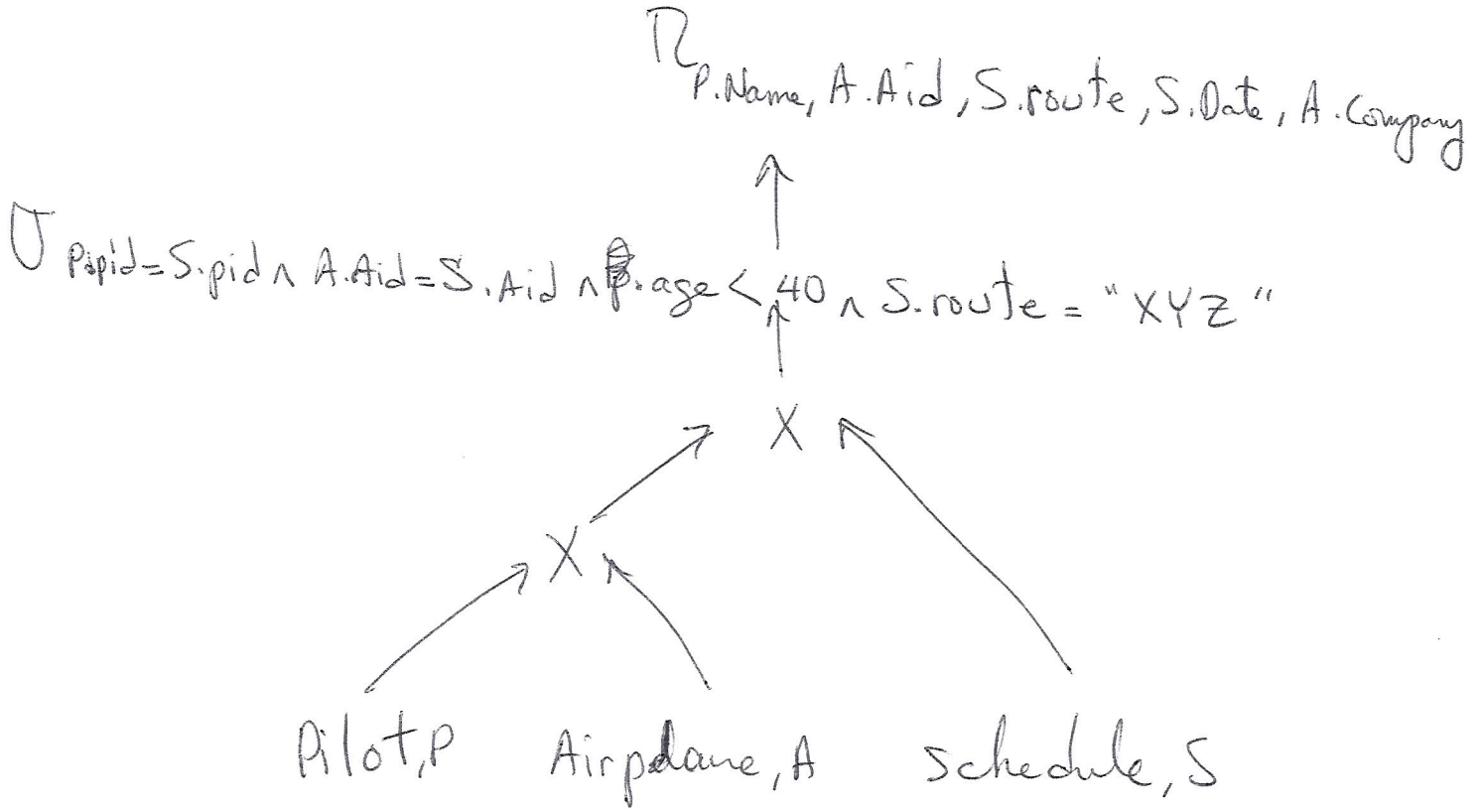
a- Basic query tree



b- The heuristic will have these objectives:

i- push selections and projections down the tree as far as possible so as to reduce the size of intermediate relations

ii- Turn cross products into join operations if possible.

iii- To join two relations that at 2 different sites, send smaller relation to site where larger relation sits.

b - cont'd.

