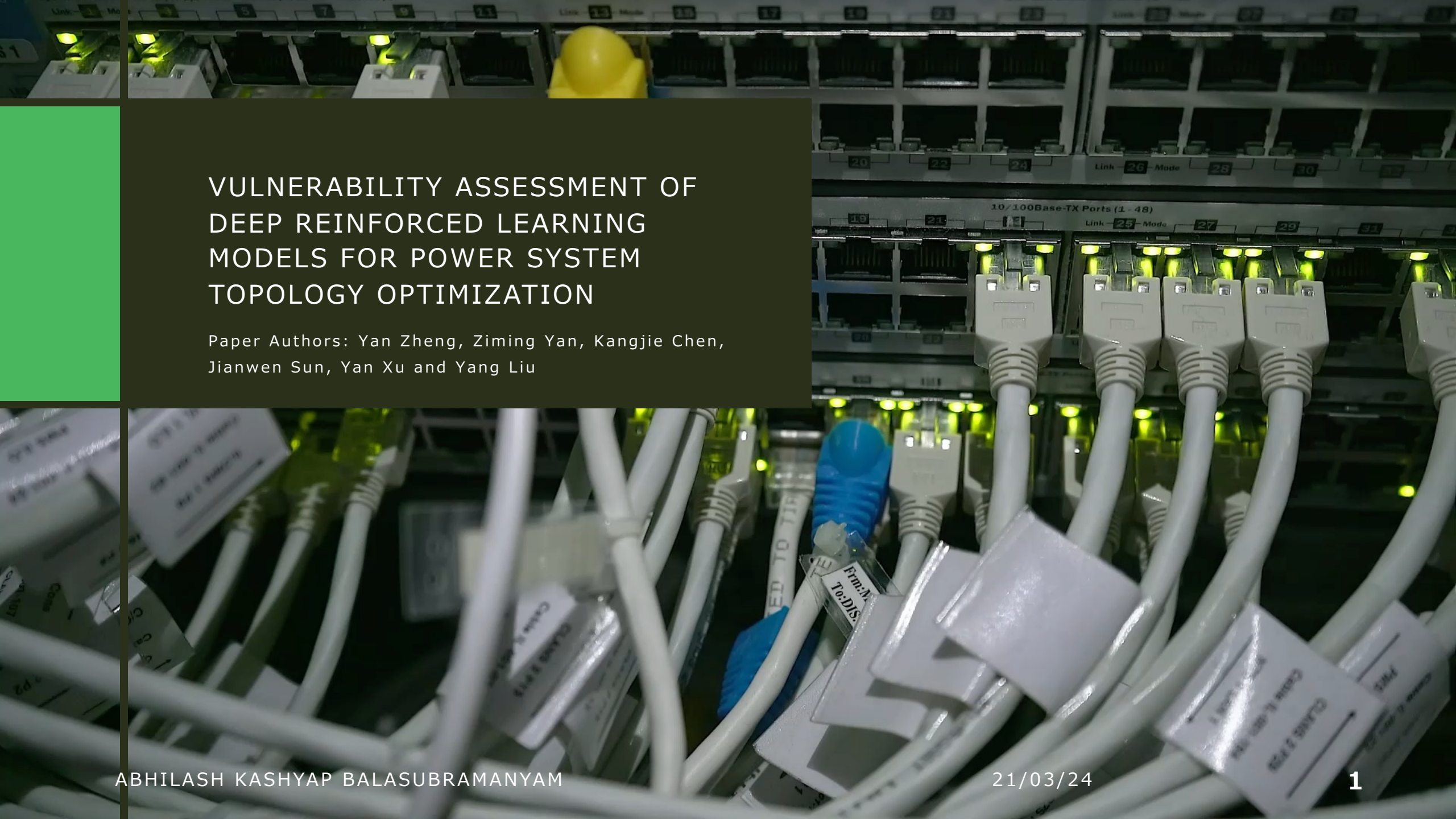
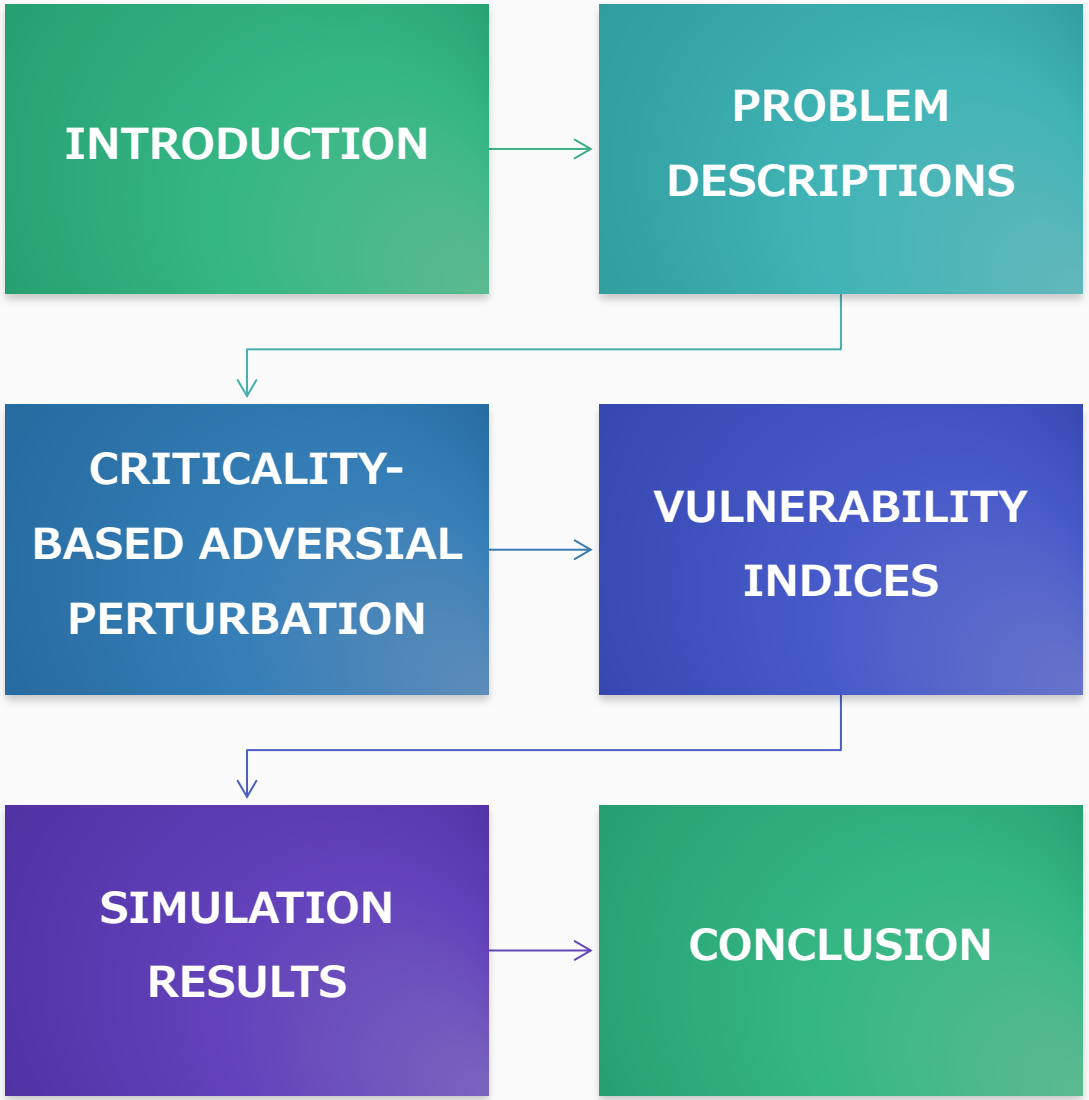


