

Online Game QoE Evaluation using Paired Comparisons

Yu-Chun Chang¹, Kuan-Ta Chen², Chen-Chi Wu¹, Chien-Ju Ho³, Chin-Laung Lei¹

¹Department of Electrical Engineering, National Taiwan University

²Institute of Information Science, Academia Sinica

³Department of Computer Science and Information Engineering, National Taiwan University



Outline

- Introduction
- Probabilistic choice modeling
- Experiment methodology
- Experiment results
- Conclusion

Introduction

- *How to measure a game's quality systematically, efficiently, and reliably in a specific environment?*
- QoE (Quality of Experience)
 - User's subjective satisfaction at a service
- QoE on game
 - Focus on **real-timeliness** and **interactivity** during game play.

Introduction (cont.)

- MOS (Mean Opinion Score)



Excellent?
Good?
Fair?
Poor?
Bad?



Good !



Excellent?
Good?
Fair?
Poor?
Bad?



Bad !

Introduction (cont.)

- Disadvantages of MOS rating
 - The concepts of f... concretely defined.
 - Users may... interpretations of the scales.
 - Difficult to verify users' scores.
- Contributions
 - We use paired comparisons and probabilistic choice model to quantify online games' QoE.
 - We apply the proposed methodology to evaluate the network robustness of three popular FPS games.
 - Halo, Unreal Tournament 20004 (UT), Alien Arena (Alien)

Paired comparisons

Outline

- Introduction
- **Probabilistic choice modeling**
- Experiment methodology
- Experiment results
- Conclusion

Paired Comparison



A (Delay 20 ms)



B (Delay 800 ms)



Which one is better?



A

Paired comparison

- Paired comparison
 - **Generalizable** across a variety of multimedia applications
 - **Simple** comparative judgment
 - Obtain **ratio-scale** QoE scores
 - Users' inputs can be **verified**
- Comparison between paired comparison and MOS rating

	Generalizable	Judgment Difficulty	Ratio-Scale Scores	Input Verifiable
Paired comparison	Yes	Low	Yes	Yes
MOS ratings	Yes	High	No	No

Input of BTL Model: A Matrix of Choice

- 4 stimuli (T_1, T_2, T_3, T_4)
- 10 experiments, each has $\binom{4}{2} = 6$ paired comparisons

	T_1	T_2	T_3	T_4
T_1		8	9	10
T_2	2		7	8
T_3	1	3		6
T_4	0	2	4	

Vote T_1 (arrow pointing to the circled 8)

Vote T_2 (arrow pointing to the circled 2)

