

Data driven competitive matchmaking across
multiple continents

W3[👑]CHAMPIONS

Talk structure

- Intro to Warcraft III, W3Champions, first project
- Matchmaking details for 2on2
- 1on1 ladder over multiple continents

Warcraft III

- Real time strategy game, release in July 2002
 - popular modes: 1on1 / 2on2 / 4on4 / FFA
- Warcraft III: Reforged released in January 2020
- Revival of the competitive (1on1) scene

W3Champions: community ladder

Why? No competitive matchmaking on release :(

When? Started March 2020, thanks @W3Pad :)

How? Chromium Embedded Framework

WARCRAFT
REFORGED

CAMPAIGN

VERSUS

CUSTOM GAMES

LOCAL AREA NETWORK

SINGLE PLAYER

COLLECTION

©2021 BLIZZARD ENTERTAINMENT, INC. ALL RIGHTS RESERVED.

W3CHAMPIONS

W3CHAMPIONS

Message of the Day

Legion TD 4v4 x3, x20 and ROC 1v1 ladders have been released. You can find them in Non-Melee section of versus screen.

[1. W3 GENERAL]:



W3CHAMPIONS

Sunday, June 6, 2021
14:18

MAP PREFERENCES

- AMAZONIA MAX PLAYERS: 2
- AUTUMN LEAVES MAX PLAYERS: 2
- CONCEALED HILL MAX PLAYERS: 2
- ECHO ISLES MAX PLAYERS: 2
- LAST REFUGE MAX PLAYERS: 2
- NORTHERN ISLES MAX PLAYERS: 2
- RUINS OF AZSHARA MAX PLAYERS: 2
- SHALLOW GRAVE V1.3 MAX PLAYERS: 2



SELECT 1 MAPS TO EXCLUDE FROM YOUR MAP POOL.

AUTOMATED TOURNAMENTS

Coming soon

SELECT RACE AND GAME MODE

Melee

Non Melee

1vs1

2vs2

4vs4

FFA



Meet same opponent?

Server

UNRESTRICTED

FLO & BNET



1VS1 MASTER

Rank 52

2026 MMR

Last Opponent

70W 71L 49.6%

7340 RP

AngryAcolyte#1835

BACK

CUSTOM

CHAT



FIND MATCH

1VS1 / NIGHT ELF

732

31



GDC

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21

What to do with the data?

- Rating system
- Ranking formula
- Matchmaking algorithm

What got me started...

- Very early adopter as a (good) player
- Unhappy with the ranking systems:
 - first: pure “XP” ranking, play most get first
 - then: pure skill, simply ordered by rating
- Gave feedback as a player...

Improving the ranking formula

Idea: balance skill (rating) and grind (activity)

- **Fair** ladder: should converge to a pure skill ranking
- **Honest** ladder: incentivize activity to improve confidence in ratings

The Ranking Formula

$$\left(\underbrace{\beta + (1 - \beta)}_{\text{skill vs activity}} \underbrace{(1 - \exp(-\alpha \#Wins))}_{\text{goes to 1 with \#Wins growing}} \right) \times \underbrace{(\mu - 2(\sigma - \sigma_{\min}))}_{\text{rating - confidence term}}$$

- β (e.g. 0.8): determines how much we weight “skill” vs “activity”
- α (e.g. 0.12): determines how “active” players need to be
- Lower bound on rating: $\mu - 2(\sigma - \sigma_{\min})$

Matchmaking

Matchmaking revisited

What's challenging about 2on2 Random Team matchmaking?

- Closely matched players does not imply fair game

Example: matched players with ratings: [1904, 1903, 1902, 1701]

- How should we model team strength?

Example: [2000, 1200] equivalent to [1600, 1600] ?

- Ratings are long tailed for very good players (lonely at the top)

Matchmaking revisited

Idea: Balance **fairness**, **uniformity** and **wait time**.

What is **fairness**? How fair a game is, given a model for win probability:

$$F(\text{Game}) = |\mathbb{P}(\text{win}) - 0.5|$$

What is **uniformity**? How closely matched players are, w.r.t. a notion of quality:

$$U(\text{Game}) = \max_{p \in G} (Q_p) - \min_{p \in G} (Q_p)$$

Matchmaking revisited

Fairness in team games: how strong is a team given its two players' ratings R_1, R_2 ?

