



## ***WHITE PAPER***

### **PACKET AGGREGATION FOR C.LINK NETWORKS**

**DOCUMENT No** 18-00WP-D008  
**VERSION** A  
**VERSION DATE** JUNE 25, 2007

## Table of Contents

1	Abstract .....	1
2	Uses Of Packet Aggregation.....	2
3	c.LINK Packet Aggregation.....	3
4	c.LINK Packet Error Handling.....	6
5	Two-Node Packet Aggregation.....	8
6	Mesh Mode Packet Aggregation.....	9
7	Summary .....	10

## Table of Figures

FIGURE 1	EXAMPLE AGGREGATED C.LINK PACKET .....	4
FIGURE 2	PACKET AGGREGATION PERFORMANCE IMPROVEMENT .....	5
FIGURE 3	EXAMPLE RESIDENTIAL GATEWAY AND ROUTER IMPLEMENTATION.....	8

# 1 Abstract

Packet aggregation serves to maximize the efficiency of a network's bandwidth by reducing the overhead associated with transmitting data. A typical packet based communications protocol requires transmission of control information that serves to delineate the start of a packet and describe the contents of its payload data. This control information is comprised of PHY preambles and MAC headers that do not usually vary in size for any given transmission. Depending on the size of the payloads, combining multiple Protocol Data Units into a single Entropic c.LINK MAC frame can greatly increase the ratio of useful payload data to control information and reduce the associated protocol overhead per transmission. By improving the ratio of payload to control information, more network bandwidth is used for application data rather than control information, resulting in improved packet throughput. Using Entropic c.LINK packet aggregation, a system operator can significantly improve their network efficiency to offer the best possible throughput and assure a satisfying user experience for their subscribers.

## 2 Uses of Packet Aggregation

Home networking traffic can consist of a mixture of small and large packet sizes. Traditional applications such as file downloads, heavily buffered web video and web page refreshes can use large packets that are transferred from the WAN in long back-to-back bursts. Applications that require real-time data flows like interactive gaming typically use small IP packets transmitted at regular intervals. Even the TCP/IP protocol makes use of small packets in the form of ACKs, of which there may be several within a 1 millisecond period. Without packet aggregation each of these packets is, regardless of their individual sizes, transmitted within a single c.LINK frame. These examples demonstrate the inefficient use of a network bandwidth due to all the overhead associated with transmission. Packet aggregation serves to maximize the bandwidth efficiency of not just small packets but large packets as well. Multiple large packets can be combined together or a combination of small and large packets can be combined to share the overhead of a c.LINK frame destined to the same c.LINK node. This is especially useful when using c.LINK for a WAN connection where virtually all packets destined from the WAN can be aggregated for reception by a c.LINK router.

For system operators to extract the greatest value from their c.LINK home networks the need to maximize network throughput without compromising robustness becomes paramount. Entropic c.LINK packet aggregation fulfills this requirement by combining multiple packets together, both large and small, into a single c.LINK frame with error checking capability. Additionally, the overall inter-frame gap (IFG) time and number of reservation requests are reduced since those packets which have been aggregated no longer need their own individual c.LINK frames. For example, a c.LINK frame that has six aggregated packets saves  $5 \times 10 = 50$  microseconds of IFG time and can potentially eliminate up to five reservation requests. All this serves to maximize bandwidth efficiency within the c.LINK network by reducing the percentage of protocol overhead relative to application data and thus increase the packet-per-second throughput.

