Measuring Mobile User Behavior and Service Usage: Methods, Measurement Points, and Future Outlook

ANTERO KIVI
Helsinki University of Technology, Networking Laboratory
P.O. Box 3000 FI-02015, FINLAND
antero.kivi(at)tkk.fi

Abstract
Having a holistic view on mobile user behavior will be increasingly difficult in the future where a fragmented user base is using multiple wireless access networks. To improve understanding on user behavior and mobile service usage, traditional survey based data collection can be complemented with a range of automated measurement methods. In this paper, the alternative data collection methods are analyzed based on a literature study and practical experiences.

Surveys and handset monitoring provide very detailed sample based data, whereas wireless access network traffic and usage accounting data is less granular but extends up to the service provider’s entire subscriber base. Server measurements give detailed data on the usage of a fairly focused user base. In the future, multi-functional mobile handsets and the servers of large Internet based service providers seem the most promising places to measure user behavior and service usage, whereas 2G/3G mobile networks will likely be less useful as traffic partly leaks to other access networks.

Keywords: mobile user behavior, service usage, traffic measurement, handset monitoring

1 Introduction
The number of people adopting mobile phones is surging globally, and accessing the Internet using advanced mobile handsets and mobile usage of laptop computers is becoming more common. The role of Internet Protocol (IP) as the platform for future mobile services is being established, while processing power and memory capacities of mobile devices and servers are also rapidly increasing. New wireless access technologies, such as WiMAX, are emerging as alternatives to existing 2G/3G mobile network technologies, and the number and coverage of WLAN hot spots is expanding. The emergence of multi-radio terminals featuring multiple radio interfaces is supporting this trend. Meanwhile, the performance of GSM/UMTS networks is increasing with the HSPA technologies, keeping 2G/3G networks as a viable alternative for wireless Internet access.

In this heterogeneous environment where both licensed and unlicensed radio spectrum is used and the network has no centralized point, understanding the system as a whole is hard. As traffic diverges to multiple networks managed by multiple access providers, having a holistic view on user behavior becomes increasingly difficult. Reliable and transparent information on true mobile service usage is of value to many stakeholders in the industry, including marketing, business development, product development, and network planning and management. Mobile user behavior is also an increasingly relevant academic research topic (see e.g. [1]). However, studies using “real” or “empiric” data seem to have very few common elements. The fundamental differences in the data collection methods must be understood to be able to position research in proper context. The purpose of this study is to clarify these differences. Helsinki University of Technology, in cooperation with industry actors, has conducted several mobile service usage measurements in Finland since 2005. This paper reflects the learnings of these measurements.

In this paper, the alternative methods for collecting information on mobile user behavior and service usage are presented and compared, after which the future outlook of each method is considered.
2 Methods for collecting data on mobile user behavior and service usage

Usage data can be obtained with many alternative methods from numerous sources (see Figure 1). The most straightforward way available to all researchers is to do a survey or a panel study on a sample of real end-users. Actual usage can also be measured directly by monitoring a sample of end-user devices, by conducting traffic measurements at the wireless access network and the core network, or by logging usage at various servers. In addition, access or service provider usage accounting systems are also a good source of information. Secondary sources can also provide explanatory information on mobile service usage. These include expert interviews, mobile industry actor quarterly/annual reports, and reports from electronics retailers/wholesalers or other related organizations (e.g. [2], [3], [4] in Finland).

Figure 1. Sources of data on mobile user behavior and service usage

What actually constitutes “mobile” usage is quite ambiguous. A user can be asked about his perceived behavior related to mobile service usage, but in a more automated data collection context accurately identifying what usage is “mobile” is complicated. True mobility can be measured relative to the wireless access network base stations, or with the means of a separate positioning system (e.g. GPS). In GSM/UMTS networks, movement of mobile terminals can be detected by observing the wireless signal between the terminal and nearby base stations, or with the changes in network cell identity the terminal is using. However, logging base station level data in a 2G/3G mobile network might not be reasonable due to its high resource requirements. In smaller WLAN networks, observing mobility at base station (wireless access point) level is more feasible with typical network management methods (e.g. syslog data or SNMP querying [5]). Data on true movement, however, is seldom easily attainable. Thus, usage of a mobile device, mobile access network, or service intended for mobile devices or for a mobile context can be used as proxies for mobility.

