

Assessing the Impact of Latency and Jitter on the Perceived Quality of Call of Duty Modern Warfare 2

Rahul Amin, France Jackson, Juan Gilbert, Jim Martin

Clemson University,
Clemson, SC

Terry Shaw

Cable Television Laboratories, Inc.,
Louisville, CO

Abstract. Jane McGonigal stated in her 2010 TED Talk that humans spend 3 billion hours a week playing video games around the planet. Americans alone devote 183 million hours per week to gaming. With numbers like these, it's no wonder why end user demands for bandwidth have increased exponentially and the potential for network congestion is always present. We conduct a user study that focuses on the question: "How much network impairment is acceptable before users are dissatisfied?" In particular, the main objective of our study is to measure a gamer's perceived Quality of Experience (QoE) for a real-time first person shooter (FPS) online game **Call of Duty Modern Warfare 2** in presence of varied levels of network congestion. We develop a Mean Opinion Score (MOS) metric to determine each gamers' QoE. We investigate the following hypothesis: The gamers' perceived QoE correlates to their skill level.

Keywords: Online Gaming; First Person Shooter Games; Network Impairment; Quality of Experience

1 Introduction

Resource allocation is the process by which network elements try to meet competing demands that applications have for network resources. While broadband cable networks have steadily increased the size of the 'pipe' available to subscribers, end user demand for bandwidth continues to grow at an exponential rate. Therefore, network operators must continue to expand the capacity of their networks in order to avoid congestion. However, it is impossible to avoid occasional network congestion. This congestion is most likely to occur at prime time when most subscribers are active and many different type of applications are simultaneously in use (with many subscribing households using multiple applications simultaneously). The response by each application to congestion differs widely. File transfer applications such as File Transfer Protocol (FTP) or Peer-to-Peer (P2P) file sharing are typically the most flexible in performance requirements. Applications that involve real-time end user interaction, such as video conferencing, VoIP (voice over IP), or online games are less flexible as

they generally require low latency. Emerging web-based video streaming applications are generally flexible; however, once performance thresholds are surpassed, perceived quality drops off quickly. Finally, over-the-top HD (high definition) voice and high quality video broadcasts (e.g. live sportscasts) are the most sensitive to fluctuation in bandwidth, latency and loss.

The research described in this paper focuses on network gaming applications. In particular, we study real-time online games. The academic community has studied the impact of latency and loss on real-time games such as first person shooter (FPS) or action games [1]-[7]. The wide range of previous results confirms that mapping network performance metrics such as latency and loss to an assessment of perceived quality is extremely difficult. The challenge is in part caused by the fact that each game responds to latency and loss in unique ways. Further, for a given game, specific modes of operations and even game play scenarios (which change quite frequently) will react differently to latency and loss. Perhaps the most problematic aspect of assessing user's quality of experience is due to the fact that a subjective assessment is a measure of how the gaming experience meets (or fails to meet) a user's expectation. Accurately quantifying expectations is very difficult as a gamer's expectation depends on past history and personal factors.

Our research is motivated by the observation that much of the prior work described in the literature has focused on objective studies. While subjective studies have been considered, there have been limited results. This is primarily because a widely accepted utility function that maps objective metric results to a perceived quality assessment has not been established. While the research described in this paper is primarily exploratory in nature, we consider the following hypothesis: *a gamer's perceived quality of experience correlates to his/her skill level.*

We focus on the widely played FPS game Call of Duty Modern Warfare II. We engage the student population at Clemson University to participate in a large scale experimental study designed to explore the impacts of network impairments on the perceived quality. We collect information from the participants including their self-assessed gaming expertise, scores achieved during the testing session, and assessment of perceived quality. Our results do support our hypothesis by clearly showing that user's level of experience defines and determines the sensitivity to latency. However, we find 'expert' gamers who are presumably the most sensitive to latency are also the most adept at compensating for impaired conditions.

This paper is organized as follows. Section 2 presents a literature survey and summarizes relevant related work. Section 3 provides a description of our experimental methodology. Section 4 documents our results and provides an analysis of our results. Finally, Section 5 provides conclusions including a discussion of the limitations of our study and possible next steps.

