Executive Summary

Government agencies have been dealing with digital evidence for many years – for example, collecting digital “images” of personal computer hard drives or taking pictures of crime scenes. The amount of collected digital devices and media has grown incredibly, both in terms of the number of units and the storage capacity of each.

Also, agencies themselves are increasingly relying on digital technologies for evidence collection, such as digital cameras and computerized video surveillance systems. And a rising number of applications that generate their own digital evidence are being used to support the digital evidence investigative process.
Digital evidence has proliferated to such a degree that many agencies find that they are fighting a losing battle against their digital evidence infrastructure needs. These agencies must process, review and store so much evidence that the old techniques of manually reviewing and storing it have largely become antiquated. The solution to the challenges created by advances in technology is to become more proficient in the use of other technologies.

Digital evidence management technologies can help agencies become more efficient with evidence collection, processing, review, presentation, archiving and disposition. In addition, digital evidence management technologies are critically important to protecting the integrity of digital evidence, such as ensuring that evidence is not accidentally or intentionally altered.

The amount of digital evidence will continue to increase, so agencies must be prepared to handle both today’s workloads and future workloads that will undoubtedly be even larger. Digital evidence infrastructure technologies, such as networking, data centers and storage, should be planned out to meet this expected growth. This white paper examines the infrastructure that underlies today’s digital evidence management technologies, offering guidance on where limited funds can best be spent.

**Forensics and Digital Evidence Technology Today**

Until recently, forensics and digital evidence technology were viewed largely as stand-alone capabilities. Conventional wisdom was that everything should be kept isolated from networks – for example, to collect a hard drive image and manually process a copy of that image on a dedicated forensics workstation. This approach, while fine for low-volume digital evidence collection, does not scale well and cannot handle the volume and variety of digital technologies currently encountered by most agencies.

Today’s digital evidence technologies are fundamentally different than those of previous generations. While they are based on the same principles (for example, ensuring that evidence is not altered), they have significantly evolved.

Digital evidence management technologies include software applications specifically designed to support digital evidence throughout its lifecycle, from initial collection of the evidence to its final disposition once it is no longer needed. These digital evidence management technologies are network-enabled and centralized, allowing multiple analysts to examine the same evidence simultaneously, and even allowing analysts from different agencies to perform their own investigations on the same instance of the evidence.

Network-enabled also means that evidence can be collected from a variety of locations, such as agents out in the field, and submitted remotely, instead of being physically carried to the forensics laboratory. For example, photographs taken at a crime scene by an investigator’s smartphone can automatically be transferred to a digital evidence management system. This saves considerable time and effort for the investigator, and it ensures that the evidence is secured as quickly as possible — literally, as soon as it is captured.

Digital evidence management technologies offer other benefits too. They are normally database-driven, with outstanding search capabilities not only for evidence metadata, but also for evidence content. This includes searching videos, audio files and other forms of media for particular information or patterns.

Such search capabilities are vitally important, given the sheer volume of media to be analyzed. Consider surveillance videos and how long it would take to manually review them for a particular event, instead of using a search capability to find frames that appear to contain a particular person’s face. The centralized nature of digital evidence management technologies also allows a lot of processing power, memory and other computing resources to be made available centrally to process evidence for many different analysts simultaneously.

A final benefit of digital evidence management technologies is their access control and auditing capabilities. These technologies are specifically designed to protect evidence, through encryption and strong authentication methods. Only individuals who are authorized to review evidence for a particular case may do so, and every action that they perform involving the evidence is carefully logged for auditing purposes.

By these actions, the digital evidence management technology carefully safeguards each piece of evidence and generates an exhaustive record of every action involving that evidence, so that no questions arise regarding the integrity and confidentiality of a piece of evidence.
The Digital Evidence Workflow

The typical workflow for digital evidence comprises six phases:

1. **ACQUIRE**: Collect digital evidence and store it securely.
2. **PROCESS**: Tag digital evidence with metadata. Convert analog evidence to a digital format, if appropriate. Also, make copies of digital evidence for the evidence owner, if necessary.
3. **ANALYZE**: Review evidence, either manually or automatically (for example, through searches), and document all findings.
4. **PRESENT**: Disclose evidence in court or to other legal parties (for example, as the prosecution would release copies of digital evidence to the defense).
5. **ARCHIVE**: Place evidence in an archival state for storage until it is no longer needed. This may be weeks, months or years, depending on the details of the individual case.
6. **DISPOSE**: Return physical evidence (computer media, etc.) to its owner and delete or otherwise destroy all stored digital evidence.

