Large-scale Profile-based Behavioral Modeling, Networking and Computing

(From *Big Data* to Behavior-based Design)

Ahmed Helmy (helmy@cise.ufl.edu), Sanjay Ranka (ranka@cise.ufl.edu)
Computer & Information Science & Engineering (CISE) Department,
University of Florida, Gainesville, FL

A significant part of future pervasive computing systems, networks and services are expected to center around human activity and behavior. Pervasive mobile and sensor networks are expected to grow significantly and integrate more closely with human societies, and hence must accommodate increasing levels of interaction and movement. In such a highly dynamic environment, networked systems need to adapt efficiently to the growth and change along several dimensions, including load, traffic, on-line activity and mobility. Understanding and realistically modeling this multi-dimensional behavioral space is essential to the design and evaluation of efficient protocols and services for the future *Internet of things*. Some of the main challenges in addressing these problems lie in gathering, analyzing and efficiently mining very large-scale mobile network and human behavioral traces, and using the knowledge garnered to design novel behavior-aware privacy preserving mobile pervasive services.

One of the essential components in this line of research is to establish a new class of realistic data-driven behavioral models for mobile users. One potential approach would be to achieve behavioral characterization and profiling based on extensive measurement and analysis of mobile networks data. This data need to be processed using novel efficient data structures and mined using novel and advanced data mining algorithms. Today's data-base and data mining technologies suffer from inherent limitations and cannot scale, both in size of the data and dimensionality of the user behavioral space, to provide an adequate framework for exploration and information retrieval in such a large-scale environment. New techniques and query processing systems that scale orders of magnitude above today's capabilities are thus necessary. Furthermore, the core of such effort would establish a set of behavioral models to enable accurate simulation and efficient design of classes of behavior- aware wireless networks and services. The required fundamental understanding to establish these models would have to be acquired through a systematic investigation of user behavior at various levels of spatio-temporal granularities. Such an effort would analyze individual, pair-wise, and emergent collective community behavior, as well as clustered behavior and social relationships - such as trust and friendship - as relates to network activity.

This white paper attempts to highlight the *grand challenges* in this area of research as well as the *promise* in making significant progress towards meeting these challenges.

Grand challenges: Three specific major research challenges or thrusts are identified, that are essential to advancing the field of large scale pervasive social computing and networking. The first is the efficient processing and mining of mobile network measurements at a very large scale. The development and application of novel and scalable data mining techniques is essential for multi-dimensional analysis and modeling to extract important features, correlations and

behavioral characteristics of wireless communities. The traces to analyze could be collected over many years from a variety of rich contexts including syslogs, MAC traps and netflow traces from campuses and municipal networks, location and encounter traces, smart-phone activity, and vehicular networks, among others. Overall the traces may include hundreds of thousands or millions of users, with data sizes that are several orders of magnitude larger than traces in existing studies. Mining mobile traces with such scale and dimensionality provides research challenges and investigation opportunities in its own right.

The second research challenge is the realistic behavioral modeling of networked mobile societies. The main network related activities include web access patterns, application usage, traffic and mobility. The spatio-temporal characteristics of these activities are correlated, requiring novel and systematic study of their interactions. Insight from the mining analysis shall inform the design of parsimonious behavioral models to achieve accuracy while maintaining tractability. Furthermore, correlations between network activity and social attributes - including interest, trust, and friendship - needs to be investigated and validated through thorough test bed experiments and large participatory deployment.

The third research challenge includes model-driven design of key classes of behavior-aware networking and computing applications that benefit directly from the proposed models. The design aspects include realistic network behavioral simulation, and new communication, computing and trust primitives that are profile-based. The introduction of similarity metrics that draw from behavioral modeling could be a key enabler for these services. The design of fully distributed architectures should also be investigated to facilitate privacy preservation and to support infrastructure-less mobile networks. The models and mechanisms introduced should provide the essential protocol building blocks for the next generation mobile social networks.

The promise: The solutions to the above challenges may rely on two main factors driven by confluence of the following evolution in mobile networking and computing. The first is the unprecedented *tight coupling* between users and mobile devices. The ubiquity of wireless Internet access is making possible the connectivity *anytime anywhere*, where many of the users' daily experiences can be captured (at least partially) by their devices. The second is the increasing capabilities of mobile devices, in terms of computation, communication, storage and sensing, which positions them as potential multi-dimensional sensors and computing nodes that can contribute significantly in participatory scenarios (and crowd sourcing). Both of these factors (tight coupling and increased device capability) enable the devices to be aware of its user's behavioral patterns in multiple-dimensions. Traditional networking and computing paradigms are essentially socially-oblivious; including machine-addressed communications (unicast, broadcast, multicast). Recent advances in social networking rely on centralized (server-based) architectures, introducing significant privacy, vulnerability and efficiency issues and concerns. There is hence a great need for the design of new socially-aware abstractions for future networking and computation architectures, protocols and services. The resources that are contributed by the users are dynamically provisioned for the benefit of the users and their underlying communities. The heart of the solution should include new paradigms for representing users in multi-dimensional behavioral spaces, and then utilizing such space to match on relevant characteristics geared towards and dictated by the applications. We refer to this paradigm as *behavior-based social computing and networking*.