Generalized p-mean:

$$R_T(p) = [0.5 \cdot (R_1^p + R_2^p)]^{1/p}$$

How to find a good "p"? Maximum Likelihood estimation on actual game history!

→ best fit for $p = -0.02$, we use the geometric mean for 2on2!

Win probability becomes:

$$\mathbb{P}(\text{win for } T_1) = g(\sqrt{R_1 R_4} - \sqrt{R_2 R_3}),$$

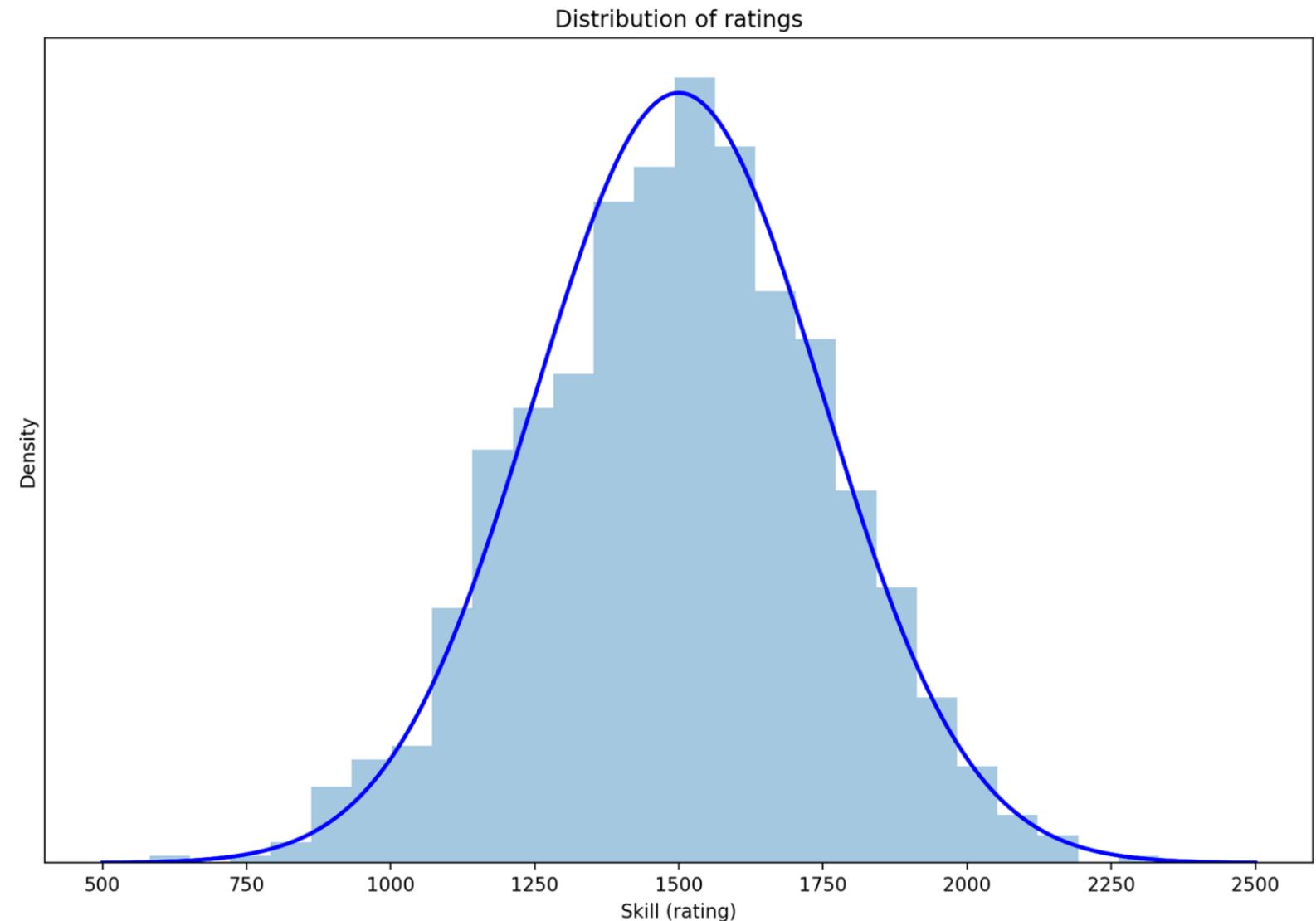
where $g(\cdot)$ is the (properly scaled) logistic function (see e.g. Elo or Glicko2).

