WAITING TIMES WITH RANDOM SERVED QUEUE

The waiting time theories reviewed in the preceding article all refer to the case that the serving of those waiting takes place in the same order as their calls came in, so that those waiting formed an ordered queue. There occurs often, however, in telephony an order of service departing entirely from the ordered queue, whose characteristic feature is, that the selection of that waiting call which is given occupation when a device becomes unoccupied takes place very much at random. If there is only one waiting when a device becomes unoccupied, the order of service has of course no significance, as the call waiting in that case will always obtain occupation on the device coming free. Now if several subscribers are waiting, it is clear that the random selection of those waiting who obtain a free device means that some subscribers may have to wait much longer than would be the case if service proceeded according to ordered queue. On the other hand, the random service may mean that some of those waiting have a much shorter waiting time than with ordered queue. The result will obviously be that the waiting time distribution with random served queue will be flatter than with ordered queue. It is true that we can show that the mean waiting time in both cases will be the same, but the greater flatness with random served queue may be expected to have a certain significance for judging the inconvenience involved for the subscribers from waiting. The question of the properties of the distribution function of the waiting times with random served queue is therefore of very great importance. In spite of this, there has so far appeared in the literature no investigation concerning this case, which probably is not due to lack of interest but to the mathematical complications involved in the treatment. It is true that there was published in 1942 a solution of the problem by *Mellor* (bibliography 1), which is incorrect, however, and gives misleading results. The solution presented in the present article was drawn up in 1938, but has not been published before[1]) in the expectation of being able to submit easier methods of numerical evaluation, a hope that has unfortunately only been partly realised.

As already indicated, the mathematical treatment of the waiting time conditions with random served queue will be appreciably more complicated than with ordered queue. The reason for this is the following. With ordered queue the subscriber's waiting time will be independent of whether further calls come while he is waiting, and it will then only be dependent on the distribution function of the congestion transition and the number of persons already waiting at the moment when the subscriber begins to wait. With random served queue, however, the possibility of a waiting subscriber being served at any moment will be dependent on how many waiting subscribers there are at the same time. Owing to this, with computation of a subscriber's waiting time, consideration must be given not only to how many others are waiting on the arrival of the subscriber's call, but one must also consider the possibility that other subsequent calls will take part in the competition for the devices coming free. As might be expected, this condition considerably complicates the treatment.

As was explained in the preceding article, with not too small groups the distribution function of the congestion transition will be very close to exponential, even if the holding times' distribution function differs appreciably from the exponential form. It has therefore been assumed in the following that the congestion transition has a purely exponential distribution function, which indeed seems to be the only case giving the possibility of profitable mathematical treatment. In this the solution is obtained from a partially differential equation of hyperbolical type with certain boundary conditions. Unfortunately it has not been possible to produce the wanted solution of this differential equation in closed form by means of known functions. It is, however, possible to find series expansions for the wanted solution, whereby the terms are determined recurrently by means of integrations. Unfortunately it is only possible to compute a few terms with a reasonable amount of work. Nevertheless, by this method it has been found to be possible to tabulate the wanted distribution function in a sufficiently large field for the

[1]) Before 1946. *(Editor's note.)*

nential function for the same mean value as with $F(t)$ comes so near to $F(t)$ that the difference cannot be traced in the scale used for the figure. Both $F(t)$ and $F_0(t)$, however, are rather steeper than the corresponding exponential functions, so that their form factors are less (though extremely little) than 2.

Bibliography.

1. *Berkeley, G. S.:* »Traffic and Delay Formulae», Post Office Electr. Eng. J., Vol. 29, 1936.
2. *Crommelin, C. D.:* »Delay Probability Formulae when the Holding Times are Constant», Post Office Electr. Eng. J., Vol. 25, 1932—33.
3. *Erlang, A. K.:* »Probability and Telephone Calls», Nyt Tidskrift for Matematik, Annual Vol. 20, Sec. B, 1909. (In Danish).
4. *Erlang, A. K.:* »Solution of Some Probability Problems of Significance for Automatic Telephone Exchanges», Elektroteknikeren, Copenhagen, 1917. (In Danish).
5. *Molina, E. C.:* »Application of the Theory of Probability to the Telephone Trunking Problems», The Bell System Technical Journal, Vol. 6, 1927.
6. *Palm, C.:* »Calcul exact de la perte dans les groupes de circuits échelonnés», Ericsson Technics, Stockholm, Vol. 4, 1936.
7. *Palm, C.:* »Inhomogeneous Telephone Traffic in Full-Availability Groups», Ericsson Technics, Stockholm, Vol. 5, 1937.
8. *Palm, C.:* »Some Investigations into Waiting Times in Telephone Plants», Tekniska Meddelanden från Kungl. Telegrafstyrelsen, Stockholm 1937, No. 7—9. (In Swedish).
9. *Palm, C.:* »Analysis of the Erlang Traffic Formulae for Busy-Signal Arrangements», Ericsson Technics, Stockholm, Vol. 6, 1938.
10. *Palm, C.:* »A Form Factor for Judging Waiting Time Distributions», Tekniska Meddelanden från Kungl. Telegrafstyrelsen, Stockholm 1943, No. 1—3. (In Swedish).
11. *Palm, C.:* »Intensitätsschwankungen im Fernsprechverkehr», Ericsson Technics, Stockholm 1943, No. 44.
12. *Pearson, K.:* »Tables of the Incomplete Gamma-Functions», H.M. Stationery Office, London, 1942.
13. *Pollaczek, F.:* »Über eine Aufgabe der Wahrscheinlichkeitstheorie», Math. Zeitschrift, Vol. 32, 1930.
14. *Pollaczek, F.:* »Theorie des Wartens vor Schaltern», Telegraphen- und Fernsprechtechnik, Vol. 19, 1930.
15. *Pollaczek, F.:* »Über zwei Formeln aus der Theorie des Wartens vor Schaltergruppen», Elektr. Nachrichtentechnik, Vol. 8, 1931.
16. *Pollaczek, F.:* »Gesprächverluste und Wartezeiten», Elektr. Nachrichtentechnik, Vol. 8, 1931.
17. *Pollaczek, F.:* »Zur Theorie des Wartens vor Schaltergruppen», Elektr. Nachrichtentechnik, Vol. 9, 1932.
18. *Pollaczek, F.:* »Kurven für die Gesprächverluste im vollkommenen Leitungsbündel», Elektr. Nachrichtentechnik, Vol. 10, 1933.
19. *Pollaczek, F.:* »Über das Warteproblem», Math. Zeitschrift, Vol. 38, 1934.
20. *Vaulot, A. E.:* »Application du calcul des probabilités à l'exploitation téléphonique», Annales des Postes, Télégraphes et Téléphones, Vol. 14, 1925.

Bibliography to the second article in this issue: *Some Observations on the Erlang Formulæ for Busy-Signal Systems* (continued from page 36).

1. *Erlang, A. K.:* »Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges», Transactions of the Danish Academy of Technical Sciences, No. 2 1948, Copenhagen.
2. *Molina, E. C.:* »Poisson's Exponential Binomial Limit», New York, 1942.
3. *Palm, C.:* »Inhomogeneous Telephone Traffic in Full-Availability Groups», Ericsson Technics, Stockholm, Vol. 5, 1937.
4. *Palm, C.:* »Analysis of the Erlang Traffic Formulæ for Busy-Signal Arrangements», Ericsson Technics, Stockholm, Vol. 6, 1938.
5. *Pearson, E. S.:* »Table of the Logarithms of the Complete Gamma-Function», Cambridge University Press, Tracts for Computers, No VIII.

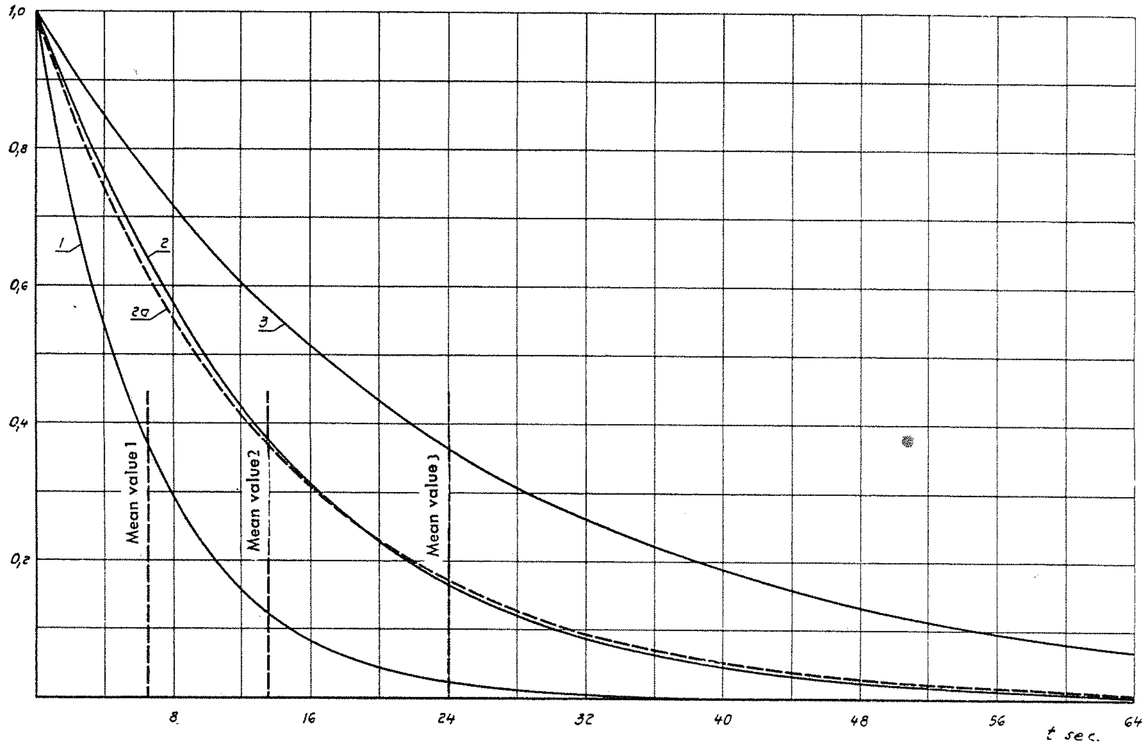

Fig. 5. Distribution function of the waiting times with voluntary departure of those waiting. Number of devices $n = 10$, traffic $A = 5$ erlangs. Mean holding time $s = 120$ secs., mean departure time $b = 12$ secs. $\alpha = 0.5$, $\beta = 1$. Serving in ordered queue and exponential distribution of holding times are assumed.

1. Distribution function $F(t)$ of the waiting times. Mean value 6.5 secs.
2. Probability of at least a t long wait $F_0(t)$. Mean value 13.54 secs.

2 a. Exponential function for the same mean value as 2.

3. Distribution function of the waiting times, when no voluntary departure of those waiting occurs. Mean value 24 secs.

and the mean waiting time (49) will be

$$b\left\{\frac{1}{1-e^{-\alpha}}-\frac{1}{\alpha}\right\}$$

Further the probability of a waiting person who does not tire of waiting having a waiting time at least t long will be

$$F_0(t) = \frac{1-e^{-\alpha \cdot e^{-\frac{t}{b}}}}{1-e^{-\alpha}}$$

and in this special case the mean value of this may be expressed in the form of a series

$$\frac{b}{1-e^{-\alpha}}\left\{\alpha-\frac{\alpha^2}{2\cdot 2!}+\frac{\alpha^3}{3\cdot 3!}-\frac{\alpha^4}{4\cdot 4!}\cdots\right\}$$

Fig. 5 shows the distribution functions for the following numerical example. Number of devices in the group $n = 10$ and incoming traffic $A = 5$ erlangs. With this we have $\alpha = 0.5$. If the mean holding time s is put at 120 secs., $\beta = 1$ means that we have the mean departure time $b = 12$ secs. We then find with the aid of the formulæ above that the mean value of the waiting times, i.e. the mean value for distribution $F(t)$ is 6.5 secs. The mean value for distribution $F_0(t)$ is 13.54 secs. This is therefore the mean value of the waiting time for a subscriber who never tires of waiting. It is remarkable that this time is approximately double the size of the mean value of waiting times actually occurring. If no voluntary departure of those waiting occurred, the mean waiting time would be 24 secs. In the example chosen the congestion is 2.37 %. If no voluntary departure of those waiting occurred the congestion would be 3.60 %, and with a busy-signal system with the same traffic and number of devices the congestion would be 1.84 %.

Fig. 5 shows $F(t)$, $F_0(t)$ and the exponential waiting time distribution that would apply if no departure of those waiting occurred. The broken curve shows the exponential function for the same mean value as with $F_0(t)$. The difference between the curves is rather slight. The expo-

But in (23), F ⇔ sojourn time. Then, how can we interpret (59) below?

By some conversion of the results presented in the above mentioned works there is obtained for the distribution function of the waiting times the expression

$$F(t) = e^{-(1+\beta)\frac{t}{b}} e^{\alpha\cdot\beta\left(1-e^{-\frac{t}{b}}\right)} \frac{W\left(\alpha\cdot e^{-\frac{t}{b}},\beta\right)}{W(\alpha,\beta)} \qquad (58)$$

Sojourn time

In this there are employed the same notations as in the previous paragraph and the formula is valid for ordered queue under the same assumptions respecting the distribution of the holding times and the departure times as were previously made. For $\beta = \infty$ (58) goes over as it should to the expression (23) for delay systems without voluntary departure for those waiting.

The distribution function (58) is now valid for the waiting times actually occurring, irrespective of whether these are terminated by those waiting receiving occupation of some device or tiring of waiting. Thus it expresses the probability that a call which is subjected to waiting will still remain after the time t as a waiting call. It should be noted that this is not the same thing as the probability for a waiting subscriber not receiving occupation during the time t, provided he does not himself tire of waiting during this time. If this latter probability is denoted $F_0(t)$, it is clearly valid that

$$F(t) = e^{-\frac{t}{b}} F_0(t)$$

seeing that the right member in this is the product of the probability that the waiting subscriber does not tire of waiting during the time t and the probability that he does not receive occupation in the same time if he waits to the end of the time. These two probabilities are independent of each other with service in ordered queue, so that their product expresses the probability of both events occurring simultaneously, which is just $F(t)$. The distribution function

Virtual waiting:
$$F_0(t) = e^{\frac{t}{b}} F(t) \qquad (59)$$

is clearly of special interest in judging the inconveniences caused to subscribers by the waiting. The same may also to some extent be said of the distribution function $F(t)$ according to (58), which moreover is that which should be obtained in measurement of the waiting times actually occurring.

The function $F(t)$ may be presented in several other forms than (58). Thus from (55) there may be obtained a presentation with the aid of the incomplete gamma function or the corresponding integral. Further, for whole number values of β a presentation with a finite number of *Poisson* expressions may be obtained from (53). In respect of the different forms of presentation attention is directed to bibliography 8.

The mean value of $F(t)$, i.e. the mean length of the waiting times actually occurring should be identical with the mean value (49) deduced before, which moreover is easy to check. For the mean value of $F_0(t)$ which obviously should be rather greater, it would appear on the other hand to be rather more difficult to set up any expression that is reasonably simple of computation. The same applies to the form factors of the distributions. In these respects the theory requires supplementing.

