Team Performance Prediction in Massively Multiplayer Online Role-Playing Games (MMORPGs)

Kyong Jin Shim  
Department of Computer Science and Engineering  
University of Minnesota  
Minneapolis, MN 55455, USA  
Email: kjshim@cs.umn.edu

Jaideep Srivastava  
Department of Computer Science and Engineering  
University of Minnesota  
Minneapolis, MN 55455, USA  
Email: srivasta@cs.umn.edu

Abstract—In this study, we propose a comprehensive performance management tool for measuring and reporting operational activities of teams. This study uses performance data of game players and teams in EverQuest II, a popular MMORPG developed by Sony Online Entertainment, to build performance prediction models for task performing teams. The prediction models provide a projection of task performing team’s future performance based on the past performance patterns of participating players on the team as well as team characteristics. While the existing game system lacks the ability to predict team-level performance, the prediction models proposed in this study are expected to be a useful addition with potential applications in player and team recommendations. First, we present player and team performance metrics that can be generalized to all types of games with the concept of point gain, leveling up, and session or completion time. Second, we show that larger or more advanced teams do not necessarily achieve higher team performance than smaller or less advanced teams. Third, we present novel team performance prediction methods based on the past performance patterns of participating players and team characteristics.

Systematic studies of team performance is expected to yield the following contributions. First, analysis of team performance in different dimensions (i.e. distribution of participating players’ demographics, archetypes, classes, sub-classes) can help game developers understand whether their games and game characters are being played as intended. Second, benefits for game players are two fold. a) Game players can not only have a view of their past and current performance but also they can have a view of their projected future performance not only in solo playing settings but more importantly, in group playing situations as many of today’s games from first person shooter games (i.e. Halo) to MMOGs (i.e. EverQuest II, World of Warcraft) involve a high level of team playing. b) A recommendation engine can be built to evaluate character or player level performance in group play settings [23], [24] and recommend players to teams seeking additional players to join in combats activities. Third, teams can regularly check on their performances as well as of other teams for the purposes of exchanging players or forming a larger combat team.

I. INTRODUCTION

Massively Multiplayer Online Role-Playing Games (MMORPGs) are personal computer or console-based digital games where thousands of players can simultaneously sign on to the same online, persistent virtual world to interact and collaborate with each other through their in-game characters. This study is concerned with forecasting of player performance in the game. While many games today provide web and GUI-based reports and dashboards for monitoring player performance, we propose a more comprehensive performance management tool (i.e. player and team scorecards) for measuring and reporting operational activities of game players and teams. This study uses operational and process-oriented performance data of game players and teams in EverQuest II, a popular MMORPG developed by Sony Online Entertainment, to build performance prediction models for teams. The prediction models provide a projection of a task performing team’s future performance based on the team characteristics and the past performance patterns of participating players on the team. While the existing game system lacks the ability to predict team-level performance, the prediction models proposed in this study are expected to be a useful addition to existing player and team performance monitoring tools. First, we present team performance metrics that can be generalized to all types of games with the concept of point gain, leveling up, and session or completion time. Second, we show that larger or more advanced teams do not necessarily achieve higher team performance than smaller or less advanced teams. Third, we present novel team performance prediction methods based on the past performance patterns of participating players and team characteristics.
tool. Previous studies [23], [24] showcased such a tool for measuring and reporting operational activities of game players. This study uses operational and process-oriented performance data of game players and teams in EverQuest II to analyze player and team behaviors in the game universe, systematically and quantitatively assess team performance, and predict and project team’s future performance. The prediction models proposed in this study are expected to be a useful addition to the performance management tool and task recommendation system for use by EverQuest II players and teams.

