



The Evolution of Information Assurance

The explosive growth in the amount of electronic information that organizations generate, and the increasing importance of that information, make protecting it with assurance processes a top priority.

Roger Cummings
Veritas Software

Modern business enterprises realize that information is their most important asset. Its importance and ever-increasing volume have led enterprises to focus their attention on the container for this precious resource, namely storage, and on ensuring the availability and integrity of their information.

The information assurance discipline integrates several techniques developed over many years for assuring the integrity and availability of stored information. At the highest level, information assurance consists of the following activities:

- creating multiple copies of information;
- migrating the copies between different storage types;
- managing the consistency of these copies, especially in light of the intermittent connectivity between them; and
- replacing the information in one copy with information taken from another.

Techniques for achieving these ends encompass various aspects of system, hardware, and software design, as well as the development of function-specific applications.

TIMELINE

A temporal approach provides a structure for describing the increasing number of available information assurance techniques and the growth in their complexity—a trend that has paralleled the increasing importance of the information itself. Using a timeline helps to put information assurance's progress into perspective. The timeline presented here is divided into several eras based on the specific information assurance techniques employed, and it extends about five years into the future. No

specific dates are associated with this timeline because it is not intended to be historically accurate. It does not imply that any specific technique is obsolete—in some variation, system administrators still use all these techniques.

Mainframes

This era covers the primitive data centers beloved of 1950s science fiction films, in which rows of reel-to-reel tape drives whirred and squealed as they processed information. In such systems, often the only method of information assurance involved copying information to and from the tape drives. This era saw the terms *backup* and *restore* coined for creating a copy and retrieving data from it, respectively. Usually, all system use ceased during the backup process, which meant system administrators tended to perform backups only at the beginning or end of the working day. Performing a restore could take hours or days, and it often required trial and error with several different tape volumes. This cumbersome and mostly manual process was subject to operator errors that were often discovered too late to prevent information loss.

The storage industry provided tape for use in this process because it was more rugged, easier to transport and store in bulk, and often significantly less expensive and more dense than disk technology. These characteristics made tolerating tape's serial access nature worthwhile, despite the obvious operational disadvantages.

Because of capacity and time constraints, not all of the information in a system could be backed up. Creating multiple tape copies could be justified only to allow storage of the most valuable information, in physically separate locations so that it would survive major calamities such as a fire or earthquake.

Distributed architectures obviated the requirement for directly connecting a tape drive to every system, thus improving device utilization.

Automation

As the volume of information increased, the information assurance process clearly required more operators and tape drives. This growth significantly affected the data center design, due to the cost of providing floor area for equipment and access for operators. In response, organizations sought to create a lights-out data center in which the same level of information assurance would require less space and no operator intervention. The confluence of two developing technologies enabled them to reach this goal:

- tape cartridges that were much denser and easier to handle than their reel-to-reel predecessors; and
- robotics systems capable of loading, unloading, and storing these tapes without human intervention.

During the automation era, programmatic control of the entire information assurance process became feasible. Software designers created specific application software that, in combination with automation, made information assurance faster and allowed more frequent backup.

Information availability increased significantly, and companies coined terms such as *NearLine* to describe how users could access copies in minutes. Multiple copies became feasible because the robotics could export tapes at defined intervals for storage at a remote site.

Developers also provided methods of ensuring the integrity of each copy. The process's efficiency allowed backing up a high percentage of the system's information, with defined policies determining the process characteristics based on the value of each information type.

Arrays

In previous eras, information assurance mostly involved creating backups of vital information that would reside outside the system. Even then, operating systems usually maintained duplicate copies of vital information on separate disks within the system, a process known as *mirroring*. High storage costs, however, meant that mirroring all information was not cost-effective.

As the volume of information—and therefore the need for storage capacity—continued to increase, however, the online storage failure rate became a significant concern. As the intervals between the use of some information stored on the disk length-

ened, it became imperative that information assurance also consider online storage because no method existed to detect the corruption of such data between backups. Hardware developers devised the *disk array* to address this problem.

A disk array separates the storage view presented to the system from the physical aspects of the storage itself. As such, disk arrays became one of the first instantiations of the virtualization techniques now commonly used throughout computer systems. The earliest arrays functioned purely as mirroring systems: They stored each piece of information written by a system on two separate disk devices and could access either copy to fulfill a read request.

