Analytic Modelling of Single Link, Multi-LAP Connections with Application to the ISDN User-Network Access

Wolfgang Fischer
University of Stuttgart
Institute of Communications Switching and Data Technics
Seidenstraße 36, D-7000 Stuttgart, F.R.G.

Abstract
For the ISDN user-network interface the layer 2 of the D-channel protocol is characterized by a set of independent Link-Access Procedures (LAPD) [5] sharing a common transmission medium. The performance investigation of this protocol which is presented in this paper takes account of two aspects of its application: signalling data transfer and packetized user-data transfer. Approximate solutions are outlined and their accuracy is shown by comparison with simulation results.

1 Introduction
The user-network interface of the ISDN is characterized by a number of B-channels for user-data transmission and a separate D-channel for the transfer of signalling information (s-data) for the use of the B-channels.

For most applications the "Basic Access" will be used with 2 B-channels and 1 D-channel with 16 kbps. The B-channels provide a circuit switched connection between a Terminal Equipment (TE) and another TE, a server module of the network providing higher layer functions of the ISDN, or a packet handler. For packet switching purposes in the ISDN a B-channel is switched to a packet handler which transmits the packetized data to the packet switching subnetwork.

Transmission of information via the D-channel occurs in a frame-oriented manner. Due to the fact that this channel will have a very low utilization on average, it offers a good opportunity for using it also for low-rate packetized user-data (p-data), as e.g. interactive data or teleaction information. The disadvantage of this approach is the fact that these data will affect the signalling performance leading to higher system response times.

This paper focuses on an analytic performance investigation of the D-channel protocol for signalling influenced by user-data, and for user-data under special consideration of the multi-LAP capability.

2 Modelling and Analysis of Signalling Performance
2.1 Modelling
Based on the well known reference configuration defined by CCITT [3] and regarding real implementations which have been published (e.g. [7,8,9]) a queueing model for the ISDN D-channel access has been developed in order to investigate the signalling performance.

Only one Basic Access has been modelled in detail whereas the other ones are represented by appropriately chosen traffic streams.

On the left hand side of Figure 1 a number of terminals (TE) is represented, each by one LAP. A LAP (Link Access Procedure) is specified according to the HDLC Balanced Classes of Procedures (LAPB). Each LAP has a pair of buffers connecting it with the standardized S-bus.

The channel access procedure of the S-bus is a kind of CSMA/CD protocol [4]. It may approximately be represented as a priority polling system without switchover times where a cyclic polling takes place for each priority class. (a more detailed analysis of this access procedure will be published in a forthcoming paper [10])

On the right hand side of Figure 1 the processor handling layer 2 functions is depicted. In most applications it is not dedicated to a single subscriber line but has to serve up to 8 Basic Access lines. In our processor model for each information type (s or p) one process will handle all traffic from the channels as well as from the layer 3 (L3). The process handling s-information has a higher nonpreemptive priority. The queues ahead of each process are served according to a cyclic polling strategy with limited-1-service.

Each of the processes contains as many LAP procedures of its information type as in all connected terminals together.
The channel is modelled by a service phase representing the transmission time of a frame $T_T$ and the frame error probability $p_P$.

The signalling traffic produced by the small number of terminals connected to a single subscriber line cannot be described by a stochastic process in a simple way because of the fixed scenario which is proceeded for instance as the result of each call setup.

For signalling purposes a window size of 1 is defined; thus, a handshake protocol is performed at layer 2. In Figure 2 such a scenario for call setup is shown where some time values are indicated which are to be determined. A more detailed explanation of this scenario can be found in [7,8,9].

All these times consist of a sum of processing times and waiting times, where the main problem is the determination of these waiting times.

2.2 Analysis

As long as we may assume that terminals connected to one basic access are signalling only independently, the arrival processes of s-data can be approximated by Poisson processes. For the priority systems (channel access and layer 2-processor) we may even assume that an arriving s-data unit only needs to wait until the service of a p-data unit is finished. Thus, we merely have to add the components (only mean values!) in order to obtain the total signalling delay. As a another approximation also the arrival process of p-data can be modelled as a Poisson process.