III. EVERQUEST II GAME MECHANICS

A. Point-Scaling System in EverQuest II

In EverQuest II, there is a concept of Ding Points, which is the amount of points one needs to obtain in order to move from one level to the next higher level [16]. For instance, to move from Level 2 to Level 3, one needs to obtain 1,000 points whereas 20,000 points are required to move from Level 73 to 74. The amount of ding points increases as one advances to the next level. As players gain more experience with the game and advance to higher levels, the types of task they can perform increase and the task difficulty also increases. The higher the task difficulty, the higher the potential point gain.

B. Tasks in EverQuest II

EverQuest II is rich in types of task players can perform with monster kills being one of the most popular. Monster kills are discussed in details in [23]. In addition to monster kills, other sources of experience points exist in the game such as alternate achievement points (AA) which can be obtained from quests, named mobs, and discovery experience. A player can gain more experience points by having another player mentor him. The mentor levels down to the level of the mentee. The mentee receives a five percent bonus to adventuring experience points.

C. Archetypes, Classes, and Sub-classes in EverQuest II

In playing MMORPGs, selection of character type (i.e., archetype, class, sub-class, and race) is considered an important decision as it defines the basis of opportunities and choices of roles and tasks within the game [18]. In EverQuest II, there are four archetypes where each archetype consists of three classes each of which in turn consists of two sub-classes [16]. Performance comparisons are discussed in details in [24].

IV. TEAM FORMATION IN EVERQUEST II

Manufacturing plants over the years have adopted the formation of work teams as a practice [42], [46], [45], [36]. Many companies have adopted team approaches to produce high quality products and services which would lead to improved customer satisfaction [32]. Additionally, huge cost savings coupled with quality improvements have been reported in numerous studies [43], [37], [40], [39], [41], [38]. A more recent study conducted empirical studies on the impact of team formations at workplaces on manufacturing performance over an extended period of time [47].

As is the case in manufacturing, team formation is a common occurrence in many MMORPGs [21], [20], [22]. The games are designed to encourage social interactions in such a way that certain quests must be done as a team. Not only that, certain quests require players each with a different set of skills. In order to successfully complete a given quest, the team members must collaborate and rely on one another. In EverQuest II’s monster kills, a player can choose to team with and collaborate with one or more players in killing monsters. Such grouping behaviors are often observed in the case of killing difficult or vicious monsters. Also, novice players can team up with more advanced players to get familiarized with the game via the game’s mentorship system. On average, over 12 million teams form monthly, making games such as EverQuest II an excellent venue for studying human and organizational behaviors such as team formation.

V. INDIVIDUAL PERFORMANCE METRICS

All of the below listed individual performance metrics can be computed over a certain time duration or over one or more player levels.

A. Efficiency Index

A previous study [23] defines player performance in EverQuest II as a function of XP point gain and play time (referred to as session time). We refer to this measure as player Efficiency Index. Given two players, the one with a larger point gain, given the same amount of time as the other player, is considered more efficient.

\[
EfficiencyIndex_{jk} = \frac{\sum_{i=1}^{N} XP_{ki}}{\sum_{j=1}^{M} ST_{kj}}
\]

where

\[XP = \text{Experience points}\]

\[N = \text{Total number of tasks completed by Player } K\]

\[ST = \text{Session time}\]

\[M = \text{Total number of sessions during which Player } K \text{ completed tasks}\]

B. Busyness Index

Player Busyness Index is a function of the total number of activities over play time (session time). Player Busyness Index is intended to show how frequently a player is involved in game play.

\[
BusynessIndex_{jk} = \frac{\sum_{j=1}^{N} ST_{kj}}{M}
\]

C. Grouping Frequency Index

This index measures what fraction of the time a given player is involved in a group play (forming a team). It is computed as the total number of tasks done as a group divided by the total number of tasks performed.
D. Soloing Frequency Index

This index measures what fraction of the time a given player is involved in a solo play. It is computed as the total number of tasks done solo divided by the total number of tasks performed.

