Issues relating to the low uptake of Networks of Workstations for commercial applications

Peter Dell

School of Information Systems

Curtin University of Technology, Australia

E-mail: dellp@cbs.curtin.edu.au

ABSTRACT

Networks of Workstations are a new approach to parallel computing which promise performance gains and more efficient use of resources, resulting in lower technology costs. Despite these benefits, there are anecdotal reports that the business community has been reluctant to use the technology and that the adoption rate of cluster computing for commercial applications remains extremely low.

These anecdotal reports are confirmed. Some concerns preventing the commercial usage of Networks of Workstations are revealed, however these concerns are contrary to the experiences of those organisations that have implemented the technology.

INTRODUCTION

New technology often takes many years before becoming widely accepted. The Internet is a case in point: it has existed in various forms for many years, however *en masse* acceptance did not occur until after 1993 (Downes and Mui 1998). Such acceptance only occurred after continuous development accommodated the needs of potential users. This illustrates the importance of identifying what is required from any new technology - it is essential to address the needs and concerns of those who must use it, otherwise it is useless.

Networks of Workstations (NOWs) are a new application of technology that is yet to receive widespread acceptance. The development of the Internet teaches that for this approach to be useful, the needs and concerns of users must be addressed. These issues must first be discovered, and that is the purpose of this research.

BACKGROUND

The quest for more is an inescapable characteristic of human nature, and computing power is certainly no exception. As Pfister (1995) observes, many people are willing to do much to have more powerful computers than anybody else, and so the cycle spirals ever upwards, as the competition for the fastest machine continues.

However, advances in sequential computing (single-CPU computing) are becoming less frequent as current engineering practices face some fundamental problems imposed on them by the laws of physics (Kuck 1997). This “speed squeeze” has resulted in a dilemma in which the continual increases in computer performance upon which we have come to rely will no longer be available (Kuck 1996).

Parallel computing is a technique that can help overcome the problems computer technology

is currently facing, and involves using multiple processors simultaneously to achieve increases in performance. Traditionally it has been confined mostly to powerful supercomputers and workstations, however it has been realised that parallel computing can deliver great benefits when applied at a microcomputer level.

Since the 1980s, supercomputers have been designed using an approach called Massively Parallel Processing (MPP), in which large numbers of commodity processors are operated simultaneously to achieve high levels of computational power.

One of the problems with MPPs is that they achieve a low level of cost effectiveness. While they are based on commodity microprocessors, the system design and implementation is based on the low-volume supercomputer market (Davis, Swanson and Parker 1997). This leads to the difference in cost between supercomputers and PCs, noted by Ridge, Becker, Merkey and Sterling (1997) as follows:

> "The PC market is two orders of magnitude larger than the workstation market, and the resulting economies of scale have allowed PC prices to decrease while sustaining dramatic performance increase."

Anderson, Culler and Patterson (1995) also refer to the difference between PCs and more powerful machines, stating that personal computers outshipped supercomputers by a ratio of 30,000:1 in the preceding five years. This trend of production of high-volume, low-powered machines over low-volume, high-powered machines has produced an economy of scale such that it is now more cost-effective to combine the processing power of a number of smaller, cheaper workstations than to invest in a single, high-powered supercomputer.

Anderson *et al.* (1995) have also identified reasons for the cost inefficiency in developing MPPs:

- The engineering time-lag between new processors being developed and then incorporated into an MPP design (MPPs typically lag behind chip development by one to two years);
- The high cost of engineering MPP systems;
- The higher cost of maintenance, in terms of cost per node;
- The high cost of operating system and application design.

All of these insights have been combined by researchers to produce a new approach to parallel computing in which multiple PCs are used in parallel, to achieve high levels of performance. This approach is called Networks of Workstations.

A NOW cannot be defined as a particular hardware configuration. Rather, it reflects the way networked computers are used. A Local Area Network (LAN) is not in itself a cluster, but a cluster can be implemented on a LAN. Similarly, a cluster is not a NOW, but a NOW can be implemented on a cluster. In a sense, it can be said that a NOW is in the eye of the beholder.