Matchmaking revisited

Uniformity in team games: how far apart are players, really?

$$Q_p = \text{Rating}_p$$

$$U(\text{Game}) = \max_{p \in G} (Q_p) - \min_{p \in G} (Q_p)$$



Matchmaking revisited

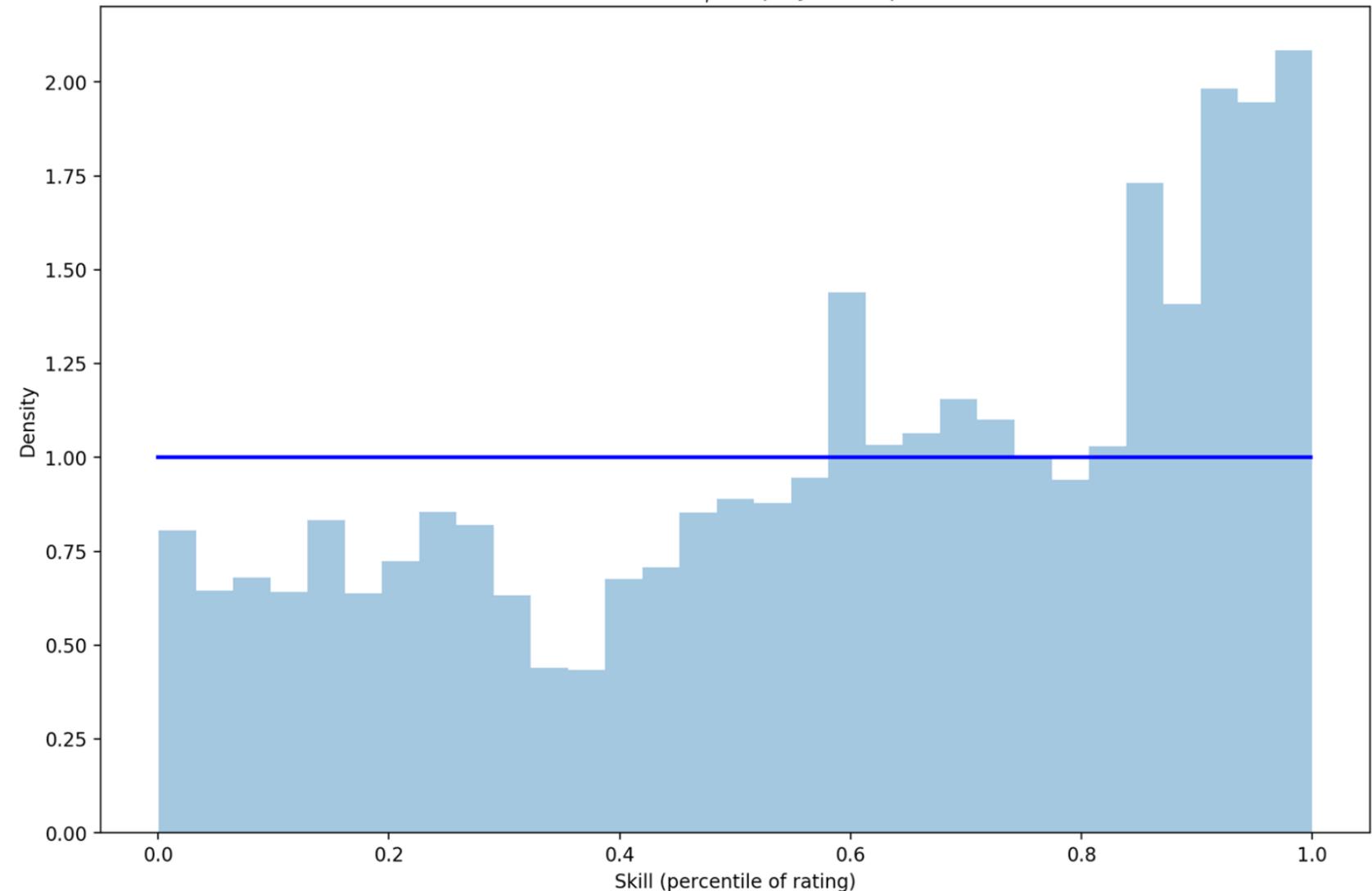
Uniformity in team games: what about the percentile of skill?