E. Seniority Index

Player Seniority Index is a measure of how advanced a given player is compared to the rest of the team members. It is computed as:

\[ \text{SeniorityIndex}_{k,j} = \text{PlayerLevel}_{k,j} - \text{GroupLevel}_j \]

The above formula computes Seniority Index value for Player \( k \) on Team \( j \). If the player is playing solo, the index value would be zero. If this value is positive, we say that this player is having more ”seniority” compared to the rest of the team members. If this value is negative, we say that this player is more ”junior” to the rest of the team members.

F. Group Size Index

Typically we compute this index as an aggregate over a time duration. A player can complete tasks as part of a team or he can solo. While Grouping Frequency Index measures how frequently a given player plays as part of a team, this index measures on average how large his team(s) get.

G. Success Index

Success Index is a measure of how successful a given player is at completing one or more tasks. In this study, we compute Success Index with respect to monster kills. It is computed as:

\[ \text{SuccessIndex}_k = \frac{MK_k}{MK_k + D_k} \]

where

\( MK \) = Total number of successful monster kills performed by Player \( K \)
\( D \) = Total number of death incidents assumed by Player \( K \)

VI. TEAM PERFORMANCE METRICS

A. Efficiency Index

Team Efficiency Index is similar to Player Efficiency Index. Given two teams, the one with a larger point gain, given the same amount of time as the other team, is considered more efficient.

B. Casualty/Survivability Index

Casualty Index is a measure of what fraction of the team has perished during an operation. Likewise, Survivability Index is a measure of what fraction of the team has survived and successfully completed a given task. In this study, we compute these Indexes with respect to monster kills.

\[ \text{CasualtyIndex}_{i,x} = \frac{DM_i}{T_i} \]
\[ \text{SurvivabilityIndex}_{i,x} = \frac{SM_i}{T_i} \]

where

\( DM \) = Total number of dead members on Team \( i \)
\( SM \) = Total number of surviving members on Team \( i \)
\( T \) = Total number of members on Team \( i \)

VII. METHODS

A. Dataset

The study uses nine months worth of player and team activity data on “Guk” server (PvE or Player-versus-Environment) from January 1, 2006 to September 9, 2006. The dataset contains over 283 million (67% solo plays, 33% team plays) player-to-task records where over 135 million (35% solo plays, 65% team plays) of them are monster kills and quest related tasks. The dataset contains 63,707 distinct players across player levels 1 through 70. Since then, Sony Online Entertainment has added an additional ten levels to the game, making 80 the maximum level one can reach. In a more recent release, Sentinel’s Fate, the game maker raised the level cap to 90. All of the characters and their activity data has been extracted from XP table in the EverQuest II database housed at National Center for Supercomputing Applications (NCSA) at the University of Illinois. The dataset contains the minimum the following information about game players and their characters: character id, character sub-class, race, task, timestamp of task completion, group size (whether a given character grouped with one or more other characters in completing a task), average group level (if a given character played with one or more other characters, this value represents the average of player levels of all characters involved in that group), experience (XP) points, and location (location in which the task was completed).

B. Session Extraction

Our preliminary analysis shows that the total amount of time between a player logs in to a game and logs out of the game does not reflect the actual amount of time that the player spent performing tasks or socializing. A player can log in and leave the game without explicitly logging out of the game, hence creating one or more chunks of what we refer to as ”inactive” or ”idle” time. In the present study and also in previous studies [23], [24], we programmatically weave one or more active sessions from the game’s performance data. Any chunk of time that exceeds 30 minutes without any activity is considered an inactive or idle time, and it is excluded from the total amount of play time computed for each player.

VIII. EXPERIMENTS AND RESULTS

In this section, we first examine the effect of solo versus group playing on Efficiency Index and Success Index. Next, we examine the effect of individual players’ past performance patterns on team performance. First, we compute individual player performance metrics for 63,707 distinct players over the nine month period (across players levels 1 through 70). Second, we compute team performance metrics. Next, we perform correlation studies on the relationship between past individual performance patterns of players working as a team and the team’s performance. Using the findings from the correlation studies, we build regression-based prediction models for team performance prediction.
A. Impact of Solo versus Group playing on Team Performance

For a given task, is it more efficient to play solo or as part of a team? Our results show that as player level increases, it becomes more efficient to play as part of a team. In our analysis, we have picked several different monsters across Monster Levels 1 through 70 and computed Efficiency Index for both solo playing and group playing. The below charts show Efficiency Index and Success Index across players levels 1 through 70, aggregated over the various monsters we analyze in this study.

