

AN EXTENDED TCP/IP PROTOCOL OVER THE LOCAL AREA NETWORK FOR DCCS

Jaehyun Park, Youngchan Yoon, and Sangchul Lee

*Department of Industrial Automation,
Inha University,
Incheon 402-751, Korea
email:jhyun@rcsl.inha.ac.kr*

Abstract: This paper proposes an extended TCP/IP protocol over local area networks, that can be used for soft real-time systems including distributed computer control systems and manufacturing automation systems. Because the proposed protocol extends the standard TCP/IP, all the existing application softwares can be used with the proposed protocol without modification. The proposed TCP/IP also provides the periodic transmission mode (PTM), which provides a very efficient transmission method of the periodical data for real-time monitoring and control systems. With PTM, the periodical data collection and updating the control signals are possible with relatively small traffic overhead. This paper includes the computer simulation results, and the prototype system has been applied and evaluated to the large scale distributed control systems and human-like robot systems.

Keywords: Ethernet, real-time communication

1. INTRODUCTION

Ethernet and TCP/IP pair is one of the most popular protocols for the MAC (Medium Access Control) (Metcalfe and Boggs, 1976) and the middle range protocol layers in industry. Because TCP/IP is ported and used on hundreds of operating systems and hardwares, there are lots of system-independent applications developed on the top of TCP/IP. Although TCP/IP on the top of Ethernet works well for the ordinary network applications such as remote login, file transfer, and remote procedure call, this is hardly used for the real-time applications such as manufacturing automation and distributed computer control applications. The major difference between the general internet applications and real-time applications is the time constraints in the message delivery. The usual internet applications permits relatively large amount of variances in the message delivery time and throughput. The real-time applications, however, have strict or very tight requirements in

the message delivery time. Failure to meet these requirements may cause a severe trouble in the overall system. Although a sophisticated real-time network protocols can be used to meet these strict real-time constraints, general softwares can not be used on the top of these new real-time network protocols without modification. This paper focuses an approach to use the popular TCP/IP - Ethernet protocols for the soft real-time systems with keeping compatibility.

There are many researches to use Ethernet for the real-time applications. Because the packet collision of Ethernet is one of the major reasons why it can hardly be used for real-time application, most of the researches concentrate on solving the packet collision problem (Tanenbaum, 1989). BRAM (Broadcast Recognizing Access Method) (Chlamtac *et al.*, 1979) and MBRAM (Modified BRAM) (Signorile, 1988) are the protocols that resolve the collision problem in Ethernet network. Although, these protocols are collision-free proto-

cols and provide the deterministic latency, there are several reasons to prevent these collision-free MAC protocols from being widely used.

- No standard contention-free MAC protocol is fixed yet. The manufacturer of control devices and softwares are not willing to support non-standard protocols. They may want to support general solution based on world wide standard such as CSMA/CD, token ring, token bus, TCP/IP, and etc. Backward compatibility must be supported so that the existing software should be reusable.
- Reprogramming the I/O drivers is required for all the existing Ethernet devices. In case of the embedded MAC drivers, such as MC68EN360 CPU, it is impossible to modify MAC layer without redesign the chip.
- All nodes connected to physical media should support the same scheduling algorithm. Even single node that use standard MAC protocol may cause unpredictable congestion problems. Even though a special conflict-free protocol such as PCSMA (Yavatkar, 1994) is developed, most of the MAC protocols can not coexist with each other.

These reasons make it difficult to use modified MAC protocol for the real-time applications.

Another approach to use TCP/IP - Ethernet protocols for the real-time applications is to improve the performance of TCP/IP layer. TCP/IP was originally designed to be used in the inter-networking environment where long latency and relatively unreliable data transmission are expected. To cope with this unreliable connection due to inter-networking, multiple steps of error controls are provided. To maintain optimal utilization of the low bandwidth channel, standard TCP/IP uses the best-effort paradigm for packet scheduling, buffer management, feedback, and end adjustment(Lefelhocz *et al.*, 1996). Recently, a series of algorithms and implementation techniques improving the performance of the transmission control protocol(TCP) have been proposed: TCP for the transactions(ISI, 1994; ISI, 1992), receiving overhead reduction technique(Clark *et al.*, 1989), and header prediction algorithm(Jacobson, 1990). While these researches are useful for the conventional internet applications, they are not efficient for the real-time applications over Ethernet-only environment that most of the real-time systems are used in.

