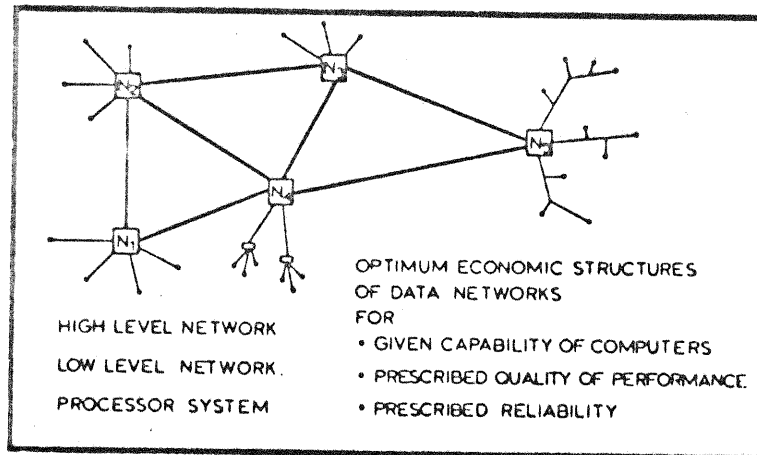


WITH COMPUTERIZED SWITCHING AND PROCESSING CENTERS

by Dr.-Ing. Ulrich Herzog
University Stuttgart, Germany

1. Problem.

The field of problems, I want to deal with in my lecture can be illustrated most easily by means of an example:
On the first slide you can see schematically a data network with five computerized centers N_1, N_2, \dots, N_5 , linked by a communication network.



The terminals, indicated by the small circles, are connected with the computers:

- directly, see node 1, 2 and 3
- via concentrators, see node 4 or
- via polled lines, as in the case of the last node No. 5.

This simple example shows already, that in modern data networks, controlled by electronic computers, we can distinguish the following 3 subsystems:

- . at first the HIGH - LEVEL - NETWORK, connecting the nodes of the network,
- . then the LOW - LEVEL - NETWORK, connecting the terminals with the HIGH-LEVEL-NETWORK
(This is only a special case of the HIGH-LEVEL-NETWORK.)

*Proc. Internat. Seminar on Telecom,
Helsinki 1970*

. and finally the PROCESSOR-SYSTEM with the computers in the nodes of the net.

This subsystem establishes the connexion between HLN and LLN and controls the traffic flow in the HLN.

In my lecture, I should like to give you a survey about the problems to be solved for network-optimization, taking into account the capability of the computers.

Of course, its not possible to discuss the manifold hardware - and software problems of electronic computers.

We shall see, however, which characteristic parameters about computer-capability are necessary for network-optimization.

A vital problem of data networks is to protect the system against breakdowns and I think, you all know, that we have ten times transmission breakdown and only one time a computer fault.

In this lecture, I have to confine myself to discuss only network configurations and operation modes guaranting a prescribed reliability in case of cable interrupts.

The interesting field of error-correcting and error-detecting-codes must be omitted.

Therefore, we will look for OPTIMUM ECONOMIC STRUCTURES OF DATA NETWORKS

for . given capability of computers
 . prescribed quality of performance
and . prescribed reliability in case of heavy load or
 interrupt of cables.

In our Institute at Stuttgart University we have been investigating such problems concerning communication networks for about ten years and for the last 2 years also problems concerning data networks.

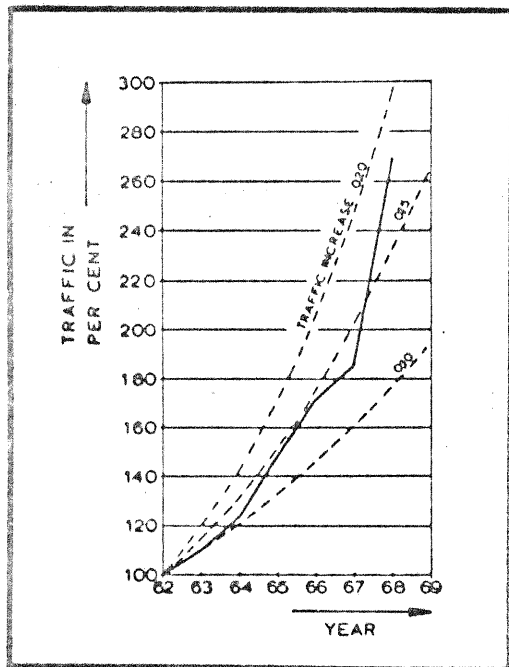
The 10 to 12 scientists require per year about 600 to 700 hours of computer time with fast digital computers for calculations and simulations, which amounts to an expenditure of about 700.000 to 900.000 DM.

