Choosing the Right Platform for SAS® Client/ Server Applications

An evaluation of factors in selecting NT and UNIX

Gregory S. Barnes Nelson
ASG, Inc.

Richard Swirski
ASG, Inc.

Introduction

Early adopters of "rightsizing" strategies in the 1990's began a swing of the pendulum toward distributed systems such as UNIX open systems and Microsoft NT as a viable alternatives to mainframe and large batch-processing platforms for handling departmental and workgroup analytical processing. The scope of this paper is not to argue the benefits of the mainframe over alternative server technologies. Rather, the focus here is to help decision-makers to understand the factors involved in choosing between NT and UNIX for running SAS applications.

Specifically, we will talk about the different ways in which SAS can be "used" in these environments. Then the discussion will move to the general topic of the hardware configurations that are important in choosing a scalable platform for deploying your application.

This paper first addresses some general hardware information and how SAS can be "used" in both the Microsoft NT and UNIX environments. We do this because there are often misconceptions about how SAS can be installed and utilized in a server environment. Secondly, we talk about the more general factors that can have an impact on how successful one platform can be over the other.

Overview of NT and UNIX platforms

Microsoft NT and various flavors of UNIX have quickly become the defacto standard for small, mid-size and even large SAS installations. As the price/ performance ratio has made it affordable, yet powerful to run terabyte-size data warehouses in these environments.

Early versions of NT were often seen as clumsy and slow but offered simple workgroup solutions. These were typically not used for high volume or mission-critical environments. Although UNIX has been around for about 25 years, early versions were not mainstream and used mostly in the technical market. In the past 10 years, UNIX servers have worked their way into corporate mainstream.

UNIX, with it's long history in scientific and academic computing environments and NT with it's adoration from the IT communities for ease of administration, the lines have become blurred as to the obvious choice for deploying SAS applications across the enterprise.

Operating Systems

In determining which factors are most important for you, it may be helpful to first talk about which environment SAS can operate. Although Microsoft NT runs on three different RISC architectures--

- the Mips Rx000 (NT no longer supported)
- the Digital Alpha
- the IBM/Motorola PowerPC--

the version that runs on x86-based systems is the most popular and consequently the only platform that SAS runs on (6.12 production). The SAS System does not currently support DEC's Alpha on NT, but support is planned in Version 7, Wave 2 of the SAS System.

The SAS System is currently supported in the following *mainstream* UNIX hardware/ operating system configurations:

- Digital UNIX
- Hewlett-Packard PA-RISC series under HP/UX
- Sun Microsystems and SPARC-compliant systems under Solaris
- IBM RS6000 and PowerPC systems under AIX

There are also several *non-standard* Unix's that SAS will run on, including:

- Data General Aviion Systems under DG-UX
- Convex Computer C-Series under ConvexOS
- MIPS ABI-compliant systems
- INTEL ABI-compliant systems
- Novell UNIXware-compliant systems

Hardware

When selecting a hardware environment, there are a variety of issues that quickly come to mind: Does the machine support multiple processors? How much disk space can I add to the machine? How much memory can I put in there? Is it a 64-bit implementation?

As Microsoft NT matures and UNIX continues to become stronger, many of the issues also become blurred. In fact, they tend to leapfrog one another adding functionality well beyond the conceivable use for anything but exceptional SAS environments. This is not to say that SMP support, 64-bit architectures, in excess of 128GB of memory, support for 4 ectabytes of storage aren't important. Rather, both NT and UNIX either currently support them or will in the very near future.

Symmetric Processor (SMP)

A commonly asked question regarding platform selection is whether a given hardware configuration will support parallelization. SAS is a single-threaded application. This means that the operating system will shift execution of individual sessions between the available processor(s). Single SAS sessions are not distributed across multiple processors.

Example: In a 4 Processor System, a single SAS instance will only run on 1 given processor. Conversely, running 10 SAS sessions on the same system will cause the different sessions to share the CPUs.

Theoretically, NT will support up to 32 processors; however, implementation of this proved that the benefits were not realized beyond 4. Currently, Intel does not support an eight-way architecture, choosing instead to support Microsoft's Wolf-pack clustering effort.

The bottom line here is that increasing the number of processors within a box is just part of the story. For example, there are some limitations in the Pentium Pro' input/output channel that is not true of RISC-based architectures, thus making these systems less suitable for handling large numbers of users or simultaneous processes.

