NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
School of Computer Science and Engineering
College of Engineering

Towards Optimal Defences on Adversarial Examples

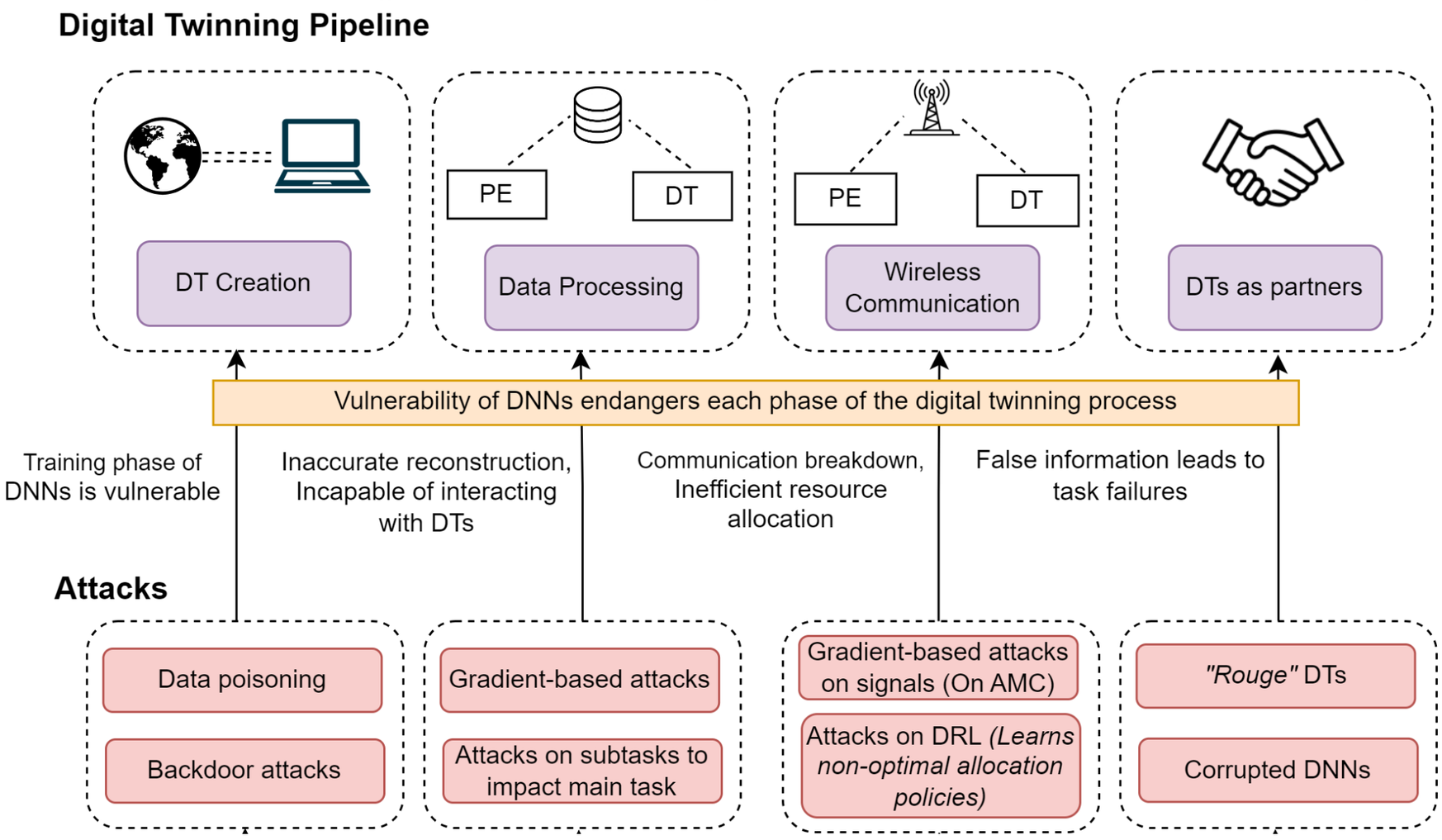
for DNN-Driven Digital Twinning

Student: Lee Yew Chuan Michael Supervisor: Asst Prof Zhao Jun

Background:

Deep-Neural-Networks (DNNs) are fundamental to each phase of the digital twinning pipeline. However, the vulnerability of DNNs to adversarial examples have been well studied. While defenses exist to tackle them, they often incur large tradeoffs and encounter resource limitations. With the increasing use of DNNs in the digital twinning process for the Metaverse, solutions are required.

Reliance of Digital twinning on DNNs and the underlying vulnerability:



Digital twinning can be split into four phases

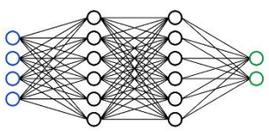
Adversarial attacks are inherent at very phase

Project Objectives:

This paper introduces a framework that uses deep reinforcement learning (DRL) as an optimizer to reduce the tradeoffs and improve the feasibility of defenses for DNNs supporting the digital twinning process. This will ensure that the digital twinning pipeline is robust and able to support real-time usage.

Our Framework:

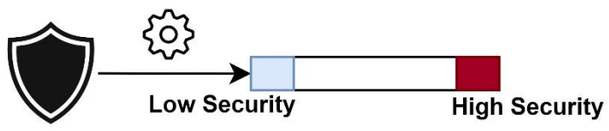
General Framework



DNNs enable digital twins

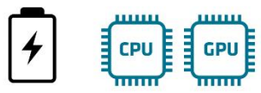
Attacks & Defences

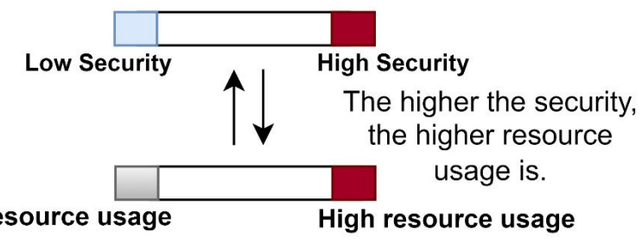
DNNs come under attack. Hence, requiring defences



Different settings in defences consume different amount of resources and yields different levels of security.

Defences, resource usage, & tradeoffs





DRL as optimisers

DRL is used as an optimiser to balance the tradeoffs. This allows defences to be used feasibly