Activity Report 2015

Team MUSE

Measuring networks for enhancing USer Experience

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Paris - Rocquencourt

THEME
Networks and Telecommunications

Table of contents

Team MUSE

Creation of the Team: 2013 October 03

Keywords:

Computer Science and Digital Science:

1.2.3. - Routing
1.2.4. - QoS, performance evaluation
1.2.5. - Internet of things
3.1.7. - Open data
3.1.8. - Big data (production, storage, transfer)
3.3. - Data and knowledge analysis
3.5. - Social networks

Other Research Topics and Application Domains:

6.3. - Network functions
6.4. - Internet of things
6.5. - Information systems

1. Members

Research Scientists

Renata Cruz Teixeira [Team leader, Inria, Senior Researcher, HdR]
Vassilis Christofidis [Inria, Advanced Research position, from Jul 2015]
Anna-Kaisa Pietilainen [Inria, Starting Research position]

Faculty Member

Timur Friedman [Univ. Paris VI, Associate Professor, on delegation]

Engineer

Michail Katsarakis [Inria, from Dec 2015]

PhD Students

Sara El Aouad [Inria/Technicolor, granted by CIFRE, from May 2014]
Diego Neves Da Hora [Inria/Technicolor, granted by CIFRE, from February 2014]
Giuseppe Scavo [Inria, granted by Alcatel-Lucent Bell Labs, from November 2013]
Stephane Wustner [Univ. Paris VI/Technicolor, granted by CIFRE, until Jan 2015]

Administrative Assistant

Anna Bednarik [Inria]

Others

Shirin Mohammad Yari [Politecnico di Torino, from Mar 2015 until Sep 2015]
Zachary Bischof [Nortwestern University, from Jun 2015 until Sep 2015]

2. Overall Objectives

2.1. Overall Objectives

Muse's research is broadly in the area of network measurements. We focus on developing new algorithms and systems to improve user experience online. In particular, we are addressing two main problems of today's Internet users:

1. Technology is too complex. Most Internet users are not tech-savvy and hence cannot fix performance problems and anomalous network behavior by themselves. The complexity of most Internet applications makes it hard even for networking experts to fully diagnose and fix problems. Users can't even know whether they are getting the Internet performance that they are paying their providers for.
2. There is too much content. The proliferation of user-generated content (produced anywhere with mobile devices and immediately published in social media) along with the vast amount of information produced by traditional media (e.g., newspapers, television, radio) poses new challenges in achieving an effective, near real-time information awareness and personalization. For instance, users need novel filtering and recommendation tools for helping them to decide which articles to read or which movie to watch.

3. Research Program

3.1. Active probing methods

We are developing methods that actively introduce probes in the network to discover properties of the connected devices and network segments. We are focusing in particular on methods to discover properties of home networks (connected devices and their types) and to distinguish if performance bottlenecks lie within the home network versus outside. Our goal is to develop adaptative methods that can leverage the collaboration of the set of available devices (including end-user devices and the home router, depending on which devices are running the measurement software).

3.2. Passive monitoring methods

This part our research develops methods that simply observe network traffic to infer the performance of networked applications and the location of performance bottlenecks, as well as to extract patterns of web content consumption. We are working on techniques to collect network traffic both at user's end-devices and at home routers. We also have access to network traffic traces collected on a campus network and on a large European broadband access provider.

3.3. Inferring user online experience

We are developing hybrid measurement methods that combine passive network measurement techniques to infer application performance with techniques from HCI to measure user perception. We will later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user's traffic observed in the network.

3.4. Content summarisation

We are working on methods to summarise a set of reviews (for example, movie reviews from Rotten Tomatoes or IMDB; or restaurant reviews from Yelp) with a set of representative tags. Each tag is a sequence of two or three words. In parallel, we are building a mobile app that allows users to directly enter tags instead of free-text reviews.

