NEW TYPE OF FILE ACCESS METHODS FOR (DISTRIBUTED) DATA BASE IMPLEMENTATION. CONSTRUCTION METHO-DOLOGY, APPLICATION TO HASH FILES.

(- + - >

W. LITWIN, IRIA



IST 663

MAP. T.003 (663

Preliminary project of a communication for the GI-Conference on Databases in Computer Networks including Small Computers, to be hold at April 11-12, 1978 in NRC of KARLSRUHE.

<u>Title</u>: New type of file access methods for (Distributed) Data Base implementation. Construction methodology, application to hash files.

Area : System Architecture or Design and Implementation.



<u>Abstract</u> : Data Base and especially Distributed Data Base design would be easier and fonctionning improved, if one could dispose of files which would hold their performances independly of their content variation. This statement is detailed below and the theory of the access methods to such files, named <u>self</u>-<u>structuring</u> files, is proposed as well as an example of such access method. A new type of hash access methods is derived from theory concepts, named <u>Virtual Hashing Access Methods (VHAM)</u> and offering to hash files the performances inenvisageable for the hash access methods existing until now.

<u>Author</u>: W. LITWIN, IRIA - SIRIUS Project, BP. 105 78150 ROCQUENCOURT, FRANCE.

Introduction.

Data Base (DB) and especially Distributed Base (DDB) and especially Distributed Base (DDB) and especially Distributed Base (DDB) and Exploit a base (DDB) and exploit a base of the exploration of the e

1

ADEL

1 6 SEP. 1379

C663

1. <u>Hash files properties analysis</u>.

The wide use results from attractive properties of this type of files, essentially due to the fact that hash function parameters can be entirely resident. Thus, there is no need of disc access in order to compute the record primary address. That's why many hash access methods have been proposed or analysed in details /GUI73/, /KNU74/, /LUM73/ and so on.

From these studies, one can see a property common to all the methods : access performances, mesured in number of file access per retrieved record, or inserted one etc, theoretically exceed the ideal performances for every load factor, provided, of course, that number of records exceed page capacity. The deterioration rate increase with load factor and this property is generally considered as principal and unavoidable handicap of hash access methods, especially cumbersome for disc files.

The deterioration is due to collisions overflowing their primary pages, which thus need more then one file access to be retrieved. To give the idea of the magnitude of the deterioration rate, one may choose the case of a file with ten records by page capacity, this value beeing quite representative of usual disc files. It can be seen then,/KNU74/, that when load factor reach 95 %, the separate chaining collisions resolution method, as models show-one of the best, has 30 % deterioration rate for successful search and 100 % for unsuccessful one. The linear probing, also widely used, is much worse, reaching 80 % for successful search and growing to 300 % when page capacity raise down to 5. records.