# VULNERABILITY ASSESSMENT OF DEEP REINFORCED LEARNING MODELS FOR POWER SYSTEM TOPOLOGY OPTIMIZATION

Paper Authors: Yan Zheng, Ziming Yan, Kangjie Chen, Jianwen Sun, Yan Xu and Yang Liu

# CONTENTS


INTRODUCTION
PROBLEM DESCRIPTIONS
CRITICALITY-BASED ADVERSIAL PERTURBATION
VULNERABILITY INDICES
SIMULATION RESULTS
CONCLUSION

# SECTION I: INTRODUCTION

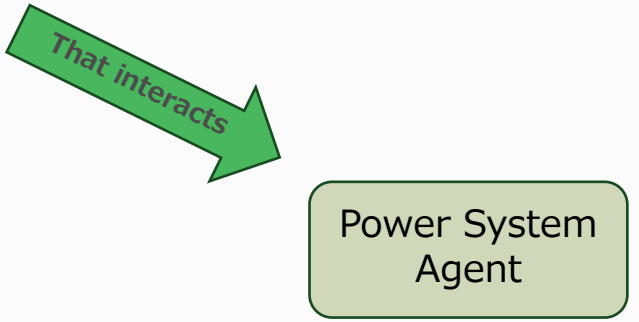
| Motivation | Problem | Proposed Solution | Targeted users |
| --- | --- | --- | --- |
| • Complex Power Grids<br>• Insufficiency of Traditional Methods<br>• DRL Vulnerable to cyber attacks and data perturbations | • Limited research on DRL Vulnerability assessment in Power systems | • Criticality-Based Perturbations<br>• Vulnerability Indices | • Grid operators - Safe grid operations |