Each node in a NOW consists of a separate computer, typically a PC or low-end workstation. The philosophy is to use multiple, cheap, mass-produced parts instead of more expensive traditional alternatives. This approach has primarily been made feasible by recent advances in network hardware, in that high-speed networking technology is now commercially

available at relatively low prices. In particular, Asynchronous Transfer Mode (ATM) technology is being explored as a useful networking environment that can allow the approach to be implemented effectively (Blair 1996).

So far, the NOW research community has headed in two main directions:

- Systems that involve taking advantage of the under-used computing resources in an organisation to achieve performance increases. This basically involves the "scavenging" of resources such as processor cycles from unused workstations (Ridge *et al*. 1997).

- PoPCs (Pile-of-PCs) can be used to achieve supercomputer performance at a fraction of the cost by using a pile of cheaper PCs and performing calculations in parallel. The NASA-sponsored Beowulf project is one such example of the PoPC approach.

The first of these approaches involves using the idle processor time of client PCs. One way this can be achieved is through the use of a special-purpose screensaver that performs cluster tasks when the PC would otherwise be idle. The second approach involves a dedicated collection of servers in a cluster.

NOWs have the potential to be able to deliver the benefits of high-performance parallel computing without suffering from these drawbacks, by building systems out of commodity machines, rather than just commodity components.

A second motivation for the use of NOWs is that most large organisations already have in place a large infrastructure of PCs which are idle for much of the time. This is supported by Piotrowski and Dandamudi (1997), who observe that "in practice, up to 80% of workstations are idle depending on the time of day". A similar estimate given by Anderson *et al*. (1995) is that "even during the daytime hours, more than 60% of workstations are available 100% of the time". This contradicts the commonly held assumption that unused capacity exists only during out-of-hours periods. This portion of idle time represents wasted computing resources, which can be harnessed by the organisation as an alternative to investing in additional technology.

Other advantages of NOWs are also present. For example, no single vendor owns the rights to any product used in a NOW. Because systems are composed of widely available commodity components such as PCs, there are many suppliers from which to choose. This increases the ability to tailor a NOW to requirements, as well as not being locked in to a particular vendor (who might raise prices, or dissolve as a company).

Despite these benefits, however, anecdotal evidence suggests that usage of NOWs for commercial applications remains low, while empirical data are scarce. A first step in furthering the development of this technology is obtaining a clearer picture of the issues relating to their use in commercial organisations.

Future Applications of Networks of Workstations

The potential applications of high-powered computers are many and varied, and any attempt to catalogue them all would be futile; this is not the intention of the author. Following are some possible uses of the technology, as suggested by other authors.

Research into the application of NOWs for enhancing database performance is being conducted by an increasing number of researchers. According to Dandamudi (1997):

> "There are several application areas that require very high transaction processing rates. For example, transactions processing rates of stock market and banking industry are very high. At the same time the size of databases is increasing."

Other areas cited by Fox, Williams and Messina (1994) where use of NOWs could be beneficial include:

- Environmental modeling, for example pollution models could be used to better enable environmental agencies to deal with pollution problems;
- Production of animated films;
- Fluid flow simulations, used in many engineering applications.

Additionally, Brewer (1996) suggests the use of NOWs for applications such as the provision of Internet services. Pfister (1995) also identifies this in his (by no means exhaustive) list of users who are able to benefit from this technology:

- Media companies, to provide interactive download of movies etc.;
- Scientists, to verify or disprove things that were previously thought too compute-expensive to test;
- Engineers, in simulation of models and designs;
- Retailers, in the examination of sales models to improve marketing;
- Financiers, in performing financial projections.

RESEARCH PLAN

This research is focused on two key areas, investigated in two phases:

- Confirmation of anecdotal evidence that NOWs are not widely used in commercial applications;

- Discovering reasons why commercial organisations are reluctant to use NOWs.

Phase 1

The purpose of this phase is to confirm anecdotal evidence that NOWs are not widely used in commercial applications. Specifically, this phase addresses the following research question:

> *"What proportion of Networks of Workstations are used for applications that are commercial in nature?"*

To this end, a brief survey of users of the technology was conducted, in which an invitation to complete a questionnaire on the WWW was sent to all the members of an e-mail distribution list for users of Beowulf, a particular NOW implementation. (Further information on both Beowulf and the mailing list in question can be obtained from www.beowulf.org).

