Measured Performance of Data Transmission Over Cellular Telephone Networks

Timo Alanko, Markku Kojo, Heimo Laamanen, Mika Liljeberg, Marko Moilanen, and Kimmo Raatikainen

Department of Computer Science, University of Helsinki

Report C-1994-53
November 1994
20 pages

Abstract. Recent developments in mobile communication and personal computer technology have laid a new foundation for mobile computing. Performance of the data communication system as seen by an application program is a fundamental factor when communication infrastructure at the application layer is designed. This paper provides results of performance measurements of data transmission over two different cellular telephone networks, a digital GSM-network and an analogue NMT-network. Since our emphasis is on performance as seen by application programs, we use the standard TCP/IP protocols in the measurements. The performance is measured using three basic operations: establishment of a wireless dial-up connection, exchange of request-reply messages, and bulk data transfer. The external conditions under which the measurements were carried out present a normal office environment when the field strength of the cellular link is good or fairly good.

Key Words: performance measurements, wireless dial-up connections, cellular telephones, GSM, NMT.

CR Classification: C.4

1. Introduction

In the computer industry mobile computing will be the revolution of 90's in the same way as personal computers were in the 80's. In mobile computing wireless networking is a key factor. Recent developments in cellular network technology and telephones as well as in portable computers are opening new possibilities for wireless data communication. Portable computers and mobile telephones have become smaller, lighter, and more powerful while cellular networks offer more advanced services. Integration of wireline and wireless networking creates new possibilities to use the services available in computer networks. However, the technical challenges in the integration are hardly trivial.

Global System for Mobile communications (GSM), which is a European digital cellular telephone network [Rah93, MP92], offers a platform for wireless data communication by specifying a variety of data services; both teleservices and bearer services. In addition to digital cellular networks, analogue cellular telephone networks can nowadays, due to recent developments in modem technology, be used as a platform for wireless data communication. The most widely used analogue cellular systems are *Advanced Mobile Phone System* (AMPS), *Total Access Communications Systems* (TACS), and *Nordic Mobile Telephone* (NMT) [PJG89].

The end users of today are accustomed to wireline LANs, in which the transfer rate is megabits per second and the bit error rate is below 10^{-9}. Many client-server applications are primarily designed to work in such environments. Compared to traditional wireline links, wireless links offered by cellular telephone networks have low throughput, low reliability, and often low quality. The maximum line speed in cellular telephone networks varies from 1,200 bps up to 19,200 bps. Radio signal is affected by noise and interference, caused, for example, by automobile ignition sparks or by adjacent channels. This implies that a wireless link has a high error rate and that disconnections are not uncommon. The integration of LAN and a cellular telephone network implies that two networking environments, having very different characteristics, are to be combined.

Since the performance of cellular wireless links is low, it is important to understand their operating characteristics, as well as to fully exploit the narrow channel offered by them. However, the actual performance of the GSM data services (e.g. throughput and reliability) is still an open question. The performance of integrated networking is affected by several factors. One of the main factors is the transmission quality (error rate, delays, etc.). Other important factors include implementation and parameterization of communication protocols, compression of data and protocol headers, error control at different protocol layers, the settings of communication equipment, and the environment in which the transfer is done. All these factors and their interactions may significantly affect the feasibility of data communication services offered to end-user applications. Last but not least, one important difference between cellular telephone networks and fixed data networks arises from the line costs. Because the billing in telephone networks is based on the connection time, long idle periods during connections should be avoided. This implies that the applications should also be able to perform operations in a disconnected mode.

We measured the data communication performance over two cellular networks as seen by application programs. The networks examined were a digital GSM-network and an analogue NMT-network. In the GSM measurements we used the asynchronous, non-transparent bearer service with a line speed of 9600 bps. NMT offers a plain physical-layer connection, where the line speed is set by modems. To accomplish an application point of view to wireless networking we used a standard TCP/IP protocol suite in our experiments.

We selected three different workloads to be examined: 1) a dial-up connection establishment, 2) exchange of request-reply messages (with a null-operation at the server site), and 3) bulk data transfer. These three operations are typical components in many applications communicating over a wireless telephone link. The performance metrics we are primarily interested in are the time needed to set up a dial-up connection, the round-trip time of request-reply messages, and the maximum transfer rate which can be sustained in bulk data transfer. These metrics characterize the performance "visible" to applications. We are also interested in factors which explain this "visible" performance, like the failure behaviour.

Through the measurements we try to find answers to questions like:

1) How do different factors and their interactions affect data communication services offered to end-user applications?
2) Will old applications still work according to their specifications?
3) How fast and reliable is the dial-up connection establishment procedure?
4) Is the use of automatic disconnect/reconnect-operations a feasible solution to the line cost problem?

The answers to such questions are important for designers who are constructing a communication infrastructure at the application level or who are designing applications to be used in cellular environments.

The rest of the paper is organized as follows. In Section 2 we describe our test environments and configurations. In Section 3 we present the measurement results. In Section 4 we discuss the effects of error control, and in Section 5 we provide the summary and propose some topics for further research.

2. Test Environments and Configurations

In this section we describe the environments and configurations in which the performance measurement study was conducted. To begin with we briefly characterize the cellular GSM and NMT networks, particularly from the performance point of view. Next, we present the structure of the data communication architecture in our experiments and discuss some factors affecting the performance. Finally, we describe the workload used in the measurements.

2.1 Cellular Networks

In wireless communication, the coverage area is an important factor for usability. Cellular telephone networks have by far the largest coverage area of landbased wireless systems. For example, the coverage area of GSM includes almost all cities and main roads in Western Europe. World wide the total number of teleoperators that have implemented or that are planning to implement a GSM network is about 80 [Mob94]. Outside Western Europe GSM is (or very soon will be) available in Australia, China, Hong Kong, Kuwait, South Africa, Russia, and some other countries. Therefore, cellular telephone network seems to be a promising platform for wireless networking.

