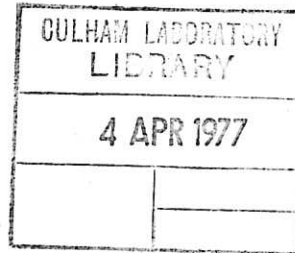


(Revised March 1977)

CULHAM LIBRARY
REFERENCE ONLYSome Observations on the Cost of On-Line Information Retrieval

by

James L. Hall
Culham Laboratory, Abingdon, Oxon. OX14 3DB, England
(EURATOM/UKAEA Fusion Association)

(To be submitted for publication in "On-Line Review")

[Note: Revised version actually published in "On-Line Information Retrieval Sourcebook"]

Abstract

A number of recently reported estimates of on-line search costs are tabulated. Search cost estimates are also derived from a generalized cost-computation approach. Generally, it appears that an "average" search falls in the range \$20-\$40. To help in the preparation of annual budget estimates a graph is derived showing the range of costs which, depending on local conditions, might apply for any given level of search traffic.

Keywords: Budgets; Cost-Use Relationship; Estimates; On-line Searching;
Search Costs.

March 1977

1. Introduction

The literature of on-line retrieval abounds in papers containing cost-of-search information. Unfortunately, these papers are widely scattered, sometimes not readily available, and are often not in a form where they provide ready guidance to the library/information manager trying to assess, for annual budgeting purposes, what estimates can be made for routine on-line retrieval as a service rather than as an experiment.

This paper aims (a) to present very briefly a tabulation of some recently reported cost study results, (b) to derive "average" costs based on a generalized cost-computation approach, and (c) based on the results so found to present a graphical relationship which users may adapt to local conditions to estimate the approximate annual cost of on-line service.

2. Some recently reported cost studies

Well over 100 papers giving search-cost information are available and can readily be traced from the extensive bibliography compiled by Hall [1]. Unfortunately, these results must be used with caution - they reflect widely differing cost-audit conditions and in many cases they are very far from true costs because the cost of equipment/maintenance, telecommunications, overheads, etc., has not been taken into account or because various subsidies apply. Additionally, until recently most reported costs were "research project" costs rather than day-to-day "library query" costs. However, a body of useful experience is now emerging and Table 1 details a number of recently reported results. Papers which are useful but apparently deal with only a very small number of costed searches, e.g. Buckley [2] and Smetana [3], have not been included in Table 1.

While there is, of course, no statistical validity in any attempt to derive an "average search" cost from data obtained under such a wide diversity of conditions, it is none-the-less of interest to note that the

"statistical" average of the search cost figures listed in Table 1 is \$28.30.

Table 1. Cost and time of on-line searching:
some typical reported estimates (1974 - 1976)

Search Cost*	Search Time*	System	Reference
\$7 - \$12	-	BASIS	Hoshovsky [4]
\$15 - \$25	30 minutes	DIALOG	Ahlgren [5]
\$13 - \$30	22 - 32 minutes	DIALOG	Brown [6]
\$14	18 minutes	DIALOG	Caponio [7]
\$47 (long search)	45 minutes (long search)	DIALOG	Elman [8]
\$13 (quick query)	13 minutes (quick query)		
\$28	23 minutes	DIALOG	Cooper [9]
\$44	30 minutes	DIALOG	Hawkins [10]
\$37 - \$77	44 minutes	DIALOG	Williams [11]
\$17	-	ESA-RECON	Page [12]
\$28	41 minutes	MEDLINE	Benenfeld [13]
\$4	28 minutes	MEDLINE	Brown [6]
\$4 - \$5	-	MEDLINE	McCarn [14]
\$7 - \$15	14 minutes	New York Times	Bachelder [15]
\$60	84 minutes	New York Times	Brown [6]
\$18 - \$26	15 - 23 minutes	ORBIT	Brown [6]
\$75	37 minutes	Various (at MIT)	Benenfeld [13]
\$32	24 minutes	Various (at Culham)	Hall [unpublished information]
\$12 - \$38	12 - 28 minutes	Various (at Exxon)	Lawrence [16]
\$42	30 minutes	Various (at NBS)	Ruhl [17]
\$24 (mean)	-	Various (NSF/SDC study)	Wanger [18]
\$17 (median)	-		

*Rounded to the nearest dollar or minute.