Some relations which are of great interest in measuring of waiting times may be shown. Thus the fraction of the full number of those waiting who cease to wait before occupation is obtained should be equal to the mean waiting time (49) divided by the mean departure time b. Further, the probability that a waiting person tires of waiting between the time t and the time $t + dt$ after waiting started is

$$\frac{1}{b} F(t)\, dt$$

This is therefore the density function for the voluntary departure of those waiting when waiting occurs in the group considered, and the validity of this is a criterion of the exponential form of the departure function. To verify this assumption by measurement one therefore only needs to investigate whether the general distribution function for those waiting and the density function for those who tire of waiting are curves with ordinates which always have a constant ratio to each other. This ratio will in such case be equal to the mean departure time b.

How to test empirically exp patience?

To demonstrate instances of the form of the distribution functions deduced we select the case $\beta = 1$, i.e. $b = s : n$. In this case (58) is simplified to

$$F(t) = e^{-\frac{t}{b}} \frac{1 - e^{-\alpha\cdot e^{-\frac{t}{b}}}}{1 - e^{-\alpha}}$$

β = 1, i.e. mean patience = b = E(service) / # servers = mean of effective service time (while waiting)

(n = 1: then ∞-server, so get simple expressions)

$\mathscr{E}_n(A,\beta)$ according to (48 a). If one has tables available which give $A \cdot E_{1,n}$ directly, it is also possible to employ the still simpler formula

$$\mathscr{E}_n = \frac{W \cdot A \cdot E_{1,n}}{n + W \cdot A \cdot E_{1,n}} \tag{56}$$

which is easy to verify.

In the table below there is given a number of values for $W(\alpha,\beta)$ computed for different values of α and β. As the congestion (48 a) for small values of $E_{1,n}$ can with reasonable accuracy be put as equal to $W \cdot E_{1,n}$, the table furnishes direct information regarding the variation of the congestion, at least for those values of congestion that are common in practice.

Finally there may be brought out an interesting special case which occurs for $\beta = n$, that is $b = s$, which in practice is a conceivable though probably rather high value for b. We then get, suitably direct from (43) and (44),

$$\mathscr{E}_n(A,n) = \sum_{\nu=n}^{\infty} \frac{A^\nu}{\nu!} e^{-A} \tag{57}$$

In this the right member may be read as the probability that in a group of devices with an infinite amount of devices one will find n or more devices occupied at one time. The formula is used in some quarters when fixing dimensions as an approximate expression for the congestion. With the assumptions stated here it is valid exactly, however.

Waiting Times with Voluntary Departure of Those Waiting.

The investigations mentioned before of delay systems with voluntary departure (bibliography 8) also contain a number of results regarding the distribution function of the waiting times with ordered queue, which will be reviewed here briefly and complemented. It is found that, unlike the case where no voluntary departure of those waiting occurs, we do not generally get any simple exponential distribution of the waiting times. On the contrary, the results will mostly be relatively complicated and numerical computations are therefore troublesome, but by no means impossible to perform.

Table for $W(\alpha,\beta)$.

α \ β	0.5	1	2	3	4	6	8	10	20	∞
0.05	1.017	1.025	1.034	1.039	1.041	$1.04\overline{5}$	1.046	1.047	$1.0\overline{5}0$	1.053
0.10	1.034	1.052	1.070	1.080	1.086	1.093	1.097	1.099	$1.10\overline{5}$	1.111
0.15	1.052	1.079	1.108	1.123	1.133	$1.14\overline{5}$	1.151	1.156	1.165	1.176
0.20	1.069	1.107	1.148	1.170	1.184	1.201	1.211	1.218	1.232	1.250
0.25	1.088	1.136	1,190	1.220	1.239	1.262	1.276	1.286	1.307	1.333
0.30	1.106	1.166	1.234	1.272	1.298	1.329	1.348	1.361	1.391	1.429
0.35	1.125	1.197	1.281	1.329	1.361	1.402	1.427	1.447	$1.48\overline{5}$	1.538
0.40	$1.14\overline{5}$	1.230	1.330	1.389	1.430	1.482	1.514	1.537	1.591	1.667
0.45	1.164	1.263	1.382	1.454	1.504	1.569	1.610	1.640	1.712	1.818
0.50	$1.18\overline{5}$	1.297	1.437	1.523	1.584	$1.66\overline{5}$	1.718	1.755	1.851	2.000
0.55	1.205	1.333	1.494	1.597	1.670	1.770	1.837	1.885	2.011	2.222
0.60	1.226	1.370	1.556	1.677	1.764	1.887	1.970	2.032	2.198	2.500
0.65	1.248	1.409	1.620	1.762	1.867	2.016	2.120	2.198	2.417	2.857
0.70	1.270	1.448	1.686	1.854	1.978	2.159	2.289	2.388	2.675	3.333
0.75	1.292	1.489	1,762	1.952	2.099	2.319	2.479	2.605	2.984	4.000
0.80	$1.31\overline{5}$	1.532	1.838	2.059	2.232	2.496	2.696	2.858	3.357	5.000
0.85	1.338	1 576	1.920	2.173	2.377	2.695	2.942	3.144	3.818	6.667
0.90	1.362	1.622	2 006	2.297	2.535	2.918	3.223	3.479	4.374	10.000
0.95	1.386	1.669	2.097	2.430	2.709	3.169	3.547	3.872	5.076	20.000
1.00	1.411	1.718	$2.19\overline{5}$	$2.57\overline{5}$	2.900	3.451	3.919	4.338	5.968	∞

The sign $\overline{5}$ indicates that the result is rounded off upwards to the five.

M/M/∞ when avg. time to abandon = avg service time.

$$n\left\{\frac{1}{W}+\alpha-1\right\}\mathscr{E}_n \tag{50}$$

This is, as it ought to be, 0 for $b = \infty$, when all incoming calls are served and it will be $A \cdot E_{1,n}$ for $b = 0$, that is the same as with busy-signal systems.

Let us now consider more closely the series (45) and state methods for its numerical computation. By comparison with the series expansion for $e^{\alpha \cdot \beta}$ it is seen at once that (45) is convergent for all positive bounded values of α and β. Again, for $\beta = \infty$, (45) has the form (47) and is convergent only for $\alpha < 1$. Further it is easily seen that $W(\alpha,\beta)$ always rises with rising α and β (here we only consider positive values for α and β). For bounded $\beta > 0$ there is then valid

$$W(\alpha, 0) < W(\alpha, \beta) < W(\alpha, \infty)$$

therefore

$$1 < W(\alpha, \beta) < \frac{1}{1-\alpha} \tag{51}$$

It is now seen, from the properties stated for $W(\alpha,\beta)$, that the congestion for bounded β is always less than one. For $\beta = \infty$ the congestion will be 1 for $\alpha = 1$, that is for $A = n$, and for greater α values, the formula has no meaning. This is fully in agreement with what we found before with delay systems without voluntary departure for those waiting, viz: that for $A > n$ there is no longer any state of equilibrium. With voluntary departure of those waiting, on the contrary, nothing like that ever happens, however great α and thus A may be. This is an extremely interesting condition. Of course, in reality there does not occur any case where one deliberately fixes the dimensions of a group of devices in a delay system, so that the offered traffic's mean value is larger than the number of devices. Nevertheless, there may arise temporary overloads of such long duration that a state of equilibrium should be able to arise. If no voluntary departure of those waiting occurs, however, no state of equilibrium can arise and gradually the waiting times approach infinity. It is evident that in such case those waiting always give up their waiting but then there will gradually set in a state of equilibrium, so that one may reckon with a congestion according to (48 a).

For numerical computations of $W(\alpha,\beta)$ one may have use for the following simple recurrence formulæ which are easily verified by insertion in (45). The formulæ are

$$W(\alpha,\beta) = 1 + \frac{\alpha \cdot \beta}{\beta+1} W\left(\frac{\alpha\cdot\beta}{\beta+1}, \beta+1\right) \tag{52 a}$$

$$\alpha \cdot W(\alpha,\beta) = W\left(\frac{\alpha\cdot\beta}{\beta-1}, \beta-1\right) - 1 \tag{52 b}$$

Furthermore, it may be noted that if β is a whole number (45) may be written in the form

$$W(\alpha,\beta) = \beta! \sum_{q=0}^{\infty} \frac{(\alpha\cdot\beta)^q}{(\beta+q)!}$$

from which we get the expression

$$W(\alpha\cdot\beta) = \frac{\beta!\, e^{\alpha\cdot\beta}}{(\alpha\cdot\beta)^\beta}\left\{1 - \sum_{\nu=0}^{\beta-1} \frac{(\alpha\cdot\beta)^\nu}{\nu!} e^{-\alpha\cdot\beta}\right\} \tag{53}$$

Thus in this case we can conveniently compute $W(\alpha,\beta)$ from a table of *Poisson* expressions. It is also possible to use the simple special case of (53):

$$W(\alpha, 1) = \frac{e^{\alpha-1}}{\alpha} \tag{53 a}$$

and then progressively apply the recurrence formula (52 b).

For the general case, where β is not a whole number, we can express $W(\alpha,\beta)$ by means of a known integral. If we derive (45) in respect of α and convert the result somewhat by means of the recurrence formulæ (52) we obtain a differential equation of the form

$$\alpha \cdot \frac{dW}{d\alpha} = \beta(\alpha - 1)\, W + \beta \tag{54}$$

and from this there is obtained the following solution

$$W(\alpha,\beta) = \frac{\beta\cdot e^{\alpha\cdot\beta}}{(\alpha\cdot\beta)^\beta}\int_0^{\alpha\cdot\beta} x^{\beta-1} e^{-x}\, dx \tag{55 a}$$

The integral in this constitutes what is called the incomplete gamma function and is to be found in table (bibliography 12). We can then write

$$W(\alpha,\beta) = \frac{\beta\cdot e^{\alpha\cdot\beta}}{(\alpha\cdot\beta)^\beta}\,\Gamma_{\alpha\beta}(\beta) \tag{55 b}$$

When $W(\alpha,\beta)$ is known it is easy with the aid of tables for $E_{1,n}$ to compute the congestion

$$[p] = \frac{E_{1,n}}{1 + (W-1)\,E_{1,n}} \cdot \frac{n!}{p!\,A^{n-p}} \qquad (46\text{ a})$$

$$[n,q] =$$

$$= \frac{E_{1,n}}{1 + (W-1)\,E_{1,n}} \cdot \frac{(\alpha\cdot\beta)^q}{(\beta+1)(\beta+2)\cdots(\beta+q)} \qquad (46\text{ b})$$

It is now seen from (45) that for $\beta = 0$ we have $W(\alpha,0) = 1$. There is then obtained from (46 a) the usual expression with busy-signal systems for the state quantity $[p]$. This is quite natural as $\beta = 0$ means $b = 0$. The mean departure time for those waiting is therefore in this case 0, i.e. none subjected to congestion troubles to wait for a device to become unoccupied. The traffic conditions in the group must then be exactly the same as with busy-signal systems. Another extreme case is also of interest, namely $b = \infty$. Then we shall also have $\beta = \infty$. If now in (45) we divide the numerator and denominator of each term by β^q, it is seen that

$$W(\alpha,\infty) = \sum_{q=0}^{\infty} \alpha^q$$

which is an ordinary geometrical series. From this we get

$$W(\alpha,\infty) = \frac{1}{1-\alpha} \qquad (47)$$

The equations (46 a) and (46 b) in this case go over to the equations (4 a) and (4 b) applying to delay systems without voluntary departure. This is also quite natural since an infinitely long mean departure time means that all those subjected to congestion continue to wait until they obtain occupation, i.e. the conditions will be the same as when no voluntary departure of those waiting occurs.

Consideration of the conditions with voluntary departure of those waiting has thus brought us to formulæ of very general validity, which contain within themselves as special cases the results presented by *Erlang* both for busy-signal systems and for normal delay systems.

We are now in a position easily to set up a formula for the congestion, that is the time of full occupation. This is obtained from the sum af all the state quantities $[n,q]$ for $q = 0, 1, 2\ldots$ and is obtained from (46 b) as

$$P_n(\text{Wait} > 0) = \mathscr{E}_n = \frac{W\cdot E_{1,n}}{1+(W-1)\,E_{1,n}} \qquad (48\text{ a})$$

(Note that W also depends on n)

For this expression there may also be used the notation $\mathscr{E}_n(A,\beta)$ with all the three parameters n, A and β set out. No special marking of α and b is required, as these are obtained from n, A and β. For $b = 0$ there is now obtained, as it should be, from (48 a)

$$\mathscr{E}_n(A,0) = E_{1,n}$$

and for $b = \infty$ we get

$$\mathscr{E}_n(A,\infty) = \frac{E_{1,n}}{1-\alpha+\alpha\cdot E_{1,n}}$$

$a = \frac{A}{n}$ $(\leftrightarrow \rho)$

which is the same thing as the recurrence formula (12 a) and therefore equal to $E_{2,n}$.

We can now also obtain a simple formula for the functioning time of the waiting arrangement, as this is defined in the left member of (6). It will be

$$\frac{W-1}{W}\,\mathscr{E}_n \qquad (48\text{ b})$$

The formulæ shown so far evidently apply irrespective of the order in which those waiting are served, as in the deduction we had no need to assume anything in this respect. The same condition applies also to the mean waiting time, this as previously shown being obtained generally from an expression of the form (18) the deduction of which, as may easily be realised, is valid with voluntary departure among those waiting, too. By means of (46 b) the following expression can be obtained for the mean waiting time

$$E[W_q \mid W_q > 0] = \frac{\beta}{y}\left\{\frac{1}{W} + \alpha - 1\right\} \qquad (49)$$

This applies to the mean waiting time of calls actually subjected to waiting. To arrive at the mean waiting time for all calls, the expression ($\times\,\mathscr{E}_n$) is to be multiplied by the congestion. (49) becomes 0 for $b = 0$, as it should be. For $b = \infty$ we have (49) assuming an indefinite form, but closer examination gives in this case

$$\frac{1}{y}\cdot\frac{\alpha}{1-\alpha}$$

which is identical with (19).

Finally, the volume of traffic is of interest, which is »lost» owing to a number of subscribers exposed to waiting not completing the waiting. For this traffic we can obtain the expression tems gives reason to suppose, however, that the assumption of an exponential departure distribution need not be feared to involve any serious restriction of the general application of the results.