2.1 Surveys and consumer panels

Surveys are the most widely used data collection method for studying mobile user behavior and service usage. Surveys can be implemented using e.g. telephone, postal mail, e-mail, web, and various face-to-face questioning methods, depending on the available resources and the objectives of the research. The used method also defines the scalability and accuracy of the survey. Time
series data can be produced by repeating a certain set of questions. Surveys are flexible as a wide range of information can be collected to study e.g. attitudes, values, beliefs, and past behavior. However, survey responses always depend on the respondents’ motivation, honesty, memory, and ability to respond. While a random sample of subjects is often selected for the survey, the actual respondents are usually self-selected, meaning that the true characteristics of the whole population cannot be obtained from the sample. Finally, formulations of survey questions and answer options could lead to different interpretations, both by the respondent and the researcher. [6]

A continuous panel study is a series of measurements on the same sample of test units over an extended period of time. Panel research is used extensively both in Europe and the U.S. to study consumer purchase patterns [7], and it has been used to study mobile service usage as well [8]. Participating panelists register usage events manually to an online or paper diary, which results in data of high accuracy and granularity. The basic limitations of the panel method are similar to those of surveys, and the continuous nature of panels significantly limits sample sizes.

2.2 End-user device monitoring

Panel studies can also be conducted at the end-user device level, where the manual registering of usage events is replaced with the logging functionality of monitoring software or hardware installed in the device. Recruiting a representative panel of people is one of the main challenges also in end-user device measurements. Explanatory background variables on device user(s) are also often collected at the beginning of the panel, with the typical survey data reliability issues. In case of devices with multiple users (e.g. PC, TV), measures need to be taken to distinguish the usage of each individual from one another.

End-user device measurements are quite common in the PC world. Commercial analysts, such as Nielsen NetRatings, have panels with hundreds of thousands of monitored participants [9]. The software records what a user does with the computer and sends that information further. While being otherwise quite similar to benign PC monitoring software, malicious spyware monitors the PC without the consent of the user. Spyware can collect information ranging from tracking the types of visited websites to recording the user’s keystrokes to intercept passwords or credit card numbers.

Television viewing can also be measured at the end-user device level. In Finland, television viewing research is conducted with a continuous measurement of a 1000 household panel. The TV sets in the panel are equipped with metering devices to monitor TV channel viewing of household members. [10]

The monitoring of advanced mobile handsets has become possible due to recent developments in handset operating systems and processing capability. A handset monitoring software can measure the usage frequencies, durations and volumes of all device features and applications. Communication-specific data (e.g. voice calls, SMSs) can be broken down between different callers/senders and recipients, and visited browsing destinations can also be logged. Usage of the handset’s different radio interfaces (e.g. GSM, WCDMA, WLAN) for packet data transmission by different handset applications can be differentiated, and the possible ad hoc connections (e.g. Bluetooth, WLAN) with other terminals can be measured as well. Usage of the handset’s offline features (e.g. camera, multimedia player, games) and broadcast multimedia reception (e.g. FM radio, DVB-H based TV) can also be captured. Moreover, location information (cell identity code, WLAN access point name, GPS coordinates) connecting user location and mobility to usage might also be collected, though the actual geographic location of access network base stations might not be known by the device. The monitoring software can be further augmented with triggered “real time” pop-up questions sent to the panelist after certain pre-specified event (e.g. at 12 o’clock, browsing session ended). While handset monitoring is not yet very common, at least two such
software applications for the Symbian S60 platform with somewhat differing functionality have previously been used by the academic community ([11] and [12], [13] and [14]). A major drawback of these studies is that their scope is limited to the users of a certain, albeit widely used, handset operating system and software platform. As the particularities of the platform might also have a large effect on usage behavior, the results are not generalizable to users of significantly differing mobile handsets. Commercial spyware specifically made for mobile handsets with some of the above functionalities has also emerged recently [15].

2.3 TCP/IP traffic measurements

TCP/IP traffic measurements (a.k.a. “network sniffing” or “packet sniffing”) are a common research approach in more technologically-oriented studies. Thus, de facto standard tools and formats exist for collecting TCP/IP traffic traces. Mobile device originated TCP/IP traffic can be measured at various places. Network architecture is in an important role, as points of convergence of mobile data traffic should be found at the network to attain comprehensive and representative measurements.

