CAUSAL ANALYSIS OF THE UNSATISFYING EXPERIENCE IN REALTIME MOBILE MULTIPLAYER GAMES IN THE WILD

Yuan Meng[1,4] *Shenglin Zhang*[2,*], *Zijie Ye*[1],
Benliang Wang[2], Zhi Wang[1], Yongqian Sun[2], Qitong Liu[3], Shuai Yang[3], Dan Pei[1,4]

[1]Tsinghua University [2]Nankai University [3]Tencent
[4]Beijing National Research Center for Information Science and Technology(BNRist)

ABSTRACT

There are anecdotal evidences that realtime mobile multiplayer games (RMMGs), which require realtime interactions, suffer from unsatisfying experience in the wild. This paper presents the first measurement results of such experience based on 12 million real game sessions from a top-tier RMMG. We observe that 13% of the game sessions suffer from at least one location resynchronization, and 7.12% have been aborted abnormally before the end of the game. This paper thus proposes *ExCause*, a general causal analysis framework to systematically analyze historical game session records to 1) obtain context factors that cause unsatisfying RMMG experience, and 2) recommend adjustments with quantified expectation of QoE improvement, by applying the *potential outcome framework*. The recommendations suggested by *ExCause* can reduce the number of location resynchronization by 95.1%, from 1.323 to 0.065 on average. Furthermore, *ExCause* enables us to rectify some misperceptions from previous correlation-based studies.

***Index Terms*—** Online Mobile Game, QoE metrics, Causal analysis

1. INTRODUCTION

Driven by the rapid development of wireless network and the popularity of smartphones, mobile gaming dominates all other types of online gaming in terms of popularity [1] . Realtime mobile multiplayer games (RMMG) (*e.g.,* VainGlory, Arcane Legend) is one of the most popular mobile game types. For example, as one of the most popular RMMG genres, First-Person Shooter (FPS) alone accounts for 25.9% of entire game sales in US in 2017 [2]. During a RMMG game session[1], multiple players explore the same virtual world and interact with each other in real time, which requires high quality of experience (QoE). Although there are anecdotal evidences that a large number of RMMG game sessions suffer from unsatisfying QoE [4], in the literature there is neither quantitative measurement of unsatisfying RMMG experience in the wild, nor the causal analysis of these unsatisfying experience.

In this paper, we first propose two QoE metrics that capture the unsatisfying experience in a RMMG session: **LRC** (the number of Location Resynchronizations experienced by a player in a session) and **AQ** (whether a player quits the session abnormally). An LR (Location Resynchronization) is an event when the player's character is forced (by the game) to "jump" instantaneously from one state/location to another on the game client in order to synchronize with the state/location on the server side. AQ (abnormal quit) happens when the player quits the game session before the end of the game, either because the game client times out (*e.g.*, when loading the virtual world map), or because the player intentionally quits the game due to the frustration from bad experience. This paper presents the first measurement results of these two metrics in the wild: in 12 million sessions from the top-tier RMMG studied in this paper, 13% of the game sessions suffer from at least one location resynchronization, and 7.12% of game sessions suffer from AQs. **This is the first contribution of the paper**.

Despite the prevalence of unsatisfying RMMG experience in the wild, their **causes** have not yet been studied. Note that there is a difference between causality and correlation. On the one hand, "A causes B" means that, everything else being equal, A's change will cause B to change. On the other hand, "A and B are correlated" means that A and B change together, but A and B's changes might be both caused by a common cause C, which means there is no causal relationship between A and B. Operators of the RMMG would like to obtain the causes of the unsatisfying experience so that they can recommend some adjustment to the players, *e.g.*, switching from 3G to WiFi, or lower the image quality, so that they obtain better experience.

This paper thus proposes, *ExCause*, a general causal analysis framework to systematically analyze historical game session records to 1) obtain context factors that cause unsatisfying RMMG experience and 2) to recommend adjustments with quantified QoE improvement expectation. For the first time in the literature, *ExCause* applies the potential outcome framework (POF) [5] (a causal inference framework) in game experience study. *ExCause* considers adjustable (thus recom-

[*]Shenglin Zhang is the corresponding author.