We introduce the same designations as before for the state quantities with delay systems, i.e. $[p]$ and $[n,q]$, to denote the mean value per unit of time of the total time when the state prevails with p of n devices occupied or with all n devices occupied and q waiting subscribers. Assuming random traffic and exponential holding time distribution there is obviously obtained the same relation as previously (1 a) for the states when there are none waiting. The same distribution function and the same termination probabilities are valid, of course, for the state p as with busy-signal systems, irrespective of the conditions for congestion. A state n,q proceeds as long as no new call comes in, none of the occupations terminates and none of the q waiting subscribers tires of waiting. The probability of a state n,q proceeding for at least the time t will then be

$$e^{-\left(y+\frac{n}{s}+\frac{q}{b}\right)t}$$

since not only the holding times but also the departure times for those waiting are assumed to have distribution functions of the exponential type. From this expression for the probability that the state n,q will persist for a given time, we can in a manner entirely analogous with the deduction of the relations for busy-signal systems determine the probabilities of the state terminating owing to new calls coming in, owing to the termination of one of the occupations proceeding or owing to the voluntary departure of one of the q waiting subscribers. It will then be possible by means of the quantities $[n,q]$ to get the expression for the mean value of the number of times per unit of time the different states change over to each other. We then obtain a general relation of the form

$$y\,[n, q-1] = \left(\frac{n}{s}+\frac{q}{b}\right)[n, q]$$

From this and (1 a) there is then obtained

$$[n, q] = \frac{(s\cdot y)^{n+q}}{n!\left(n+\frac{s}{b}\right)\left(n+2\frac{s}{b}\right)\cdot\cdots\cdot\left(n+q\frac{s}{b}\right)}[0] \quad (43)$$

The quantity $[0]$, at last, is determined from the fact that the sum of all state probabilities must be 1, so that we get

$$\frac{1}{[0]} = \sum_{p=0}^{n}\frac{(s\cdot y)^p}{p!} +$$

$$+\frac{(s\cdot y)^n}{n!}\sum_{q=1}^{\infty}\frac{(s\cdot y)^q}{\left(n+\frac{s}{b}\right)\left(n+2\frac{s}{b}\right)\cdots\cdot\left(n+q\frac{s}{b}\right)} \quad (44)$$

To simplify the final formulæ we shall introduce somewhat different notations. First, we set $A = s\cdot y$ = the volume of traffic offered to the group, as before. In this case this is slightly greater than the traffic handled in the group, as some of those waiting are assumed to abandon their calls. Further we set $\alpha = \frac{s\cdot y}{n}$ = the ratio between the traffic offered and the number of devices. Finally it seems appropriate to set the mean fatigue b and the mean holding time s in relation to each other. We therefore introduce $\beta = \frac{n\cdot b}{s}$. The constant β is then the ratio between the mean departure time and the mean interval between the congestion transitions. The second series in the right member of (44) will now, with these notations, have the form

$$\sum_{q=1}^{\infty}\frac{(\alpha\cdot\beta)^q}{(\beta+1)(\beta+2)\cdots(\beta+q)}$$

It appears advisable to introduce a separate notation for this series. We put

$$W(\alpha,\beta) = \sum_{q=0}^{\infty}\frac{(\alpha\cdot\beta)^q}{(\beta+1)(\beta+2)\cdot\cdots\cdot(\beta+q)} \quad (45)$$

In this series there occurs for convenience also a term for $q = 0$, which is set equal to 1, so that we have $W(0,\beta) = 1$ for $\alpha = 0$. There is now obtained from (44)

$$\frac{1}{[0]} = 1 + A + \frac{A^2}{2!} + \cdots\frac{A^{n-1}}{(n-1)!} + \frac{A^n}{n!}W(\alpha,\beta)$$

By means of the notation $E_{1,n}$ introduced on page 19 formula (7) for loss in busy-signal systems, this may be written in the form

$$\frac{1}{[0]} = \frac{A^n}{n!}\left\{W - 1 + \frac{1}{E_{1,n}}\right\}$$

We then get, finally, the following expressions for the state quantities $[p]$ and $[n,q]$:

tion that the holding times follow an exponential distribution function and that the calls are distributed at random during the times when the subscribers are not engaged in conversation. Thus no accumulating of call needs is considered as occurring. If the number of subscribers is denoted by N and the incoming traffic from each subscriber is a, the mean value of that part of the whole time when the state p prevails is represented by

$$[p] = K\binom{N}{p}a_1^{\,p} \qquad (41\text{ a})$$

This is valid for $p \leq n$. If there is congestion and the number of waiting calls is q there is obtained instead

$$[n, q] = K\binom{N}{n+q}\frac{(n+q)!\,a_1^{\,n+q}}{n!\quad n^q} \qquad (41\text{ b})$$

The constant K in these formulae is determined by the condition

$$\sum_{p=0}^{n-1}[p] + \sum_{q=0}^{N-n}[n, q] = 1 \qquad (41\text{ c})$$

and the traffic quantity a_1 is obtained from

$$N a = \sum_{p=1}^{n-1} p\,[p] + n\sum_{q=0}^{N-n}[n, q] \qquad (42\text{ a})$$

For small congestions there is valid with good accuracy

$$a_1 = \frac{a}{1-a} \qquad (42\text{ b})$$

Numerical computations according to the above equations will in general unfortunately be quite troublesome, so that the formulae are hardly likely to find any great employment in computation of dimensions in practice.

Congestion with Voluntary Departure of Those Waiting.

The formulae presented earlier in this article for delay systems have all been based on the assumption that the subscribers when subjected to waiting have always continued to wait until they are served. This is probably not always the case in reality; it is a wellknown experience that with markedly long waiting times the subscribers will often become impatient and replace the receiver, renewing the call immediately or after a little while only. In the former case there should not be caused any effect on the size of the congestion; the result of a subscriber ceasing to wait and immediately trying a new call will only be that he will be placed further back in the queue, in case the serving of those waiting is done in ordered queue. If on the other hand the subscriber, after tiring of waiting, does not renew the call until after a certain time, the call may be regarded as »lost» in the same way as a call meeting with congestion in a busy-signal system, so that the renewed call is suitably counted among the normal random distributed calls. In this case there evidently arises owing to the voluntary departure of waiting subscribers a reduction of the congestion and of the waiting times for those who continue to wait. It has been found possible with certain assumptions theoretically to compute the traffic conditions with such voluntary departures among those waiting and there are then obtained formulae of a fairly general character, which comprise as special cases the *Erlang* solutions both for busy-signal and for delay systems. Respecting the deduction attention is directed to bibliography 8. Here we shall only present some complementary formulae and tables, which facilitate employment of the results in practice. In this connection there will be given a brief survey of the most important of the results obtained earlier.

To allow of a reasonably simple mathematical treatment it must be assumed that the probability that a waiting subscriber gets tired of waiting may be expressed as a simple exponential function of the time he has already been waiting. We therefore start out from

$$e^{-\frac{t}{b}}$$

as expressing the probability that a subscriber who has to wait has still not ceased to wait after a time t. We denote this function the departure function. The constant b, which is the mean value of the distribution, may be denoted the mean departure time. Now it is true that both measurements and theoretical considerations which cannot, however, be gone into more closely in this connection, show that the departure function has in reality a more flat type than that represented by the exponential distribution. Experience in respect of the influence of the holding time distribution on traffic conditions with delay sys- CV ∨ 1

Patience ~ exp(1/b) (mean = b)

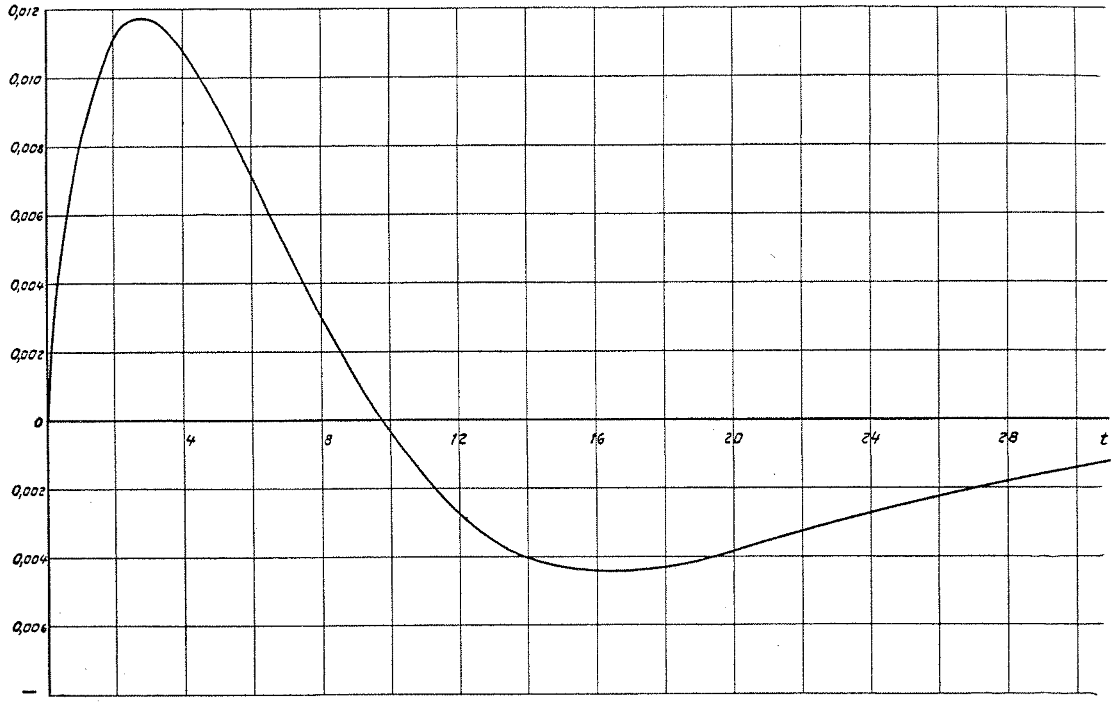

Fig. 4. The difference $\left(1 - \frac{t}{100}\right)^{19} - e^{-\frac{t}{5}}$ as a function of t.

function should quite closely agree with the purely exponential form. To show how good this agreement is in reality, the function (33) has been tabulated here for $n = 20$ and $s = 100$ secs. There is obtained from (33) in this case

$$F(t) = \left(1 - \frac{t}{100}\right)^{19}$$

valid for all values of t up to 100. Beyond that the function has the value 0. The difference between this function and the exponential function for the same mean value has a maximum of around 0.01. With graphical presentation, therefore, the scale must be quite large for the curves to be distinguished from each other. Some idea of how near together the curves come is given by fig. 4, which shows the difference

$$\left(1 - \frac{t}{100}\right)^{19} - e^{-\frac{t}{5}}$$

as a function of t.

As a final example, we may consider the distribution introduced on page 54 for two exponentially regulated stages. We then have

$$f(t) = \left(1 + \frac{2}{s}t\right)e^{-\frac{2}{s}t}$$

and get

$$f_1(t) = \left(1 + \frac{t}{s}\right)e^{-\frac{2}{s}t}$$

The form factor according to (40) will then be

$$\frac{n!}{(2n)^n}\left\{1 + 2n + \frac{(2n)^2}{2!} + \cdots + \frac{(2n)^n}{n!}\right\}$$

It can be shown that this expression has 2 as limit with growing n. For $n = 2$ we get the form factor 1.625. It may be of interest to consider the congestion values in the table on page 55 in the light of the form factor values for the distribution function of the congestion transition. With constant holding time this form factor will be 1.333, with two exponentially regulated stages there is obtained the value 1.625 and finally for purely exponential distribution the value will be 2.000.

In conjunction with the general discussion of the applicability of the *Erlang* solution for delay systems with exponential holding time distribution, something may also be said about the conditions with delay systems when the number of subscribers is regarded as limited. Some formulae for this case will be found in bibliography 7, chap. 5 (unfortunately partly incorrect owing to printing errors). They are deduced on the assumpthe latter case should therefore be fully adapted to general application. Naturally, this rule must be employed with a certain amount of discrimination, as it is known that conditions with very small groups can be very different. This should be particularly observed for $n = 1$, a case which is quite often found with service stations (e.g. booking offices, counter service etc.).

To obtain an idea of how good the agreement is with the formulæ applying to exponential holding time distribution, it may be worth while in special cases to compute the function (33), or at least determine its form factor which may be expected to furnish a good indication of how closely (39) is satisfied. The form factor for any distribution function $f(t)$ is defined by the equation (3) in the first article of this number, page 3. In that expression s is the mean value of the distribution, which for the function (33) above is equivalent to $s:n$. Thus the form factor for this is expressed by

$$\frac{2n^2}{s^2}\int_0^\infty t \cdot F(t)\, dt$$

If in (33) we introduce the function $f_1(t)$ according to (34), there is obtained from this

$$\frac{2n^2}{s^2}\int_0^\infty t\,\{f_1(t)\}^{n-1} f(t)\, dt$$

Now $f_1'(t) = -\frac{1}{s} f(t)$ so that the expression may be written

$$-\frac{2n}{s}\int_0^\infty t \cdot d\,\{f_1(t)\}^n$$

From this there is obtained by partial integration

$$-\frac{2n}{s}\Big[t\{f_1(t)\}^n\Big]_0^\infty + \frac{2n}{s}\int_0^\infty \{f_1(t)\}^n\, dt$$

Now it can always be assumed that

$$\lim_{t \to \infty} t\{f_1(t)\}^n = 0$$

which, in fact, is necessary if the form factor is to be finite. The square bracket in the last expression but one will then be zero, and we obtain as final expression for the form factor

$$\frac{2n}{s}\int_0^\infty \{f_1(t)\}^n\, dt \qquad (40)$$

With exponential distribution we have $f(t) = e^{-\frac{t}{s}}$, from which we get $f_1(t) = e^{-\frac{t}{s}}$. There is then obtained, as there should, for the form factor the value 2, irrespective of n. As an example of a flat distribution we may consider the completely monotone function

$$f(t) = \frac{8}{(2+t)^3}$$

which is to be found in curve 2, fig. 1, page 5. In this case $s = 1$ and the form factor has the value 4. We now find

$$f_1(t) = \frac{4}{(2+t)^2}$$

From (40) there is then obtained

$$2n\,4^n\int_0^\infty (2+t)^{-2n}\, dt$$

or, evaluated

$$\frac{2}{1-\frac{1}{2n}}$$

In this case then the form factor for the distribution function of the congestion transition for increasing n approaches the value 2, that is the value it has with purely exponential distribution. For a group with 20 devices, therefore, the form factor will have the value 2.051, whereas with the original function $f(t)$ it is 4. Even with 2 devices it has gone down to the value 2.67.

As yet another example, we will consider the distribution with constant holding time, with which the form factor has, of course, the minimum value 1. Here $f(t) = 1$ for $0 \le t \le s$ and $f(t) = 0$ for $t > s$. We then get $f_1(t) = 1 - \frac{t}{s}$ for $0 \le t \le s$ and $f_1(t) = 0$ for $t > s$. In this case there is obtained from (40)

$$\frac{2n}{s}\int_0^s \left(1-\frac{t}{s}\right)^n dt$$

or

$$\frac{2n}{n+1}$$

Thus for a group of 20 devices the form factor for the distribution function of the congestion transition will be 1.905, which shows that the means that the probability that the $n-1$ occupations which are still proceeding after the first congestion transition will all continue after the time t is expressed by

$$\frac{1}{s^{n-1}}\left\{\int_0^\infty f(x+t)\,dx\right\}^{n-1}$$

In addition, the fresh occupation which started at the congestion transition and which derived from one of the waiting calls will continue at least the time t, and the probability of this is $f(t)$, this being the distribution function of the holding times. The product of $f(t)$ and the above probability, thus expresses the probability that none of the n occupations will terminate during the time t after the congestion transition. But this product is identical with $F(t)$ according to formula (33). We may now repeat the reasoning and determine the probability that after the next congestion transition, if such occurs, all the occupations will last at least the time t and will then again obtain the function $F(t)$. This function therefore is preserved throughout the whole congestion state and gives the probability that after each congestion transition taken at random none of the n occupations will terminate during the continued time t. The formula (33) therefore gives the general expression for the *congestion transition's distribution function,* for which we found earlier in (24) a special expression, valid for exponential holding time distribution. If we introduce $f(t) = e^{-\frac{n}{s}t}$ it is also easy to see that the two expressions agree.