In the case where it is natural that the signalling procedures of more than one terminal are triggered by a common event as indicated by the scenario we have to analyse the behaviour of the system for a single batch arrival and all the arrival processes resulting from it. This behaviour is highly dependent on the relationship between the different service times and we have to investigate the sequence of events occuring after that batch has arrived.

Let us for instance consider the call setup scenario: After waiting for the end of the transmission of a p-data unit, all SABM-frames are transmitted in an uninterrupted sequence to the ET, if we assume that all TEs are reacting with the same delay. Reaching the ET these frames follow a deterministic arrival process. Whether this sequence is served uninterruptedly depends upon the residual service time the first frame meets, the relation between the service time of one SABM-frame and its transmission time, and the probability that s- or p-data units from other basic accesses arrive within that time.

Whether the UA-frames are created directly by the layer 2 as response to the SABM-frames, or triggered by a layer 3 process will determine the further behaviour.

As can easily be seen, these considerations are highly dependent upon the actual configuration and the timing relationships, thus no simple approach can be given which is valid for the general case.
2.3 Numerical examples

In order to show the accuracy of the above decomposition approach Figure 3 depicts the mean call-delay for incoming calls as seen from the ET versus the channel load by p-data. The simulation results are given with their 95%-confidence intervals. The analytic results are in good agreement with the simulations.

The following parameters for a fully symmetric system have been assumed:

- 4 Basic accesses served by one layer 2 processor
- Exponentially distributed frame length for p-data with 200 bytes mean
- 4 ms processing delay (constant) for each frame
- Frame lengths of s-data according to [5,6]
- Negligible error probability.

3 Analysis of User-Data Performance

3.1 General approach

In order to investigate the basic mechanisms of our multiple-LAP system we use a simplified model where processing delays are aggregated into the propagation delay. But as can easily be seen from the following derivation, also arbitrary processor models may be included in the analysis. Here for each direction the model consists of an admission queue, the protocol machine, the channel access, the transmission phase and the propagation phase. For the terminals the channel access is modelled as a cyclic polling system without switchover times, whereas for the exchange termination a single FIFO-queue is located ahead of the channel.

We assume a Poisson arrival process into each of the admission queues. A detailed description of the protocol mechanisms can be found in [2,5]. Here, we only mention the following facts:

- The first transmission error of one information frame (I-frame) will be corrected using "Reject-recovery" if no timeout occurs before. All other retransmissions of this frame are invoked by "Timeout-recovery".
The retransmission timer will be started each time an I-frame is put into a channel access queue.

The number of retransmissions of I-frames is not limited.

Each I-frame will be acknowledged by a supervisory frame (S-frame) as long as the protocol machine is not in a recovery state.

The "inter-error-time" for single bit errors is distributed exponentially.

The error probability of an S-frame is neglected due to its short framlength of 56 bits.

The flow time through the layer 2 consists merely of admission-delay + channel-access-delay + effective transmission time + propagation-delay, where the "effective transmission time" is the time an I-frame needs from the first time leaving the access queue until it is successfully transmitted and leaves the transmission phase.

For the analysis we use the notation given in Table 1.

### 3.2 Waiting times

In a first step we calculate the waiting times in each of the access queues. In order to be able to do this it is necessary to compute the effective arrival rate of I- and S-frames. This may be done by the following approximate formulas which take into account the recovery mechanisms of the protocol.

---

Figure 3: Call delay for incoming calls

![Call delay for incoming calls](image)