SAS Client/ Server Model

There is no "typical" SAS installation that we can cheerfully compare our application against, pull some handy configuration chart out, and order our out-of-the-box SAS application server. In fact, rarely are we treated with such options in today's technology supermarket. The important factors in helping to determine the optimal solution may not be all that allusive either.

In the following pages, we make use of the following types of applications as classifications:

- Defined applications (such as CFO Vision, HR Vision, IT Service Vision)
- End-user query and analysis tools (Enterprise Reporter, designed AF applications, reporting interface applications, SQL query window)
- Power user applications (either defined such as Enterprise Miner or open access tools such as SAS Program Manager, SAS/Assist or SAS/Insight)

As applications crop up around using and accessing data around SAS' data, applications that use visual basic and other Microsoft end-user tools to access SAS data sets or MDDB data cubes will have interesting effects on these definitions. In addition, we are likely to see SAS being used as a tool for delivering data via web and internet-based requests that will present unique challenges to our selection of the "right" platforms. However, since we are using the term **server** in the *client/server* formula as really the back-end, behind-the-scenes number crunching, data storage facility, I don't suspect that the information presented here will change significantly.

Here, I use the term server as we define them in the SAS client/ server model. In both the NT and UNIX environments, you can run SAS in a variety of ways.

Methods of running SAS in server environments

NT

In the Microsoft NT environment, SAS can run on the following versions:

- Windows NT 4.0 Server and Workstation
- Windows NT 3.51 Server and Workstation

And can be run as "Fat Clients" such as:

- Using NT as a file server
- Launching SAS from the server
- Using SAS/Connect to remote submit jobs for processing on the server

Or "Thin Clients" such as:

- NT Emulators (Citrix WinFrame)

UNIX

In the UNIX environment, we don't really refer to a "fat client" as we typically either logon to the UNIX machine and run SAS (telnet, x-windows) or submit jobs from a "client" machine such as Windows 95 for processing on the UNIX machine (rsubmit).

How well will the application run

One of the most critical factors in selecting a platform for a client/ server SAS application is the basic question "how well will the application run in this environment?" Since we are concerning ourselves mostly with the server portion of the client/ server equation, the focus can be averted from the end-user interface issues like: x-windows vs. SAS/AF application vs. thin client.

Although discussions about these strategies are not trivial, they really fall into the realm of the application development strategy. Impacting server performance is what we are referring to here, we make the assumption that you've already established some guidelines for client/ server architecture whether that be thin client, or 2- and 3- tier configurations.

Sample Configurations

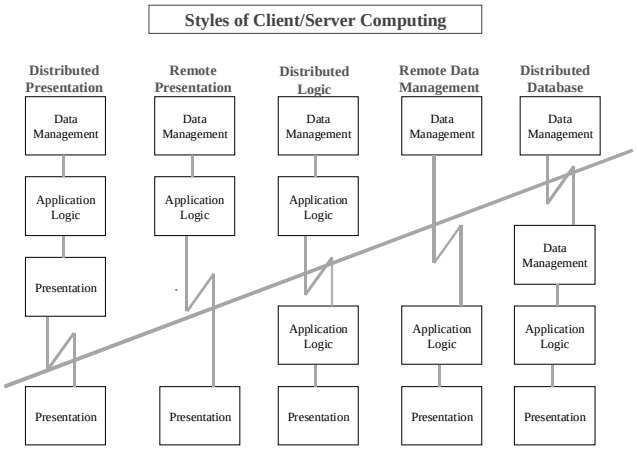
The SAS client/ server model assumes that we have some workstation, with local processing capabilities, attached to another platform usually a compute or data server. The client requests services, usually data, and the server retrieves resources and delivers the requesting information.

The client usually handles input and screen management. Servers have some built-in intelligence. Multiple clients are served by a single server.

The Gartner Group developed a segmented model of client/ server computing that divided the processing into three distinct areas:

- Data management
- Application logic
- Presentation

The data management component deals with the data repository, indexing and data aquisition. Application logic refers to applying the business rules, data consolidation and analysis. Finally, presentation is the component that has to do with capturing end-user input, displaying the reports, charts, tables and other output, providing the end-user drill down and other graphical user interfaces.