[1]A RMMG session is a process including matching other players and PvP (player versus player) playing until the end of the game [3].

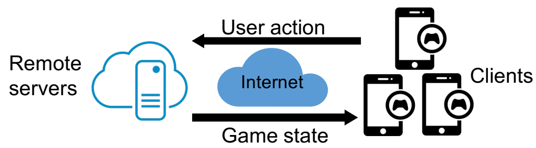

Fig. 1. The framework of RMMG

mendable) factors (*e.g.*, WiFi vs. 3G) as causes and unadjustable factors (*e.g.*, number of CPU cores in a smartphone) as confounding factors, which removes the recommendation effect bias induced by the unadjustable factors. Such a framework, we believe, is general enough to be applicable to other game genres or mobile Apps. **This is the second contribution of the paper.**

We apply *ExCause* to a top-tier RMMG game and observe that *ExCause*'s recommendation can have very good expected improvement. For example, adjusting the Android OS version can reduce LRC by 95.1% from 1.323 to 0.065 on average. In addition, we observe that OS version and pixel density are the top causes that lead to unsatisfying QoE in the studied RMMG, instead of access network type, commonly suspected as top causes by previous correlation-based studies [3, 6, 7]. This highlights the necessities of using causal analysis instead of correlation analysis when recommending adjustments. **Above empirical results are the third contribution of the paper.**

The rest of the paper is organized as follows. Section 2 gives an introduction to QoE metrics and context factors. The framework of *ExCause* is depicted in Section 3. Section 4 shows the results, followed by the related work in Section 5 and conclusion in Section 6.

2. QOE METRICS AND CONTEXT FACTORS

This section first defines two QoE metrics for RMMG, and then shows the measurement result of unsatisfying RMMG QoE in the wild. We then depict the adjustable and unadjustable context factors that most affect RMMG QoE.

2.1. QoE metrics

A RMMG is a type of online game that allows multiple mobile players to cooperate and compete with each other in a virtual world, and sometimes to interact meaningfully with people around the physical world. It includes a variety of gameplay genres, including role-playing, first-person shooter, real-time strategy, simulations of sports or racing, *etc.* Figure 1 shows the architecture of RMMG. In a game session, the remote servers maintain the states (*e.g.,* locations/poses/actions in the virtual world) of every player, conduct complex calculations based on the game logic, and send the latest information to every player. Each client receives information from the remote servers, updates game states, renders new game videos, and sends user actions to the remote servers.

An LR is a situation where a player's state on the remote server is greatly different from that on the client. If an LR occurs, a player's game character on his game client is forced to suddenly "jump" from one state/position to another, in order to synchronize with the state on the servers. LR is definitely an unsatisfying experience in RMMG.

A player usually quits the client when the entire game session finishes. An AQ happens when the player quits the game session before the end of the game, either because the game client times out (*e.g.*, when loading the virtual world map), or because the player intentionally quits the game due to the frustration by bad experience such as image stutter. All these cases will be treated as AQ events. Usually, an AQ of a player can in turn affects the QoE of his/her teammates in the same cooperative game session. Because AQs can be the symptoms of malicious behaviors as well, game operators typically penalize frequent AQs by taking away some credits from the corresponding players. AQ thus is definitely an unsatisfying experience in RMMG.

For a game session, we define the number of LR event experienced by a player as LR count (LRC). As the two QoE metrics, LRC and the abnormal quit state AQ capture the unsatisfying experience in a game session.

2.2. Dataset and unsatisfying user experience in the wild

The dataset used in this paper is collected from the FPS game of a top-tier global RMMG company. The SDK implemented in each client continuously collects the game information including QoE information (LR and AQ), and the levels of context factors, *e.g.,* ISP, OS version. After that, it uploads this information to a centralized server for data aggregation. We collected all the records of 1.2 billion game sessions within 30 days. Because analyzing such a great number of records would consume too much computational resources, and 1% of the dataset (12 million records) is large enough to be statistically significant, we randomly sampled 1% from the dataset, which is used throughout the paper.