GSM/UMTS networks provide a centralized point for packet data traffic measurements, as traffic to and from the mobile terminals is routed to external networks via just a few places. Measuring traffic at an access point to an external packet data network (at the GGSN Gi interface), such as the Internet, effectively captures traffic of all of the operator’s subscribers [11]. Traffic measurement in WLAN networks is somewhat more problematic due to the small size of individual WLAN hot spots and lack of centralized routing in larger city-wide implementations. Traffic should be captured at the backhaul connection near the router connecting several WLAN access points to the wired network. On the other hand, if device-specific MAC addresses are used to identify individual users, measurement should be conducted nearer the access points before the first router. [5] WiMAX access networks featuring larger geographical coverage and smaller number of base stations than WLAN networks could provide more convenient points for traffic measurements.

Traffic measurements are sometimes also conducted in IP core networks (see e.g. [16], [17]). However, the enormous traffic volume in the core network necessitates rather short measurement periods or sampling, and essentially does not allow capturing application layer data. Furthermore, identifying mobile device originated traffic from PC traffic is more difficult. Finally, traffic measurements can also be conducted at the server side by various service providers. This is, however, rarely done as other more potent methods exist for measuring usage at the server level (see chapter 2.5).

Traffic of mobile handsets can be distinguished from other device types such as laptops by identifying the end-user device operating system (OS) software. This can be done from TCP/IP traffic with at least two methods. In the HTTP protocol header, the User-Agent field provides information on the used browser, the underlying operating system, and in some cases even the used device model. However, using application layer data should be avoided as it introduces various privacy issues, considerably increases the resource needs in traffic trace collection, and simply does not work with encrypted traffic. A solution to these problems is to use a method known as TCP OS fingerprinting, which recognizes different OSs by identifying idiosyncrasies in the implementation of their respective TCP/IP stacks. No application layer data is needed, as the method only uses certain TCP and IP header fields. However, as proxy servers establish separate TCP connections with the traffic destination hosts, the originating OS cannot be identified after a proxy with TCP fingerprinting. [18] While HTTP User-Agent fields and TCP OS fingerprints can be forged by malicious users, this is not likely to happen with mobile devices in a large scale.
Application protocols can also be identified from TCP/IP traffic traces. The most accurate method to identify application protocols is to observe their headers directly from application layer data. However, using application layer data introduces the same difficulties as presented above in the case of OS identification. The most straightforward identification method is to use transport protocol (TCP, UDP) port numbers. This method works for most mainstream applications (e.g. web, e-mail), but cannot be accurately used to detect some emerging application types (e.g. VoIP, P2P) due to multiple potential inaccuracies (see e.g. [17]). Traffic characterization utilizes traffic behavior, such as relative packet and byte volume per flow shares, in addition to the transport protocol port numbers to classify traffic to service classes. This method can improve the accuracy of the simple port number based method. [19]

Straightforward methods exist for aggregating raw traffic traces into time-stamped traffic flows, with the associated uplink and downlink byte volumes and packet counts. Using the methods described above, each flow can be categorized by originating operating system, application protocol, destination host IP addresses (e.g. for web servers), as well as by weekday and time of usage. Individual usage sessions can be separated from typical traffic traces, but they cannot be directly accounted to individual users if IP addresses are allocated to user terminals dynamically. At the access network level users can be distinguished by logging the allocation of IP addresses at a centralized authentication server (based on e.g. RADIUS [20] or DIAMETER [21] protocols) or at another specific network element (e.g. GSM/UMTS network SGSN [22]), or possibly by studying the application layer authentication protocol data directly from the captured traffic. Moreover, with such effort the subscriber identifiers (e.g. login name, IMSI code) associated with each individual user and his traffic could also be linked to access provider’s subscriber register data.

2.4 Usage accounting systems, registers

For any service provider the accounting system that registers the usage of chargeable services by individual users is a great source of information on user behavior. For an IP network access provider, the accounting function is a part of the AAA (Authentication, Authorization, and Accounting) system, whereas in GSM/UMTS networks the function is typically called charging and billing. Furthermore, service provider customer/subscriber registers often provide additional explanatory background variables on individual users.

In GSM/UMTS networks, the time-stamped charging data records (CDRs) register the mobile subscriber (by IMSI code), the used mobile terminal (by IMEI code), and the used service (e.g. voice call, SMS, packet data traffic to/from different networks). The billing system combines aggregated CDR data with tariff information. Register data on subscribers can also be linked to usage data, depending on the specific implementation of the reporting functionalities of the operator’s information systems. Unrestricted access to the CDR databases or customer registers also enables the use of sophisticated data mining techniques to uncover usage patterns (e.g. [23]). However, in this type of data the subscriber only refers to the bill payer of the subscription, not the actual end-user. Similarly, the mobile terminal is actually the SIM card equipped terminal connected to the network, not necessarily the actual end-user device such as a PC.

