Online Incident Response Planning under Model Misspecification through Bayesian Learning and Belief Quantization

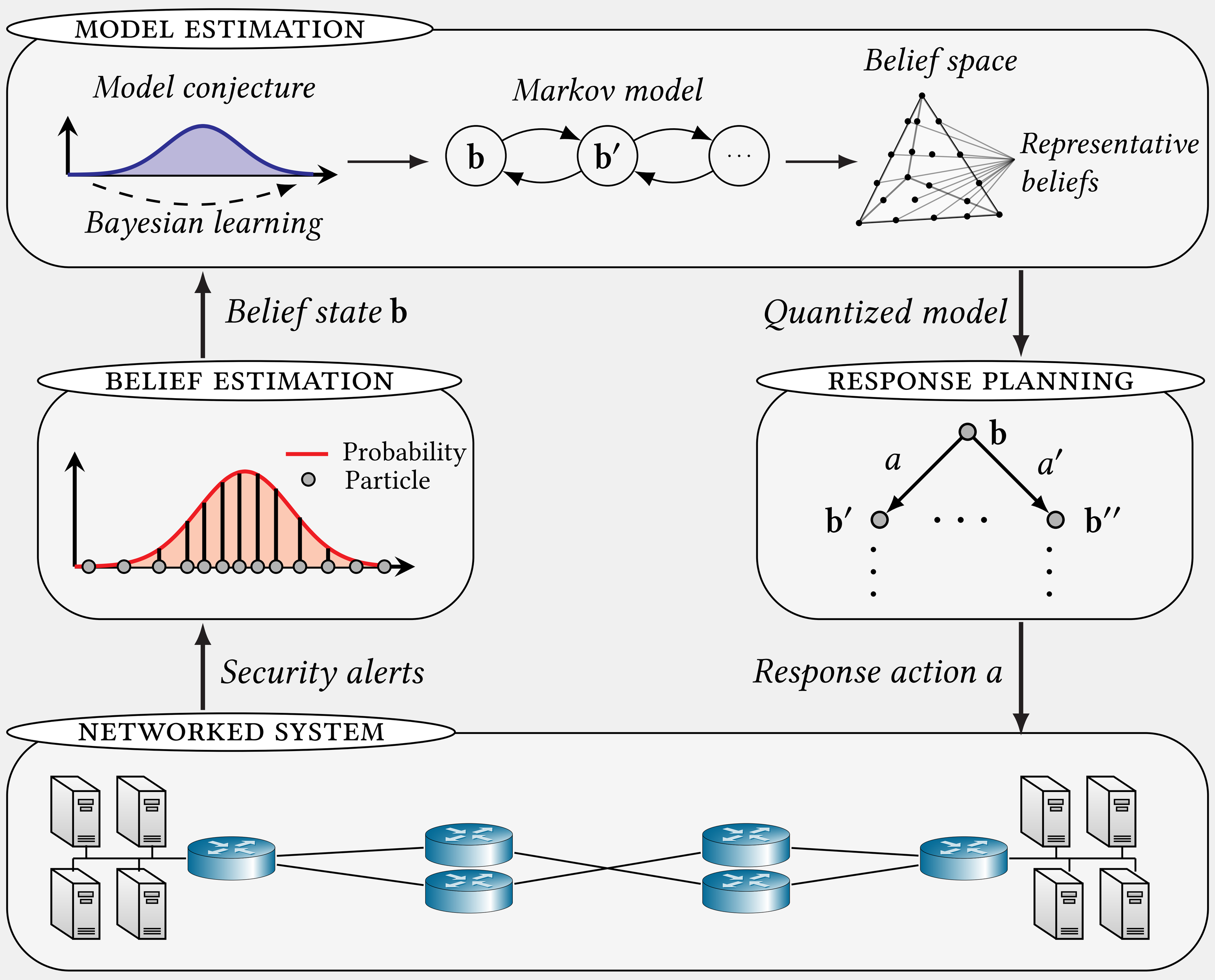
18th ACM CCS Workshop on Artifical Intelligence and Security
Taipei, Taiwan, October 17 2025

Kim Hammar (kimham@kth.se) Tao Li (li.tao@cityu.edu.hk)
KTH Royal Institute of Technology and City University of Hong Kong