From the cumulative distribution function (CDF) in Figure 2, we can see that over 13% of game sessions suffer from at least one LR, and 1.3% of game sessions suffer even more than five LRs. If we can decrease the ratio of game sessions that experience LR, the QoE of a great number of players will be improved. In addition, players quit abnormally in 7.12% of the game sessions. Above quantitative results show that LRs and AQs are indeed prevalent in the wild.

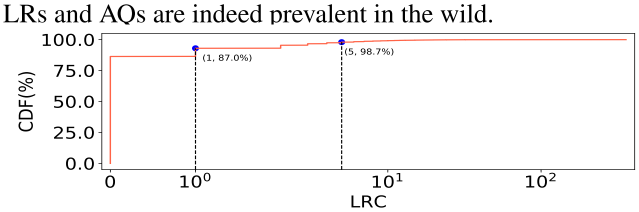

Fig. 2. The CDF of the LRC.

2.3. Context factors

Diverse types of context factors can impact the QoE of a game session [3]. In this work, we analyze the context factors that are most likely to impact players' QoE based on operators' experience, as summarized in Table 1.

According to operators' experience, some context factors are adjustable to players, and the others are unadjustable. The adjustable context factors can be used to make reasonable recommendations to players. That is, we can recommend players to adjust context factors in order to greatly improve QoE. On the other hand, we should consider the impact of unadjustable context factors when we try to locate the causal adjustable context factor. The adjustable context factors are justified as follows. (1) **TAN**: Usually, a player can change the TAN of the mobile device via system preferences. For example, a player can switch from 2G to WiFi for better QoE. (2) **IQ**: The IQ level of each game session can be configured. Generally, a higher level of IQ leads to rendering a larger size of image file every time. (3) **PD**: An Android player can configure the PD of the mobile device in the system preferences. Thus this factor is adjustable for the Android players but unadjustable for the iOS players. PD can be classified into the following five levels [8]: LDPI, MDPI, HDPI, XHDPI, XXHDPI. (4) **OSV**: In general, a player can upgrade the OSV of the mobile device in the system preferences, and downgrade it with third-party tools [9]. (5) **ISP**: The ISP of the network can be easily changed for the mobile devices that have dual SIM cards. In addition, a player can also change the ISP without having to change the phone number.

3. RECOMMENDATION FOR THE ADJUSTABLE SYSTEM SETTINGS VIA CAUSAL ANALYSIS

As aforementioned, a great number of game sessions suffer from unsatisfying QoE, and multiple context factors can impact QoE [3]. If a player who is unsatisfied with the QoE adjusts the context factors to the "right" ones, the QoE will be improved. In this work, we propose *ExCause*, a framework to automatically give recommendations to the players who suffer from unsatisfying QoE. *ExCause* includes two components: (1) identifying the critical sets with unsatisfying QoE, and (2) making recommendations based on causal analysis.

3.1. Identifying the critical sets with unsatisfying QoE

Each game session has a combination of context factors, *e.g.,* {TAN = 4G, IQ = HIGH, PD = XHDPI, OSV = iOS 12, ISP = China Mobile}. If two game sessions have the same combination of context factors, we group them together in the same set. Considering the noises in data collection and the cost of making recommendations, in this work we analyze the sets with more than 100 game sessions.

A game session is unsatisfying if its $LRC > 0$ or there occurs an AQ. If a set has relatively high number and ratio of unsatisfying game sessions, we call it a *critical set*. Formally, a critical set is a set with $\frac{\#unsatisfying\ game\ sessions}{\#game\ sessions} > \frac{\#all\ the\ unsatisfying\ game\ sessions}{\#all\ the\ game\ sessions}$, and if the OS is iOS, then $\frac{\#unsatisfying\ game\ sessions}{\#all\ the\ unsatisfying\ game\ sessions\ of\ iOS} > 0.1\%$, and $\frac{\#unsatisfying\ game\ sessions}{\#all\ the\ unsatisfying\ game\ sessions\ of\ Android} > 0.1\%$ if the OS is Android. Because the total game sessions on Android greatly outnumber those on iOS, we divide game sessions based on OS when defining critical set.

