I. Introduction

THIS IS THE SUPPLEMENTARY MATERIAL FOR OUR UNDER REVIEW PAPER

II. Evaluating alternative reward structures for MAHT

We study the proposed reward structure of our work with the two alternative reward structures for confidence maximisation or entropy minimisation. We will train the PPOGC algorithm with all three reward structures on MAHT problems with binomial observations. We will limit our demonstration to the unconstrained case. Since all algorithms achieve an empirical error probability lower than ϵ, we will only present the different stopping times in table I.

As you can see when the PPOGC algorithm is rewarded for error minimisation it achieves a significantly shorter stopping time. It performs significantly worse for confidence maximisation We are not sure why that happens but we speculate it is due to the large values the confidence may take making the already computationally difficult multiagent training procedure even more unstable.

setting	error	entropy	confidence
$\epsilon = 0.1$ first scenario	**9.247**	10.49	11.25
$\epsilon = 0.05$ first scenario	**10.799**	11.35	11.48
$\epsilon = 0.1$ second scenario	**13.743**	17.63	18.69
$\epsilon = 0.05$ second scenario	**14.14**	19.38	25.97
$\epsilon = 0.1$ third scenario	**13.86**	17.34	23.33
$\epsilon = 0.05$ third scenario	**14.928**	17.78	28.13

TABLE I: The average stopping time of PPOGC for different reward structures

III. Additional Figures for the unconstrained MAHT problem

Fig. 1: Accuracy and average stopping time for the first decentralised scenario: synthetic Gaussian data

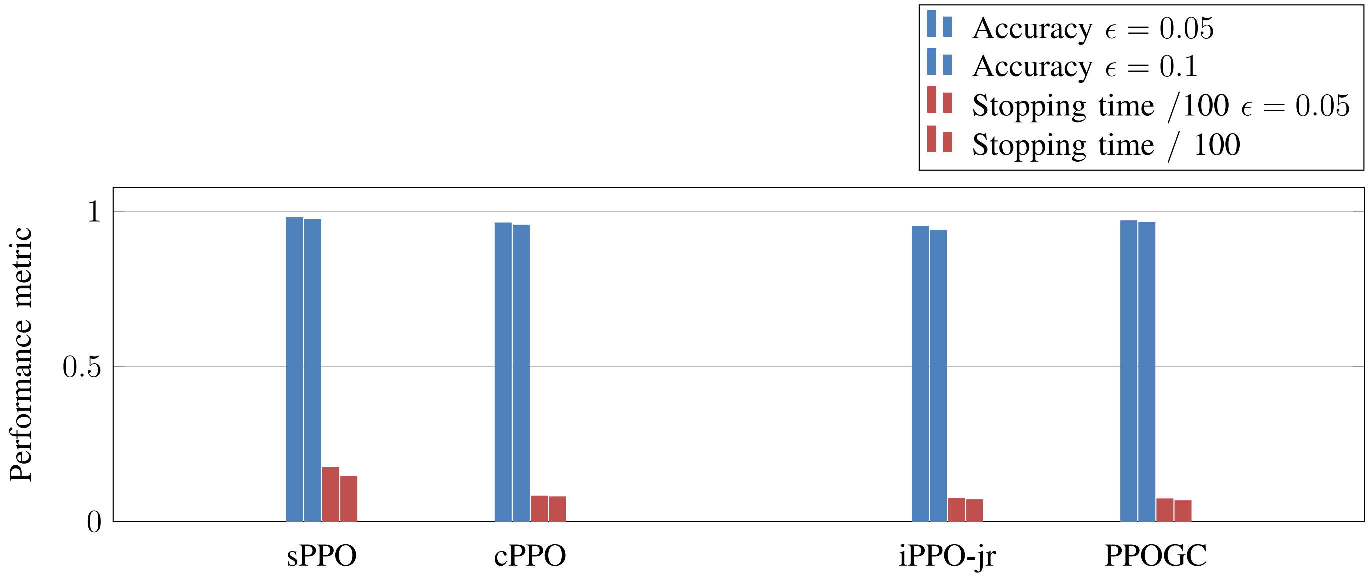


Fig. 2: Accuracy and average stopping time for the second decentralised scenario: synthetic Gaussian data