Soon, however, developers devised techniques that allowed retrieving information in the event of any single disk's failure, without first requiring that the system store each piece of information twice. These redundant array of independent disks (RAID) techniques used parity schemes to protect the information.¹ They consumed only about 20 percent more storage than the data itself. Later arrays also let systems write at performance levels higher than any one physical disk by *striping* data: storing successive units of information on different physical disks.

Distributed architectures

With the growth of *local area networks* (LANs) and the proliferation of departmental servers, the single-system information assurance applications of the previous eras gave way to complex distributed applications. These applications created and managed multiple copies of information by employing several systems in different roles.

Common roles included a *media server* with attached backup tape drives, a *data transfer engine* that performed the I/O-intensive task of transferring information between a departmental server being backed up and the media server, and a *management console* that supervised and controlled the entire process.

Such distributed architectures obviated the requirement for directly connecting a tape drive to every system, thus improving device utilization. They also decreased the period during which the departmental server was unavailable to clients and applications or was operating at reduced performance levels. The LANs that these architectures used raised a new set of security concerns, leading to the introduction of features such as encrypted data sets and dedicated access mechanisms in information assurance applications.

Instant backup

Continued growth in the number of applications required to be online and available 24/7 meant that networks could no longer tolerate information assurance procedures that adversely affected performance. Thus, a new technique developed that captured an *image* of the information in the system at a specific time so that the assurance process could be performed against that image rather than the live information. This approach avoided causing lengthy interruptions to applications that rely on continuous access to information. These imaging techniques have a variety of names, including *snapshot*, *mirror*, *frozen image*, and *point-in-time copy*.

Imaging techniques also employ several underlying schemes—such as log-structured location determination—to ensure that an image remains stable while applications continue to use its original instance. Several different computer system components, such as a disk array or software agents embedded within an operating system, can create such stable images. We can view this assurance process as breaking off one of the mirrored information copies, performing the assurance process, then resynchronizing the copies—a task known as *resilvering*.

Storage networks

Initially, designers viewed storage as an integral part of a computer system, accessible only to and by the operating environment running on the system's processor. Given the short operating distance of the storage interfaces, and the limited amount of space available in the system cabinet, the system had only a relatively small storage capacity.

Over time, however, a trend toward the separation of storage and computer systems developed. Storage interfaces transformed from point-to-point connections into a more complex topology reminiscent of LANs, and this became known as a storage area network. A SAN supported a much longer operating distance than its predecessor and allowed multiple computer systems to access storage simultaneously, within the constraints of different formats and sharing rules defined by the systems' operating environments.

Today, SANs are based largely on Fibre Channel and SCSI interface technology, but the TCP/IP-based iSCSI and FC/IP will also find employment in tomorrow's SANs.²⁻⁶

The development of SANs affected information assurance techniques in two ways. First, multiple systems could share tape drives, even if they had markedly different operating environments. Information assurance applications therefore

required sophisticated access control schemes to ensure that multiple systems did not try to access a single device at the same time. Some applications went further and incorporated sophisticated caching strategies that could spool the data bound for tape to disk on an intermediate system. With such a cache, the backup could proceed without waiting for a physical tape drive to become available.

Second, SANs greatly increased the scalability of the storage that could be connected to a single computer system. As a result, information assurance processes evolved that did not use tape at all, but rather made copies of information and stored it on other disks. Rather than backups, these processes functioned as *replication* schemes that maintained active copies of data at two or more locations.

Developers devised two types of replication schemes: *synchronous* and *asynchronous*. Synchronous replication writes all data to all copies and prohibits the application from proceeding until all copies confirm they have been successfully updated. Asynchronous replication lets one copy proceed without waiting for confirmation from all other copies, and it transfers data between the copies at regular intervals in the background. Asynchronous replication is thus less precise because copies may be out of date with one another by that regular interval. On the other hand, this scheme places less stress on the bandwidth of the between-copy interface and causes no performance reduction for any application that accesses a copy.

Redundant infrastructure

As SANs became more advanced and stable, a greater percentage of the data center's information resided on SAN-attached storage. Meanwhile, the growth of Web services and e-commerce greatly increased the number of organizations that needed 24/7/forever operation. Both the restore window and the scheduled maintenance period, therefore, effectively shrank to zero.

This development placed a new set of requirements on both SANs and information assurance techniques. First, the SAN infrastructure needed to incorporate redundancy so that it never completely failed. This infrastructure supported a pool of servers, some of which could be taken offline for maintenance or switched to execute information assurance applications with no decrease in total system performance.