3.2. Making recommendation based on causal analysis

In this section, we explore how to make recommendations to improve the QoE for critical set the most. According to the theory of causal and effect [10], people usually make decisions based on causal intuition. Intuitively, context factors can be the causes of QoE's degradation or improvement, which can be used to give recommendations to improve QoE.

A primary technique to find the causal context factor is to design a controlled experiment. In order to design a true experiment in our context, players have to be randomly assigned to different levels (values) of context factors (*e.g.,* access Internet with 3G or WiFi) and observe the resultant QoE of game sessions. However, conducting such an experiment at scale is prohibitively hard and expensive, or even impossible. Moreover, it also has legal, ethical, and other issues if one would intentionally degrade the QoE of a set of players.

Therefore, we make recommendations for critical sets based on analyzing a large collection of historical records of game sessions. Specifically, we locate the causal context factors for critical sets using Potential Outcome Framework (POF) [5], which has been widely applied in social and biomedical sciences. To the best of our knowledge, this is the first time that POF is used in analyzing the QoE of mobile games. Such a framework, we believe, is general enough to be applicable to other game genres or mobile apps.

3.2.1. Causal analysis based on historical dataset

In POF [5], only one of the potential outcomes ($Y_i(1)$ or $Y_i(0)$) will be obtained because of the absence of random controlled trials. Intuitively, we cannot get an unbiased estimate of the Average Treatment Effect (ATE) by only comparing the outcomes of the treatment group ($Z_i = 1$) and those of the control group ($Z_i = 0$). That is because the confounders (X_i) can affect each game session's both treatment state (Z_i) and outcome (Y_i).

However, the causal effect can be calculated with confounders under the two following assumptions: (1) **The stable unit treatment value assumption (SUTVA)**. The potential outcome observation on one game session should be unaffected by the particular assignment of the treatments to the other game sessions. (2) **Unconfoundness**. Given the observation confounders X, the treatment assignment is independent with the potential outcome, which can be formally denoted as $Z_i \perp \{Y_i(1), Y_i(0)\} \Big| X_i$, then $ATE = E[E(Y|X_i, Z_i = 1)] - E[E(Y|X_i, Z_i = 0)]$.

Our context basically follows the above two assumptions, and thus we can use the following framework. Based on the above two assumptions, Rosenbaum and Rubin defined the propensity score $e(X_i)$, which is the probability that the

Table 1. The context factors that are most likely to impact a player's QoE.

Type	Name	Description
Adjustable	TAN	The type of the access network, including WiFi/4G/3G/2G.
	ISP	The Internet Service Provider of the access network.
	IQ	The image quality level of the game session, including very low, low, high, very high.
	OSV	The OS version of the mobile device, like Android8, Android7, iOS10, *etc*.
	PD*	The pixel density of the mobile device, which is only adjustable for the Android users.
Unadjustable	RAM	The random access memory of the mobile device.
	# CPU Cores	The number of the CPU cores of the mobile device.
	Game Type (GT)	The type of the game, including PVP (player versus player) or PVE (player versus environment).
	Device Model (DM)	The model of the mobile device.
	Province	The province where the player is located.
	OpenGL Version (OGLV)	The version of the Open Graphics Library (OpenGL) of the mobile device.

sample is assigned to the treatment group [11] as $e(X_i) = P(Z_i = 1|X_i)$ Essentially, $e(X_i)$ is the result of the dimensionality reduction of X. We estimate $e(X_i)$ using logistic regression. Therefore, we can get $Z_i \perp \{Y_i(1), Y_i(0)\} \Big| e(X_i)$.

After that, we use the inverse probability weighting method to estimate ATE and the average treatment effect for the contro(ATC) [12, 13] as

$$\widehat{ATC} = \frac{\sum_{i=1}^{N} Z_i Y_i (1-e(X_i))/e(X_i)}{\sum_{i=1}^{N} Z_i (1-e(X_i))/e(X_i)} - \frac{\sum_{i=1}^{N} (1-Z_i) Y_i}{\sum_{i=1}^{N} (1-Z_i)} \quad (1)$$

Note that ATE is the effect of moving the entire population of game sessions from a control group to a treatment group. In addition, ATC is the average treatment effect on the game sessions which are initially in the control group, which is denoted formally as $E[Y_i(1) - Y_i(0)|Z_i = 0]$.