Now, the practical significance of the reasoning here advanced lies in the fact that the function (33) for n values of any size comes very close to the exponential function (24). To show this we introduce the designation

$$f_1(t) = \frac{1}{s}\int_0^\infty f(x+t)\,dx \qquad (34)$$

We further introduce a function $K(t)$, defined by

$$f_1(t) = 1 - \frac{t}{s}\{1 - K(t)\} \qquad (35)$$

By deriving (34) and (35) we get

$$f'_1(t) = -\frac{1}{s}f(t) = -\frac{1}{s} + \frac{1}{s}K(t) + \frac{t}{s}K'(t)$$

from which

$$f(t) = 1 - K(t) - t\cdot K'(t) \qquad (36)$$

If we introduce the variable $\tau = \frac{n}{s}t$ and expand $\ln f_1(t)$ in series according to (35) we get

$$\ln f_1(t) = -\frac{\tau}{n}\left\{1 - K\left(\frac{s\tau}{n}\right)\right\} - \\ -\frac{1}{2}\frac{\tau^2}{n^2}\left\{1 - K\left(\frac{s\tau}{n}\right)\right\}^2\cdots$$

From (33) we then get

$$\ln F\left(\frac{s\tau}{n}\right) = \ln f\left(\frac{s\tau}{n}\right) - \frac{n-1}{n}\tau\left\{1 - K\left(\frac{s\tau}{n}\right)\right\} - \\ -\frac{1}{2}\frac{n-1}{n}\cdot\frac{\tau^2}{n}\left\{1 - K\left(\frac{s\tau}{n}\right)\right\}^2\cdots \qquad (37)$$

We shall now determine the limit of this expression when $n \longrightarrow \infty$. We then note first that $f(0) = 1$. From (34) it further follows that $f_1(0) = 1$. From (35) it is then seen that we have

$$\lim_{t\to 0} t\cdot K(t) = 0$$

From this it follows, however, that

$$\lim_{t\to 0} K'(t)$$

must be bounded. It then follows from (36) that

$$\lim_{t\to 0} K(t) = 0$$

From (37) it then follows, when we let $n \longrightarrow \infty$ that

$$\lim_{n\to\infty} \ln F\left(\frac{s\tau}{n}\right) = -\tau \qquad (38)$$

which means that

$$F(t) \longrightarrow e^{-\frac{n}{s}t} \qquad (39)$$

for great values of n.

Since the distribution function of the congestion transition wholly determines both the waiting times and the durations of the congestion states, (39) now shows that the congestion conditions in large groups will be independent of the distribution function of the holding times and the same as in groups with exponential holding time distribution. The comparatively simple formulae for justification for the opinion that the formulae valid for exponential holding time distribution are also applicable for arbitrary distribution function for holding times, at least with small congestions and large groups, may be obtained by comparison with conditions in busy-signal systems since in this latter case we know the expressions for the distribution functions of the various states with arbitrary distribution function $f(t)$ for the holding times. (Regarding this, see bibliography 9, particularly formulae (37) and (43).) These distribution functions for the states determine uniquely all traffic conditions with busy-signal systems, provided there is statistical equilibrium, which means that the traffic must have proceeded, theoretically, for an infinitely long space of time. It is found, however, both from experience and from theoretical investigations (bibliography 11, chap. 7), that equilibrium is restored very quickly after small disturbances. If we now consider a delay system at an instant when a congestion state ceases, then it may be assumed that the state that follows with $n - 1$ devices occupied has a distribution function different from the corresponding one in a busy-signal system. If now the congestion in the delay system is small and the number of devices large, the mean value of the time between two successive congestion states is very great in relation to the mean interval between the incoming calls. Now, during the time between two congestions a delay system operates on exactly the same laws as a busy-signal system, since it is only in respect of the treatment of congested calls that the systems differ. During this interval between successive congestion states the traffic conditions should therefore seek to attain the same state of equilibrium as in a busy-signal system and if this interval is long the ages of the occupations proceeding on the occurrence of a fresh congestion state should distribute themselves in approximately the same manner as with a busy-signal system, which means that the probability that none of the occupations proceeding terminates in the time t should be approximately the same as in a busy-signal system. This last is expressed according to bibliography 9 by

$$F(t) = \frac{f(t)}{s^{n-1}} \left\{ \int_0^\infty f(x+t)\,dx \right\}^{n-1} \qquad (33)$$

In this s is the mean holding time and n the number of devices in the group. The distribution function of the holding times has an arbitrary form $f(t)$.

One may obtain an idea of the quantities referred to in the reasoning by means of a numerical example. We consider a delay system with the number of devices $n = 20$ and the incoming traffic 11 erlangs. The mean holding time may be 126 secs., which is that usual with us for local calls. If we apply the formulae that are valid for exponential holding time distribution we find as value for the congestion 0.010 and for the mean duration of the individual congestion states 14 secs. Thus per hour congestion prevails on the average 36 secs., which means 2.6 congestion states on the average per hour. The mean value of the interval between two successive congestion states will then be 1400 secs., i.e. approximately 23 mins. During this time there arrives an average of 122 calls and a like number of occupations terminate. Between two successive congestion states there occurs then an average of 244 changes of state. This should be sufficient to cancel out even a very great disturbance of the statistical equilibrium.

We shall now show a remarkable property in the function $F(t)$ above. Assume that the expression (33) for this function is also valid in a delay system on the occurrence of a congestion state, which from what is stated above should be approximately correct if the intervals between the congestion states are of any length. The expression (33) then gives the probability that none of the n occupations terminates during the time t. While all the n occupations are proceeding therefore there can come in fresh calls, which are then subjected to delay. When later one of these n occupations terminates, there supervenes immediately a fresh occupation and the congestion continues. We shall now determine the probability that after such a *congestion transition* none of the occupations terminates during the remainder of the time t. This may easily be done by means of the formulae deduced in bibliography 9 since the expression (43 a) on page 52 in that paper gives the probability that all p occupations occurring in a state p arising out of a state $p + 1$ will proceed for a further time t. As we have assumed that (33) above is valid, the said formula may be applied directly to our case for $p = n - 1$. This

$$S = \sum_{r=0}^{\infty} \{[20]_r + [11]_r + [02]_r\}$$

As the sum of all the state quantities is one, we have further the relation

$$[00] + [10] + [01] + S = 1$$

During the time when one of the states [10] or [01] prevails, one occupation is dealt with in the group. During the time that congestion prevails two occupations in the group are dealt with. The traffic dealt with per unit of time in the group will then be $[10] + [01] + 2S$ and this must be equal to the traffic offered to the group, which clearly is $s \cdot y$. We then get the relation

$$[10] + [01] + 2S = s \cdot y$$

We have thus obtained 6 equations which only contain the 6 unknowns [00], [10], [01], $[11]_0$, $[02]_0$ and S. If from this we solve S, there is obtained the following expression, in which we have set the incoming traffic $s \cdot y = A$:

$$S = \frac{A^2}{2 + A}\left\{1 - \frac{(2 - A)\,A}{64 + 56\,A + 12\,A^2 + A^3}\right\} \quad (32)$$

If we compare this expression with the congestion with purely exponential holding time distribution, i.e.

$$E_{2,2} = \frac{A^2}{2 + A}$$

it is seen that (32) always gives less congestion except for the maximum value $A = 2$. The table below furnishes a comparison between some congestion values obtained with constant holding time, with the distribution treated here for exponentially regulated stages and finally with purely exponential holding time distribution.

Traffic	Congestion with the form factor		
A	1.0	1.5	2.0
0.0	0.000	0.000 0	0.000 0
0.2	0.018	0.018 1	0.018 2
0.4	0.064	0.066 2	0.066 7
0.6	0.134	0.137 3	0.138 5
0.8	0.221	0.226 7	0.228 6
1.0	0.323	0.330 8	0.333 3
1.2	0.439	0.447 1	0.450 0
1.4	0.565	0.573 6	0.576 5
1.6	0.702	0.708 7	0.711 1
1.8	0.847	0.851 2	0.852 6
2.0	1.000	1.000 0	1.000 0

The form factor 1.0 corresponds to constant holding time, the form factor 1.5 to the distribution examined above with exponentially regulated stages and the form factor 2.0 to the purely exponential distribution. The values for constant holding time are taken from *Erlang's* tables, which only go to 3 decimal places.

Despite the fact that the congestion values in the table refer to a small group, the differences between the cases are relatively small. For small congestion values the differences are very small, growing with rising A to a maximum around $A = 1.3$. The congestion with the form factor 1.0 is then 0.011 less than with the form factor 2.0, which is only 2.5 % of the congestion value. The relative deviations always decrease with rising A and are zero for $A = 2$. It is extremely interesting to note that the congestion with the form factor 1.5 always lies between the congestion for the form factors 1.0 and 2.0, and considerably nearer the congestion value for the latter form factor. For the general utility of the formulae applying to exponential holding time distribution, this is an exceedingly favourable circumstance.

The discussion presented appears to show that in all steep holding time distributions occurring in reality one can with good accuracy apply the formulae for exponential distribution. Another question of equal importance is how conditions will be with flat distributions which, as stated, have a form factor greater than 2. Thus, measurements performed lately of holding times with local traffic have indicated that here we have relatively flat distributions, with form factors between 3 and 4. For the investigation of conditions in such cases the author treated in an earlier work delay systems with arbitrary completely monotone holding time distributions (bibliography 7). The result was obtained that the same formulae as for exponential distribution also apply for an arbitrary completely monotone distribution, at any rate where the congestion is concerned. Unfortunately, in a scrutiny recently undertaken, the author has found a number of gaps in the proof, so that this result, otherwise extremely important, cannot yet be regarded as fully established.

The reasoning advanced up to now has mainly been based on numerical comparisons in special cases and can therefore not be considered as sufficiently general. A considerably more general

In measurements of 16000 register holding times there was obtained a distribution curve with the form factor 1.23, which means a considerable departure from the conditions with absolutely constant holding time. The measurement results are mentioned in more detail in bibliography 10. Now it appears very natural that the traffic conditions with holding time distributions whose form factor lies between 1 and 2 should differ less from the conditions with purely exponential holding time distribution than what is the case for traffic conditions with absolutely constant holding time, i.e. with the form factor 1. With steep distributions occurring normally in actual practice it should therefore be possible with still greater safety to apply the formulæ valid for exponential distribution than is the case with absolutely constant holding time. With a view to examining this condition more closely the author has tried to introduce distributions with exponentially regulated stages. In the first article of this number, »Some Propositions Relating to Flat and Steep Distribution Functions», it was shown that such distributions always have a form factor between 1 and 2. As an example of the results obtained in these investigations, we may go through a simple case, the computations for which will not be too complicated, but which all the same allow of some numerical comparisons that are of interest.

The law of distribution of the holding times is assumed to be such that each occupation must pass through two stages before it is terminated. For either stage there applies the distribution $e^{-\frac{2t}{s}}$. The distribution function of the holding times will then be according to equation (30) in article 1 of this number

$$\left(1+\frac{2}{s}t\right)e^{-\frac{2t}{s}}$$

The mean holding time is s and the distribution's form factor is 1.5. The appearance of the distribution function is seen from curve 2 in fig. 2 on page 17 (of that article).

Let us now consider a group with 2 devices in a delay system. The calls are assumed to come in random distributed, their mean number being y per unit of time. For the mean durations per unit of time for the states possible in the group the following notations are introduced:

[00] : no occupation present.

[10] : only one occupation in stage 1 present.

[01] : only one occupation in stage 2 is present.

$[20]_r$: two occupations in stage 1 are present. In addition there are r waiting calls.

$[11]_r$: one occupation is present in either stage. In addition there are r waiting calls.

$[02]_r$: there are two occupations in stage 2 and r waiting calls.

The distribution functions for each of these states can now easily be set up and from them one can determine the mean number of times per unit of time that each state is present and goes over into some other state. As regards method, reference may be made to the usual deduction of the *Erlang* formulæ (see, e.g. bibliography 6). By then setting the mean number of times per unit of time that a state prevails as equal to the number of times per unit of time that it arises out of another state, there is obtained a linear equation for each state. Thus, the following equations, among others, are obtained:

$$y\,[00]=\frac{1}{s_2}[01]$$

$$\left(\frac{1}{s_1}+y\right)[10]=\frac{1}{s_2}[11]_0+y\,[00]$$

$$\left(\frac{1}{s_2}+y\right)[01]=\frac{1}{s_1}[10]+\frac{2}{s_2}[02]_0$$

$$\left(\frac{2}{s_2}+y\right)[02]_0=\frac{1}{s_1}[11]_0$$

In each equation the left member represents the mean number of times per unit of time the state considered prevails. The right member gives the mean number of times per unit of time that the state arises from other states. The simple but long-winded deduction may be omitted here, as it would be trivial for those who are acquainted with the methods in the preceding works.

As there exists an infinite number of states one can obviously set up an infinite number of equations. However, to determine the congestion it is sufficient with the four equations given above. To show this we introduce the notation S for the congestion, i.e. for the sum of all the states when both the devices are occupied and an incoming call must therefore wait. We then have

M/E2/S

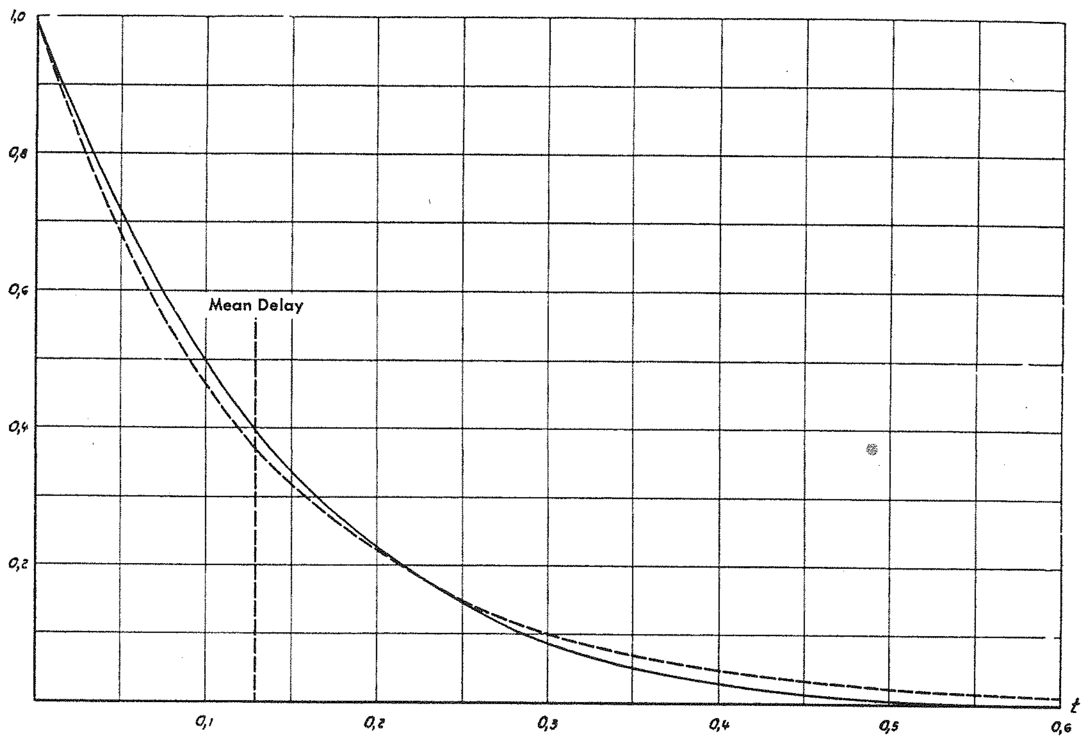

Fig. 3. Distribution function of the delay with constant holding time. Number of devices $n=10$, traffic $A=4$ erlangs. Holding time equal to unit of time. The mean delay is 0.129. The broken line curve shows the exponential function for the same mean value.

ficance, particularly in respect of the occurrence of long delays.

