

Network Diagnostics

What tools are available?

Tue, Jan 27, 1998

The local station/IRM system provides a number of network diagnostic tools. This note describes many of the choices available.

Network Frames Page

The system maintains a record of network frames received or transmitted inside each node. Each record includes the time-of-day, the network node, the size of the frame, and a pointer to the frame contents as it resides in the network frame buffer. The structure that holds these records is a data stream, with room for about 237 records before it wraps. The network buffers are 128K bytes long, in the case of token ring, and 85 full-size frames long, in the case of ethernet. So, if the contents of the data stream are sampled before 237 records have been written into the data stream, and the network buffers have not wrapped, one can actually view the contents of the frame itself as it lies in the frame buffer. The NETF Page Application is normally installed on Page F. Even when the contents of the frame are lost, this diagnostic tool has been proven itself useful countless times.

Network Frame Monitor Local Application

This local application program is called FMON. It captures 56 bytes of frames received and/or transmitted from a specified node number. For use with IP, one must determine the pseudo node number for this purpose before making the test. Note that if an hour passes without any frame activity regarding the selected node, it's ARP table will be released, so that any future communications with that node may cause a different ARP table entry to be put into use, meaning that a different pseudo node number would apply. This is a weakness of the implementation. But when it can be used, it allows selection of a portion of the frame to be sampled, so that one might skip the frame header, or one might in addition choose to skip the IP or IP/UDP headers, thus permitting more of the actual message to be captured for analysis. One can turn on a Bit to enable capture, turning the Bit off to disable further capture, in order to permit more careful analysis of the results without fear of the results being overwritten. Besides the frame portion, the record includes the time-of-day.

Broadcast diagnostics

An area of memory is used for holding special ethernet broadcast/multicast-related diagnostics. This area is from 00005000–000053FF. An 8-byte header is followed by up to 127 eight-byte entries. The header is as follows:

BCKEY (2) BCCNT (2) BCTOTAL (4)

The BCKEY word is used to select which optional diagnostic is desired. There are four options, each of which is specified by a 2-character ascii key code.

BC	Six-byte physical addresses that broadcast, <i>not</i> including ARP requests.
AR	IP addresses sending ARP request frames.
11	Target IP addresses in ARP requests received from 131.225.1.1
MC	Multicast 6-byte physical addresses invalid for this node

For each mutually-exclusive option, counts of occurrences are kept. The data to be matched is in the first 6 bytes of each entry, with the count in the last two bytes. By using this tool, one can view how much broadcast activity exists on the network and how much of it consists of ARP requests. Multicast addresses that hash to the same 6-bit code (and therefore accepted by the ethernet controller chip) can also be tabled.

Each option is selected by manually writing the appropriate two-character code in the first word of the table. By default, the table is all zeros, so no diagnostic is enabled after system reset. The BCCNT word is the count of entries that have been filled and are therefore eligible for searching to match against the next 6-byte key. If no match is found, then the table is extended by one entry, and this word is incremented. After a count of 127 entries has been reached, no more entries can be added, but counts are still made for entries whose key matches incoming frames. The BCTOTAL longword is the total count of all entries. It should be the sum of all received frames that were placed into the table, or whose matching entry was counted.

Ethernet interrupt times

When an interrupt occurs, the elapsed time of processing is measured in microseconds. The status word is also recorded, so that one can identify which entry is a received frame and which is a transferred frame. The table containing this data is a circular buffer that provides room for 128 eight-byte entries before it wraps. Each entry consists of four words:

STATUS (2) CYCNT (2) MSEC (2) MICROS (2)

The status word is taken from the System Control Block at 00160000. It should have the value of 2040 or 4040. The value 2040 represents a transmit interrupt, and the value 4040 represents a receive interrupt. (The 4 in the third nibble is the Receive Unit status, which should always be 4, signifying Ready.) The cycle counter is actually the low word of the cycle counter global longword. The millisecond word specifies the time within the specified cycle (in half-milliseconds) when the interrupt occurred. These two words allow knowing whether one entry occurred at nearly the same time as the next entry. The last word gives the elapsed time in microseconds of processing the interrupt.

Ethernet Receive Unit diagnostics

When the ethernet controller chip encounters a frame that cannot fit into the buffer space given it by the driver, the Receive Unit enters a No Resources state that prevents any further frame reception. The software must notice this occurrence and restore the Receive Unit to the Ready state. After several years, this effect was noticed, when a frame appeared on the network that seemed to be 2 bytes longer than the ethernet standard frame size limit of 1500 bytes (Beyond the 14-byte frame header). It was 2 bytes longer because two extra bytes were inserted before the frame header, and the value of those two bytes happened to make the first six bytes appear to be a multicast address that by odd chance happened to pass the chip's multicast hash filter. Nonetheless, even as unusual as this occurrence is, the software must protect against it. The software was modified to detect the condition and recover from it. But the software also captures some part of a frame that caused this condition to occur, so that it can be

analyzed to determine the source of the frame. To that end, there is a diagnostic table at 00006000–000063FF that includes room for 16 records, each 64 bytes in size. The contents of the record is the time-of-day, followed by the first 56 bytes of the contents of the frame buffer. This is enough to include the frame header, the IP header, and the first 6 bytes of the UDP header. (If an extra 2 bytes is inserted before the real frame header, we would capture only the first 4 bytes of the UDP header.)

Foreign Node Capture diagnostics

Mindful of the potential for outside Internet nodes to gain access to local stations/IRMs, a diagnostic is included that captures network activities initiated by nodes outside Fermilab, or outside the local network. (Since Fermilab is considered a Class B network, its subnet mask is FFFF0000, so that “foreign” means any source IP address that does not begin with 131.225.

The information captured includes the time-of-day of the first frame and the last frame, the contents of the first and last frames, and the number of frames received. After 30 seconds of no activity, the record is closed. The buffer is located at 00006800–00006FFF and allows for 16 such records of foreign node activity.

TFTP Transaction Log

A data stream can be defined in a node to hold records of TFTP transactions that are recorded by the TFTP server local application called LOOP_TFTP. Each record is 16 bytes long and holds the time-of-day, the number of records transferred, the client node that performed the transaction, and the elapsed time needed for completion of the transaction. Node0562 is often the node from which IRMs boot, so it has such a data stream defined. A companion Page application called PAGE_TFTP is used to display the recent contents of the TFTP Log data stream.

Miscellaneous diagnostic variables

A large number of counters and captured information can be found in the ethernet variables area at 00160000–001601FF. In addition, a number of variables are maintained local to the SNAP Task, which processes the IP protocols for received frames. At present, this area is from 0004C612–0004C6D9.