3.5. Filtering real-time Web streams

The Web is rapidly transformed into a real-time information system forcing us to revise both how to effectively assess relevance of information for a user and how to efficiently implement information retrieval or dissemination functionality. To capture various contextual aspects of user needs and information shared in the Real-time Web, besides content and social relevance, we also have to consider implicit (e.g. pageviews) or explicit user feedback (e.g., like, retweet or reply events). To accommodate high arrival rates of information items (e.g., 100 millions of tweets per day) and user events (e.g., billions of pageviews per day) we are exploring a publish/subscribe paradigm in which we index queries and update on the fly their results each time a new item and relevant events arrive. In this respect, we need to process continuous top-k text queries combining query-dependent (as text similarity) and query-independent (as social relevance or user attention) scores with time decay functions.

3.6. Flexible online drift detection

Monitoring streaming content is a challenging big data analytics problem, given that very large datasets are rarely (if ever) stationary. In several real world monitoring applications (e.g., newsgroup discussions, network connections, etc.) we need to detect significant change points in the underlying data distribution (e.g., frequency of words, sessions, etc.) and track the evolution of those changes over time. These change points, depending on the research community, are referred to as temporal evolution, non stationarity, or concept drift and provide valuable insights on real world events (e.g. a discussion topic, an intrusion) to take a timely action. In this paper, we adopt a query-based approach to drift detection and address the question of processing drift queries over very large datasets. To the best of our knowledge, our work is the first to formalize flexible drift queries on streaming datasets with varying change rates.

4. Application Domains

4.1. Home Network Diagnosis

With the availability of cheap broadband connectivity, Internet access from the home has become a ubiquity. Modern households host a multitude of networked devices, ranging from personal devices such as laptops and smartphones to printers and media centers. These devices connect among themselves and to the Internet via a local-area network—a *home network*— that has become an important part of the "Internet experience". In fact, ample anecdotal evidence suggests that the home network can cause a wide array of connectivity impediments, but their nature, prevalence, and significance remain largely unstudied.

Our long-term goal is to assist users with concrete indicators of the causes of potential problems and—ideally—ways to fix them. We intend to develop a set of easy-to-use home network diagnosis tools that can reliably identify performance and functionality shortcomings rooted in the home. The development of home network diagnosis tools brings a number of challenges. First, home networks are heterogenous. The set of devices, configurations, and applications in home networks vary significantly from one home to another. We must develop sophisticated techniques that can learn and adapt to any home network as well as to the level of expertise of the user. Second, there are numerous ways in which applications can fail or experience poor performance in home networks. Often there are a number of explanations for a given symptom. We must devise techniques that can identify the most likely cause(s) for a given problem from a set of possible causes. Third, even if we can identify the cause of the problem, we must then be able to identify a solution. It is important that the output of the diagnosis tools we build is "actionable". Users should understand the output and know what to do.

We are conceiving methods for two application scenarios: (i) when the end user in the home deploys our diagnostic tools either on the home gateway (the gateway often combines a DSL/cable modem and an access point; it connects the home network to the ISP) or on devices connected to the home network and (ii) when ISPs collect measurements from homes of subscribers and then correlate these measurements to help identify problems.

Assisting end users. We are developing algorithms to determine whether network performance problems lie inside or outside the home network. Given that the home gateway connects the home with the rest of the Internet, we are designing an algorithm (called *HoA*) that analyzes traffic that traverses the gateway to distinguish access link and home network bottlenecks. A measurement vantage point on the gateway is key for determining if the performance bottleneck lies within the home network or the access ISP, but we also need to deploy diagnosis tools in end-devices. First, some users may not want (or not know how) to deploy a new home gateway in their homes. Second, some problems will be hard to diagnose with only the vantage point of the gateway (for example, when a device cannot send traffic or when the wireless is poor in certain locations of a home). We can obtain more complete visibility by leveraging *multiple* measurement nodes around the home, potentially including the home gateway, all participating jointly in the measurement task. We have an ongoing project to realize a home network analyzer as a web-based measurement application built on top of our team's recently developed browser-based measurement platform, *Fathom*. To integrate the home gateway in the analyzer, we plan to engage the BISmark Project. BISmark already provides a web server as well as extensive configurability, allowing us to experiment freely with both passive as well as active measurements. We must develop a home network analyzer that can first discover the set of devices connected to the home network that can collaborate on the diagnosis task. We will then develop tomography algorithms to infer where performance problems lie given measurements taken from the set of available vantage points.