# SECTION II: PROBLEM DESCRIPTIONS

**1. DRL for Network Topology Optimization**

Focus on Optimizing Power flow by switching transmission lines or disconnecting loads

Achieved using DRL agents

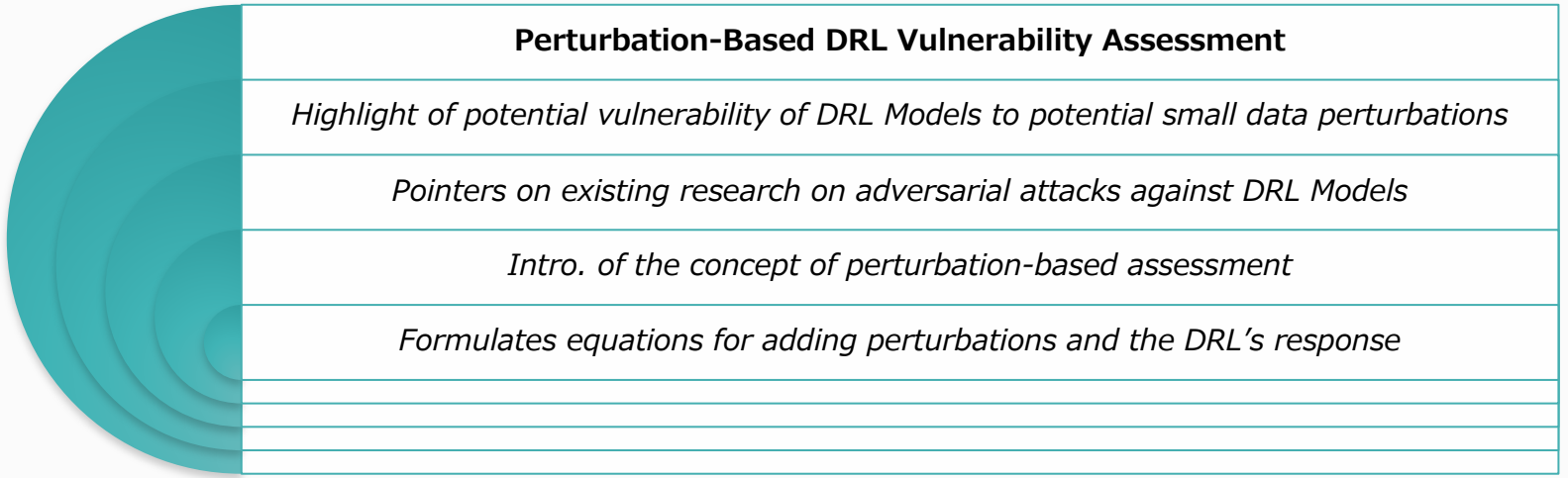


# SECTION II: PROBLEM DESCRIPTIONS

## 2. Technical Details

- *Describes objective function* → Maximizes → Remaining Transfer Capabilities
- Power flow solution using power flow equations
- Markov Decision Process (MDP) is used to formulate the decision-making process
- Reward function → Penalizes → Power Flow divergence and satisfies transmission line limits
- Deep Q Neural Network (DQN) → Train → DRL Agent to find Optimal Policy