The present author has in quite a large number of cases compared the congestion values computed by *Erlang*, *Crommelin* and *Pollaczek* for constant holding time with those which apply for equivalent traffic and number of devices with exponential holding time distribution. It was then found that the differences in general are so small that they must be considered as lacking significance in the work of dimensioning. The differences will be the smaller the larger the groups are. Only with very small groups will the difference be of such a magnitude that one possibly should take them into account. In the case $n = 1$, there is no difference, as the equation (7), as stated earlier, is valid generally.

Pollaczek, for the congestion with constant holding time, has deduced an asymptotic formula (see bibliography 19) valid for great n values:

$$\frac{1}{1-\alpha} \cdot \frac{(\alpha e^{1-\alpha})^n}{\sqrt{2\pi n}} \qquad (31)$$

in which $\alpha = A:n$. Now it was shown earlier, see equation (17), that

$$\frac{n}{n-A} \cdot \frac{A^n}{n!} e^{-A}$$

for great n values constitutes a good approximate value for the congestion with exponential holding time distribution. If now in this expression we introduce the wellknown *Stirling* approximate value for the gamma function, i.e.

$$n! = \sqrt{2\pi n} \cdot n^n e^{-n}$$

then the expression (31) is obtained in this case too. From this it is seen that both for constant and exponential holding time there are obtained the same expressions in the limit for the congestion in large groups. *Pollaczek* has also stated asymptotic expressions for the mean delay and the distribution function of the delays, which show that these, too, will be equal for constant and exponential holding time distribution for great n values.

The numerical and theoretical results stated all appear to show that with groups not too small it is possible with good approximation to count with the formulæ for exponential holding time distribution for constant holding time, too. Now there hardly occur in reality any traffic cases in which the holding time may be regarded as completely constant. The case in telephony in which one comes nearest to the constant holding time would seem to be with register occupations.

The General Dependence of Congestion Conditions on the Law of Distribution of Holding Times.

The condition that congestions and delays in waiting systems are dependent on the distribution law valid for the holding times constitutes, of course, a considerable inconvenience, the like of which does not exist with the busy-signal systems with which, as previously shown, the congestion and the other traffic values are entirely independent of the form of the holding time distribution function. Moreover, conditions with delay systems are so complicated that the theoretical investigations so far have only furnished definite results for some special types of holding time distributions. Even if it were possible to perform computations for every kind of distribution function for the holding times, the application would be found extremely troublesome, as in fixing dimensions one would always have to take into account which kind of occupations were present in the individual case. In such conditions it may appear as if the value of the results so far gained were rather limited. Fortunately such does not appear to be the case, and the reason is that the influence of the distribution function of the holding times as regards congestion and delays seems in general to be slight from the purely numerical point of view. In consequence of this it would appear that in most cases one may employ with confidence the formulæ, which are valid for exponential holding time distribution and which for numerical computations are the most convenient. There are many motives for such an attitude and these will be more closely examined in this section.

State	Constant holding time	Exponential holding time
[0]	0.0183	0.0183
[1]	0.0730	0.0732
[2]	0.1464	0.1464
[3]	0.1951	0.1952
[4]	0.1952	0.1952
[5]	0.1565	0.1562
[6]	0.1044	0.1041
[7]	0.0596	0.0595
[8]	0.0297	0.0298
[9]	0.0140	0.0132
Congestion	0.0078	0.0088

To arrive at an estimate of the numerical difference between the results with constant and exponential holding time, we will first consider a case computed by *Crommelin* with constant holding time and compare the same with the equivalent values obtained from *Erlang*'s formulæ for exponential holding time. In the example the number of devices $n = 10$ and the incoming traffic $A = 4.0$ erlangs. The table below shows the values that are valid for the different states $[p]$ and for the congestion with constant and exponential holding time.

For the state quantities [0] up to and including [8] the agreement is practically exact, and the small differences there are would seem owing to the irregular occurrence to be entirely attributable to the circumstance that the 4th decimal is not exact as regards the values for constant holding time. On the other hand, the congestion is distinctly less with constant than with exponential holding time and the difference will be seen to have mostly have lain in the state 9, which is the highest state where no congestion occurs. That the congestion should be smaller with constant than with exponential holding time is fairly natural and appears to be a condition generally valid (except for $n = 1$), as all comparisons between numerically computed values show this. It is also rather natural that this decrease in the congestion should be compensated by an increase of the durations of just the highest states.

In the example given it must be considered that the difference between the congestion values is so slight that it is of no significance in practice if in deciding dimensions one reckons with one value or the other.

In respect of mean delay for the calls that have to wait, *Crommelin* found for the same example the value 0.129, the unit of time being set equal to the constant holding time. With exponential holding time distribution there is obtained the mean delay 0.167. The difference should be considered here as having a certain significance though in percentage it is not particularly great.

In fig. 3 will be seen the distribution function for the delays in the above-mentioned example with constant holding time as compared with the exponential function for the same mean value. The difference between the curves may be considered as having quite appreciable signi-

$$Q(z) - a_n z^n =$$

$$= -(n-y)\frac{(z-1)(z-\lambda_1)\cdots(z-\lambda_{n-1})}{(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_{n-1})} \qquad (24)$$

This is identically valid, so that the coefficients for the different powers of z must be equal in both members. By means of the definitions (18) and (22) all $[p]$ for $p < n$ will then be uniquely determined. For example we find

$$1 - \sum_{p=0}^{n-1}[p] = 1 - \frac{n-y}{(1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_{n-1})} \qquad (25)$$

This expresses the congestion, that is the part of the whole time when all n devices are held.

The function $\psi(z)$ is now fully determined and by expanding the right member of (23) according to rising powers of z and comparing the coefficients for the different powers of z, there will be determined all $[p]$ even for $p \geq n$. Thus it has been shown by the introduction of the generating function that the system (19) has a single solution, which satisfies the condition that all $[p]$ are non-negative and that the sum of them is 1.

For computation of the roots λ it is convenient to introduce $\beta = \alpha \cdot \lambda$, $\alpha = y:n$ being the mean value of the load per device. By putting the denominator in (23) equal to zero there is obtained the equation

$$\beta \cdot e^{-\beta} = \alpha \cdot e^{-\alpha}\sqrt[n]{1} \qquad (26)$$

For each of the n roots of 1 there is obtained a value for β, the absolute value of which is $\leq \alpha$. Not more than 2 roots are real, the others are conjugately complex.

What is stated above shows the fundamental reasoning behind the *Crommelin* presentation of the problem. It is possible then from the generating function to produce expressions both for the mean waiting time and for the distribution function of the waiting times. The latter has a broken form, in that its derivative is discontinuous for $t = 1, 2, 3$ and so on, i.e. for integral multiples of the constant holding time. If we put $t = T + \tau$, with T a whole number or zero and $0 \leq \tau \leq 1$, then there is obtained for the probability that a random selected call will be subjected to at least a t long delay the following series expansion

$$F(t) =$$

$$= \sum_{\mu=1}^{\infty}\sum_{\nu=0}^{n-1} a_\nu \frac{\{y(\mu-\tau)\}^{(\mu+T)n+n-1-\nu}}{\{(\mu+T)n+n-1-\nu\}!} e^{-y(\mu-\tau)} \qquad (27)$$

Erlang has also given an asymptotic formula

$$F(t_1) \approx e^{-y(r_0-1)(t_1-t_2)} F(t_2) \qquad (28)$$

by means of which for great t values one can compute $F(t_1)$ out of $F(t_2)$, with $t_1 > t_2$. The formula will be the more accurate the greater α and n are. The constant r_0 has been tabulated by *Erlang* (bibliography 4) and is also given in *Berkeley*'s above-named article (bibliography 2).

The formulæ stated above are valid for delays with ordered queue. The delay conditions with random served queue and constant holding time do not appear to have been the object of investigations.

The determination of the congestion according to (25) will be troublesome owing to the necessity of computing the roots λ. *Pollaczek* and *Crommelin* have therefore also deduced a somewhat more convenient expression for computing the congestion. According to this

$$\sum_{\mu=1}^{\infty}\frac{e^{-\mu y}}{\mu}\sum_{\nu=\mu n}^{\infty}\frac{(\mu y)^\nu}{\nu!} \qquad (29)$$

is equal to the natural logarithm for 1 minus the congestion. There has also been produced a series expansion for computation of the mean delay. According to this the mean delay for all calls (i.e. not only those which have to wait) is

$$\sum_{\mu=1}^{\infty} e^{-\mu y}\left\{\sum_{\nu=\mu n}^{\infty}\frac{(\mu y)^\nu}{\nu!} - \frac{n}{y}\sum_{\nu=\mu n+1}^{\infty}\frac{(\mu y)^\nu}{\nu!}\right\} \qquad (30)$$

As has been seen from this brief exposé the theoretical treatment of the traffic conditions with delay systems and constant holding time gives relatively complicated final formulæ, which are little suited to numerical computations. Nevertheless, a number of these have been performed by *Erlang, Pollaczek, Crommelin* and *Berkeley*. Quite a number of conclusions of great interest can be obtained from the results of the computations, and these will be touched on in more detail below.

We see that the model described holds good even if $p > n$ and therefore there are waiting calls also at the end of the interval. The probability that at the beginning of an interval of time selected at random of the kind under consideration one of the states 0, 1, .. n will prevail is now according to the notation (18) equal to a_n. Further the probability that the state p will prevail at the beginning of the interval is $[p]$ and finally the probability that during an interval of time with length 1 there will come in μ calls but not more is, according to the wellknown laws for random traffic,

$$\frac{y^{\mu}}{\mu!} e^{-y}$$

The probability $[p]$ that at the end of the interval the state p will prevail must now be equal to the sum of the probabilities for the different contingencies described above for the occurrence of the state p at the end of the interval. On the basis of this we get the relation

$$[p] = a_n \frac{y^p}{p!} e^{-y} + \sum_{\mu=1}^{p} [n+\mu] \frac{y^{p-\mu}}{(p-\mu)!} e^{-y} \quad (19)$$

which is valid for all values of p, i.e. also for $p > n$. For $p = 0$ the right member is reduced to the first term only.

By means of (19) there is obtained a linear equation system with an infinite amount of unknowns. According to *Crommelin*, the system was first stated by *Fry*, though not published. It is found that this system in conjunction with the conditions that all the unknowns are probabilities, i.e. non-negative quantities, and that the sum of all the quantities $[p]$ is one, has a unique solution. To show this we define with the aid of the probabilities $[p]$ a function of a variable z:

$$\psi(z) = \sum_{p=0}^{\infty} [p] z^p \quad (20)$$

Such an expression is designated a *generating function*, as the separate probabilities $[p]$ constitute the coefficients to the powers of z. Since the sum of all the probabilities must be one, there is valid

$$\psi(1) = \sum_{p=0}^{\infty} [p] = 1 \quad (21)$$

Since all the quantities $[p]$ are non-negative, the series is absolutely convergent. The function $\psi(z)$ is therefore limited for each z (also complex) whose absolute value $|z| \leq 1$.

If now we multiply both the members in (19) by z^p and add the relations for all p values thus obtained, we get

$$\psi(z) = a_n e^{y(z-1)} + e^{-y} \sum_{p=1}^{\infty} \sum_{\mu=1}^{p} [n+\mu] z^p \frac{y^{p-\mu}}{(p-\mu)!}$$

If the summation order in the double sum is reversed, the sum over p can be performed directly and we get

$$\psi(z) = a_n e^{y(z-1)} + e^{y(z-1)} \sum_{\mu=1}^{\infty} [n+\mu] z^{\mu}$$

The remaining sum can clearly be written in the form

$$z^{-n} \left\{ \psi(z) - \sum_{p=0}^{n} [p] z^p \right\}$$

If we introduce the designation

$$Q(z) = \sum_{p=0}^{n} [p] z^p \quad (22)$$

we obtain finally

$$e^{y(1-z)} \psi(z) = a_n + z^{-n} \{\psi(z) - Q(z)\}$$

or

$$\psi(z) = \frac{Q(z) - a_n z^n}{1 - z^n e^{y(1-z)}} \quad (23)$$

The denominator in this expression has a zero place for $z = 1$, and *Crommelin* has shown that altogether it has n zero places, the absolute values of which ≤ 1. Now since $\psi(z)$ from what was just stated is to be limited for all $|z| \leq 1$, the numerator must also contain these roots. Now the numerator is a polynomial of the nth degree in z and will therefore be fully determined through these roots except for a constant factor. If the roots are $1, \lambda_1, \lambda_2, \cdots \lambda_{n-1}$, then we have

$$Q(z) - a_n z^n = k(z-1)(z-\lambda_1) \cdots (z-\lambda_{n-1})$$

The constant k is obtained from

$$\lim_{z \to 1} \psi(z) = 1$$

The limit is obtained in the usual way by deriving the denominator in $\psi(z)$ and setting $z = 1$. We then get

Pollaczek published in 1930—1934 (bibliography 13—19) a number of investigations into these problems, the early ones being based on conditions which are not present for applications in telephony. The final solutions were published at about the same time by *Pollaczek* and *Crommelin,* whose works appeared 1932—1933 (bibliography 19 and 2).

Pollaczek in his investigations considers to begin with a finite space of time T, during which a given number of calls come in randomly distributed. The times of the holdings caused by these calls obey an arbitrary distribution law. By means of an extensive mathematical apparatus there are then obtained particularly complicated expressions for congestions and delays. To make these manageable there is considered the case $T \rightarrow \infty$ with which the assumptions change into those applying to ordinary random traffic. No solution that could be used for numerical computation was produced, however, for the general holding time distribution function but only if this is exponential, when the *Erlang* results given above are obtained, or constant, when the results are still complicated yet possible to evaluate.

Crommelin's procedure is of quite a different nature. He starts out direct from random traffic, in which all holding times have the same length, and by means of relatively simple mathematical aids arrives quickly at general expressions for the solutions, for which afterwards series expansions of the same kind as *Pollaczek*'s are obtained.

Respecting the literature dealing with delay systems with constant holding time, there may finally be mentioned a paper by *Berkeley* (bibliography 1), in which some of *Crommelin*'s formulæ are reproduced and explained in some detail and a number of comparative computations performed. The results are compared with a number of tests on »artificial» traffic, which are incorrectly conceived however.

Here we shall review some of the concordant *Pollaczek* and *Crommelin* results, without going into the details of the deductions. Nevertheless, a brief survey will be given of *Crommelin*'s fundamental equation partly because through it there is obtained an idea of the essential elements in the method of treatment and partly because it gives a particularly fine example of the employment of generating functions, a method often employed in the theoretical treatment of traffic problems. Otherwise attention may be directed to *Crommelin*'s work (bibliography 2) and to *Pollaczek*'s latest work (bibliography 19) which give a final summary of the whole problem.