Assisting Internet Service Providers (ISPs). Our discussions with several large access ISPs reveal that service calls are costly, ranging from $9–25 per call, and as many as 75% of service calls from customers are usually caused by problems that have nothing to do with the ISP. Therefore, ISPs are eager to deploy techniques to assist in home network diagnosis. In many countries ISPs control the home gateway and set-top-boxes in the home. We plan to develop more efficient mechanisms for home users to report trouble to their home ISP and consequently reduce the cost of service calls. This project is in collaboration with Technicolor and Portugal Telecom. Technicolor is a large manufacturer of home gateways and set-top-boxes. Portugal Telecom is the largest broadband access provider in Portugal. Technicolor already collects data from 200 homes in Portugal. We are working with the data collected in this deployment together with controlled experiments to develop methods to diagnose problems in the home wireless.

4.2. Quality of Experience

Understanding how users react to different levels of network performance presents two main challenges:

1. User perception is subjective and contextual. Different users may have different tolerance levels to network performance and the same user may have different expectations under different circumstances. Take for example the round-trip time (RTT), a typical network performance metric. If RTTs are larger than usual, a user who is doing remote login may feel that the connection is unusable, whereas another who is watching YouTube may notice no problem (because YouTube has a playout buffer to mask some network delay). Take another example of a user downloading her email. This user may tolerate some delay when she is leisurely checking her email at home, but she may become extremely frustrated with the same delay if she is in an airplane and needs to download her email just before takeoff.
2. It is challenging to "measure" users. We must develop methods to measure the user perception of network performance as users perform their routine online tasks. It is hence important that these methods are not too intrusive. Otherwise, users are unlikely to participate in the experiment. In addition, we must capture user perception at different levels of performance and in a variety of scenarios.

We will develop tools that run on end systems to collect network performance data annotated with the user perception. These tools will adopt a hybrid measurement methodology that combines network measurement techniques to infer application performance with techniques from HCI to measure user perception. We will later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user's traffic observed in the network. Models of user perception of network performance can be used to detect when performance is poor to trigger diagnosis or to adapt network/application performance to better serve users.

4.3. Crowd-sourced content recommendation

The Internet today serves as a large content distribution platform (online content varies from traditional news, TV series, and movies to specialized blogs and family pictures shared over social networks) as well as a platform for users to exchange opinions about practically everything (from movies to services and restaurants). The amount of information available online today overwhelms most users and selecting which content to watch or what do has become a challenge. We are applying passive measurement methods and content summarisation techniques to help users to identify relevant content in two scenarios. First, we are developing a system called WeBrowse that passively observes network traffic to extract user clicks (i.e., the URLs users visit). A user click is a good measure of interest, as users often have an idea of the type of content they are about to access (e.g., because they saw a preview or because a friend recommended it). Intuitively, the more users click on a URL, the higher the interest in the content on the corresponding page. WeBrowse then promotes "hottest" and most popular content to users of a network. We have a deployment of WeBrowse in a campus network. Second, we are working on techniques to summarise user feedback (for example, movie or restaurant reviews) with semi-structured feedback. Today reviews are either free-form text or star rating. Star rating is too coarse to capture the nuances of why a user likes or dislikes something, whereas free text is hard for users to parse and extract a clear opinion. We are instead working with semi-structured reviewing where users enter *tags* (a short sequence of words describing the user experience). We are working with Technicolor on the summarisation of movie reviews and on building a mobile app (called TagIt) where users can review movies directly with tags.

5. New Software and Platforms

5.1. Fathom

Fathom - browser-based network measurement platform
KEYWORDS: Internet access - Performance measure - Network monitoring
FUNCTIONAL DESCRIPTION

Fathom is a Firefox browser extension that explores the browser as a platform for network measurement and troubleshooting. It provides a wide range of networking primitives directly to in-page JavaScript including raw TCP/UDP sockets, higher-level protocol APIs such as DNS, HTTP, and UPnP, and ready-made functionality such as pings and traceroutes.

