Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021)

Data analytics & forensics for emerging cloud native converged networks

Achilles Petras, Paul Veitch @ BT Converged QoE Research

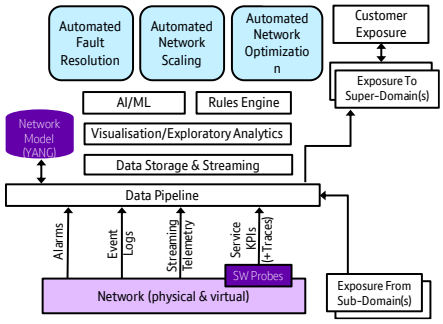
The focus of this whitepaper is on performance of both emerging converged networks and how data collection of various measurement techniques and platforms can support the robust deployment of new technologies (especially in the context of 5G future architecture) and correlate layers of infrastructure, applications or user experience.

From a network operator's perspective, the following trends underpin the research questions we are asked to provide for WOMBIR 2021:

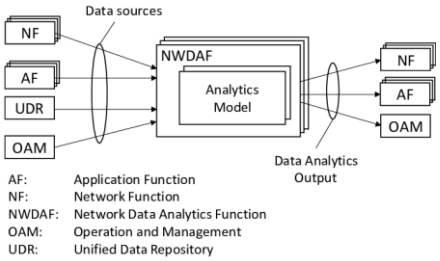
- Over the last 2 years key stakeholders of network operators (business analysts, strategists, marketeers, architects & in-house researchers) have come to a solid realisation that network instrumentation, QoE monitoring and data analytics is at the heart of their success in a highly competitive market. In addition, 2020's pandemic accentuated the imperative reliance on networks to support digital transformation across all verticals (industries) that expect their workforce to be able to work and deliver effective services and solution from home, office or on the move
- NfV has matured and has become a pre-requisite to build cloud native 5G network cores. Thus, IT and networks are undoubtedly intertwined. This adds additional 'layers' that produce valuable data in order to do be able to deliver effective root cause analysis (RCA) across the boundaries of compute infrastructure, virtualisation, traditional IP networking and application transport and presentation layers

Therefore, the main research questions are:

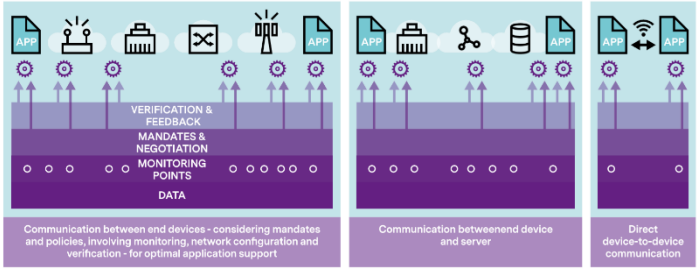
- how information from distinct data sources and systems (eg network node telemetry, passive packet analysis, RAN KPIs, Service KPIs, active testing, crowdsourcing measurements etc) can be correlated and interpreted ("joining the dots") and deliver network and customer experience forensics
    - We could embrace Big Data platforms packaged with data processing, pipelining, analytics and ML tooling (eg Google Cloud); Platform); would this be our route to actionable insights?



- due to softwarisation of networks performance data gathering, correlation and analysis from both compute (eg System statistics using data collection agents such as collectd) and networking layers via physical but mostly virtual probes (passive packet analysis) and agents (active synthetic testing) are required
- How effective Closed Loop Automation (CLA) can be achieved via the Network Data Analytics Function (NWDAF), introduced in 5G core architecture? The purpose of the NWDAF is to provide centralized data collection and analytics functions to other Network Functions (NFs); looks reasonable on paper, would it work in real life?
- CLA depends on ML/AI to cope with a multi-dimensional problem space which brings together a potentially very large number of data sources and mitigating actions. But ML is not a single approach, there are many categories and flavours of ML algorithms suited to very distinct applications.
    - Challenge in choosing and using the appropriate Clustering Analysis or/and Anomaly detection on time series



- Classic bottom up QoE validation by correlating network KPIs and metrics with Applications Score (eg MOS) is not straightforward or fully modelled especially for new emerging Augmented/Virtual/Mixed Reality applications
- Based on emerging 6G documentation[1] we need to consider top to bottom e2e Mandate-Driven Architecture (MDA), which is steered by the needs of applications and users. A "mandate" can be seen as the collection of network services required from underlying network(s) to comply with dynamic end-to-end QoS needs from a specific application, hereby respecting profiles and preferences of users and machines.
- Initial stepping stone is 5G slicing provisioning and troubleshooting where the challenge is to have suitable instrumentation to identify non-compliance for each slice's SLAs: How do we instrument different domains that the e2e service traverses ie in the Radio Access Network (RAN), shorthaul, backhaul or network core which may be managed by different operators, content providers, edge compute providers?
    - Challenge to agree on common best practices, metrics definition in order to have clear accountability on where the problem is and getting it fixed



[1] www.6gchannel.com