

# END-TO-END ARGUMENTS IN SYSTEM DESIGN

**J.H. Saltzer, D.P. Reed and D.D. Clark\***

**M.I.T. Laboratory for Computer Science**

This paper presents a design principle that helps guide placement of functions among the modules of a distributed computer system. The principle, called the end-to-end argument, suggests that functions placed at low levels of a system may be redundant or of little value when compared with the cost of providing them at that low level. Examples discussed in the paper include bit error recovery, security using encryption, duplicate message suppression, recovery from system crashes, and delivery acknowledgement. Low level mechanisms to support these functions are justified only as performance enhancements.

## Introduction

Choosing the proper boundaries between functions is perhaps the primary activity of the computer system designer. Design principles that provide guidance in this choice of function placement are among the most important tools of a system designer. This paper discusses one class of function placement argument that has been used for many years with neither explicit recognition nor much conviction. However, the emergence of the data communication network as a computer system component has sharpened this line of function placement argument by making more apparent the situations in which and reasons why it applies. This paper articulates the argument explicitly, so as to examine its nature and to see how general it really is. The argument appeals to application requirements, and provides a rationale for moving function upward in a layered system, closer to the application that uses the function. We begin by considering the communication network version of the argument.

In a system that includes communications, one usually draws a modular boundary around the communication subsystem and defines a firm interface between it and the rest of the system. When doing so, it becomes apparent that there is a list of functions each of which might be implemented in any of several ways: by the communication subsystem, by its client, as a joint

---

\* Authors' addresses: J.H. Saltzer and D.D. Clark, M.I.T. Laboratory for Computer Science, 545 Technology Square, Cambridge, Massachusetts 02139.; D.P. Reed, Software Arts, Inc., 27 Mica Lane, Wellesley, Massachusetts 02181.

This research was supported in part by the Advanced Research Projects Agency of the U.S. Department of Defense and monitored by the Office of Naval Research under contract number N00014-75-C-0661.

Revised version of a paper from the Second International Conference on Distributed Computing Systems, Paris, France, April 8-10, 1981, pp. 509-512.; Copyright 1981 by The Institute of Electrical and Electronics Engineers, Inc. Reprinted with permission.

Published in ACM Transactions in Computer Systems 2, 4, November, 1984, pages 277-288.

Reprinted in Craig Partridge, editor Innovations in internetworking. Artech House, Norwood, MA, 1988, pages 195-206. ISBN 0-89006-337-0. Also scheduled to be reprinted in Amit Bhargava, editor. Integrated broadband networks. Artech House, Boston, 1991. ISBN 0-89006-483-0.

Scribe/FinalWord source: <http://web.mit.edu/Saltzer/www/publications/>