This paper proposes a modified transmission control protocol(TCP) to reduce communication latency over Ethernet link without modification of MAC layer. For this, this paper introduces another type of service called least-effort. Even though least-effort is not able to be applicable to the low bandwidth link, internet, it shows a

relatively high performance for a special purpose network such as manufacturing networks based on Ethernet that has a higher bandwidth than ordinary internet. This paper also proposes a special service called PTM(Periodic Transmission Mode). This mode increases utilization and results in lower transmission delay.

Next section describes the key features of the proposed extended TCP protocol. Implementation issues and evaluation results follows in the subsequent sections.

2. KEY FEATURES OF LAN-TCP

This paper extends the standard TCP protocol for the Ethernet-only environment, and the proposed protocol is called LAN-TCP(TCP for local area network only). The LAN-TCP has three major features : least-effort strategy, structure oriented protocol, and periodic transmission mode. In this section, these features are described in detail.

2.1 Least-effort strategy

The standard TCP follows the best-effort strategy to transmit a packet over a physical network link that is usually slow. TCP sliding window is a good example of explaining best-effort delivery (Comer, 1995; Stevens, 1994). It makes stream transmission efficient in wide area network, and it is an important feature of TCP for internetwork between high speed network and lower one. The best-effort strategy, however, increases the possibility of packet collision when used with high speed link like Ethernet. This packet collision is one of the major reasons of unexpected latency of packet delivery over Ethernet. To reduce this packet collision without modification of MAC layer, the least-effort strategy, contrary to the best-effort strategy, is proposed in this paper. The main idea of the least-effort protocol is that a certain limitation of the network bandwidth is reserved to single node, and the operating system is responsible for managing this limitation. The bandwidth limitation reserved to each node depends on the system architecture, the operating systems, and the transfer rate of media. The minimum bandwidth can be set to default or required amount requested by applications when applications open or bind SAP(Service Access Point) of transport layer or socket layer in BSD socket compatible environment. To maximize the effect of the least-effort strategy, LAN-TCP does not use the sliding window algorithm that is derived from the best-effort paradigm and developed for the inter-networking purpose. In the LAN-TCP, to disable the sliding window, the window size is limited to MSS(Maximum Segment

Size:(Comer, 1995; Stevens, 1994)). If one of two machine has a LAN-TCP facility, bandwidth limitation can be achieved by pseudo window size which is described in the next section.

Least-effort strategy is applied only to TCP or socket layer in LAN-TCP. There are four reasons for this restriction.

- Port de-multiplexing is applied only above the IP layer and the output data of IP layer are transmitted or queued to IP buffer without further de-multiplexing. Generally IP fragmentation is not required in local area network because TCP process already knows the size of MTU(Maximum Transfer Unit). If the bandwidth is not managed by TCP process, de-multiplexed TCP segments stored at IP buffer must be re-arranged to distribute transmission rate of each application processes. This rearrangement will be another overhead of protocol layering and make the buffer management complex.
- Application process should be pended until sending system call is terminated if multiple segments are sent, and TCP process passes segments to the IP process at intervals for pending duration.
- LAN-TCP should be compatible and coexistable with TCP. If least-effort delivery is not required, the standard TCP connection should be explicitly used. IP can not negotiate connection type with remote machine due to its connection-less feature. Compatibility is an important feature of LAN-TCP. If LAN-TCP can not support conventional TCP protocol, the problem of difficulty of implementation remains same as collision-free MAC protocol.
- IP or TCP process must know which application process is sending packet and how much bandwidth is reserved. IP process can be modified to deal with these information, but TCP process already has these facility known as TCB(Transmission Control Block) (Comer and Stevens, 1994; Wright and Stevens, 1995). TCB can be easily modified to implement least-effort mechanism.

2.2 Structure oriented protocol

TCP is a stream oriented protocol, and TCP's stream is unstructured. This unstructured stream makes programming inefficient. On the contrary, UDP uses structured datagram delivery, but UDP doesn't have transmission control facility. Application process using UDP is responsible for controlling transmission. LAN-TCP is datagram and reliable stream oriented protocol. In the LAN-TCP, stream is not divided into bytes but struc-

tured datagram. The destination machine passes to the receiver exactly the same sequence and size of blocks that the sender passes to it on the source machine. This feature makes buffer management of LAN-TCP easy and structural.