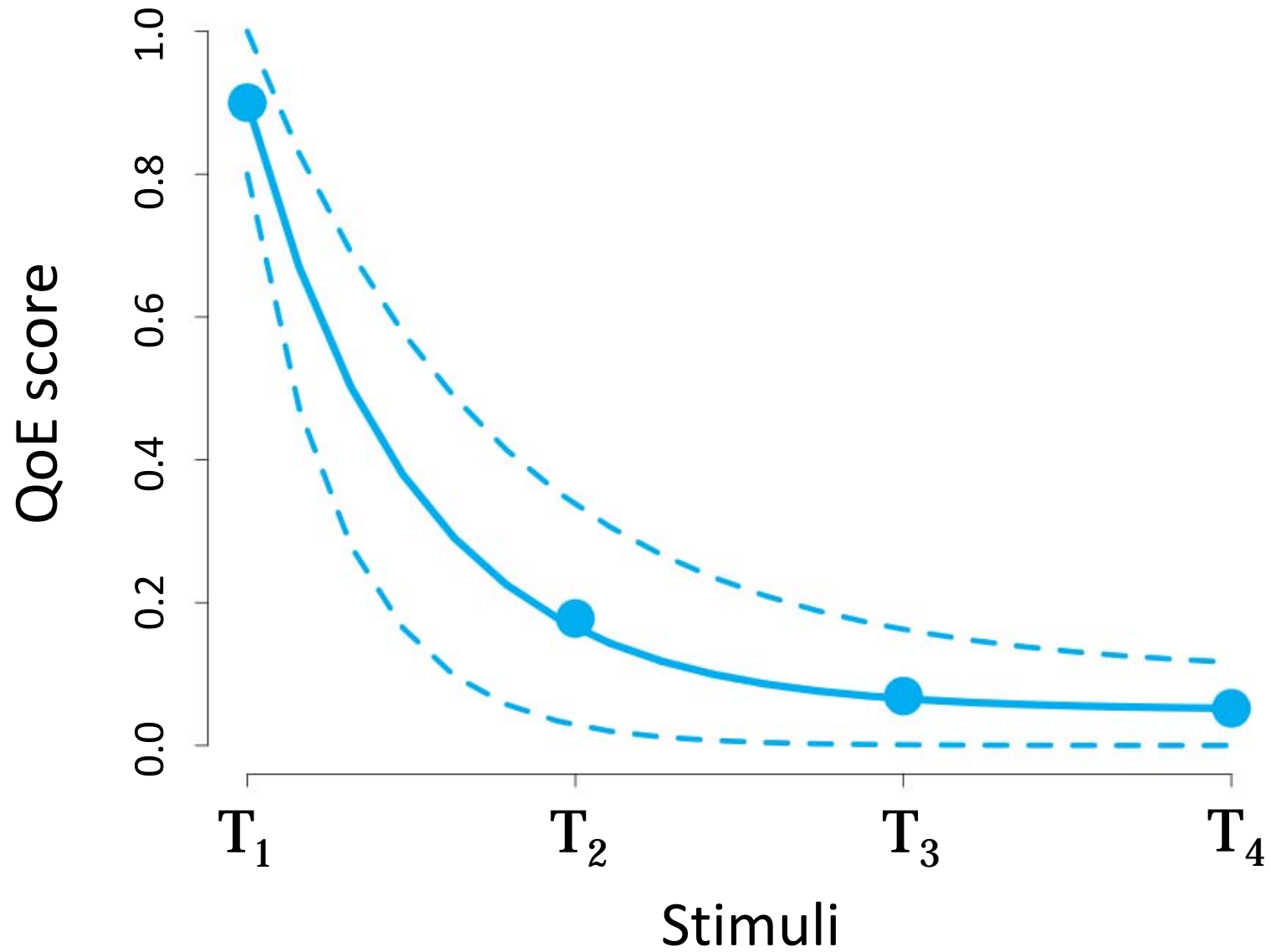
BTL Model

- Bradley-Terry-Luce (BTL) model

$$P_{ij} = \frac{\pi(T_i)}{\pi(T_i) + \pi(T_j)} = \frac{e^{u(T_i) - u(T_j)}}{1 + e^{u(T_i) - u(T_j)}}$$

- n stimuli (T_1, T_2, \dots, T_n)
 - $\pi(T_i)$ is the estimated score of the stimulus T_i
 - $u(T_i) = \log \pi(T_i)$
 - P_{ij} is the choice frequency of choosing T_i instead of T_j
- Input: a matrix of choice frequencies
 - Output: the estimated QoE score of stimulus T_i
(based on maximum likelihood estimation)

Output of BTL Model



Verification of User's Inputs

- Stochastic transitivity properties
 - WST, MST, SST
 - If $P_{ij} \geq 0.5$ and $P_{jk} \geq 0.5$, then

$$P_{ik} \geq \begin{cases} 0.5 & \text{(WST),} \\ \min\{P_{ij}, P_{jk}\} & \text{(MST),} \\ \max\{P_{ij}, P_{jk}\} & \text{(SST),} \end{cases}$$

- Kendall's u-coefficient

$$u = \frac{2 \sum_{i \neq j} \binom{a_{ij}}{2}}{\binom{m}{2} \binom{n}{2}} - 1.$$

Verification of User's Inputs

- Goodness-of-fit of the BTL model
 - Compare the likelihood L_0 of the fitted model and the likelihood L of the unrestricted model.
 - Test statistic
$$-2 \log(L_0/L)$$

Outline

- Introduction
- Probabilistic choice modeling
- **Experiment methodology**
- Experiment results
- Conclusion