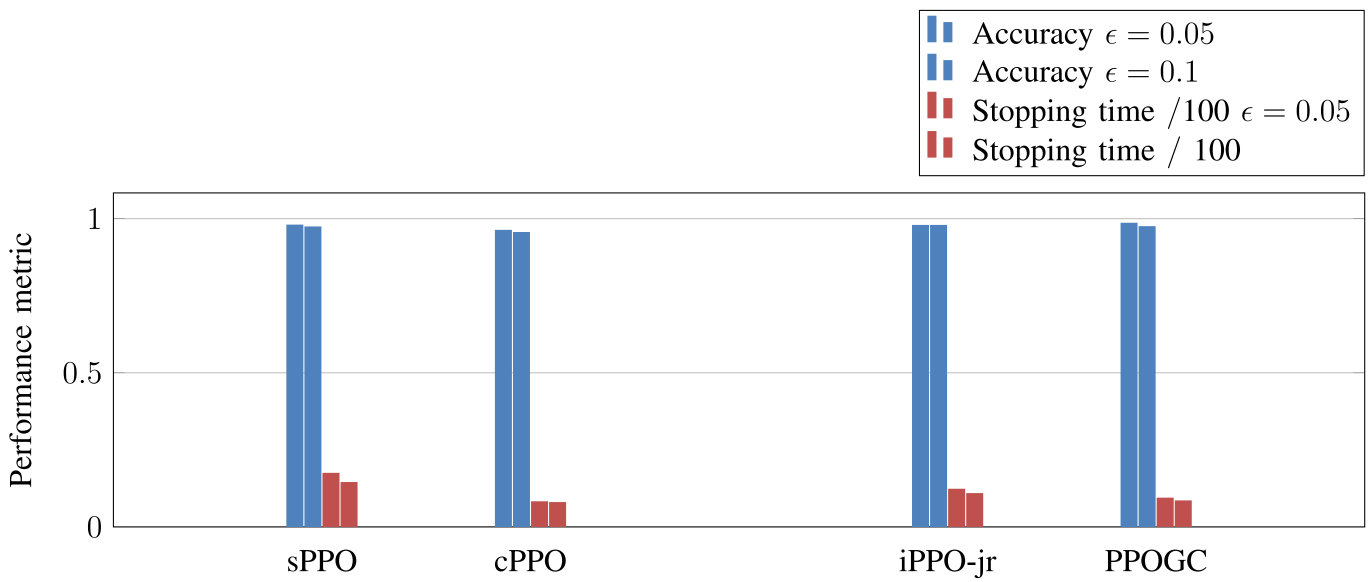


Fig. 3: Accuracy and average stopping time for the third decentralised scenario: synthetic Gaussian data