2.3 Periodic transmission mode

The periodic data transmission is one of the fundamental communication method for the real-time systems including distributed computer control systems. When a packet is transmitted periodically, the receiver is able to predict next arrival time in local area network. LAN-TCP supports a special transmission mode for the periodic transmission called PTM(periodic transmission mode). To identify the type of segment data, TCP header contains 6-bit fields: URG, ACK, PSH, RST, SYN, and FIN. Each bit is used to initiate transition of TCP state machine(Comer, 1995; Stevens, 1994). Among them, ACK bit is important for the PTM mode. To the TCP acknowledgment scheme, ACK is sent by destination machine after segment from source machine was received. Although destination machine can generate only last ACK for acknowledgment of multiple segments in sliding window algorithm, the receiver always generate ACK when segment is transmitted periodically because each segments are transmitted every the fixed interval. Generation of series of ACK can be avoided in PTM mode. ACK packets are not sent after PTM connection is established. ACK packet is used only for negative acknowledgment when destination machine didn't receive segment until next arrival time. From the queuing analysis of CSMA/CD, PTM can reduce the media traffic up to half or less(Rom and Sidi, 1990; Tobagi and Hunt, 1980).

3. IMPLEMENTATION

This section describes major issues to implement the LAN-TCP. The various techniques for optimizing TCP is not implemented in the prototype LAN-TCP yet.

3.1 Connection management

Only a slight modification of standard TCP is required for implementing the LAN-TCP. The connection establishing and closing scheme of LAN-TCP is almost same as that of the standard TCP except the PTM mode. Generally, TCP options follow only SYN segment while the connection is established. Standard TCP header contains 6 reserved bits for future use. Two unused bits are used for negotiating connection of LAN-TCP.

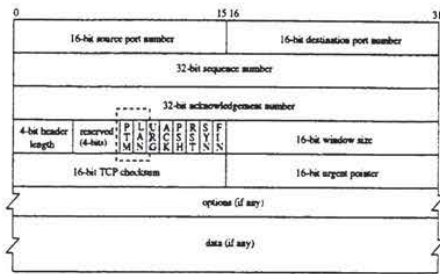


Fig. 1. LAN-TCP header

One is for LAN-TCP connection and the other is for PTM connection. The position of two bits is shown in Fig. 1, enclosed by dashed box. 'LAN' is used for LAN-TCP connection and 'PTM' is used for PTM connection. According to IAB internet standard, these bits must be zero(ISI, 1981).

There are some congestion control mechanism for best-effort delivery such as slow start algorithm, round-trip time measurement, Nagle algorithm. These do not need to be implemented in LAN-TCP because LAN-TCP is only for local area network especially inside subnet.

3.2 Additional timers for LAN-TCP

TCP process requires various timers to control transmission: retransmission timer, persist timer, and keep alive timer(Wright and Stevens, 1995). Some of them are disabled while LAN-TCP connection is established. Besides these standard timers, two additional timers are added for LAN-TCP: in-band timer and PTM timer. In-band timer is used by sender side for bandwidth limitation, and PTM timer is used by receiver side to send ACK for negative acknowledgment while PTM connection is established. Before a TCP segment is sent, TCP process must check whether in-band timer is expired or not. If in-band timer is already expired, segment is sent immediately, otherwise, TCP process is delayed until the timer is expired. After a segment was sent, in-band timer is initiated. The operation of two timers are depicted in Fig. 2 and Fig. 3. If PTM connection was established, all of TCP timer must be disabled in PTM connection requester except keep alive timer, in-band timer, and PTM timer. In most operating systems, the resolution of TCP timer is around several tens of mili-second unit. This resolution is sufficient to manage control mechanism of the standard TCP, but higher resolution is required for LAN-TCP. It is somewhat easy to obtain higher resolution in RTOS(Real Time Operating System). POSIX standard supports a function 'nanosleep' which has higher resolution, but real resolution is bounded up system clock and depends on operating system and processor. To fix this variation, the resolution of LAN-TCP timer is