To achieve this goal, specialized clustering software developed that migrated applications between

Performing assurance processes against an image of the information avoids causing interruptions to applications that rely on continuous access.

processing resources. In addition, information assurance applications had to extend the access control schemes used for tapes to support disk access.

Second, the infrastructure needed to be seamless, spanning multiple physical locations so that system operation could continue without any interruption detectable by users even in the event of a major disaster. Often, such an infrastructure used an inter-switch link extender that employed a high-performance wide-area network to link switches located in each of the SAN's multiple physical locations.

These developments fundamentally affected how enterprises viewed both SANs and information

assurance techniques. SANs became a key business tool, as the "Storage Area Networks for Business Continuity" sidebar describes. Further, information assurance acquired the additional role of creating a quasilegal system operation record for identifying problems, resolving disputes, and planning and training, while also preventing loss of significant information.

These changes have increased the likelihood that information backed up on one server would be restored on a different server. Information assurance applications must, therefore, be able to tolerate changes in the hardware configuration and

Storage Area Networks for Business Continuity

Derek Granath, Brocade Communications Systems

Business continuity plans have become a necessity for small and large enterprises alike. Not having a plan in place that ensures 24/7 availability can be more expensive than putting one in place. If a disaster recovery plan is needed, it had better work right the first time. According to the Fibre Channel Industry Association, system downtime can cost a firm from \$14,500 per hour in lost automatic teller system fees to as much as \$6.45 million per hour from disrupted stock brokerage operations.

To improve business continuity, firms are adding storage area networks to enterprise systems that incorporate a combination of redundant components, connections, software, and configurations to minimize or eliminate single points of failure. By reducing or eliminating these failure points, SANs help improve the overall availability of business applications.¹

SANs achieve this high availability through a comprehensive, fault-tolerant system design that includes all components and supports 24/7 uptime requirements. The design must eliminate vulnerabilities to any source of downtime including equipment failures, human errors, environmental disasters, or even sabotage. Delivering a high-availability environment through a SAN requires establishing availability objectives, creating fault tolerance, and implementing an intelligent SAN infrastructure and fabric management.

Availability

Internet and global e-business application requirements demand that companies increasingly implement computing infrastructures specifically designed for at least 99.999 percent availability—the equivalent of less than 5.3 minutes of downtime a year. Availability is a function of outage frequency caused by unplanned failures or scheduled maintenance and upgrades, plus the time to recover from those outages. Companies must identify specific availability requirements and predict potential failures to create a high-availability solution that meets the organization's needs. Objectives vary widely among and within companies: Some can tolerate no disruption, while short outages may only minimally affect others.

To address this uptime issue, many companies implement flexible SANs that incorporate fault tolerance through redundancy, mirroring, hot-plugging capabilities, and avoiding single points of failure. They also speed recovery through simplified fault monitoring, diagnostics, and nondisruptive server-storage maintenance and repair. Using intelligent routing and rerouting, coupled with dynamic failover protection, minimizes the need for human intervention during failover events. The design of the SAN is key to delivering a "beyond five nines" solution for application availability and business continuity.

Fault tolerance

Implementing fully redundant SANs consisting of alternate devices, data paths, and configurations can increase system availability. Ensuring dual paths through separate components is particularly important, especially when physical location and disaster tolerance are concerns: A single device cannot adequately address these issues.

For better availability, the focus shifts from servers to applications. Mission-critical applications should reside on redundant, highly available servers and storage devices to allow access to data even during a failure. Sophisticated software avoids application or host failover by moving workload to a secondary server. Likewise, clustering technology transfers the workload to multiple active servers without disrupting data flow.

To further improve availability, servers should include redundant hardware components with dual power supplies, network connections, and mirrored system disks. Each server should have multiple connections to alternate storage devices through multiple switches, with a minimum of two independent connections to the SAN. In addition, servers should feature dual-active or hot-standby configurations with automatic failover capabilities.

The path between the server and storage, interface cards, cabling, fabrics, or storage connections is another likely failure point. Dual-redundant interface-card configurations help ensure path availability and boost performance with additional SAN connectivity.

For true fault tolerance, multiple paths must be connected to

operating environment. Further, the server to which the information is restored may have never operated before, requiring a single process—known as a *bare metal restore*—to reestablish an operating environment and all applications and data.

Networked storage

Where SANs stretched the interface between servers and directly attached devices, *network-attached storage* exported not just the storage device from the server but part of the file system as well. Thus, systems that used a block-based paradigm to access storage via a SAN had to use a file-

based paradigm to access NAS via a LAN, an approach that offers ease of use and management.