At the time when analogue cellular telephone systems were designed, data transmission was not considered a service worth supporting. However, recently specifications of data services in analogue networks have been completed, e.g., *Cellular Digital Packet Data* (CDPD) [Con93]. In addition, the developments in modem technology, particularly in modulation, compression, and error correction techniques, have improved data transmission capabilities.

Data transmission capabilities have, from the very beginning, been a design criterion of the GSM system. Therefore, GSM specifies almost the same services as *Integrated Services Digital Network* (ISDN) does. In particular, GSM specifies two modes of asynchronous bearer service: transparent and non-transparent. The transparent mode of transmission is derived from the ISDN specifications, primarily from V.110. The (available) throughput is constant and the transmission delay is fixed. The bit error rate at the line speed of 9600 bps is assumed not to exceed the level of 10^{-3}. The non-transparent mode is implemented using the *Radio Link Protocol* (RLP), which is a modification of the HDLC protocol. The RLP protocol includes error correction mechanisms which reduce (according to the specifications

Table 1: Hardware Used in the Measurements

Mobile Node	DECpc 325 SL, 386SX 25 MHz, 8 Mbytes of main memory
Fixed Host	AMBRA Sprinta 486, 486SX 25 MHz, 8 Mbytes of main memory
GSM Data Adapter	DTP-2 Cellular Data Card, (PCMCIA Type II Version 2)
Card Modem (MN)	Nokia PC Card Modem (PCMCIA Type II Version 2), V.32bis V.42bis
Modem (FH)	AT&T Paradyne (COMSPHERE 3810), V.32bis V.42bis
Modem Pool (IWF)	Nokia ECM Fast SW (DS61660), V.32bis V.42bis
NMT Phone	Nokia 121
GSM Phone	Nokia 2110

[ETS90]) the average bit error rate down to 10^{-8}. The price of the low bit error rate is variable throughput as well as variable delays in transmission. Since the radio signal is affected by noise and interference, rapid changes in the quality of data transmission and even disconnections are typical phenomena in cellular links. At the time of this study, GSM does not include any data compression services; however the specification work is in progress.

2.2 Test Environments

The data processing configuration in our experiments is similar to traditional client-server environments. The only difference is that the computers used in our tests, a *Mobile Node* (MN) and a *Fixed Host* (FH), are connected via a cellular telephone network. The mobile node is a typical "notebook PC" and the fixed host is a "desktop PC", both running MS-DOS*, version 6.2, and Windows*, version 3.1. A detailed list of the hardware used in our tests is given in Table 1.

We used two different cellular networks in our experiments — a digital GSM-network and an analogue NMT-network. The structure of data communication in the GSM environment is shown in Figure 1. The mobile node is connected to the GSM network through a data adapter and a GSM phone. The main components of the GSM network include a *Base Transceiver Station* (BTS) for each cell in the coverage area and a *Mobile Services Switching Center* (MSC). An interface to external networks is provided by the *Interworking Function* (IWF) at the MSC. In our experiments, the *Public Switched Telephone Network* (PSTN) is the external network used to connect to the fixed host. The PSTN transmission requires a

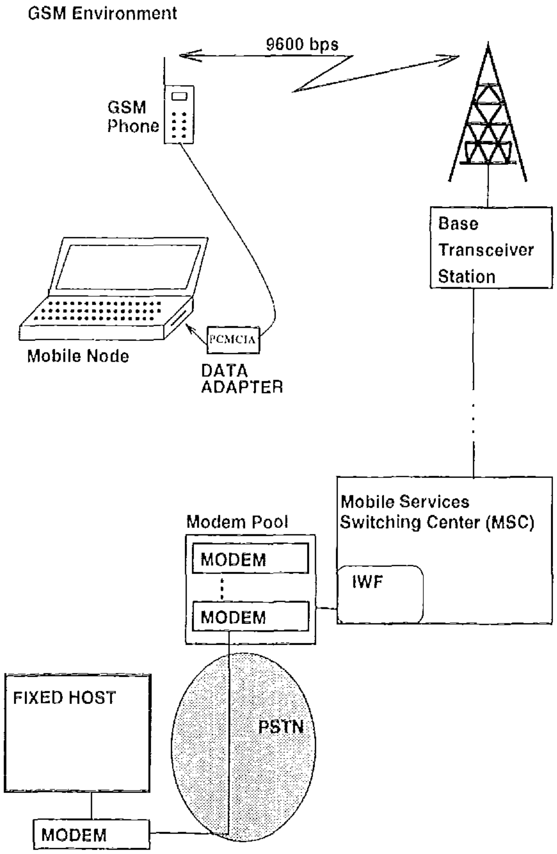

Figure 1: Communication Architecture in the GSM environment

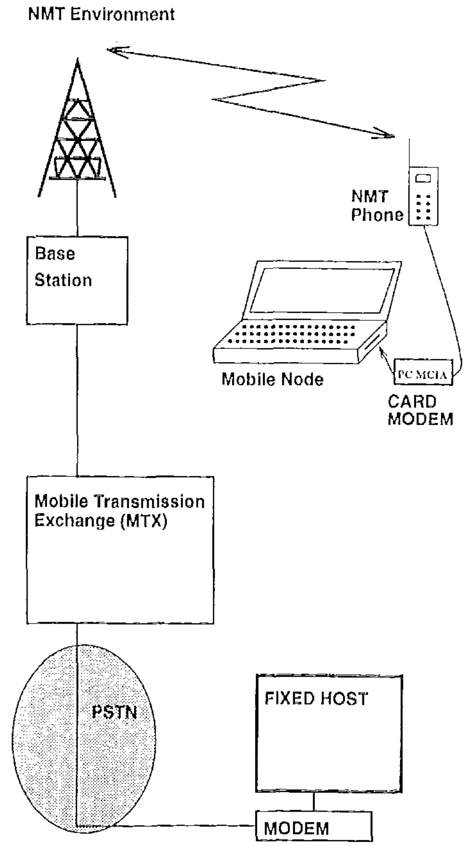

Figure 2: Communication Architecture in the NMT environment

pair of modems: one is selected, during the GSM data-call set up, from a pool of modems controlled by the IWF, the other one is connected directly to the fixed host.