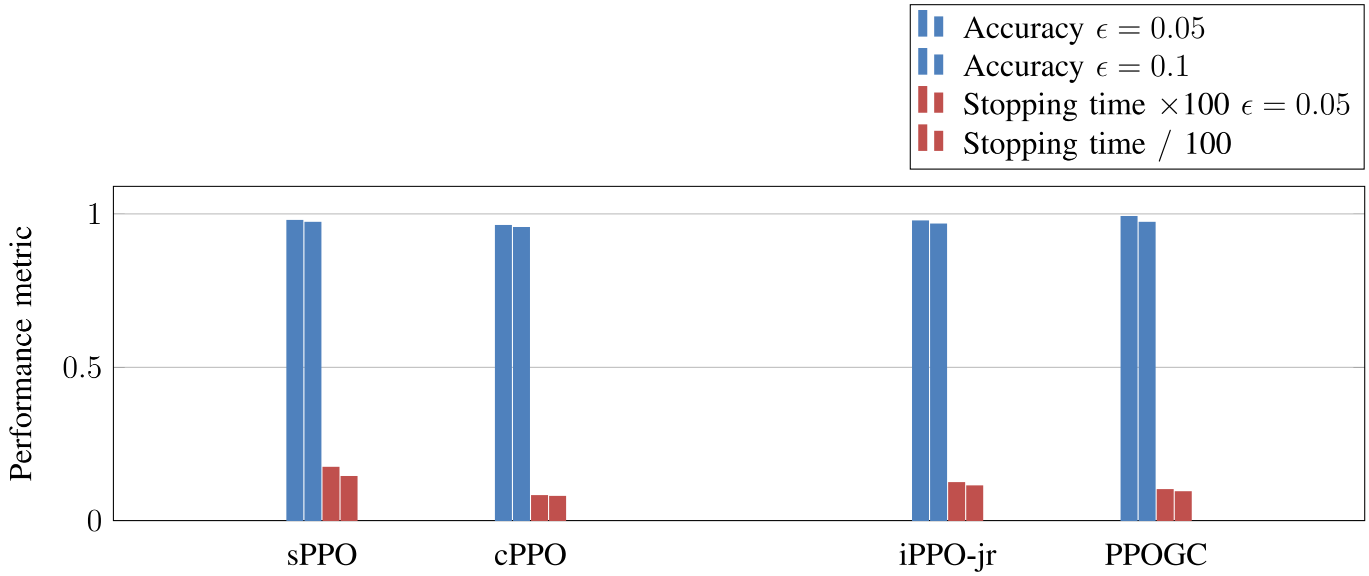


Fig. 4: Accuracy and average stopping time for the first decentralised scenario: TON_IOT data

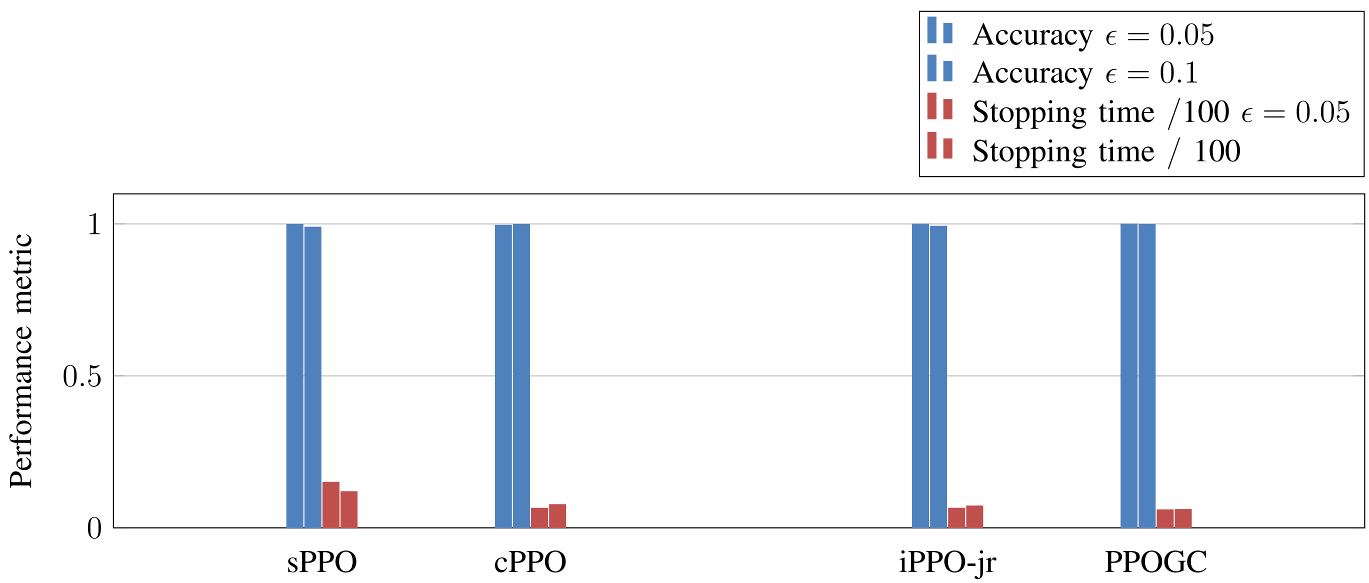


Fig. 5: Accuracy and average stopping time for the second decentralised scenario: TON_IOT data

