Table 6: The hyperparameter settings used during the evaluations.

Name	Description	IEEE 14 Value
DRQN hidden sizes	The number of neurons at each hidden layer of the Q Networks	[128, 128]
DRQN activation function	The function applied at each neuron in the hidden layers	Leaky Relu
DRQN dropout rates	The percent of dropout used between each layer	[0.7, 0.5, 0.5]
DRQN optimizer	The type of optimizer used to update the Q network weights	Adam
DRQN learning rate	The learning rate of the optimizer	0.00005
DRQN batch size	The number of experiences used to compute each gradient update	1024
DRQN sequence length	The number of past observations fed through the DRQN when performing a policy update for a sample	50
DRQN target update frequency	The number of learning steps between updating the target model’s parameters from the main model’s parameters	75
Gamma	The discount factor used when computing the total discounted, future rewards	0.9
Operator Boltzmann starting temperature	The initial temperature value for the operator’s Boltzmann policy exploration value	2.7
Operator Boltzmann temperature decay	The amount subtracted from the operator’s Boltzmann temperature to reduce exploration over the course of training	0.05
Operator Boltzmann decay frequency	The number of policy updates between Boltzmann temperature decay steps	250
Operator Minimum Boltzmann temperature	The smallest temperature value that is allowed during training of the operator agent	0.1
Detection epsilon start	The beginning value for the attack detection agent’s exploration amount	0.4
Detection epsilon decay	The amount subtracted from the detector’s epsilon each update	0.1
Detection epsilon decay frequency	The number of policy updates between epsilon decay steps	200
N_Simulations	The number of actions that the simulation policy test for each step in the real environment	10
N_Actors	The number of processes, per trainer that are spawned to generate experiences in the simulator	1
c_{GC}	The operator reward coefficient for the load served reward	4.5
c_{PL}	The operator reward coefficient for the powerline capacity reward	-4.5
c_{GD}	The operator reward coefficient for the generator dispatch reward	0.25
c_1	The coefficient of the immediate reward in the operator agent’s simulation policy	15
c_2	The scaling factor of the false positive reward of the detection agent	0.017
c_3	The scaling factor of the false negative reward of the detection agent	0.019