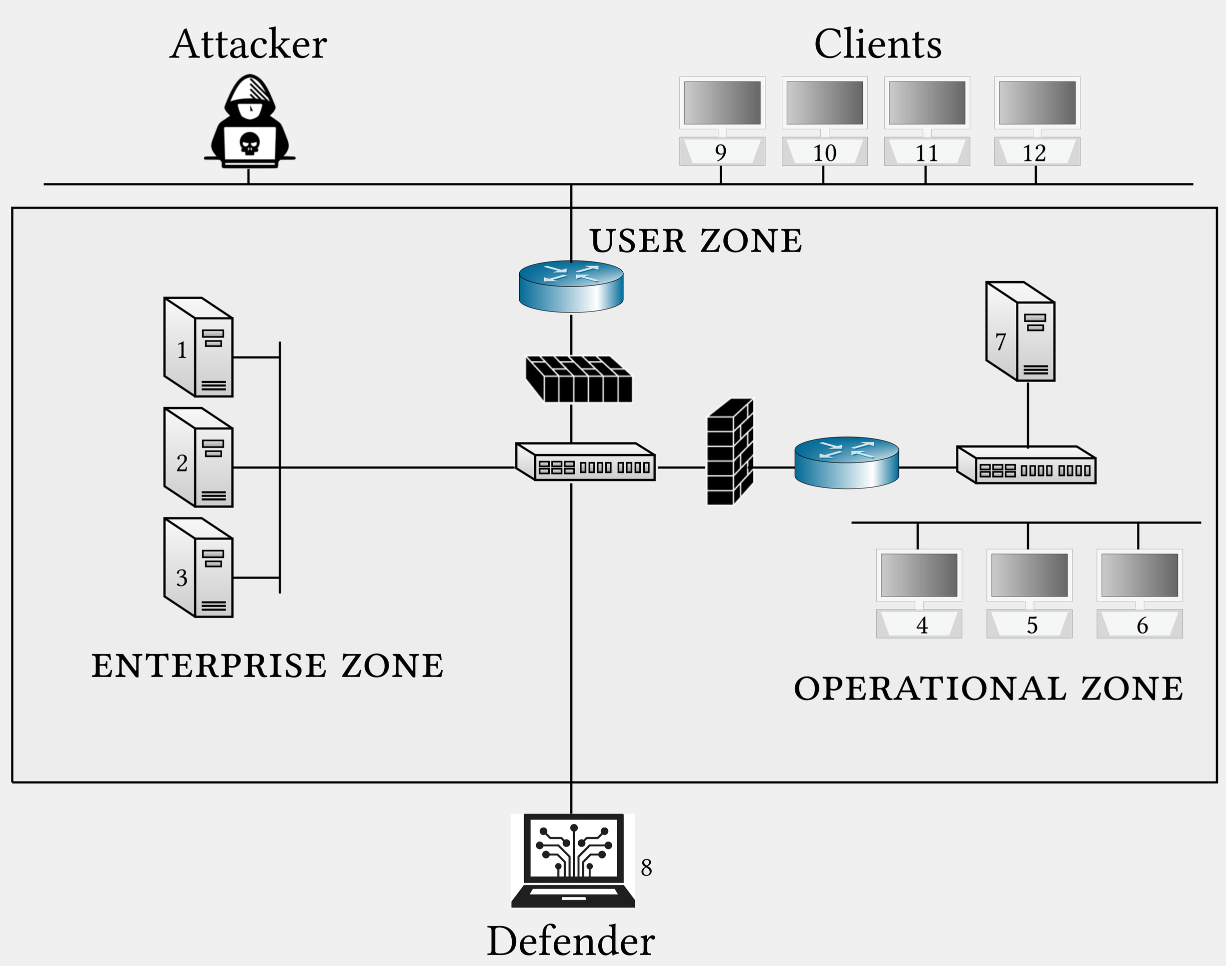
Contributions

1. We present MOBAL, an online method for incident response planning.
2. We establish bounds on misspecification and quantization errors.
3. We show that MOBAL obtains **state-of-the-art performance** on CAGE-2.

Misspecified Online Bayesian Learning (MOBAL)



POMDP Model of Incident Response



We formulate incident response planning as a POMDP and seek to find a near-optimal response strategy π that maps belief states to response actions.

Theoretical Results (Informal)

Proposition 1 (Consistent conjectures). The model conjecture learned by MOBAL is asymptotically consistent with respect to the information feedback.

Proposition 2 (Misspecification error bound). The difference between the conjectured optimal cost function $\overline{J}^{\star}$ and the true optimal cost function $J^{\star}$ is bounded as

$$\|\overline{J}^{\star} - J^{\star}\|_{\infty} \leq \frac{\gamma \alpha c_{\max}}{(1-\gamma)^2},$$

where γ is the discount factor, α quantifies the difference between the transition probabilities in the conjectured model and the true model, and $c_{\max}$ is the maximum stage cost.