2 Background

The broad genre of multi-player, online games includes non-real-time games such as board games or fantasy games as well as real-time games such as first person shooter,

sports, or real-time strategy. The work in [5] presents a useful taxonomy for classifying real-time online games. The authors correlate the game's relationship between latency requirements and 'precision' (either movement or shooting) requirements. Further attributes include the interaction model and the game perspective. The interaction defines how a player interacts with the game world. The authors indicate the interaction is either through an avatar or with an omnipresent model. The avatar model exists at a particular location in the game context and is only involved with local interaction. The omnipresent model provides players a global view of the game. The game perspective is tied to the interaction model as it defines how players view the game world. For avatar interaction, such as in shooter games, a first or third person shooter perspective is common. For omnipresent interactions, such as in role-playing games, the perspective might be variable. A player might have access to aerial views as well as first or third person views.

Because real-time online games are sensitive to network and system performance, there has been a significant amount of prior academic research in this area. Many of these studies share our goals and methods. Works such as [1]-[7] all seek to understand the impact that loss and latency have on real-time online games. A commonly used evaluation approach is to emulate loss and latency in a controlled test-bed environment. Either objective measures such as game results or subjective tests based on Mean Opinion Scores (MOS) are used. The impacts of loss are arguably not as significant as those caused by latency because most games are designed to tolerate some amount of loss either by sending game state or control commands frequently or by utilizing error correction. Therefore, much of the research presented in the literature has focused on the impacts of latency.

In [1], the authors found that the ability to carry out 'precise' movement and shooting tasks is not significantly affected by loss rates or latency that are typical for the FPS game Unreal. They did observe players scores (in terms of kills and deaths) in some game modes drop as the latency increased beyond 100 milliseconds (ms). In [2], the authors focus on the effects of latency on World of Warcraft. They find that because the game is real time strategy (rather than first person shooter), a high level of induced latency (well beyond 100 ms) is noticeable but has negligible effect on the outcome. Other works have used methods involving subjective tests where gamers were asked about their perceived quality. In [7], as round trip times exceed 150 ms, the performance of Halo gamers decreased by about 50% while the perceived quality was still high. Further, games with over 500 ms of delay and 4% loss were considered unplayable. The work in [3] shows that in some scenarios (i.e., play situations of a specific FPS game), even small amounts of latency and jitter (as small as 60 ms) caused users to notice a drop in perceived quality.

The work examined in [5] attempted to group the games in prior studies based on model (avatar or omnipresent), perspective (first or third person or varied), and genre (FPS, Racing, Sports, RPT, RTS) to one of three levels of latency sensitivity. The thresholds of induced latency for high, medium and low sensitivity settings were estimated to be 100 ms, 500 ms, and 1,000 ms. More recent work [6] has addressed this by exploring the effects of latency on classes of games. Using the open source third-person shooter game BZFlag, the authors illustrate that the range of tolerated latency

can range by an order of magnitude (from less than 100 ms to 1.0 seconds) depending on the required precision for game play.

The work in [8] provides a useful framework for describing and assessing game quality of experience through the use of ‘perceptual view inconsistency’. The framework assumes that 1) Quality of Experience (QoE) is based on an overall assessment that is based on individual (and decomposable) perceptions; 2) Inconsistencies result from the disconnect between the gamer’s view and the server’s (or global) view; 3) An overall QoE assessment requires each objective inconsistency to be mapped to corresponding subjective values. The authors refer to prior measurement work and show that the degradation of various metrics of perceived quality, each of which are specific to certain segments of specific games, can be modeled as a half normal (Gaussian) curve.

Based on the wide range of results that exists in the related work, we draw the following motivating conclusions:

- The impacts of latency and loss depend not just on the class of game, but also on particular modes of play and on specific gaming contexts.
- Establishing the relationship between latency and loss network measures with perceived quality is the most challenging issue that remains largely unsolved.

Our study develops a methodology for exploring the complex interaction between network impairment, the impacts on FPS online games, and the quality of experience perceived by end users (i.e., the gamers).

3 Methodology

3.1 Experimental Setup

Fig. 1 represents the experimental setup. For this setup, there are two sets of game-players. The first set consists of two players, referred to as “Participants”, who play on a controlled network environment. Looking at Fig. 1, we emulate network impairment with the Linux *netem* capability¹ operating on our router. We add latency and jitter to packets arriving at the router destined for the ‘congested participant’. The other participant is connected directly to the university’s network and is referred to as the ‘uncongested participant’. The second set of players, referred to as “Gamers”, play from their home or dormitory room on campus during each session to maintain a controlled but realistic level of background competitors. The location of the game host is an experimental parameter. We designate either the congested participant or the uncongested participant as the host.