In order to support these phases, agencies performing digital evidence–based investigations need an infrastructure that can store, secure and otherwise manage digital evidence. Major components of this infrastructure include storage, networks and a data center. The rest of this white paper examines these components in greater detail, providing guidance on where limited resources should be focused.

Storage

Storage needs can be extreme for digital evidence management. Three factors drive these increasingly high storage needs: the number of devices from which digital evidence is collected, the potential storage size of each of these devices and the use of digital technologies for collecting evidence (such as digital cameras and video recorders).

Another factor is that storage is needed for several purposes. Storage may be needed to collect the original digital evidence from a computer, mobile phone or other digital device. Storage may be needed to hold photos, video recordings and other documentation of an incident taken by investigators.

Storage may also be needed to hold copies of digital evidence for analysis and presentation purposes, and to give copies of digital evidence to other parties (such as prosecution, defense, evidence owners or agencies in other jurisdictions). Let’s consider these two primary forms of storage: direct-attached storage and network-based storage.

Direct-attached Storage for Individual Analysts and Investigators

Direct-attached storage involves one or more storage drives, such as traditional external hard drives or flash drives, that are directly attached to a local computer.
In several circumstances, it makes sense for individual investigators and analysts to have direct-attached storage. For example, an agency with only one or two analysts may have limited digital evidence needs. Or, an investigator or analyst may collect disk images from computers out in the field, without high-speed Internet access.

The default media option for many analysts has been traditional hard disks. However, traditional hard disks are largely being supplanted by solid-state drives (SSD), also known as flash drives. Unlike traditional hard drives, SSDs do not have moving parts, so they are more durable and resistant to physical trauma.

Another significant advantage of SSDs is that they use considerably less power than traditional hard drives. SSDs are also physically much smaller than traditional hard drives of the same capacity. Finally, SSDs tend to provide better performance than their traditional counterparts.

However, one primary disadvantage of using SSDs is that they are more expensive than traditional hard disks with the same capacity. But if the increased efficiency of SSDs leads to shorter data acquisition times, then the savings in labor may more than make up for the higher purchase costs.

Another disadvantage of using SSDs is that they are not available in capacities as large as traditional hard disks. Depending on an agency’s data acquisition needs, SSDs may simply be too small to be feasible for use. However, for other uses, SSDs are generally preferable to older hard-disk technologies.

Regardless of whether traditional hard disks or SSDs are used, using direct-attached storage can lead to significant problems. Direct-attached storage typically provides no redundancy, so if a hard disk fails, the data it holds is lost. Also, direct-attached storage offers no easy sharing of stored digital evidence, because the storage generally is available only to one user at a time. Sharing evidence generally means disconnecting the storage, physically transferring it to another user and connecting it to that user’s computer.

**Network-based Storage for Analysts and Investigators**

Most agencies find that they need network-based storage to meet the demands of their analysts and investigators. Investigators collect too much digital evidence to rely solely on direct-attached storage devices (a single case might necessitate the use of dozens or even hundreds of direct-attached storage devices), plus agencies need analysts and investigators to seamlessly share digital evidence with each other. Many agencies find that they need centralized network-based storage to achieve the scalability and shareability required to meet today’s digital evidence needs.

A common form of network-based storage is known as network-attached storage. NAS storage devices each have a separate, dedicated IP address. Another form of network-based storage is known as a storage area network. In a SAN, the whole network has a single IP address.

NAS is generally used in a smaller environment than a SAN, which is more scalable and provides higher performance. Unified storage, a hybrid form of network-based storage that uses both NAS and SAN, supports a variety of storage devices and provides a consistent, easy-to-use interface to manage them.