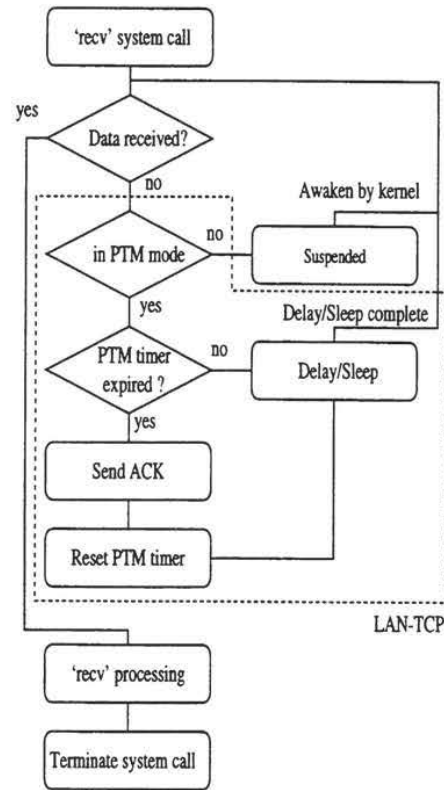


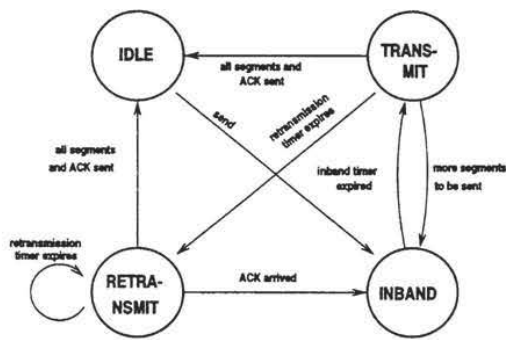
Fig. 2. LAN-TCP recv processing

set to the system clock. With this resolution, the maximum bandwidth depends on system clock. For example, if system clock is 60Hz and media MTU is 1500, then burst transfer rate of single process can be up to 90 kbytes.

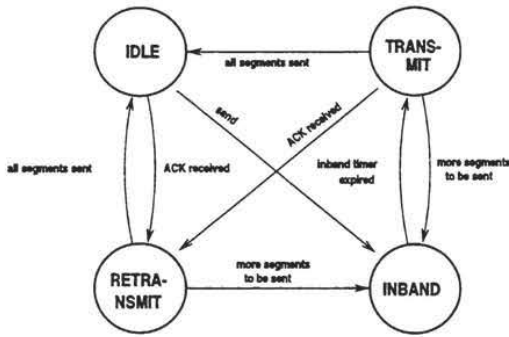
3.3 Periodic transmission mode

The PTM connection is not bi-directional. When LAN-TCP connection is established, only sender set 'PTM' bit in SYN packet, thus uni-directional PTM connection is established. Transmission interval is automatically detected by receiver side, and it decreases from initial value to mean value gradually.

In PTM mode, PTM timer notifies that next packet is not arrived yet, and ACK, which means negative acknowledgment in PTM mode, must be sent(see Fig. 2). If sender side receives ACK, sequence number of ACK is checked, and corresponding segment must be sent again. Standard TCP process already has a facility, retransmission buffer, enough to solve this problem. Mostly ring-style buffer is used with retransmission buffer. In typical TCP process, acknowledged segment in retransmission buffer is removed, but it must be left in PTM mode. Segment in ring buffer is not removed but destroyed by head pointer of ring buffer step by step. If receiver side can not process



(a) Normal mode



(b) PTM mode

Fig. 3. finite state machine used for LAN-TCP send processing

periodic packets in time by unexpected disturbance, then series of segments can be dropped. If dropped packets are out of ring buffer, these packets can not be retransmitted. In this case, the receiver must send source quench ICMP message, or the connection will be aborted. ICMP is only for internet protocol, but it is the only way to notify alert status.