$$Q_p = \text{ECDF}(\text{Rating}_p)$$

ECDF (empirical cumulative distribution function): proportion of players with a rating smaller than player p's

$$U(\text{Game}) = \max_{p \in G} (Q_p) - \min_{p \in G} (Q_p)$$

Distribution of Q_p for players in queue



Matchmaking revisited

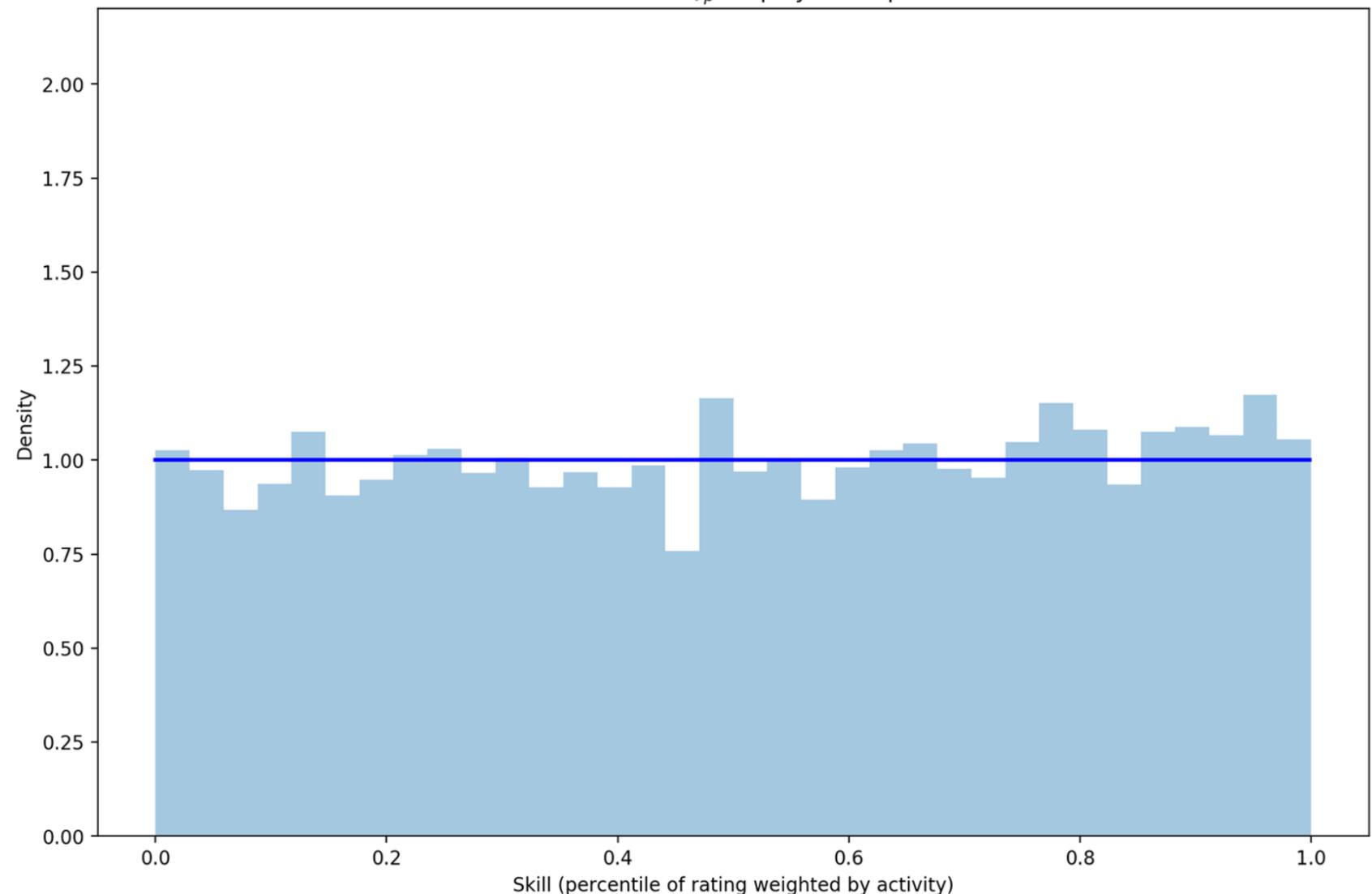
Uniformity in team games: activity weighted percentile of skill!

