**Scalable Databases for Big Data Cellular Networks Analytics**

**PIs**: Profs. Shasha, Rangan, Panwar

**Students:** Marco Mezzavilla (postdoc). The project will need one other student with a database / distributed computing background. Does Dennis or Jinyang know anyone who could do this? Another possibility is to ask Torsten?

**Introduction of Team:** Marco is a recent PhD from Michele Zorzi’s group at the University of Padova, one of the top researchers in wireless networking. Marco did his PhD in various 3GPP networking issues including detailed simulations in ns3. He has worked with Aleks Damnajovic, Qualcom’s MAC-layer lead. The other student would complement Marco’s cellular knowledge with knowledge of the computer science aspects, esp. distributed computing and databases.

**Affiliate Sponsors with direct interest**: This project came out of a discussion with Erik Koster at Ericsson Research, Sweden, where he and Prof. Rangan were panelists in BigData in communications in CTW. He runs a data analytics groups at Ericsson and said this project would have considerable interest in Ericsson where it could ultimately lead to a new product line. I believe the work can also be pitched to the carriers such as AT&T, and potentially VZ should they join.

**Overview Research Statement**: The project will be to develop scalable, distributed database architectures for tracking packet flows, MAC-layer transmissions and channel conditions in large cellular networks.

**Research Thrusts and Improvements for NYU WIRELESS:** Tools for network diagnostics are essential for managing large cellular networks, to identify fault conditions, monitor network performance and plan future network expansion. Current network monitoring generally only captures large scale aggregate data (e.g. bandwidth, cell associations, etc.). The purpose of this proposal is to develop a distributed, scalable database and logging tools that can capture every MAC and networking layer event for every packet amongst a large set (e.g. millions) of users. Many associated PHY-layer reports (e.g. CQI, RRSP, neighbor strength) would also be captured to correlate the data. As networks move to 5G, these sorts of diagnostics will become particularly essential since there is currently little known about data applications that will run on these networks, the devices that will connect to the networks and the sorts of channel conditions that will be encountered. The key technical challenge in the sheer volume, variety and velocity of the data – the three ingredients of BigData : a large carrier such as Verizon can have 100 million subscribers – potentially 1 to 10 million of which are in active mode. UEs in active mode will be generating MAC-layer events every 1ms with PHY-layer reporting at similar time scales. The migration to 5G will increase all these dimensions by orders of magnitude.

Since acquiring data from carriers will be difficult, the first phase of the project will run on simulated data from ns3. Our group has made considerable progress in distributed implementations of ns3 toward this end.

**Existing support and additional justification for the Project**: Marco is being funded partially via CATT funds.

**Justification for this project :** Carriers have always stated that network analytics is their main interest, but there has been little work in academia. The work could thus help bring in new affiliates from the carriers, esp. Verizon. The area is also completely new – at least in published work in academia. The work could therefore be a new area that NYU WIRELESS can lead in.