### 3 c.LINK Packet Aggregation

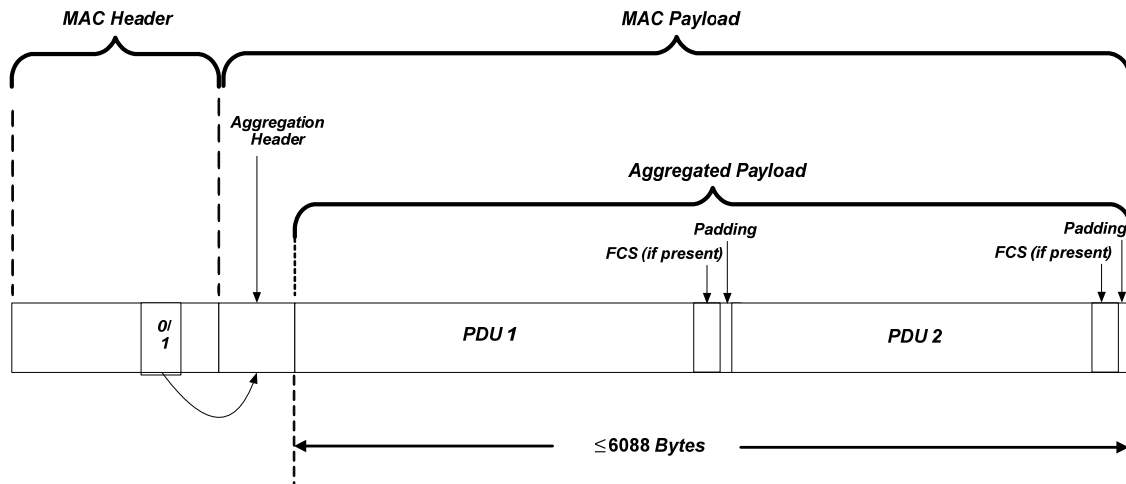
Entropic has carefully studied the requirements for implementing packet aggregation on c.LINK networks in order to painlessly introduce this highly beneficial feature into system operator deployments with minimal disruption. Entropic's c.LINK packet aggregation accomplishes this principal goal by using a combination of efficient software and flexible hardware.

c.LINK packet aggregation uses efficient packet queuing and aggregation algorithms working in conjunction with highly flexible DMA controllers that require minimal CPU attention. As a result, packet aggregation can be enabled on preexisting EN2210 based products without requiring any hardware changes. Systems can use the same crystal and boot code memory devices that are already used with existing c.LINK EN2210 based products. An additional benefit of c.LINK's economical and self-contained implementation is that it allows packet aggregation to work across a range of products, from low-cost Layer 2 c.LINK bridges having no host controller to advanced c.LINK routers and gateways using network processors and offering advanced network services.

In order to minimize the impact to higher network layers within the c.LINK network, Entropic's c.LINK architecture relies only on MAC layer packet aggregation. The aggregation is completely transparent to the higher network layers and is an autonomous function within the c.LINK function of a node. Packet aggregation is automatically performed within the c.LINK devices and only occurs between nodes which are aggregation capable.

During the admission process, a c.LINK node declares its capabilities to the network coordinator which in turn broadcasts those capabilities to the other nodes in the network. Because the network coordinator is not involved in the packet aggregation decision process, packet aggregation can occur even in networks where the network coordinator is not aggregation capable. This capability allows operators to incrementally deploy and utilize this advanced feature into existing MoCA 1.0 networks.

Packet aggregation is made possible since a c.LINK frame currently has a maximum packet size of approximately 6K bytes, including the packet header. Into one of these c.LINK frames, multiple Protocol Data Units (PDU) can be combined. The PDU consists of an actual unit of application data such as an Ethernet packet and can accommodate a full length Ethernet packet of 1522 bytes. Currently, a single c.LINK frame can aggregate up to 6 variable length PDUs or 4 full sized Ethernet (1522 byte) PDUs that are destined for the same node. A transmitting node requests that the NC allocate time slots equivalent to the actual size of an aggregated packet. All packets transmitted by c.LINK use the reserved field in the MAC header to indicate whether a c.LINK packet is aggregated. An aggregated packet will have an aggregation header following the MAC header. Figure 1 shows the basic structure of an aggregated c.LINK packet.



**Figure 1: Example Aggregated c.LINK Packet**

The aggregation header contains information on the number of packets aggregated, size of each individual PDU and a header checksum. Tradeoffs must be evaluated in determining an optimum number of PDUs to aggregate. Aggregating too many small PDUs can end up increasing the overall system cost. For example, the performance of the system processor or network processor is likely one factor that could be limiting small packet throughput performance. Aggregation of 6 PDUs already enables a sustained packet rate of 50,000 to 60,000 packets-per-second which may be difficult to sustain on some processors. However, before using a more powerful and expensive processor the profile of traffic, mix of large versus small packets, should be understood in order to determine whether that much small packet throughput is truly required. Otherwise, the result can end up being a costlier system, over-designed for an unrealistic usage scenario.