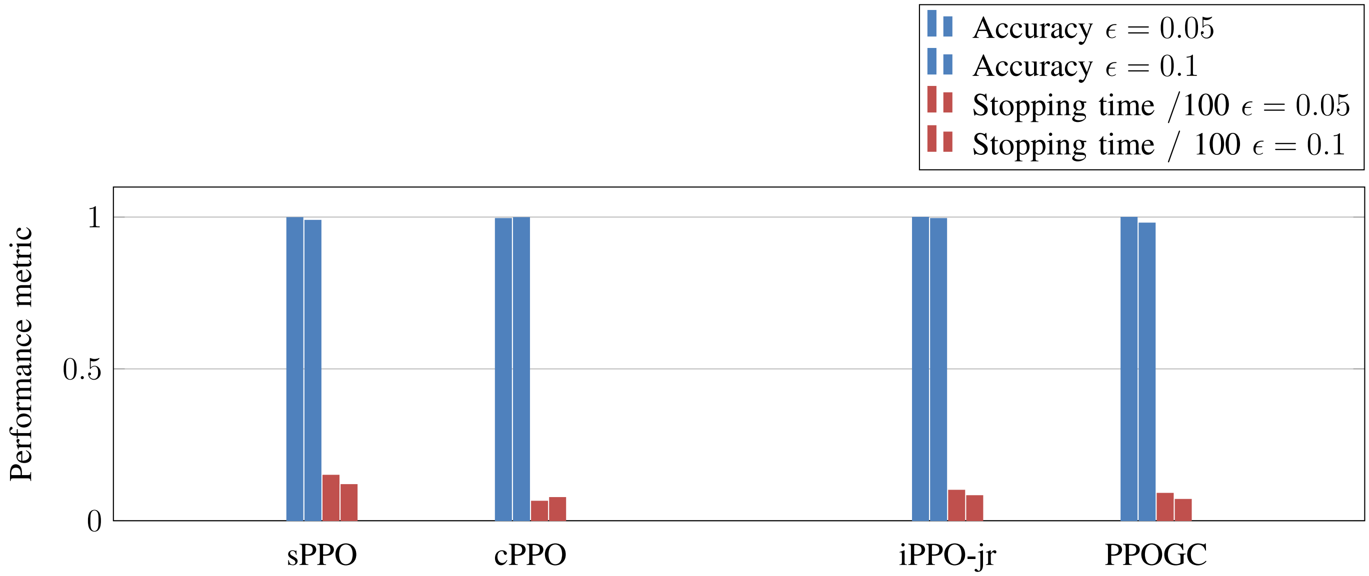
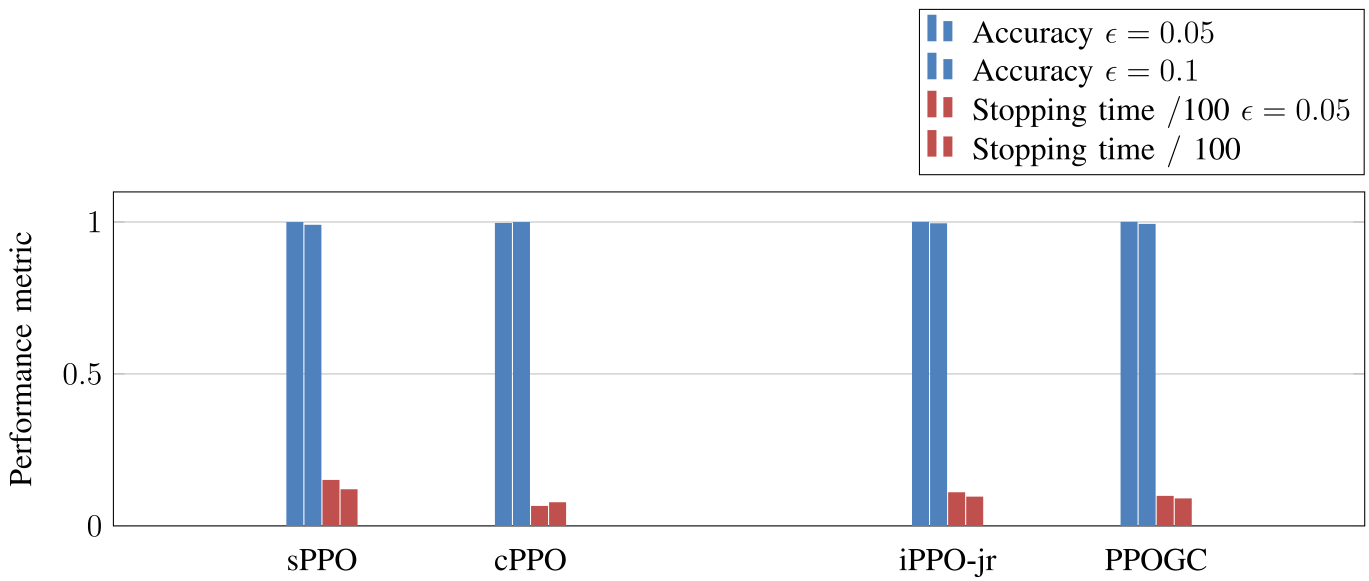
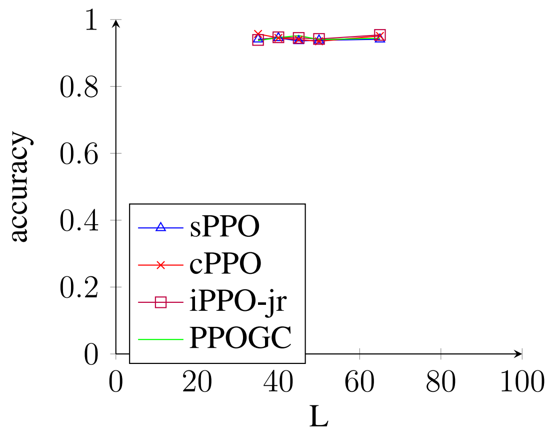