Transformation rules needed are:

$$R \bowtie_C S \Leftrightarrow \sigma_C (R \times S)$$

~~$R \times S \Leftrightarrow S \times R$~~

$$(R \times S) \times T \Leftrightarrow R \times (S \times T)$$

$$\sigma_{p_1(A_1)} (\sigma_{p_2(A_2)} (R)) \Leftrightarrow \sigma_{p_1(A_1) \wedge p_2(A_2)} (R)$$

$$\sigma_{p(A)} (R \times S) \Leftrightarrow (\sigma_{p(A)} (R)) \times S$$

$$\pi_C (R \times S) \Leftrightarrow \pi_A (R) \times \pi_B (S)$$

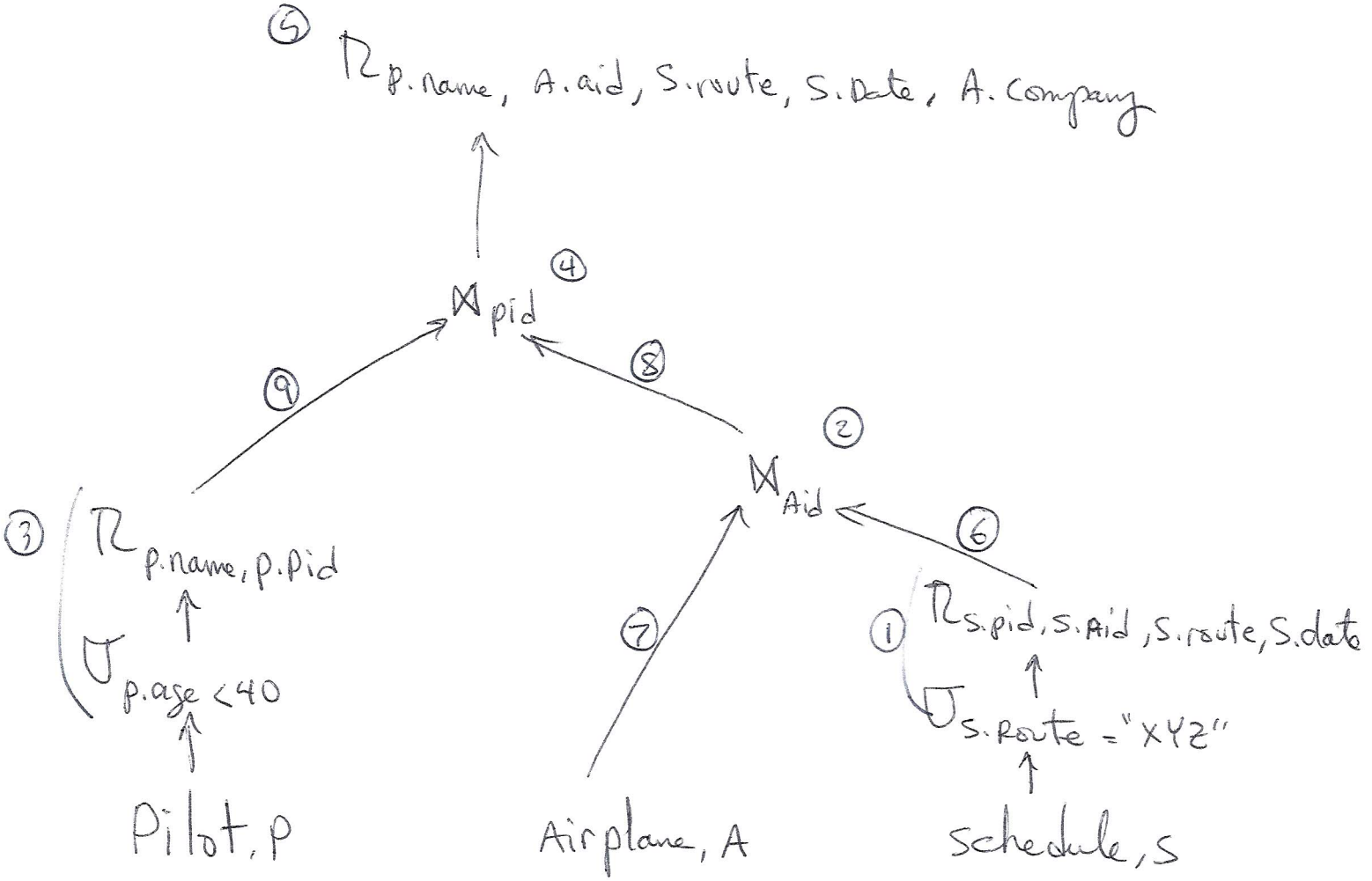
where $C = A \cup B$

C -

Airplane record length = 10 + 60 = 70

Pilot " " = 10 + 60 + 4 + 2 = 76

Sched " " = 10 + 10 + 8 + 8 + 20 + 8 = 64



Costs

① Assumption :- Selection & Projection done in one operation (one pass over the sched rel.)

~~Cost of U+R~~ Have an index structure on Route

- Cost of moving thru index structure is ignored. Only I/O cost of fetching the tuple is calculated

I/O Cost of U+R = $1/50 * 10{,}000 * 64 * 1 = \boxed{12{,}800}$ ①

Size of records after projection =

$10 + 10 + 20 + 8 = 48$

Size of intermediate sched rel. relation after U ~~&~~ +R =

$48 * 1/50 * 10{,}000 = \boxed{9{,}600}$

Size of Airplane rel. = $500 * 70 = \boxed{35{,}000}$

$35{,}000 > 9{,}600$

⇒ It is cheaper to ~~set~~ send sched's over to site 1 to do the join between Airplanes & sched's.

⑥ Cost of sending sched's ~~between~~ = ⑥

$9{,}600 * 10 = \boxed{96{,}000}$

② Cost of join at site 1

Assumption: index structure exists on Airplane relation for field AId.

~~Cost of join~~ = Cost of going thru index structure
* is ignored.

No. of records in sched' = 1/50 * 10000 = 200

200 * (48 + 70) * 1 = 23,600 ②

48: size of sched' rec's
70: size of Airplane rec's

no. of records after join = 200

size of each record after join = 48 + 60 = 108