On the other hand, NAS devices often lack the ability to load and execute external applications—a limitation that constrains information assurance to use the same file-level access as servers and other applications. This method severely limits applications that have been developed over many years to use the block-level paradigm of a directly attached storage device.

The Networked Data Management Protocol, a common method for accessing NAS data,⁷ provides a solution to this quandary. NAS devices that sup-

alternate physical locations within the SAN, to different switches in a multiswitch fabric, or to different blades within a core fabric switch. To provide full redundancy, some companies choose a dual SAN configuration. Server-based path-failover software typically allows a dual-active configuration for dividing workload between multiple host bus adapters (HBAs). The software monitors the health of available storage products, servers, and physical paths and automatically reroutes data traffic to an alternate path when failures occur, as Figure A shows.

Organizations that cannot tolerate any application interruption replicate their complete data-center environments in different physical locations. SAN extension capabilities using metropolitan area networking technologies, such as dense wave division multiplexing, enable full bandwidth synchronous replication of the data center up to 100 km away from the primary data center. While they do not deliver full bandwidth performance, SAN extensions using wide area networking technologies such as Sonet or IP offer redundancy beyond the radius of virtually any physical disaster.

Many of today's storage devices feature multiple SAN connections for fault tolerance in storage solutions. Multiple connections guard against failures from a damaged cable, controller, or SAN component such as an optical interface module. Mirrored storage subsystems that function on a peer-to-peer basis across the fabric also create highly available storage connections for fault tolerance. Combining mirroring with switch-based routing algorithms creates a resilient, self-healing environment that provides an alternate access point to data regardless of path conditions.

System availability

Achieving higher availability through redundancy and fault tolerance requires first understanding the system's specific uptime requirements, then designing a solution tailored to meet them. Complete system outages can be avoided only by eliminating all potential single points of failure through complete component, device, connection, and path redundancy.

Multiple connectivity paths, clustering techniques, and dual

fabrics all contribute to a fault-tolerant solution. Physically separating devices to protect against localized physical disasters helps to create fault-tolerant systems. Additionally, networks of switches are less vulnerable to localized disasters that might affect the entire system's fault tolerance. Together, these measures help organizations become more efficient, reliable, and resilient.

Reference

1. P. Massiglia et al., *The Resilient Enterprise*, Veritas Press/John Wiley & Sons, 2002.

Derek Granath is director of product marketing for Brocade Communications Systems. He received a BSEE from Stanford University and an MBA from Santa Clara University. Contact him at dgranath@brocade.com.

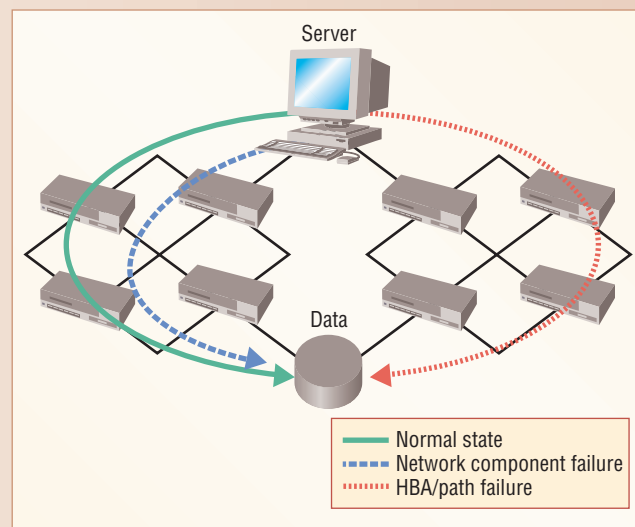


Figure A. Dual SAN configuration for fault tolerance. The SANs software monitors the health of available system components and reroutes data traffic to an alternate path when failures occur.

Security Concepts

Five concepts underlie the definition of most security technologies.

Authentication

This security measure establishes the validity of a transmission, information set, or originator. In its most straightforward form, authentication ensures that the information originated from a specific known source. Often, to enable bidirectional information transfer, protocols require authenticating two parties to one another.

Authorization

Authentication is a necessary condition of authorization: actions that a specific party is permitted to execute. Authorization ensures that a requestor is allowed to receive a service or perform an operation. Access control is an example of authorization.