3.4 Backward compatibility

Because the LAN-TCP is compatible with the standard TCP, the connection between the LAN-TCP nodes and the ordinary TCP nodes can be established, and two protocols work well without conflict at all. TCP has a facility known as window size advertisement to manage sliding windows. Window size in TCP header(see Fig. 1) is the size of free TCP buffer in remote machine's TCP process. Local machine sends full size of window which can be received and stored in remote machine. To prevent remote machine from sending more than one segment of window, pseudo window size is used in LAN-TCP. If one of two machine adopts LAN-TCP, bandwidth limitation can be achieved easily by pseudo window size. LAN-TCP uses pseudo window size rather than actual window size when the remote machine doesn't support bandwidth limitation. The maximum pseudo

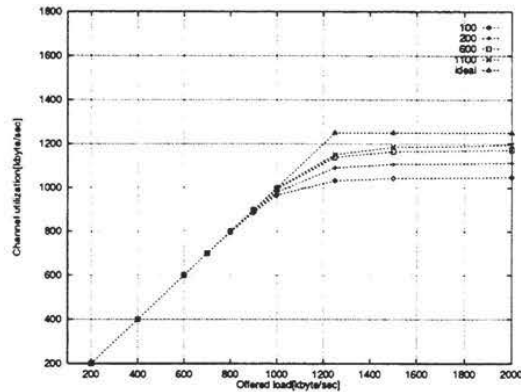


Fig. 4. Evaluated Performance of CSMA/CD with 20 nodes

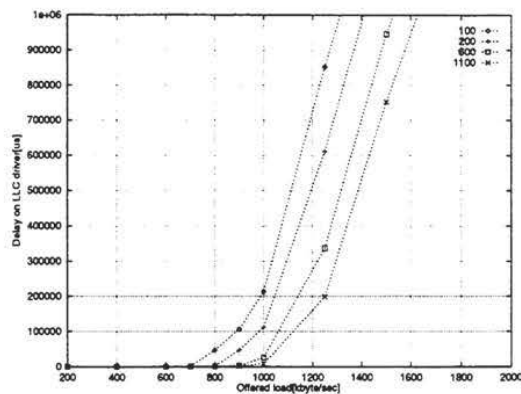


Fig. 5. Delay on LLC with 20 nodes

window size will be limited to MSS(Maximum Segment Size).

If applications use PTM, they need to use special socket system call rather than standard socket system call. These applications pass parameters different from normal LAN-TCP connection when create the end point of connection. For example, 'socket' system call can be called with new parameter 'SOCK_PSTREAM' rather than 'SOCK_STREAM', if PTM connection is required. If remote machine doesn't support LAN-TCP, PTM connection can not be established, thus the connection will be aborted. In this case, normal LAN-TCP connection is established and works well without conflict with conventional TCP.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed protocol, a computer simulation package is developed. In the CSMA/CD protocol, if two stations detect an idle state and begin transmission at the same time, a collision will occur. Any station detecting a collision stops its transmission and generates jam signal, and then waits a random

period of time before retransmission. To avoid overloading the channel, the range of retransmission interval increases using a binary exponential back off algorithm. The complete analysis of binary exponential back off algorithm is very complex. Even though the simplified models were proposed(Tobagi and Hunt, 1980; Boggs *et al.*, 1988), the precise model with complete binary exponential back off algorithm should be used to evaluate the performance of the proposed protocol. Additionally, to evaluate the effect of physical environment, simulation model should contain a facility of handling propagation delay.

The first simulation is evaluating the basic utilization under heavy load to validate the computer simulation package itself. Utilization of the CSMA/CD shown in Fig. 4 is almost same as the measured utilization by Shoch and Hupp(Shoch and Hupp, 1980). Evaluated utilization is up to 95% with 20 nodes and 1100bytes of packet length. Average transmission delay from network layer to the destination node is shown in Fig. 5.

4.1 Simulation model

The network configuration simulated consists of four components(two periodic servers, one burst server, and one burst client), and two groups of periodic sources. Periodic server receives packets from the periodic data sources and sends acknowledgment packet to them in standard TCP model, but in the LAN-TCP model, acknowledgment packet is not sent.

Burst server and burst client exchange series of segments and acknowledgment packets as sliding window scheme. Burst client in this simulator does not generate packets by probability function but follows actual transmission scheme which exchanges one acknowledgment packet per several segments. Burst client does its best to transmit packets as soon as possible in conventional TCP model, but in the LAN-TCP model, transmission interval is never exceeded a preset limitation reserved at each node.