- Participants: Anna-Kaisa Pietilainen and Stéphane Archer
- Contact: Anna-Kaisa Pietilainen
- URL: https://muse.inria.fr/fathom/

5.2. HostView

FUNCTIONAL DESCRIPTION

End-host performance monitoring and user feedback reporting

- Participants: George Rosca, Anna-Kaisa Pietilainen and Renata Cruz Teixeira
- Contact: Renata Cruz Teixeira
- URL: https://muse.inria.fr/hostview/

5.3. Online HoA

Online implementation of home and access throughput bottleneck detection algorithm 'HoA'
FUNCTIONAL DESCRIPTION

Implementation of HoA as collectd plugin for OpenWRT.

- Contact: Renata Cruz Teixeira
- URL: https://github.com/inria-muse/browserlab

5.4. SimilarityExplanation

Prototype implementation for explaining a set of similar and recommended movies.
FUNCTIONAL DESCRIPTION

In this web-based prototype for similar movies explanation, we propose two types of browsing for : personalized browsing and non personalized browsing. In the non personalized browsing we suppose that we don't have the user profile. Similar movie sublists are ordered only according to their similarity to the selected movie. For the personalized browsing , we select users that have different profiles from our dataset. We give these users names of actors, according to the types of movies they watch. For each user, we compute the predicted ratings using the matrix factorization model. We select pairs of genres to display to each user based on the preferred genres for the user. In our prototype we identify the preferred genres per user based on the most frequent movie genre pairs that the user has already seen. We then organize the recommended movies with a high rating prediction in sublists, according to the user most preferred genre pairs. When a user selects a movie from the sublists of recommended movies, our application suggests the similar movies presented under four sublists with the added list of words. The sublists are personalized for each user by reordering the movies according to the users predicted ratings.

- Contact: Sara El Aouad
- URL: http://muse.inria.fr/tagit

5.5. UCN

User-Centric Networking
FUNCTIONAL DESCRIPTION

The User-Centric Networking (UCN) project is seeking to understand how people consume various kinds of content when using computer networks. Within this project we are undertaking a detailed user study across a range of environments in order to understand the practices involved in consuming media and other content according to context.

- Participants: Renata Cruz Teixeira and Anna-Kaisa Pietilainen
- Contact: Anna-Kaisa Pietilainen
- URL: https://muse.inria.fr/ucn

5.6. WeBrowse

FUNCTIONAL DESCRIPTION

WeBrowse is the first passive crowdsource-based content curation system. Content curation is the act of assisting users to identify relevant and interesting content in the Internet. WeBrowse requires no active user engagement to promote content. Instead, it extracts the URLs users visit from traffic traversing an ISP network to identify popular content. WeBrowse contains a set of heuristics to identify the set of URLs users visit and to select the subset that are interesting to users.

- Contact: Giuseppe Scavo
- URL: http://webrowse.polito.it/

6. New Results

6.1. Home Network or Access Link? Locating Last-mile Downstream Throughput Bottlenecks

Participants: Srikanth Sundaresan (ICSI), Nick Feamster (Princeton), Renata Teixeira

As home networks see increasingly faster downstream throughput speeds, a natural question is whether users are benefiting from these faster speeds or simply facing performance bottlenecks in their own home networks. In our paper recently accepted for publication in PAM'16, we studied whether downstream throughput bottlenecks occur more frequently in their home networks or in their access ISPs. We identified lightweight metrics that can accurately identify whether a throughput bottleneck lies inside or outside a user's home network and developed a detection algorithm that locates these bottlenecks. We validated this algorithm in controlled settings and characterized bottlenecks on two deployments, one of which included 2,652 homes across the United States. We found that wireless bottlenecks are more common than access-link bottlenecks—particularly for home networks with downstream throughput greater than 20 Mbps, where access-link bottlenecks are relatively rare.

6.2. On the Reliability of Profile Matching Across Large Online Social Networks

Participants: Oana Goga and Krishna Gummadi (MPI-SWS), Patrick Loiseau (EURECOM), Robin Sommer (ICSI), Renata Teixeira