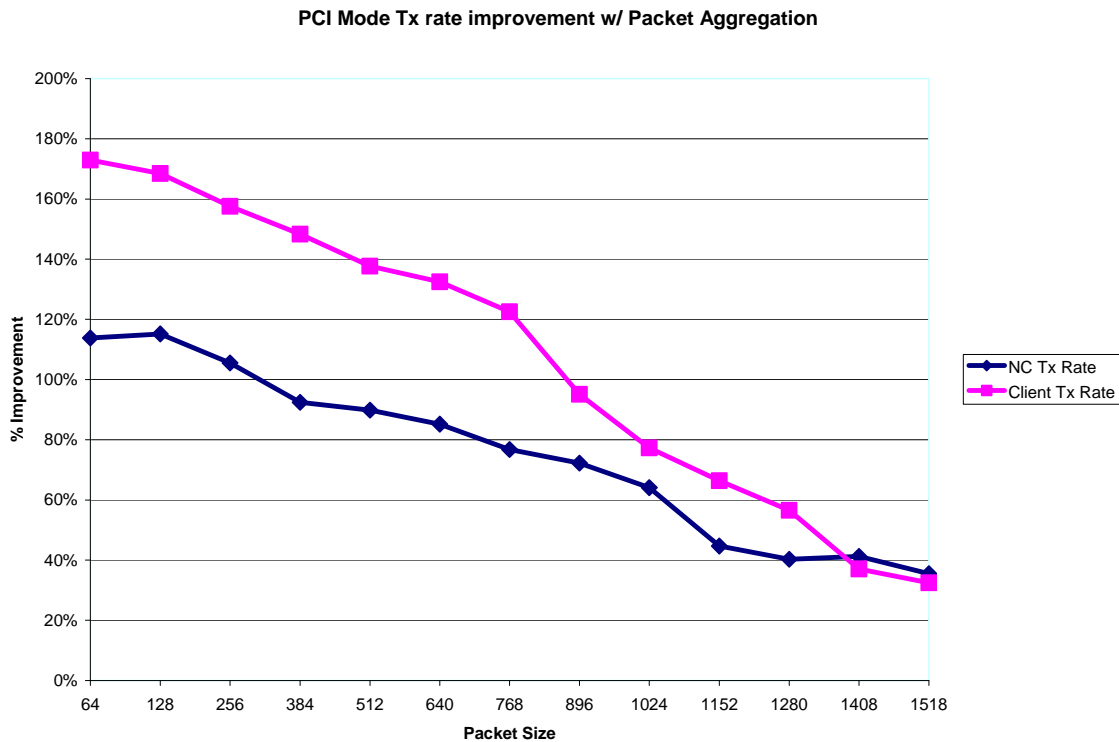
c.LINK packet aggregation occurs automatically and requires no intervention from higher network layers. As the network approaches capacity, multiple packets begin to queue up at the transmitter. When this occurs, packets that are normally waiting for their individual c.LINK frames in order to be transmitted over the same network segment are instead automatically aggregated together. Since aggregation does not occur when network bandwidth frees up there is no penalty in increased packet latency for aggregated or non-aggregated packets.

Preliminary results are provided in Figure 2 and show the dramatic improvement in packet-per-second throughput that packet aggregation can yield on a c.LINK node using the PCI interface. Typically, the network coordinator has better throughput performance for all packet sizes but with c.LINK packet aggregation, client node performance comes close to matching the NC performance over all packet sizes. Packet aggregation not only dramatically improves small packet throughput on client nodes, but in addition both client nodes and NCs will experience overall throughput increases with large packets as well. Lab results have shown rates approaching 185 Mbps for full sized packets (1522 bytes) in a two node network compared to 125 Mbps without packet aggregation.

For situations where a PDU is too large to fit in the remaining space of an aggregated c.LINK frame, packet fragmentation was considered. For example, if an aggregated frame had space remaining for 1000 bytes, a full-size 1522 byte Ethernet packet could be fragmented into two packets, one 1000 byte fragment and one 522 byte fragment. The 1000 byte fragment could be aggregated in the current frame and the 522 byte fragment aggregated with the following frame. However, implementing fragmentation could require additional memory in order to store the fragmented packets as they arrive and wait to be processed for reassembly. Additionally, in the event that transmission errors occur, retransmission algorithms were considered to resend corrupted fragments. The disadvantages of implementing retransmission are discussed in section 4 and a simpler more economical approach to mitigating transmission errors is utilized. Only 3% to 5% of additional throughput can be achieved with fragmentation but with significantly increased cost and complexity. On its face packet fragmentation holds the promise of higher network throughput. The

reality is that it provides minimal gains while increasing device costs and network complexity. Actual test data show that c.LINK packet aggregation in of itself provides 95% of the potential throughput increase with absolutely no cost impact, unnecessary complexity, or backward compatibility issues.