4.2 Simulation results

Fig. 6 through Fig. 8 show a delay, average load, and number of collisions using the proposed protocol and standard TCP for the same network configuration. Fig. 6 shows the transmission delay observed at the network layer under the various load condition generated by periodic sources with 10 % of variance of transmission interval. The efficiency of PTM connection can be measured in this figure. There is significantly higher delay under conventional TCP than proposed protocol when

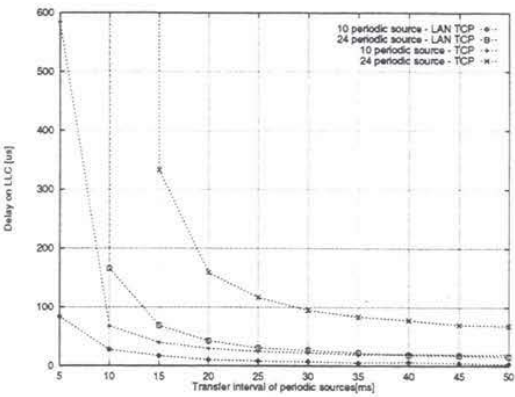


Fig. 6. Delay on LLC under loads generated by periodic sources (packet length is 258bytes)

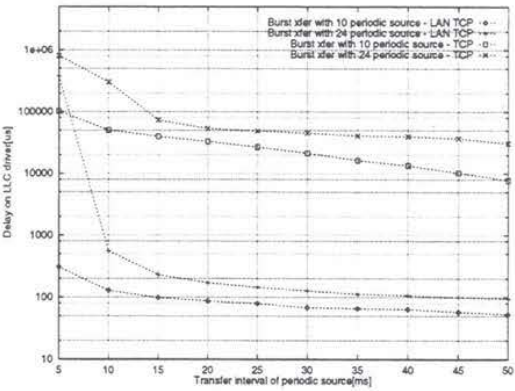


Fig. 7. Transmission delay caused by transient burst transmission

burst transmission is performed(Fig. 7). In a TCP burst transmission, average load can not exceed by unity because the transmission is based on packet exchange(Fig. 9). Even though this type of traffic is not continued long, it can cause critical problem in industrial network, especially very long delay, even transmission failure on MAC layer. From the simulation result, many transmission failures caused by 16 contiguous collisions are occurred as the number of collisions is increased (Fig. 8). The reason of huge delay is best-effort transmission which makes the average load around unity abruptly. Fig. 9 shows the effect of best-effort transmission under low load. For example, offered load is 348kbytes/sec with 25 mili-second of transfer interval by 24 TCP periodic sources. The load is increased to 1044kbytes/sec after burst transmission is started. This load results in significant transmission delay(Fig 7) and collisions(Fig 8). By the simulation result, 20~220 packets are dropped by transmission failure.

Of course, this number of collisions is not practical because burst transmission is never continued long during tens second. The number of collisions and transmission delay are varied by the start time of periodic transmission. Consider the

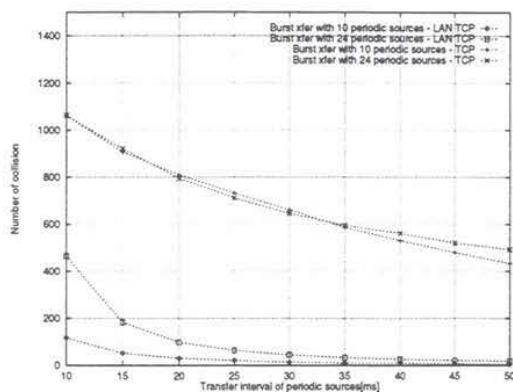


Fig. 8. Number of collisions under burst transmission

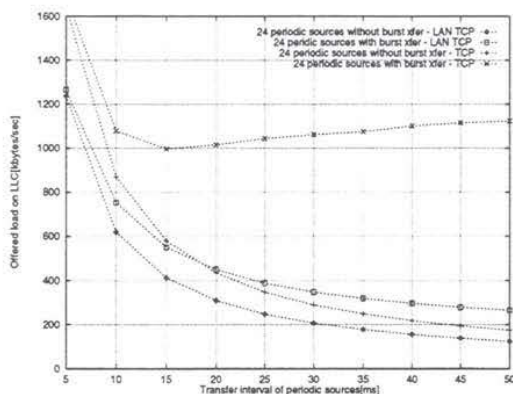


Fig. 9. A variation of channel utilization

worst case, if all nodes start periodic transmission and the transmission interval is not varied, then extreme transmission delay is occurred without burst transmission. On the other hand, if the start time is well distributed, then no collision is detected at all. So, periodic transmission which is based on system clock has potential problem.