Figure 1 shows that in lower player levels (1 through 7), solo players achieve higher Efficiency Index than group players. Beyond Player Level 7, Efficiency Index for group playing starts exceeding that of solo playing. The results indicate that beyond a certain level, grouping with one or more other players in killing a monster leads to overall higher Efficiency Index.

Figure 2 shows that it becomes increasingly unsafe to play solo as player level increases. In higher levels, often times players encounter monsters too vicious that as their health deteriorates, they need other players on the team to impose damage onto the monsters while they recuperate and regain their health. In absence of such support or enough support, small size teams or solo players may individually have to take on more damage on themselves which may lead to death.

B. Impact of Team Size on Team Performance

Is it necessarily true that the more the better? Do large teams necessarily achieve high efficiency? Our results indicate that this statement is true up to a certain point. The below charts summarize our results.

The red line in Figure 3 shows that Efficiency Index increases up until group size of six and then it starts declining. The blue line plots the Efficiency Index computed for solo players for the same tasks that the group players reflected in the red line performed. The chart shows that some of the tasks can be done with higher efficiency by playing solo.

Figure 4 shows that up until group size of six, it becomes increasingly safer to play as part of a group than playing solo. However, beyond the group size of six, the Success Index declines dramatically. The blue line plots the Success Index computed for solo players for the same tasks that the group players reflected in the red line performed. The chart shows that overall, it is still safer to play as part of a group, but as the group size becomes large beyond the size of six, we observe more occurrence of death(s) on the team.
C. Impact of Level Diversity on Team Performance

A team can consist of players of varying levels. Level Diversity within a team is a measure of the average level difference amongst the team members. We compute Level Diversity as the average of the absolute values of the Seniority Index values of all the members on a team.

Figure 5 plots on the x-axis the Level Diversity and on the y-axis the Team Efficiency Index (segmented by group level). The graph does not show any strong relationship between the two variables.

D. Impact of Task Difficulty on Team Performance

In our analysis, we examine monster kills and the task difficulty is defined as a function of monster kills. Figure 7 and Figure 8 show monster kills at various monster levels in the first week of September 2006. Figure 8 shows that the mean group level in team plays is in majority close to the level of the monster. As players level up, the monsters that they encounter become more vicious. Suppose that a player levels up to Level 40 and encounters Level 40 monsters. Our findings indicate that for the same task difficulty, smaller teams can take slightly longer session time to kill a given monster than larger teams, but the difference is negligible. Figure 7 shows that the majority of the Level 40 monster kills are performed by groups of size four, however, we see a good number of solo players. A further look into this group of solo players reveals that a majority of them are players of levels more advanced than 40.

Another investigation reveals that a majority of the higher level solo players that attempt similar level monsters are of fighter/warrior classes with heavy armors that often plays tanks in organized raid combats. We do not have many data points that show very high level players attempting to kill monsters whose levels are way below their levels. In the few cases we have observed such data points, we have found that due to the nature of the game’s point scaling system [23], the XP point gain in such cases would be very minimal. Perhaps because of the low challenge level, subsequently low XP point gain, and lack of entertainment in attempting mediocre monsters, players and teams do not target tasks whose difficulty is way lower than their levels. The implication of this finding on our prediction models is that there is not much variation in Efficiency Index values due to task difficulty and that task difficulty, given the dataset we have in this analysis, would not be a good independent variable to use in our prediction models.