Confidentiality

This security measure protects against the disclosure of information to parties other than the intended recipients. Often, administrators use cryptography to ensure confidentiality by encoding the information with a defined algorithm and some secret information known only to the originator and the intended recipients. Alternatively, *steganography* ensures confidentiality by disguising the type of information, for example, by hiding a message in a bitmapped image.

Integrity

This security measure lets the receiver or reader determine that an information set has not been altered since creation or while in transit. Integrity schemes often use some of the same underlying technologies as confidentiality schemes, but they usually involve adding information to form the basis of an algorithmic check rather than encoding all the information.

Nonrepudiation

This security measure prevents later denial that an action happened, a communication took place, and so on. In communication terms, nonrepudiation often involves the interchange of authentication information combined with some form of provable time stamp.

Going forward, information assurance applications must include techniques involving all five concepts if they are to avoid becoming weak points in secure systems, and if the information that they produce is to be accepted as a legal record.

port NDMP provide the types of access that information assurance applications specifically require.

Mobile devices

Much client-side information has migrated from fixed-desktop systems to mobile devices such as laptops and personal digital assistants. In this environment, connectivity between mobile devices and any equipment capable of providing the appropriate storage for information assurance processes is only intermittently available, of inconsistent bandwidth, and often employs a publicly available infrastructure. Yet information assurance is all the more

vital because of the higher probability of physical component loss or damage inherent in the mobile environment.

Much of the information stored on each mobile device, such as its operating system, is common to many devices. Given the remoteness of the mobile device from any support organization, information assurance still must protect this data so that small failures will not have a large impact. Thus, information assurance processes must support multiple clients of several types, and the processes themselves must operate on a self-service basis with little or no user training.

Bandwidth limitations have led to the development of information assurance applications capable of detecting and adapting to various levels of interconnection performance. In some cases, such applications identify changes to the information that occurred after the previous transfer, then they process only information about those changes.

The duplication of information across devices has led to the creation of sophisticated techniques, based on approaches such as checksums, that can identify common information and produce consolidated backups across multiple mobile devices. These devices' dependency on public networks has led to communication processes that incorporate confidentiality protection and methods of client authentication. Using automated scheduling and event-driven techniques helps to minimize the need for human intervention.

Intelligent infrastructure

As they refine SANs, developers are seeking ways to transfer information directly between the disks that departmental servers access and the tapes a media server manages without using a LAN. Clearly, such applications require even more sophisticated access control schemes to isolate the selected information from interference by other systems connected to the SAN. Removal of the LAN bandwidth restrictions and server processing bottlenecks makes such applications more scalable than their predecessors.

The most recent developments in this area have taken this approach a step further by supporting *server-free* schemes. This approach uses a copy engine in the SAN to directly transfer information between the SAN-attached disks and tapes. The information assurance application manages the copy engine, but the majority of the information that it transfers does not flow through any of the servers on which that application resides.

LAN-free applications have only recently achieved general deployment, while server-free schemes are

being deployed only in restricted configurations so far, where testing has shown their performance to justify the cost.

FUTURE TRENDS

I expect the development and specialization of information assurance applications to continue and even accelerate over the next five years. Several trends will likely change and extend the requirements placed on information assurance applications, including

- Internet growth,
- advances in storage density and SAN scalability,
- increasing use of wireless client devices with limited processing power and memory, and
- broader availability of high-performance wide area networks.

In response to these trends, information assurance will have to address both increased security and support for truly global replication.

Increased security

Ultimately, replacing paper records and manual processes with interactive applications and electronic information will require treating all data in whatever form as a legal record of activity. Further, increasing use of the Internet and remote access schemes by mobile users who deal with confidential corporate information has made the security of such data an increasing concern.

Information security must ensure far more than the confidentiality of information in transit. It must also ensure that only authorized parties have access to such information, a task that will require abuse-resistant methods for identifying such parties.

In the information security context, these *authentication schemes* implement one of five key security concepts described in the “Security Concepts” sidebar. These techniques will play a central role in shaping future information assurance applications and ensuring that the information they manage achieves the status of legal records. Further, as information processing systems incorporate more security techniques, developers will need to use some or all of these concepts to ensure that their information assurance applications do not compromise system security.

Although applications alone may never be capable of addressing all security requirements, incorporating security features into a device can further enhance system security, as the “Self-Securing Devices” sidebar describes.

Self-Securing Devices

Greg Ganger, Carnegie Mellon University

Crackers, e-mail viruses, self-propagating worms, and denial-of-service attacks make security compromises a fact of life. No single defense can adequately guard against these threats, so security functionality should be distributed among physically distinct components. Inspired by siege warfare, individual devices erect their own security perimeters and defend their own critical resources, such as a network link or storage media.