Figure 4: Queueing model for user-data performance investigations

![Queueing model for user-data performance investigations](image)
\[ N \]
Number of terminals, each handling one LAP

\[ \lambda_{i,i} \]
original arrival rate of I-frames

\[ \lambda_{i,\text{corr},i} \]
increased arrival rate of I-frames

\[ \lambda_{S,i} \]
arrival rate of S-frames

\[ T_{\text{out},i} \]
Timeout value of retransmission timer

\[ p_{F,i} \]
Frame error probability of I-frames

\[ T_{T,I,i} \]
Transmission time of an I-frame

\[ T_{T,S} \]
Transmission time of an S-frame

\[ T_{A,i} \]
original interarrival time of I-frames

\[ w_i \]
Window size

\[ T_e \]
"inter-error-time"

\[ T_{\text{prop}} \]
Propagation delay

\[ T_{w,i} \]
Waiting time in channel access queue

\[ \rho \]
Total occupancy of a service phase

\[ \rho_i \]
Occupancy of a service phase by frames of LAP \( i \)

\[ F_X(t) \]
Distribution function of random variable \( X \)

\[ \Phi_X(s) \]
Laplace-Stieltjes transform of \( F_X \)

\[ E[X^k] \]
kth moment of random variable \( X \)

\[ \text{VAR}[X] \]
Variance of random variable \( X \)

\[ c_i^2 \]
Squared coefficient of variation of random variable \( X \)

(index \( i \) means LAP \( i \), * indicates partner LAP)

Table 1: Notation used throughout the analysis

\[ \lambda_{I,\text{corr},i} \approx \lambda_{i,i} \left( 1 + p_{F,i} \left( (1 - p_{F,i}) \sum_{k=0}^{w-2} p_{F,i}^k (k + 1) + 1 + (1 - p_{F,i})(1 + \min(w - 1, \lambda_{i,i} T_{\text{out}})) \sum_{k=1}^{2} p_{F,i}^k k \right) \right) \]  \hspace{1cm} (1)

\[ \lambda_{S,i} \approx \lambda_{i,i} \left( 1 + p_{F,i} + \frac{1 + \lambda_{i,i} T_{\text{out}}}{1 - p_{F,i}} \right) \]  \hspace{1cm} (2)

\[ p_{F,i} = P\{T_e < T_{T,I,i}\} = \int_{t=0}^{\infty} (1 - e^{-t/T_e}) dF_{T,T,I,i}(t) \]  \hspace{1cm} (3)

From \( \lambda_{I,\text{corr},i} \) and \( \lambda_{S,i} \) the total arrival rates into each queue and the effective service time distribution for each of the transmission phases may be determined.

If we assume that each I-frame is immediately acknowledged by an S-frame, the fraction of the propagation phases' output process representing I-frames is serving &I input process for the following channel access queue. Assuming a Poisson process for the remaining part of the input process we are able to determine the variability of each input process using Kühn's method [13].

The mean waiting time at ET-side is computed using the formula of Krämer and Langenbach-Bels [12] whereas for the polling system we use an approximation recently published by Boxma and Meister [1].

\[ E[T_{\text{w},i}] = E[T_{\text{w}}] \frac{1 - \rho + \rho_i}{1 - \rho + \rho_i \sum_{j=1}^{N} \rho_j} \]  \hspace{1cm} (4)

Here \( E[T_{\text{w}}] \) is the mean waiting time in the corresponding GI/GI/1- FIFO system with the same total arrival rate. (Boxma and Meister have set up this approximation only for Poisson input, but since in most cases we are not too far from that, it appears reasonable to use this extension.)

For the total output process of the polling system we assume that it is identical to that of the corresponding GI/GI/1-FIFO system. The arrival processes into all the queues are now computed iteratively.

Since we are not able to determine the second moments of the waiting times, we will approximate them by the second moment of an exponential distribution.
3.3 Effective Transmission Time

In the following step the effective transmission time is determined according to the concept of the "virtual transmission time" by Bux, Kümmel and Truong [2]. This method uses a phase type representation of an effective server which is characterized by two moments of its service time distribution function. There are 3 different time values contributing to this effective service time:

- $T_{T,1}$: Transmission time without errors
- $T_1$: Additional time needed for the first correction
- $T_2$: Additional time needed for subsequent corrections