The structure of data communication in the NMT environment is shown in Figure 2. The data access from the mobile node to the cellular phone is through a modem which is implemented on a PCMCIA card. In the NMT network, the access point to PSTN is offered by the *Mobile Transmission Exchange* (MTX). The fixed host is connected to the PSTN through a modem as in the GSM environment.

In both test environments, several protocols are involved in the data transmission path between the applications on the mobile node and the fixed host. These protocols are shown in Figures 3 and 4. We used a TCP/IP protocol suite over the *Serial Line Internet Protocol* (SLIP) [Rom88]. TCP was used as the transport protocol for request-reply messages as well as for bulk data transfer. The TCP/IP protocol suite was Digital's RoamAbout* Transporter

*Digital and RoamAbout are trademarks of Digital Equipment Corporation.

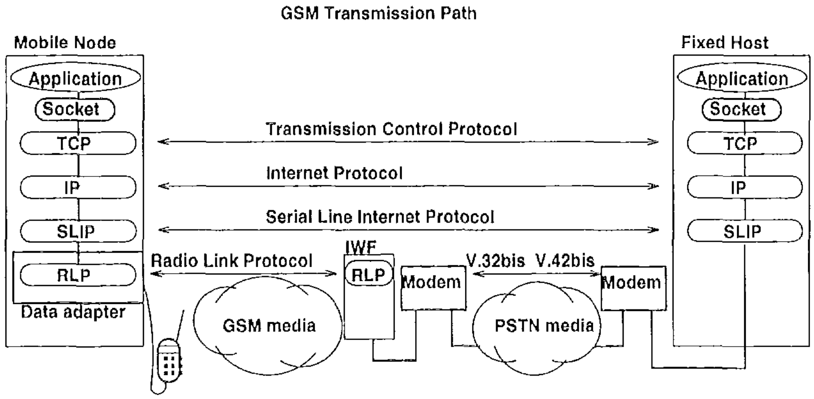

Figure 3: Protocols used over the GSM Path

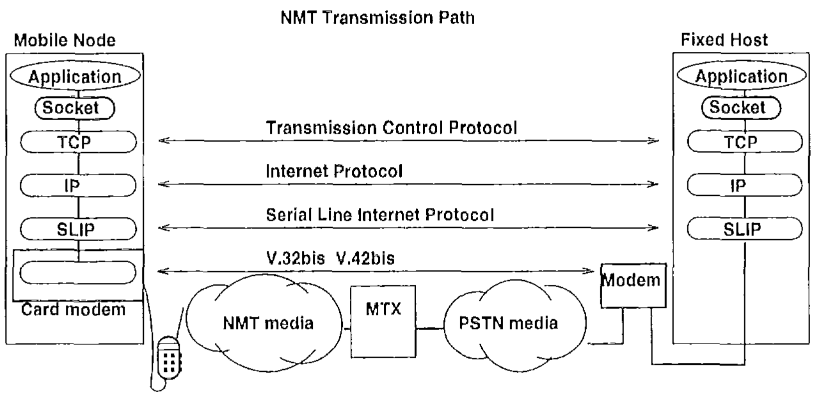

Figure 4: Protocols used over the NMT Path

version 2.0, which provides both the Berkeley and the Windows socket interfaces to TCP.

The reader should notice, that the implementation of the TCP/IP protocol suite affects the results, as does the processing power of the computer running the TCP/IP stack and performing the I/O operations. However, we did not measure nor analyze these factors. Furthermore, the protocol overhead of TCP/IP can be significant, especially over low-speed lines, if the maximum segment size (MSS) of TCP is short. The MSS used in all our tests was 536 bytes, as fixed by the implementation of the stack.

At the physical layer and the link layer there are several factors which affect the performance: nominal transfer rate, modem settings, parameters of the RLP protocol, frame size and use of data compression. Table 2 gives a detailed description of the parameters used in the measurements.

Achieving the best performance over the air path requires a trade-off between the transfer rate, the error rate, which is a function of the radio quality, and the frame size. In

Table 2: Parameters Used in the Measurements

	NMT	GSM	PSTN‡
Line speed	4800–7200 bps	9600 bps	14,400 bps
Window size	15 (LAPM)	61 (RLP)	15 (LAPM)
Frame size	16 octets (LAPM)	240 bits (RLP)	16 octets (LAPM)
Compression	V.42bis	none	V.42bis
Retrain	enabled	na	enabled

‡: PSTN part of the GSM path.

an NMT network, which is a plain analogue transmission medium, most of the setup is either preset or negotiated by the modems. We performed some preliminary experiments and observed that if the error control mechanism of modems is used, then the error rate remains reasonable with transfer rates not exceeding 7200 bps. Without any error control mechanisms, the error rate seems to be too high for TCP to operate efficiently.

2.3 Workload in Measurements

In a mobile environment it is appropriate to extend the concept of workload from the traditional one. We define the workload to consist of two parts: "what is done" and "under which external conditions the work is done".

The "what is done" part corresponds to the traditional concept of workload. In this study we are interested in three basic operations: establishment of a wireless dial-up connection, exchange of request-reply messages, and bulk data transfer. The performance of the dial-up connection establishment has to be considered in at least three important cases. The first one is related to savings in the channel costs: during long computations, is it profitable to tear down the wireless dial-up connection for the idle periods? The second case arises in data retrieval services: in a disconnected state, should missing data be retrieved automatically, i.e., without end-user intervention, over the wireless link? In both cases the end-user is confronted with an unexpected delay, which, if being too frequent or lasting for too long, may cause frustration. The third case is associated with fault-tolerance: the link-layer can be required to re-establish a broken connection without notifying the transport layer about the failure. The natural performance metric is the time needed for setting up the dial-up connection. Of some interest is also the frequency of the unsuccessful attempts and how the time is divided among the different phases of the call set-up procedure.

The performance in the request-reply type of communication is of vital interest when interactive applications are used in distributed environments. The end-user using a mobile workstation is most likely interested in response times of small tasks involving a modest amount of data exchange. In our experiments we measured the round-trip times of request-reply message exchanges with various message sizes: from 1 byte up to 2048 bytes. In these tests, no operation was performed at the server site; the server simply sent a reply message of a predefined size.

