



Implementing Address Resolution Using M70X0 Network Search Engine Technology in Multi-Gigabit IP Network Interfaces

INTRODUCTION

Address Resolution Protocol (ARP)

ARP is the standard protocol in IP networks for the conversion of Media Access Control (MAC) addresses into Internet Protocol (IP) addresses and back again when forwarding IP packets to the next hop of their route between two IP hosts. The principles of operation of ARP are defined in RFC826¹. In the layer 2 switching algorithm, when a switching device receives a packet on an interface, the IP address of the destination host is extracted from the packet header. This is passed to the Address Resolution module which looks the IP address up in the ARP table and if found returns the MAC address to which the packet should be forwarded on the next hop to the final destination. If the address is not found then the packet is either discarded, or forwarded to a default port. The network interface then assembles an ARP packet to request the required address and broadcasts this to all the segments on the device. When a reply is received to the ARP request, the information in the packet is used to update the ARP table. All network interfaces have a finite limit to the size of the ARP table, so a LRU algorithm is used to discard / replace items that are not encountered frequently. A timeout also needs to be implemented for entries that are not used for a period to prevent incorrect information persisting in the ARP table if a host is moved or changes IP address.

Solutions

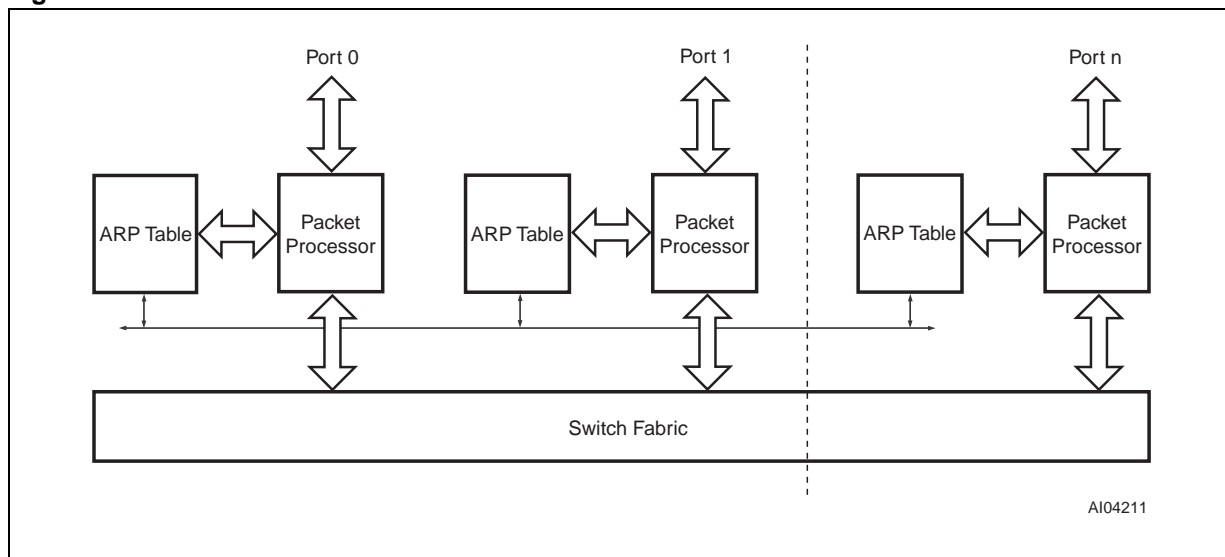
It is clear when analyzing the process required to forward the packet, that the ARP table will need to be queried each time a packet is received by the network interface host. If the network interface is attached to a host system then the ARP table will have only a few records as each host will only normally communicate with a small number of other hosts. However if the network interface is a port on a switch or router within the cloud of the internet the ARP table could contain many thousands of entries. As the number of hosts increases, with the growth of the internet, backbone devices will need to be able to support table sizes of up to a million entries to avoid packet loss and increased latency. A networking device transporting 64 byte packets using Gigabit Ethernet implementing IPv4, will have 512ns available for address resolution if the packets are to be forwarded at wire speed. Microprocessor based solutions for ARP lookup are possible for network devices with a small number of ports, however for the multi-port devices required in the internet backbone, alternative techniques are needed to maintain wire-speed forwarding.