A centrally controlled network function, such as authentication or network management, is required for efficient WLAN network usage measurements as usage accounting can then also be conducted in a centralized manner. WiMAX networks will most likely have these functions under centralized management. AAA systems typically provide similar data to CDR data, i.e. time, duration and volume of usage for each individual subscriber. By definition, authentication data also entail a linkage between the user and subscriber register.
2.5 Server side measurements

At the server side, in addition to the above-mentioned TCP/IP traffic and usage accounting system based methods, service usage and behavioral patterns can be studied by collecting log files at various servers. These include e.g. portals and individual web/wap sites/servers, search engines, and proxy servers. Background data on the registered users of a service might also be available. The scope of any server level measurement is naturally limited to the users of the service in question.

Usage of web portals and individual sites can be monitored in a similar manner. A typical method includes placing small pieces of JavaScript code on all pages of a web site. Each time a page is loaded by a user, the code executes and sends data to another server that is generating the log files. New and repeated visitors can be distinguished using browser cookies. At best, the method enables the identification of individual users and their detailed usage patterns. By identifying the used browser version (from the HTTP header) the use of a mobile device can be identified as well, although currently only the most advanced mobile browsers adequately include the required JavaScript and cookie support. Another way of obtaining web site usage data is to use server software (e.g. Apache web server) log files. Log file data is typically less detailed, but can cover several sites, as web hotel and hosting services often locate multiple web sites on the same server machine. [24], [25]

Analyst companies (e.g. TNS [26], Nielsen/NetRatings [9]) providing web site usage analysis services can also combine the data measured on multiple individual web sites. When such data is published, if published at all, it typically covers the service providers in one country regardless of the origin of the actual users (see [26]). No mobile device specific data has yet been published. The representativeness of these studies is somewhat questionable, as they only cover the clients of the analyst company who also allow the publication of their own data. This means that all web sites are (by far) not included in the data, as analysis services are typically purchased only by well-established service providers. Thus, providers of services such as advertising and adult content, for instance, are typically not covered by the method. Other analytics software (e.g. Google Analytics [27]) could be used similarly to aggregate browsing data.

Search engine companies have another source of usage data at their disposal, as analysis on the most used search words provides information on service popularity. Mobile usage can be identified from accesses to mobile-adapted search sites as well as from the use of services specifically made for searching the mobile web (e.g. Google Mobile [28]). Moreover, a mobile device accessing standard PC search site can also be identified, again by its browser type, and the searches of individual users can again be distinguished with the help of browser cookies. As search companies typically also offer a range of other Internet services they might be able to relate the background data on registered users to search behavior, regardless of which device is used as long as browser cookies in the device identify the individual in question. The potential of such data was demonstrated in autumn 2006, when AOL released about 2.2 gigabytes of search logs to the general public with the seemingly good intention of providing the research community with hard search engine usage data. While the identities of individual people were not revealed by AOL, many of them could be easily deduced from the used search words. [29] In another much less controversial case, search engine data has been used to analyze mobile search patterns [30].

Traffic of multiple users converges also at proxy servers. Caching web proxies, for instance, can be used to measure web site popularity. The Opera Mini java browser for mobile devices serves as a specifically mobile related example of proxy based usage measurement. As Opera Mini fetches all requested content through an Opera proxy, detailed statistics on the browsing behavior of all Opera Mini users are available to Opera Software ASA (see e.g. [31]).
3 Comparison of mobile usage data collection methods

Mobile usage can be adequately covered only with some of the data collection methods, including:
1. Surveys and panels
2. Handset monitoring
3. Wireless access network TCP/IP traffic measurements
4. Wireless access network usage accounting systems
5. Server measurements

These methods have fundamental differences regarding research scope and characteristics of collectable data. Moreover, privacy issues and other limitations also hinder data collection efforts.

3.1 Research scope

The data collection methods differ regarding the typical researcher, method scope, and sample size. A summary of the characteristics of each method is presented in Table 1.