¹ Please refer to this site for further information on *netem*: <http://www.linuxfoundation.org/collaborate/workgroups/networking/netem>

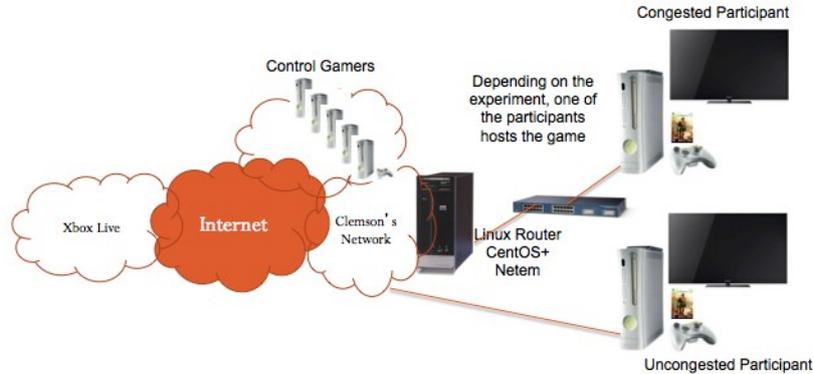


Fig. 1. Experiment Setup

3.2 Experimental Procedure

Before each game session, the two participants are asked to fill out a pre-survey questionnaire geared to determine their expertise level and to estimate their network performance expectation based on their normal gameplay environment. Each game session lasts for ten minutes where all participants and gamers play in the “Free for All” game mode, and then the participants answer a post-survey questionnaire that consists of four Likert Scale survey questions. The four post-survey questions deal with each gamer’s perception on following experiences: (i) Gameplay Satisfaction (ii) Gameplay Frustration (iii) Impact of Lag on Gameplay (iv) Likelihood to change network service providers. A MOS metric is computed for each participant using a weighted-sum approach where each of the four post-survey responses is given equal importance. The MOS metric derived for each participant is averaged to assess overall MOS (or QoE) for each network impairment setting. A detailed description of pre-survey questionnaire, post-survey questionnaire and MOS calculation is provided in [9].

3.3 Study Design

The study is divided into two phases: (i) Calibration phase (ii) Actual Experimentation phase. The ‘calibration’ phase involved a series of experiments designed to identify the set of network congestion settings that would be used in the actual study. The impact of jitter is quite different depending on which participant serves as the host. We conducted a pilot study in which we invited experienced Call of Duty Modern Warfare 2 players to play on several combinations of latency and jitter settings. In all experiments, 100 ms latency and a range of jitter in increments of 50 ms is added to all packets sent or received by the congested user. The results of the pilot study, presented in Fig 2, suggested that when the uncongested participant serves as the host, jitter in the range of [0, 300] ms is appropriate and when the congested participant serves as the host, jitter in the range of [0, 150] ms is appropriate. Moreover, the func-

tionality of *netem* was verified in the calibration phase of the study. A detailed description of these calibration phase findings is presented in [9]. Using the network congestion settings presented in Fig 2, a total of 56 experiments were conducted using university students as test subjects to obtain 7 MOS results per network congestion setting during the actual experimentations phase.

