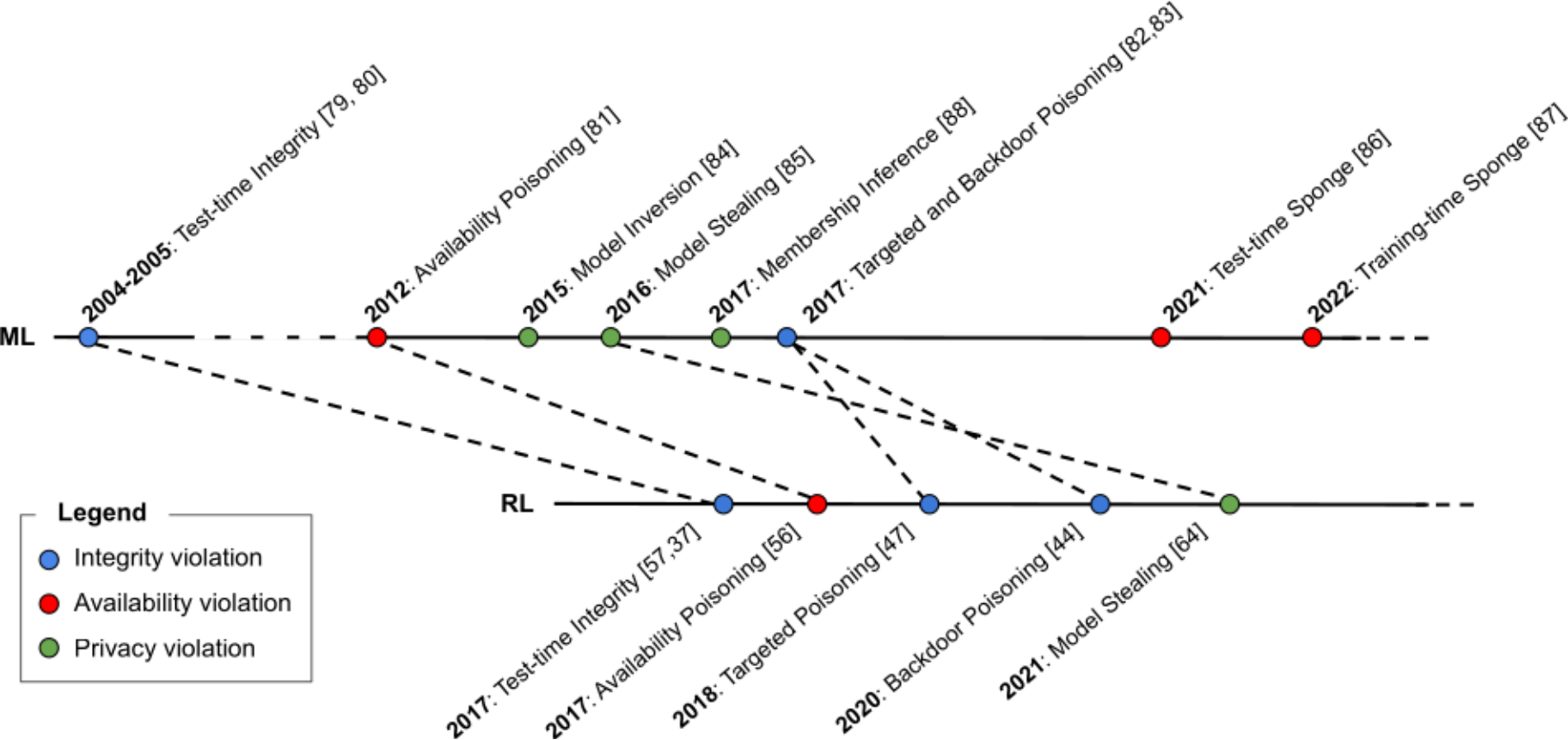

ML
RL
2004-2005: Test-time Integrity [79, 80]
2012: Availability Poisoning [81]
2015: Model Inversion [84]
2016: Model Stealing [85]
2017: Membership Inference [88]
2017: Targeted and Backdoor Poisoning [82,83]
2021: Test-time Sponge [86]
2022: Training-time Sponge [87]
2017: Test-time Integrity [57,37]
2017: Availability Poisoning [56]
2018: Targeted Poisoning [47]
2020: Backdoor Poisoning [44]
2021: Model Stealing [64]
Legend
Integrity violation
Availability violation
Privacy violation