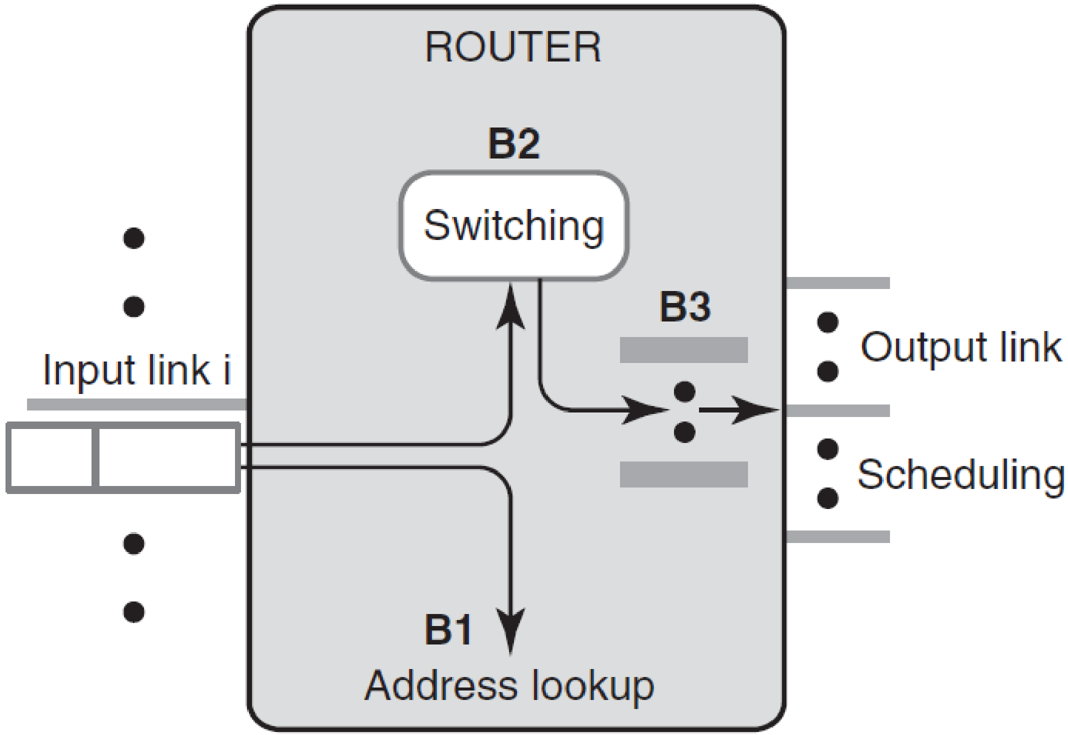
Switches

Exact Match

A model of a switch



Exact-Match Lookups

Used in network nodes when address space is flat

- Ethernet
- MPLS

Simplest form of database query

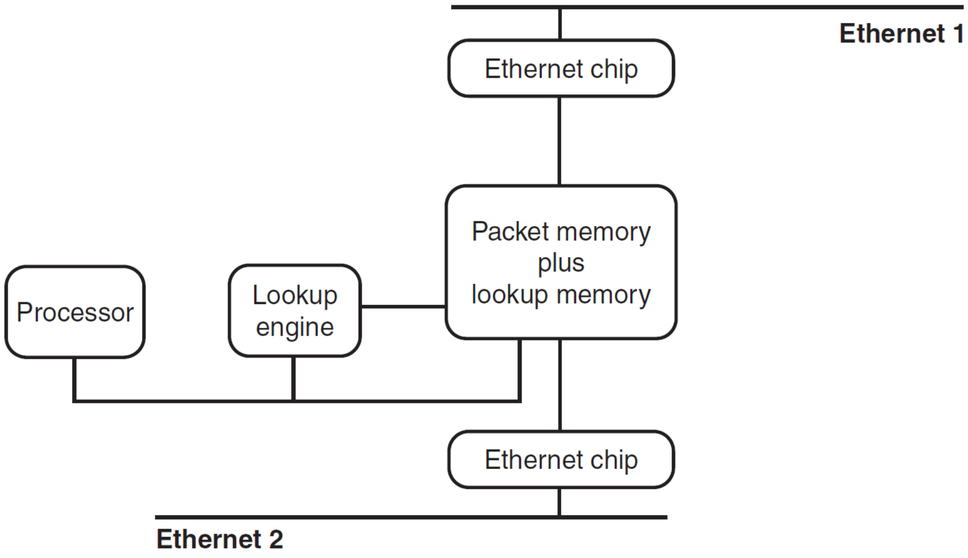
- Query specifies key K.
- K is fixed-length.
- Retrieve state associate with K

Lots of well-studied techniques, but problem complicated by wire-speed lookup requirement

Ethernet Bridge

An Ethernet bridge is a filtering repeater with learning.

- By looking at source addresses, the bridge fills a mapping table
- By looking at the destination address and the table, the bridge decides to forward or to drop the frame



Ethernet Bridge Lookup

A new 64-byte frame arrives every: 64*8/ (10 Mbit/s) = 51.2 μs

To avoid queues the bridge must perform 2 lookups (the source and the destination) per port every 51.2 μs.

With two ports we need a lookup every 12.8 μs.

The first DEC bridge used cheap DRAM with access time of 100 ns.

The table should store 8000 entries.

Solution:

- Binary search gives deterministic lookup times
- Implementing the search in hardware moves the bottleneck to the memory access
- Binary search in a table of 8000 entries requires log2(8000) = 13 accesses
- Final lookup time is 1.3 μs

FDDI Lookup

FDDI (Fiber Distributed Data Interface)

- Obsolete, we do no study the details of the protocol
- Part of the IEEE 802 family, uses the same MAC addresses as Ethernet
- Speed is 100 Mbit/s
- Minimum frames are 40 byte (so a frame arrives every 3.2 μs)

The binary search approach does not scale

DEC Gigaswitch (~1994)

To reduce memory accesses Gigaswitch uses hash tables

- Average → constant time
- Worst case → linear time

The trick is using a variant of perfect hashing

- Perfect hashing guarantees constant time because there are no collisions, but requires that the keys are known in advance
- Gigaswitch uses a parametrized hash function. If the number of collisions becomes too large, the hash parameters are changed and the table is rehashed. This is infrequent, because the table changes slowly over time.

Gigaswitch hash table

The 48-bit MAC address is represented as a polynomial with binary coefficients A(x)

The hash function is H(x) = A(x) M(x) mod G(x)

- G(x) is a 48-degree irreducible polynomial (fixed)
- M(x) is a 47-degree polynomial (the function parameter)

The result H(x) is a 48-bit string

- Property: if A(x), A'(x) are different, their hashes are different
- The bottom 16 bits are used as an index in a 64K entry hash table
- In case of collision, the remaining 32 bit are used to disambiguate using a balanced binary tree of maximum depth 3

Any search requires at most 4 memory accesses.

In case there are more than 7 collisions in a bin, the overflowing keys are stored in a CAM, looked-up in parallel.

Initially the hash parameter M(x) is chosen randomly.

In the (very rare) case the CAM stores too many addresses, the processor chooses a new M(x) and rehashes