Source: Gartner Group

The following scenarios show some sample configurations of what we are talking about:

Client/Server Applications

Processing: both the client and the server

Data stored on server (or client or both)

The GUI resides on the client

Server setinit required if server processing occurs

SAS/CONNECT: communication between the client and the server

File Server

Processing occurs only on the client.

Data is stored on the server.

The SAS System is installed on the server or the client.

The GUI resides on the client.

A server setinit is not required.

NT: Clients must create "shares" to gain access to the data

Unix: Clients must use some communication method such as NFS (Network File System) or samba to gain access.

Data Server

Processing occurs only on the server.

Data is stored on the server.

The SAS System in installed on the server. SAS/SHARE*NET or the SAS Universal ODBC Driver is installed on the client machine.

The GUI is provided by a third party ODBC compliant application.

A Server setinit is required.

Terminal Emulation

Processing occurs only on the server.

Data is stored on the server.

The SAS System is installed on the server. The GUI resides on the server, but is displayed by a third party emulation package on the client.

NT: Thin clients such as WinFrame by Citrix Systems, Inc. (http://www.citrix.com) for Microsoft NT Server.

Unix: X-window servers such as Hummingbird Communications Ltd. Exceed (http://www.hummingbird.com) or common telnet applications for vt100 emulation.

A server setinit is required.

Tips on Client/ Server Configurations

Whenever possible, keep data traffic off the local area network as it is usually faster to access data from a disk local to the machine running the SAS system

Generally it is best to only send results and summarized data over your network. (Note: the File Server configuration does not follow these general rules)

Factors in Sizing the Machine

Although this paper certainly does not purport to be the definitive guide to platform selection, we hope that this will serve as a guide to people. In general,

- The amount and storage location of data,
- the number of users accessing the data,
- the type of processing anticipated,
- and the patterns of usage

can all have an impact on the environment. These can generally be broken down into these basic components:

- Memory: Physical and virtual
- CPU
- Disk and other I/O subsystems
- Network

To varying degrees these factors can influence how a SAS application will perform. The following section discusses how these components can have an impact on platform selection. We will discuss the following:

Criteria	*Description*
Cost of ownership	Not only what does the machine cost "out-of-the-box", but what factors influence the cost of ownership and how does the number of users, IT support, ad personnel costs impact overall cost of ownership?
Application interfaces	What other applications/ interfaces does the SAS application need to interact with as well as what other general applications are required? Does this have an impact on the platform you select?
Support for external data	Does the application need to be "close" to a data source outside of SAS such as Oracle, Sybase or OLE?
High availability	How critical is it that the machine is available 7 by 24 and automatic hot-swapping of components to take over machine failures?

Scalability	What provisions need to be considered when talking about the growth of the application?
Manageability	How does this machine and the SAS application fit into the overall information technology management framework? Is global administration necessary?
Security	How do you want to manage users access to resources?
Host features	Are there specific features of one operating environment that make it easier or are critical to your application such as electronic mail or fax support?

Cost of Ownership

Price/ performance

Price/performance is simply a breakdown of what you are getting for the money. Obviously, lower price, higher performance is always the goal.

Unfortunately, in a SAS environment, there is no clear measure or "benchmark". The idea of a "SASstones" benchmark was considered but never received vendor support. When sizing machines, some measures that can be used are TPC-C for OLTP applications and TPC-D for decision support systems. An independent breakdown of numerous systems can be found at http://www.tpc.org.

Each of these benchmarks has both a performance metric for determining the overall performance of the system and a price/performance metric for seeing the actual cost of the type of transaction being tested. These can be good tools for sizing/comparing systems.

Experience with the environment

Learning curves are often an unseen cost when implementing a new system. When considering a new environment, close decisions should be biased on what expertise is available in house as well as through vendor support.

Personnel Costs

Gartner Group:

- UNIX Administration: $60,000
- NT Administration: $55,000

Number of users and patterns of usage

Any reference in this paper to number of users is referring to users actually working on the system at a given time. When sizing the machine, plans have to be made to accommodate a "worst-case" scenario where the most users will be doing the most work. Often it Is possible to schedule large jobs to run "off-line" thus leaving the performance available for the actual on-line users. Or even "staggering" workload to fully utilize the system at all times.

Support

There is much to be said regarding having a single vendor solution. Sometimes it is advantageous to pick "best of breed" for all components of a system, however this can and has lead to "finger-pointing". All too often, support professionals have been heard to say "it's a hardware problem", while another claims "it's a software/configuration problem". Meanwhile, you are stuck in the middle without resolution. Having a single vendor can eliminate some of this.