Matching the profiles of a user across multiple online social networks brings opportunities for new services and applications as well as new insights on user online behavior, yet it raises serious privacy concerns. Prior literature has showed that it is possible to accurately match profiles, but their evaluation focused only on sampled datasets. In our KDD'15 paper [2], we study the extent to which we can reliably match profiles in practice, across real-world social networks, by exploiting public attributes, i.e., information users publicly provide about themselves. Today's social networks have hundreds of millions of users, which brings completely new challenges as a reliable matching scheme must identify the correct matching profile out of the millions of possible profiles. We first define a set of properties for profile attributes–Availability, Consistency, non-Impersonability, and Discriminability (ACID)–that are both necessary and sufficient to determine the reliability of a matching scheme. Using these properties, we propose a method to evaluate the accuracy of matching schemes in real practical cases. Our results show that the accuracy in practice is significantly lower than the one reported in prior literature. When considering entire social networks, there is a non-negligible number of profiles that belong to different users but have similar attributes, which leads to many false matches. Our paper sheds light on the limits of matching profiles in the real world and illustrates the correct methodology to evaluate matching schemes in realistic scenarios.

6.3. Exploiting crowd sourced reviews to explain movie recommendation

Participants: Sara El Aouad, Christophe Dupuy, Francis Bach, and Renata Teixeira (Inria), Christophe Diot (Technicolor)

Streaming services such as Netflix, M-Go, and Hulu use advanced recommender systems to help their customers identify relevant content quickly and easily. These recommenders display the list of recommended movies organized in sublists labeled with the genre or some more specific labels. Unfortunately, existing methods to extract these labeled sublists require human annotators to manually label movies, which is time-consuming and biased by the views of annotators. In our work [6], we design a method that relies on crowd sourced reviews to automatically identify groups of similar movies and label these groups. Our method takes the content of movie reviews available online as input for an algorithm based on Latent Dirichlet Allocation (LDA) that identifies groups of similar movies. We separate the set of similar movies that share the same combination of genre in sublists and personalize the movies to show in each sublist using matrix factorization. The results of a side-by-side comparison of our method against Technicolor's M-Go VoD service are encouraging.

6.4. Characterizing Home Device Usage From Wireless Traffic Time Series

Participants: Katsiaryna Mirylenka (IBM Research - Zurich), Vassilis Christophides, Themis Palpanas (Paris Descartes University), Ioannis Pefkianakis (Hewlett Packard Labs), Martin May (Technicolor).

The analysis of *temporal behavioral patterns* of home network users can reveal important information to Internet Service Providers (ISPs) and help them to optimize their networks and offer new services (e.g., remote software upgrades, troubleshooting, energy savings). Our study [4] uses time series analysis of continuous traffic data from wireless home networks, to extract traffic patterns recurring within, or across homes, and *assess the impact of different device types (fixed or portable) on home traffic*. Traditional techniques for time series analysis are not suited in this respect, due to the limited stationary and evolving distribution properties of wireless home traffic data. We propose a novel framework that relies on a *correlation-based similarity* measure of time series, as well as a notion of *strong stationarity* to define recurring motifs and dominant devices. Using this framework, we analyze the wireless traffic collected from 196 home gateways over two months. Our framework goes beyond existing application-specific analysis techniques, such as analysis of wireless traffic, which mainly rely on data aggregated across hundreds, or thousands of users. It enables the extraction of recurring patterns from traffic time series of individual homes, leading to a much more fine-grained analysis of the behavior patterns of the users. We also determine the best time aggregation policy w.r.t. to the number and statistical importance of the extracted motifs, as well as the device types dominating these motifs and the overall gateway traffic. Our results show that ISPs can exceed the simple observation of the aggregated gateway traffic and better understand their networks.

6.5. On Continuous Top-k Queries with Real-Time Scoring Functions

Participants: Nelly Vouzoukidou (Google, France), Bernd Amann (LIP6), Vassilis Christophides.

Modern news sharing and social media platforms allow millions of users to *produce and consume information in real-time*. To assess relevance of published information in this new setting, batch scoring based on content similarity, link centrality or page views is no longer sufficient. Instead, streams of events like “replies” (for posting comments), “likes” (for rating content) or “retweets” (for diffusing information) explicitly provided by users represent valuable online feedback on published information that has to be exploited in order to adjust in real-time any available score of information items. Note that in the future Internet of Things (IoT), not only digital, but also physical objects will be expected to be ranked in a fully automated way with respect to real-time human activities (viewing concentration), vital signals (emotional arousal), etc.