3. A generalized cost-computation approach

The economics of on-line retrieval can be very complex when considered from the point of view of a service operator or a regional network authority. A service operator may have to take many factors into account, often on an intuitive basis during the early months/years of service while building up a valid body of statistics on actual computer/data base use, staff costs, etc.; costing policy is then quite likely to be developed on a heuristic basis. A network authority may have to deal with many similar problems, but may additionally have other problems, e.g. trying to ensure that far-removed users are not unduly penalised in terms of telecommunications costs.

Fortunately, the individual library/user-unit is not normally directly faced with these problems since major service operators now operate on the basis of definite advertised charges per connect-hour with a small per-item charge for any items supplied as line-printer output. The charge per connect-hour varies according to the data base. The average connect-hour cost of the 63 data bases for which access costs are given in Hall's study [1] is \$70. The full list given by Hall does, however, contain a number of very expensive and somewhat specialised data bases. Exclusion of these leads to an average of \$60 per hour; this figure is used in the following calculations although it is appreciated that some user-units whose searches are concentrated on the many data bases rated at \$25 - \$50 may be in the fortunate position of operating at average connect-hour costs of \$30 - \$40 or even lower.

Given a connect-hour charge (either as an average or for a particular data base) a graph of connect-hours v. cost p.a. can readily be drawn and used for budget estimates. However, the connect-hour cost is not the only cost involved and there are advantages in using the cost-computation formula given by Elman [8]:

$$\text{Cost}_{\text{total per search}} = (T \times C_{\text{sum}}) + P$$

where

T = on-line time in minutes

C_{sum} = sum of all costs per minute of operation [i.e. connect-time, staff costs, telephone charges and terminal/coupler/printer cost]

P = cost of off-line printed citations.

A simple example (based largely on 1976 costs) illustrates the use of this approach. Consider a search carried out under the following arbitrary conditions:

- (i) Data base: rated at \$60 per connect-hour (= \$1.00 per minute)
- (ii) Staff costs: say \$10 per hour (= \$0.17 per minute). [This figure includes a notional component of \$3 - \$4 for overheads.]
- (iii) Telephone (to network node) and network charge: say \$12.50 per hour (= \$0.21 per minute). [Charges may be much lower or higher, of course. For example, network access to DIALOG by US users at January 1977 through TYMNET is \$8/hour and through TeleNet is \$5/hour; European access through TYMNET is c.\$20. Cost of dialled connection to the network node must of course be added to these figures.]
- (iv) Terminal/coupler equipment lease or amortization: say \$100 per month. (= \$0.01 per minute, but assessed at \$0.10 per minute thus properly accounting for non-use time if true utilization of terminal is 10%. Costs may, of course, be lower when an old or slow 10 cps terminal is used although connect-time cost may then be increased.)
- (v) Printout cost of an off-line citation: say \$0.10.

For a short 10 minute search with a few references printed at the terminal:

$$\begin{aligned} \text{Cost} &= (10 \times [\$1.00 + \$0.17 + \$0.21 + \$0.10]) \\ &= (10 \times \$1.48) \\ &= \$14.80 \end{aligned}$$

For a 20 minute search with an off-line printout of 50 references:

$$\begin{aligned} \text{Cost} &= (20 \times \$1.48) + (50 \times \$0.10) \\ &= \$29.60 + \$5.00 \\ &= \$34.60 \end{aligned}$$

For a longer 30 minute search with 100 printout references:

$$\begin{aligned}
\text{Cost} &= (30 \times \$1.48) + (100 \times \$0.10) \\
&= \$44.40 + \$10.00 \\
&= \$54.40
\end{aligned}$$

Elman's formula, while not perfect (e.g. it ignores pre-search preparation costs), has the great merit of simplicity and it can easily be used to calculate the approximate cost of a search in any local situation.

The "mix" of searches, in terms of time-per-search and data base(s) used, will naturally vary from unit to unit. To take one possibility, consider the quite arbitrary case where the mix of searches taken over a long period averages two short 10 minute searches, one 20 minute search with off-line printout, and one relatively long 30 minute search with off-line printout. The cost of the four components would then be \$14.80 + \$14.80 + \$34.60 + \$54.40 = \$118.60, which leads to an average search cost of \$29.65. The average search may involve several "sessions", i.e. approaches to one or more data bases; the search cost is taken here to be the sum of the various session costs.