<table>
<thead>
<tr>
<th>Methods Attributes</th>
<th>Surveys and panels</th>
<th>Handset monitoring</th>
<th>Wireless access network</th>
<th>Server measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Researchers with access to data</td>
<td>Anybody (analysts, consulting firms, investment banks, academics…)</td>
<td>Those with a monitoring client and ability to recruit the sample</td>
<td>Wireless access operators</td>
<td>Service providers (mobile operators, 3rd party providers, search engines…)</td>
</tr>
</tbody>
</table>
| Method scope and sample size | Survey respondents: $10^{-1} - 10^3$  
Panel participants: $10^{-1} - 10^3$ | Sample of panelists using one handset OS and software platform: $10^{-1} - 10^3$ | Operator’s subscriber / terminal base: $10^2 - 10^7$  
Operator’s subscriber / terminal base: $10^2 - 10^7$ | All users of the measured service(s): $10^2 - 10^7$ |

The selection of a usage data collection method depends very much on its availability, i.e. on the researcher’s access to data. Surveys are by far the most used method as they are available to practically everyone. Implementing a handset monitoring panel is harder, as a functioning monitoring client for mobile devices is needed in addition to the capability of recruiting a representative panel. Only the operators are able to do measurements in the wireless access networks, as direct access to operator information systems or the network itself is required. Server measurements can only be conducted by the service provider in question, or someone on their behalf.

The selected research method dictates the scope and size of the studied sample. While survey sample sizes vary a lot depending on how the survey is implemented (e.g. doorstep vs. web surveys), panel studies typically have smaller sample sizes. Wireless access network measurements have varying sample sizes, as the measurement scope can range from the users of a small WLAN hot spot up to the entire subscriber base of a large 2G/3G mobile operator. Server side measurements can also have varying samples, at maximum encompassing all users of the service(s) measured.
3.2 Characteristics of data

The nature of collectable data varies by method. The method largely determines which specific usage variables and independent variables explaining the usage patterns can be obtained. A summary of the characteristics of data provided by each method is presented in Table 2.

Table 2. Summary of data characteristics by method

<table>
<thead>
<tr>
<th>Methods Attributes</th>
<th>Surveys and panels</th>
<th>Handset monitoring</th>
<th>Wireless access network</th>
<th>Server measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of data</td>
<td>Subjective data on perceived aggregate service usage, diary method could sort out individual events</td>
<td>Accurate and objective data on handset application and feature usage</td>
<td>Quite accurate and objective profile of TCP/IP traffic in the network</td>
<td>Objective data on service-specific usage, accuracy depending on the method</td>
</tr>
<tr>
<td>Explanatory / independent variables</td>
<td>Any explanatory variables on the respondents Perceived time and context of usage</td>
<td>Any background variables on the panelists Handset model and access network used Time of usage Location of usage (cell ID, WLAN access point name, GPS coordinates)</td>
<td>No data on individual users without extra effort, different terminal operating systems identifiable Time of usage No location data</td>
<td>Some background variables on the subscribers (type, tariff), no data on real end-users Terminal model in GSM/UMTS Time of usage No location data</td>
</tr>
<tr>
<td>Usage / dependent variables</td>
<td>Perceived amount and frequency of service usage</td>
<td>Volume, frequency, and duration of handset application and feature usage</td>
<td>Volume of usage (bytes, flows) per application protocol and traffic destination</td>
<td>Depends on the used method</td>
</tr>
<tr>
<td></td>
<td>Usage volume Usage frequency</td>
<td>Additional measurements could</td>
<td>Volume and frequency of service usage (e.g. calling, SMS, data transfer)</td>
<td>1) link usage accounting data and traffic data 2) add location data to both measurements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1) link usage accounting data and traffic data 2) add location data to both measurements</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The main difference between surveys and different automated data collection methods is that surveys provide subjective data on usage as perceived by the respondents, whereas measurements generally provide more objective data. The accuracy and granularity of data vary by the method. Survey respondents are likely able to provide information on aggregate usage of different services, whereas diary based panel and actual measurements can attain higher granularity by registering individual usage events and transactions.

The ability to obtain variables potentially explaining usage varies a lot by method. Survey studies can include any background variable (e.g. gender, handset type, pricing scheme) or other more abstract explanatory variable (e.g. respondent attitude and values) the respondent is capable of providing. Data on the perceived time (e.g. evening) and context (e.g. home) of usage can also be obtained, while diary based methods might achieve higher level of accuracy in this respect. Handset monitoring typically begins with a survey study collecting relevant background data on the panelists. The monitoring software itself is aware of the handset model and features, the used access network, as well as the time of each usage event. The location of usage might also be collected (cell ID, WLAN access point name, GPS coordinates). In wireless access network TCP/IP traffic measurement individual users cannot be separated from each other without complementary measurements. However, analysis methods enable the separation of different terminal operating
systems, and e.g. the separation of mobile handsets from laptop originated traffic. The exact time of usage is also registered in TCP/IP traffic measurements. Wireless access network usage accounting system based data can be associated with subscription information (subscription type, service tariffs), although no data on the actual end-users is typically available. In GSM/UMTS networks, the used terminal is also identifiable from charging data. A timestamp of each transaction is registered as service pricing might depend on it, and other background variables might also be available. For server level measurements, everything depends on the type of method used. Individual users might be separated and identified, and previous registration could also provide explanatory background variables for each user. Mobile handset originated usage can be distinguished from PC usage in some cases, as necessary if the method is used for measuring mobile users. The time of service usage is also often obtainable. User location cannot be known at the server, though the operator and country of the user can be derived from terminal IP addresses.

