SaTC: CRII: Backdoor Detection, Mitigation, and Prevention in Deep Neural Networks

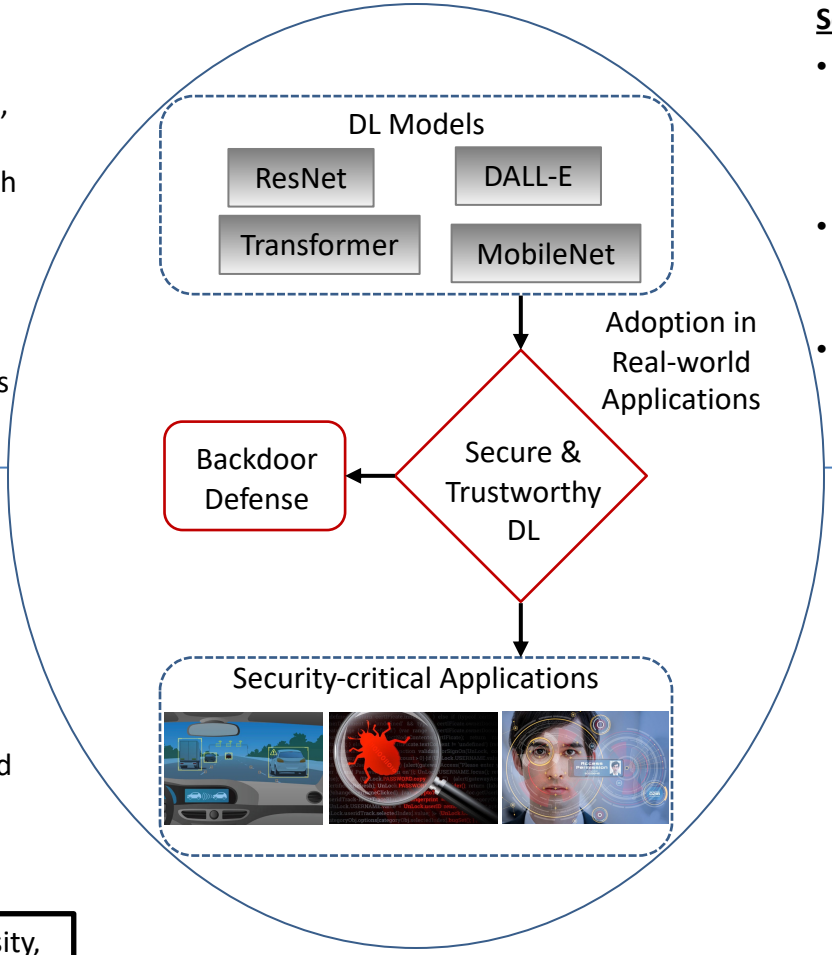
OLD DOMINION UNIVERSITY

Challenge:

- While Deep Learning (DL) is embraced as an important tool, it is becoming an increasingly attractive target for attacks such as neural backdoor
- There is an urgent need for developing solid solutions to detect and mitigate backdoor attacks in deep learning models

Solution:

- Investigate fundamental and comprehensive understanding of neural backdoor attack
- Design efficient and scalable detection schemes
- Devise effective eradication and prevention approaches to neural backdoor attacks



Scientific Impact:

- Unveil the generalizable principles and theories that enable a deep understanding of various neural backdoor attacks
- Based on the principles and theories, new schemes will be derived for backdoor defense
- Develop a testbed to facilitate and standardize experimental neural backdoor research

Broader Impact:

- Lead to enabling technologies to secure deep learning systems, accelerating their development adoption in various applications
- Make a substantial contribution to cybersecurity workforce development by offering training opportunities to students
- The developed testbed will be a first-of-its-kind development and experimental platform to support secure AI research

CNS-2153358, Old Dominion University, PI: Rui Ning