Most forms of network-based storage offer features to optimize the amount of data stored. For example, data deduplication looks for duplicated chunks of data (such as 10 copies of the same photo) and stores only one copy, noting the existence of the rest. Data deduplication is a special form of data compression, which is the general term for taking advantage of patterns in data to store it in a smaller amount of space than would otherwise be used.

It is important to note that if encryption is being used to protect digital evidence files, any form of data compression should be applied first. The compressed data should then be encrypted, as compression is largely ineffective when applied to encrypted data.

Agencies are still developing standards and best practices to increase the effectiveness of digital evidence systems. The National Institute of Justice, an agency of the U.S. Department of Justice, works with other organizations to establish standards for the use of digital evidence. These standards address measures agencies can take
for the safe handling and subsequent analysis of digital evidence collected at a crime scene. They also include criteria for how digital evidence equipment is used and for certifying examiners.

**Backup and Archival Needs**

Like any other data, digital evidence must be backed up so that it exists on at least two instances of media, with one preferably stored securely offsite to support disaster recovery. This need for backup media is intensified for digital evidence because agencies frequently need to archive data for months or years, and backup devices are a common way of achieving this.

Some agencies choose to leave materials on centralized storage even when they arearchivable, depending on their storage costs. But for most agencies, centralized storage is too expensive to be used to hold archived materials. Instead, archived materials can be written to backup tapes, DVDs or other media and stored securely.

The appropriate backup/archive solution for a particular agency will vary widely based on the amount of data to be archived, the length of time that the archived data is to be kept, the costs of the various backup/archive solutions and the network-based storage technologies in use.

**Removable Media for Investigators**

Many digital cameras, video recorders and other audiovisual devices use removable media, such as memory cards, secure digital (SD) memory cards and picture cards, for storing the photographs, videos and audio clips that they take. However, because these forms of media are rewritable, investigative agencies may have some concerns about how well they support the integrity of evidence. For example, a photograph stored on a rewritable card could be altered, either inadvertently or intentionally, and saved back onto the card, thus compromising the legitimacy of the evidence.

To combat this, special SD memory cards have been developed that are write-once, read-many (WORM) cards. Digital cameras, video recorders and other audiovisual devices that support these cards can write to them, but they cannot overwrite any data that has already been written on them. Every photograph, video file and audio clip taken is stored permanently on the memory card and cannot be modified, deleted or otherwise tampered with. And these cards can be read by any device that reads SD memory cards, which are ubiquitous.

WORM SD cards are more expensive than regular SD cards. Agencies should weigh the cost of WORM SD cards against the risks of evidence being inadvertently or intentionally altered and the costs of controls put into place to prevent such alterations.

**Networks**

The storage needs for digital evidence can be massive. Similarly, digital evidence technologies demand a network infrastructure that supports secure, rapid access to data on demand — with proper authorization, of course.

**Network Security**

An investigative agency should have a network for digital evidence management systems that is physically separated from the rest of the agency’s networks. If Internet access is needed on this digital evidence management network, it should have its own dedicated Internet feed instead of sharing Internet connectivity with other agency networks.

Internet access is needed in most cases — for example, to acquire and apply operating system and application patches to analyst computers and digital evidence management servers. Isolating the digital evidence management network from the agency’s other networks helps to minimize the exposure of the network to attackers, particularly when paired with a strongly configured next-generation firewall.

A next-generation firewall is most readily distinguished from traditional firewalls by its understanding of applications. This is particularly helpful for web-based applications. Instead of seeing HTTP as a single application, a next-generation firewall can distinguish each web-based application using HTTP from the others and apply separate security policies to each application.

A next-generation firewall also provides intrusion-prevention capabilities, helping to protect the agency’s digital evidence management servers and workstations from malicious attacks. All of the devices on a digital evidence management network should be strongly protected by firewall policies that allow only the minimum access required to each device.

Another helpful network security control is network access control. NAC, which is primarily geared toward the security of workstations, governs access to a network by checking and verifying the security state of the workstations attempting to gain access to that network. NAC typically verifies that anti-virus software is installed, enabled and up to date, and that all operating system and major application patches have been applied.

Any workstation that fails one or more of the security checks is prohibited from accessing the primary network, although it may be granted access to a virtual local area network (VLAN) strictly for the purposes of remediating
problems, such as installing missing patches. NAC is also an effective way to protect workstations from malware and other threats by ensuring that they are well secured before granting them network access.