The survey contained used both closed and open questioning techniques. Closed questions were used to gather descriptive data such as the number of nodes, hardware and software configurations, while open questions were used to obtain data on areas such as for what applications the NOW is used and other comments regarding its use.

The applications for which respondents' NOWs were used were classified into three main categories:

- Scientific applications

  This category included the wide range of scientific applications reported such as chemistry simulations, computational fluid dynamics, computational physics applications and so on.

- Non-scientific IT applications

  This category included generic technical operations that were not scientific in nature, such as software development and document retrieval.

- Commercial applications

  This category included applications that are commercial in nature, such as large database operations, data mining and so on.

Phase 2

Although NOWs have the potential to be applied to a range of commercial applications, they have not become widely used in business organisations. This section addresses the factors limiting commercial uptake of NOWs. For practical reasons it was decided to investigate the local business community, and thus the following research question was asked:

> *"What reasons are given by relevant people in the business community to explain why NOWs are not widely diffused throughout industry in general in Perth, Western Australia?"*

Examining this question in detail, it contains four key phrases:

- what reasons are given
- widely diffused
- industry in general
- in Perth

The first phrase, *what reasons are given*, suggested that some form of research which involved soliciting the opinions of others was required. Such methods could have been survey or interview based, or possibly a case study.

The second phrase, *widely diffused*, indicated that some form of wide-ranging research method was appropriate; this signalled that a case study would not have been advisable. Interviews could have been employed, however the number required would have resulted in a time-frame beyond the scope of this research. Thus, the appropriate methods were narrowed down to survey research.

This leads to the third and fourth phrases observed above: *industry in general* and *in Perth*. These phrases identified the target which was to be sampled.

Phase 2 Questionnaire Design

The research topic is a new area of research. This fact, and the nature of the research question, meant that a suitable questionnaire to be either adapted or used verbatim had not been previously developed. Therefore, it was necessary to create a questionnaire specifically for the purpose of this research. The questionnaire can be viewed on the WWW at www.dssrg.curtin.edu.au/~dellpt/appendix1.html.

The target demographic was composed of 61 public sector organisations in Western Australia, ranging from large government ministries to small technical colleges. An additional 36 Western Australian commercial enterprises, all within the top 500 Australian companies, were also included. Three different staff members from each organisation were surveyed, representing IS/IT Management, IT Support and Software Development staff.

Thus, it was felt that a wide variety of people were represented, ranging from those in charge of wide-ranging, highly sophisticated IS/IT organisations, to support staff in small, organisations. This was important to maximise generalisability of results.

The purpose of the questionnaire was two-fold. The first purpose was to gain some raw data on how much is known about NOWs and how much they are actually used in the business community. The second purpose was to determine issues relevant to their use, as perceived by those in the business community. It is felt by the author that for the technology to provide any significant benefit, it has to achieve relevancy with the majority of those with the potential to use it. This implies a need to determine the issues by which those potential users will judge the technology.

Because of the two-fold purpose of the questionnaire, two types of questions were posed. The first related to simple statements of fact, and involved closed questions with simple Yes/No answers. The second, open-ended type of question involved written answers and invited respondents to express their own views on the topic without being limited by category-based answers. This was considered to be important as it was not possible to foresee (and thus provide categories for) the range of responses which might be given.

This tactic was repeated three times in three separate sections of the questionnaire, allowing the areas of IS/IT Management, IS/IT Support, and Software Development to be targeted separately.

RESEARCH RESULTS

This section presents the results from the two research phases.

Phase 1

Results in Table 1 clearly reinforce anecdotal reports that NOWs are not used in commercial applications. 87% of responses reported a wide range of scientific applications, while only 4% reported commercial or potentially commercial uses.

Table 1
Usage rates of Networks of Workstations for different applications

Application type	Usage level
Scientific	87%
Non-scientific IT	7%
Commercial	4%
Other	6%

Results total more than 100% due to some responses reporting multiple applications

However, there is no reason why the technology cannot be applied in a business setting. Factors explaining the reluctance of commercial organisations to use NOWs were investigated in the next phase.

Phase 2

The two principle variables used for this analysis were Staff Level of respondent, and Issues Raised by the respondent. While Staff Level was precoded, it was necessary to postcode Issues Raised as it was not possible to predict the range of responses, due to the open nature of the questions. The resulting codes for Issues Raised were either positive (benefits likely to be gained from the technology) or negative (problems likely to occur from use of the technology).