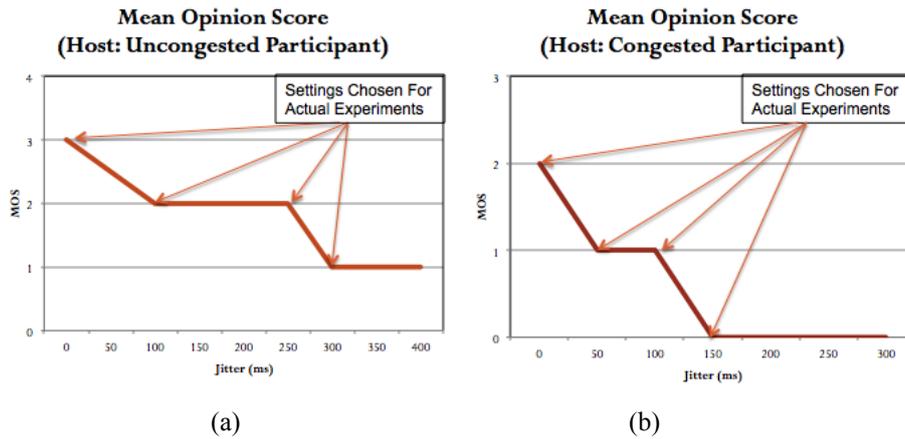


Fig. 2. Pilot Study Results

4 Results and Analysis

We performed 56 total experiments, which resulted in 112 survey responses. Figs 3, 4, and 5 profile the participants. Fig 3 indicates that the participants are reflective of the undergraduate population at Clemson University. The majority rely on wired rather than wireless connectivity. Fig 4 suggests that the participants represent a broad range of FPS gaming experience, ranging from novice to expert. Finally, further analysis of the pre-survey responses suggests that 10% (about 10) of the participants were complete novices to any online games and another 10% were experts with Call of Duty Modern Warfare as shown in Fig 5. The rest of the participants had some level of experience with FPS games.

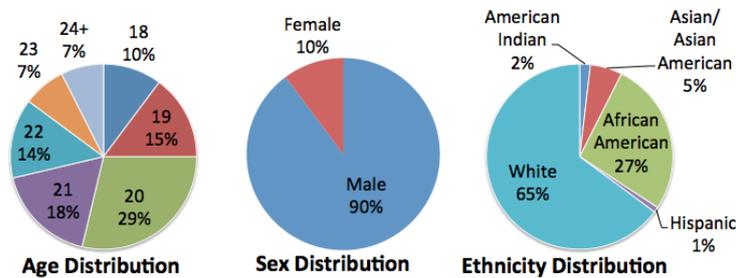


Fig. 3. Participant Demographics

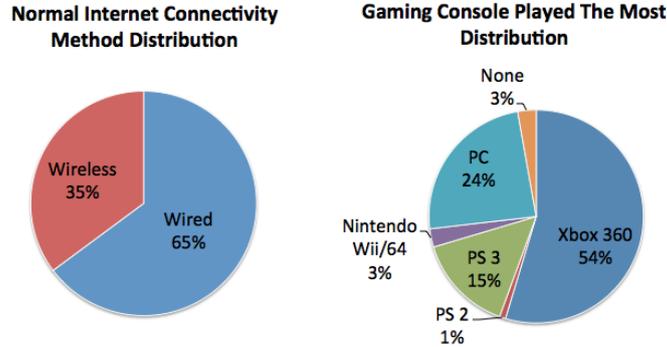


Fig. 4. Participant Internet Access Method

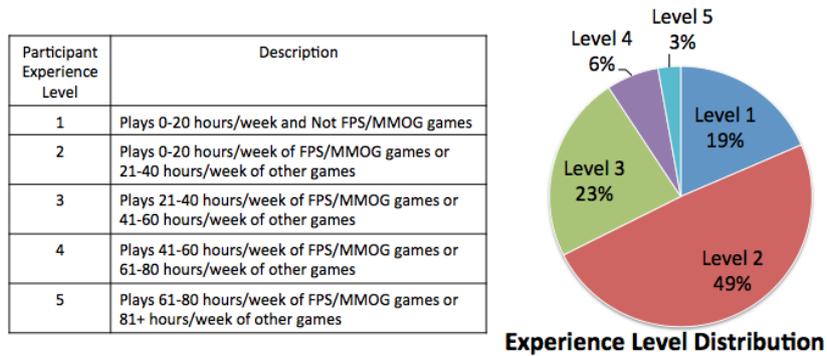


Fig. 5. Participant Gaming Experience

The results from actual experiments are shown in Fig 6(a) and 6(b) when the uncongested participant and congested participant are game hosts respectively. Note that only the results derived from post-survey responses of congested participants are presented in this study. The results derived from post-survey responses of uncongested participants are random in nature as these participants do not play under any network impairment and are presented in the extended version of this paper [9]. Fig 6(a) suggests that as jitter approaches 250 ms, the QoE degrades to a level (MOS < 3) that make the gamers' experience unacceptable when the game host is the uncongested participant. Fig 6(b) suggests that as jitter approaches 100 ms, the QoE degrades to a level (MOS < 3) that makes the gamers' experience unacceptable when the game host is the congested user. We observe that the MOS in Fig 6(b) drops as jitter is increased but not as definitively as in Fig 6(a). This happens because the congested participant is the game host in Fig 6(b). For this setting, the command and control data that the congested participant sends is actually avoiding the *netem* latency effects since the server is located at the local host. We conjecture that the outbound traffic stream is more sensitive to latency than the inbound traffic stream.