Agencies should also consider enabling network auditing through the use of network flow information, such as the Internet Protocol Flow Information Export (IPFIX) protocol or the Netflow protocol. Protocols such as these are used to collect information about IP-based network flows, such as connections between remote digital evidence management clients and the agency’s digital evidence management servers. This logging demonstrates not only connectivity between hosts, but also the amount and type of data transferred between those hosts and is yet another way to audit access to sensitive digital evidence.

**Network Performance**

Every data center and every analyst workstation setup will contain local area network (LAN) components. Another interesting element of network performance is the performance of wide area networks. WANs are increasingly used for digital evidence management:

- To transfer digital evidence to the agency’s centralized location from remote locations
- To grant access to digital evidence for partner agencies
- To facilitate backup/archiving for disaster recovery purposes

As WAN use has increased, so has the need to make WAN communications more efficient. Specialized devices called WAN optimization controllers (WOCs) are dedicated to optimizing the use of WAN bandwidth. WOCs use many techniques to improve WAN performance, including data compression, data deduplication, network latency reduction, and traffic prioritization and shaping (by user, protocol, application, host, etc.).

Another technology associated with WAN efficiency is called Wide Area Ethernet (WAE), which is basically a way of providing WAN services over distributed Ethernet networks. WAE allows multiple Ethernet networks to be connected over virtual private network-like links, so that all the disparate Ethernet networks are treated as a single Ethernet-based WAN.

This can make network management and usage easier, while also enabling the ability to use traffic shaping techniques and other WAN-based methods of enhancing network performance. Another advantage of WAE is that it offers higher throughput than other WAN technologies.

Balancing the WAN load across servers used to be accomplished by dedicated devices known as load balancers. The primary purpose of load balancing is to split client requests across multiple servers in a seamless way, so that the clients don’t need to know the location (for example, the IP address) of each server. The load balancer automatically manages those connection requests and chooses the appropriate server for the desired connection.

Load balancing is still a commonplace technique, but increasingly it’s being performed by devices other than...
load balancers. A common example is application delivery controllers. These devices are placed between the firewall and the application servers.

Not only do the devices perform load balancing among the application servers behind them, but they also provide data compression, traffic shaping and Secure Sockets Layer offloading (so that the application delivery controller, not the application server, handles the encryption and decryption load associated with SSL-protected application traffic, such as HTTPS).

Additionally, many application delivery controllers provide web application firewall capabilities, further improving the security provided to the application servers. Agencies with a significant digital evidence infrastructure should give serious consideration to acquiring application delivery controllers to enhance performance and increase security.

Digital evidence networks also may employ the Spanning Tree Protocol, which lets IT administrators design a network that includes redundant links that provide automatic backup paths. The protocol eliminates the need for IT staff to manually enable or disable these backup links.

**Data Center**

The glue that binds the storage and network infrastructure together is the data center itself. The data center makes the storage and networks available to the applications running on servers and workstations in a digital evidence system.

The data center also provides the processing horsepower behind those applications. Three elements of the data center are of greatest interest in terms of supporting digital evidence technologies: high-density servers, interconnectivity protocols (network protocols, storage communication protocols, etc.) and interconnectivity hardware (switches).

**High-density Servers**

High-density servers, also sometimes referred to as blade servers, are stripped-down servers that contain only the essential hardware, so as to minimize the use of space, the consumption of power and the complexity of the supporting infrastructure. In data center terminology, a blade enclosure contains numerous blades.

Each blade has the minimal server components necessary, including one or more (typically multiple) processor cores, with the blade enclosure itself providing the functions that the individual blades do not. The details of this vary from server to server, but typical functions provided by the enclosure for the blades include cooling, power, networking, input devices (keyboard, mouse, etc.) and output devices (monitors).

Centralizing these functions provides several advantages. One is improved manageability: Having all of these “servers” integrated into a single server makes them significantly easier to deploy, monitor and maintain. Troubleshooting is often simplified because fewer components can fail (for example, one primary power supply as opposed to many primary power supplies).