The usage volume and frequency of different services can be obtained at different levels of granularity. Surveys and panels provide data on the amount and frequency of service usage, as perceived by the respondent. Handset monitoring registers the volume, frequency and duration of usage per application for each panelist. TCP/IP traffic measurements provide byte and flow counts on the volume of usage, accountable to used application protocols and traffic destinations (Internet hosts/servers). Usage accounting systems can produce data on service usage (e.g. voice calling, SMS, data transfer) typically aggregated to subscriber or terminal type level. The type of data obtainable on the server level depends completely on the used measurement method.

A GSM/UMTS network is aware of the cells covering terminals with active packet data connections, and usage of specific WLAN access points can be obtained from e.g. AAA data. Moreover, the access operator can also relate cell identities or base station sites to geographical locations. However, combining location information to usage accounting data let alone traffic data might prove too laborious in large networks. While such efforts are not common in GSM/UMTS networks (see [22]), more examples from smaller WLAN networks can be found (e.g. [5], [32]).

**3.3 Privacy issues**

The legislation designed to ensure the protection of privacy sets strict limitations on usage data collection. As there are differences in national legislation, the specific limitations might vary by country. In general, Finnish privacy laws are relatively strict, especially concerning employers’ rights to monitor the employees. In the future, legal convergence (e.g. EC Directive on Privacy and Electronic Communications [33]) should lead to similar laws in the European countries. In Finland, two laws stipulate by whom, how, and for what purposes identification information (Act on the Protection of Privacy in Electronic Communications 516/2004 [34]) and personal data (Personal Data Act 523/1999 [35]) can be used. Interpretations of the laws can be made for the context of end-user behavior and service usage research. Identification information can be used for technical development of products or services, and for marketing purposes with the subject’s consent. Personal data may be processed for the purposes of historical, scientific, or statistical research, as well as for direct marketing and market research purposes unless prohibited by the data subject. In Finland, the responsibility in controlling the compliance is shared by the Finnish Communications Regulatory Authority and the Office of Data Protection Ombudsman, depending on whether the matter concerns identification information (the former) or personal data (the latter). However, it is not always evident which law is to be applied and which authority is concerned in a measurement effort, even in the Finnish context. Thus, making a generic comparison of the legal implications of each data collection method is not feasible.
Collection of usage data also raises many privacy concerns. At the level of an individual, users might change their behavior when aware of being studied. This is particularly relevant in participative methods such as surveys and handset monitoring studies. On the other hand, the “big brother” effect referring to pervasive surveillance and the reduction of personal privacy takes place when users are not certain what data is collected on them and by whom. These concerns have been recently raised by governments’ and corporations’ increased data collection activity for both anti-terrorism (e.g. USA PATRIOT Act [36], EC directive on data retention [37]) and commercial purposes (e.g. Google). At an organizational level, companies providing usage data (e.g. operators) for research purposes might have sensitivity concerns as well, as company-specific data should not be divulged to competitors.

3.4 Other limitations of data collection methods

The major limitation of any type of usage measurement is the fact that one can only measure what a user has done, not why the user has done something or what he/she actually intended to do. Thus, determining the motivation of usage or the real end-user need is hard based on measurement data only. Surveys are more flexible in this regard, as data on both perceived usage and motivation of usage can be collected, provided that the survey questions are formulated accordingly to obtain valid data. Handset monitoring facilitating the pop-up question mechanism provides a promising new method to uncover the true reasons of usage, as the panelist can be asked triggered questions immediately following a usage session. However, such an intrusive method must be used very carefully. While traffic measurements and usage accounting systems can’t basically obtain any data on the motivation of usage, server level measurements have some means to understand it for instance by analyzing the user’s site-specific browsing or search patterns.

