HEWLETT PACKARD
SYSTEM PERFORMANCE CENTER 11 April 1986

"Making Distributed Systems Work" position statement

John Wilkes

The Advanced Commercial System project at Hewlett Packard Laboratories is engaged in a mixture of applied research and advanced development in the area of large, loosely-coupled, distributed systems. Our goal is to produce a prototype of a distributed system for the commercial data processing marketplace of the 1990s. To that end, we are aggressively pursuing a number of areas of work:

1. *Distributed system structure.* Each node has (or is able to load) a local instance of an interface to any system resource; we think of the resource as being represented by a "distributed resource manager" comprised of all its instances. Individual resource implementers decide for themselves whether (and which of) those instances are completely functional (e.g. for the distributed process management intrinsics), simply remote procedure call stubs (e.g. for printing services provided by another node), or some middle form, such as a storage system interface for a discless node that handles data caching and all the requisite consistency algorithms, but contains no device access code.

2. *A very wide range of scalability* in the numbers of nodes and delivered resources, of order 1 : 100. This is a considerable challenge to existing resource management and selection algorithms, and we expect this to be a significant area of contribution. In order to meet our performance goals, we conduct a form of load balancing we term *work sharing* (parcelling out process-sized portions of jobs to the best available processor on the fly, but not relocating them thereafter). We anticipate that these same mechanisms will be able to help automate the traditionally painful processes of system capacity planning and hardware configuration.

3. *The use of massively parallel computation*, by taking advantage of the inherent parallelism resulting from the large numbers of processor nodes. Some of this parallelism will be available through special efforts in particular subsystems (e.g. a distributed database), but much of it will have to be extracted from existing serially-structured applications. We plan work in compiler technology to apply some of the existing results in the numeric supercomputer area to the commercial environment.

4. *High system availability* in the face of component (e.g. node) failures. Part of this work is the simple provision of redundant systems; most of it is in taking advantage of their existence to construct algorithms robust to failures. Because of the relatively cheap availability of processing power in the systems we are planning, we expect to be able to expend reasonable runtime effort in providing this resiliency.

5. *Security.* We intend to make the resulting distributed system at least as secure (and private) as equivalent single-machine systems.

6. Some parallel efforts are taking place in hardware packaging and a very high speed fibre-optic interconnect to make all this feasible.

Items 3 and 4 above are at once both the most challenging and the ones of greatest potential reward. This is particularly so given our emphasis: to produce a level of robustness and functionality appropriate for a product release not too long after the prototype demonstration and analysis is complete. We think of this project as providing *very* large amount of computing power, in easily acquired increments,

while retaining (and in some cases improving on) the well understood benefits of the traditional single-machine timesharing model. As such, it must contain *all* the components and functions of a traditional single-machine system in the distributed form.

Our project is currently (early April) in the analysis and specification stage, with some work advancing into the areas of preliminary design and early prototyping. We are working in the medium of HP's newly announced commercial operating system (MPE/XL), running on the SPECTRUM program hardware, the first products of which perform in the 4.5–6.7 MIPS range. A considerable challenge here is the rather disturbing 'reality' of this system, which is roughly an order of magnitude larger (in most of the interesting dimensions) than the traditional distributed system testbeds (such as UNIX™). Much of our effort is directed towards the very immediate complexity of "making distributed systems work".

My own background includes working on and with the Cambridge Distributed System for my Ph.D., and since then conducting research in the area of distributed systems for workstation-sized machines at Hewlett Packard, as well as work in the areas of processor, I/O and disc architectures (for the SPECTRUM program). I was directly responsible (amongst other things) for initiating and bringing to fruition a project to produce an easily-retargetable, table-driven remote procedure call system, aimed at the heterogenous computing systems Hewlett Packard sells, and the even more mixed ones that it uses internally.

™ UNIX is a trademark of AT&T Bell Laboratories.