Together with conventional OS and firewall defenses, such self-securing devices¹ offer greater flexibility for dealing with intrusions. By having each device erect an independent security perimeter, the network environment gains many outposts from which to act when under attack. Devices not only protect their own resources, but also can observe, log, and react to the actions of other nearby devices. Infiltration of one security perimeter will compromise only a small fraction of the environment—other devices can work dynamically to identify the problem, alert still-secured devices about the compromised components, raise the environment’s security levels, and so on.

Given that storage devices persistently store data, one natural intrusion-survival extension is for such devices to protect stored data by preventing undetectable tampering and deletion. Self-securing storage devices do this by managing a storage space from behind the device’s security perimeter, keeping an audit log of all requests, and keeping clean versions of data later modified by attackers.¹

A storage device cannot distinguish between legitimate and compromised user accounts, so protecting legitimate users requires keeping all versions of all data. Finite capacities limit how long such comprehensive versioning can be maintained, but recent 100 percent per year storage-capacity growth will let modern disks store several weeks’ worth of all versions. If detection mechanisms reveal an intrusion within this window, security administrators can use this valuable audit and version information for diagnosis and recovery.¹

Reference

1. J.D. Strunk et al., “Self-Securing Storage: Protecting Data in Compromised Systems,” *Proc. 4th Symp. Operating Systems Design and Implementation*, Usenix, Berkeley, Calif., 2000, pp. 165-180.

Greg Ganger is director of the Parallel Data Lab and an associate professor at Carnegie Mellon University. Ganger received a PhD in computer science and engineering from the University of Michigan. Contact him at ganger@ece.cmu.edu.

Global replication

The usefulness of synchronous and asynchronous replication techniques is a function of the available network bandwidth between disparate physical locations. The growth of high-performance WANs based on optical data transmission and all-optical switching, however, promises to make increased network bandwidth available between many more locations, over greater distances, at improved cost-performance.

When this trend combines with growing storage densities and improved SAN scalability, economi-

cally supporting many copies of the same information spread across an entire country, or even the globe, could be feasible. Such a scheme could support low-latency access and a high degree of availability by automatically routing a mobile user to the closest physically available copy.

It may even become possible to permit concurrent access to different copies of the same information by different users, with a background process making any changes and consolidating all copies to create a consistent version. Storage will then become a true utility, available in all locations and situations, removing user concerns over availability and accessibility.

In many ways, these replication techniques offer advances mirroring the evolution that the Uniform Resource Name framework, defined for the World Wide Web, brought to document availability.

The spectrum of information assurance techniques is continually broadening as the industry responds to the ever-increasing value and volume of information contained in storage. These techniques have evolved from comparatively simple offline backups and are headed toward the complex features that tomorrow's enterprises will require to support always-available, location-inde-

pendent information, and ultrareliable, infinitely scalable storage resources. ■

References

1. P. Massiglia, *Highly Available Storage for Windows Servers*, John Wiley & Sons, New York, 2001.
2. Fibre Channel Industry Association, <http://www.fibrechannel.org>.
3. Fibre Channel standards, <http://www.t11.org>.
4. SCSI Trade Association, <http://www.scsita.org/>.
5. SCSI standards, <http://www.t10.org>.
6. iSCSI, <http://www.ietf.org/html.charters/ips-charter.html>.
7. Network Data Management Protocol, <http://www.ndmp.org>.

Roger Cummings is a senior staff software engineer for Veritas Software, where he participates in the development of storage software products. Cummings received a BSc (Eng) in electronic engineering from Queen Mary College, University of London, UK. He is a member of the Technical Council of the Storage Networking Industry Association, co-chair of SNIA's Security Technical Working Group, and chair of iNCITS Task Group T11.5. Contact him at roger.cummings@veritas.com.

Read articles on these diverse topics in *Computer* in 2003

outlook: looking ahead to future technologies
january

commercial workload evaluation
february

grid computing
march

hardware/software codesign
april

the changing face of networking
may

agile software development
june

nanotechnology
july

piracy and privacy
august

mobile systems
september

web services
october

safety-critical systems
november

power-aware computing
december

To submit an article for publication in *Computer* on these or other topics, read our author guidelines at <http://computer.org/computer/author.htm>.

Innovative Technology for Computer Professionals
Computer

IEEE
COMPUTER SOCIETY