$$Q_p = \text{WECDF}(\text{Rating}_p)$$

WECDF (**weighted** empirical cumulative distribution function): proportion of players with a rating smaller than player p 's, **weighted by how much they play**

$$U(\text{Game}) = \max_{p \in G} (Q_p) - \min_{p \in G} (Q_p)$$

Distribution of Q_p for players in queue



Matchmaking revisited

Loop over all possible games and create a game if and only if:

$$\text{Uniformity}(G) + c \cdot \text{Fairness}(G) < \tau - m \cdot \text{AVG}(\text{Wait time})$$

Find parameters τ , c and m by running experiments on historical data

c	m	tau	avg fairness	avg q uniformity	avg wait	median fairness	median q uniformity	median wait	q25 fairness	q75 q unif	q75 wait
CURRENT	CURRENT	CURRENT	0.089	0.397	3.253	0.074	0.375	2.897	0.853	0.559	3.924
2.000	0.120	0.120	0.062	0.239	2.661	0.044	0.202	2.108	0.853	0.309	3.446
2.000	0.120	0.250	0.067	0.270	2.292	0.049	0.239	1.746	0.853	0.350	3.005
2.000	0.120	0.300	0.069	0.280	2.153	0.052	0.250	1.615	0.853	0.365	2.833
1.500	0.120	0.200	0.069	0.262	2.340	0.049	0.229	1.827	0.853	0.340	3.075
1.500	0.120	0.250	0.072	0.273	2.194	0.053	0.244	1.701	0.853	0.357	2.892
1.500	0.120	0.300	0.072	0.285	2.089	0.054	0.259	1.587	0.853	0.370	2.749
3.000	0.100	0.120	0.055	0.222	2.983	0.037	0.189	2.361	0.853	0.286	3.812
2.500	0.100	0.050	0.054	0.208	3.201	0.037	0.172	2.582	0.853	0.269	4.037
3.000	0.100	0.050	0.052	0.206	3.224	0.036	0.171	2.597	0.853	0.268	4.065
2.500	0.100	0.100	0.055	0.219	2.994	0.039	0.186	2.398	0.853	0.284	3.856
2.500	0.100	0.080	0.055	0.214	3.078	0.037	0.179	2.476	0.853	0.278	3.887
3.000	0.100	0.080	0.054	0.211	3.137	0.037	0.177	2.503	0.853	0.272	3.922
3.000	0.100	0.150	0.056	0.227	2.902	0.038	0.197	2.246	0.853	0.292	3.732
3.000	0.100	0.100	0.054	0.217	3.064	0.038	0.184	2.446	0.853	0.282	3.861
2.500	0.100	0.120	0.057	0.223	2.916	0.038	0.192	2.309	0.853	0.289	3.751
2.500	0.100	0.150	0.058	0.228	2.808	0.040	0.197	2.185	0.853	0.293	3.618

Competitive ladder over multiple continents

Problem: our 1on1 player base is spread over all continents, but only Europe and maybe NA can sustain themselves.

Official servers only in Chicago, Amsterdam and Seoul :(

We (@fluxxu) created Flo, which allows us to host games on our own servers anywhere we want.



GDC[®]

GAME DEVELOPERS CONFERENCE | July 19-23, 2021 | #GDC21



Fair global ladder for 1on1

Deterministic server selection:

Pick from list of single servers according to utility function of pings for both players:

$$\text{Utility}_s = 0.9 \cdot \max(p_1, p_2) + 0.1 \cdot \min(p_1, p_2)$$

May lead to systematic disadvantage for one of two given players, for example US East vs Europe is consistently hosted on US East (NYC).

Fair global ladder for 1on1

Fair server pairs:

Each server “S” is associated with a bias for player 1 (positive if $p_1 > p_2$):

$$\beta_S = \pm \sqrt{\max(0, |p_1 - p_2| - 20)}$$

For two servers A and B, if β_A and β_B have opposite signs, then that pair of servers is fair if we pick server A with probability:

$$\delta = \frac{-\beta_B}{\beta_A - \beta_B}$$

Loop over fair pairs, pick with same utility function:

$$\text{Utility}_{A+B} = \delta \cdot \text{Utility}_A + (1 - \delta) \cdot \text{Utility}_B$$

Conclusion

- Over 3 million games, all the top chinese and korean players on our ladder
- Replicated our melee ladder for community custom maps, support multiple kinds of gamemodes (e.g. 3v3v3v3, battle royale FFA)
- Abstracted out all logic from the game, ready to be applied to your game :)

Go to MatchBox.gg for more information!