We consider a full availability group with n devices in a delay system, with the incoming traffic random and y the mean value of the number of calls per unit of time. The holding times are always of equal length. For the sake of simplicity we make the unit of time equal to the constant holding time, so that $s = 1$ and the traffic flow will be equal to y. As before, we introduce the notation $[p]$ for the probability that p but not more of the n devices are occupied at one time. In this case it is appropriate to retain the expression $[p]$ even though $p > n$, in which case by $[p]$ we mean the probability that all devices are occupied and $p - n$ calls waiting, that is what we previously denoted by $[n, p-n]$. Making use of *Crommelin*'s notation we introduce now

$$a_\nu = \sum_{p=0}^{\nu} [p] \qquad (18)$$

We consider an interval of time of the length 1. Since all holding times have the length of 1, holdings proceeding at the beginning of the time interval must all have terminated at the end of the interval. Any delayed calls that may be present at the beginning of the interval and all the calls coming in during the interval of time considered must at the end of the interval either form occupations in progress or wait (here we consider only the case that all calls subjected to delay continue to wait until occupation is obtained). If a state p exists at the close of the interval, one of the following cases must have arisen:

at the beginning of the interval there prevailed one of the states $0, 1 \ldots n$ and during the interval p new calls have come in,

at the beginning of the interval there prevailed the state $n + 1$, i.e. there was one waiting call, and during the interval $p - 1$ new calls have come in,

etc. and finally

at the beginning of the interval there prevailed the state $n + p$ and during the interval no new calls have come in.

the waiting queue being always served in order there is no need when deducing the distribution function (21) for the qth waiting call to take any account of the further calls coming in during the delay to take up places further along the queue. It is this condition which, among other things, makes itself apparent by (21) not containing y, which makes the deduction for ordered queue relatively simple. With random served queue there occurs on the other hand an influence on the delay from calls coming in later, which makes conditions more complicated.

Of fundamental importance for the whole delay theory is the circumstance that the dropping out of the occupations is governed by an exponential function of the form

$$e^{-\frac{n}{s}t} \tag{24}$$

which gives the probability that none of the n occupations proceeding during a congestion state terminates during a time t after an arbitrarily selected point of time. Through this function being exponential the probability of any of the waiting calls occupying a device will be independent of the earlier course of events. We may designate (24) the distribution function of the *congestion transition,* meaning by the congestion transition the circumstance that when all devices are occupied one of the occupations terminates and another, originating in a waiting call, immediately takes its place. A congestion state during which calls arrive and must wait will thus display one or more congestion transitions. The distribution function of the congestion transition (24) thus gives the probability that during the time t no congestion transition will occur. It may here be noted that though both this distribution and the distributions for the states n,q are exponential, the same does not apply to the distribution for the durations of the actual congestion states. This, which is of the utmost importance for judging the quality of service, will be treated in more detail in a later article.

In conclusion, we shall refer to an interesting circumstance. In the deduction of the distribution (21) for the qth waiting, we did not need to take into account the magnitude of the state quantities $[p]$ or $[n,q]$. These did on the other hand come in when determining $F(t)$ according to (22). This is not, however, a necessity, but was only employed to simplify the presentation. Basing solely on the congestion transition's distribution (24) and the random distribution of the calls, it is possible to deduce $F(t)$ without making use of the previous results concerning the values of the state quantities. This will, however, make the deduction more complicated. This circumstance does nevertheless have an interest, as it explains why the group's number of devices n does not have any significance for the delays, but only the quantity $s:n$ which is the mean value of the time between two congestion transitions.

Congestions and Delays with Constant Holding Time.

The exponential holding time distribution, for which the results presented earlier in this article are valid, was formerly considered as agreeing well with actual conditions for local traffic. But with trunk traffic and also as regards conditions with a number of connecting devices, which only are involved in the establishing of connexions, such as registers, it was considered more correct to reckon with holding times always equally long. The assumption of constant holding times would moreover appear purely formally to be the simplest possible, so that it should lead to a simple mathematical treatment. This is now found to be by no means the case; on the contrary, conditions will be appreciably more complicated than with exponential holding time distribution. Despite this, several scientists have devoted a considerable amount of work to investigating the conditions with delay systems when the holding times are constant. *Erlang* produced as early as 1909 (bibliography 3) formulæ for delays with the number of devices $n = 1$, a special case which is of interest for all service arrangements where only one person handles the service, e.g. booking offices and the like. Later (1917) *Erlang* published (bibliography 4) solutions also for the cases $n = 2$ and 3. Unfortunately no proof was furnished for the results, which were fairly complicated in form, a circumstance that has perhaps contributed to their seeming not to have been given proper attention. Moreover, it seems quite probable that *Erlang* was in possession of methods also for the treatment of cases with greater n values. It was not until around 1932 that there appeared, thanks to the investigations of *Pollaczek* and *Crommelin,* general solutions for the congestions and delays for a general n-value.

when the state $n,q-1$ is prevailing and therefore there were $q-1$ waiting before, will not obtain an occupation within the time t will then be

$$f_q(t) = \left\{1 + \frac{n}{s}t + \frac{1}{2!}\left(\frac{n}{s}t\right)^2 + \cdots + \right.$$
$$\left. + \frac{1}{(q-1)!}\left(\frac{n}{s}t\right)^{q-1}\right\} e^{-\frac{n}{s}t} \qquad (21)$$

We term this the distribution function for the qth waiting. Now an average of $y[n, q-1]$ calls are offered to the group per unit of time, these occurring when the state $n, q-1$ prevails, and of these calls the fraction $f_q(t)$ will have to wait the time t. There will then occur on the average per unit of time a total of

$$\sum_{q=1}^{\infty} y[n, q-1] f_q(t)$$

calls, which are subjected to at least the delay t. Now there arise on the average per unit of time

$$\sum_{q=1}^{\infty} y[n, q-1]$$

calls which are subjected to delay. The quotient between these two expressions, i.e.

$$F(t) = \frac{\sum_{q=1}^{\infty}[n, q-1] f_q(t)}{\sum_{q=1}^{\infty}[n, q-1]} \qquad (22)$$

will then give that part of the total number of waiting calls, for which the delays are at least t long. $F(t)$ according to (22) is therefore the distribution function of the delays and expresses the probability that a call selected at random which has to wait will then have a delay that is at least t long.

The general expression (22) can now be appreciably simplified by inserting $f_q(t)$ according to (21) and $[n,q]$ according to (2 b). We get first

$$F(t) = \frac{e^{-\frac{n}{s}t} \sum_{q=0}^{\infty} \sum_{\mu=0}^{q} \left(\frac{A}{n}\right)^q \frac{\left(\frac{n}{s}t\right)^{\mu}}{\mu!}}{\sum_{q=0}^{\infty}\left(\frac{A}{n}\right)^q}$$

The denominator in this constitutes a geometrical series with the sum $1 : \left(1 - \frac{A}{n}\right)$. The double sum in the numerator can, by reversing the summation order, be written

$$\sum_{\mu=0}^{\infty} \frac{\left(\frac{n}{s}t\right)^{\mu}}{\mu!} \sum_{q=\mu}^{\infty}\left(\frac{A}{n}\right)^q$$

The sum over q is a geometrical series with the sum

$$\frac{\left(\frac{A}{n}\right)^{\mu}}{1 - \frac{A}{n}}$$

There then remains of the whole double sum

$$\frac{1}{1 - \frac{A}{n}} \sum_{\mu=0}^{\infty} \frac{\left(\frac{n}{s}t\right)^{\mu}\left(\frac{A}{n}\right)^{\mu}}{\mu!}$$

which is

$$\frac{e^{yt}}{1 - \frac{A}{n}}$$

After these simplifications there is obtained from the expression for $F(t)$:

$$F(t) = e^{-\left(\frac{n}{s} - y\right)t} \qquad (23)$$

Sojourn Time ~ ~ exp(mean = $\frac{1}{n\mu}\frac{1}{1-\rho}$ = $\frac{1}{n\mu - \lambda}$ = $\frac{s}{n-A}$, as (19).

The distribution function of the delays has thus proved to be a simple exponential function. It is easily seen that the mean delay will be equal to the expression (19) deduced before. We again note the necessity of having $A < n$, as otherwise the series will not be convergent.

The expression (23) was given first by *Erlang* (bibliography 4). A deduction for it appears to have been first published by *Molina* (bibliography 5). In the same way as in respect of mean delay it occurs occasionally that one means by the distribution function of the delays the probability that a random selected call, irrespective of whether it must wait or not, will have at least the delay t. This distribution function will obviously be equal to

$$E_{2,n} \cdot F(t)$$

in all points, except when $t = 0$, when it is equal to 1. In this point therefore the function has a negative leap of the magnitude $1 - E_{2,n}$.

As regards the deduction shown above, some conditions may be specially mentioned. Owing to

The sum in the denominator is a common geometrical series, of which the value is

$$\sum_{q=0}^{\infty}\left(\frac{A}{n}\right)^{q}=\frac{1}{1-\frac{A}{n}}$$

If we derive both the members of this in respect of $A:n$ we get

$$\sum_{q=1}^{\infty}q\left(\frac{A}{n}\right)^{q-1}=\frac{1}{\left(1-\frac{A}{n}\right)^{2}}$$

The left member in this is except for the factor $A:n$ equal to the sum in the numerator in the expression just obtained for mean delay. After some conversion there is then obtained for the mean delay the simple expression

$$\frac{s}{n-A} \qquad (19)$$

As the whole of the traffic offered to the group is with delay system served within the group, $n - A$ constitutes the traffic reserve for delay system. For the case here considered, when all those waiting continue to wait until they obtain occupation, the mean delay will be equal to the mean holding time divided by the traffic reserve.

The expression shown above for mean delay was first stated by *Erlang* (bibliography 4). It is valid irrespective of what rules and what order apply for the serving of those waiting, seeing that in the deduction no conditions in this respect were introduced.

It should be noted that the expression (19) is valid for the delays actually occurring. It thus expresses the probable delay for a call subjected to delay. In contrast with this, it is sometimes usual to imply by mean delay the probable delay for a call selected at random and offered to the group. The denominator in (18) is then replaced by y alone. It is easily seen that the mean delay for all calls is obtained from the mean delay (19) valid solely for the calls which are really subjected to delay by multiplying the latter by $E_{2,n}$ which, as seen, expresses the probability that a call selected at random is compelled to wait. The mean delay for all calls offered to the group is then

$$\frac{s}{n-A}E_{2,n} \qquad (20)$$

In deciding dimensions of plant, both the expressions (19) and (20) are of interest. It seems, however, appropriate to reserve the term mean delay for (19), which indeed gives the mean value of the delays actually occurring and thereby gives a better idea than (20) of their magnitudes. Let us for example consider a group with 20 devices and the offered traffic of 11 erlangs. If the mean holding time is 126 secs., the mean delay will be 14 secs. according to formula (19). In this case the congestion is 0.010 and the mean delay for all calls according to (20) will be 0.14 secs., a quantity which can hardly in itself give any idea of the real magnitude of the delays.

Let us now determine the distribution function of the delays in the case with ordered queue, i.e. when those waiting are served in the same order as the calls came in. We then consider first a call offered the group when the state n, $q-1$ prevails. There will then be $q - 1$ waiting, who have priority, and only after all these have occupied devices does the call considered occupy the next device becoming free. Each time one of the occupations proceeding terminates all those waiting in the queue move forward one step. A person waiting who has $q - 1$ waiting before him in the queue must therefore go through q stages before his call is served. During the first stage he has $q - 1$ waiting before him, during the second stage $q - 2$ waiting before him and so on, so that in the qth stage he is at the head of the queue. The probability that a stage will not terminate in the time t after it started is now equal to the probability that none of the n occupations in progress will terminate in the same time, and is thus equal to

$$e^{-\frac{n}{s}t}$$

as we had assumed that the holding times follow the exponential distribution $e^{-\frac{t}{s}}$. The delays therefore follow a distribution built up of the exponentially regulated stages of the kind treated in the first article of this number. In addition, we have the special case when the mean durations of all stages are equally great and equal to $s:n$ according to the formula just above. The distribution function for such a case is given by equation (30) in the article just referred to. Here s_0 in that formula must obviously be replaced by $s:n$. The probability that a call, arriving there is still another circumstance which influences to a high degree the distribution function of the delays, and that is the order in which the waiting calls are served. Sometimes the serving is so arranged that when there are several waiting calls at one time that which has been waiting the longest always receives the first device becoming unoccupied. Thus those waiting are served in the order in which their calls arrived.

It seems appropriate for systems where the selection among those waiting takes place according to this rule, to speak of *ordered queue*. Such systems with the formation of ordered queues are to be found, besides in telephony, in most of the places where people are served in everyday life, so that the delay conditions in this case have great general interest. Mostly there occurs, however, with delay systems in telephony an arrangement of service differing altogether from the ordered queue. For technical reasons it is then so arranged that when there are several waiting and some device becomes unoccupied, it is more or less an accident which of those waiting obtains occupation of this device. If therefore, the selection of the one waiting who is first served may be considered to be entirely random, one speaks of a *random handled queue*. In most cases, of course, it would seem that the random selection was not exactly realised, though it appears that usually one may calculate as if a purely random selection were present. In some cases, however, there are such arrangements that the system must be treated as a mixture of ordered and random queue. It is worth pointing out that service arrangements corresponding to random handled queue formations also occur in other fields besides telephony. Such a case which should be of interest for the science of economics is offered, for example, by the redemption of a bonded loan when this is done by periodical drawing of the bonds by lot.

For the case where the holding times follow an exponential function and the serving of those waiting is done in ordered queue, *Erlang* has already stated the general solution which will be treated in this section. If with the same holding time distribution the serving is done by random selection of those waiting, the mathematical treatment immediately becomes considerably more complicated. This case will be examined more closely in a later article.

Though the distribution function of the delays, as stated, turns out different according to the rules on which the selection among those waiting is made, this does not hold good for the mean delay. Moreover it is quite natural that this should be independent of the manner of serving the queue. The state quantities $[p]$ and $[n,q]$ were in fact fully determined without any conditions regarding the manner of service being required. Now the total duration of the different states n,q determine the total delay for all those waiting, and with this the mean delay is also determined. We shall show here how this can be computed with the usual assumption of exponential holding time distribution.