Hearing such figures one must consider whether such a high expenditure is really justified.

Now, the economic importance of such intensive investigations can be seen immediately, persuing traffic statistics of the last years.

- . The increase of data traffic is - like teletraffic - exponentially and there is no symptom for beginning saturations.
- . We can persue, that the average yearly increase of data-traffic is about 10 to 20 percent, and sometimes even more.

The next slide demonstrates this tendency:



The abszissa shows the period between 1962 and 1969 and the ordinate the amount of traffic, defining the traffic of 1962 as 100 per cent.

The dashed lines are theoretical curves for a yearly increase of 10, 15, and 20 per cent.

There between, you can see the traffic development of SITA-Computer-Centers:

At first we have a traffic increase of about 10 per cent, then 15 per cent.

Between 1967 and 68 we have a violent development, reached by automation of some SITA-Centers.

If we want to manage such enormous traffic increases, an important expense for traffic control and leased lines will be necessary.

Therefore, the aim of our research team is, to investigate data networks and to find optimum economic structures.

In the second part, I want to give you a survey about the STRUCTURE AND OPERATION MODE OF DATA NETWORKS, then I will go on to the ANALYSIS.

And in the 4th part a survey is given about the SYNTHESIS OF OPTIMUM STRUCTURES.

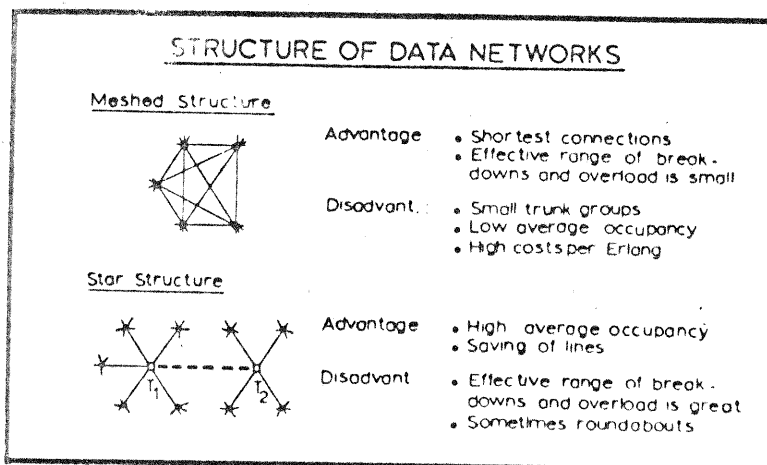
The practicability of analysis and synthesis depends considerably on the chosen traffic model: The more realistic the description of traffic, the better the results!

There is given a short survey on the stochastic behavior of traffic flow in part 5. In addition it is shown, which characteristics of real traffic and computers have to be measured or estimated.

Finally, it is shown, by what means network optimization can be done.

Now, having given an abstract of my lecture, we can start with the second part,

the STRUCTURE AND OPERATION MODE OF DATA NETWORKS.



Two basic structures are known for network configurations:

The meshed structure and the star structure.

At first the meshed structure: Each node corresponds to a computer center with several terminals. Every node is connected with each other.

- The advantages of this structure are, that the traffic is handled on the shortest connection. And the breakdown of a trunk group does not influence the traffic of the other directions, if no variable routing is allowed.
- The disadvantages are, that the traffic between two nodes is relatively small and hence the average occupancy low. Therefore cost per traffic unit are very high.

