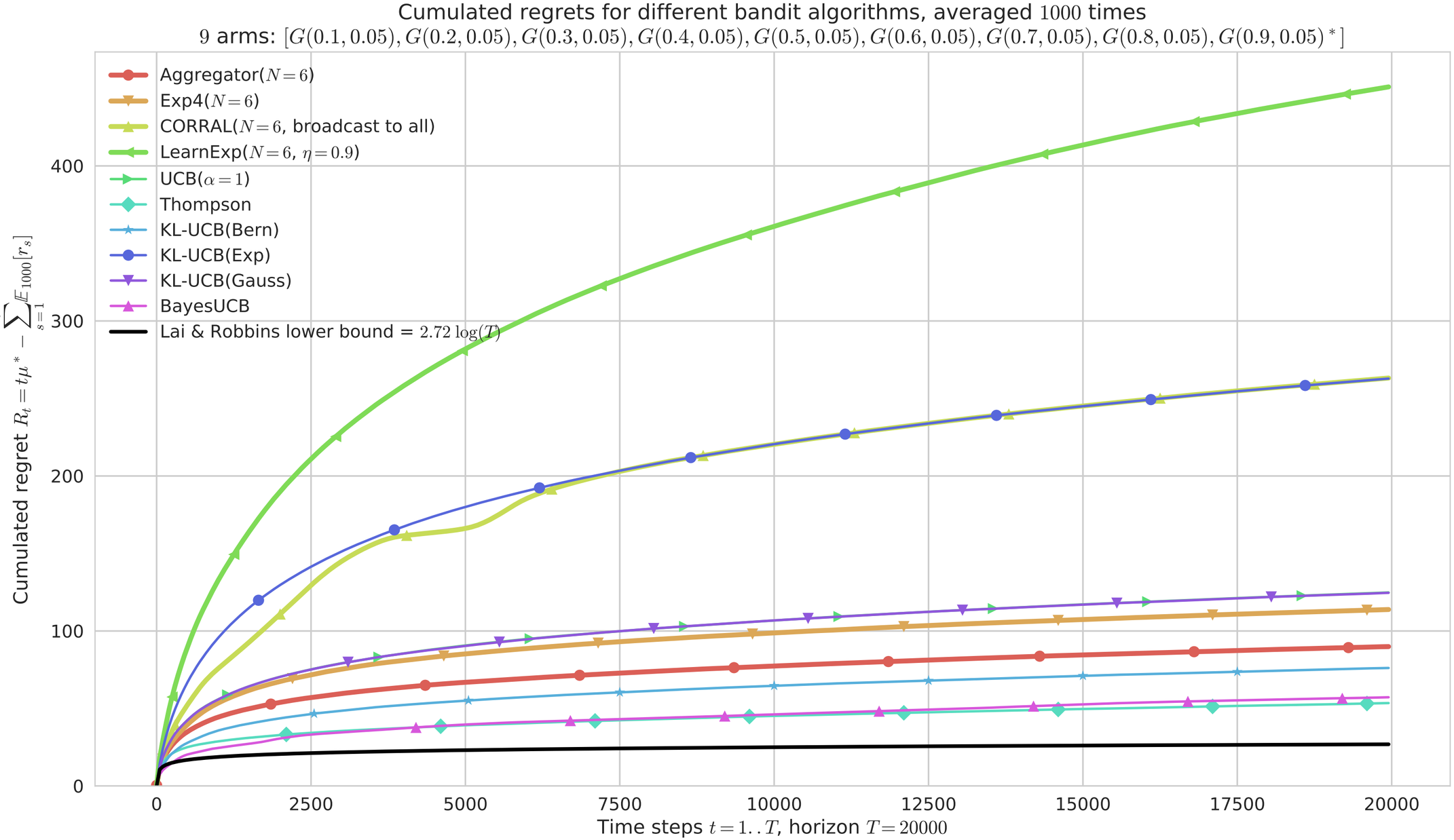
Cumulated regrets for different bandit algorithms, averaged 1000 times
9 arms: [G(0.1, 0.05), G(0.2, 0.05), G(0.3, 0.05), G(0.4, 0.05), G(0.5, 0.05), G(0.6, 0.05), G(0.7, 0.05), G(0.8, 0.05), G(0.9, 0.05)*]
Aggregator(N=6)
Exp4(N=6)
CORRAL(N=6, broadcast to all)
LearnExp(N=6, η=0.9)
UCB(α=1)
Thompson
KL-UCB(Bern)
KL-UCB(Exp)
KL-UCB(Gauss)
BayesUCB
Lai & Robbins lower bound = 2.72 log(T)
Cumulated regret $R_t = t\mu^* - \sum_{s=1}^{t} \mathbb{E}_{1000}[r_s]$
0
100
200
300
400
0
2500
5000
7500
10000
12500
15000
17500
20000
Time steps t = 1..T, horizon T = 20000