参会记录：“互联网基础行为测量与分析”课题四研讨会

1 日程安排

时间：2021 年 6 月 11 日 9:00-12:00

地点：线上会议、清华大学深圳国际研究生院大楼 1619 会议室（深圳会场）、清华大学 FIT 楼 3-225 会议室（北京会场）

日程安排：

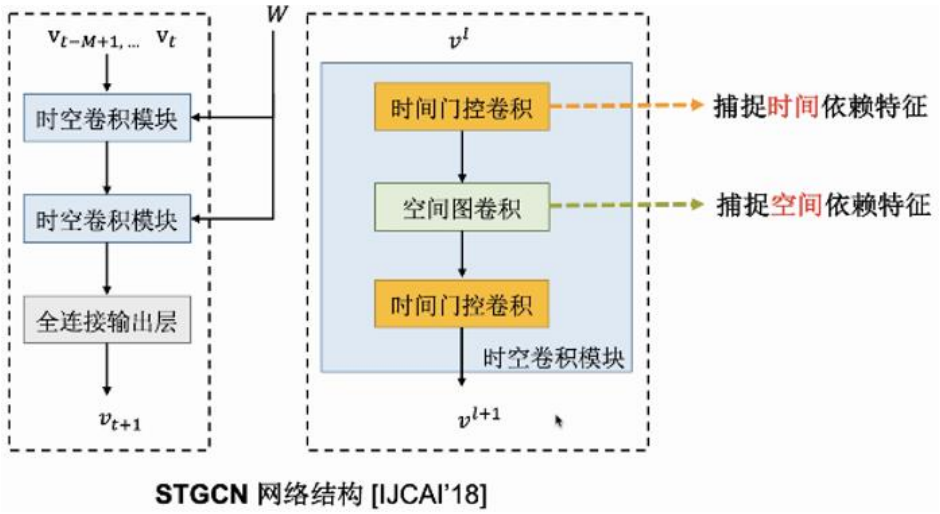
序号	内容	报告单位/报告人
1	项目负责人致辞	清华大学 杨家海教授
2	课题负责人致辞	清华大学深圳国际研究生院 夏树涛教授
3	子课题 1-3 汇报	清华大学深圳国际研究生院
		中国科学院计算所
4	子课题 4 汇报	清华大学
5	子课题 5 汇报	中国科学院信工所
6	下一步工作研讨，包括但不限于： 1、原型系统工作细分和完成时间表 2、与课题 5 对接、对接测试流程和完成时间表	所有单位
7	会议总结	所有单位

2 主要内容

主要旁听了各单位围绕课题开展的科研工作。

2.1 数据中心网络流量测量分析

研究工作一基于时空图卷积神经网络对 DC-WAN 流量进行预测，测量数据中心流量，分析数据中心的业务流量特征，相对移动平均方法，88%的 DC 对的流量预测更准确，对流量较大的 DC 对，流量预测的准确度较高。



STGCN 网络结构 [IJCAI'18]

建议与问题：

（1）对数据中心流量做观测能够产生什么研究点？由于使用 NetFlow，只能做粗粒度的，可能可以做细粒度的测量；新兴业务流模式与传统业务流模式的区别。

（2）是什么数据中心？大厂内部数据中心

（3）仅与移动平均比，没有与其它方法进行比较？用张量做流量的异常检测，是不是也需要进行对比，目前对比的太简单。

（4）为什么使用图神经网络？基于对时序和空间关系的观察，数据中心流量既有时序特征又有空间特征

（5）NetFlow 数据中提供的端口这些服务只有 80 和 443 吗？不清楚。

（6）工作最后发表在哪？SIGCOMM，IMC 在投。

（7）根据业务来做有什么思路：现在仅通过 QoS 区分业务质量，下一步可能进一步细分；不同业务流量特征波动较大，根据特征差异进行分组。

（8）业界没有别人做过别的工作吗？（数据中心内部流量测量）没有数据中心之间的利用流量预测可以做的下游任务是什么？网管？网络安全？QoS？流量工程

（9）工程也是为了管理业务服务，具体是什么呢？不清楚。

2.2 多源异构测量数据的特征挖掘

（1）提出基于字节标签联合注意力网络的流量分类方法

（2）收集微信行为数据集、Telegram 行为数据集、WhatsApp 行为数据集

（3）在网络基础行为多维度关联分析中，提出了稀疏稠密时空卷积溜预测方法、基于多模态数据融合的带宽预测。

建议与问题：

（1）应当提出用了什么方法，该方法具有有一定通用性，这些方法在流量分析中有一定的作用。

（2）创新点在哪里，体现的不够清晰。

2.3 异常网络基础行为的在线检测与预测

（1）提出一种基于异构集成学习的网络流量异常检测算法 HELAD，结合无监督 Autoencoder、有监督 LSTM、以及集成学习的优势。

（2）挖矿流量检测（特征提取，多维关联，基于统计模型进行检测）、IoT 设备流量分类（提出半监督的精确分类方法、提出基于 VAE 的无监督设备识别方法）、恶意域名检测（提出图节点分类方法：HinDom、HGDom、DeepDom）

（3）提出基于动态密度反馈机制的活跃 IPv6 地址探测方法，提出基于深度学习的网络安全事件预测算法；解释网络安全异常检测中深度学习模型。

2.4 异常网络基础行为追踪溯源与定位研究进展

（1）网络流量解析与行为标识技术、网络用户与终端标识技术

（2）提出基于预处理降噪模型的匿名流量关联方法

3 附件

旁听 PPT 截图。