NT

In it's vanilla form, NT is simple to setup.

Wealth of third party support for hardware and software.

All x86's are not created equal. Although there are numerous choices for add-on products, whether they are hardware or software, most products come "as-is" and claim to work on an x86 System. The selection and implementation often come at a large cost of man-hours, which are often overlooked in the cost.

Seldom used in its vanilla form. NT used in a true server environment has the same administrative challenges that exist with any other server.

UNIX

Systems usually come pre-loaded with O/S and majority of devices "configured".

Add-on hardware solutions are often limited by vendor offerings.

There is no longer a wealth of products to select from, however, the products to select from are those which have proven track records. Eliminates the "guessing game" of how/if this product will perform on a given server. Solutions typically work the first time due to the fact that there are so few variables

Seldom used "as-is" out of the box. More "variables" that can add to the overall performance but at the price of knowing which ones and how to "tweak" them.

Same administrative tasks as with any server environment

Application interfaces

The fact that you are reading this paper clearly illustrates that your plans are to use SAS. However, chances are, there are other business needs that may or may not be SAS related. Rather than have separate entities to solve every business problem, it is often advantageous to utilize what assets you have or are planning.

In addition, if there is data that you consistently receive from clients, vendors or partners a system to match the lowest common denominator offers a better solution

If a single system can handle all the needs, this is clearly better than disparate systems all doing their own job. This offers a "cleaner" environment rather than sending data from system to system.

Support for External Data

Although there are a variety of data sources available directly to SAS, there are many that are not available as direct interfaces/ APIs on certain platforms.

Interface methods available (SAS 6.12)

Database/ Platform	*NT*	*AIX*	*Digital UNIX*	*Solaris*	*HP-UX*
DB2		X			
Oracle	X	X	X	X	X
Ingres		X	X	X	X
Informix		X	X	X	X
Sybase/ SQL server	X	X	X	X	X
PC Files	X	X	X	X	X
ODBC	X				X
OLE	X				
DDE	X				

High availability

A basic definition of High Availability is service meets or exceeds the expectations of the users. High availability comes at a price and so it is important to consider what the application needs are. If the system MUST be up 7x24, some form of high availability must be built in. There becomes a certain amount of "weighing" what risks vs. cost you are willing to take in the implementation.

Considering a typical environment, there are several items which can lead to lost service:

CPU Failure

Disk Failure

Network Failure

Software Failure

Software and network failure are beyond the scope of the host operating system, however, CPU and disk issues can be addressed.

CPU Failure

In the event of a failed CPU, other CPU(s) in a cluster need to "absorb" the additional applications. NT offers "Wolf-pack" which currently implements a 2-node NT Cluster. Future releases are slated to support up to 16 nodes. Most Unix vendors have some type of "clustering" solution which go far beyond the 2 node limitation of NT. These solutions often offer "Load Balancing" in addition to CPU failover.

Disk Failure:

In the event of a failed disk(s), other disk(s) continue operation until the failed component can be replaced. Several types of RAID (Redundant Array of Inexpensive Disks) solutions are available for both Unix and NT. These solutions can be provided by the equipment manufacturer or via third party. Although software implementations of RAID are available on both platforms, large performance gains can be achieved by configuring RAID at the hardware level.

Scalability

How well does a platform support planned and unplanned growth. Give the users what they want, provide the performance they need and they will want more. Performance is a relative term so what seems fast today will not be tomorrow. When planning a system, it will be in your best interest to plan for expansion. Select a platform that has the ability to expand both in performance and capacity.

Manageability

Feature	*UNIX*	*WINDOWS NT*
Text and graphical management tools	Most	N
Remote administration and diagnostics	Y	Optional
Graphical volume	Optional	Y

management		

Global administration

Both NT and Unix will be "aware" of other servers on the network out of the box. By design, Unix has the advantage in that you can "connect" to that server (either in graphical or character mode), run jobs, monitor system activity, reboot and anything else you could do from the console. Add this with the "built-in" scripting ability and all administrative tasks can be accomplished without ever leaving your desk. NT does have some of this functionality and third-party software will allow you to connect to the server, however still does not match the tools standard to Unix.

