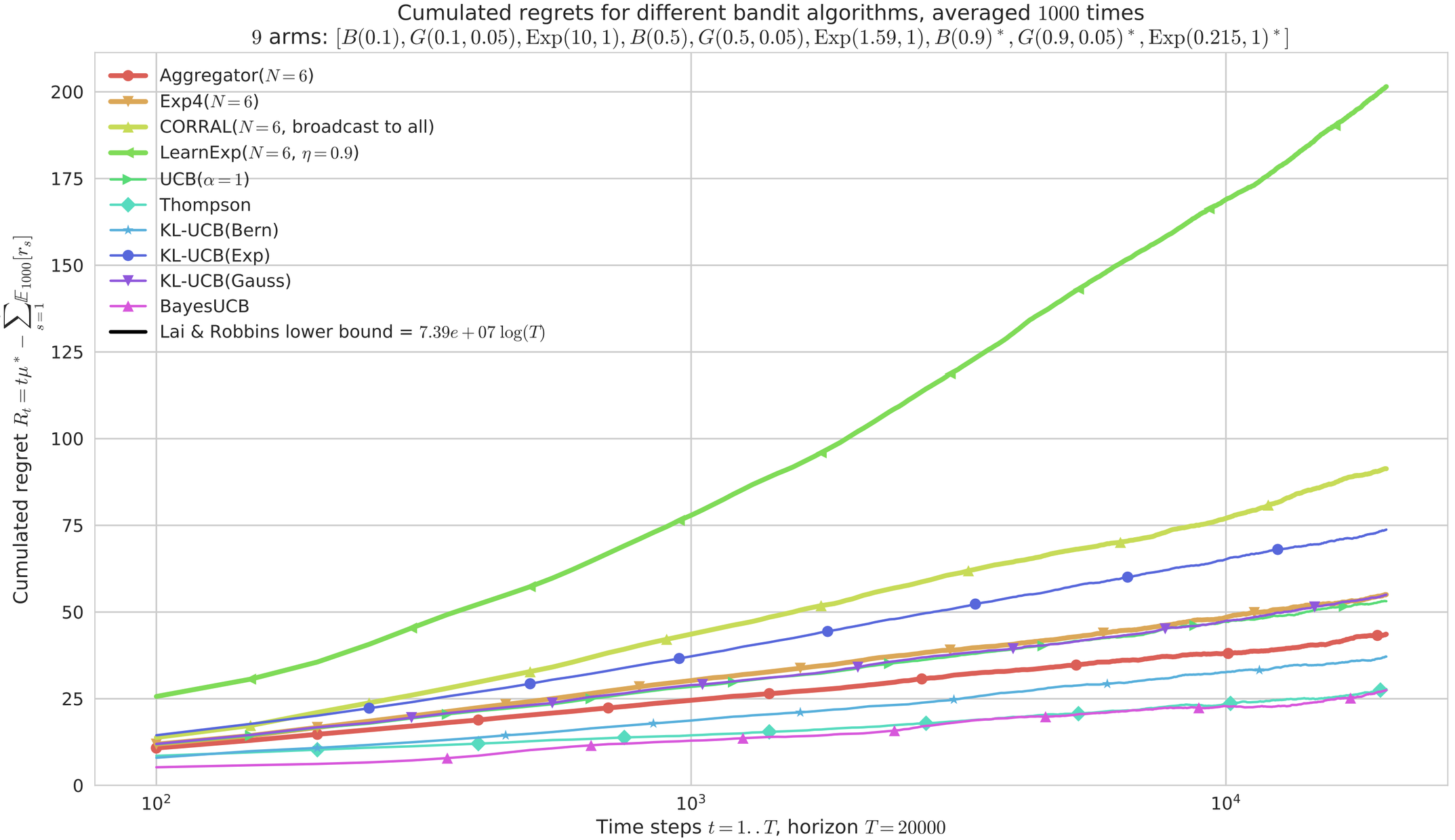
Cumulated regrets for different bandit algorithms, averaged 1000 times
9 arms: [B(0.1), G(0.1, 0.05), Exp(10, 1), B(0.5), G(0.5, 0.05), Exp(1.59, 1), B(0.9)*, G(0.9, 0.05)*, Exp(0.215, 1)*]
Aggregator(N=6)
Exp4(N=6)
CORRAL(N=6, broadcast to all)
LearnExp(N=6, η=0.9)
UCB(α=1)
Thompson
KL-UCB(Bern)
KL-UCB(Exp)
KL-UCB(Gauss)
BayesUCB
Lai & Robbins lower bound = 7.39e+07 log(T)
Cumulated regret R_t = tμ* − Σ_{s=1}^{t} E_1000[r_s]
0
25
50
75
100
125
150
175
200
10^2
10^3
10^4
Time steps t = 1..T, horizon T = 20000