Scientific Report

First name / Family name	Andrea Hess
Nationality	Austrian
Name of the *Host Organisation*	Swedish Institute of Computer Science (SICS Swedish ICT)
First Name / family name of the *Scientific Coordinator*	Daniel Gillblad
Period of the fellowship	14/04/2014 to 14/04/2015

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

I.A. Cellular network data analytics

The main focus of my work in this domain was the study of mobility and communication patterns extractable from a country-wide 3G dataset. In particular, we proposed a user profiling model aiming to capture temporal variations in user behavior, published in [1]. Moreover, several aspects related to content caching have been investigated by means of the same dataset: in [2] application usage was characterized; in [6] dimensioning of a network cache based on request patterns was discussed.
Another research question tackled is the scalable modeling of common movement routes based on cellular network data. With the goal of modeling routes derived from large-scale network data in a compact way while also preserving the privacy of individual travellers, we developed route aggregation methods based on clustering and frequent sequence mining. The approach was first applied to the data from this year's Data for Development (D4D) challenge, described in [3]. It has then been further elaborated, leading to [5].

I.B. Data-driven in-network fault detection

We developed a decentralized link monitoring approach with the objective to support troubleshooting mechanisms in Software-Defined Networks (SDN) by providing highly accurate measurements in a scalable and timely manner. Estimations of link delay and packet loss based on measurement data are computed locally at network nodes. This allows to detect network faults quickly and to send measurement results in compact form to network controllers, and thus, reduces the communication overhead significantly. This work led to [4], a second publication is under preparation.

I.C. Theoretical work in mobility modeling

As continuation of my PhD research, I worked on the definition of guidelines for building and validating representative human mobility models, leading to a journal submission [7].

I.D. Scientific peer reviewing

During the fellowship period, I provided reviews for the IEEE Transactions on Mobile Computing and the Elsevier Vehicular Communications journal.

II – PUBLICATIONS DURING YOUR FELLOWSHIP

II.A. Accepted publications

[1] *Exploring Communication and Mobility Behavior of 3G Network Users and Its Temporal Consistency.* Andrea Hess, Ian March, Daniel Gillblad. IEEE International Conference on Communications (ICC) 2015. (to be published in IEEE Xplore Digital Library)

Abstract: Over the past decade, telecommunication network operators have more and more realized the added value of data analytics for their network deployment efficiency. Early studies targeted the global network perspective by localizing peak loads, both in terms of area and time period. Due to their higher granularity and

information richness, current telecommunication datasets allow increasingly deeper insights into the network activities of the users. Existing network traffic classification studies tend to divide users into groups without considering the transitions between different groups caused by individual behavioral traits, which we expect to show observable regularities. Our approach defines a profiling model that characterizes the user behavior as well as its temporal dynamics from two perspectives: w.r.t. *(i)* the network load the users generate, and *(ii)* their mobility patterns. The model is evaluated with two unsupervised clustering algorithms of different complexity (namely, XMeans and EM) by means of a 3G trace dataset from a European operator.

[2] *Understanding Usage and Activity in Cellular Networks by Investigating HTTP Requests.* Fehmi Ben Abdesslem, Anders Lindgren, Andrea Hess. IEEE Consumer Communications & Networking Conference (CCNC) 2015. (to be published in IEEE Xplore Digital Library)

Abstract: The number of mobile devices is estimated to now exceed the world's population, using more and more cloud services, and hence generating more and more traffic. Smartphones generate 95% of the total global handset traffic, and while approximately half of this traffic is sent to cellular networks, other handsets such as tablets are also using increasingly the cellular networks. This paper provides a closer look at the traffic generated on cellular networks by exploring billions of HTTP requests sent by millions of users to a nation-wide cellular network during 41 days. We confirm that – as in many other contexts – 20% of the users are responsible for more than 80% of the requests and provide a deeper analysis of the cellular network usage. Furthermore, we characterise the activity of users on their mobile device and which cloud services they use. For instance, almost 30% of the users use the cellular network frequently, mainly using search services and social networks, but 20% of their requests are sent to advertisement and tracking systems.

