Creating Synthetic Data for Network Game User QoS Sensitivity Trials

Sebastian Zander, Grenville Armitage
Centre for Advanced Internet Architectures, Technical Report 160422B
Swinburne University of Technology
Melbourne, Australia
s.zander@murdoch.edu.au, garmitage@swin.edu.au

***Abstract*—In 2003 we conducted trials with the interactive online Xbox game Halo where we subjected the players to different emulated network conditions in order to find out how the different conditions impacted on their play. We varied the emulated network delay and packet loss between different games and measured how the different network conditions affected the players subjectively (mean opinion score) and objectively (in-game kill/death statistics). Results of the analysis were published in [1]. The ethics approval under which the trials were conducted does not allow us to publish the original data collected during the trials. In this report we describe a simple technique to create an unidentifiable completely synthetic dataset with similar properties as the original dataset (at least as far as simple analysis is concerned). The synthetic dataset is publicly available at XXX.**

***Index Terms*—Network Games, User QoS Sensitivity, Synthetic Data**

I. Introduction

In 2003 we conducted trials with the interactive online Xbox game Halo where we subjected the players to different emulated network conditions in order to find out how the different conditions impacted on their play. We varied the emulated and network delay and packet loss between different games and measured how the different network conditions affected the players subjectively (mean opinion score) and objectively (in-game kill/death statistics). The results of the analysis were published in [1].

The ethics approval under which the trials we conducted does not allow us to publish the original data collected during the trials, despite the fact that the data is unidentifiable. In this report we describe a simple technique to create an unidentifiable and completely synthetic dataset with similar properties as the original dataset. The synthetic dataset is publicly available at http://caia.swin.edu.au/genius/files/CAIA-DATA-160422B.tgz. Our synthesisation technique is very simple and so not all types of analysis on the synthetic data will yield results similar to analysing the original data. However, a number of the analyses in [1] can be applied on the synthetic dataset and will produce similar results.

The primary motivation for releasing the synthetic dataset is to make it available for teaching in courses that introduce undergraduate students into research methods or data analysis methods. In this setting the limitations of the synthetic dataset are less relevant, as the teacher can define the analysis questions that students should answer. The synthetic dataset may also be useful for other research purposes.

If you use the dataset for teaching or research purposes you should cite this report as [2].

Section I describes the dataset format. Section III describes the technique we used to create a synthetic dataset. Section IV concludes the report.

II. Dataset Format

Note that in each trial/experiment we had eight players playing on four Xboxes (one acting as server, the other acting as clients). Further details of the trials are described in [1]. There are two datasets within http://caia.swin.edu.au/genius/files/CAIA-DATA-160422B.tgz, one for the experiments with network delay (`xbox_games_delay.txt`) and one for the experiments with packet loss (`xbox_games_loss.txt`). Each dataset is a comma-separated value (CSV) ASCII text file. The first line of each dataset file identifies the columns. Table I list the column names in the file and describes the columns.

III. Creating the Synthetic Data

The approach used for creating the synthetic data was very simple. For each combination of (delay/loss, server/client) we computed the mean and standard deviations for the quality, stay/leave, kills and deaths statistics

Table I
DATASET FORMAT

Column	Short name	Explanation
1	night	Evening/night the trial was carried out. You can also interpret this as experiment number. Four experiments were carried out.
2	player	Player number. Each experiment had 8 players.
3	game	Game number for each experiment and player. Each experiment with packet loss had 6 games, each experiment with network delay had 8 games.
4	delay/loss	Delay in milliseconds or loss rate in percent.
5	server/client	If set to 1, indicates that player played on Xbox acting as server, if set to 2 indicated player played on Xbox acting as client.
6	quality	Mean opinion score of the perceived game quality recorded by a player after the game ranging from 1 (very bad) to 5 (very good).
7	stay/leave	If set to 1, a player would like to continue playing in these network conditions, if set to 2 a player would like to leave the game (recorded by a player after the game).
8	kills	The number of times the player character killed another player character as indicated by the after-game statistics screen.
9	deaths	The number of times the player's character died as indicated by the after-game statistics screen.

from the original data. We then assumed that the original distributions are Gaussian (which is not entirely true) and created the synthetic data by drawing random numbers from the Gaussian distributions (with mean and standard deviation set to the empirical values) and rounding these. Our synthetic dataset was created for the same number of experiments and players as in the original experiments.

IV. CONCLUSIONS AND FUTURE WORK

We have used a simple technique to create a synthetic dataset based on an original dataset collected for trials with human players, which played the Xbox game Halo over a network with varying emulated network conditions. The dataset is available at http://caia.swin.edu.au/genius/files/CAIA-DATA-160422B.tgz and allows to reproduce a number of analyses in [1] with similar results. However, due to the simplicity of the synthesisation method not all analysis techniques can be expected to deliver useful results with the synthetic data. We have made the synthetic dataset available primarily for teaching purposes where this is not an issue, as in this setting in a teacher can control what analysis students should carry out.

REFERENCES

[1] S. Zander and G. Armitage, "Empirically Measuring the QoS Sensitivity of Interactive Online Game Players," in *Australian Telecommunications Networks & Applications Conference 2004 (ATNAC 2004)*, Sydney, Australia, 8-10 December 2004, pp. 511–517. [Online]. Available: http://caia.swin.edu.au/pubs/ATNAC04/zander-armitage-ATNAC2004.pdf

[2] S. Zander and G. Armitage, "Creating Synthetic Data for Network Game User QoS Sensitivity Trials," Centre for Advanced Internet Architectures, Swinburne University of Technology, Melbourne, Australia, Tech. Rep. 160422B, 22 April 2016. [Online]. Available: http://caia.swin.edu.au/reports/160422B/CAIA-TR-160422B.pdf