# SECTION II: PROBLEM DESCRIPTIONS



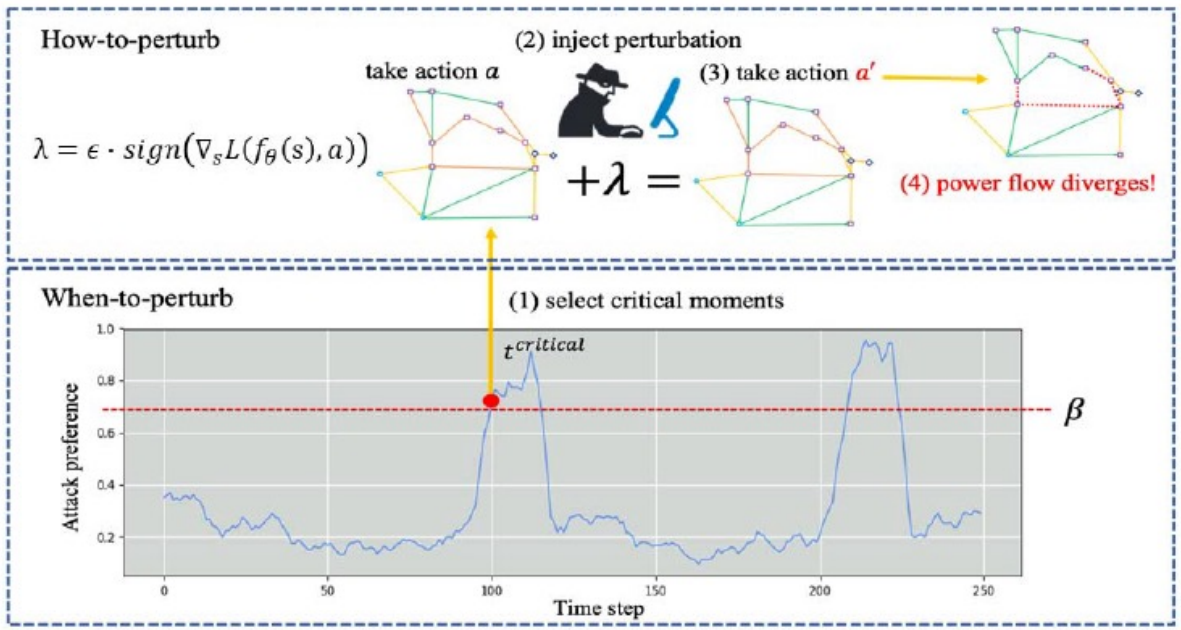



Fig. 2. Illustration of criticality-based adversarial perturbation.

# SECTION III: CRITICALITY-BASED ADVERSARIAL PERTURBATION

### WHEN TO PERTURB

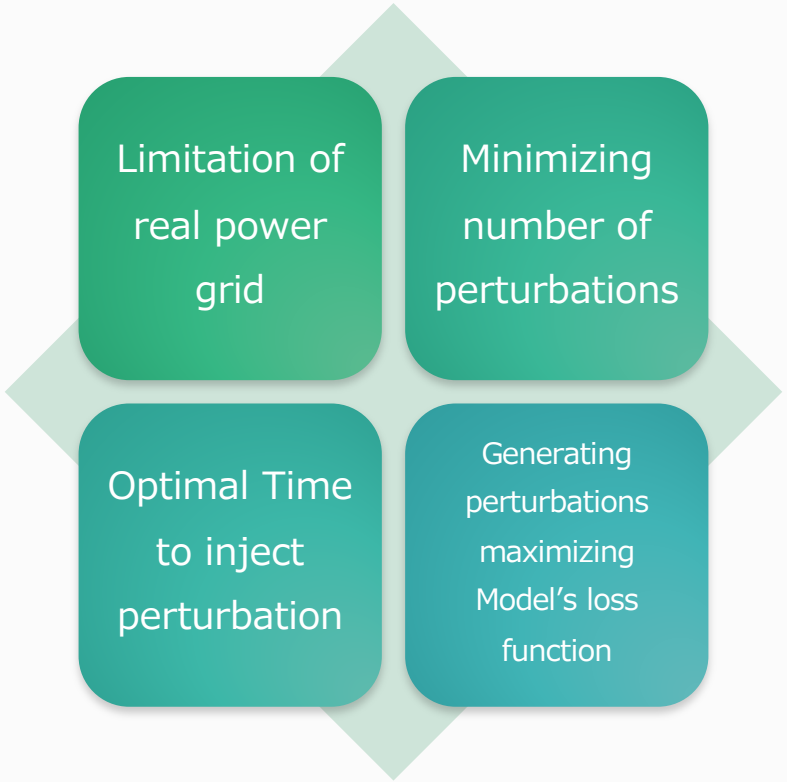
- Critical moment identification
- Model -> Calculating attacker-preference value
- Higher preference values