Positive issues largely revolved around three factors:

- Increased Processing Power
- More Efficient use of Resources
- Cost Advantages

These closely reflect the benefits that can be derived from the use of NOWs. Negative concerns were grouped into five categories, shown in Table 2.

Table 2
Negative concerns regarding Networks of Workstations

Category	Description
Administration	Those issues affecting administrative tasks within the organisation. For example, concerns involving system and network administration, support of systems, staff levels, knowledge requirements, and so forth.
Network Problems	Issues relating directly to the performance of relevant computer networks. For example, effects on network speed and availability.
Data Concerns	Issues relating to the organisation's data, such as security of information, and so on
Technology Problems	Concerns regarding the capabilities of the technology, such as concern about the unproven nature of the technology
Development Problems	The affect on the software development process
Miscellaneous	Any other negative concern.

Most respondents focused on problems which may occur as a result of using NOWs, rather than on the benefits that might be obtained. It is the author's opinion that this may be a

reflection of the risk-averse nature of most businesses. The two areas that received the most concern were Administration and Technology Problems. Table 3 shows the proportion of responses citing each area of concern.

Table 3
Proportion of concerns raised by survey respondents

Area of Concern	Proportion of responses
Administration	56%
Technology Problems	50%
Network Problems	36%
Data Concerns	14%
Development Problems	7%
Miscellaneous	21%

Results total more than 100% due to some responses reporting multiple applications

It can be seen that over half of the respondents suggested administration related problems as a concern. Typical of many of these responses was a view that the increase in cost due to a more complex administration task would outweigh the cost savings from using a NOW.

A lack of faith in NOW technology was also identified, with 50% of respondents indicating the sentiment summed up by one respondent: “the business case needs to be made clear”.

Effects on the network, such as those related to congestion, were also cited as a concern by just over one third (36%) of respondents. This is perhaps somewhat unfounded, depending on the implementation environment of the NOW. Almost certainly measures could be taken to avoid this category of problem.

Very little concern about other areas was shown by respondents. The three areas already discussed were the most prominent. The responses also showed IS/IT Management staff had more concerns about the technology than IT Support and Software Development staff, as can be seen in Table 4.

Table 4
Breakdown by staff level of concerns raised

Area of Concern	Staff Level		
	IS/IT Management	IT Support	Software Development
Administration	60%	67%	43%
Network Problems	40%	44%	29%
Data Concerns	20%	11%	0%
Technology Problems	60%	44%	43%
Development Problems	40%	0%	29%
Miscellaneous	0%	0%	14%

Apart from the notable exception of development problems, every category was raised as a concern by more IS/IT Management staff than by the other staff types. This can be interpreted in at least two ways:

- The first is that management staff are more pessimistic about the technology than those more closely related with technical aspects.

- The second is that management staff typically have more of a "big picture" view of things, and consider issues in a wider range of contexts than other staff members.

It is not the intention of the author to debate which of these, or other possible explanations, is the most likely, but merely to note the difference in responses from different levels of staff.

DISCUSSION

It is likely that NOWs can deliver significant benefit to commercial organisations in a number of ways. Research into the application of NOWs for enhancing database performance is being conducted by a number of researchers. According to Dandamudi (1997):

> "There are several application areas that require very high transaction processing rates. For example, transactions processing rates of stock market and banking industry are very high. At the same time the size of databases is increasing."

Pfister (1995) also identifies various commercial applications for which the technology might be applied, including retailers, in the examination of sales models to improve marketing, financiers, in performing financial projections, and media companies, to provide interactive download of movies and so forth.

However, this research has confirmed anecdotal evidence that usage of NOWs is not widespread within commercial organisations. In fact, nearly all NOWs in use are being used for scientific applications.

The two principle issues raised by the business community to explain their reluctance to use the technology are that NOWs will result in an increase in administrative workload, and that the technology is unlikely to deliver the benefits it claims to.