1. IEEE Internet Working Group RFC 826 see www.rfc-editor.org

AN1342 - APPLICATION NOTE

Parallel Tables. Each port within our networking device acts independently of each other over the task of address resolution. Each port maintains its own ARP table and processor so that resolution can be performed at the full data rate. The table held by each port will be a copy of all the hosts known to the networking devices. Figure 1 shows the architecture of such a network device.

Figure 1. Network Device Architecture for Parallel Tables



Advantages

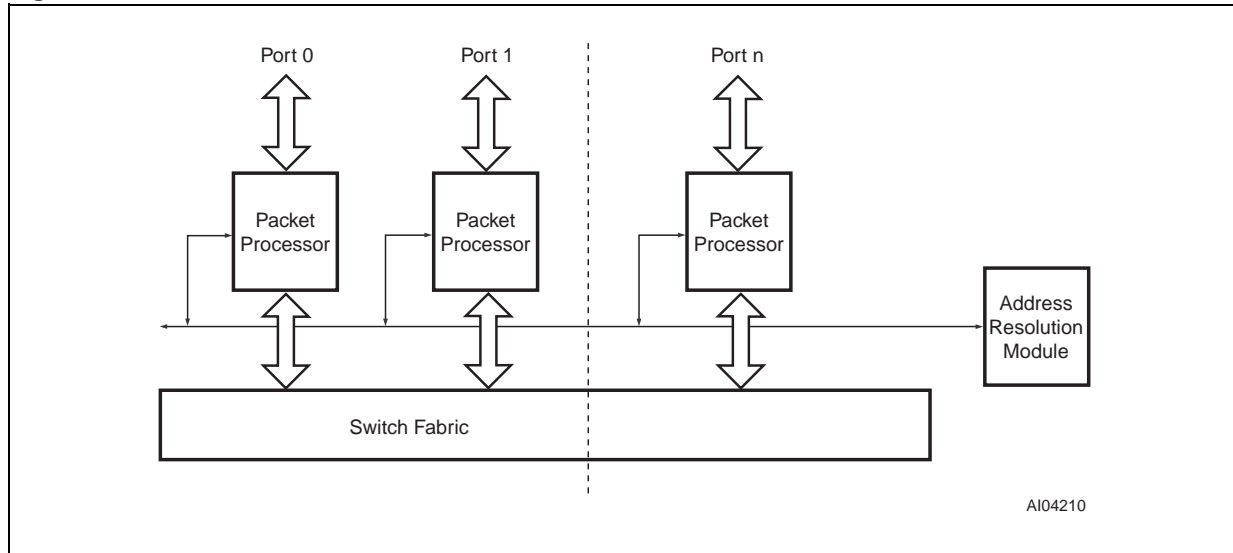
- As each port is independent, high volume traffic on one port will not affect any other port.
- The number of ports implemented in the device is only limited by the capabilities of the switching fabric.

Disadvantages

- Relatively high cost per port as each port needs a processing element and a storage element to hold the table.
- A mechanism needs to be implemented to propagate changes and new table entries through the device.
- As the table size grows, all ports need to be upgraded to support larger tables.

Centralized Address Resolution Module. The network device has a centralized Address Resolution Module which hold the ARP table for all ports in the device. Figure 2 shows the architecture of such a network device. The centralized ARM is often implemented using a Network Processor optimized for such tasks. This improves the number and speed of the ports that can be supported.

Figure 2. Network Device Architecture for Centralized Address Resolution Modules



Advantages

- Simplified table management as only one table is maintained.
- Efficient use of resources as there is no duplication.
- Lower cost per port.
- Upgrade is practical as host numbers increase as only the central module needs to be expanded.