Proposition 3 (Approximation error bound). The difference between the cost function approximation $\tilde{J}$ obtained through quantization and the conjectured optimal cost function $\overline{J}^{\star}$ is bounded as

$$|\tilde{J}(\mathbf{b}) - \overline{J}^{\star}(\mathbf{b})| \leq \frac{\epsilon}{1-\gamma},$$

where γ is the discount factor and ϵ is the maximum variation of $\overline{J}^{\star}$ within each belief space partition.

Proposition 4 (Asymptotic (conjectured) optimality). The cost function approximation $\tilde{J}$ obtained through quantization converges to the conjectured optimal cost function $\overline{J}^{\star}$ as $r \to \infty$, where r is the quantization resolution.

Theorem 1 (Sub-optimality bound of MOBAL). The sub-optimality of the cost function approximation $\tilde{J}$ obtained through MOBAL is bounded as

$$\|\tilde{J} - J^{\star}\|_{\infty} \leq \frac{\epsilon}{1-\gamma} + \frac{\gamma \alpha c_{\text{MAX}}}{(1-\gamma)^2}.$$

Evaluation Results on the CAGE-2 Benchmark

Method	*Offline/Online compute time (min)*	*Cost (↓ better)*
No misspecification		
MOBAL	0/8.50	15.19 ± 0.82
CARDIFF	300/0.01	$\mathbf{13.69} \pm 0.53$
PPO	1000/0.01	119.02 ± 58.11
C-POMCP	0/0.50	$\mathbf{13.32} \pm 0.18$
POMCP	0/0.50	29.51 ± 2.00
Misspecification		
MOBAL	0/8.50	$\mathbf{35.91} \pm 9.01$
CARDIFF	300/0.01	94.28 ± 33.27
PPO	1000/0.01	124.38 ± 55.49
C-POMCP	0/0.50	92.71 ± 27.67
POMCP	0/0.50	91.51 ± 28.23

(C-POMCP and CARDIFF are state-of-the-art methods.)

Online Response Planning, Belief Estimation, and Bayesian Learning

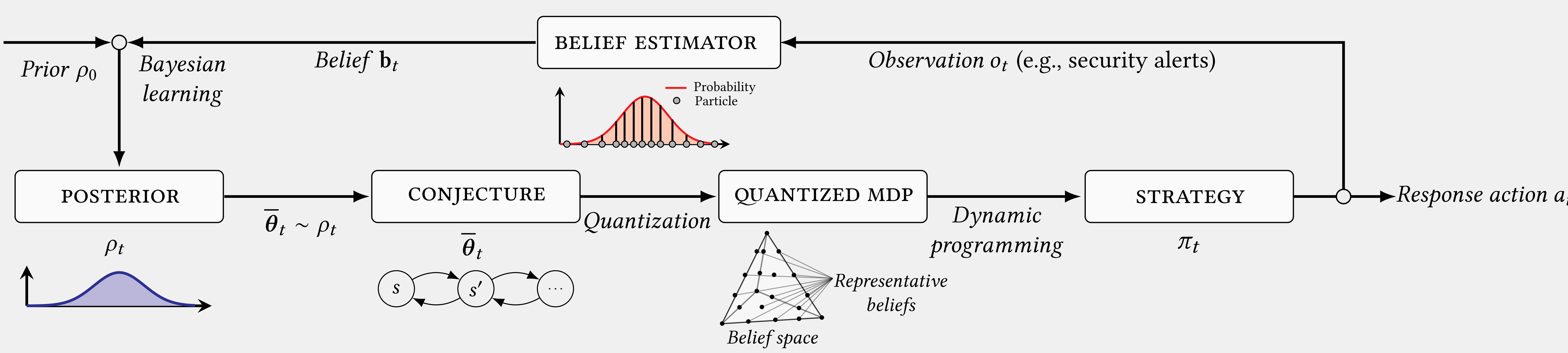


Figure: MOBAL: an iterative method for online learning of incident response strategies under model misspecification. The figure illustrates a time step during which (*i*) the posterior distribution over possible system models is updated via Bayesian learning based on feedback from the system; (*ii*) a conjectured model is sampled from the posterior and quantized into a computationally tractable MDP; and (*iii*) a response strategy is computed using dynamic programming. **Preprint:** https://arxiv.org/pdf/2508.14385.