Another factor impacting the ultimate throughput of packet aggregation is network latency. In a c.LINK network the network coordinator has latency advantages over the other nodes when it comes to putting its own data packets onto the network. In order for a c.LINK node to schedule time on the wire it must first submit a request to the network coordinator. This step is eliminated when the network coordinator schedules its own packets resulting in a lower latency. Lower latency means the transmitter can put data on the network faster resulting in higher throughput for a given memory allocation. Therefore, the ability to dictate who assumes the network coordinator role is critical when trying to maximize the transmission bandwidth of a specific node. Note that the data in Figure 2 shows the performance improvement percentage, raw packet-per-second throughput is typically greater for the NC when packet aggregation is not utilized. To favor lower latency of a particular node, Entropic has introduced a feature called Preferred NC which allows the operator to specify a node to be the network coordinator (NC). With packet aggregation and preferred network coordinator, the efficiency of the c.LINK network is driven toward its theoretical limits.



**Figure 2: Packet Aggregation Performance Improvement**

## 4 c.LINK Packet Error Handling

One potential drawback to packet aggregation is that because there are multiple PDUs encapsulated in a single c.LINK packet, an error during transmission of a packet could have a greater effect on the overall PDU packet error rate. However, preliminary lab results showing PDU packet error rates still on the order of  $1e-6$  demonstrate the PDU packet error rates are not significantly impacted with packet aggregation. This result is backed by a theoretical analysis.

In the c.LINK system, the error probability is the same for every byte transmitted and received. Therefore the probability of losing a packet is directly proportional to the number of bytes that must be correctly received in order to fully recover the packet. In the non-aggregated case, a maximum sized packet consisting of 1522 bytes has the highest probability of error, with each byte having a probability of error.

When multiple PDUs are aggregated, the maximum PDU size is still 1522 bytes so the probability that one particular aggregated PDU has an error is the same as in the non-aggregated case. This should not be confused with the probability that ANY PDU in an aggregated packet is in error. Certainly if four 1522 byte packets are aggregated, the total length is 4 times longer and therefore the probability of an error somewhere in the total packet is 4 times larger. However, any individual PDU in the total aggregated packet is still 1522 bytes long and the probability of an error is still the same as a single non-aggregated 1522 byte packet. In order to take advantage of this fact, the c.LINK aggregation scheme allows for a frame check sum (FCS) for each aggregated PDU. By having individual FCSes, individual PDUs can be independently checked for errors so even if another PDU in the aggregated packets has errors, good PDUs are preserved and recoverable.

The only significant increase in probability of errors due to packet aggregation results from the transmission of an additional aggregation header. If the aggregation header is in error, all the individual PDUs in an aggregated packet may not be recoverable. As explained below, the increase in error probability caused by this is less than 4%.

The aggregation header used for aggregating four 1522 byte packets is 14 bytes including a header FCS. As discussed, the error probability is proportional to the length of the message so the probability of the aggregation header being in error is  $1522/14 = 108.7$  times less than a 1522 byte packet being in error. This means for every 108.7 maximum size PDUs that has an error, the system will have one header error which will result in up to 4 full-size packets lost. So the increase in overall error rate of maximum size packets due to the exposure of the aggregation header is  $112.7/108.7$ , which is less than 4%. Therefore where there was  $1e-6$  error rate before, there will be  $1.04e-6$  error rate instead. It is important to note that this additional error rate is a theoretical maximum and will only be experienced when the network is being fully utilized with a significant portion of the traffic being aggregated.

Another approach to mitigating the issue could be to use a retransmission technique such as automatic repeat-request (ARQ) transmission whereby the packet is sent repeatedly in the absence of an acknowledgment from the receiver. However, ARQ transmissions are not ideally suited for applications in which strict jitter and latency requirements must be maintained.

By nature, retransmission schemes introduce network latency and consume extra bandwidth. For streaming video applications in which the timely delivery of data is just as important as guaranteed delivery, this added latency can result in data arriving at the receiver too late to be of any use, ultimately causing video degradation. Furthermore, a proper retransmission implementation requires additional storage elements. Memory is required for storing the transmitted packet till an acknowledgment is received or the retransmission times out. Additional memory is also needed at the receiver to reorder packets in the event a retransmission is necessary. This added memory simply drives up solution costs for an approach that can have negative consequences on the network.