Experiment Methodology

Halo: Combat Evolved



Unreal Tournament 2004



2010/6/8

Alien Arena 2008

15



Experiment Methodology

Game server



Change in 5 seconds



FreeBSD w/
dummynet



A

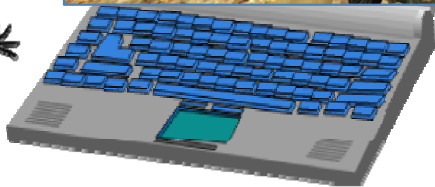
Experiment
Controller



Client B



Client A



Experiment Methodology

- 2 players with 6 bots in each game
- Deathmatch mode

Outline

- Introduction
- Probabilistic choice modeling
- Experiment methodology
- **Experiment results**
- Conclusion

Data Summary

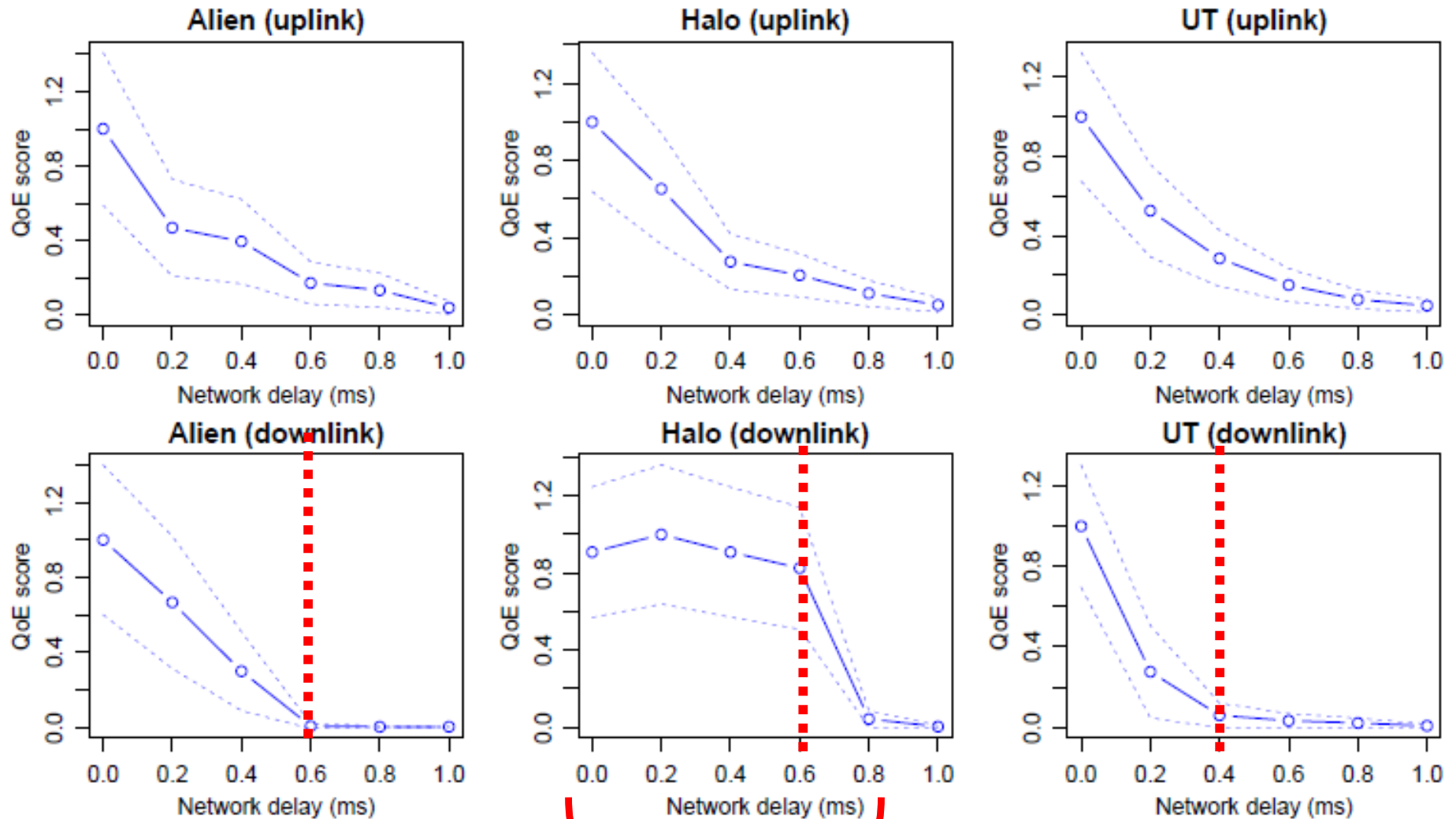
	Settings	# Comparisons	
Delay	0 ms, 200 ms, 400 ms, 600 ms, 800 ms, 1000 ms	Uplink	600
		Downlink	690
Loss rate	0%, 10%, 20%, 30%	Uplink	288
		Downlink	252
Delay jitter	0 ms, 250 ms, 500 ms	Uplink	72
		downlink	78