Bulk data transfer is an essential part of communication in many applications. Bulk data transfer, e.g., file transfer, usually involves a lot of clerical work. However, in this study we are only interested in the ability of a wireless system to transfer data in big amounts. Hence, in our experiments just raw data was sent, ASCII text to be precise, from the sender's memory to the receiver's memory; no disks were involved. We measured the arrival rate of incoming bytes received by the application program. In addition, the influence of transfer direction on the performance was examined.

The "under which external conditions the work is done" part of the workload specifies the conditions in which the mobile workstation is used. It is generally known that communicating over a mobile phone is sensitive to the environment where the phone is used. Moreover, it may even be sensitive to the time of day. For example, pre-lunch-hour peaks may cause problems in contention for bandwidth and resources. The field strength depends on the physical environment: the distance from the base station, geographical surroundings, buildings, and so on. It should also be noticed that the surroundings may also create reflections that affect the link quality even when the field strength is good.

The radio signal is sensitive to interferences with electromagnetic phenomena. These interferences may cause bursts of transmission errors and even cause call disconnections. When the mobile workstation is moving, handovers can cause errors and disconnections. It is also possible that the workstation moves out of the coverage area, which naturally causes at least a temporary disconnection. Hence, the failure behaviour in wireless cellular data transmission is very different from that in wireline transmission.

What we want to model is a "normal office environment" in which the field strength is good or fairly good. During the measurements the field strength, as displayed by the phones, was typically at levels 3 and 4 on a scale of 0–5. This "normal office environment" forms a baseline for the behavioral model of a mobile workstation. It is in a sense "the best possible situation" that an end-user may expect to have in his or her office. The effects of degrading conditions are easy to figure out, at least in qualitative terms: increasing frequency of interferences, which leads to longer delays and eventually to disconnections.

3. Measurement Results

As stated in the Introduction, we are interested in the performance that data transmission in current NMT and GSM telephone networks provides for the application programs. Therefore, we examined three different workloads which represent typical components of application programs that need communication services over a wireless telephone network.

3.1 Dial-Up Connection Establishment

Figure 5 shows the distributions of observed dial-up connection establishment times over NMT and GSM links. The connections were initiated from the mobile node to the fixed host. The time is measured from issuing the call command to receiving the "connect" message. The figure is based on tests of 400 successful connection establishments.

Over the NMT link a successful establishment typically took 31–35 seconds. A small fraction of establishments (c. 1%) took more than 35 seconds; the main reason for the extended delay was found to be in modem handshaking. When interferences degrade the wireless link, several retries may be needed before the link connection is established.

In addition to the total establishment time, we also examined the times spent in the major phases of the connection establishment. In the NMT environment the dominating time components in a typical connection establishment were roughly as follows: 1) the time to transfer the call command from the modem card to the telephone (typically c. 5 sec.), 2) the telephone call connection time to the modem of the fixed host (typically c. 11–14 seconds), and 3) the modem handshaking time (typically c. 15 sec., or more).

The modem handshaking time in phase 3) includes the modem specific delays, i.e., the delay before the beginning of the actual modem handshaking (c. 2 seconds) as well as the delay between the completion of the handshaking and receiving the "connect" message at the mobile node (c. 1 sec.). During the NMT test, in which we collected the 400 successful connection establishments, we observed 10 unsuccessful ones. This means that the failure probability in connection establishment over the NMT link was about 2–3 %. However, in poor conditions the failure probability can be much higher. The primary problem with these failures was the long detection time, typically 90–95 seconds.

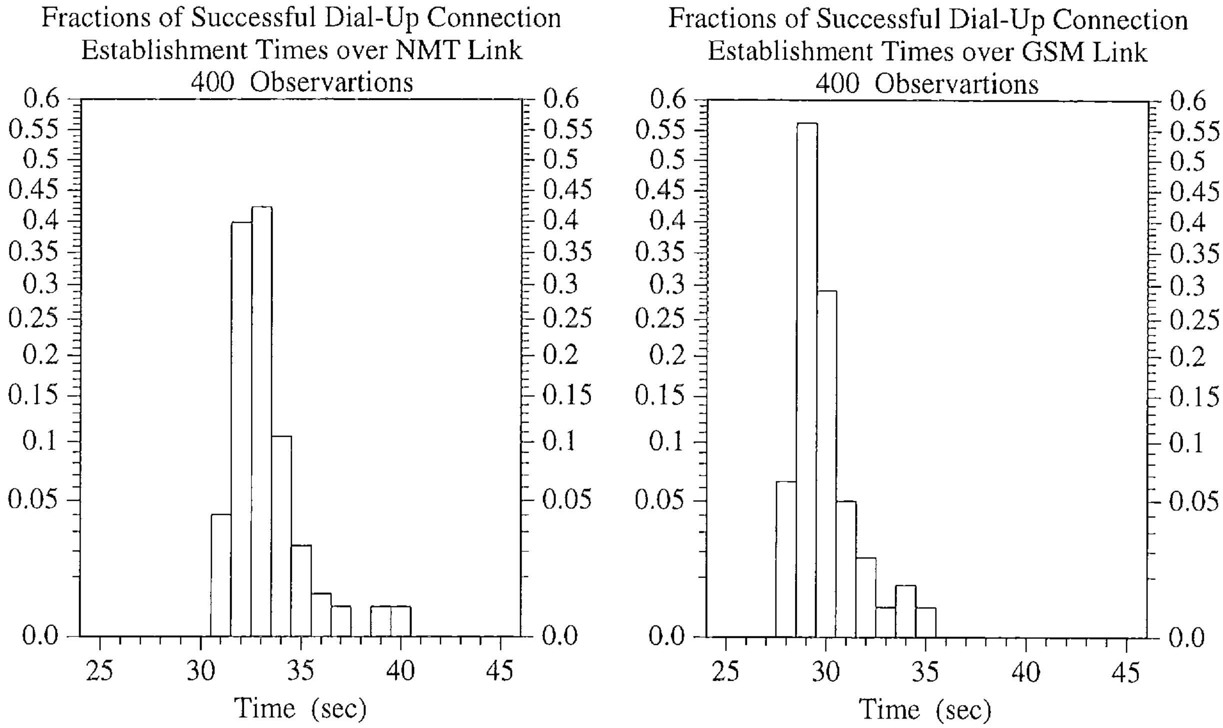

Figure 5: Distribution of Dial-Up Connection Establishment Times from the Mobile Node to the Fixed Host over NMT and GSM Links Based on 400 Successful Connection Establishments

The nonlinear scale on the y-axis is based on the arcsin-transform which stabilizes the variance of bar heights in empirical histograms.