Both of these fears are contradicted in comments from respondents in Phase 1. Comments such as the following were typical:

- "Easy from user point-of-view, no change required to code"
- "Easy to maintain"
- "System has been exceptionally stable and has proven to be very cost effective."
- "Near linear speed-up"
- "Linear improvement seen for program run on cluster"
- "Greater processing performance, ease of use"
- "Best price/performance. Scalability."
- "Speed and cost efficiency"
- "Great price/performance ratio"
- "Cheap computational power"
- "Supercomputer performance at $1/10^{th}$ the cost"
- "Met performance with minimal hardware cost"
- "Much more computational power than would be available from a traditional commercial "big iron" vendor".

The last comment listed suggests that clusters not only compete with supercomputers, but provide computational power that actually exceeds traditional supercomputers. This is something that could be investigated in future research.

Implications

A list of "a number of strategically important forces [that] affect the pace and effectiveness of progress in using IS/IT" observed by Ward and Griffiths (1996) includes:

- the capabilities of the technology
- the economics of the technology
- the applications that are feasible
- the skills and abilities available to develop and use the applications

NOWs have the capability to increase available computer power with proportionally minor increases in cost. Thus, at least the first two factors in the above list are relevant. It follows that the technology can have a significant impact on the "pace and effectiveness of progress in using IS/IT".

CONCLUSION

This research has presented quantifiable evidence that the business community has indeed been reluctant to adopt NOWs for applications requiring high performance computer technology.

Several issues of concern have been identified, including the perception that NOWs lack support from commercial vendors, doubts about the capability of the technology, and a belief that using NOWs will impose an burden to administrative and development tasks.

More needs to be done to allay these fears for NOWs to deliver their potential benefits. Further investigation is required into how these fears can be addressed, so that the technology's benefits can be realised.

REFERENCES

Anderson, T., Culler, D. & Patterson, D. (1995) A Case For NOW (Networks of Workstations), *IEEE Micro*, Vol. 15, No 2, pp. 54-64.

Blair, G. S. (1996) A Convergence of Parallel and Distributed Computing?, *Abstract Machine Models for Parallel and Distributed Computing*, IOS Press, Amsterdam, Netherlands, pp 1-11.

Brewer, E. (1996) The Inktomi Experience, Invited Talk at the *Second Workshop on Networks of Workstations*, Cambridge USA. Cited in: Culler, D. E., Arpaci-Dusseau, A., Arpaci-Dusseau, R., Chun, B., Lumetta, S., Mainwaring, A., Martin, R., Yoshikawa, C. & Wong, F. (1997) *Parallel Computing on the Berkeley NOW*, [http://now.cs.berkeley.edu/Papers2/Postscript/jpps.ps].

Dandamudi, Sivarama P. (1997) Using Networks of Workstations for Database Query Operations, In: *Proc. Int. Conf. Computers and Their Applications*, Tempe, Arizona, USA, pp 100-105.

Davis, A., Swanson, M. & Parker, M. (1997) Efficient Communication Mechanisms for Cluster Based Parallel Computing, *Communication and architectural support for network*

based parallel computing: first international workshop; Proceedings / CANPC '97, Springer-Verlag, Berlin, Germany, pp 1-15.

Downes, Larry & Mui, Chunka (1998) *Unleashing the Killer App*, Harvard Business School Press, Boston, Massachusetts, USA.

Fox, G. C., Williams, R. D., & Messina, P. C. (1994) *Parallel Computing Works*, http://www.npac.syr.edu/copywrite/pcw/.

Kuck, D. J. (1996) *High Performance Computing: Challenges for Future Systems*, Oxford University Press, New York, USA.

Kuck, D. J. (1997) Facing Up to Software's Greatest Challenge: Practical Parallel Processing, *Computers in Physics*, Vol. 11, No. 3.

Pfister, Gregory F. (1995) *In Search of Clusters: The Coming Battle in Lowly Parallel Computing*, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.

Piotrowski, A. and Dandamudi, Sivarama P. (1997) A Comparative Study of Load Sharing on Networks of Workstations, *Int. Conf. Parallel and Distributed Computing Systems*, New Orleans, USA, pp 458-465.

Ridge, D., Becker, D., Merkey, P. & Sterling, T. (1997) Beowulf: Harnessing the Power of Parallelism in a Pile-of-PCs, *Proceedings, IEEE Aerospace*.

Ward, John & Griffiths, Pat (1996) *Strategic Planning for Information Systems* (2nd edn), John Wiley and Sons, Chichester, England.