From Eq. (1), we can see that first term of ATC is the mean outcome of the game sessions in the treat group with the weighting of $(1 - e_i)/e_i$, and the second term is the mean outcome of the game sessions in the control group.

3.2.2. Recommendation via propensity score weighting

In this work, each critical set is a control group ($Z = 0$), and the other sets can be used as treatment groups ($Z = 1$). To minimally change the context factors, for each critical set we change the value of only one context factor at one time to construct the treatment group. For example, Set A = {**TAN = 3G**, IQ = HIGH, PD = XHDPI, OSV = iOS 12, ISP = China Mobile} is a critical set. Set B = {**TAN = 4G**, IQ = HIGH, PD = XHDPI, OSV = iOS 12, ISP = China Mobile} is one treatment group which only changes Set A's TAN from 3G to 4G.

The estimator of ATC in Eq. (1) is used to evaluate the potential QoE improvement of each treatment group (*i.e.,* recommendation set). Eventually, we will find the the treatment group that can most improve the QoE of the critical set, which thus has the largest ATC. In the critical set, we have considered all the five adjustable context factors. However, as aforementioned, in addition to adjustable context factors, unadjustable ones (see Table 1) can also impact QoE. Moreover, the unadjustable context factors can have causal impact on adjustable context factors. For example, device model (DM) can impact OS version (OSV) and pixel density (PD). Therefore, all six unadjustable context factors are used as confounders in POF.

In this study, there may be other unobservable factors impacting the QoE, which are not included in the confounders. This is a general caveat that holds for all science domains that attempt to infer causality from observational data. We did our best to avoid the negative impact of unobservable confounders by trying to observe as many factors as possible.

4. RESULTS

For each of the identified critical sets, *ExCause* recommends one adjustable factor (change factor) that can improve the QoE of the critical set the most. Table 2 shows the average recommendation effect per (metric, OS, change factor) tuple for all the identified critical sets. For example, the first of four rows in bold shows that, for 40 Android critical sets, the recommended change factor is OSV, which can reduce the average LRC by 95.1% from 1.327 to 0.069.

From this table, we can observe that OSV is the most frequent recommendation change factor among the five adjustable factors. The second is the PD for the Android users. The improvement ratio is larger than 90% for Android users with the change of OSV or PD according to both LRC and AQ. It means that the client device factors are the most guilty for the unsatisfying user experience of online mobile game, instead of TAN, which is commonly suspected as top cause by previous correlation-based studies [3, 6, 7].

Table 3 shows the average (across all adjustable factors and critical sets) effect by top recommendations. We can see that the average of *LRC* in one game session can be reduced by 93.2% (92.8%) from 1.301 (2.66) to 0.089 (0.19) for Android (iOS) devices. Similarly, the *AQ* ratio is reduced by 82.6% (36.8%) from 7.5% to 1.3% for Android (iOS) devices.

We also expand the four bold rows in Table 2 into Table 4 to investigate the details of the recommendation about the OSV and PD for the Android users, who have more freedom to change OSV and PD than iOS users. From Table 4,

Table 2. The recommendation result for the critical settings. IR is the improvement ratio.

Performance	OS	Change factor	Sets Count	Original	ATC	IR
Avg LRC	iOS	IQ	13	3.216	-3.025	94.1%
		OSV	6	1.424	-1.246	87.5%
		TAN	4	8.833	-8.563	96.9%
		ISP	2	0.743	-0.534	71.9%
	Android	**OSV**	**40**	**1.323**	**-1.258**	**95.1%**
		PD	**17**	**1.327**	**-1.234**	**92.9%**
		IQ	7	1.275	-1.101	86.4%
		ISP	6	0.766	-0.589	76.9%
		TAN	6	1.556	-1.343	86.3%
AQ ratio	iOS	OSV	26	0.253	-0.090	35.5%
		TAN	13	0.261	-0.070	26.8%
		IQ	12	0.226	-0.142	62.8%
		ISP	4	0.257	-0.074	28.8%
	Android	**OSV**	**14**	**0.075**	**-0.068**	**90.7%**
		PD	**12**	**0.077**	**-0.057**	**74.0%**
		TAN	5	0.074	-0.046	62.2%
		ISP	1	0.081	-0.046	56.8%
		IQ	1	0.076	-0.057	75.0%

Table 3. The recommendation effect result of critical sets.