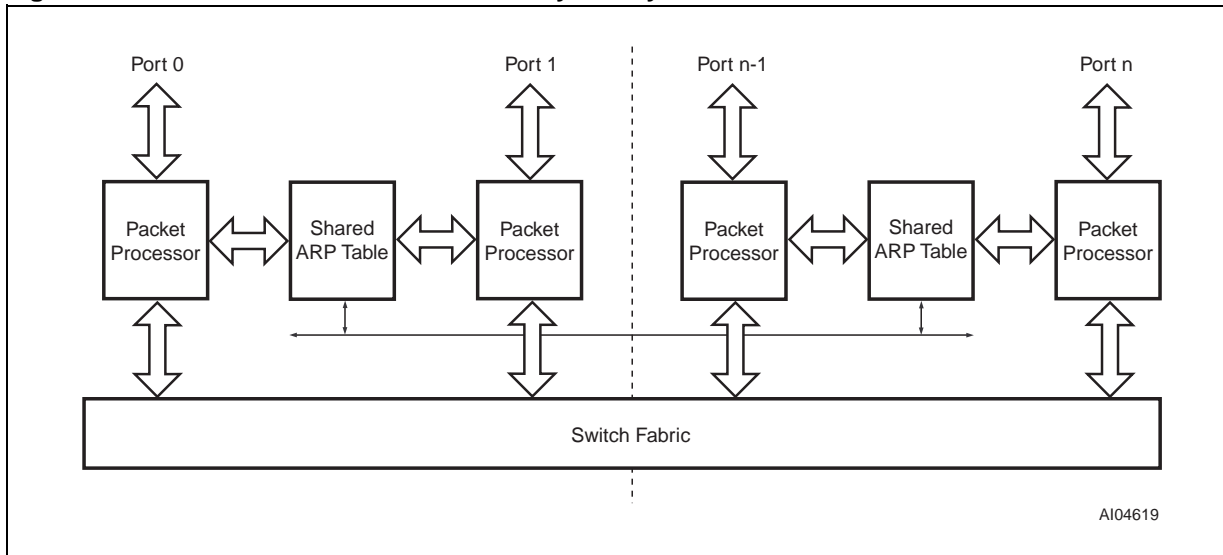
Disadvantages

- Performance is limited by the capabilities of the central module.
- Only a limited number of ports can be supported by the central module.

AN1342 - APPLICATION NOTE

Hybrid Approach. To reduce the cost per port of the parallel processing architecture, but support a larger number of ports than the centralized module architecture, some systems employ a hybrid where groups of ports share a common ARP table. Such an architecture is shown in Figure 3.

Figure 3. Network Device Architecture for Hybrid Systems



Advantages

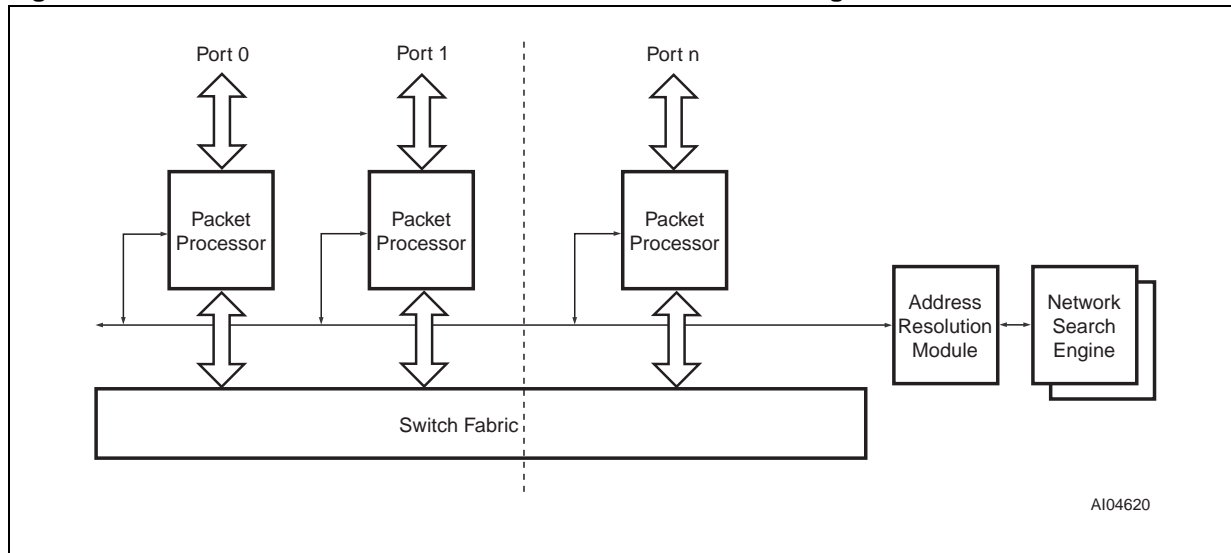
- Reduced cost per port compared to parallel processing architecture.
- Expansion capabilities only limited by switching fabric.