Third party tools are available for both platforms to further add in managing the enterprise.

Reliability

Feature	*UNIX*	*WINDOWS NT*
Per-process memory protection	Y	Y
Recoverable file system	Y	Y
Remote diagnostics	Y	Optional

Integration

Feature	*UNIX*	*WINDOWS NT*
Standard application installation (network and local)	N	Y
Automatic detection of hardware	Some	Y
Multiple network protocols	Y	Y
Windows SMB file sharing	Optional	Y
Macintosh file sharing	Optional	Y
Unix NFS file sharing	Y	Optional
Vendor device driver support (PC)	Poor	Good

Security

Feature	*UNIX*	*WINDOWS NT*
User log-on required	Y	Y
File-level access permissions	Y	Y(1)
File-access control lists	Few	Y(2)
Security auditing	Most	Y
Role-based access	Few	Y

(1) Windows NT and Unix both offer read, write, and execute permissions on each file. NT adds "take ownership" and "change permission" to these. (2) Windows NT access-control lists apply not only to files but to all objects managed by the OS.

Host Features

Scheduling

Earlier, we spoke about scheduling large jobs to run off-hours. A host scheduler allows "batch" jobs to be run at certain intervals unattended. Systems typically have a low-usage time and by using a scheduler jobs that would normally run during "peak" hours can be batched to run at night. In addition schedulers can be used to implement accounting, system checks, backups, database refreshes etc.

NT

- AT command - Service must be running, Command line interface only
- Third Party Schedulers (e.g., Argent Software http://www.argent-nt.com)

UNIX

- AT command
- crontab (Chronological timer) Very flexible and can be *driven* from external programs such as SAS.
- Third Party Schedulers

Scripting

This is the ability for the OS to run a defined set of instructions. These can either be run interactively or via a scheduler.

NT

- Batch (CMD) files allow sequential execution of commands. In addition simple tests/logic operators can be added

Unix

- "Shell" scripts allow execution of commands and offer powerful programming functions such as if/then/else, case statements, functions, redirection and others. In fact, the majority of system startup/configuration files are written using "Bourne" Shell Scripts.

Both platforms support implementations of other "script-type" languages such as Perl which can be a great aid in system management.

Operating system advantages for SAS

There may be obvious choices for selecting one platform over another that have to do with the specific advantages a computing environment has over another as it relates to how it interfaces/ interacts with the SAS System. Here are just a few examples of some of the advantages of each:

UNIX

- Use of pipes and named pipes for interprocess communication
- Easy integration of UNIX environment variables into SAS macro variables (i.e., $USER, $HOME)
- Easy integration with the UNIX shell scripting environment.

NT

- Ability to communicate with other Microsoft Windows applications via OLE (ActiveX support in 6.12) and DDE
- DCOM support which allows for the invocation and coordination of objects across the network

Bottom Line

Both offer robust environments

Both offer levels of scalability

Both have graphical interfaces for administrative purposes. NT has the edge here because it is commonly known interface

Both can deliver text and graphical applications.

Both OSes give applications a protected virtual address space in which to run.

Both support multiple CPUs and lightweight processes.

Both run on a variety of platforms, although UNIX runs on far more.

Both support advanced file systems with long filenames.

Both offer powerful peer file sharing and other network services.

The case for UNIX

For medium to large instances, UNIX offers multi-user access built in, much greater support for processors, memory and disk as well as greater throughput for disk-intensive applications. A large "single-entity" or an even larger "multiple-entity"

"Mature" O/S (25 years)

- Established clustering platform
- Competition among vendors
- High Scalability
- Fault tolerance
- Large file systems
- Reliability
- Performance throughput
- Support for large number of sessions

The case for NT

For small to medium size instances (1-4 concurrent SAS users, up to 20 GB of data), NT offers a good solution.

Although NT can be setup in a distributed environment using many NT (Servers), this will add a significant cost of ownership both in administration as well as individual SAS Licenses.

- Application portability (closer to the client configuration)
- Desktop integration
- Single implementation (Intel vs. various flavors of UNIX)
- Easy to setup in vanilla form
- Inexpensive hardware
- Price/performance ratio
- Easier to maintain (locally)
- Homogeneous environment
- Lots of third party support

Today

NT - accepted as a "corporate solution"

Servers can offer up to 4-way symmetric multi-processors (SMP) and can be clustered into 2 server scenarios

Single user-multi-tasking environment

With third-party products, can be made multi-user

NT is winning more mind share among users. It's new, it's hot, it's from Microsoft, and it has "Windows" in its name. UNIX suffers from discrimination against old age and from disunity among vendors. People who cut their teeth on Windows are moving into authority and will increasingly look toward Microsoft for solutions.