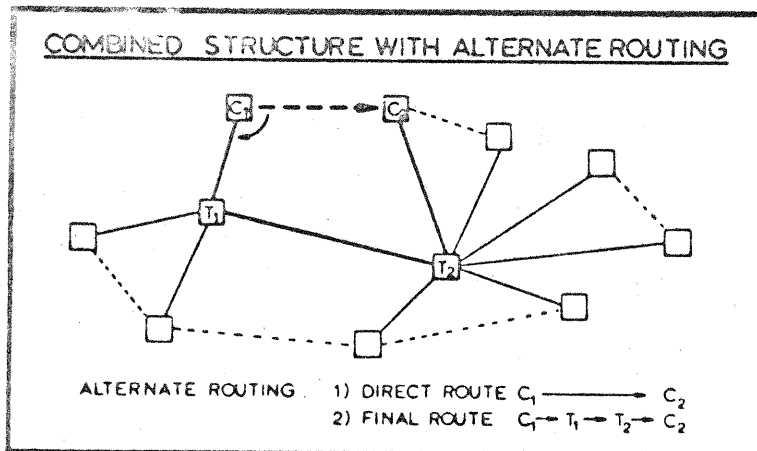
The second basic structure is the star structure: In this case all traffics from left to right in the slide are pooled in the so called TRANSIT CENTER T_1 .

We have only one long distance trunk group from T_1 to T_2 and the traffics are distributed in the transit center T_2 to the several destinations.

- . The advantage of this structure is the high average occupancy of long distance trunk groups and therefore the saving of trunks and money.
- . The disadvantages are, that the breakdown of a transit group is catastrophic and round about ways are often not avoidable.

Large data networks are mainly star networks in their basic structure. However, to avoid the disadvantages of the star net, a network with some direct routes and alternate routing is very advantageous.

The next slide shows an example:



The traffic from Computer Center C_1 to C_2 is handled for the most part like mesh-trunking directly, i.e. 70 - 90 per cent of the traffic is carried on a very cheap and economic way. The remainder part is pooled with other traffics as in the star-net.

Therefore, messages from Computer Center C_1 to C_2 are offered at first to the direct route.

If this route is occupied, the message is directed via T_1 , T_2 and then to the destination C_2 .

Hence the network with alternate routing combines the advantages of both, meshed net and star net. The reliability in case of heavy load or interrupts is very good.

Furthermore the additional expense for common control in the Computer Center is not too high, because the routing strategy only takes into consideration the traffic distribution in one node. A second method, even more suitable with respect to the traffic flow is the method of ADAPTIVE ROUTING.

In that case, traffic distribution of the neighbored nodes is taken into account, too.

At fixed points of time the traffic distribution from all groups of one node is communicated to the neighbored centers.

This additional information allows a more flexible routing strategy. Local overload and breakdown of transmission ways are taken into account automatically.

However, the additional expense to transmit these informations and to control the traffic flow is considerably higher than in the case of alternate routing.

May I summarize: Combining the advantages of the basic structures - meshed net and star net - we have found the METHOD OF ALTERNATE ROUTING. In that case, only traffic distribution of one node is used for routing.

Even more suitable with respect to the traffic flow is the METHOD OF ADAPTIVE ROUTING, because we have more information about the instantaneous traffic distribution of the whole net.

A logical continuation of these routing strategies would be a fully centralized network with common control, finding the very optimum strategy for each message.

However, the technical and economic difficulties are too high, nowadays and I think also in the next years.

ALTERNATE ROUTING has pushed his way.

However, in my opinion, the METHOD OF ADAPTIVE ROUTING will be the most effective and economic routing strategy for data networks in near and far future. The additional expense will be tolerable and the method essentially improves the grade of service and the reliability in case of heavy load or interrupt of cables.

After we have seen the importance of routing strategies, we can go on to study the ANALYTICAL METHODS.

There are two possibilities to analyze the quality of performance of data networks:

SIMULATION and MATHEMATICAL CALCULATION.

Simulation is a powerful tool to analyze complex systems.