### HOW TO PERTURB

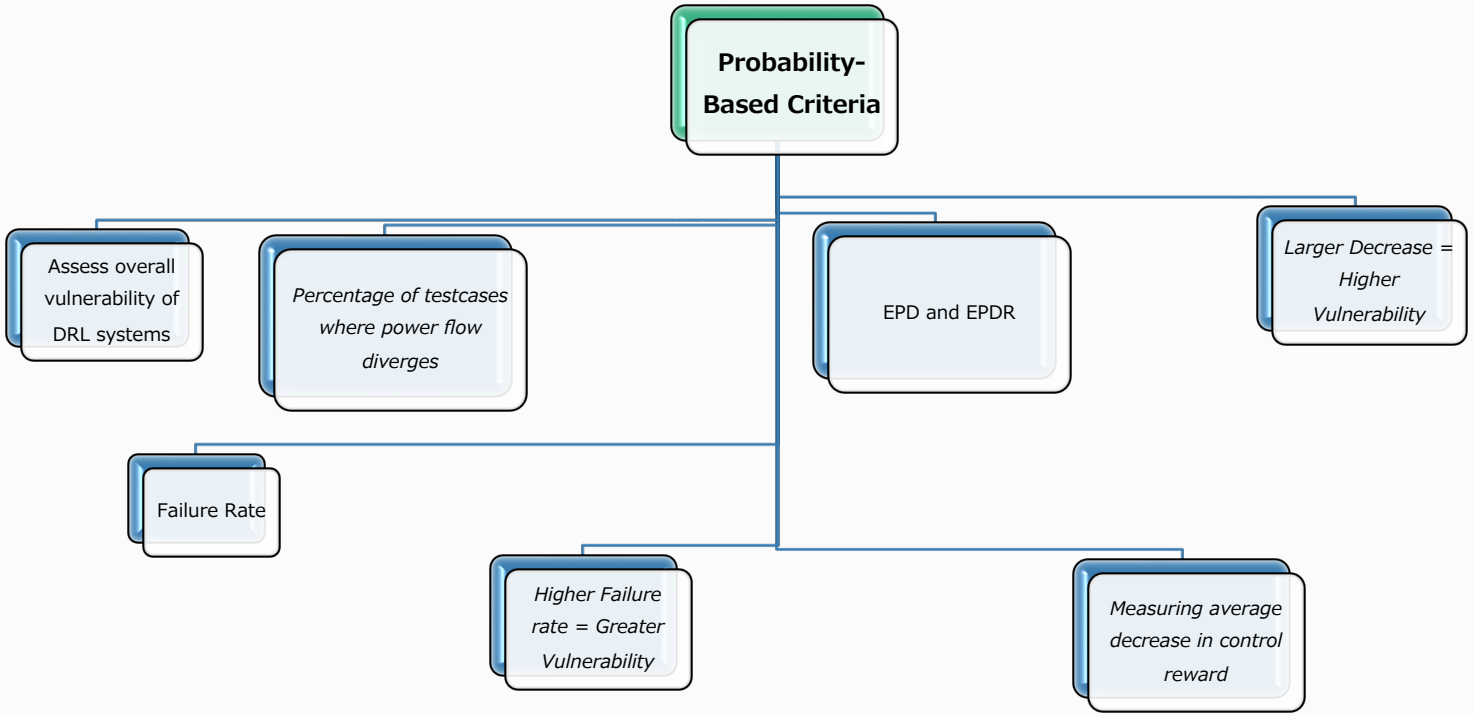
- Focus on crafting effective perturbations
- Model leverages Fast Gradient Sign Method (FGSM)

