网络流量生成的表示层级与目标一致性综述

王碧莹[1]，王宝生[1]，赵双[1,2]，苏金树[1,3]，陈曙晖[1]，王敏欣[1]，李正芃[1]，魏子令[1]
[1]国防科技大学计算机学院，中国长沙市，410073
[2]湖南警察学院信息技术（网监）系，中国长沙市，410138
[3]军事科学院，中国北京市，100091

摘要：网络流量研究依赖于大规模、高质量的流量数据。然而，由于隐私限制、采集成本、类别不平衡以及数据持续更新等原因，获取此类数据仍然十分困难。这些挑战增加了研究人员对流量生成的关注。尽管目前已提出许多生成方法，现有研究和综述往往忽略了两个关键问题：生成何种形式的流量以及它能支持哪些实际目标。基于2019年至2026年间发表的113份候选记录，本综述从表示层级和目标一致性两个视角，对其中39项代表性网络流量生成研究进行详细分析。我们将现有方法归纳为4个表示层级，并分析了生成数据与使用场景间的关系。我们发现，许多方法保留了有利于分类和入侵检测等下游任务的信息，但任务有用性并不能保证其在真实网络环境中的可重放性或可用性。高层表示虽然更容易建模，但往往会丢弃协议语义、数据包依赖关系和通信逻辑。因此，我们区分了任务一致性与协议一致性，并指出后者目前仍缺乏足够探索。我们进一步总结了评估实践，讨论了各层级特有的评估指标，指明了未来研究方向，包括可控流量生成、协议感知的状态一致合成以及面向工程的评估方法。

关键词：网络流量生成；流量表示层级；任务一致性；协议一致性
https://doi.org/10.1631/ENG.ITEE.2026.0095