Over the GSM link a successful establishment typically took 28–31 seconds. None of the establishments exceeded 35 seconds. The times spent in the major phases of connection establishment were roughly as follows: 1) the time to transfer the call command from the computer to the telephone (very short, less than 0.5 second), 2) the telephone call connection time (typically c. 12–14 seconds), and 3) the modem handshaking time (typically c. 16–18 seconds).

The modem handshaking time in phase 3) includes the actual handshaking between a modem in the modem pool and the modem connected to the fixed host, as well as the delays associated with the handshake. The modem specific delay (c. 2 seconds) before the beginning of the modem handshaking is the same as in the NMT environment. However, it takes roughly 4 seconds to deliver the "connect" message from the modem pool to the mobile node. During the GSM test, in which we collected the 400 successful connection establishments, we observed 46 unsuccessful connection establishments. Hence, in our experiments the failure probability in connection establishment over a GSM link was about 10–15 %.

Our measurements indicate that the tradeoff analysis of intentional disconnection/reconnection operations is far from trivial. There are two conflicting primary factors. The line costs are relatively high, which suggests that the connection should not be left open for long idle periods. On the other hand, a long connection reestablishment time, about half a minute, irritates the end-user. A feasible solution would be to make the reconnection to take place as a background operation started when communication is anticipated in the near future. Unfortunately, this kind of predictability can hardly be found in many applications. However, a tool allowing the end-user to control reconnection operations might be helpful. Another problem that must be taken into account is the failure probability in the connection establishment: although quite low in our measurements, it may raise notably high, especially if conditions are poor.

The reader should notice that the connection establishment time — especially the telephone call connection phase — depends on the structure of the PSTN. With a different structure within the PSTN, the connection establishment times could be different in both environments. Furthermore, we would like to emphasize that these tests were carried out in normal circumstances. When end-user activities increase, some components in the wireless environment will eventually become saturated.

3.2 Exchange of Request-Reply Messages

In the exchange of request-reply messages we are primarily interested 1) in typical round-trip times and 2) in the fraction of operations that take an exceptionally long time. The application in the mobile node sends a request to the application in the fixed host which immediately sends a reply. In our experiments we used five different pairs of message sizes. It should be remembered that in the NMT environment the modem compression was used whereas in the GSM environment no compression was available. Therefore, the results in the two environments can not be directly compared with each other.

The distributions of the round-trip times for the NMT environment are summarized in Table 3. The upper half of the table gives the 50^{th}, 90^{th}, 95^{th}, and 99^{th} percentiles of the round-tip times as well as the shortest and longest observed round-trip times. The lower half summarizes the long round-trip times by giving the fractions of outliers based on various criteria. We say that a round-trip time is an outlier if it exceeds the median round-trip time by more than *tol* seconds. The criteria used in this study are: tol = 0.2, 0.5, 1, 2, 5, and 10 seconds.

Table 3: Summary of 500 Round-Trip Times (seconds) in Exchange of Request-Reply Messages over NMT Link

Req./Rep. Size	Percentiles					
	Min	50%	90%	95%	99%	Max
1/1	.27	.44	.49	.60	6.10	17.64
32/512	.71	.93	.99	1.07	4.67	18.02
32/1024	.99	1.21	1.32	1.43	15.44	18.52
1024/32	1.26	1.48	1.54	1.65	11.29	18.52
32/2048	1.54	1.79	1.98	2.03	2.66	8.13
Req./Rep. Size	Fraction of Outliers†					
	0.20	0.50	1.00	2.00	5.00	10.0
1/1	.032	.018	.012	.012	.012	.006
32/512	.042	.014	.012	.012	.010	.006
32/1024	.058	.030	.030	.026	.014	.012
1024/32	.042	.032	.030	.022	.010	.010
32/2048	.086	.020	.008	.008	.006	

†: Observation is an outlier if it exceeds the median by more than the seconds given in the column label

Table 3 supports the following two main conclusions. Firstly, most round-trip times are close to the median, which varies from c. 0.4 to c. 1.8 seconds in our tests. With each pair of message sizes, more than 95% of the round-trip times are within 0.5 seconds. Secondly, only few round-trip times are exceptionally long. About 1% of the round-trip times exceeds the median by more than 5 seconds.

The typical round-trip time with one byte messages is 0.4–0.5 seconds. This is obviously a lower limit to response times in normal conditions. However, the time can be reduced: When both the header compression and the modem compression were on, we observed typical round-trip times of 0.1–0.2 seconds with one byte messages. For an end-user, the longest acceptable delay in echoing a character is of the order of 0.2 sec [Sch87]. Therefore, our results suggest that applications in which the characters are echoed over an NMT link are convenient to use only if all compression possibilities are in use.