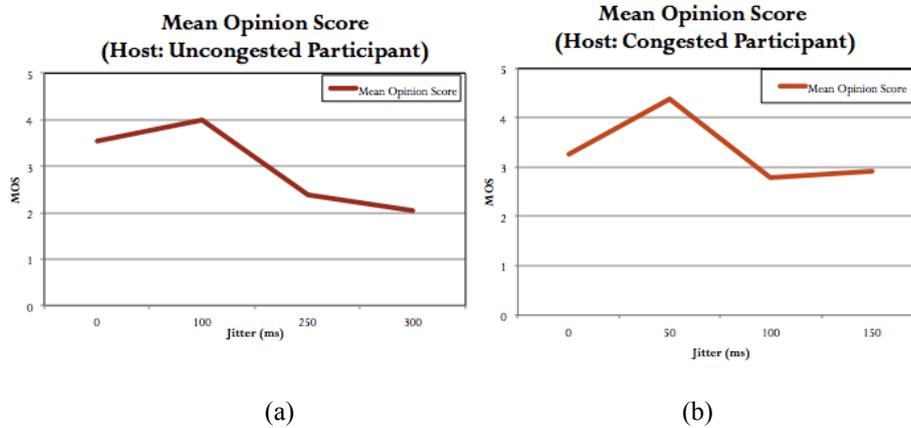
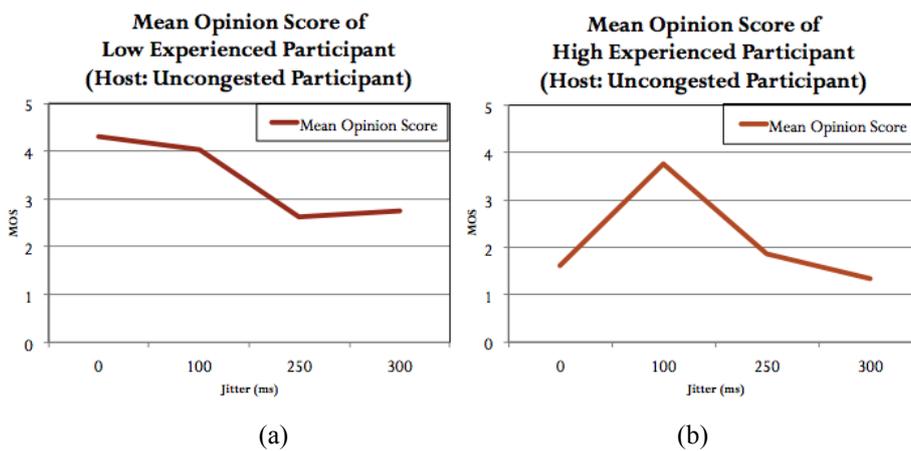


Fig. 6. Overall Actual Experimentation Results

The results that compare the responses provided by experienced gamers vs. inexperienced gamers are presented in Figs 7(a)-(d). For both cases, when the game host is uncongested or congested participant, the results support our hypothesis by clearly suggesting that participant's level of expertise defines and determines the sensitivity to network congestion. The more experienced the participant (Figs 7(b) or 7(d)), the less tolerant he/she is to network congestion compared to the less experienced participants (Figs 7(a) or 7(c)). Also, the 'expert' or highly experienced gamers significantly berate even the lowest network congestion setting of 0 jitter/100 ms latency as seen in Figs 7(b) and 7(d) and have a hard time quantifying the level of congestion when the congestion settings are low.