E. Prediction of Team Performance

We have earlier stated that there appears to be some correlation between Efficiency Index and Group Size and also between Efficiency Index and Group Level. First, we conduct a correlation analysis to examine which of the individual performance metrics variables have any association with team performance. Although we have also earlier reported that there appears to be some correlation between Success Index and team performance, this aspect of team performance prediction is omitted in this paper and will be a future addition to the current study. Additionally, we focus on teams of size one (solo playing) through six in the following analysis. Our initial findings indicate that the behaviors of teams whose size is beyond six are less predictable compared to smaller teams.

Based on the findings so far, we report three different regression models. The game mechanics appear to be driving
solo and team plays in such a way that variables of interest such as level diversity, task difficulty, and individual player performance metrics appear to be statistically insignificant in predicting team performance. Hence, in this section, we build regression models based on two prominent variables, group size and group level. The first linear regression model predicts a team’s performance given its group size. The second linear regression model predicts a team’s performance given its group size and level. The third multiple linear regression model predicts a team’s performance given its group size and level. The following tables list correlation analysis results.

Table I shows that of all the individual performance metrics, the current group size appears to be statistically the most significant variable. Our results indicate that team performance in terms of Efficiency Index cannot be accurately predicted from individual level Efficiency Index values. Additionally, our results indicate that level diversity has very little association with team performance.

Table II shows that of all the individual performance metrics, the current group size appears to be statistically the most significant variable. Our results indicate that team performance in terms of Efficiency Index cannot be accurately predicted from individual level Efficiency Index values alone. Additionally, our results indicate that level diversity has very little association with team performance.

Table IV shows overall improved correlations between individual performance metrics and Team Efficiency. Given the above three findings, for each of the three Team Efficiency Indexes, we build a regression model.

<table>
<thead>
<tr>
<th>Individual Performance Metric</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Level</td>
<td>0.1707209</td>
</tr>
<tr>
<td>Group Size</td>
<td>0.9303044</td>
</tr>
<tr>
<td>Average Individual Efficiency Index</td>
<td>-0.02565945</td>
</tr>
<tr>
<td>Average Individual Busyness Index</td>
<td>0.2500358</td>
</tr>
<tr>
<td>Average Individual Group Level</td>
<td>0.1428135</td>
</tr>
<tr>
<td>Average Individual Group Size</td>
<td>0.2752545</td>
</tr>
<tr>
<td>Level Diversity</td>
<td>-0.006854777</td>
</tr>
</tbody>
</table>

**TABLE I**

CORRELATION ANALYSIS ON INDIVIDUAL PERFORMANCE METRICS AND TEAM EFFICIENCY (SEGMENTED BY GROUP SIZE)

<table>
<thead>
<tr>
<th>Individual Performance Metric</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Level</td>
<td>0.9227762</td>
</tr>
<tr>
<td>Group Size</td>
<td>0.1320354</td>
</tr>
<tr>
<td>Average Individual Efficiency Index</td>
<td>0.09665737</td>
</tr>
<tr>
<td>Average Individual Busyness Index</td>
<td>0.04035929</td>
</tr>
<tr>
<td>Average Individual Group Level</td>
<td>0.764814</td>
</tr>
<tr>
<td>Average Individual Group Size</td>
<td>0.2516168</td>
</tr>
<tr>
<td>Level Diversity</td>
<td>0.1030059</td>
</tr>
</tbody>
</table>

**TABLE II**

CORRELATION ANALYSIS ON INDIVIDUAL PERFORMANCE METRICS AND TEAM EFFICIENCY (SEGMENTED BY GROUP LEVEL)

<table>
<thead>
<tr>
<th>Individual Performance Metric</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Level</td>
<td>0.6853188</td>
</tr>
<tr>
<td>Group Size</td>
<td>0.3418811</td>
</tr>
<tr>
<td>Average Individual Efficiency Index</td>
<td>0.07418273</td>
</tr>
<tr>
<td>Average Individual Busyness Index</td>
<td>0.1366675</td>
</tr>
<tr>
<td>Average Individual Group Level</td>
<td>0.5571977</td>
</tr>
<tr>
<td>Average Individual Group Size</td>
<td>0.2575587</td>
</tr>
<tr>
<td>Level Diversity</td>
<td>0.07210109</td>
</tr>
</tbody>
</table>