[3] *Travel Demand Analysis with Differentially Private Releases.* David Gundlegard, Clas Rydergren, Jaume Barcelo, Nima Dokoohaki, Olof Görnerup, Andrea Hess. Contribution to the D4D Senegal challenge, accepted for presentation at NetMob 2015. (to be published on NetMob conference website)

Abstract: When analyzing large quantities of human mobility traces, the aspects of sensitivity of traces to be analyzed, and the scale at which such analysis can be accounted for is of high importance. The sensitivity implies that identifiable information must not be inferred from the data or any analysis of it. Thus, prompting the importance of maintaining privacy during or post-analysis stages. We aggregate the raw data with the goal to retain relevant information while at the same time discard sensitive user specifics, through site sequence clustering and frequent sequence extraction. These techniques have at least three benefits: data reduction, information mining, and anonymization. Further, the paper reviews the aggregation techniques with regard to privacy in a post-processing step. The approaches presented in the paper for estimation of travel demand and route choices, and the additional privacy analysis, build a comprehensive framework usable in the processing of mobile phone data for transportation planning.

[4] *Observing Software-defined Networks Using a Decentralized Link Monitoring Approach.* Rebecca Steinert, Andrea Hess. IEEE Conference on Network Softwarization (NetSoft) 2015. (to be published in IEEE Xplore Digital Library)

Abstract: Scalable and automated monitoring processes for testing, debugging, and operation of VNFs and service-chains are crucial components towards achieving the aims of network softwarization – i.e., cheaper, faster, and shorter service deployment and network management processes. In this paper we present a decentralized monitoring approach aimed at supporting automated deployment and operation of VNFs and service-chains. The approach is inspired by network tomography and is designed to address observability limitations and scalability issues that arise from performing measurements from an SDN controller. From successive end-to-end measurements link metrics are derived via in-network parameter estimation with no need of forwarding raw measurements to the controller, which significantly reduces the measurement overhead compared to when monitoring individual links explicitly from an SDN controller.

II.B. Pending publications

[5] *Caching Analysis and Control of Audiovisual Content in a 3G Network*. Ian Marsh, Nima Dokoohaki, Andrea Hess, Daniel Gillblad.

[6] *Scalable and Privacy-preserving Mining of Frequent Routes in Cellular Network Data.* Olof Görnerup, Andrea Hess, Nima Dokoohaki.

[7] *Data-driven Mobility Modeling: A Survey and Engineering Guidance for Mobile Networking.* Andrea Hess, Karin A. Hummel, Wilfried Gansterer, Günter Haring.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- June 4-5, 2014 – Project meeting on UNIFY (Unifying Cloud and Carrier Networks) in Stockholm, Sweden
- September 3-5, 2014 – Project meeting on UNIFY in Budapest, Hungary
- October 7/9, 2014 – SICS Cloud and Big Data Day / Smart Sustainable Cities Day in Kista, Sweden
- October 23-24, 2014 – 4th ERCIM ABCDE fellow seminar in Pisa, Italy
- March 5, 2015 – National workshop on "Mobile network data for traffic applications" organized by Sweco in Stockholm, Sweden
- *(planned)* June 2015 – Presentation of [1] at IEEE ICC 2015 in London, UK

IV – RESEARCH EXCHANGE PROGRAMME (REP)

One REP visit took place at INRIA Lille-Nord Europe in France, December 15-19, 2014. The visit was hosted by Dr. Nathalie Mitton, who leads the group on self-organizing Future Ubiquitous Networks (FUN). During the visit we had fruitful discussions about current research activities in mobile ad-hoc networking and application of the results in smart city systems, which broadened my view on research problems in the domain. Furthermore, I was introduced to the remotely accessible FIT IoT-LAB testbed that enables experiments of algorithms for large-scale wireless networks and seems to be suitable to evaluate parts of my current and previous work beyond network simulators.