Typical round-trip times with the other message sizes are 1–2 seconds. When we compare the round-trip times for reply messages of 0.5, 1, and 2 kilobytes, we observe that each additional kilobyte in the reply increases the round-trip time by about 0.6 seconds. These measurements indicate that applications which have a textual user interface can typically be

Table 4: Summary of 500 Round-Trip Times (seconds) in Exchange of Request-Reply Messages over GSM Link

Req./Rep. Size	Percentiles					
	Min	50%	90%	95%	99%	Max
1/1	0.77	0.99	1.04	1.04	1.10	1.15
32/512	1.37	1.54	1.59	1.65	1.65	1.70
32/1024	1.98	2.14	2.20	2.20	2.25	2.31
1024/32	2.09	2.20	2.31	2.31	4.15	4.56
32/2048	3.13	3.24	3.35	3.35	3.35	4.56
Req./Rep. Size	**Fraction of Outliers†**					
	0.20	0.50	1.00	2.00	5.00	10.0
1/1						
32/512						
32/1024						
1024/32	.034	.028	.014	.008		
32/2048	.004	.002	.002			

† Observation is an outlier if it exceeds the median by more than the seconds given in the column label

used over an NMT link. However, round-trip times taking 5–20 seconds are quite possible. They can create timing problems in many applications designed for traditional LANs and WANs.

The round-trip times measured in the GSM environment are summarized in Table 4. The transmission delays seem to be rather long: the round-trip time of one byte is about 1 second. Each additional kilobyte in the reply increases the round-trip time by about 1.1 seconds.

The conclusions over the distributions of round-trip times are almost the same as in the NMT environment. However, instead of being just close to the median, most round-trip times are *very* close to it. In each case more than 95% of observations are within 0.3 seconds. In a GSM environment round-trip times which are exceptionally long are *rare*.

Figure 6 summarizes the typical round-trip times for exchange of request-reply messages over NMT and GSM links. The figure shows the minimum, the median and the 90^{th} percentile. As we have already mentioned, the comparison between the environments is strongly affected by the modem compression used in the NMT environment. However, the figure indicates that the GSM environment is more stable.

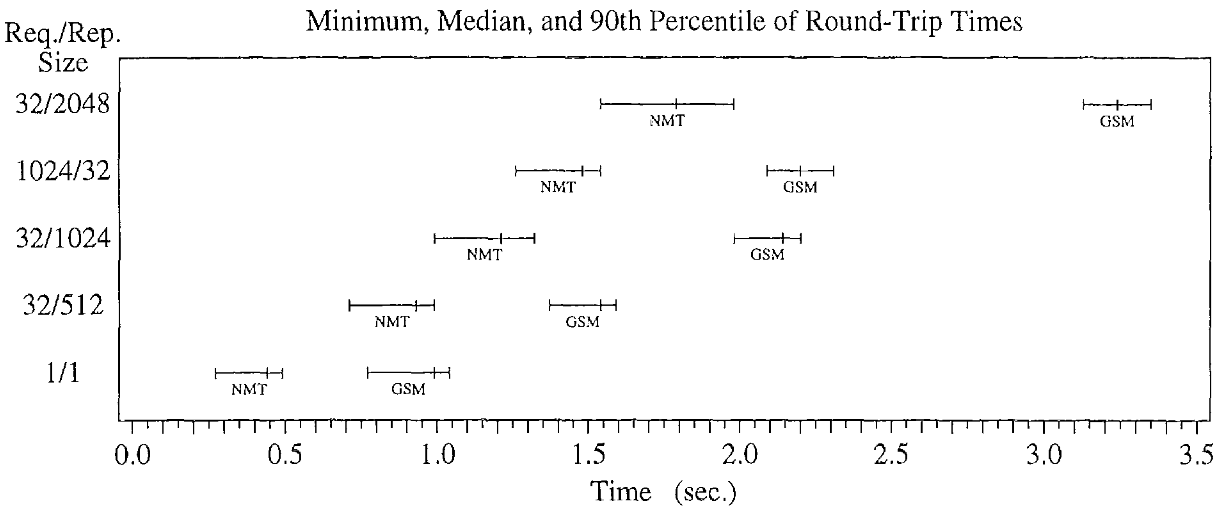

Figure 6: Summary of Round-Trip Times in Exchange of Request-Reply Messages over NMT and GSM Links

3.3 Bulk Data Transfer

In the bulk data transfer we are primarily interested in 1) maximum sustainable transfer rate in very good conditions and 2) the effective transfer rate in different operating conditions. We measured the times between blocks of bytes received by the application program. In both environments we repeated the bulk data transfer test ten times from the fixed host to the mobile node and ten times from the mobile node to the fixed host. In the NMT environment the total amount of bytes transferred was 2,200,000 and in the GSM environment 3,400,000.

Figure 7 illustrates the behaviour of bulk data transfer when the fixed host sent 40,000 bytes to the mobile node. The effect of the radio link problems becomes clearly visible. During the period of interference the transfer stagnates for a while to continue later at full rate. The frequency and length of these stagnations depend on the location of the mobile node, on the changes in its environment and on its session-time mobility.

The maximum sustainable transfer rate can be estimated using robust regression in the following way. Let t_i denote the time when the i^{th} block of bytes is received. Let x_i denote the cumulative amount of bytes received at time t_i. Now $r_{ij} = (x_j - x_i)/(t_j - t_i)$, $i \neq j$, is an observed transfer rate. The estimated maximum sustainable transfer rate is the median of all possible r_{ij}'s.

In the NMT environment the estimate is 1,270 bytes per second when the data is sent from the mobile node and 1,380 bytes per second when the data is sent from the fixed host.