It is interesting to note that while no statistical significance is placed on the "average" (\$28.30) of the search costs noted in Section 2, this figure is not far removed from the cost (\$29.65) derived above. It would be unwise, however, to place too much reliance on these two figures beyond noting that an average search probably falls in the range \$20 - \$40, with an overall figure of \$30 not being unreasonable.

4. The general cost-use relationship

With the growing acceptance and use of on-line retrieval as a routine search tool it is of increasing importance to try to derive a relationship, albeit approximate, between the total cost (\$) p.a. and the number of "average" searches p.a. Such a relationship expressed in graphical form could allow provisional budget estimates to be made quickly, and indeed could form the basis for more exact estimates when all relevant local

conditions have been taken into account. Many different factors will of course affect different user units, e.g. different traffic mix, possession of a fully amortized terminal, particularly heavy use of cheaper/dearer data bases, etc.

4.1 Establishing "lower limit" costs

It is possible to establish "lower limit" costs, i.e. the case where, for instance, (a) existing staff carry out searches on-line rather than manually so that for all practical purposes there is no extra staff cost component, (b) the terminal and associated equipment is supplied and maintained by the central computing service, and (c) telephone charges to the nearest network "node" are borne by the central telephone account (the network charges themselves may well be included in service suppliers' monthly invoices and so will be likely to be paid from library funds). These "lower limit" costs for "average" data bases at \$60 per hour (= \$1.00 per minute) with a network charge at \$8.00 per hour (= \$0.13 per minute) are:

For a 10 minute search with no off-line prints:

$$\begin{aligned}\text{Cost} &= (10 \times [\$1.00 + \$0.13]) \\ &= (10 \times \$1.13) \\ &= \$11.30\end{aligned}$$

For a 20 minute search with 50 off-line prints:

$$\begin{aligned}\text{Cost} &= (20 \times \$1.13) + (50 \times \$0.10) \\ &= \$22.60 + \$5.00 \\ &= \$27.60\end{aligned}$$

For a 30 minute search with 100 off-line prints:

$$\begin{aligned}\text{Cost} &= (30 \times \$1.13) + (100 \times \$0.10) \\ &= \$33.90 + \$10.00 \\ &= \$43.90\end{aligned}$$

The average "mix" cost is then $\$11.30 + \$11.30 + \$27.60 + \$43.90 = \$94.10$, leading to an average search charge of \$23.50 and from this is derived the

lower line ("Invoiced cost" line) in Figure 1. The shaded area below this line is intended to indicate that even lower invoiced costs are in fact possible where the traffic mix contains many short queries or much use of data bases with a connect-hour cost of less than \$60. While this cost-line is manifestly artificially low, it none-the-less probably represents the actual extra cost to the library in a number of organizations where the non-library costs are absorbed in other service budgets (just as users' general library costs are often absorbed in the library budget and no specific budget action need be taken by individual users or user-groups).

4.2 Establishing "upper limit" costs

Similarly, "upper limit" costs can also be easily estimated. Taking an "average" search cost from Section 3 of \$29.65 = c.\$30 the "Overall cost" line shown in Figure 1 can be derived. Again, there is a shaded area above the line to reflect unusual conditions. It will be noted that these "upper limit" figures are based on calculations in which the whole cost of the computer terminal is charged to on-line searching and the level of terminal use is taken as 10%, i.e. 90% idle time. In practice many different conditions may apply, e.g. the terminal may also be used for other library or non-library tasks, or alternatively library searches may be run during "spare" time available on a nearby shared terminal. At medium-high annual search loads on a library-owned terminal intensive use (exceeding 10%) makes little difference to the "overall" search cost line. At low traffic levels it is unlikely that a library would hold its own terminal unless this happened to be an older machine already largely or completely amortized; the "overall" search cost line shown at low traffic levels assumes that the library is paying only a modest charge either for its own terminal or for access to a shared terminal.