Rather than indexing as quickly as possible information items to re-evaluate *snapshot queries*, publish/subscribe systems index *continuous queries* and update on the fly their results each time a new matching item arrives. Existing publish/subscribe systems rely on two alternative continuous filetring semantics, namely *predicate-based* filtering or *similarity-based top-k* filtering. In predicate-based systems, incoming items that match the filtering predicates are simply added to the result list of continuous queries, while in similarity-based top-k publish/subscribe systems, matching items have also to exhibit better relevance w.r.t. the items already appearing as the top-k results of the continuous query. In top-k publish/subscribe systems the relevance of an item remains constant during a pre-specified time window, and once its lifetime exceeds the item simply expires. Only recently, information recency has become part of the relevance score of continuous queries. Clearly, when information relevance decays as time passes both (a) results lists maintenance and (b) early pruning of the query index traversal are challenged. While these problems have been studied for (textual or spatio-textual) content scoring functions with time decay, non-homogeneous scoring functions accommodating various forms of *query-dependent* and *query-independent* information relevance with time decay is supported only by MeowsReader. In this work we are going beyond this general form of *time-decayed static scores* and consider continuous queries featuring *real-time scoring functions* under the form of *time decaying positive user feedback* for millions of online social media events per minute and millions of user queries.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- “Improving the quality of recommendation using semi-structured user feedback” CIFRE contract with Technicolor for thesis of Sara el Aouad from May 2014 to April 2017.

- "Crowdsourced Home Network Diagnosis" CIFRE contract with Technicolor for thesis of Diego da Hora from February 2014 to January 2017.
- "Exploiting Network Content-awareness to provide novel added value services" contract under the Inria-Alcatel Lucent Bell Labs common Lab (ADR ICN) to fund the doctoral thesis of Giuseppe Scavo from November 2013 to October 2016.

7.2. Bilateral Grants with Industry

- "Collaborative Home Network Troubleshooting", Comcast grant, from December 2015.

8. Partnerships and Cooperations

8.1. European Initiatives

8.1.1. User-Centric Networking (UCN)

Type: FP7

Instrument: Specific Targeted Research Project

Duration: October 2013 - September 2016

Coordinator: Technicolor

Other partners: Eurecom, Fraunhofer FOKUS, Intamac, University of Cambridge, University of Nottingham, Martel, NICTA, Portugal Telecom

Inria contact: Renata Teixeira

Abstract: This project introduces the concept of User Centric Networking (UCN), which is a new paradigm leveraging user information at large to deliver novel content recommendation systems and content delivery frameworks. UCN recommendation and content delivery systems will leverage in-depth knowledge about users to help them find relevant content, identify nearby network resources and plan how to deliver the actual content to the appropriate device at the desired time. These systems will additionally account for influences from users' social networks on their content consumption. The goal of this project is to design a UCN system architecture for user-centric connected media services. We will build UCN upon three complementary research pillars:

1. understanding user context: This data can be broadly categorized into three groups. First, the physical and environmental context A second category of data is that which can be extracted from social network interactions. The third category of data is behavioural
2. profiling and predicting user interests: By gaining a deep understanding of the user, we may be able to cast a much wider net in the content ocean and locate a richer catalogue of interesting content for the user
3. personalizing content delivery: Rather than the user (or the service provider) having to worry about the mode of connectivity, device, service, location, etc., the network intelligently directs and adapts the transport stream, or perhaps pre-fetches and replicates content chunks, to the particular and immediate needs of the user.

See also: http://usercentricnetworking.eu/

8.2. International Initiatives

8.2.1. Inria International Partners

8.2.1.1. Informal International Partners

- Princeton (Prof. Nick Feamster): We have a long-term collaboration on measuring the performance of residential broadband Internet access networks and more recently on home network diagnosis.

- ICSI (Dr. Srikanth Sundaresan, Dr. Christian Kreibich, Dr. Robin Sommer): With C. Kreibich, we have been developing Fathom, a browser-based network measurement platform. We are now adding home network diagnosis capabilities to Fathom. We are collaborating with S. Sundaresan on detecting last-mile bottlenecks. In addition, with Robin Sommer we are working on the potential of matching the profiles of a user across multiple online social networks.
- Northwestern University (Prof. Fabian Bustamante and his doctoral student Zachary Bischof): we are working on identifying user activity from network traffic.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

8.3.1.1. Internships

- S. Mohammadyari, master intern, Politecnico di Torino, Italy, March to July 2015.
- Zachary Bischof, doctoral intern, Northwestern University, USA, July to September 2015.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

- Program chair of the 5th PhD School on Traffic Monitoring and Analysis (TMA), 2015.