# SECTION III: CRITICALITY-BASED ADVERSARIAL PERTURBATION

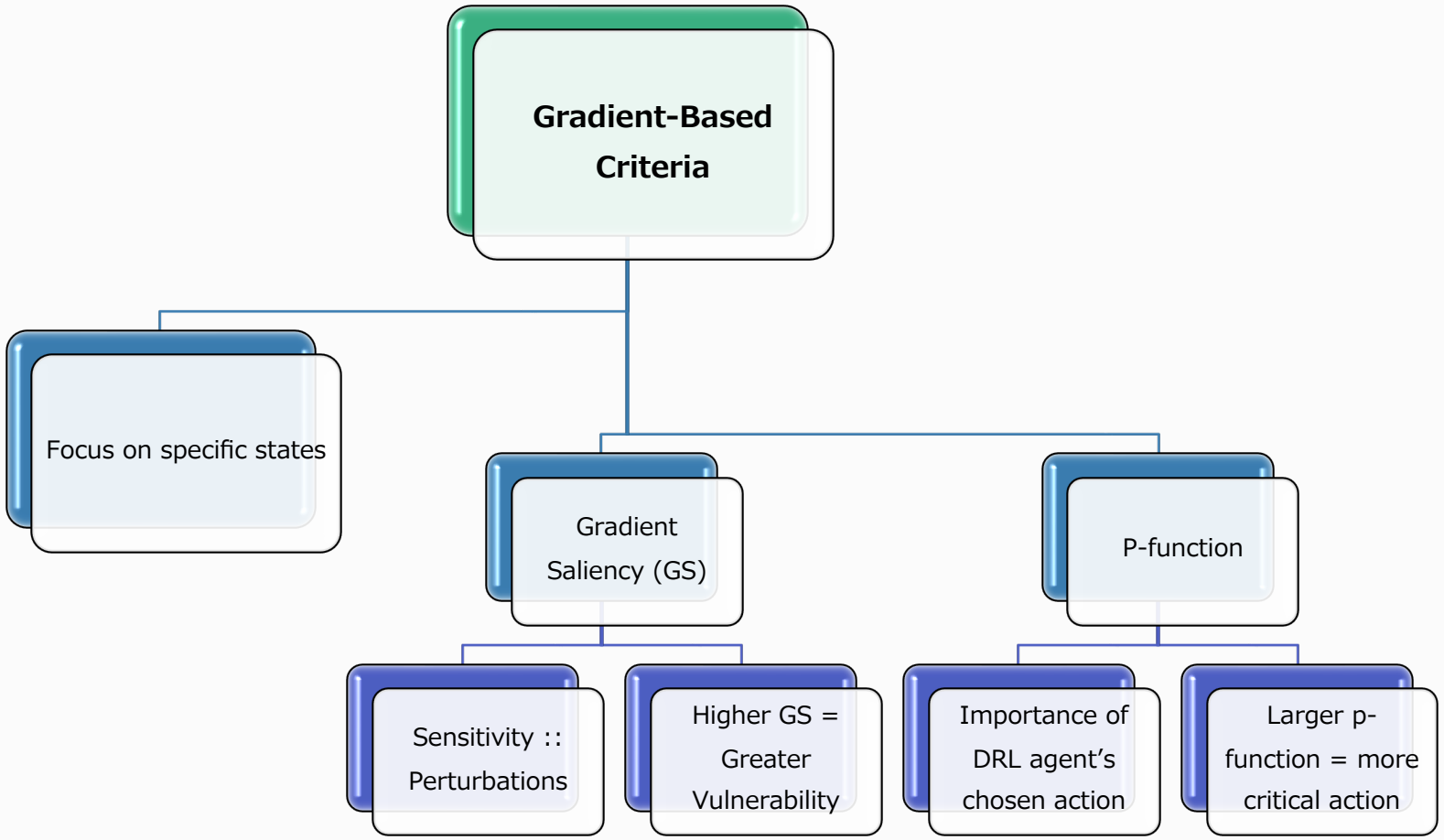
## KEY POINTS




# SECTION III: CRITICALITY-BASED ADVERSARIAL PERTURBATION

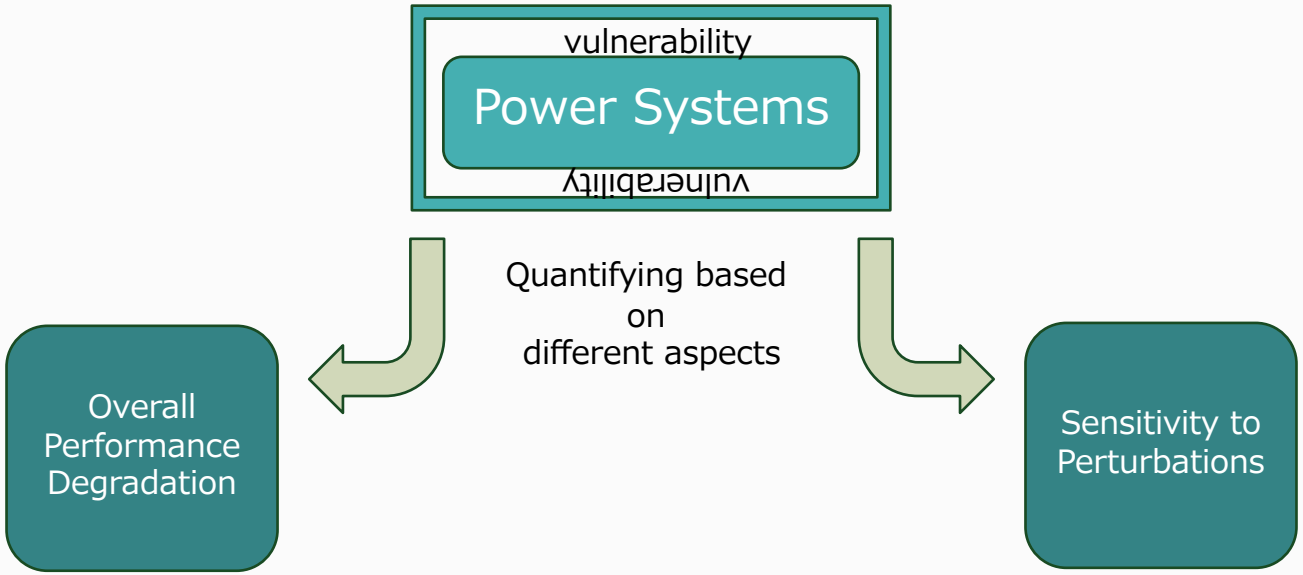



# SECTION IV: VULNERABILITY INDICES



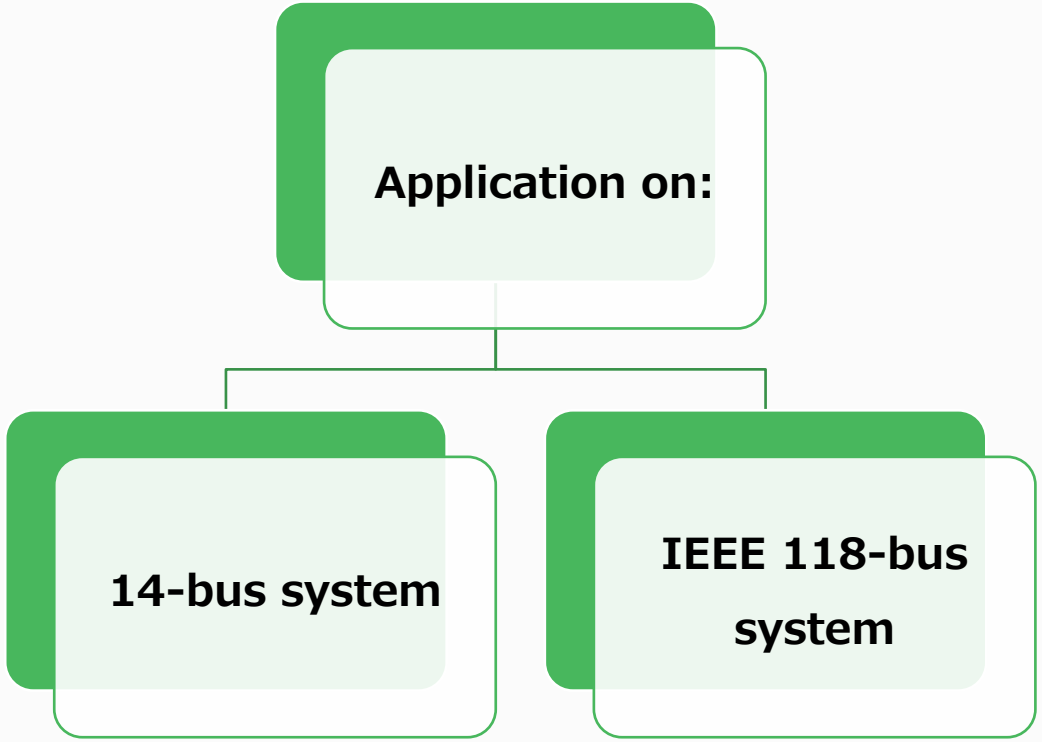

# SECTION IV: VULNERABILITY INDICES

## In Summary




# SECTION IV: VULNERABILITY INDICES

# SECTION V: SIMULATION RESULTS



# SECTION V: SIMULATION RESULTS

## Evaluation Method

### DRL Controller Performance

- No Perturbations
- Random Noise Perturbations
- Targeted Attacks (FGSM and Criticality-based)

### Metrics - Accessing Vulnerabilities

- Critical Attack Rate (CAR)
- EPD and EPDR
- Action Preference (p-function)
- Gradient Saliency (GS)

# SECTION V: SIMULATION RESULTS

## KEY FINDINGS

### DRL CONTROLLER PERFORMANCE

- Degrades with FGSM and Criticality based attacks
- No degrade with no perturbations and random noise