A program is written on an electronic computer, to act as a model of the system in question. If the model is accurate, it behaves in the essential aspects like the real system and allows to study the grade of performance.

The main-advantage of simulation composed to mathematical methods is, that new and unsolved problems can be investigated in relatively short time.

It's possible to estimate fairly well the period for solution, while a success of mathematical methods is guaranteed not in each case. Therefore, simulation should be done always as a first step, if the problems were complex, and hence in data networks, too.

Today we are no longer thrown on the normal computer languages as ALGOL or FORTRAN. These languages are rather unhandy for simulation.

Special languages for simulation permit the system to be prescribed with relative ease. They can be highly flexible and can simulate almost any system configuration. A basic model, written in such a language can be steadily increased in complexity or detail until it represents very accurately the behavior of a real system.

There are several such languages available. Commonly used one's are GPSS, SIMULA, SIMSCRIPT.

The problem of a planning engineer is not only to analyze given structures. The main problem is to find optimum economic networks and for that purpose simulation is unsatisfactory. It is true that we can find by systematical variation of the structure and traffics fairly well solutions. Then, however, simulation time on computers may increase enormously and are beyond prize.

Real Optimization with reasonable expense is only possible with MATHEMATICAL METHODS, treated now.

A lot of investigations has been done in calculating the basic elements of data networks, i.e. in single server systems and multi-server systems. These basic elements have been investigated exactly with respect to various traffics and various handling disciplines.

In addition, some exact results are available for "rowed waiting", i.e. for series of successive lines, connected by switching centers.

For complex structures as data networks there are known up to now no exact solutions, which can be evaluated. Therefore, we are thrown on good approximate solutions and a good deal of our investigations are directed to this object.

Good approximate results are known for the basic structures - star net and meshed net - without variable routing. However, we have seen, that these configurations are not suitable for large data networks.

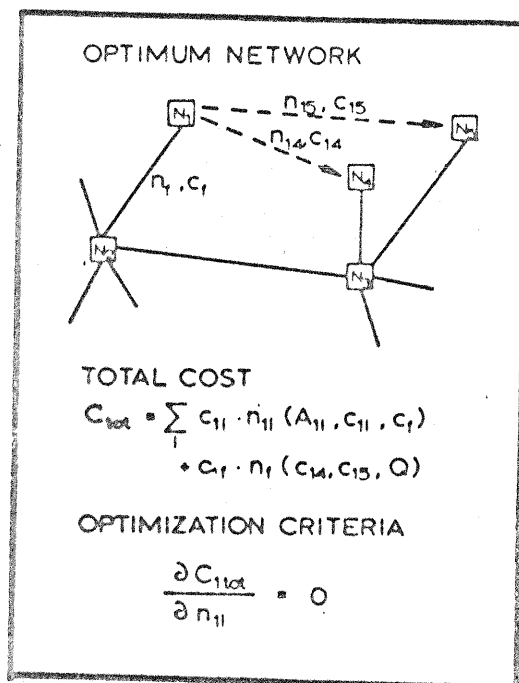
Some results are also available for systems with alternate routing, but the problem of optimum routing in a node is not solved to my knowledge and a lot of my investigations are directed of this aim.

Finally, I do not know any good approximate method for networks, operated by adaptive routing and I think, that will be a profitable problem for all traffic engineers in near and far future.

After we have seen, which network configurations can be treated mathematically up to now and when simulation is advantageous or necessary, we can go on to the

SYNTHESIS OF OPTIMUM STRUCTURES.

As we have seen before, systems with alternate or adaptive routing are superior to networks without these facilities, with respect to economy and reliability. Therefore, I want to show you the principle of synthesis for networks, operating in one of these modes. The next slide shows once more the principle structure.



Starting point for optimization is a basic network, hierarchically organized with a minimum length of lines, indicated by the solid connections Node 1 to Node 2, Node 2 to Node 3, etc.

If there is no connection between nodes, having fairly high cross traffic, we take care for additional direct routes.

In our example, two direct routes are indicated by the dashed lines Node 1 to Node 5 and Node 1 to Node 4.