Microsoft's slow embrace of the Internet has worked to the advantage of UNIX. But by bundling Internet software with NT Server and making its leading applications Internet-aware, Microsoft can establish NT as the no-brainer choice for companies that are setting up new Web sites and intranets.

UNIX - has been/is a "corporate solution"

Servers offer SMP well beyond 32 processors. Numerous servers can be clustered together to offer an even larger highly available solution.

Multi-user, multi-tasking environment

Proven track record for performance and mission-critical applications

UNIX is still the best solution for large databases and other enterprise-scale jobs. That won't change until Microsoft radically improves the scalability of NT on SMP machines with six or more processors. Also, NT has no timeserver, which may rule it out for large on-line transaction processing (OLTP) systems.

The future

1996 prediction: NT would capture 37 percent of the market. In 1995 NT had obtained only 1 percent.

64-Bit computing and introduction of "Merced" chip will advance both platforms.

NT 5.0 will only be able to efficiently use 8 processors

UNIX will still have the processor, throughput and scalability advantage

UNIX will NOT dramatically change in the next version thus eliminating any new "learning curves", while:

NT 4 new user interface and other internal changes

NT 5 new web-based user interface, introducing NDS and other changes

Both offer high speed, high availability storage sub-systems.

Conclusion

When selecting a corporate infrastructure, consider not just the pieces, but rather the final solution the pieces will create.

The goal is not just the most inexpensive out of the box, but the best performance, growth and most inexpensive for the next x years.

For large-scale situations, UNIX has often been the obvious choice. However, due to rising performance improvements, reliability and availability, NT now offers an alternative in the small to mid-range server market. At the same time, the UNIX arena is not stagnant. In order to remain competitive, UNIX servers HAVE to improve performance and add to their value.

In the end, there is no one-size-fits-all answer to which OS is better. Experts who want to craft the best possible solution for a given business problem must be knowledgeable enough and open-minded enough to adopt either OS--or both.

Server(s) selection should be based **primarily** on what platform will solve the application, **performance and scalability goals.**

Close decisions can be biased on "learning curves".

Neither platform is a "wrong" solution.

Acknowledgements

We would like to acknowledge the assistance of ASG, Inc., our employer, for providing us the time and resources for developing this paper.

Bibliography

Cates, M. W. (1997), "Taking Advantage of the SAS System on Windows NT". *Proceedings of the Twenty-second annual convention of the SAS Users Group International Conference.*

Clegg, J. & Riggsbee, C. (1997), "OLE and the SAS System for Windows Release 6.12", *Proceedings of the Twenty-second annual convention of the SAS Users Group International Conference.*

Mehler, G. (1997), "Integrating Windows Clients and the SAS System into the Enterprise". *Proceedings of the Twenty-second annual convention of the SAS Users Group International Conference.*

CTC (1997), "Can a SAS Session Take Advantage of Parallel Servers?" CTC Support Documentation from SAS Institute.

Crevar, M. et al. (1997), "What Size NT Server Do I Need" CTC Support Documentation from SAS Institute.

Garner, C. & Reinard, S. "The SAS System: A complete client/server solution"

Lutter, D. "Evolving SAS Environments: The right architecture for the right job"

Lyons, M. (1997), "HP-UX and Windows NT: Platform selection Guide for Servers in Mission Sensitive and Mission Critical Commercial Environements".

Arthur, T. (1996), "Experiences with Remote Library Services: Choosing the right platform", *Proceedings of the Twenty-first annual convention of the SAS Users Group International Conference.*

Address

Questions pertaining to this article should be addressed to:

Gregory S. Barnes Nelson
ASG, Inc.
2000 Regency Parkway, Suite 355
Cary, NC 27511
919-467-0505 (Voice)
919-467-1195 (Fax)
gbn@asg-inc.com

Richard Swirski
ASG, Inc.
2000 Regency Parkway, Suite 355
Cary, NC 27511
919-467-0505 (Voice)
919-467-1195 (Fax)
rs@asg-inc.com