From this representation the first two moments of the effective transmission time can be found to be (the indices $i$ are dropped for reasons of simplicity)

$$E[T_{eff}^1] = E[T_{T,1}] + p_F E[T_1] + \frac{p_F^2}{1-p_F} E[T_2]$$  \hspace{1cm} (5)

$$E[T_{eff}^2] = E[T_{T,1}^2] + p_F E[T_1^2] + \frac{p_F^2}{1-p_F} E[T_2^2] + 2p_F E[T_{T,1}] E[T_1] + 2 \frac{p_F^2}{1-p_F} E[T_1](E[T_{T,1}] + E[T_1] + \frac{p_F}{1-p_F} E[T_2])$$ \hspace{1cm} (6)

$$2 \frac{p_F^2}{1-p_F} E[T_1](E[T_{T,1}] + E[T_1] + \frac{p_F}{1-p_F} E[T_2])$$ \hspace{1cm} (7)

3.3.1 Derivation of $T_1$

Mainly in the low-load region we have to distinguish between two cases: Reject-recovery and Timeout-recovery. Which of both actually takes place depends on whether the timer runs out before the next frame arrives. For Reject-recovery:

$$T_1 \geq T_{rej} = T_{A,cut} + 2T_{prop} + T_{T,S} + T_{T,1} + T_w + T'_w$$  \hspace{1cm} (8)

$T_{A,cut}$ is the interarrival time of I-frames where the original interarrival time distribution is truncated at $T_{out}$. Using the external arrival process the $k$-th moment of $T_{A,cut}$ follows from

$$E[T_{A,cut}^k] = \frac{1}{F_{T_A}(T_{out})} \int_{t=0}^{T_{out}} t^k dF_{T_A}(t)$$  \hspace{1cm} (9)

We have to use this truncated distribution since we know that if we have to deal with Reject-recovery the timer cannot have run out. For Timeout-recovery:

$$T_1 \geq T_{TO} = T_{out} + 2T_{T,S} + 2T_{prop} + T_w$$  \hspace{1cm} (10)

Assuming independence between all time values forming $T_{rej}$ and $T_{TO}$ we may add their means and variances in order to obtain the first two moments of $T_{rej}$ and $T_{TO}$.

For each disturbed I-frame which has arrived after the one which was the reason for $T_1$ an additional $T_{A,cut}$ adds to $T_1$ if the timer does not run out before.

Defining $p_{TO}$ as $P\{T_{out} < T_A\}$ we may write

$$T_1 = p_{TO} \cdot T_{TO} +$$

$$(1 - p_{TO})(1 - p_F) \cdot T_{rej} + $$

$$(1 - p_{TO})p_F p_{TO} \cdot (T_{TO} + T_{A,cut}) + $$

$$(1 - p_{TO})^2 p_F (1 - p_F) \cdot (T_{rej} + T_{A,cut}) + $$

$$... + $$

$$(1 - p_{TO})^{w-1} p_F^{w-1} \cdot (T_{TO} + (w-1)T_{A,cut}) + $$

$$(1 - p_{TO})^w p_F^w (1 - p_F) \cdot (T_{rej} + (w-1)T_{A,cut})$$  \hspace{1cm} (11)

From the following Laplace-Stieltjes transform the first two moments of $T_1$ can be derived.

$$\Phi_{T_1}(s) = (p_{TO} \Phi_{T_{TO}}(s) + (1 - p_{TO})(1 - p_F) \Phi_{T_{rej}}(s)) \sum_{j=0}^{w-1} (p_F(1 - p_{TO}) \Phi_{T_{A,cut}}(s))^j$$  \hspace{1cm} (12)
3.3.2 Derivation of \( T_2 \)

For the second and all further errors of the same I-frame we have to deal only with Timeout recovery. Here \( T_2 \geq T_{TO} \). Since the timer is restarted with every I-frame put into the channel access queue it may happen that a series of further frames (up to \( w - 1 \)) will be transmitted before the timer is able to run out.

Thus we may write

\[
T_2 = T_{TO} + (1 - p_{TO})p_{TO} \cdot T_{A, cut} + 2(1 - p_{TO})^2 p_{TO} \cdot T_{A, cut} \\
+ \ldots + (w - 2)(1 - p_{TO})^{w-2} p_{TO} \cdot T_{A, cut} + (w - 1)(1 - p_{TO})^{w-1} \cdot T_{A, cut}
\]  

(13)

We do not need to worry about the possibility that these subsequent frames are disturbed since they are discarded by the partner LAP anyway.