Another advantage is a huge reduction in cabling. With so many functions consolidated into the server, this environment has minimal need for cabling, and may require only one network cable instead of dozens, or hundreds.

High-density servers are designed for efficiency as well as flexibility. An agency can deploy a partially filled, high-density server to meet initial digital evidence needs, then add more blades to that server as needs increase. It can scale further if the agency finds itself in need of far more servers quickly. Adding blades or servers is generally transparent from the point of view of both users and applications. The extra blades or servers are managed as part of the whole.

Agencies are likely to use high-density servers to support their digital evidence technologies because of the sheer processing power that these servers can provide. Because of the number of processor cores that can be aligned to work simultaneously, high-density servers provide massively parallel processing and can be extremely effective at supporting computationally intensive applications, such as widespread data searches and image processing and recognition technologies.

**Interconnectivity Protocols**

The pieces of a high-density server need to use one or more interconnectivity protocols in order to work together. Interconnectivity is particularly important in two respects: storage and networking.

In the most basic architectures, traditional protocols such as Ethernet (for networking) and Fibre Channel (for storage) are used, with network and storage interfaces embedded in each blade. However, increasingly high-density servers are using newer technologies that often combine storage and networking interconnectivity in a single protocol.

This is most often accomplished through use of the Fibre Channel over Ethernet protocol. FCoE is essentially an implementation of Fibre Channel that works over
Ethernet instead of the traditional Fibre Channel networks (which typically used fiber optic cables or twisted pair copper wire). The Ethernet connections between blades or between blade servers can support both network communications and storage communications.

FCoE supports the use of higher-speed Ethernet networks, such as 10 Gigabit Ethernet (10 Gig-E), providing high throughput for both networks and storage. However, it is important to note that FCoE is not IP-routable, so while it works quite effectively for a truly local network, it is not intended for use in broader network connections.

FCoE works most effectively through the use of converged network adapters (CNAs), which are essentially a hybrid of an Ethernet network interface card (NIC) and a Fibre Channel host bus adapter (HBA). FCoE can work through normal Ethernet NICs, without an HBA present, by having software simulate the HBA function, but this is significantly less efficient than using native CNAs.

Other advantages of using CNAs are reductions in the number of network adapters and interface cards needed, as well as reductions in the amount of cabling and network equipment.

**Interconnectivity Hardware**

Networking equipment is needed to connect local networks, such as FCoE-based networks, to larger IP-based LANs and WANs in order to enable access to the applications and data stored on high-density servers. A term commonly used in data centers is top-of-rack switch, which can be thought of as a networking switch deployed at the rack level.

A rack can contain several high-density servers, so a top-of-the-rack switch brings together those servers at a single network junction. This is the appropriate place to connect the local (for example, FCoE-based) networks and to interconnect them with other IP-based networks. FCoE is not IP-routable, so it typically passes through the switch from one FCoE network to another, but does not traverse standard IP-based LANs or WANs.

Because of the consolidation achieved through the use of FCoE, these top-of-rack switches typically don’t need to support a huge number of ports, compared with the equivalent in traditional servers and interconnection protocols. Performance is a major consideration when selecting a top-of-rack switch. Other important factors are power consumption and cooling.

**Other Interconnectivity Protocols**

Agencies have many interconnectivity protocols to choose from, not just FCoE. Examples include Fibre Channel itself, PCI Express, Serial ATA and InfiniBand. All of these protocols are designed for high-speed storage access for servers and their processors. InfiniBand has received particular attention in recent years because of the quality of service (QoS) properties that it supports.

QoS is normally associated with WANs; using the InfiniBand protocol allows QoS principles to be applied to storage devices within a single rack or within a group of racks in a data center. This can improve efficiency and supports the availability of critical communications within the storage network.

InfiniBand provides other beneficial characteristics as well. For example, it supports the exchange of cryptographic keys to provide additional security assurances through authentication. This may be of particular interest to agencies with high-security digital evidence needs that want to provide even greater levels of security for their digital evidence.

There is no “right answer” for every agency as to which interconnectivity protocol it should use. Agencies considering the selection of interconnectivity protocols should review all the options and choose the protocol that best meets their needs for performance, scalability, security and cost.