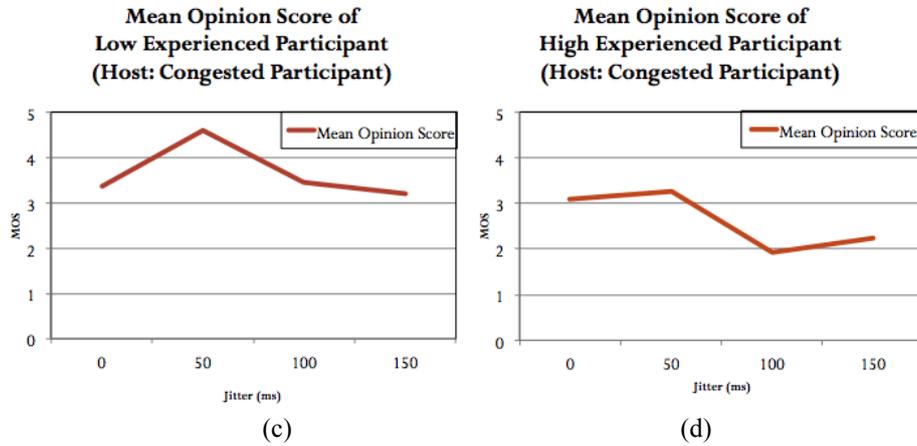


Fig. 7. Results based on Participant's Expertise Level

5 Conclusions

We have studied the impact of latency and jitter on the perceived quality of the FPS game Call of Duty Modern Warfare 2. Our experimental methodology involved college students (gamers) playing a closed session of "Free for all" game mode. A total of seven gamers were involved in each game. Two gamers (referred to as participants) were under observation, one of whom had his/her network impaired in a controlled manner. The participants under observation provided a numeric assessment of their satisfaction, level of frustration, impact of lag and likeliness to change network service providers.

Our results suggest that, for the cases when the host is the uncongested user and the user under study is the congested user, a random jitter that is in a range of [0, 250] ms leads to a MOS less than 3. For the case when the host is the congested user and the user under study is the congested user, a random jitter in the range [0, 100] ms leads to a perceived MOS less than 3. Our original hypothesis was that *the perceived quality correlates to a gamer's skill level*. We did observe different thresholds of tolerance of network impairment based on the level of experience and skill. The more experienced the participant, the less tolerant he/she is to network congestion. The 'expert' or highly experienced gamers significantly berate even the lowest level of network impairment (with constant 100 ms latency and no jitter). However, these 'expert' gamers who are presumably the most sensitive to latency/jitter are also the most adept at compensating for impaired conditions as observed from their final game scores (which are not presented in this paper).

Our methodology has several limitations. First the number of user's in the study was not sufficient. While most data points in the results curves were based on 7 samples, some results (in particular those shown in Figs 7(a)-(d)) are based on only 2-3 samples. Second, our methodology did not precondition the test subjects. In other words, ideally, we would have a participant take our survey in his/her usual gaming

environment. This provides a familiar context with which a user can base his/her assessment. In our next steps we would like to take a small set experienced Call of Duty gamers and first train their expectations and game play on ‘ideal’ conditions. Then, each gamer will periodically be asked their opinion allowing us to fully map study participants to their perceived quality based on controlled network settings.

References

1. T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, M. Claypool, “The Effects of Loss and Latency on User Performance in Unreal Tournament 2003”, Proceedings of 3rd ACM SIGCOMM Workshop on Network and System Support for Games (NetGames-04), August, 2004.
2. N. Sheldon, E. Girard, S. Borg, M. Claypool, E. Agu, “The Effect of Latency on User Performance in Warcraft III”, In *Proceedings of ACM Network and System Support for Games Workshop (NetGames)*, Redwood City, California, USA, May 2003.
3. P. Quax, P. Monsieurs, W. Lamotte, D. Vleeschauwer, N. Degrande, “Objective and Subjective Evaluation of the Influence of Small Amounts of Delay and Jitter on a Recent First Person Shooter Game”, Proceedings of the SIGCOMM 2004, August, 2004.
4. R. Woundy, Y. Lee, A. Veiga, C. Williams, “Congestion Sensitivity of Real-time Residential Internet Applications”, Proceedings of the SCTE Cable-Tec Technical Workshop, November 2010.
5. M. Claypool, K. Claypool, “Latency and Player Actions in Online Games”, ACM Communications of the ACM, November 2006.
6. M. Claypool, K. Claypool, “Latency Can Kill: Precision and Deadline in Online Games”, Proceedings of Multimedia Systems Conference (MMSys’10), February, 2010.
7. T. Lang, G. Armitage, “A Ns2 Model for the Xbox System Link Game Halo”, Proceedings of the Australian Telecommunications Networks and Applications Conference, 2003.
8. P. Chan, M. Zarki, “Perceptual View Inconsistency: An Objective Evaluation Framework for Online Game Quality of Experience (QoE)”, ACM NetGames ’11 Workshop, October, 2011.
9. R. Amin, F. Jackson, J. Gilbert, J. Martin, T. Shaw, “Assessing the Impact of Loss and Latency on the Perceived Quality of Call of Duty Modern Warfare 2,” Online. Available at <http://people.cs.clemson.edu/~jmarty/projects/GamingAssessments/GamingStudyReportV6.0.pdf>, last accessed January 3, 2013.