For the arrival process of these subsequently transmitted frames the original arrival process of I-frames is taken although we know that there may also be frames from the retransmit buffer of the LAP resulting in a somewhat different arrival process.

From the following transformation again the first two moments of \( T_2 \) are derived

\[
\Phi_{T_2}(s) = \Phi_{T_{TO}}(s) \left( p_{TO} \sum_{i=0}^{w-2} (1 - p_{TO})^i \Phi_{T_{A, cut}}(s) + (1 - p_{TO})^{w-1} \Phi_{T_{A, cut}}^{w-1}(s) \right)
\]  

(14)

3.4 Admission Delay

If \( w \) messages have been sent by the LAP considered, and no acknowledgement has been received yet, the transmission of arriving I-frames is stopped. These I-frames which cannot immediately be transmitted have to wait in the admission queue ahead of the protocol machine.

For the analysis of the admission delay we use a model which can be derived from that published by Reiser [15].

![Figure 5: Mapping of an admission station](image_url)

In the left hand sindmodel of Figure 5 the 'diamond' puts a customer from the admission queue into the transmission queue if there is at least one credit in the credit queue. The customer takes the credit and sends it back when leaving the server. In the closed network formed by the transmission queue, the server and the credit queue, exactly \( w \) credits are circulating.

This can be mapped into an equivalent system with a queue ahead of the server which is subdivided into a series of an infinitely large queue (the admission queue) and a pad with \( w - 1 \) places (the transmission queue).

Thus for a GI-server and Poisson input we obtain a standard M/GI/I-FIFO system with a partitioned queue.

In order to evaluate the waiting time in the admission queue we first establish an approximation for the state probabilities in the system, if the GI-server is given only by its first two moments [11].

Observing a waiting place \( k \) in the M/GI/1 system the mean duration \( t_k \) an arbitrary customer spends there is

\[
t_k = p_k r_k + E[H] \sum_{i=k+1}^{\infty} p_i = p_k r_k + E[H] \left( 1 - \sum_{i=0}^{k} p_i \right)
\]  

(15)

where

- \( p_k \): probability of state \( k \)
- \( r_k \): mean residual life time of the customer in service at the arrival instant
- \( H \): service time

With Little’s law we obtain the occupancy \( \rho_k \) of the waiting place \( k \)

\[
\rho_k = \lambda t_k
\]  

(16)
\[ p_k = \rho_{k-1} - \rho_k = \frac{1 - \rho + \lambda k \prod_{i=0}^{k-1} r_i}{\prod_{i=1}^{k} (1 - \rho + \lambda r_i)} \]  

where \( \rho = \lambda E[H] \) is the total server occupancy.

From the fact that \( \rho_k = \sum_{i=k+1}^{\infty} P_i \), as a side result we obtain an expression for the conditional mean residual life time which has also been found by Mandelbaum and Yechiali [2] using a more formal derivation.

\[ r_k = (1 - \rho) \frac{1 - \sum_{i=0}^{k} P_i}{\lambda P_k} \]  

Since in general we do not know the service time distribution in order to obtain \( r_k \) we approximate \( r_k \) by the forward recurrence time of the service time and obtain an approximate state distribution.

\[ p_k \approx \begin{cases} 1 - \rho & \text{if } k = 0 \\ \frac{2(1 - \rho)}{1 + \frac{c_H}{\lambda}} \left( 1 - \frac{c_H}{1 - \frac{c_H}{\lambda}} \right)^k & \text{if } k > 0 \end{cases} \]  

This approximation is exact for M/M/1 and yields the exact mean waiting time for M/GI/1.

By some standard arguments we obtain the mean waiting time in the admission queue

\[ T_{w, \text{adm}} \approx \frac{E[H]}{1 - \rho} \cdot \frac{(\frac{2}{\lambda} (1 + c_H))^{w}}{1 - \frac{c_H}{1 - \frac{c_H}{\lambda}}^{w-1}} \]  

What remains to do is to establish the first two moments of the service time distribution function for this virtual server.