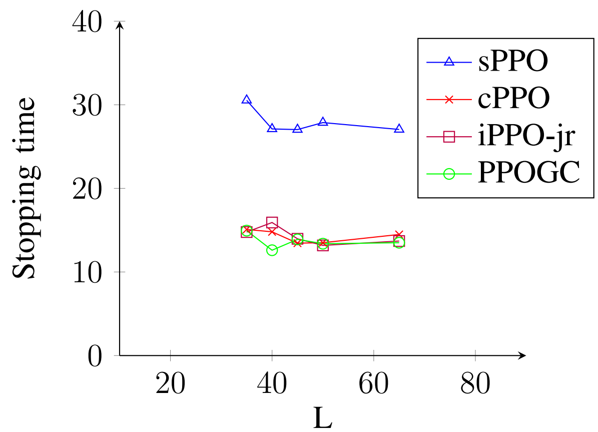
Fig. 6: Accuracy and average stopping time for the third decentralised scenario: TON_IOT data



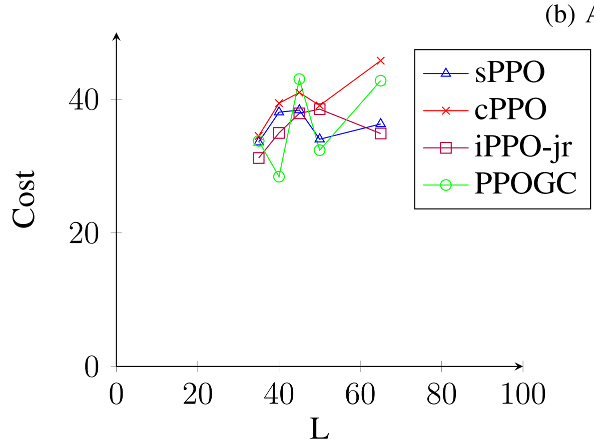
IV. Figures for the constrained MAHT problem