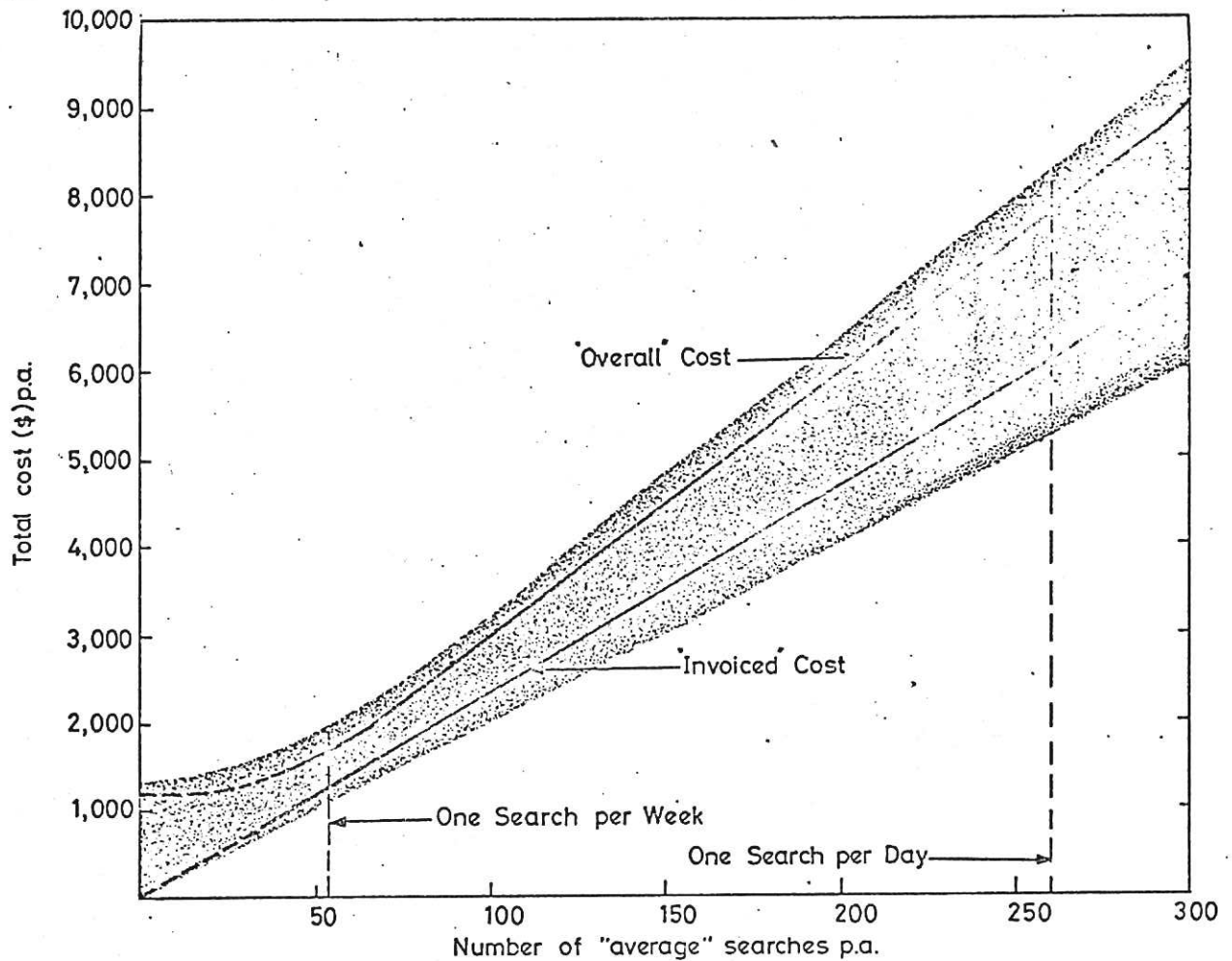


Figure 1. The cost-use relationship, showing how the total cost p.a. can vary with search traffic. Actual cost will depend on local conditions but is likely to fall within the shaded area. The lower line represents costs as they may be invoiced by the service operator(s). The upper line shows "overall" cost. The "average" search and cost estimates are based on the mix as defined in Section 3. Where the average search differs greatly from that defined in Section 3 it would be advisable to plot cost-use lines appropriate to the local traffic.