9.1.1.2. Member of the organizing committees

- R. Teixeira, Workshop co-chair of ACM SIGCOMM 2015.

9.1.2. Scientific events selection

9.1.2.1. Chair of conference program committees

- R. Teixeira, Co-chair of the program committee of ACM CoNEXT 2015.
- V. Christophides, Area Chair for "Semi-structured, Web, and Linked Data Management" in the 18th International Conference on Data Engineering (ICDE'16) , Bali, Indonesia 2016 (internationalscienceindex.org/event/2016/10/bali/ ICDE)
- V. Christophides, Workshop co-chair of the 2nd International Workshop on Preservation of Evolving Big Data, co-located with EDBT 2016 with G. Papastefanatos and S. Viglas March 15, 2016 Bordeaux, France (www.diachron-fp7.eu/2nd-diachron-workshop.html)

9.1.2.2. Member of the conference program committees

- R. Teixeira, ACM IMC 2015
- V. Christophides, Int'l Conference on Information and Knowledge Management (CIKM) 2015
- V. Christophides, Int'l Conference on Extending Database Technology (EDBT) 2016

9.1.2.3. Reviewer

- A.-K. Pietilainen, CHI 2016

9.1.3. Journal

9.1.3.1. Member of the editorial boards

- R. Teixeira, Editor of IEEE/ACM Transactions on Networking (Feb. 2011 – Feb. 2015)

9.1.4. Invited talks

- R. Teixeira, “HostView: Measuring Internet quality of experience on end-hosts”, Journée thématique - Métrologie des réseaux - de la DGA, Rennes, November 2015.
- R. Teixeira, “UCN user study”, panelist at ACM SIGCOMM workshop on crowdsourcing and crowdsharing of (big) Internet data, London, August 2015.
- R. Teixeira, “HostView: Measuring Internet quality of experience on end-hosts”, Princeton University, USA, July 2015.
- R. Teixeira, “HostView: Measuring Internet quality of experience on end-hosts”, Journée du Conseil scientifique de l’Afnic, Paris, July 2015.
- R. Teixeira, “Home network diagnosis”, Workshop of “Aux frontières du système” of the “Conférence d’informatique en Parallélisme, Architecture et Système”, Lille, June, 2015.
- R. Teixeira, “Measuring Internet Experience from Home Networks", invited talk, Workshop on Active Internet Measurements, San Diego, USA, April, 2015.
- R. Teixeira, “Home network performance diagnosis”, Microsoft Research Devices and Networking Summit, Paris, May 2015.
- R. Teixeira, “Home network performance diagnosis”, workshop on Information and Communication Systems and their application to vertical sectors, Montevideo, Uruguay, March 2015.
- R. Teixeira, “Fathom: A Browser-based Network Measurement Platform”, invited talk at Federal University of Rio de Janeiro, Brazil, March 2015.
-

9.1.5. Leadership within the scientific community

- R. Teixeira, Vice-chair of ACM SIGCOMM (since Jun. 2013).
- R. Teixeira, chair of the ACM SIGCOMM Industrial Liaison Board.
- R. Teixeira, member of the steering committee of the ACM Internet Measurement Conference (since Nov. 2009)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: Renata Teixeira, “Methodology for research in networking”, 10h eqTD, M2, UPMC, France.

Master: Timur Friedman, Renata Teixeira, “Network Metrology”, 24h CM, M2, UPMC, France. Created this new master’s level class.

E-learning

Mooc: Timur Friedman, Renata Teixeira, “Network Metrology”, preparation of a five-week Mooc to start in May 2016 in the platform FUN, supported by Inria.