Disadvantages

- Increased complexity of overall solution.
- Inter table synchronization required.
- Upgrade of table size requires multiple modules to be upgraded.

Network Search Engine. The introduction of Network Search Engines and Co-processors enhances the capabilities of the centralized ARM architecture eliminating the need to adopt the complex hybrid approach. Using a Network Search Engine for address resolution raises the bar for network devices as performance is limited only by the capabilities of the switch fabric backbone. The architecture of such a solution is shown in figure 4.

Figure 4. Network Device Architecture for the Network Search Engine



Advantages

- Reduced cost per port.
- Expansion capabilities only limited by switching fabric. ARM performance up to 83 million searches per second, supporting backbones up to 20Gbit for 64 byte IP packets.
- Simplified table management as only one table is maintained.
- Efficient use of resources as there is no duplication.
- Upgrade is practical as host numbers increase as only the central module needs to be expanded. Additional Search Engines can be added to give capacities of up to 1 million entries.
- As table sizes exceed 64K records, Search Engines provide a component cost advantage over similar SSRAM based solutions.

Disadvantages

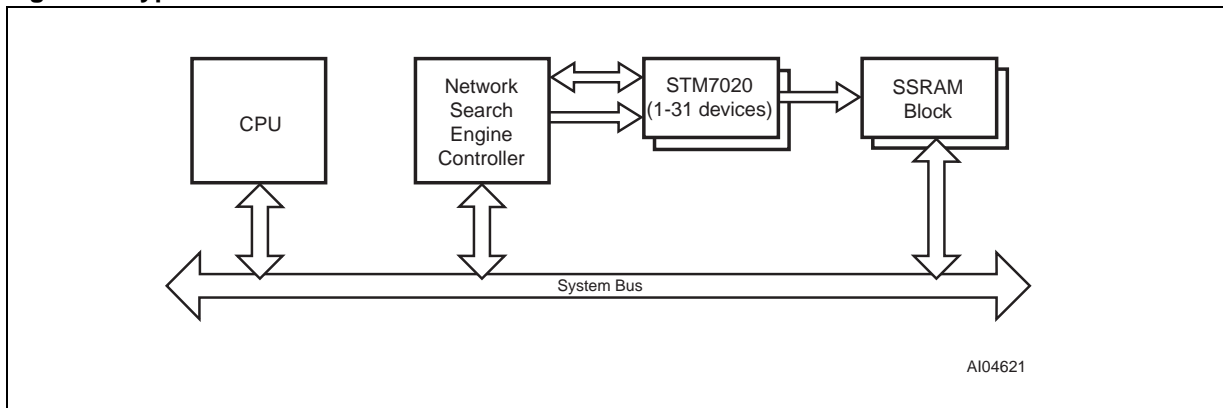
- New techniques need to be learned to implement search engine based devices.