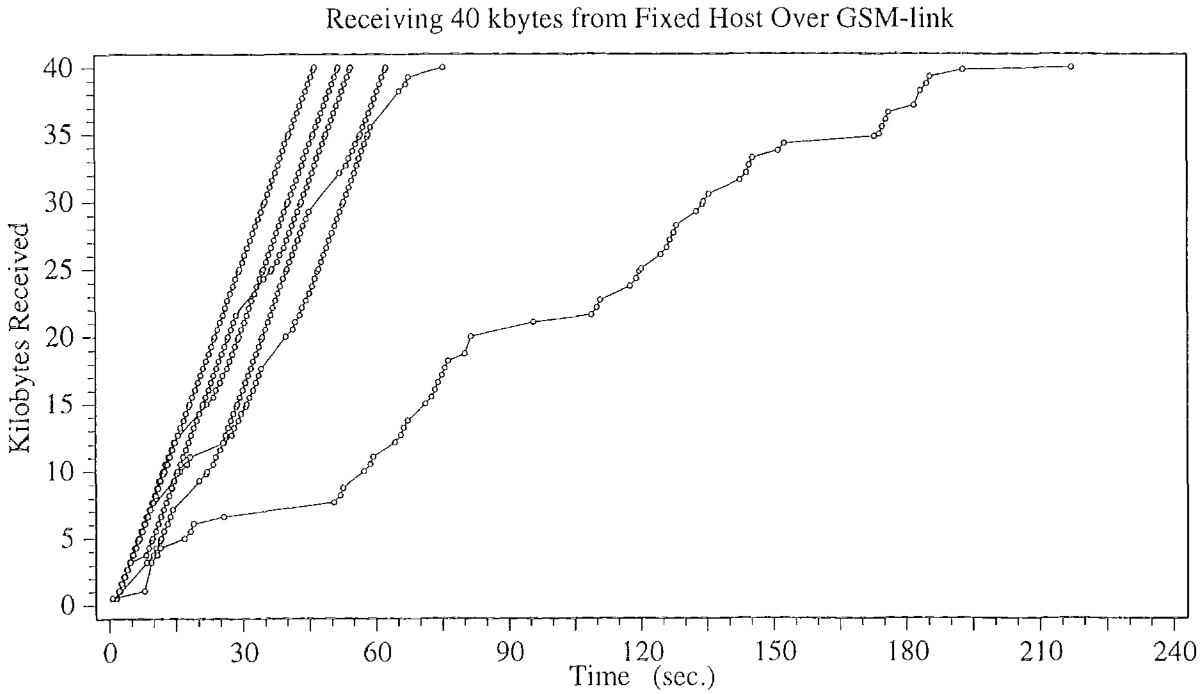

Figure 7: Example Traces of Bulk Data Transfer over GSM Link

In the GSM environment the estimates are 880 and 890 bytes per second, respectively. The difference between the environments is primarily due to the modem compression. In the NMT environment the transfer direction has a notable effect on the maximum sustainable transfer rate. Transferring data from the mobile node to the fixed host is slower than transferring data in the opposite direction. An explanation can be sought in the heterogeneity of implementations used in various parts of the system: data communication, buffering techniques, etc. The power of the sender also has some influence on the failure rate. Some kind of asymmetry is to be expected in future systems too. However, the amount and direction cannot be predicted. In the GSM environment the transfer direction had no influence of practical importance.

The effective transfer rate depends on the quality of the wireless link. In the GSM environment we have observed that in good conditions (field strength 3–4, as displayed by the GSM phone) the effective transfer rate is 80–100 % of the estimated maximum sustainable transfer rate. When the quality degrades (field strength 1–2), the effective rate falls down to 10–50 % of the maximum. According to our experience, the corresponding percentages in the NMT environment are 70–95 % and 10–30 %. Based on these observations we give in Table 5 estimated amount of kilobytes that can be transferred in 15 minutes.

Table 5: Estimated Amount of Kilobytes that Can Be Transferred in 15 Minutes

Environ-ment	Link quality		
	Ideal	Good	Poor
NMT	1240	870	310
GSM	800	640	240

4. Effects of Error Control

When a wireless link is involved in data communication, the most important single factor affecting the performance is the failure behavior. Hence, we next summarize some aspects of importance.

In an NMT network the error control is vital for the data transfer to succeed at all. Without error control the response times, which normally are only a couple of seconds, can grow to several minutes. In a GSM network, when the non-transparent transfer mode is in use, the RLP protocol is used for error control. With the RLP the bit-level error rate decreases to 10^{-8}. In a wireless environment, the origins of a typical link-layer problem are the often recurring but intermittent distortions in the radio connection. Due to the fault-tolerance features of GSM and to the error correcting modems in NMT, typical transmission failures become visible only as exceptionally long delays.

The variability of response times is caused by retransmissions and, in an NMT environment, also by modem retrainings. A retraining procedure takes typically at least some 5–10 seconds; retransmissions are needed as long as the exceptional conditions in the air prevail. For an end user, a sudden exceptionally long waiting time is irritating. Notice, that even an improvement of the conditions can lead, in NMT, to modem retraining, now in order to increase the transfer speed. Unfortunately, in a request-reply operation the retraining delay will always be noticed — but not the increased transfer speed. In bulk data transfer, the failure behavior is usually not that irritating. Due to errors, the effective transfer rate decreases, but an end user will scarcely notice it until a general deterioration takes place. Infrequent delays, even long ones, have a relatively small effect on the total transfer time.

The controlling functions of the transport layer are typically designed with the failure behavior of the fixed networks in mind. For example, the TCP flow control assumes that a loss of a TCP segment is caused by a congestion problem. Therefore, actions are taken by the transport layer to reduce congestion [Jac88]. In a wireless environment the explanation for the retransmission timer going off is usually different from that in fixed networks: the TCP segment is not actually lost but delayed, or it is lost but not due to congestion rather than transient disruption of the wireless link. Therefore, the recovery actions taken by TCP may even decrease the performance. For example, if the TCP segments are only delayed retransmissions are unnecessary, since the original packets will get through anyway as soon as the radio link has recovered. Moreover, if the TCP segments are lost due the temporary distortion on the wireless link, then TCP should not take actions to exponentially back off the retransmission timeout, because after the disappearance of the distortion the wireless link is immediately capable of working at full speed. In theory the TCP congestion algorithm may even slow down the recovery [CI94]. However, as long as the radio link is so slow (9600 bps), this slow-down is not significant in practice.

Temporary disconnections of the radio link cause problems to traditional TCP/IP implementations. In case of disconnection, the link layer should be prepared to recover through reconnecting without notifying the transport layer. However, the long reconnection time may trigger the transport-layer timers. And even if these timers could be adjusted, the application-layer timers would start creating problems. Hence, even a successful recovery at the link layer may not save the application execution. The problem has been discussed in [KRA94].