If the simulation results(Fig 6, Fig 7, and Fig 8) are acceptable in real world, 20 periodic sources generating 200bytes of control data at each 10 milli-second and some LAN-TCP hosts can be feasible configuration.

5. CONCLUSION

This paper describes an effort to use industry standard TCP/IP - Ethernet for soft real-time systems including distributed computer control systems and manufacturing automation systems. To maintain the backward compatibility, the standard TCP header is extended to reduce traffic overhead and include periodic transmission mode. Because the proposed protocol extends the standard TCP/IP, all the existing application softwares the proposed TCP/IP can be used without modification. Besides soft real-time charac-

teristics, the proposed TCP/IP includes periodic transmission mode(PTM), which provides a very efficient transmission method of the periodical data for real-time monitoring and control systems. With PTM, periodical data collection and updating the control signal are possible with relatively small traffic overhead. Computer simulation shows the proposed TCP/IP has relative higher performance compare to the ordinary TCP/IP protocol.

The prototype system has been applied and evaluated to the large scale distributed control systems and human-like robot systems. Optimization of the proposed protocol is required to be used for a large scale distributed system and fine calibration may be necessary to maximize real-time performance to the base operating system.

6. REFERENCES

- Boggs, D.R., J.C. Mogul and C.A. Kent (1988). Measured capacity of an ethernet: Myths and reality. *Proceeding of ACM SIGCOMM '88* pp. 222-234.
- Chlamtac, Imrich, William R. Franta and K. Dan Levin (1979). BRAM: The broadcast recognizing access method. *IEEE Transactions on Communications* 27(8), 1183-1189.
- Clark, David D., Van Jacobson, John Romkey and Howard Salwen (1989). An analysis of tcp processing overhead. *IEEE Communication Magazine* pp. 23-29.
- Comer, Douglase (1995). *Internetworking with TCP/IP Volume I : Principles, Protocols, and Architecture*. Prentice Hall, Inc. Englewood Cliffs, New Jersey.
- Comer, Douglase and David L. Stevens (1994). *Internetworking with TCP/IP Volume II : Design, Implementation, and Internals*. Prentice Hall, Inc. Englewood Cliffs, New Jersey.
- ISI (1981). Transmission control protocol. *RFC793*.
- ISI (1992). Extending TCP for transactions - concepts. *RFC1379*.
- ISI (1994). T/TCP - TCP extensions for transactions functional specification. *RFC1644*.
- Jacobson, V. (1990). 4BSD TCP header prediction. *Computer Communication Review*.
- Lefelhocz, Christopher, Bryan Lyles, Scott Shenker and Lixia Zhang (1996). Congestion control for best-effort service: Why we need a new paradigm. *IEEE Network* pp. 10-19.
- Metcalf, R.M. and D.R. Boggs (1976). Ethernet: Distributed packet switching for local computer networks. *Communication of the ACM* 19(7), 395-404.
- Rom, Raphael and Moshe Sidi (1990). *Multiple Access Protocols*. Springer-Verlag. New York.

- Shoch, John F. and Jon A. Hupp (1980). *Measured Performance of an Ethernet Local Network*. Xerox Corporation. California.
- Signorile, Robert P. (1988). MBRAM-A priority protocol for PC based local area networks. *IEEE Network* 2(4), 55-59.
- Stevens, W. Richard (1994). *TCP/IP Illustrated, Volume 1*. Addison-Wesley Publishing Company. Massachusetts.
- Tanenbaum, Andrew S. (1989). *Computer Networks*. Prentice Hall, Inc.. Englewood Cliffs, New Jersey.
- Tobagi, F.A. and V.B. Hunt (1980). Performance analysis of carrier sense multiple access with collision detection. *Comput. Networks* 4, 245-259.
- Wright, Gary R. and W. Richard Stevens (1995). *TCP/IP Illustrated, Volume 2 : The Implementation*. Addison-Wesley Publishing Company. Massachusetts.
- Yavatkar, Rajendra (1994). A reservation-based CSMA protocol for intergrated manufacturing networks. *IEEE Trans. on Systems, MAN, and Cybernetics* 24(8), 1247-1258.