Practical resource limitations also restrict data collection efforts. The resource requirements of different data collection methods vary, and are also implementation-specific. Generic research phases can nevertheless be identified. A planning phase is naturally present in each method. Access to data has to be organized when using any of the methods. In surveys and panel studies this includes the recruitment of respondents and panel participants, whereas in access network and server measurements access to the measurement points (i.e. network elements, information systems) has to be negotiated with the operator or service provider running the systems. The setup and data collection phase actually implements the data collection. Surveys entail the highest degree of manual work, whereas the setup of appropriate measurement equipment (hardware and software) has to be conducted in automated measurements. Finally, in processing and analysis the data is typically processed manually or with some separate software/script, and then input into a software tool capable of statistical analysis. The first time implementation of any data collection effort is naturally the most laborious, depending on the initial level of research know-how. However, subsequent data collection rounds can be facilitated by careful documentation and there is also potential for automation within all of the methods, especially with the automated measurement methods. On the other hand, measurements in live operational systems will inevitably involve operator or service provider personnel, not just the researcher. The associated research is always subject to cost-benefit considerations, especially when a separate measurement setup is needed for collecting the data. From this perspective, usage accounting systems and server measurements with built-in data collection mechanisms for e.g. billing purposes provide the most feasible data sources.

4 Future outlook on the data collection methods

Many powerful trends are currently shaping the mobile industry, as described in the introductory chapter. These trends will also have a profound effect on the future potential of the methods to study mobile user behavior and service usage.
At a general level, increases in processing power and memory capacity of mobile devices and service provider information systems will make various automated data collection methods more common. As each separate method has its advantages and disadvantages, using multiple methods can give further insight on usage. Combining several methods is recommendable also from the theoretical point of view, as method triangulation can improve both internal and external validity of the data [38]. Therefore, empirical measurement data will also be increasingly used to complement traditional survey based data. An example of a research problem that would benefit from the use of multiple data collection methods is the study of the market-wide adoption and diffusion dynamics of mobile data services. When collecting longitudinal data with such a research topic, properly implemented sample based methods give insights into the use of emerging services while wireless access network measurements describe more general market-wide usage patterns.

Traditional survey studies will lose some of their potential in explaining mobile user behavior. The rapidly growing number of mobile device users consists of people from highly varying backgrounds, which leads to a more fragmented user base and, thus, to challenges in representative sampling of survey respondents especially when studying the mass market. The number of wireless access alternatives will also have an effect on surveys. If mobile devices are capable of changing wireless access networks automatically, the users might not know which network they are using at any given time. Relating usage behavior to any characteristics of a particular access network will therefore be hard in survey studies.

As mobile device convergence (phone, PDA, camera, MP3 player, TV…) is currently taking place, a single mobile device could eventually be the point of convergence for the majority of usage of an individual user. Thus, the terminal is a promising place for measuring mobile user behavior in the future. Moreover, as the device monitoring software is aware of changes in network usage, behavior can also be associated with the used access network. The motivation of usage and user experience can also be studied with triggered pop-up questions.

The role of any one wireless access network among several alternatives will be smaller in the future, especially in the case of GSM/UMTS networks that currently take the bulk of mobile device originated traffic. While HSPA traffic is still measurable in GSM/UMTS networks, the spreading WLAN hot spots offer an attracting access alternative to multi-radio terminal users. Moreover, the IP everywhere trend threatens to reduce GSM/UMTS operators to mere providers of bit pipe access service, highlighted by emergence of access operator independent mobile VoIP service providers. The 3GPP (3rd Generation Partnership Project) has tried to counter these developments by specifying the IMS (IP Multimedia Subsystem) and UMA/GAN (Unlicensed Mobile Access / Generic Access Network), which could keep some of the IP based services still in the scope of operator usage accounting systems. IMS allows GSM/UMTS operators to provide the services typically provided at the Internet and charge for individual service use, while UMA/GAN allows roaming and handover between WLAN hot spots and 2G/3G mobile networks.