Implementing ARP using Network Search Engine Technology

The Address Resolution Module could be implemented using a Network Search Engine such as the M7020 from STMicroelectronics, Inc. The M7020 part is a 16K x 136bit Search Engine and can be cascaded to up to 31 devices, supporting tables of up to 992K entries for this application. The M7020 can be configured to support tables of 34, 68, 136 or 272 bits wide and can support multiple tables of differing sizes, however in this application a single table is appropriate.

Architecture. A typical Address Resolution module architecture is shown in Figure 5:

Figure 5. Typical Address Resolution Architecture



The Network Search Engine controller could be either a custom ASIC designed specifically for this application or it could be the Search Engine Controller LNI8010 from Lara Networks, Inc. A high performance RISC CPU or Network Processor would be used with a system bus clock of up to 200MHz to achieve maximum performance.

The Network Search Engine will be used to implement a single table containing entries for both 32 bit IP addresses and 48 bit MAC addresses. The corresponding IP address or MAC address result will be stored in the correct location of the SSRAM block. One bit of the record will be used to indicate the address type. Thus we will need to implement records that have 49 bits or 33 bits of data. Our table will be constructed of 68 bit wide records in the search engine as this the nearest configuration we have to our required size. The unused bits may be used in a router or switch design to indicate the port on which the address is located. A global mask register will be loaded with the value 0x0 0001 FFFF FFFF FFFF to be used to mask the unused bits during our search operations. The IP address records will have the bits 32-47 of zero so that the same mask register can be used for both search operations. The structure of the records will be as shown below:

Table 1. Records Structure

Search Engine				SSRAM		
68	47			68	47	0
Spare	0	48-bit MAC Address		Spare	0x0000	32-bit IPv4.0 Address
Spare	1	0x0000	32-bit IPv4.0 Address	Spare	48-bit MAC Address	

To configure the Search Engine to a single 68 bit wide table, we need to set the CFG bits of the Command Register to the value 00000000. For full details see the M7020 Network Search Engine Data Sheet.

As we are only looking here for "exact match" results, the mask array will not be used in this application, and the data array and SSRAM will be programmed with the known IP - MAC address pairs. Each address pair will have two entries in our table so that we are able to perform search operations with either a known MAC address or a known IP address.

Search Performance. The maximum performance that will be achieved by the Address Resolution Module will be when the Network Search Engine is operated with a full pipeline. Given that the system bus is 64 bits wide, one cycle is required to load the compare word into the search engine. After the search latency has passed, the match address will be presented on the SSRAM address bus interface and after the access delay of the SSRAM device, the result word will be read from the SSRAM data bus. If our host CPU operates with a system bus of 200MHz, then this can be interleaved into the search words that are presented to the Search Engine.

The performance that can be achieved for a 64K record table made up from 8 of 83MHz M7010 devices (or 4 of M7020 devices) is summarized below:

Table 2. Performance Summary

Operation	Non-pipelined	Pipelined
Write Search Word	12ns	12ns
Latency	60ns	n/a
Read Result Word	5ns	5ns
Total Response Time	77ns	17ns

AN1342 - APPLICATION NOTE

The CPU will also need to receive and interpret the address received from the packet processors which will add to the response time. While the traffic rate is low (e.g., below 5M packets / sec) then most operations will be resolved using the search engine non-pipelined. The performance enhancement of the pipeline comes into action when it is most important at high traffic densities.

The following table details the number of ports and the data rates that may be supported by this Address Resolution Module in the centralized architecture:

Table 3. Number of Ports, Data Rates for M7010 Address Resolution Module

Average Packet Size	#1 Gbps Ports	#2.4 Gbps (OC-48) Ports	#9.6 Gbps (OC-192) Ports
64 bytes (e.g., VoIP)	25	10	2.5
128 bytes	50	20	5
512 bytes (e.g., www)	200	80	20
1K bytes (e.g., FTP)	400	160	40

The results in this table demonstrate that when using Network Search Engine technology, it is possible to provide centralized Address Resolution for high performance, multi-network devices, even up to OC-192 data rates.