4.3 Application of the cost-use relationship graph

The library/information manager can easily see from Figure 1 what the

range of costs could be for any given level of annual on-line searching, and knowing the local conditions (e.g. with respect to terminal and telephone charges) can interpret where within the range the expected local cost might be expected to fall. The search traffic estimate must of course include a component for test and training searches not just "real user" searches. It is important to note that Figure 1 is intended for general guidance only - it does not claim to give definitive cost estimates and must be interpreted in the light of local circumstances. Despite this limitation, however, it is presented on the basis that for librarians new to the field of on-line retrieval "rule of thumb" guidance is better than no guidance at all.

5. Other factors

Many other factors deserve attention in any full study of the cost of on-line searching, e.g. the recent practice of allowing "user group" and "high use level" discounts [19]. It is particularly important that on-line search costs should not be considered in isolation but must be compared with the cost of a comparable or acceptable (to the user) manual or computer batch search, or indeed with the cost of not providing a search service at all (in which case the user may carry out a very expensive search). These aspects are not dealt with here but will be examined in a related text by Hall [20].

It must also be borne in mind that the cost of on-line searching has been falling rapidly - Burchinal [21] quotes Lockheed data indicating a drop from an index of 100 in 1970, to 40 in 1973, to 30 in 1974, and to 20 in 1975. Against this, of course, the recent steep rise in general costs may for the near future offset further decreases in actual computer searching costs leading to a levelling-off of overall search cost.

Among many recent comments on the cost(ing) of on-line searching, four deserve particular mention here. Wanger [18] presents an extremely

useful discussion on the cost of using on-line services. Hindle [22] notes that where there are no labour cost savings and on-line service in fact results in a real "extra" cost, then justification must rest on the ability to handle a larger number of requests and an improved search performance. Wilde [23] notes that to exclude staff costs is to ignore basic business concepts. Hall [20] argues that while a "reasonable" staff cost component should be taken into account, it should always be borne in mind that attempts to determine an "exact" cost for on-line searches are futile since few, if any, other library costs are known in exact terms, i.e. generally accepted as applicable to most/all libraries.

6. Conclusion

An examination of some recently-reported on-line search costs together with a generalized cost-computation approach suggests that the cost of an "average" on-line search falls in the range \$20 - \$40. A graph is derived, relating annual search traffic to approximate annual cost; this graph when modified to suit local conditions may be helpful to those with a responsibility for the funding of routine on-line search service.

7. References

- [1] Hall, J.L. "On-Line Retrieval 1965-1976: A Bibliography with a Guide to On-Line Data Bases and Systems", Aslib Bibliography No.8: London: Aslib: 125pp. (March 1977).
- [2] Buckley, J.S. "Planning for Effective Use of On-Line Systems", Journal of Chemical Information and Computer Sciences, -15 (No.3): 161-164 (August 1975).
- [3] Smetana, F.O., Furniss, M.A. and Potter, T.R. "A Study of the Relative Effectiveness and Cost of Computerized Information Retrieval in the Interactive Mode", NASA: CR 142206: report N75-18116 (1974).
- [4] Hoshovsky, A.G., "TRISNET - A Full Service Network", Bulletin of the American Society for Information Science, 2 (No.3): 28-29 (September/October 1975).
- [5] Ahlgren, A.E., "Providing On-Line Search Services Through the Public Library", (in) Proceedings of the American Society for Information Science, 38th Annual Meeting, 26-30 October 1975, Boston, Massachusetts, Vol.12 'Information Revolution', C.W. Husbands and R.L. Tighe (eds.): Washington, D.C.: ASIS: 156-157 (1975).
- [6] Brown, C.P., Yeates, E.J. and Van Hoesen, M.J., "Use and Cost of On-Line Systems at the National Bureau of Standards Library", (in) Proceedings of the American Society for Information Science, 38th Annual Meeting, 26-30 October 1975, Boston, Massachusetts, Vol.12 'Information Revolution', C.W. Husbands and R.L. Tighe (eds.): Washington, D.C.: ASIS: 132-133 (1975).
- [7] Caponio, J.F. and Moran, L., "CAIN: A Computerized Literature System