When a state n,q prevails there are q waiting. On the average per unit of time this state will prevail for the time $[n,q]$ and there will then be obtained during the same time a total waiting time of the magnitude $q[n,q]$. Altogether for all the different states there is then obtained on the average per unit of time the total delay

$$\sum_{q=1}^{\infty} q\,[n,q]$$

Now all calls, which come in when all devices are occupied must wait. On the average per unit of time there is therefore obtained

$$y \sum_{q=0}^{\infty} [n,q]$$

waiting calls. The mean value of the delay for a call which is obliged to wait is now equal to the total delay for all calls divided by the total number of waiting calls and will thus be

$$\frac{\sum_{q=1}^{\infty} q\,[n,q]}{y \sum_{q=0}^{\infty} [n,q]} \tag{18}$$

This expression is valid quite generally. For exponential holding time distribution there now applies the formula (2 b) for the state quantities and the above expression then changes to

$$\frac{\sum_{q=1}^{\infty} q \left(\frac{A}{n}\right)^{q}}{y \sum_{q=0}^{\infty} \left(\frac{A}{n}\right)^{q}}$$

$$P_\mu = \frac{A^\mu}{\mu!} e^{-A}$$

Since

$$\sum_{\mu=0}^{\infty} P_\mu = 1$$

(5) may be written in the form

$$E_{2,n}\left(1 + \frac{A}{n-A} P_n - \sum_{\mu=n+1}^{\infty} P_\mu\right) = \frac{n}{n-A} P_n$$

or

$$\frac{\frac{n}{n-A} P_n}{E_{2,n}} - 1 = \frac{A}{n-A} P_n - \sum_{\mu=n+1}^{\infty} P_\mu \quad (16)$$

The right hand member in this is always positive, which can easily be shown by maximising the sum therein by means of a geometrical series. From this it follows that

$$\frac{n}{n-A} P_n > E_{2,n}$$

If we drop the sum in the right member in (16) we then obtain the difference

$$\frac{\frac{n}{n-A} P_n}{E_{2,n}} - 1 < \frac{A}{n} \cdot \frac{n}{n-A} P_n \quad (17)$$

Thus $\frac{n}{n-A} P_n$ may be used as approximate value for $E_{2,n}$ the left member in (17) expressing the relative error maximised by the right member. The approximation is only usable for relatively small values of $E_{2,n}$ up to 0.2 to 0.3, but is very much better than the corresponding approximation for $E_{1,n}$ that was shown in the preceding article. Moreover in the latter case there is obtained a smaller value than the real one, whereas for $E_{2,n}$ according to (17) one obtains a greater value than the real one and is therefore on the safe side.

As an example of the serviceability of the approximation we may take $A = 15$ and $n = 21$. We then have $E_{2,n} = 0.10232$, while with the approximate value we get 0.10453. The relative error in this is only 2.2 %, whereas according to (17) the maximum limit for the error will be 7.5 %. In general one can reckon on the accuracy being appreciably greater than (17) gives one to suppose.

For purposes of comparison there have also been plotted in fig. 1 curves for the approximate value $\frac{n}{n-A} P_n$ as a function of A. As may be seen, a noticeable deviation from the $E_{2,n}$ curves only appears for over 10 % congestion.

It is clear that we can also employ P_p as an approximate value for the state magnitudes $[p]$ with delay systems. As all state quantities have the same denominator, the relative error in this case will be equally as great as for the approximate value for $E_{2,n}$ and it is thus maximised by the right member in (17).

Finally it may be noted that in the fields where the approximation formulæ are valid both for busy-signal and for delay systems we also have

$$E_{2,n} \approx \frac{n}{n-A} E_{1,n}$$

the agreement, however, not being then so good as with the approximate value maximised by (17).

Delays with Ordered Queue.

In judging the quality of service in a delay system the size of the delays is of importance as well as the congestion. A representative measure for the waiting times occurring is constituted by the mean delay, which nevertheless cannot in itself be considered as giving sufficient information for an all-round judging of the inconveniences to the subscribers associated with the delay. For this purpose it is desirable also to know how many of the delays are so long that they are regarded by the subscribers more or less as an inconvenience. It is therefore necessary to know the distribution function of the delays, that is how many of the whole number of delays occurring are longer than different times t.

The calculation of the distribution function of the delays is generally found to be a highly troublesome problem, though for a given case, which will be treated in this section, it proves easy to perform. The form of the distribution function will be dependent on what distribution function applies to the holding times. Only if the latter is exponential will the mathematical treatment be relatively simple. Nevertheless, the case with constant holding time has also been attacked by several scientists, an account of whose results will be given later on. Besides the nature of the distribution function of the holding times, however,

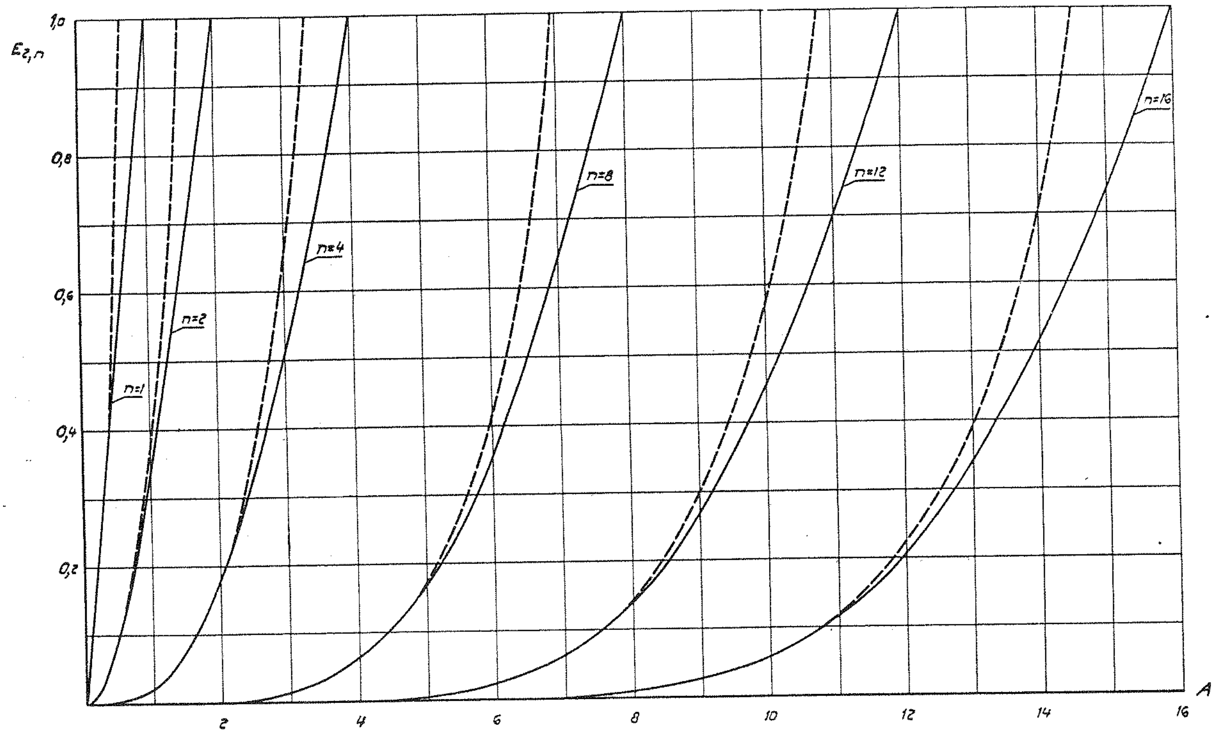

Fig. 1. The congestion $E_{2,n}$ as a function of A for different n.

——— $E_{2,n}$ — — — $\frac{n}{n-A} \cdot P_n$

busy-signal system, which is done with the help of the recurrence formula (11) in the preceding article. To compute individual congestion values it is often an advantage to use an approximate formula, obtained by means of the *Poisson* expressions mentioned in the preceding article

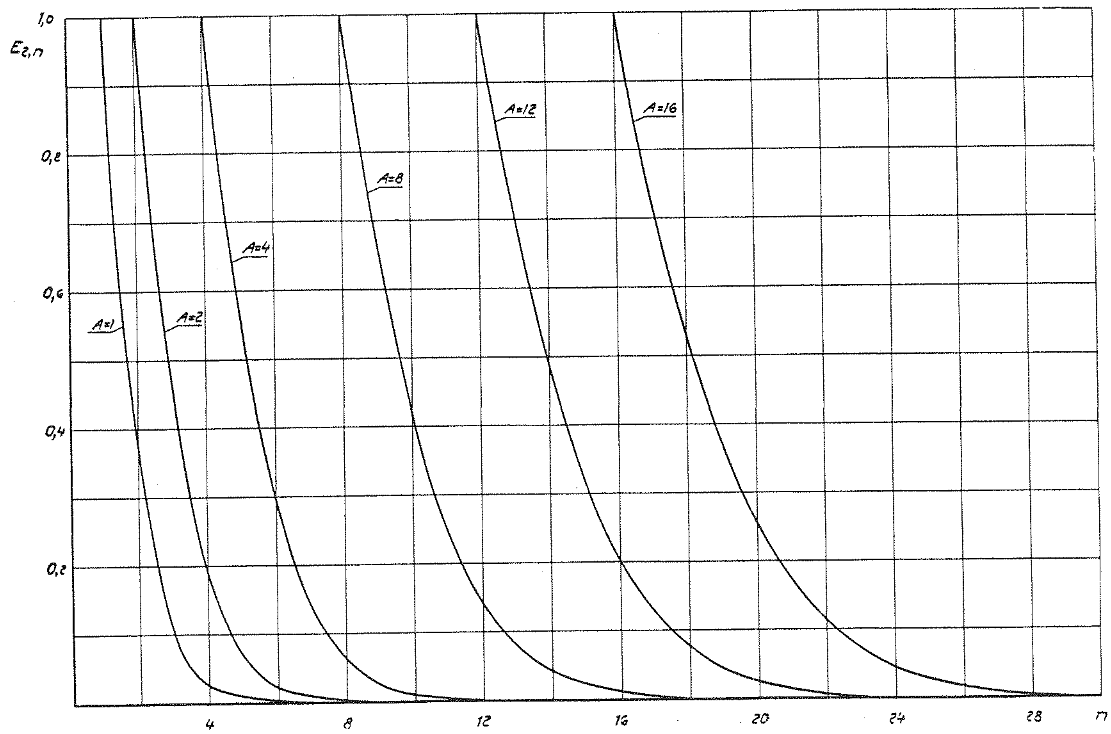

Fig. 2. The congestion $E_{2,n}$ as a function of n for different A.

of (5) or one of the expressions (9) it is therefore easy to verify the validity of the formula

$$E_{2,n} = \frac{A(n-1-A)E_{2,n-1}}{(n-1)(n-A) - A \cdot E_{2,n-1}} \qquad (10a)$$

from which, inversed, there is obtained

$$E_{2,n-1} = \frac{(n-1)(n-A)E_{2,n}}{A(n-1-A+E_{2,n})} \qquad (10\,b)$$

These formulæ are rather complicated, however. More interesting for the discussion are then the relations which may be obtained between $E_{2,n}$ and the loss $E_{1,n}$ in busy-signal systems. For the latter there is valid according to formula (7) in the preceding article

$$E_{1,n} = \frac{\frac{A^n}{n!}}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^n}{n!}}$$

It is then easy to verify the following relation

à la Massey

$$E_{2,n} = \frac{A \cdot E_{1,n-1}}{n - A + A \cdot E_{1,n-1}} \qquad (11\,a)$$

which in form reminds us of the recurrence formula (11) in the preceding article. By inversion of the formula above there is also obtained

$$E_{1,n-1} = \frac{n-A}{A} \cdot \frac{E_{2,n}}{1 - E_{2,n}} \qquad (11\,b)$$

Moreover, the following relation may be deduced

$$E_{2,n} = \frac{n \cdot E_{1,n}}{n - A + A \cdot E_{1,n}} \qquad (12\,a)$$

from which by inversion

$$E_{1,n} = \frac{(n-A)E_{2,n}}{n - A \cdot E_{2,n}} \qquad (12\,b)$$

Provided that $n > A$ it is seen from (11 a) that we always have $E_{2,n} < 1$. The denominator in (12 a) is moreover equal to the traffic reserve in a busy-signal system with n devices and therefore always less than n. From (12 a) there then follows the difference

$$E_{2,n} > E_{1,n} \qquad (13)$$

In the preceding article it was demonstrated that $E_{1,n}$ always grows with rising A. The numerator in (12 a) will therefore grow with rising A. Moreover the denominator in (12 a) must decrease with rising A, as it expresses the traffic reserve in a busy-signal system, which of course decreases with growing A. From this it follows that $E_{2,n}$ always grows with rising A.

For $A = 0$ we have $E_{2,n} = 0$. For $n = 1$ we have $E_{2,n}$ a straight line in accordance with (7). For greater n values we see, e.g. from (9 a), that $E_{2,n}$ has an $(n - 1)$-fold tangent point with the A axis in origin. For the derivative of $E_{2,n}$ in respect of A we can obtain the expression

$$\frac{dE_{2,n}}{E_{2,n}} = \left\{ n - A + \frac{A}{n-A}(1 - E_{2,n}) \right\} \frac{dA}{A} \qquad (14)$$

which is somewhat more complicated than the similar expression for the derivative of $E_{1,n}$ (see preceding article, formula (13 b)). It can be shown that the derivative of $E_{2,n}$ always grows with growing A. Its maximum value then occurs for $A = n$. A simple formula can be deduced from this maximum value. From (9 b) there is obtained in fact

$$\frac{1 - E_{2,n}}{n - A} =$$

$$= \frac{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^{n-1}}{(n-1)!}}{(n-A)\left(1 + A + \frac{A^2}{2!} + \cdots + \frac{A^{n-2}}{(n-2)!}\right) + n\frac{A^{n-1}}{(n-1)!}}$$

If now, taking this into consideration, we let $A \to n$ in (14), there is obtained

$$\frac{d}{dA} E_{2,n}(n) = \frac{1}{n \cdot E_{1,n}(n)} \qquad (15)$$

Fig. 1 shows how $E_{2,n}$ varies with A for some different n values. It is also of interest how $E_{2,n}$ varies with n for constant A. Unfortunately we do not have for this case any equally easily discussed difference formula as with busy-signal systems. If we join up by curves the points obtained for different integral values of n then we get charts of the appearance shown in fig. 2. The curves start for integral values of A in the point $n = A$, where $E_{2,n} = 1$. If A does not have an integral value, the curves only start at the next higher integral value of n.

For tabulating $E_{2,n}$ the recurrence formula (11 a) is found to be particularly convenient. It is advisable to make the computation at the same time as the computation for the loss $E_{1,n}$ in a of the whole time when there is no waiting call. During this part of the time the load conditions, as observed above, are the same as in a busy-signal system and thus the pth device is occupied on the average the time α_p. During the time when there is no waiting call there is therefore carried by the pth device in a waiting system per unit of time the traffic

$$\alpha_p \left(1 - \frac{A}{n} E_{2,n}\right)$$

During that part of the whole time, $\frac{A}{n} E_{2,n}$, when there are waiting calls, all devices are occupied. The addition to the traffic on the pth device for this time is thus $\frac{A}{n} E_{2,n}$. Altogether then there is handled by the pth device on the average per unit of time the traffic

$$\alpha_p \left(1 - \frac{A}{n} E_{2,n}\right) + \frac{A}{n} E_{2,n}$$

If we write this expression for the load on the device hunted pth in order in a group of n devices in a delay system in the form

$$\alpha_p + \frac{A}{n}(1 - \alpha_p) E_{2,n} \qquad (8)$$

it is seen that, since α_p constitutes the load on the pth device in a busy-signal system with the same incoming traffic and the same number of devices, then the second term in the expression (8) gives the increase in the load which arises in a waiting system owing to the serving of the delayed calls. If the congestion $E_{2,n}$ is rather small, as it normally is, this increase will be very slight except for the very highest devices, where its value in relation to α_p can be appreciable.

The formula (8) given above is, as already stated, valid only for the case that the holding times are exponentially distributed. It was produced in 1925 by *Vaulot* (bibliography 20), but appears to have attracted little attention, as it has never been referred to previously.

Certain analogies have been pointed out above between the traffic conditions in busy-signal and delay systems, which moreover will be further exemplified and enlarged upon later. There are, however, also highly essential differences in the two systems and as an expression for what is perhaps the most important difference the following may be observed. In a group with ordered hunting in a busy-signal system any successive devices whatever can be treated as a detached group in a busy-signal system. This condition is due to the fact that the lost calls never return. A group of n devices in a delay system can, on the other hand, in no circumstances be divided up into smaller detached groups. Which device a waiting call will gradually come to occupy depends in fact upon the whole group's state and can therefore not be computed solely from knowledge of the traffic conditions in a part of the group's devices.

