource-efficient by applying our proposed zone-based server consolidation strategy. Using a set of real-life player count statistics of World of Warcraft, we conduct a simulation study and show that the zone-based server consolidation strategy can reduce the number of required servers by 52% and energy consumption by 62%, while maintaining user-experienced latencies.

In the future, we plan to incorporate longer real-life traces in order to investigate the combined effect of multiple-game hosting and long-term variations of subscribers (normally a decreasing trend for most games). In addition, how to achieve optimal zone reallocations at appropriate time intervals and how to manage user-perceived gaming experience by efficiently allocating resources among virtual machines on a physical server are also part of our future researches.

TABLE III
Summary of benefits introduced by zone-based server consolidation strategy.

	Single game		Multiple games	
	$D_d^\dagger$	$D_h^\ddagger$	D_d	D_h
Server Investment	84%	53%	91%	48%
Energy Consumption	89%	43%	84%	38%

†: D_d denotes the Dynamic (day) strategy.
‡: D_h denotes the Dynamic (hour) strategy.

References

[1] Blizzard Entertainment Inc., "World of Warcraft subscriber base reaches 11.5 million worldwide," http://us.blizzard.com/en-us/company/press/pressreleases.html?081121, Nov 2008.
[2] C. Chambers, W.-c. Feng, S. Sahu, and D. Saha, "Measurement-based characterization of a collection of on-line games," in *IMC'05*, 2005, pp. 1–14.
[3] V. H.-h. Chen and H. B.-L. Duh, "Understanding social interaction in World of Warcraft," in *ACE'07*, 2007, pp. 21–24.
[4] U. Environmental Protection Agency, "Server and data center energy efficiency," *Final Report to Congress*, 2007.
[5] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," in *ISCA'07*, 2007, pp. 13–23.
[6] T. Henderson and S. Bhatti, "Modelling user behaviour in networked games," in *SIGMM'01*, 2001, pp. 212–220.
[7] J. Heo, X. Zhu, P. Padala, and Z. Wang, "Memory overbooking and dynamic control of Xen virtual machines in consolidated environments," in *IM'09*, 2009, pp. 630–637.
[8] J. Kim, J. Choi, D. Chang, T. Kwon, Y. Choi, and E. Yuk, "Traffic characteristics of a massively multi-player online role playing game," in *NetGames'05*, 2005, pp. 1–8.
[9] M. Kwok and G. Yeung, "Characterization of user behavior in a multi-player online game," in *ACE'05*, 2005, pp. 69–74.
[10] D. A. Menasce, "Virtualization: Concepts, applications, and performance modeling," *CMG'05*, 2005.
[11] R. Miller, "WoW's back end: 10 data centers, 75,000 cores," http://www.datacenterknowledge.com/archives/2009/11/25/wows-back-end-10-data-centers-75000-cores/, 2009.
[12] Ministry of Culture of the People's Republic of China, "2009 White paper of China MMOG market," http://big5.gov.cn/gate/big5/www.gov.cn/gzdt/2010-01/19/content_1514798.htm, 2010.
[13] MMORPG.com, "Total game list," http://www.mmorpg.com/gamelist.cfm, 2010.
[14] V. Nae, A. Iosup, R. Prodan, and T. Fahringer, "The impact of virtualization on the performance of massively multiplayer online games," in *NetGames'09*, 2009, pp. 1–6.
[15] B. Nardi, "Collaborative play in World of Warcraft," in *LA-WEB'06*, 2006, p. 3.
[16] D. Pittman and C. GauthierDickey, "A measurement study of virtual populations in massively multiplayer online games," in *NetGames'07*, 2007, pp. 25–30.
[17] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for Internet-scale systems," in *SIGCOMM'09*, 2009, pp. 123–134.
[18] ScreenDigest, "Subscription MMOGs: life beyond World of Warcraft," http://gmedia.glos.ac.uk/igd301/docs/PR-LifeBeyond WorldOfWarcraft-240309.pdf, 2009.
[19] A. Spellmann, K. Erickson, and J. Reynolds, "Server consolidation using performance modeling," *IT Professional*, vol. 5, no. 5, pp. 31–36, 2003.
[20] P. Svoboda, W. Karner, and M. Rupp, "Traffic analysis and modeling for World of Warcraft," in *ICC'07*, 2007, pp. 1612 –1617.
[21] Vazirani and V. Vijay, *Approximation Algorithms*. Berlin: Springer, 2001.
[22] VMware Inc., "Virtualization overview," http://www.vmware.com/pdf/virtualization.pdf.
[23] Voig.com, "MMOGData: Charts, subscriptions," http://mmog-data.voig.com/.
[24] B. S. Woodcock, "An analysis of MMOG subscription growth version 21.0," http://www.mmogchart.com/.
[25] WorldofWarcraft.com, "What is a realm," http://www.worldofwarcraft.com/info/basics/realmtypes.html.
[26] R. Wright, "Expert: 40 percent of World of Warcraft players addicted," http://www.tomshardware.com/forum/96296-13-expert-percent-world-warcraft-players-addicted, 2006.
[27] Z. Zhou, E. Wan, and X. Xu, "2006 online game report," Pacific Epoch Red Innovation Report Series, 2006.