Performance	OS	Original	ATC	IR
Avg LRC	iOS	2.66	-2.47	92.8%
	Android	1.301	-1.212	93.2%
AQ ratio	iOS	0.225	-0.083	36.8%
	Android	0.075	-0.062	82.6%

we can see there are 32 and 10 sets where the recommendation is upgrading OSV according to the *LRC* and *AQ*, but still 8 and 4 cases whose recommendation is to downgrade the OSV (highlighted in bold), respectively. For the PD, almost all critical settings are recommended to lower the PD. It is consistent with the fact that the lower PD, the smoother the frames are, and the less *LR*. The result shows that a too high PD may also lead to a high *AQ* ratio, which may be caused by the overload of the mobile devices.

Table 4. The details of the recommendation with the change of OSV and PD.

Performance	Change factor	Details	Sets Count
Avg LRC	OSV	Upgrade OSV	32
		Downgrade OSV	8
	PD	Higher PD	1
		Lower PD	16
AQ ratio	OSV	Upgrade OSV	10
		Downgrade OSV	4
	PD	Higher PD	1
		Lower PD	11

Benefits of causal analysis: in Table 4, we can see there are some critical sets with the recommendation of downgrading the OSV. It seems to contradict with common sense that a higher OSV has better performance than a lower OSV. We investigate the reason through the case study. For the critical set: {TAN=3G, IQ=Low, PD=XHDPI, OSV=Android 8, ISP=China Union}, which has 1.526 times of *LR* event in one play on the average. Under the recommendation of downgrading the OSV from Android 8.0 to Android 5.0, the estimated average *LRC* is nearly 0 (ATC=1.526), which is better than the recommendation of replacing the TAN 3G with WiFi (ATC=0.515). We found that the plays in this critical set mostly happened in some provinces where low-end devices are much more used. The unsatisfying experience in this critical set is because higher version OSes are running on low-end devices, consistent with *ExCause*'s recommendation (downgrading the OSV).

However, should we conduct a correlation-based study as some of previous studies did? As shown in Figure 3, the correlation between the Android OSV and *LRC* in the whole population is negative. That is, the higher OSV means the less *LRC*. This misleading results are due to the lack of causal analysis and not considering the Province and the Device Model as confounders. This case highlights that the causal analysis is necessary for decision making based on the data-driven analysis.

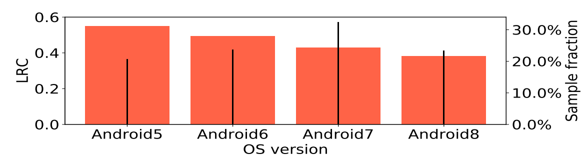

Fig. 3. The correlation between *LRC* and Android OSV.

5. RELATED WORK

PC and Mobile game QoE: The analysis of *PC* game QoE in studies such as [6, 7] are via data visualizations. [3] uses frame latency performance to study *mobile* game QoE, but frame latency might not be observable by the players, in contrast to the definitely observable location resynchronizations and abnormal quits in this paper. Furthermore, [3] uses the Kendall Correlation and the Information Gain to quantify the relationship potential factors and QoS metrics. All of the above studies are essentially correlation-based. Our results in Section 4 highlight the necessities of using causal analysis instead of correlation analysis when recommending adjustments because correlation does not equal causality.

Causal analysis in the QoE study: Causal analysis is popular in the field of social and biology [5, 14]. However, to the best of our knowledge, [15] is the only causal analysis study on application QoE. It uses *quasi-experiment design* to investigate the causes of online video streaming QoE. More specifically, [15] uses the *exact matching confounders method* to construct the treat and control datasets with similar confounder distributions, but drops the samples that cannot exactly match confounders in either dataset, inevitably introducing some bias. The above method is not applicable to our scenario, because our goal is to estimate the recommendation effect under *original* confounder distributions. Instead, we use the *propensity score weighting* method to solve the unmatched confounder problem.