Form of the Congestion Function. Methods for Numerical Computations.

The congestion function $E_{2,n}$ according to (5) has in part a similar build up as the function $E_{1,n}$ for busy-signal systems (see preceding article, formula (7)). Despite this there prevails a rather essential difference in respect of the course of the two functions, which is due to the occurrence in (5) of the quantity $n:(n - A)$, which for great A values has a pronounced effect on $E_{2,n}$. By multiplying numerator and denominator by $(n - A):n$ we can now bring (5) to the form

$$E_{2,n} = \frac{\dfrac{A^n}{(n-1)!}}{\displaystyle\sum_{\nu=0}^{n-1} (n-\nu)\frac{A^\nu}{\nu!}} \qquad (9\text{ a})$$

or

$$E_{2,n} = \frac{\dfrac{A^n}{(n-1)!}}{(n-A)\left(1 + A + \dfrac{A^2}{2!} + \cdots + \dfrac{A^{n-2}}{(n-2)!}\right) + n\dfrac{A^{n-1}}{(n-1)!}} \qquad (9\text{b})$$

From this it is seen that the numerator is of the power n and the denominator of the power $n - 1$ in A. In the expression for $E_{1,n}$ on the other hand, both numerator and denominator are of the power n. For $A = n$ we have according to (9 b) the congestion $E_{2,n} = 1$ and for greater A values $E_{2,n}$ has no physical significance.

As we did when dealing with the loss in busy-signal systems we can for $E_{2,n}$ too, set up recurrence formulæ for successive n values. By means

A consequence of the condition shown is that for $n = 1$ we obtain

$$E_{2,1} = A \tag{7}$$

which of course can also be obtained from the general expression (5). The formula (7) has, however, a more general range of application than the equation (5). When a group in a delay system consists only of one single device, the whole of the traffic A offered, must be carried by that device, which will then altogether be occupied the time A per unit of time. As the group only comprises this device, congestion prevails when it is occupied, which means that A also expresses the congestion. This reasoning must obviously be valid irrespective of the general deduction for (5), and we then find the interesting proposition that the special formula (7) is valid irrespective of whether the hunting traffic is random and whether the holding times follow some given distribution function.

In conclusion we shall show some formulæ for the individual devices' loading in a group with an arbitrary number of devices. For busy-signal systems there were obtained in the preceding article the formulæ (9) and (10). With delay systems the traffic carried is equal to the traffic offered A. The grade of utilisation, i.e. the mean load on all devices, will then be $\alpha = \frac{A}{n}$, with n representing the number of devices in the group. If hunting in the group is random, so that all devices are utilised equally, each individual device will on the average be occupied the part $\frac{A}{n}$ of the whole time. Again if hunting in the group takes place in fixed order, the different devices will be loaded to an unequal degree. Though for computation of the loads we find no formula so directly apparent as with busy-signal systems, yet results may be attained by means of a fairly simple discussion. To carry this out we shall first note a significant relation between the traffic conditions with busy-signal and delay systems. This relation is derived from the fact that, provided there is exponential holding time distribution, a difference between busy-signal and delay systems can only arise when congestion prevails.

Let us consider a delay system at a moment when the state $n - 1$ has arisen out of a state $n,0$ and thus a congestion state has just been interrupted. In the period up to the next time the state $n,1$ arises and a call is thus compelled to wait, all incoming calls are dealt with in the same manner as if the group belonged to a busy-signal system. If now in addition, and this is a necessary condition, the holding times follow an exponential function, then as is known the probabilities of the continued durations of the $n - 1$ occupations which were proceeding when the state $n,0$ ceased, are independent of how long the occupations have already lasted. On account of this the statistical laws which are valid both for the traffic conditions during the interval of time under consideration, that is between the end of a congestion state and the next time a call has to wait, will be the same as in a busy-signal system. From this it follows that for such intervals of time one may count on the same relative values for states and loads as with busy-signal systems.

In respect of the states the condition mentioned can be shown direct from the formulæ. We consider the expression

$$\frac{[p]}{[n,0] + \sum_{p=0}^{n-1} [p]}$$

which obviously gives the relation between that part of the whole time when state p prevails and the part of the whole time when no call is delayed. From the equations (2 a) and (2 b) there is obtained for this expression

$$\frac{\frac{A^p}{p!}}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^n}{n!}}$$

which is the same as the state quantity $[p]$ in a busy-signal system with n devices (see preceding article).

We shall now employ the relation demonstrated to compute the load on a device in a delay system with ordered hunting. According to the formula (10) in the preceding article the load on the pth device in a busy-signal system is

$$\alpha_p = A\,(E_{1,p-1} - E_{1,p})$$

Now (6) expresses that part of the whole time when there is at least one waiting call in the delay system. Therefore $1 - \frac{A}{n} E_{2,n}$ gives that

The sum of all waiting states $[n,q]$ when all devices of the group are occupied may be denoted $E_{2,n}$. We find

$$E_{2,n} = \frac{\frac{A^n}{n!} \cdot \frac{n}{n-A}}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^{n-1}}{(n-1)!} + \frac{A^n}{n!} \cdot \frac{n}{n-A}} \tag{5}$$

The solution indicated was published by *Erlang* in 1917 (see bibliography 4). The quantity $E_{2,n}$ expresses the *congestion* or, as is also sometimes said, the *full occupation time*. The subscript 2 is introduced here to distinguish between the *Erlang* expressions for busy-signal and for delay systems. Another quantity which is of interest in making calculations for dimensions is the sum of all the waiting state quantities $[n,q]$, when $q > 0$ and thus at least one waiting call is present. For this sum there is obtained

$$\sum_{q=1}^{\infty} [n, q] = \frac{A}{n} E_{2,n} \tag{6}$$

This expression may, as a suggestion, be designated the *delay arrangement's functioning time.*

For proof of the equations given, attention may be directed to bibliography 7. The assumptions valid for the deduction may be summarised as:

that the group is fully available,

that all congested calls wait until they obtain occupation,

that the traffic hunting the group is random,

that the holding times constitute a purely exponential distribution,

that the hunting traffic A is less than the number of devices n.

As regards the last-named condition it may be observed that for $A \rightarrow n$ one gets $E_{2,n} = 1$ in the limit. For $A > n$ the hunting traffic cannot be wholly served by the group, which means that no statistical equilibrium is ever attained. Generally it may be observed, that among the conditions above nothing is said regarding the order in which the devices are hunted, or concerning the rules which, with several calls waiting, apply for the selection of the waiting call which is to be the first to obtain occupation. The formulæ stated above, therefore, are valid irrespective of the conditions prevailing in those respects.

The expression (5) for the congestion leads one to suppose that this is greater than with a busy-signal system having the same number of devices and the same incoming traffic. Later it will be proved by purely mathematical means that this is always the case. This circumstance implies nothing surprising, as in delay systems all the hunting traffic is served within the group, so that the traffic carried, other conditions being equal, is greater than with busy-signal systems. Moreover, the serving of that traffic which if the delay arrangement were not adopted would be rejected by the group, takes place of course for the most part during the time when all devices are occupied. In consequence of this, compared with a busy-signal system, the congestion time with delay system must be greater and all states when congestion does not prevail must be smaller. The condition may also be illustrated in the following manner. If we have a busy-signal system, in which we imagine that all subscribers whose calls meet congestion continually repeat their calls at theoretically infinitely small intervals, then there are obviously obtained the same conditions in the group as with a delay system and the congestion will be expressed by $E_{2,n}$, instead of by $E_{1,n}$. One has then assumed the maximum possible *reaction of the congestion.* We can draw from this the valuable conclusion that the congestion reaction with a busy-signal system will produce an increase of the congestion beyond $E_{1,n}$, which cannot ever be greater than $E_{2,n}$, however.

It may be of interest to show by purely mathematical means that the traffic carried in the group with a delay system really is equal to the traffic A offered to the group. This can easily be done. While a state p is prevailing, there is obviously carried a traffic load of p times the state's duration. While an arbitrary waiting state n,q is prevailing there is further carried a traffic load of n times the state's duration. The total traffic carried in the group per unit of time is then expressed by

$$\sum_{p=1}^{n-1} p\,[p] + n \sum_{q=0}^{\infty} [n, q]$$

and it is easily seen that by means of the equations (2 a) and (2 b) this may be written

$$A \left\{ \sum_{p=0}^{n-1} [p] + \sum_{q=0}^{\infty} [n, q] \right\}$$

which on account of (3) is reduced to A.

able for practical computation. Fortunately there is reason to suppose that the conditions with other distribution functions for the holding times quite closely resemble those occurring with exponential distribution. The conditions with delay systems, therefore, thanks to research in recent times, may be considered as mainly cleared up, even though many important questions require still further investigation.

As regards the treatment of traffic conditions with delay systems, too, *Erlang*'s contributions are of fundamental importance. His results in respect of delay systems are published in a number of articles together with the treatment of busy-signal systems and comprise as well the conditions with exponential holding times distribution, for which he draws up formulæ for losses, states, occupations and the distribution of delays in a special case, as also a number of special formulæ applying to constant holding time. The results will be found in bibliography 4, and they are as always where *Erlang* is concerned entirely correct, though the proofs given are hardly convincing and partly omitted.

To describe *Erlang*'s results we consider a full availability group of n devices in a delay system. The traffic hunting the group is assumed to be random and to have the intensity $A = s \cdot y$, with s the mean holding time and y the mean number of calls offered per unit of time. The holding times are assumed to follow the distribution function $e^{-\frac{t}{s}}$, which therefore expresses the probability that an occupation will last at least the time t. It is further assumed that all subscribers who cannot owing to congestion be served immediately wait until they obtain occupation of a device.

As with busy-signal systems we now introduce expressions for different states in the group, though we have here to consider two different kinds of state, depending on whether waiting calls are present or not. With the state p, in which $p < n$, there is meant the condition that p, but none more of the group's n devices are occupied simultaneously. The mean value per unit of time of the total time when the state p prevails may be denoted $[p]$. Further there is meant by the state n,q the condition that all n devices are occupied and at the same time q calls are waiting to obtain occupation. For this q may be any integral value or zero. The mean value per unit of time of the total time when such a waiting state prevails may be denoted $[n,q]$. It can now be shown that the following relations exist between the different state quantities:

$$[p] = \frac{A}{p}[p-1] \qquad (1\,a)$$

$$[n,0] = \frac{A}{n}[n-1] \qquad (1\,b)$$

$$[n,q] = \frac{A}{n}[n,q-1] \qquad (1\,c)$$

It should be noted that the coefficient in (1 c) is independent of the value of q. From the system (1) there are now obtained the following formulæ:

$$[p] = \frac{A^p}{p!}[0] \qquad (2\,a)$$

$$[n,q] = \frac{A^n}{n!}\left(\frac{A}{n}\right)^q[0] \qquad (2\,b)$$

In addition there holds good the obvious relation that the sum of the probabilities for all states existing is one, thus

$$\sum_{p=0}^{n-1}[p] + \sum_{q=0}^{\infty}[n,q] = 1 \qquad (3)$$

The second sum in the left member in this, as may easily be seen, from (2 b), constitutes a geometrical series which is convergent for $A < n$. On summation there is then obtained

$$\sum_{q=0}^{\infty}[n,q] = \frac{n}{n-A}\cdot\frac{A^n}{n!}[0]$$

From the system (2) and the relation (3) all state quantities will be explicitly determined. We get: for $p = 0, 1, 2, \ldots n-1$

$$[p] = \frac{\dfrac{A^p}{p!}}{1 + A + \dfrac{A^2}{2!} + \cdots + \dfrac{A^{n-1}}{(n-1)} + \dfrac{A^n}{n!}\cdot\dfrac{n}{n-A}} \qquad (4a)$$

for $q = 0, 1, 2, \ldots$ ad inf.

$$[n,q] = \frac{\dfrac{A^n}{n!}\left(\dfrac{A}{n}\right)^q}{1 + A + \dfrac{A^2}{2!} + \cdots + \dfrac{A^{n-1}}{(n-1!)} + \dfrac{A^n}{n!}\cdot\dfrac{n}{n-A}} \qquad (4b)$$

CONTRIBUTIONS TO THE THEORY ON DELAY SYSTEMS

Delay systems differ from busy-signal systems only in the way in which calls meeting with congestion are dealt with. With busy-signal systems a subscriber whose call is met with congestion must make a fresh call to obtain the wanted communication. With delay system a subscriber who makes a call when all the devices of a group reached are occupied requires only to wait with the receiver off in order to come finally into occupation of a device in the group. Owing to this difference in manner of operating, the inconvenience to which subscribers are subjected with congestion will be of entirely different types in busy-signal and in delay systems. This condition must be carefully considered when judging grade of service. Obviously the manner of operating has also a strong effect on the occupation conditions in a group. In this respect, however, the boundaries between busy-signal and delay systems will be more fluid, seeing that the traffic conditions in a group are not only dependent on the system by which operation of calls in the group is arranged but also on the subscribers' reaction to congestion. If the subscribers in a busy-signal system repeat the calls met with congestion at very close intervals the traffic conditions will be very like those arising in a pure delay system. Again, if the subscribers in a delay system prefer on congestion to break the waiting at an early stage and instead make fresh calls after a fairly long wait, the traffic conditions will closely resemble those arising with a purely busy-signal system. On account of these circumstances the traffic conditions will in reality, with both busy-signal and delay systems, be represented by intermediate conditions between the boundary values that arise if on the one hand, with a busy-signal system, no reaction from the lost calls occurs and on the other hand, with a delay system, no waiting calls disappear before they have obtained occupation. It will be shown later how a mathematical description of such intermediate conditions can be obtained.

As stated in the preceding article the Telegraph Administration's present standards for dimensioning groups of devices are based on the losses in busy-signal systems. This is applied also to delay systems, and the reasons appear to be the following. As stated above, the traffic conditions both with busy-signal and with delay systems will be affected by the subscribers' reactions in the direction of a compromise between the two systems. It should then be possible to count approximately with the same formulæ in both cases, and it then seems more convenient to use the simple expressions for purely busy-signal systems. Since those standards were drawn up, however, knowledge of the nature of congestion has been appreciably extended and strong criticism of the reasoning advanced may now be raised. The compromise stated would appear generally to lie nearest the delay system's traffic conditions, so that there would rather be reason to employ the formulæ for this throughout. In this way there would also be obtained valuable possibilities of varying the dimension prescriptions to take into account the varying conditions of operation in different groups of devices. For these and other reasons, the problems respecting traffic conditions with delay systems have acquired greater interest in recent years. In the present article there will be given some new results applying to full availability groups in delay systems. Some of the older results will also be subjected to scrutiny. A couple of special problems of greater range will be dealt with in the next two articles.

Erlang's Solution with Exponential Distribution of Holding Times.

The theoretical treatment of traffic conditions with delay systems will in general be considerably more complicated than in respect of corresponding problems with busy-signal systems. Thus it is found that the loss conditions with delay systems are no longer, as with busy-signal systems, independent of the distribution of the holding times. Only if this is exponential is it possible to resort to fairly simple methods of treatment and to obtain final formulæ service-