Table Management. As in this application the search is always for an exact match, the table management process is simplified as there is no requirement to sort the table in any particular order. The requirement is to implement the following processes:

Add Entry

This process will be requested after an ARP packet has been received by an interface indicating the identification of a new host on a segment. Two entries will need to be added to the data table for search by either IP or MAC address. Entries are added to the Search Engine by executing a LEARN command. This command writes the data provided to the data array of the search engine to the next available free location, and then asserts a write cycle on the SSRAM which we use to write the SSRAM portion of the data record. The process is summarized in the following steps:

- Learn command on the Search Engine to write the search word to the next free location.
- Wait for the latency of the Learn operation.
- Write cycle on the SSRAM to store the related data.

Delete Specific Entry

The search engine maintains an array of registers that hold the address of successful search operations. This is used to identify an entry that we wish to delete. Bit 0 of the data array has is used in the Search Engine to indicate if the entry is used or free. To delete an entry we need to write a 0 value to this bit in the data array for the record we wish to delete. The sequence of step to delete an entry are as follows:

- Search command to find the required record.
- Wait for the latency of the search operation.
- Perform a write command of 0 into bit 0 using the Successful Search Register returned by the search operation to provide the write address.

In this application, the delete operation will have to be carried out twice for the MAC and the IP address.

Entry Ageing

An ageing algorithm for entries that are not used for a period of time can be implemented by using the spare bits in the data array. The 16 spare bits could be used to time-stamp the use of each record after each search with a counter value that is incremented by the host processor every second. Every second, the management process would then perform a search on these bit to find records that were last stamped 30 seconds ago (if the timeout is 30 seconds) and then delete these records. This process would have to be repeated until no further successful matches are found.

CONCLUSION

Address Resolution is a key issue in the implementation of Network Interface Devices, and as data rates increase, the performance of lookup in the ARP table becomes a critical issue, particularly with multi-port devices such as switches and routers. A number of architectures have evolved to address this problem, the simplest being the Centralized Address Resolution Module architecture. To date this architecture has been unable to keep up with the increasing performance requirement and so Hybrid architectures have been evolved. The introduction of Network Search Engines allows us to implement multi-port devices at data rates of up to OC-192 and beyond using the simpler centralized architecture, thus avoiding the complexities of distributed and hybrid architectures. This will result in reduced engineering effort and quicker time-to-market for those willing to deploy this technology. As the number of hosts addressed by network devices in the Internet cloud increases, devices based on Network Search Engine technology will be able to expand to meet the demand with no impact on performance. With current generation parts, table sizes of up to 992K entries can be created supporting up to 496K hosts.

CONTACT INFORMATION

If you have any questions or suggestions concerning the matters raised in this document, please send them to the following electronic mail addresses:

<i>apps.nvram@st.com</i>	(for application support)
<i>ask.memory@st.com</i>	(for general inquiries)

Please remember to include your name, company, location, telephone number, and fax number.

Information furnished is believed to be accurate and reliable. However, STMicroelectronics assumes no responsibility for the consequences of use of such information nor for any infringement of patents or other rights of third parties which may result from its use. No license is granted by implication or otherwise under any patent or patent rights of STMicroelectronics. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all information previously supplied. STMicroelectronics products are not authorized for use as critical components in life support devices or systems without express written approval of STMicroelectronics.

The ST logo is registered trademark of STMicroelectronics
All other names are the property of their respective owners.

© 2001 STMicroelectronics - All Rights Reserved

STMicroelectronics GROUP OF COMPANIES
Australia - Brazil - China - Finland - France - Germany - Hong Kong - India - Italy - Japan - Malaysia - Malta - Morocco -
Singapore - Spain - Sweden - Switzerland - United Kingdom - U.S.A.

www.st.com



LittleDiode supplies new, hard to find or obsolete electronic components and semiconductors all over the world.

With over two million different components listed you are sure to find the part you need.

Feel free to visit us today at our online store:

LittleDiode.com

Looking forward to providing you with the best possible service.