Therefore, the traffic from Center N_1 to N_4 is offered at first to the direct primary channels. All traffic, not servable on this direct route will be lead to the final route via Node 2 - Node 3 to the destination N_4 .

As we know already, alternate or adaptive routing is only possible if the traffic can be handled at least on two ways. And this demand - decisive for reliability - is satisfied in our example, too! But how we can find the real optimum structure, i.e. the optimum number of channels for the direct- and final ways?

For optimum configurations we have to take into consideration

- . the cost of channels
- . the line speed and
- . the cost for processing and switching in the computer centers.

A very skilful idea, which is commonly used in telephone switching theory and which we may employ accordingly, is - to determine the cost of one channel

- . with regard to its length
- . with regard to the line speed and
- . with regard to cost for switching and processing.

Therefore, the total cost of one channel on the direct route from Node 1 to Node 4 are C_{14} , in the next direction C_{15} and at least for the final way C_f .

The total cost for this part of the net are given by the first equation, shown on the slide:

We have to add the cost of each direct route and those of the common final route.

As we can see from this equation, the number of channels n in the direct routes is a function of

- . the traffic offered A and

- the cost ratio, which depends on the cost of the direct route and the final route.

The number of channels in the final route n_f is a function of

- all previously hunted direct routes and
- the prescribed quality-parameter, e.g. the average waiting time.

By partial derivation we then get the optimization criteria.

Optimization must be done for each node of the net, i.e. we have to find the real economic optimum for the whole network by an iteration process.

L + G, the practicability of the optimization method depends considerably on the chosen traffic model:

the more realistic the description of stochastic behavior of traffic flow, the better the results, compared to reality.

Therefore, it's worth the trouble, to study in the next part the TRAFFIC MODEL.

The arrival and departure of messages correspond to a special stochastic process, which can be prescribed by methods of the probability theory.

The stochastic behavior of the messages is determined fully by means of an

- input - and termination process
- or - in an equivalent manner -
- by all moments of the traffic distribution, i.e. mean value, variance, skewness, etc.

The next slide shows the most important input - and termination processes. Of greatest use and simplicity for input- and termination processes is the EXPONENTIAL DISTRIBUTION FUNCTION, shown on the top of the slide with the mean value a_m .

However, in data networks, the termination process is often better represented by a constant message length.

These two functions are extremes of the so called Erlangian family of distribution functions, which is used to describe input- and termination process neither exponential nor constant, but in between. Stated briefly, the density functions have smaller tails than the exponential.

COMMONLY USED INPUT- AND TERMINATION PROCESSES	
EXPONENTIAL	$P(a \leq t) = 1 - e^{-\frac{t}{a_m}} \quad t \geq 0$
CONSTANT	$P(a \leq t) = \begin{cases} 1 & 0 \leq t \leq a_m \\ 0 & t > a_m \end{cases}$
ERLANG-k	$P(a \leq t) = 1 - e^{-k \cdot \frac{t}{a_m}} \sum_{\xi=0}^{k-1} \frac{(k \cdot \frac{t}{a_m})^\xi}{\xi!} \quad t \geq 0$
HYPEREXPON.	$P(a \leq t) = 1 - p e^{-2p \frac{t}{a_m}} - (1-p) e^{-2(1-p) \frac{t}{a_m}}$

A family without limitation is the mixed Erlangian. This is a weighted sum of Erlangian functions.

The simplest case is a mixture of two exponentials, called hyperexponential and also shown on the slide.