- **Uplink** : client → server
- **Downlink** : server → client

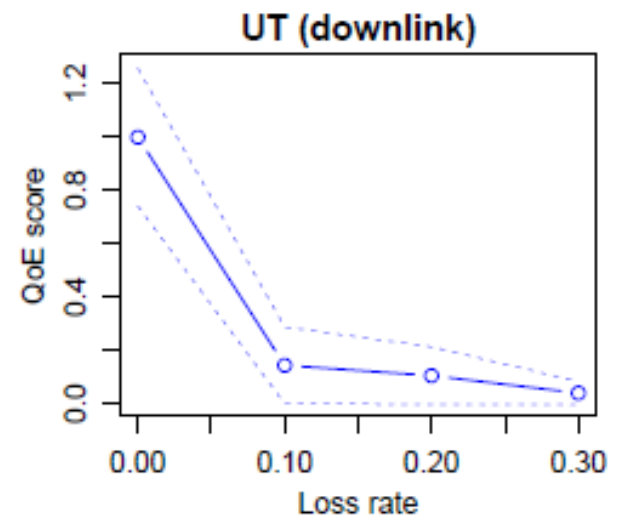
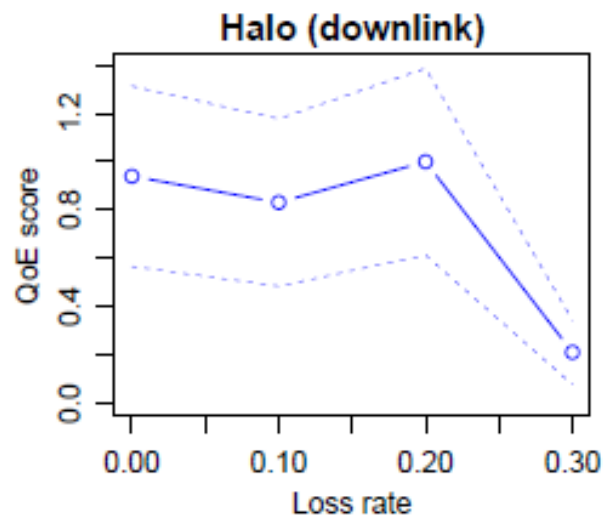
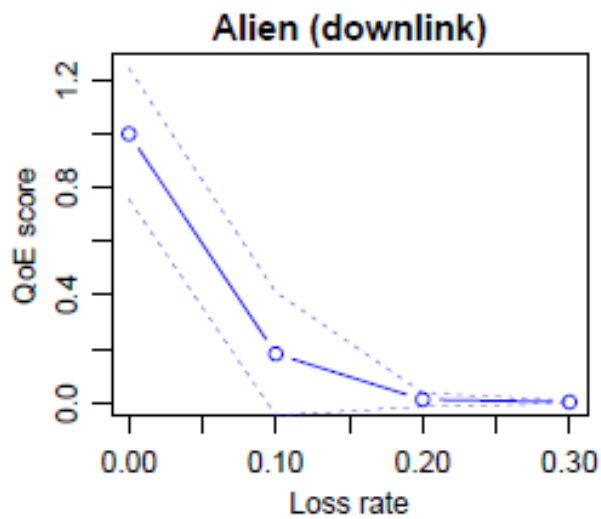
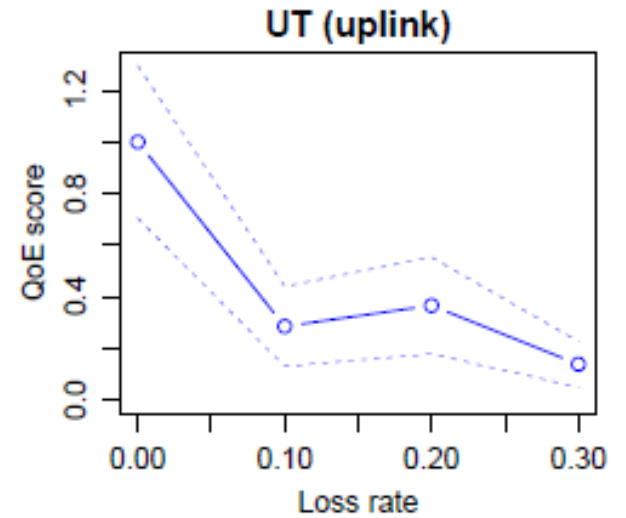
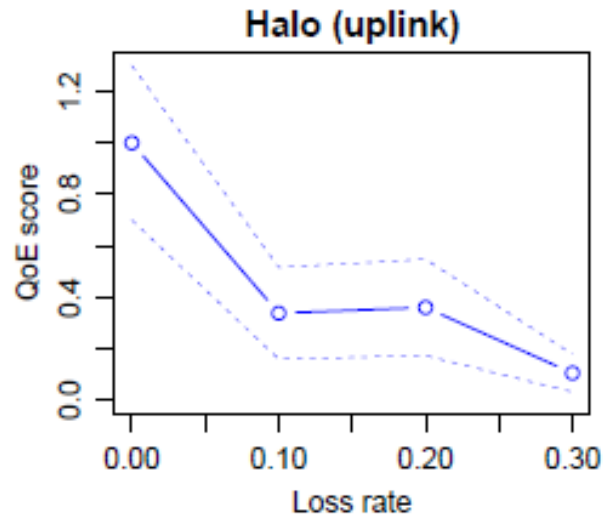
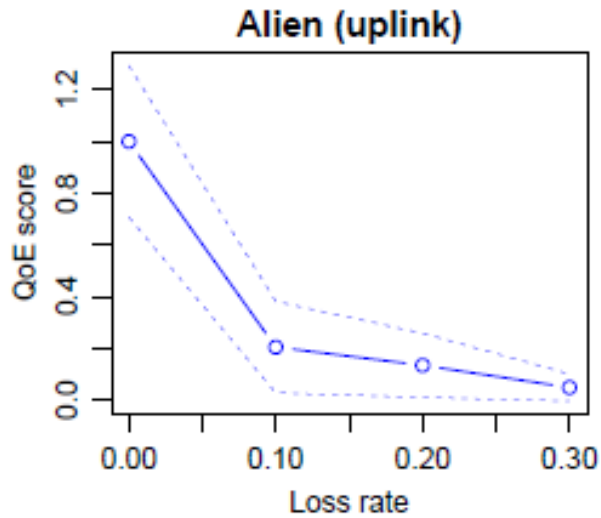
Verifications

Game	Factor	Link	WST	MST	SST	Kendall	p-BTL
Alien	delay	uplink	0	2	7	0.35	0.33
		downlink	0	0	3	0.62	0.80
	loss	uplink	0	0	1	0.40	0.12
		downlink	0	0	0	0.74	0.86
	jitter	uplink	0	1	1	-0.04	0.26
		downlink	0	0	0	1.00	0.02
Halo	delay	uplink	0	0	6	0.36	0.35
		downlink	2	2	5	0.53	0.15
	loss	uplink	0	0	0	0.25	0.98
		downlink	0	1	3	0.22	0.03
	jitter	uplink	0	0	0	0.55	0.44
		downlink	0	1	1	0.32	0.05
UT	delay	uplink	0	1	10	0.42	0.01
		downlink	0	2	5	0.56	0.11
	loss	uplink	0	1	1	0.28	0.01
		downlink	0	0	2	0.43	0.17
	jitter	uplink	0	0	0	1.00	0.02
		downlink	0	0	0	-0.08	1.00