Most server side measurements should function in a similar manner regardless of the used access technology. As mobile handsets are converging with PCs in functionality and will be capable of using all the services previously used exclusively by PCs instead of the mobile-specific/adapted versions, distinguishing the two might become more difficult. This also applies to some handset applications (e.g. Opera Mini browser) whose implementation has enabled certain server side measurements. On the other hand, handset development (e.g. JavaScript support) will also enable some methods that previously functioned only with PCs. Some of the methods in identifying mobile clients will in any case be applicable, but the used access method might also be of less interest to service providers in the future when adapting the service to the lesser capabilities of the mobile devices is no longer necessary.
5 Conclusions

A range of alternative methods for collecting data on mobile user behavior and service usage have been presented in this paper. Each of the methods has its advantages and disadvantages, and the applicability of a particular method depends on the research objectives as no method is suitable for all purposes. Thus, multiple methods are often used. In summary, different methods entail a trade-off between sample size and data granularity, if issues such as legislative and resource limitations are not considered. Surveys and handset monitoring provide very detailed sample based data, whereas wireless access network data is less granular but can be based on the entire subscriber base of an operator. Server level measurements are a compromise between the above, as detailed usage patterns can be uncovered from a fairly focused user population.

Considering the future, all automated data collection methods provide increasingly accurate and granular data on user behavior compared to the traditional survey method. Nevertheless, as surveys can provide data on issues that are not directly measurable, automated data collection and surveys are likely to complement each other in the future. Mobile handsets emerge as the most promising place to measure mobile user behavior and service usage, while the mechanism of pop-up questioning also presents a potential new research approach. Server side measurements will also be increasingly potent, especially in the case of large providers of Internet based services (e-mail, search...), such as Google. 2G/3G mobile networks are likely to become less useful as mobile phone traffic will partly diverge to other wireless access networks.

Some suggestions for further research can be made based on the analysis made in this study. Current measurement methods facilitate the collection of time stamps for individual service usage events. While a geographic “location stamp” is needed for the implementation of some future mobile services, a “mobility stamp” depicting the speed of user movement could reveal even more about behavior of moving users (e.g. car, train, airplane). A “context stamp” (e.g. work, free-time) would also be an interesting addition in explaining the user motivations, and the potential of the pop-up questioning associated with handset monitoring should be further explored in this respect. Finally, a “price stamp” detailing the use of money would be essential for studying e.g. business models and user preferences. As electronic use of money also always entails user authentication and usage accounting, financial institutions (e.g. banks, credit card companies) could emerge as one new point of convergence for studying mobile user behavior.
Appendix

Explanations for the most important technical terms used in the paper:

- **CDR (Charging Data Record, formerly Call Detail Record).** A formatted collection of information about a chargeable event (e.g. time of call set-up, duration of the call, amount of data transferred, user identity, terminal identity…) used in billing and accounting.
- **DVB-H (Digital Video Broadcasting - Handheld).** A technology for bringing broadcast services to handheld receivers.
- **GGSN (Gateway GPRS Support Node).** An element of the GSM/UMTS network packet-switched core responsible for providing gateway access to external packet data networks and generating charging data records.
- **GSM/UMTS (Global System for Mobile communications / Universal Mobile Telecommunications System).** A 3GPP specified 2G/3G mobile telecommunication system.
- **HSPA (High-Speed Packet Access).** A performance upgrade to UMTS radio access.
- **HTTP User-Agent.** A field of the HTTP (Hypertext Transfer Protocol) protocol header identifying the software program used by the HTTP client.
- **IMEI (International Mobile station Equipment Identity).** A unique code identifying each mobile terminal in the GSM/UMTS system.
- **IMS (IP Multimedia Subsystem).** An architectural framework of 3GPP for delivering IP multimedia services to mobile end users using various wireless and wired access networks.
- **IMSI (International Mobile Subscriber/Station Identity).** A unique code identifying each mobile subscriber in the GSM/UMTS system.
- **MAC (Medium Access Control) address.** A unique serial number assigned to each network card/interface.
- **SGSN (Serving GPRS Support Node).** An element of GSM/UMTS network packet-switched core responsible for mobility management, session management, tunneling and routing, compression, authentication, encryption, and the generation of charging data records.
- **SNMP (Simple Network Management Protocol).** A protocol used by network management systems to monitor network-attached devices.
- **Syslog.** A standard for forwarding log messages in an IP network.
- **UMA/GAN (Unlicensed Mobile Access / Generic Access Network).** A telecommunication system allowing seamless roaming and handover between local area networks (e.g. WLAN) and wide area networks (e.g. GSM/UMTS) using the same dual-mode mobile handset.
- **VoIP (Voice over Internet Protocol).** Routing of voice conversations over the Internet or through any other IP-based network.
- **WCDMA (Wideband Code Division Multiple Access).** A solution for the radio interface in 3G mobile telecommunications systems.
- **WiMAX (Worldwide Interoperability for Microwave Access).** A technology based on the IEEE 802.16 standard enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL.
References