6. CONCLUSION

This paper presents the first measurement results of unsatisfying RMMG experience in the wild and quantitatively shows that location resynchronizations and abnormal quits are prevalent in a top RMMG game. Then this paper proposes *ExCause*, a general causal analysis framework to systematically analyze historical game session records to 1) obtain context factors that cause unsatisfying RMMG experience, and 2) recommend adjustments with quantified expectation of QoE improvement, by applying the *potential outcome framework*. The recommendations suggested by *ExCause* can reduce the number of location resynchronization by 95.1%, from 1.323 to 0.065 on average. Furthermore, *ExCause* enables us to rectify some mis-perceptions from previous correlation-based studies. We believe that *ExCause* framework can be applied to other game types and other Apps.

7. ACKNOWLEDGEMENT

We thank the anonymous reviewers for their valuable feedbacks. We thank Juexing Liao and Ruchun Wu for their helpful suggestions and proofreading. This work has been supported by the Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology, the National Natural Science Foundation of China (NSFC) under grant No. 61872215, the Fundamental Research Funds for the Central Universities under grant No. 63191427, the Okawa Research Grant, CERNET Technological Innovation Project for Next Generation Internet under grant No. NGII20180121, and SZSTI under Grant No. JCYJ20180306174057899.

8. REFERENCES

[1] Newzoo, “Number of mobile game market,” https://newzoo.com, Accessed July 16, 2018.

[2] Statistics, “Number of mobile gamers forecast,” https://www.statista.com, Accessed July 16, 2018.

[3] Chou Mo, Guowei Zhu, Zhi Wang, and Wenwu Zhu, “Understanding gaming experience in mobile multiplayer online battle arena games,” in *Proceedings of the 28th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video - NOSSDAV ’18*. pp. 25–30, ACM Press.

[4] Reddit, “Game stutter during an active fight,” https://www.reddit.com/r/FortNiteBR/comments/73vnek/game_stutter_during_an_active_fight, Accessed July 16, 2018.

[5] Carol Joyce Blumberg, “Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction: Book Reviews,” vol. 84, no. 1, pp. 159–159.

[6] Oliver Hohlfeld, Hannes Fiedler, Enric Pujol, and Dennis Guse, “Insensitivity to Network Delay: Minecraft Gaming Experience of Casual Gamers,” in *2016 28th International Teletraffic Congress (ITC 28)*. pp. 31–33, IEEE.

[7] Maryam Amiri, Hussein Al Osman, and Shervin Shirmohammadi, “Game-aware bandwidth allocation for home gateways,” in *2017 15th Annual Workshop on Network and Systems Support for Games (NetGames)*. pp. 1–3, IEEE.

[8] Android Developers, “The pixel density level for the android phone,” https://developer.android.com/guide/topics/resources/providing-resources#AlternativeResources, Accessed July 16, 2018.

[9] Jerry Hildenbrand, “How to downgrade your android phone to a previous version,” https://www.androidcentral.com/downgrade-android-help, Accessed July 16, 2018.

[10] Judea Pearl, *Causality, Models, Reasoning, and Inference*, Cambridge University Press, 2009.

[11] Paul R. Rosenbaum and Donald B. Rubin, “Reducing bias in observational studies using subclassification on the propensity score,” *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 516–524, 1984.

[12] Younathan Abdia, K. B. Kulasekera, Somnath Datta, Maxwell Boakye, and Maiying Kong, “Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study: Propensity scores based methods for estimating average treatment,” vol. 59, no. 5, pp. 967–985.

[13] Peter C. Austin, “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies,” vol. 46, no. 3, pp. 399–424.

[14] Edward K. Kao, “Causal Inference Under Network Interference: A Framework for Experiments on Social Networks,” *arXiv:1708.08522 [cs, math, stat]*, Aug. 2017.

[15] S. Shunmuga Krishnan and Ramesh K. Sitaraman, “Video Stream Quality Impacts Viewer Behavior: Inferring Causality Using Quasi-Experimental Designs,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 2001–2014, Dec. 2013.