The first moment is merely the mean service time of the slowest server in the acknowledgement chain consisting of both channels. In our case this is the effective service time which is increased by the service of all frames other than the considered class of I-frames arriving within this increased service time.

\[ E[H] = \frac{E[T_{eff}]}{1 - \rho + \lambda_{I,I} T_{T,I,I}} \]  

where \( \rho_{T} \), is the total occupancy of the transmission phase. In order to obtain the second moment we look at the residual time an arriving customer observes until the next acknowledgment message leaves the chain.

Let us define

\[ \begin{align*}
    P_{\text{prop}} &= 1 - e^{-\lambda_I E[T_{\text{prop}}]} \\
    P^* &= 1 - e^{-\lambda_I E[T_{\text{adm}}]} \\
    \rho T_s &= \lambda_{I,I} E[T_{T,s}] \\
    P_{T,I,I} &= \frac{1}{\lambda_{I,I} E[T_{T,I,I}]} \\
    T_X &= \text{Forward recurrence time of } T_X \end{align*} \]

Assuming independence between the probabilities mentioned above we obtain the mean forward recurrence time of our service time distribution function

\[ E[H^*] \approx \frac{1}{k} \left( P_{\text{prop}} \cdot E[T_{\text{prop}}] + (1 - P_{\text{prop}}) P_{T} \cdot (E[T_{\text{prop}}] + E[T_{T,s}]) + (1 - P_{\text{prop}})(1 - P_{T}) P^* \cdot (E[T_{\text{prop}}] + E[T_{T,s}] + E[T_{\text{prop}}]) + (1 - P_{\text{prop}}) P_{T} \cdot (E[T_{\text{prop}}] + E[T_{T,s}] + E[T_{\text{prop}}] + E[T_{T,s}]) + (1 - P_{\text{prop}})(1 - P_{T}) P^* \cdot (2E[T_{\text{prop}}] + E[T_{T,s}] + E[T_{T,s}] + E[T_{T,I,I}]) \right) \]  

where \( k \) is a normalising constant defined as the sum over all coefficients of the time expressions above.

Then the squared coefficient of variation is given by

\[ c_H^2 = \frac{2E[H^*]}{E[H]} - 1 \]
3.5 Numerical Examples

In order to show the accuracy of the analysis approach we have chosen the following parameters for fully symmetric systems:

- Transmission rate: 16 kbps
- Bit error probability: $10^{-8}$ and $10^{-5}$
- Constant length of information field: 256 bit and 1024 bit
- Constant length of S-frames: 56 bit
- Number of terminals: 1 and 4
- Window size: 3 and 7
- Propagation delay: 5 ms (constant)

![Graph](image)

**Figure 6: Mean transfer delay**

(a) $p_{bit} = 10^{-8}, w = 3$

(b) $p_{bit} = 10^{-5}, w = 7$

In the Figures 6(a,b) the results obtained by the numerical analysis are compared with results from simulations. The mean transfer delay is plotted versus the total throughput for one direction.

We observe that the results are quite accurate over a wide throughput range. Inaccuracies are mainly due to 3 effects:

1. The waiting times in the access queues are very sensitive with respect to the values of the arrival rates $\lambda_{i,corr}$, $\lambda_{S,i}$ and the relation between them. Further work is needed to characterize these rates more accurately for high error rates.

2. Mainly for one LAP the independence assumption when calculating the residual service time for the admission delay does not hold very accurately. Thus in this case a better approximation for $E[H^T]$ should be established.

3. The arrival processes from the admission queues are no Poisson processes any more. Their variance tends to be smaller.

4 Conclusion

For two different applications of the D-channel of the ISDN user-network interface approximate performance investigations have been presented. They appear to be quite helpful in assessing the influence of parameter choices for real implementations. In few cases further studies are required in order to increase the accuracy for some parameter ranges.
5 Acknowledgements

We wish to thank U. Deininger and U. Häsberle for their programming effort during the development of the simulation programs. The helpful discussions with H. Kröner are especially greatly acknowledged.

References