Network Delay

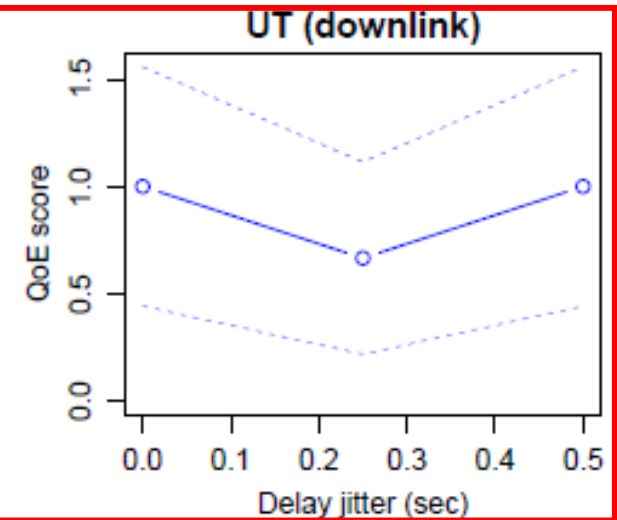
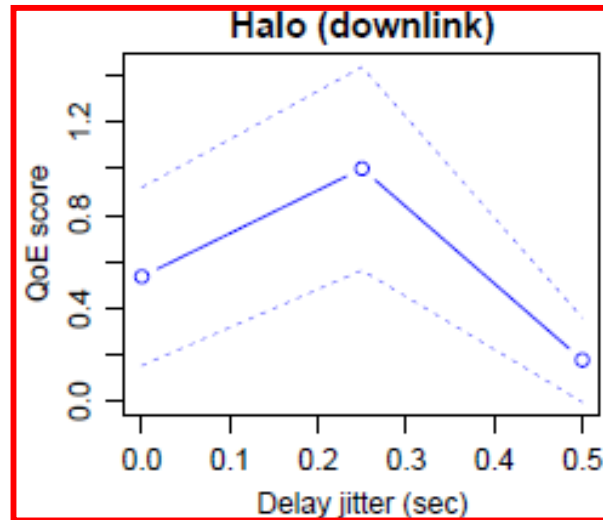
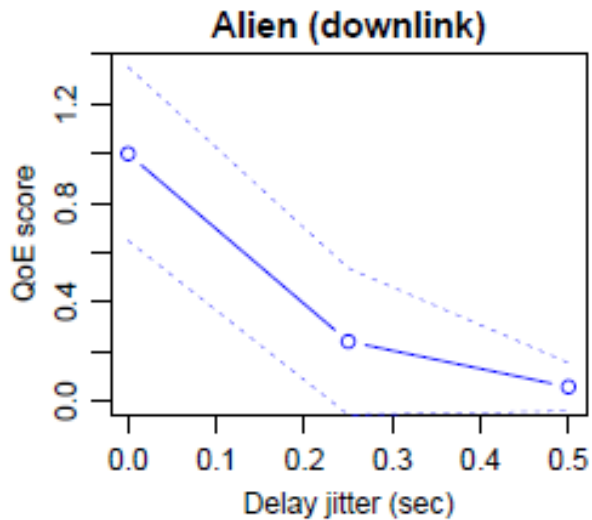
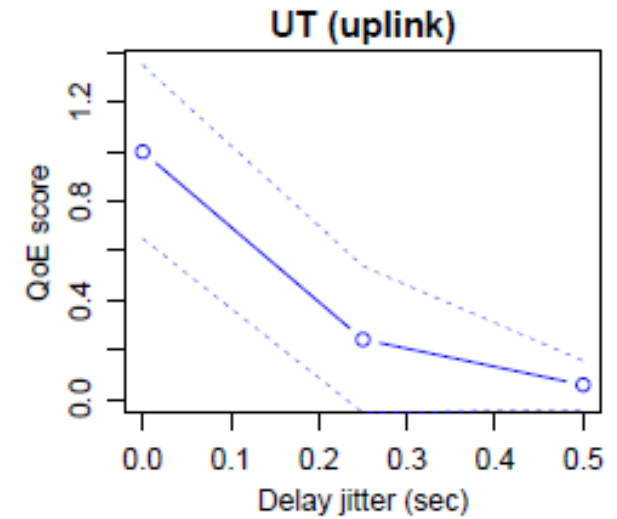
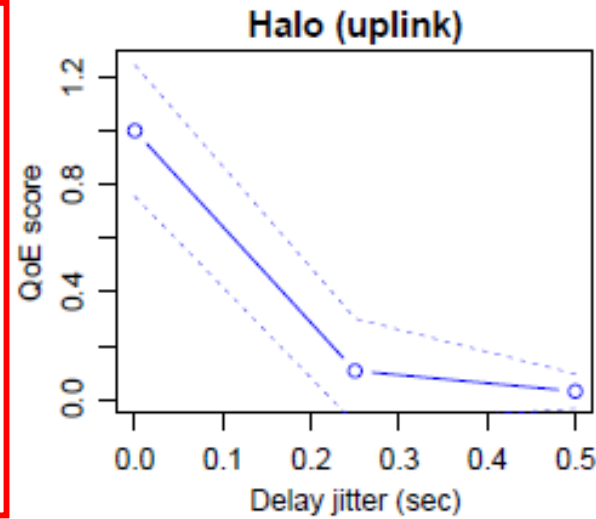
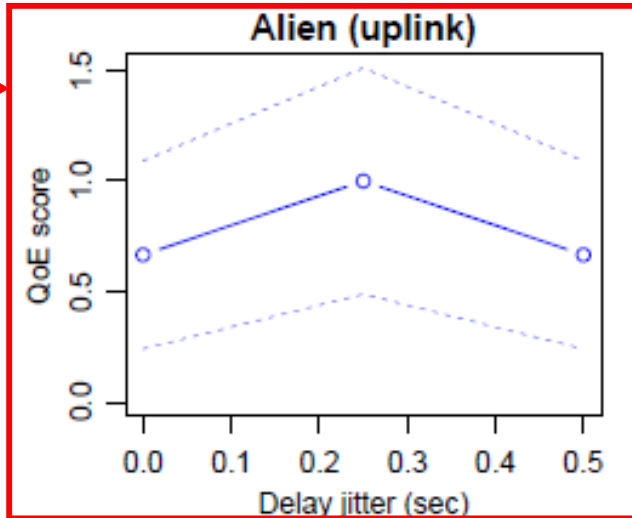


Loss Rate



Delay Jitter

Time synchronization in the **client**



Time synchronization in the **server**

Conclusion

- We use paired comparisons and probabilistic choice model to evaluate game's QoE.
- We apply the proposed methodology to evaluate the network robustness of three popular FPS games.
- Future work
 - Conduct more experiments.
 - Evaluate games' QoE among different game genres







Quadrant of Euphoria

Researchers

 Image	 Audio	 Video
Register	Register	Register
Login	Login	Login

Note to researchers: If you want to try Quadrant of Euphoria out, we provide demo profiles for all three experiment types above. Please login any one of them with the name **demo** and the password **qoedemo**.

Experiment Participants

Type	Experiment	Description	Reward	Link
	jpg2000	JPEG 2000 Quality Study.	\$1.0	go
	new_jpg	We want to test our new compression method.	N/A	go
	compression	Audio VBR compression level.	\$1.5	go
	mp3_lossless	Verify the loss-less MP3 codec	\$1.5	go

<http://mmnet.iis.sinica.edu.tw/proj/qoe/>

Documents

- Manuals
 - For researchers
 - For experiment participants
- Publications
 - A Crowdsorceable QoE Evaluation Framework for Multimedia Content

Thank you for your attention!