# Frame Types and Protocol Extensibility

Tony Jeffree
01 September 1998

This note was prompted by the discussion at the July meeting of the 802.3 Link Aggregation Task Force following the proposal made by Jeff Lynch, *et al* regarding the potential use of 802.3 MAC Control frames as the vehicle for carrying Link Aggregation Control Protocol PDUs (LACPDUs). The main focus of that presentation was not the detail of the PDU contents (LACP parameter values, flags, etc.); rather, the nature of the wrapper in which the contents is to be carried.

## 1. Introduction

The principal objectives in choosing an appropriate format for LACPDUs would seem to be:

1. That the format should be capable of carrying the required payload;
2. That LACPDUs shall use a (multicast) destination address that is not forwarded by (non-Link Aggregation aware) Bridges;
3. That PDUs carried in that format should be readily identifiable as an LACPDUs; i.e., that there should be a clear protocol identification mechanism;
4. That the format chosen should preferably not be incompatible with existing hardware that might otherwise be enhanced/modified by to make use of the LACP;
5. That the format should preferably allow extensibility, to accommodate such future enhancements to the protocol as may be developed.

The first three of these requirements are relatively easy to address.

The first requirement essentially boils down to a PDU size issue. The protocol mechanisms that have been discussed to date are relatively abstemious in terms of the size of payload that they would require, with the possible exception of extending the protocol to cover use on shared media, which is in any case a controversial area of discussion.

The second requirement is straightforward. If MAC control frames are used, this requirement is already met. If some other solution is chosen, one of the spare addresses in the set defined in 802.1D can be requested and allocated.

The third requirement can readily be met by approaches that are already well-known. The choice effectively boils down to the following three options:

- Use of an Ethernet protocol type (as is done for MAC Control and VLAN Tag Headers, for example);
- Use of a specifically allocated LLC address (as is done for MAC Bridge BPDUs, for example);
- Use of a SNAP identifier (as is done for bridging Ethertypes onto non-802.3 media, for example).

Given the recent history, and the prevalence of use of protocols identified by Type values on 802.3 LANs, the first of these choices is probably the most acceptable.

The fourth and fifth requirements are rather more difficult to address, and are dealt with in the following sections.

## 2. Compatibility with existing equipment

Link Aggregation can potentially be used with a wide variety of existing equipment:

- Servers, or high end workstations, that might aggregate a number of links using existing NICs with modified driver software;
- Switches, routers etc. based on existing equipment that could potentially be upgraded by firmware/software means;
- Etc.

The advantages of maintaining compatibility with existing devices include:

- Rapid development and deployment of “new” aggregation-capable devices based on existing equipment & silicon;
- Potential to upgrade existing client equipment by replacing drivers, firmware, etc.

In short, if this kind of backward capability is maintained, there is a good basis for rapidly deploying aggregation-capable devices. Conversely, if we choose a mechanism that cannot be used with existing technology, not only will it be necessary for vendors to immediately upgrade their now obsolete technology, but also there will be a relatively lengthy period (1 year? 2 years? 3 years?) during which there will be inbuilt resistance to deployment of the technology due to the existence of non-upgradeable devices on client sites.

The primary requirement for such upgrading to be possible is that, whatever frame format is used, it must be possible for the equipment concerned to both *recognise* the protocol and *act upon it*.

Unfortunately, in the case of MAC control frames, we have a mechanism that has potential problems with existing hardware; particularly if that hardware has been designed to conform with the letter of the 802.1x standard. The MAC control mechanism makes use of a Type value, followed by an opcode; the latter defines the type of MAC control function. Currently, one such opcode is defined, to be used by the 802.3x flow control PAUSE operation. Although one of the remaining, reserved, opcode values could be allocated to be used by LACP, 802.3x clearly states that MAC control frames that carry unsupported opcode values are discarded, and are neither passed to the MAC Client, nor interpreted, nor acted upon by the MAC Control sublayer (see 31.5.1 for example.) This means that any truly 802.3x-compliant silicon in which flow control has been implemented as an embedded function (as opposed to handled by software) is highly likely to have been implemented such that all MAC control frames that do not carry the PAUSE opcode are discarded and can not be made available for interpretation by software.

This leads to the following conclusions about the use (or potential use) of the unused opcode values in MAC Control frames:

1. If we choose to use MAC control frames to carry LACPDUs, then we will be discarding the potential for retrofitting Link Aggregation into some existing devices;
2. If we wish to avoid this kind of issue with potential future use of these opcodes for other protocols, then we had better fix the 802.3x specification. The fix required would be something along the lines of requiring that all MAC Control frames with unknown opcodes be passed to the MAC Client. However, it would still be some significant length of time before we could be reasonably sure that the majority of 802.3 hardware in the marketplace conforms to the fix;
3. If nothing changes in the 802.3x spec, the only cases where it will be straightforward to use any of the unused MAC Control opcode space is in cases where compatibility with existing hardware is not an issue; for example, on some as yet unknown LAN medium where there will never be a requirement to interwork on that medium with existing 802.3 devices, and where it is reasonable to assume that deploying the new technology will involve the development of new silicon.

From the above, the conclusion is that the use of MAC control frames is not a desirable mechanism for use in LACP, and that a new multicast address and Ethernet Type value should be allocated for use by the protocol.

## 3. Protocol extensibility

The above discussion gives a simple example of the kind of problems that can occur when designing (or attempting to design) protocols that are extensible. It is therefore worth considering at an early stage in this design effort:

1. To what extent we want to (or are able to) design LACP to allow for future extension; and
2. How best to achieve the extension capability if we decide that it is desirable.

The fundamental problems with extending a protocol are:

- Making sure that devices that implement earlier versions of the protocol don't do the wrong thing with information aimed at later versions;
- Making sure that later versions of the protocol can interoperate with earlier versions.

Clearly, doing a perfect job of this requires that you already know exactly what the extensions will be when you design the initial version. However, there are a small number of principles that may help to make any extension to the protocol more possible:

- Do not discard PDUs that are too long, or that use fields, flag values, etc. that are "reserved" as far as you are aware. Simply interpret the bits you know about in the way you know how.
- Where possible, leave "hooks" that will make it easy, later on, to hand such stuff over to be processed by something that has a better idea as to what should be done with it.
- In the case of LACP, the best you can do may be to record all reserved field values (including stuff that appears to be outside the legal size of an LACPDU) and to reflect them back, unchanged, in PDUs you transmit.