48: size of sched' rec's
60: size of Company feed

size of relation after join at step ② = 108 * 200 = 21,600

③ Cost of selection & projection on Pilot.

Assumptions: selection & projection are done in one step.

Age range = 65 - 25 = 40

Age range of interest = 40 - 25 = 15

ratio of pilots in this range = $\frac{15}{40} * 100\% = 37.5\%$

Since many pilots (>>5%) are in the range will use scanning to read all pilots (no indexing)

Cost = 1000 * 76 * $1 = 76,000 ③

↑ size of pilots rec.

Size of pilots' rec's after this op = 10 + 60 = 70

no. of tuples in pilots' = 1000 * 37.5% = 375

Size of the whole pilots' rel = 70 * 375 = 26,250

26,250 > 21,600

So it is more efficient to send the results of join from step ② to site 2 to perform 2nd join.

(8) Cost of transfer of results of ~~second~~ 1st join (from step (2))

$= 21{,}600 * 10 = \boxed{216{,}000}$ (8)

(4) Cost of join at step (4)

Assumption: join ~~&~~ at ~~sel~~ step 4 & projection at step 5 are done in one step.

we have an index structure on pilot for field PId.

Cost of join $= \$200 * (108 + 70) * 1 = \boxed{35{,}600}$ (4)

108 ↑ size of sched' recs

70 ↑ size of pilots' recs

Size of each record after join & projection =

$60 + 10 + 20 + 8 + 60 = 158$

Size of final results $= 158 * 200 = \boxed{31{,}600}$

$31{,}600 > 26{,}250$

It would have been cheaper to send both pilots' & sched' to site 4 and do the join at site (4)!!

Summary of Costs

~~calc.~~ evaluate sched' at site 3 :	12,800	I/O
Send sched' to site 1 :	96,000	Comm.
evaluate pilots' at site 2 :	76,000	I/O
~~Send sched results~~ perform 1st join at site 1 :	23,600	I/O
Send results of 1st join to site 4 :	216,000	Comm.
Send pilots' to site 4 :	262,500	Comm.
cost of join & π at step ④ :	35,600	I/O
Total cost	722,500	