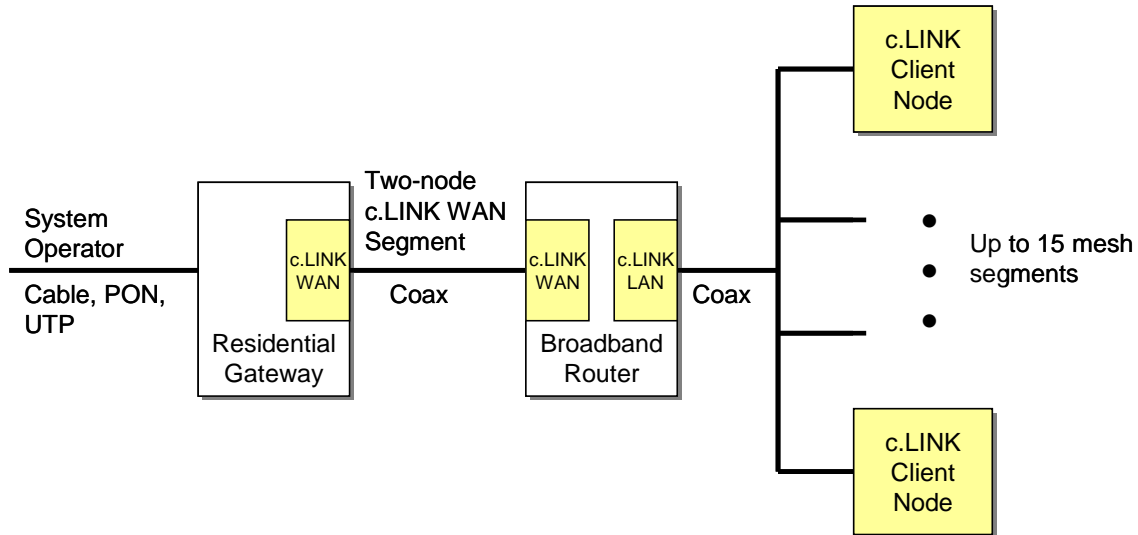
Using an ARQ transmission scheme is also problematic in that it only addresses future product deployments. Complex changes to the basic protocol like this require significant modifications to the

system and is not backwards compatible with existing deployments. Products currently deployed by operators would not yield any measurable benefits and careful consideration must be given to how a node with retransmission interoperates with nodes not supporting retransmission. Retransmission techniques add a lot of complexity and cost to correct a statistically insignificant problem created only under heavy conditions.

As shown above, a simpler and less obtrusive approach to mitigating increased packet error rates which may be caused by aggregation can be achieved by utilizing FCSes in the aggregated PDUs of a c.LINK packet. By using this approach system operators gain the benefits of increased packet rate throughput with PER mitigation without impacting network latency or increasing system costs and the ability to maintain backward interoperability.

## 5 Two-Node Packet Aggregation

Two-node packet aggregation is a special instance of packet aggregation appropriate for WAN applications where for example, a residential gateway connects to a broadband home router through a c.LINK segment. Figure 3 illustrates the various c.LINK segments in a typical home network deployment. This usage model expects that the downstream traffic from the gateway to the broadband router is significantly higher compared to the upstream traffic from the router to the gateway. In this scenario the gateway needs to be efficient on transmit path while the router needs to be efficient on receive path.



**Figure 3: Example Residential Gateway and Router Implementation**

The gateway can maximize its transmission efficiency by assuming the network coordinator (NC) role and reducing latency. This can be accomplished by having the gateway act as a preferred c.LINK network coordinator. As discussed earlier this has the added benefit of increasing network bandwidth for a given memory allocation. The Preferred NC capability is controlled by software configuration and the new advanced capability bits.

## 6 Mesh Mode Packet Aggregation

In a 16-node fully meshed network each c.LINK client node has the ability to send traffic directly to any of the other 15 nodes. If each node was given an opportunity to schedule transmission of packets every scheduling cycle in a round-robin fashion, the overall network bandwidth would be impacted reducing the benefits provided with packet aggregation. To improve packet throughput, c.LINK has implemented a heuristic scheduling algorithm to address scheduling latencies caused when there are many nodes on the network but only a few are actually sourcing data. In this instance, the network coordinator will schedule reservation requests more frequently to nodes that are actively transmitting data compared to the nodes that are not transmitting, as is typical in most video network topologies.

## 7 Summary

Packet aggregation is a significant feature essential for maximizing c.LINK network throughput. Packet throughput can be increased anywhere from 30% to 170% depending on PDU size and the role as client or network coordinator. Moreover, as subscribers increasingly use more real-time applications where the network carries a mixture of small and large packets these significant performance gains produce improved robustness at the application level which in turn translates into a superior user experience. Packet aggregation yields gains for both the system operator and the subscriber. By employing Entropic's intelligent packet aggregation in its c.LINK solutions, system operators can rest assured they are wringing out the best performance available from their home networks.