Homebase: Developing a Corpus of Domestic Network Usage

Tom Rodden

Computing Department

Nottingham University

Jubilee Campus

Nottingham NG8 1BB

tar@cs.nott.ac.uk

Paul Tennent

Computing Department

Nottingham University

Jubilee Campus

Nottingham NG8 1BB

pxt@cs.nott.ac.uk

Matthew Chalmers

Computing Science Department

Glasgow University

matthew@dcs.gla.ac.uk

Beki Grinter

School of Interactive Computing

College of Computing,

Georgia Institute of Technology

Keith Edwards

School of Interactive Computing

College of Computing,

Georgia Institute of Technology

Andrew Crabtree

Computing Department

Nottingham University

Jubilee Campus

Nottingham NG8 1BB

tar@cs.nott.ac.uk

Abstract

We present a new collaboration aiming to develop strategies for the capture, filtration, representation, and analysis of home network data leading to the redesign of home network infrastructure. A central feature of this is the development of a corpus of digital domestic data based on a combination of network traffic monitoring and ethnographic studies situated in domestic environments. A significant portion of the evolving corpus will be stored, managed, filtered, represented and explored using the Digital Replay System.

Keywords

Home Networks, Corpora, Replay, Domestic Studies

Introduction

Domestic environments have been a source of inspiration for future digital technologies for some time. We like many other researchers are interesting in gaining insight into the nature of domestic environments through a range of studies and exploiting these to realize future technologies. A growing area of interest has being understanding the work that inhabitants undertake in order to manage and maintain the underlying infrastructures, in order to design a future generation of home networks.

To this end, we are about to undertake a series of studies of domestic networks and the everyday activities inhabitants undertake in order to understand these networks and their roles within these activities. We seek to understand these activities from multiple perspectives: users' activities in the home, systems effects in the computer in the home, and the traffic in the underlying networks. As part of the project we will construct a corpus constituted of both system-oriented network measurement data, and more human oriented ethnographic data accounting the usage and integration of those networks in everyday domestic life.

Network Data Capture

The homework project will exploit a series of techniques to capture underlying network data. This will include the capture of data within the domestic router and the development of a measure plane akin to existing internet measurement planes. These approaches will result in the collection of a broad range of systems and traffic data. These will normally take the form of a series of time stamped system logs.

Empirical Studies of Home Networks

As well as collecting systems and network data our corpus will include people oriented data captured in the domestic environment by researchers including video, audio and field notes. The project team has a strong background in conducting ethnographic studies and particular experience of making empirical studies of domestic network management techniques. It is our intention that the combination of these heterogeneous data types will allow a deeper exploration of the management and integration of home networks into

everyday life, and support the process of conducting further study into this area.

Challenges

Beyond the technical challenges of data capture, the key issues we face as a project that we would like to discuss at this workshop include:

Filtration – the sheer volume of data generated by capturing network traffic will require techniques to filter irrelevant data, and to allow levels abstraction across tasks and transactions. Additionally the subjects of a study must be able to filter recorded data to remove any captured information deemed private.

Representation and interaction – information visualization techniques are needed to support the analysis and use of the data. Approaches must be found to make understandable, accountable and relevant representations of network traffic at various levels of abstraction—for researchers' and for inhabitants' varied purposes.

Storage and Synchronization – The captured data must be sensibly stored and synchronized. As the data will come from several sources and in several formats, a suitable technique for achieving synchronization must be realized.

Digital Replay System

To address the challenges described above, one of our key objectives is to develop generic tools to support the recording and replaying of captured ubiquitous computing data. The heterogeneous nature of this data makes presentation a significant challenge, so to that end we intend to make use of the facilities provided by

the Digital Replay System (DRS), developed at the University of Nottingham's Digital Records For e-Social Science research node to support the storage, filtration, representation and exploration of data in ubiquitous computing settings.
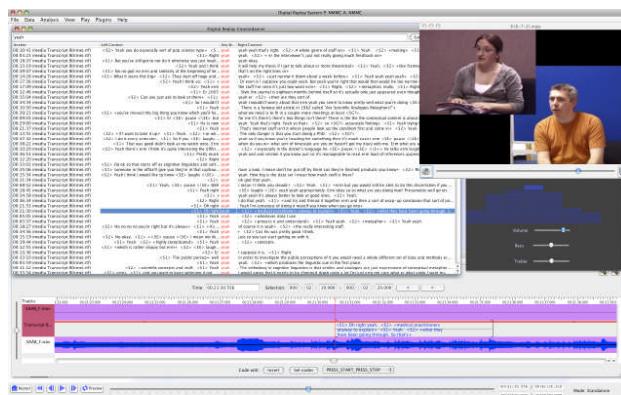

Figure 1 – Screenshot of the Digital Replay System showing synchronized audio, video and textual data.

DRS is a cross-platform multi-user qualitative analysis toolkit developed to support the analysis of heterogeneous media types. DRS treats heterogeneous data types with equal prominence, that is, no one media type is dominant. It supports the use of video, audio, images, text, annotations, coding and system log data in a similar fashion, presenting them all in a unified and synchronized manner and thus allowing any one to serve as a cue for any other. A researcher may therefore begin her exploration of a dataset from any one of these areas.

DRS is defined by two major areas: storage and display. As a data repository, DRS uses an RDF based ontology to allow rich metadata describing media type, synchronization information, annotations and much more to be applied to a particular dataset. Additionally DRS supports the construction of databases from system logs. A tool called the logfile workbench allows a user to specify the layout of a log file, apply appropriate metadata (for example defining the layout of time stamps), then parse that file to construct a database which can then be viewed synchronously with other data types.

DRS displays data in a series of synchronized viewers. Any given media type can have one or more associated viewers, which can coexist to support the display of several media items simultaneously. For example, video has a synchronized player allowing the user to view several synchronized videos. This can be combined with a transcription viewer which shows a textual transcript of the video. The effect of the synchronization allows the analyst to search for and select a particular utterance and the video will automatically jump to that point.

The viewer system becomes particularly powerful when exploring system log data. Several different simple graph types are available, allowing data to be examined alongside synchronized videos that offer a context for that data. For example an analyst wishing to examine a particular remote access to a configuration interface could isolate and examine the interaction, and use video recorded as the configuration instructions were being sent (and synchronized within DRS) to see how that interaction with domestic infrastructure was managed in context.

DRS is now publicly available to download through DReSS node at Nottingham and has been used in a number of studies of ubiquitous computing experiences (http://web.mac.com/andy.crabtree/NCeSS_Digital_Records_Node/Welcome.html). Source code for DRS is available under the BSD license via sourceforge (http://sourceforge.net/projects/thedrs).

Legacy Data

One area of interest for us is the possibility of making the facilities available in DRS more broadly available to the community interested in studying domestic settings. One goal for participation in this workshop is to explore how we might exploit DRS to increase the availability of domestic data, and whether existing data can be incorporated into an emerging DRS-supported corpus to enhance the applicability of our data.

Conclusions

At this early stage in the homework project we hope to use this workshop to present initial data collected as part of our study of domestic networks. With the advantages offered by the existing Digital Replay System, we hope to show that a rich multimedia corpus taking both system *and* user oriented data and using each to give context to the other will allow us to develop a deeper understanding of the emerging management processes that people use to integrate the increasing amount of varied interacting technologies permeating the domestic environment.