### PROPOSED METHOD

- Higher efficiency over FGSM
- Since, fewer attacks achieve similar performance

### CRITICALITY-BASED APPROACH

- Higher Critical Attack Rate compared to FGSM

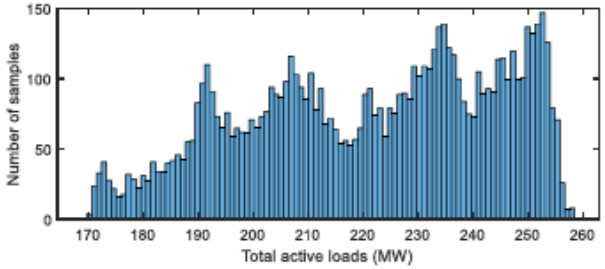
### P-FUNCTION > GS

- Prediction of potential DRL malfunction before they occur
- GS - not effective

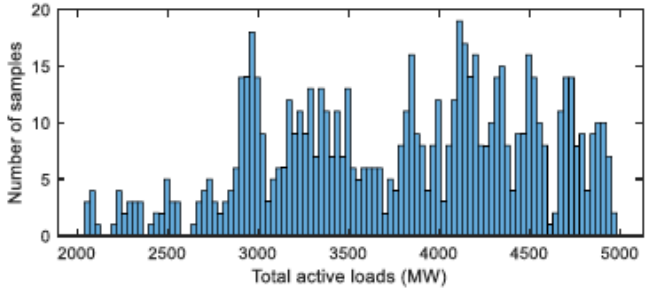
**Impact of attacks is more severe with the 14-bus system vs 118-bus system**

# SECTION V: SIMULATION RESULTS

**Impact of attacks is more severe with the 14-bus system vs 118-bus system**




(*a*) Histogram of the employed 14-bus system total load

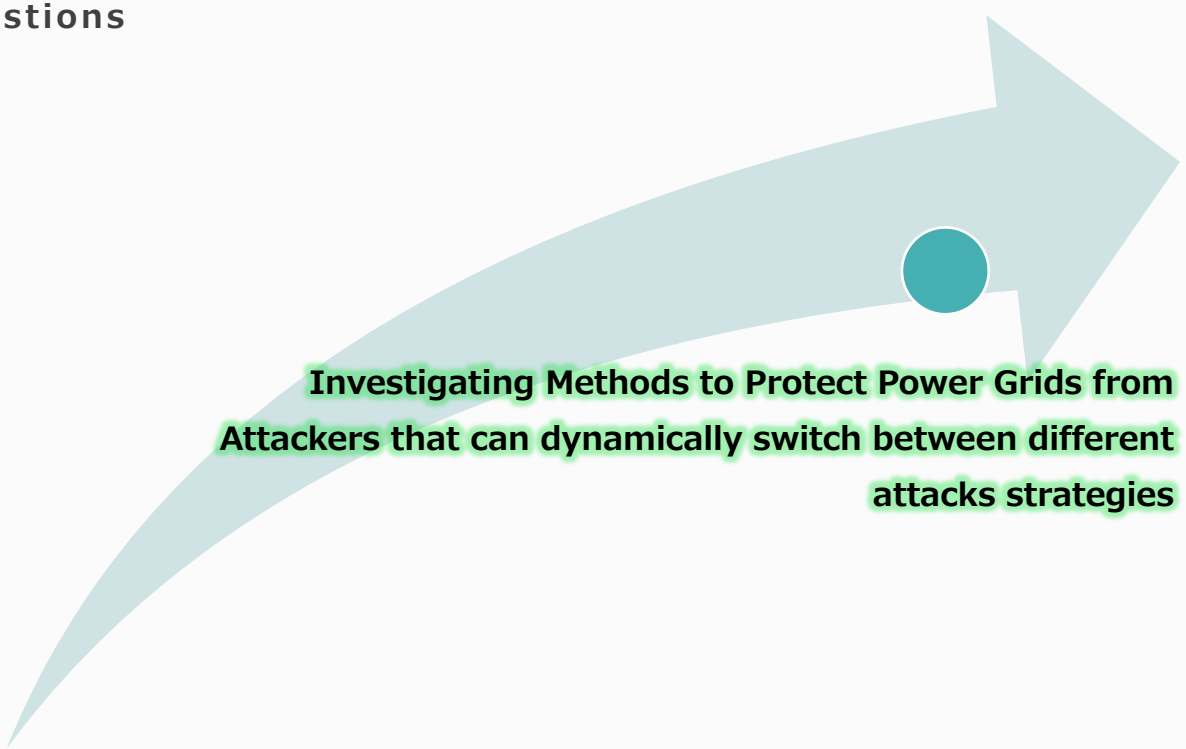



(*b*) Histogram of the employed IEEE 118-bus system total load

# SECTION VI: CONCLUSION

# SECTION VI: CONCLUSION

**Future Work suggestions**




THANK YOU
FOR YOUR
PATIENCE ☺