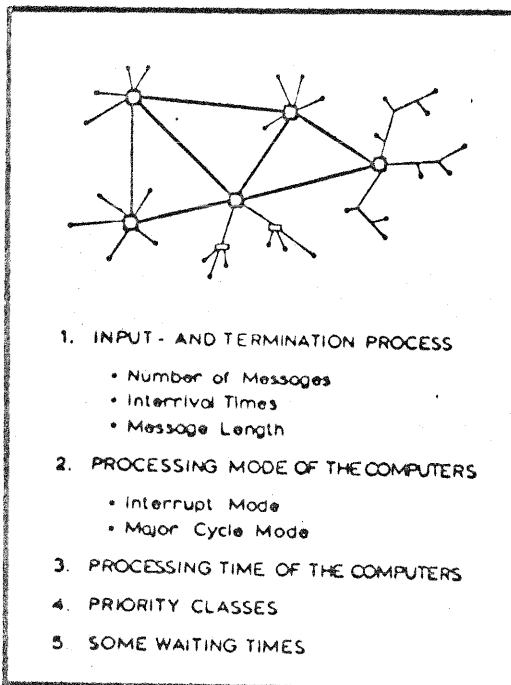
Now having seen the most important input - and termination processes, we can declare, which characteristic values of the real traffic and computers have to be determined.

At first we have to determine the INPUT - AND TERMINATION PROCESS of the data traffics, originated by the terminal devices.

We have to measure the number of messages in each direction,

- . the interarrival times
- . and the message length!

It is nearly sufficient in each case, to measure the first two moments of the distribution functions, i.e. the mean value and variance.



Numerous investigations have shown, that realistic traffic models can be found by that means.

Of great influence for traffic flow is the second point, the processing mode of the computers!

Operating in the INTERRUPT MODE, we do not have additional waiting times before the computers. If we have the MAJOR CYCLE MODE, then all input- buffers are scanned only at fixed points of time. Therefore, additional waiting times have to be expected.

Of course, PROCESSING TIME OF COMPUTERS
and PRIORITY CLASSES FOR MESSAGES
have to be taken into account, too.

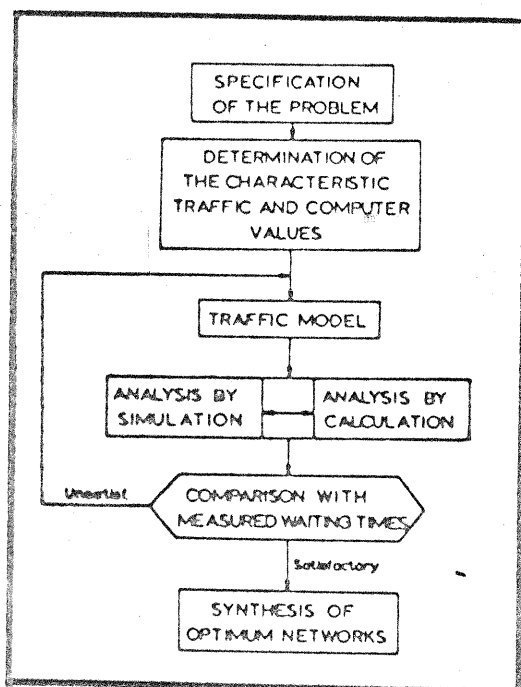
The last point is:

SOME WAITING TIMES FOR REAL TRAFFIC FLOW!

Of course, that is one of the problems to be solved. However, it's greatly helpful to have some real dates of the existing system, to check the chosen traffic model.

In the previous parts there was given a survey on the main problems of network optimization.

Finally, I want to show you, how the solution of this optimization problem should be managed:



At first we have to specify the problem, i.e., we have to determine, what demands we have to fulfil

— with respect to the quality of performance and

- with respect to reliability in case of heavy load or interrupt of cables. At this point, we have to decide already about the routing strategy

- fixed routing or variable routing
- alternate routing or adaptive routing.

Next, we have to determine the characteristic traffic and computer values, i.e. we have to measure or to estimate the traffic values and computer capability.

By means of the traffic values it is now possible to fix the traffic model and to start analytical investigations for network structures. As mentioned before, the first step will be in most cases the analysis by simulation and in a second step we use mathematical methods. Of course we have to compare both methods, if no exact calculation can be done.

The results of analysis depend considerably on the chosen traffic model: The more realistic the model, the better the results, compared to reality. Therefore, we have to compare the simulated or calculated waiting times with the measured ones.

If the comparison is unsatisfactory, we have to find a new and more realistic traffic model and we have to repeat the procedure of analysis.

If the comparison is satisfactory we can go on to find the real optimum, which is the base for an economic putting in of data networks.