We should not forget that the end user also has some possibilities to affect the quality of communication service. The error rate is connected with the field strength and with reflections, which may be quite sensitive to external circumstances (see Chapter 2.3). Hence, there are situations where the user should pay some attention to where and when he or she tries to use a mobile system. In some coverage areas the field may be weak, or there may exist strong reflections, so that moving the mobile equipment to some other place can have a substantial effect on the performance of the system. For example, in the bulk data transfer tests with the GSM system we could observe a decrease of transfer rate from 880–890 bytes/second to 750–780 bytes/second by moving the mobile phone during the call setup a few meters in the same office. NMT seems to be essentially more sensitive to the physical environment than GSM. In our steel-and-concrete departmental building we had no problems with voice calls (GSM or NMT). However, we observed some clear differences between the offices in the quality of service of GSM data calls. Moreover, NMT data calls were possible only in certain offices.

There are also some performance factors which are out of reach for a normal end user. These include the differencies in the efficiencies of the telephone switching centers and exchanges, the capacities of the base stations, bulk arrivals of customers — for example after the arrival of an airplane — and other similar phenomena.

5. Summary and Further Research

In this study we have characterized the behavior of data communication where a mobile workstation is connected to the fixed network through a cellular telephone system. We performed two series of measurements, one using the analog NMT system, the other with the digital GSM. We assumed that the mobile workstation was located in a "normal office environment".

The most important performance results with corresponding conclusions can be summarized as follows:

- the connecting time is rather long, usually half a minute, hence, use of disconnect/reconnect operations to achieve savings in line costs should be considered carefully;
- the round-trip time is rather long (NMT: normally about 0.5 seconds, GSM: about 1 second), hence, echoing over the wireless link should be discouraged (a possible exception: NMT with all possible compressions on);
- the response time in an interactive operation with a modest amount of data exchange is short, about 1–3 seconds, hence, communication patterns from traditional network environments can be adapted;
- files can be transferred, for the time being, at a rate of 1300–1400 bytes/second (NMT with modem compression) or 700–900 bytes/second (GSM without compression), hence, only files of a modest size (say, some hundreds of kilobytes) should be transferred over the wireless link and in any case, use of data compression is advisable;
- the radio link is reliable and stable in its behavior, but only in "normal circumstances"; it is also very sensitive to noise and interference, and this sometimes leads to exceptionally long delays; hence, the traditional control mechanisms, like the flow control and the application–level timers, may prevent the use of traditional applications in a mobile environment; an example: transporting large files with the ftp may turn out to be painful or even impossible;
- the sensitivity of the radio link should have an effect on the design of applications aware of mobility (new fault-tolerance features, new functionality) and on the end-user behavior (what can be done, where and when).

As we have pointed out several times, the measurements were taken in a "normal office environment". It is easy to figure out what deviations from "normality" can cause: the rate and length of the "exceptionally long delays" increase, and so does the probability of radio link disconnections. To what extent this effect realizes is impossible to say in general — it depends on the local circumstances. The only general remark we are able to give is that the GSM system seems to be fairly robust: if the phone is located in an area with a low field strength, then the data transfer takes place rather fluently.

Finally, we want to present some remarks about future research in this area. First of all, measurements using a "physically mobile" workstation should be undertaken, for example, in a car or in a train, moving 50–150 km/h. However, in these kinds of studies it is very difficult to specify under which circumstances the results have been obtained, otherwise than, for example, "on a highway which, according to the operator, is covered by GSM".

A second extension concerns the network configuration. Our measurements were all done with a configuration where the mobile node and the fixed server were directly connected to each other (see Figures 1 and 2). However, in a "normal" configuration the server host resides somewhere else in the fixed network. The addition of different kinds of network components will obviosly change the client-server communication behavior.

In our tests there was only one active connection over the wireless link at a time. The behavior of the link may, however, be different, if there are several connections simultaneously active. As a matter of fact, the capacity of a wireless link is so low, that trying to sustain several active connections may easily deteriorate the performance to an unacceptable level.

Acknowledgments

The Mowgli research project is supervised by professor Martti Tienari and funded by Digital Equipment Corporation, Nokia Cellular Systems, Nokia Mobile Phones, Telecom Finland, and the Finnish Ministry of Education.

References

[CI94] R. Cáceres and L. Iftode. The Effects of Mobility on Reliable Transport Protocols. In *Proc. IEEE 14th International Conference on Distributed Computer Systems*, pages 12–20, Poznan, Poland, June 1994. IEEE, IEEE Computer Society Press.

[Con93] CDPD Consortium. *Cellular Digital Packet Data System Specification, Rel 1.0 Books I–V.* CDPD, 1993.

[ETS90] ETSI. Quality of Service. GSM specification 02.08, version 3.0.0., ETSI/TC GSM, March 1990.

[Jac88] V. Jacobson. Congestion Avoidance and Control. In *Proc. ACM SIGCOMM '88 Symposium on Communications Architectures and Protocols*, pages 314–329, Stanford, Calif., August 1988.

[KRA94] M. Kojo, K. Raatikainen, and T. Alanko. Connecting Mobile Workstations to the Internet over a Digital Cellular Telephone Network. Report C-1994-39, Univ. of Helsinki, Dept. of Computer Science, Helsinki, Finland, September 1994. Published also in Proc. Workshop on Mobile and Wireless Information Systems (MOBIDATA), Rutgers University, NJ, Nov 1994.

[Mob94] *Mobile Communications International, Issue 18*, Summer 1994.

[MP92] Michel Mouly and Marie-Bernadette Pautet. *The GSM System for Mobile Communications.* Mouly and Pautet, 1992.

[PJG89] J.D. Parsons, D. Jardine, and J.G. Gardiner. *Mobile Communication Systems.* Blackie, Glasgow, UK, 1989.

[Rah93] M. Rahnema. Overview of the GSM System and Protocol Architecture. *IEEE Communication Magazine*, 31(4):92–100, April 1993.

[Rom88] J. Romkey. A Nonstandard for Transmission of IP Datagrams over Serial Lines: SLIP. Request for Comments 1055, Network Information Center, June 1988.

[Sch87] B. Schneiderman. *Designing the User Interface.* Addison-Wesley, 1987.