9.2.2. Supervision

PhD: Nelly Vouzoukidou (October 2011-September 2015): “Continuous top-k Queries over Real-time Web Streams” University Pierre et Marie Curie (UPMC). Advisors: B. Amann and V. Christophides

PhD in progress:

- Sara el Aouad, “Improving the quality of recommendation using semi-structured user feedback”, May 2014, advisors: C. Diot and R. Teixeira
- Diego da Hora, “Crowdsourced Home Network Diagnosis”, February 2014, advisors: C. Diot and R. Teixeira

- Giuseppe Scavo, “Exploiting Network Content-awareness to provide novel added value services”, November 2013 advisors: Z. Ben-Houidi and R. Teixeira
- Sofia Kleisarchaki (October 2013-now): “Temporal Analytics of the Social Web” cotutelle between University of Crete & University of Joseph Fourier - Grenoble. Advisors: S. Amer-Yahia and V. Christophides)
- Vassilis Efthimiou (February 2013-now): “Entity resolution in the Web of Data” University of Crete. Advisor: V. Christophides

9.2.3. Juries

V. Christophides: member of the supervision committee of the Ph.D. Thesis of Charalampos Nikolaou, Univ. of Athens, 2015

10. Bibliography

Publications of the year

International Conferences with Proceedings

[1] V. EFTHYMIOU, K. STEFANIDIS, V. CHRISTOPHIDES. *Big Data Entity Resolution:*, in "2015 IEEE International Conference on Big Data (IEEE BigData 2015)", Santa Clara, CA,, United States, October 2015 [*DOI :* 10.1109/BIGDATA.2015.7363781], https://hal.inria.fr/hal-01199399

[2] O. GOGA, P. LOISEAU, R. SOMMER, R. TEIXEIRA, K. P. GUMMADI. *On the Reliability of Profile Matching Across Large Online Social Networks*, in "KDD'15: ACM SIGDD Conference on Knowledge Discovery and Data Mining", Sydeny, Australia, August 2015 [*DOI :* 10.1145/2783258.2788601], https://hal.inria.fr/hal-01162402

[3] S. KLEISARCHAKI, S. AMER-YAHIA, A. DOUZAL-CHOUAKRIA, V. CHRISTOPHIDES. *Querying Temporal Drifts at Multiple Granularities*, in "CIKM 2015", Melbourne, Australia, October 2015, https://hal.archives-ouvertes.fr/hal-01208397

[4] K. MIRYLENKA, V. CHRISTOPHIDES, T. PALPANAS, I. PEFKIANAKIS, M. MAY. *Characterizing Home Device Usage From Wireless Traffic Time Series*, in "19th International Conference on Extending Database Technology (EDBT)", Bordeaux, France, March 2016, https://hal.inria.fr/hal-01249778

[5] S. SUNDARESAN, N. FEAMSTER, R. TEIXEIRA. *Measuring the Performance of User Traffic in Home Wireless Networks*, in "Passive and Active Network Measurement Conference", New York, United States, Lecture notes in computer science, March 2015, https://hal.inria.fr/hal-01100675

Conferences without Proceedings

[6] S. EL AOUAD, C. DUPUY, R. TEIXEIRA, C. DIOT, F. BACH. *Exploiting crowd sourced reviews to explain movie recommendation*, in "2nd Workshop on Recommendation Systems for Television and Online Video", Vienna, Austria, September 2015, https://hal.inria.fr/hal-01193308

Scientific Books (or Scientific Book chapters)

[7] V. CHRISTOPHIDES, V. EFTHYMIOU, K. STEFANIDIS. *Entity Resolution in the Web of Data*, Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool, August 2015, vol. 5, n^{o} 3, pp. 1-122 [*DOI :* 10.2200/S00655ED1V01Y201507WBE013], https://hal.inria.fr/hal-01191691

[8] V. CHRISTOPHIDES, T. PALPANAS. *Report on the First International Workshop on Personal Data Analytics in the Internet of Things (PDA@IOT 2014)*, SIGMOD Record, ACM, March 2015, vol. 44, n^{o} 1 [*DOI :* 10.1145/2783888.2783905], https://hal.inria.fr/hal-01253709

Other Publications

[9] S. KLEISARCHAKI, S. AMER-YAHIA, A. DOUZAL-CHOUAKRIA, V. CHRISTOPHIDES. *Querying Temporal Drifts at Multiple Granularities*, September 2015, working paper or preprint, https://hal.inria.fr/hal-01182742