- for the Agricultural Sciences", *Journal of Chemical Information and Computer Sciences*, 15 (No.3): 158-161 (August 1975).
- [8] Elman, S.A., "Cost Comparison of Manual and On-Line Computerized Literature Searching", *Special Libraries*, 66 (No.1): 12-18 (January 1975).
- [9] Cooper, M.D. and DeWath, N.A., "The Cost of On-Line Bibliographic Searching", *Journal of Library Automation*, 9 (No.3): 195-209 (September 1976).
- [10] Hawkins, D.T., "Impact of On-Line Systems on a Literature Searching Service", *Special Libraries*, 67 (No.12): 559-567 (December 1976).
- [11] Williams, P.W. and Curtis, J.M., "The Use of On-Line Information Retrieval Services", *Program*, 11 (No.1): 1-9 (January 1977).
- [12] Page, J.R.U., "Interrelation of Technical and Policy Aspects of Scientific Technical Information Networking in Europe: An Overview of an OECD Information Policy Group Study", *Tidskrift for Dokumentation*, 31 (No.1): 8-18 (1975).
- [13] Benenfeld, A.R., Pensyl, M.E. and Marcus, R.S., "User Receptivity to Fee-for-Service Computer-Based Reference in a University Community", (in) *Proceedings of the American Society for Information Science, 38th Annual Meeting, 26-30 October 1975, Boston, Massachusetts, Vol.12 'Information Revolution'*, C.W. Husbands and R.L. Tighe (eds.): Washington, D.C.: ASIS: 151-152 (1975).
- [14] McCarn, D.B., "Trends in Information", (in) *Proceedings of the 37th Annual Meeting of the American Society for Information Science, Atlanta, Georgia, October 13-17, 1974. Vol.11 'Information Utilities'*,

- P. Zunde (ed.), Washington, D.C.: ASIS: 145-150 (1974).
- [15] Bachelder, S., "The New York Times Information Bank: A User's Perspective", (in) Proceedings of the 1975 Clinic on Library Applications of Data-Processing: The Use of Computers in Literature Searching and Related Reference Activities in Libraries. Lancaster, F.W. (ed.): Urbana-Champaign Ill.: University of Illinois: Graduate School of Library Science: 17-30 (1975).
- [16] Lawrence, B., Weil, B.H. and Graham, M.H., "Making On-Line Search Services Available in an Industrial Research Environment", Journal of the American Society for Information Science, 25 (No.6): 364-369 (November-December 1974).
- [17] Ruhl, M.J. and Yeates, E.J., "Introducing and Implementing On-Line Bibliographic Retrieval Services in a Scientific Research and Development Organization", Journal of Chemical Information and Computer Sciences, 16 (No.3): 147-151 (August 1976).
- [18] Wanger, J., Fishburn, M. and Cuadra, C.A., "On-Line Impact Study: Survey Report of On-Line Users, 1974-1975". Final Report on a National Science Foundation Study: System Development Corporation: 307pp. (1976). (Chapter 8, "Costs of Using On-Line Services", pp.146-164, is particularly valuable.)
- [19] Lockheed's DIALOG discount scheme is detailed in the monthly newsletter Chronolog, 5 (No.2): 2 (February 1977). The System Development Corporation (ORBIT) and Bibliographic Retrieval Services (BRS) are also offering discount schemes.
- [20] Hall, J.L., "On-Line Information Retrieval Sourcebook". (In preparation, 1977.)

- [21] Burchinal, L.G., "Bringing the American Revolution On-Line: Information Science and National R & D", Bulletin of the American Society for Information Science, 2 (No.8): 27-28 (March 1976).
- [22] Hindle, A. and Raper, D., "The Economics of Information", (in) Annual Review of Information Science and Technology (Vol.11), M.E. Williams (ed.): Washington, D.C.: ASIS: 27-54 (1976).
- [23] Wilde, D.U., "Generation and Use of Machine-Readable Data Bases", (in) Annual Review of Information Science and Technology (Vol.11), M.E. Williams (ed.): Washington, D.C.: ASIS: 267-298 (1976).

LEGEND

Figure 1. The cost-use relationship, showing how the total cost p.a. can vary with search traffic. Actual cost will depend on local conditions but is likely to fall within the shaded area. The lower line represents costs as they may be invoiced by the service operator(s). The upper line shows "overall" cost. The "average" search and cost estimates are based on the mix as defined in Section 3. Where the average search differs greatly from that defined in Section 3 it would be advisable to plot cost-use lines appropriate to the local traffic.