**TABLE III**

CORRELATION ANALYSIS ON INDIVIDUAL PERFORMANCE METRICS AND TEAM EFFICIENCY (SEGMENTED BY BOTH GROUP SIZE AND GROUP LEVEL)
Figure 9 shows a linear regression model of Team Efficiency Index (segmented by group size). Heteroscedasticity is clearly evident in the residual plot in Figure 10, which shows that the vertical scatter is quite different in different vertical strips (large in some slices and small in others).

Figure 11 shows a linear regression model of Team Efficiency Index (segmented by group level). Figure 12 shows the residual plot for the built linear regression model. There is a visible amount of heteroscedasticity. The scatter in the residuals for large values of Team Efficiency (the range 3.0 and above) is a bit larger than the scatter of the residuals for smaller values of Team Efficiency.

Lastly, we examine the residual plot of the third regression model, for Team Efficiency Index including both group size and group level. Figure 13 shows the residual plot for the multiple linear regression model built for Team Efficiency Index including both group size and group level. Between 1.3 and 3 on the x-axis, there is a huge single cluster of points, indicating that the data are not randomly scattered above and below the x-axis.

The visible amount of heteroscedasticity and clusters in the residual plots raise a concern. We discuss further in the next
Table IV summarizes the three regression models built in this study.

### IX. Conclusion

In this paper, we examine team performance in EverQuest II. First, we report that as player level increases, it becomes more efficient to play as part of a team. Second, we report that large teams do not necessarily achieve high efficiency. Our results show that six is the golden number for the number of team members, as team efficiency starts degrading beyond the team size of six. Third, we report that the level diversity within a team, a measure of the average level difference amongst the team members, do not greatly affect the team performance. Our findings indicate that for the same task difficulty, smaller teams can take slightly longer session time to kill a given monster than larger teams, but the difference is negligible. Lastly, we build team performance prediction models using regression approach. We report three different regression models. The first linear regression model predicts a team’s performance given its group size. The second linear regression model predicts a team’s performance given its group level (average across individual players on the team). The third multiple linear regression model predicts a team’s performance given its group size and group level. The regression models report a good start of 86.55%, 85.15%, and 52.62% coverage. Additionally, based on our initial analysis, we report that a team’s performance is not correlating with the average over individual performances of participating players on the team.

### X. Future Directions

An extension to the current work involves investigating team performance beyond team size of six. Our initial findings show that performance of teams beyond size six are much less predictable. The team performance prediction models reported in this study lack the ability to integrate model dynamics over time. For instance, instead of computing the mean individual player performance over a certain duration in both solo playing and group playing, we look to model the ups and downs of the player’s performance patterns and incorporate it into the team performance prediction models. The residual plots reported in Experiments and Results section raise a concern due to visible heteroscedasticity and clusters (or lack of randomness in the data scatter). We look to either further segment the teams based on other attributes prior to model building. Additionally, we look to find more datasets that will allow us to build team performance prediction models using task difficulty as an independent variable.

### Acknowledgment

The research reported herein was supported by the National Science Foundation via award number IIS-0729421, the Army Research Institute via award number W91WAW-08-C-0106, the Intelligence Advanced Research Projects Activity via AFRL Contract No. FA8650-10-C-7010, and the ARL Network Science CTA via BBN TECH/W911NF-09-2-0053. The data used for this research was provided by the SONY corporation. We gratefully acknowledge all our sponsors. The findings presented do not in any way represent, either directly or through implication, the policies of these organizations. Acknowledgements are also due to the members of the University of Minnesota, Northwestern University, University of Southern California, University of Illinois, University of Michigan, and University of Chicago.

### References


