M.I.T. Laboratory for Computer Science April 15, 1981

Computer Systems Groups Request for Comments No. 205

Trip Report: INRIA Workshop on Distributed Systems
Second International Conference on Distributed Systems
Versailles, April 6-10, 1981

by J. H. Saltzer and D. P. Reed

1. A two-day workshop, sponsored by the Institute National de Recherche en Informatique et en Automatique (INRIA), and chaired by Liba Svobodova, attracted about 50 participants, mostly from Europe. The sessions were notable for a large number of reports describing actual working systems and experimental projects, with little emphasis on theory or performance analysis. For the American participants who had attended the Pala Mesa distributed systems workshop last November, this meeting seemed like a continuation of that one.

2. Mike Schroeder reported operational experience with Grapevine, a distributed mail forwarding and name registration service with several servers now running. Debugging and related operational issues seem to be occupying much of their attention. A paper is being readied for the 8th Symposium on Operating System Principles. This system appears to be worth studying.

3. The closing session of this workshop was devoted to listing all the properties of distributed systems that make them different and hard to deal with. Liba Svobodova's record of this list follows:
   1. partial failure
   2. lack of a universal frame of reference
   3. you can't single-step it to debug, etc.
   4. physical dispersion: remote button pushing
   5. heterogeneity: instruction set, operating system, but also difference of scale
   6. attempts to hide the distribution, e.g., interface crossing much more expensive
   7. performance cost of message sending
   8. resources used to manage the distributed nature

9. latency range
10. representation of distributed applications
11. can't capture usual methods of modularity
12. the parallelism is real and every one can see it
13. lack of notation to deal with parallelism
14. semantics of programming environment
15. higher expectation of service integration possibility
16. failure and concurrency--hard to do both at once
17. problem of a timeout

David Reed then summarized the list with two general properties:

- need for a better representation for people to describe the system
- need for a better representation for the system to describe itself.

4. The Second International Conference on Distributed Systems was, unfortunately, quite weak. Although only one out of every three papers submitted was accepted, the average quality level was still low, and the majority of papers simply described some project without giving much insight. One might conclude that there are too many, too frequent conferences and symposia on this subject, and not enough good quality work going on to fill them.

5. Jerry Saltzer visited the research laboratory of Thomson-CSF, the largest French electronics company, and elicited much interest in M.I.T./L.C.S. activities (which were almost unknown there previously) and proposals for further contact. One group there is trying to develop a microprocessor architecture for company-wide use, and another group has developed a standard backplane bus for use with the Motorola 68000. That group was very interested in the existence of the M.I.T. CLU compiler for the M68000 since they have a commitment to object-oriented system design for the future. The organization seemed quite interested in Swallow, with its append-only storage, because Thomson-CSF seems to be responsible for French optical disk products. They have under development a variety of products, ranging from consumer-level read only analog disks (for TV) to a read-write digital disk for computer storage. That latter product is slated to hold 10" bits per platter, have access time of a few dozen milliseconds (through use of galvanometer mirrors) and cost in the vicinity of $10000. It is scheduled for volume production in the last quarter of this year.

6. Examples of the hardware of Polynet, the Logica, Ltd., commercial version of Cambridge ring network, were on display at a small exhibition that accompanied the conference. Shown were a PDP-11 Unibus DMA interface, an LSI-11 Q-bus programmed I/O port, a multibus interface, the ring repeater itself, and the control/monitor station. Prices were not mentioned (I forgot to ask) but complexity of the hardware appeared comparable to the V.2 ring. They are awaiting arrival of an LSI chip that performs most of the ring function. This ring has two disadvantages when compared with ours:
   - effective data rate between two stations of under 1 Mbit/sec. despite 10 Mbit/sec signalling on two twisted pairs.
   - need for a central monitor station

   The largest ring so far assembled operates with about 30 nodes; so far all experience with the ring has been very satisfactory. The British Science Research Council has ordered 10 copies of the ring for installation in British university laboratories.

7. Victor Lesser of the University of Massachusetts at Amherst reported an experience with the "Hearsay" speech understanding system that illustrates the end-to-end argument with some force. Curious about the cost of synchronizing and coordinating primitives in the multiprocessor implementation of Hearsay, they tried a version of the system with all synchronization removed--processes could use one another's results without setting locks. Apparently the time performance improved somewhat and the quality of the result did not change. Higher-level integration of partial results and hypotheses concerning the sounds being recognized is done expecting inconsistency. Therefore, having a little extra noise in the partial results because of mis-synchronization doesn't really make much difference.

8. Another end-to-end argument was unexpectedly provided by the Grapevine system, which tolerates slow propagation of updates and transient inconsistency. The application is fairly tolerant of minor inconsistency, and overnight database comparisons are adequate to discover and repair inconsistencies. There seems to be no need for a "hard" atomic action mechanism to support this system.

9. Olivetti telecommunications has built a 10Mbit/sec Ethernet to the Xerox-Intel-Dec specification. Future intentions are not clear.

10. David Reed visited CII-Honeywell Bull at Louvciennes (outside Paris), for I.L.P. The principal persons visited were Jean Rohmer and Xavier Roban. Most of the afternoon was spent describing the Swallow distributed data storage system being built by the CSS group under Reed. Roban has developed a distributed transaction system on the Honeywell level 64 computer which uses some of the same ideas (particularly using multiple object versions) that are used in Swallow. We also discussed the way that Swallow uses optical disks, as append-only storage. Since Thomson-CSF is developing such disks to be sold in late 1982, Honeywell is interested in developing ways to use them.

11. While at INRIA, we saw a demonstration of the KAYAK system being developed for office automation applications. KAYAK consists of a workstation similar to the Xerox STAR or Alto, and an Ethernet-like network. The workstation has an 8086, up to 1 Mbyte main memory, a high-density (1000 x 1000) bitmap display, the net, a mouse, and audio input/output. The network transports at 1 Mbit/sec. All of the hardware was working. We saw network test software and picture drawing software demonstrated. Programming is done in Pascal (a French standard).

12. Copies of the proceedings of the conference and a short activity description of each of the workshop participants can be borrowed from either of us.