(a) Accuracy

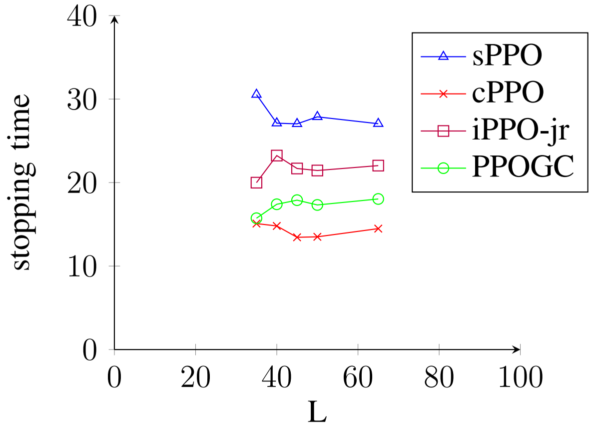


(b) Average stopping time

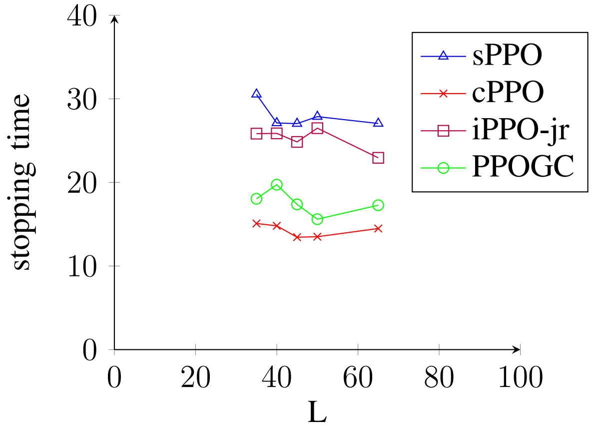


(c) Average cost

Fig. 7: Accuracy average stopping time and average cost for different values of L: first decentralised MAHT problem



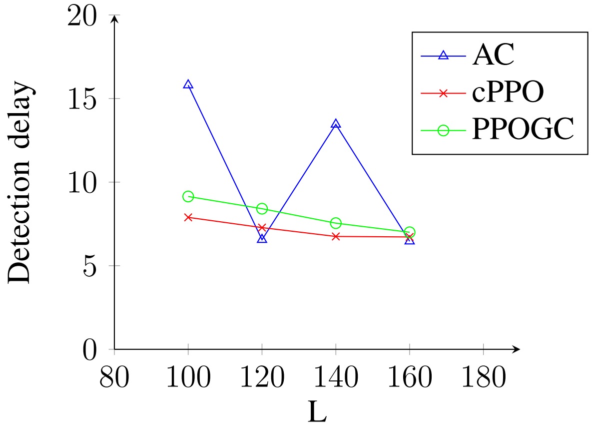
(a) Second decentralised problem



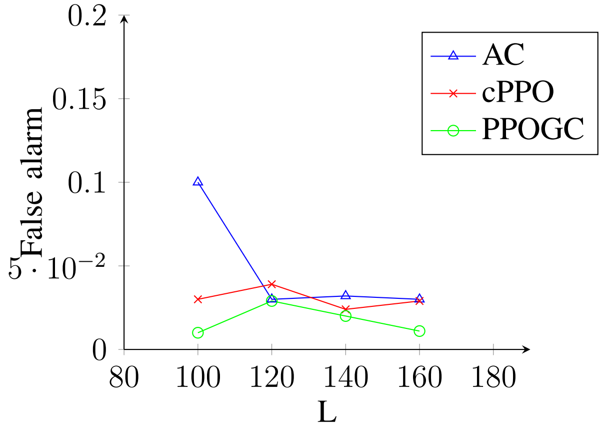
(b) Third decentralised problem

Fig. 8: Average stopping time for the second and third decentralised MAHT problems for different values of L. Since the Accuracy is greater than $1-\epsilon$ and the cost constraint is satisfied we did not include those measures in the plots.

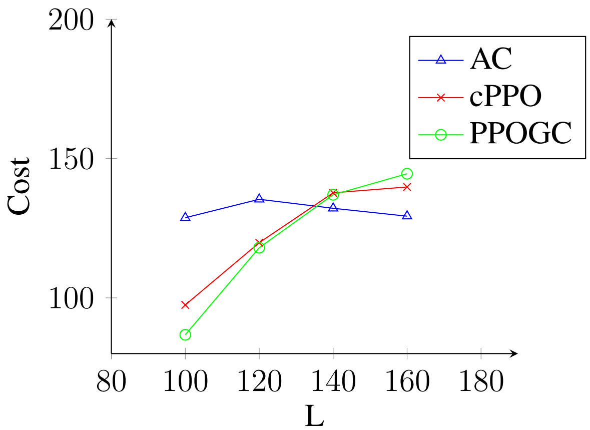
V. Figures the temporal monitoring problems



(a) Detection delay

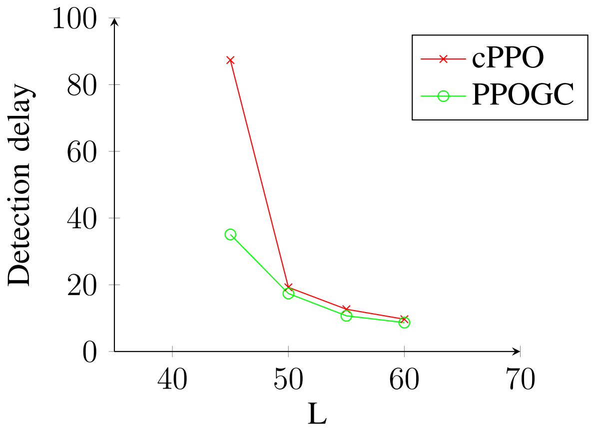


(b) False alarm rate

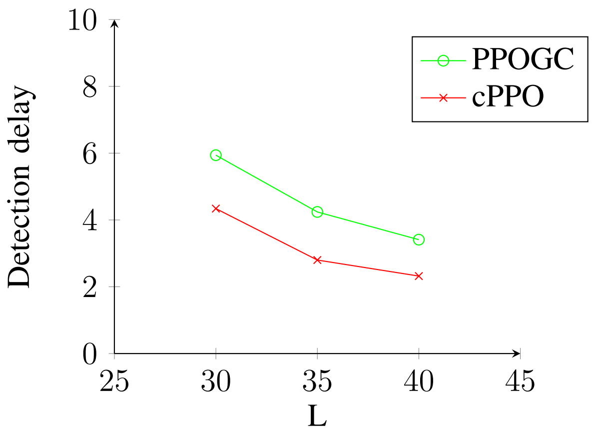


(c) Episode cost

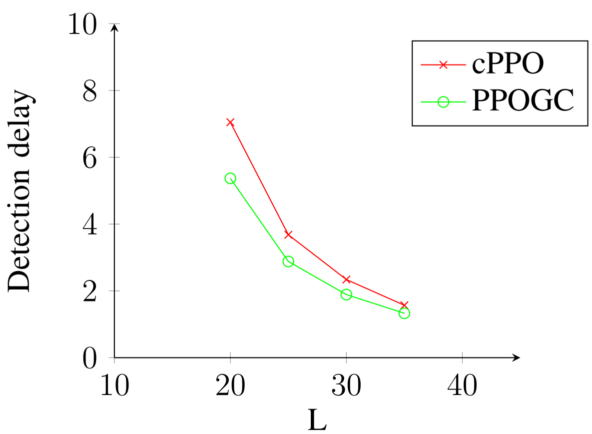
Fig. 9: Average detection delay, False alarm rate and episode cost for the MATMAD problem: Binomial observations.



(a) Detection delay:Binomial Data



(b) Detection delay: Gaussian data



(c) Detection delay: TON_IOT data

Fig. 10: Average detection delay for the MATMMAD problem. Since the cost constraints are satisfied by both algorithms and the false alarm rate is always les that 0.01 we did not include these plots.