How to Implement Trusted Computing

A Guide to Tighter Enterprise Security

By Roger L. Kay
Endpoint Technologies Associates

Introduction

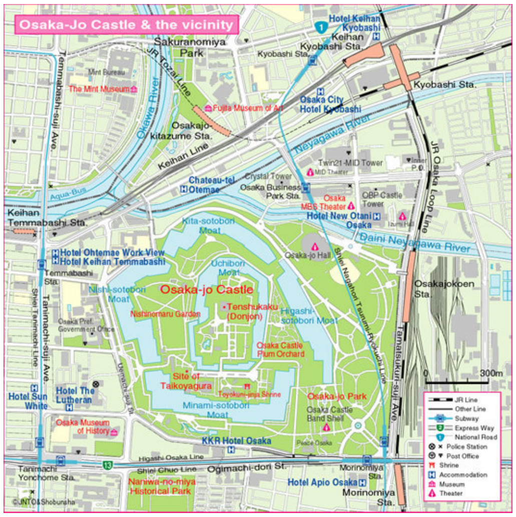
Philosophically, one can imagine the security problem two ways: as a perimeter or as a set of secure connections.

Perimeter Defense

The failure of a perimeter is starkly illustrated by the Maginot Line, famous from World War II. The French built a strong, secure wall against potential attack by the Germans, but neglected to make the perimeter complete, leaving out, among other things, an extension into Belgium of the network of trenches, bastions, walls, redoubts, mines, barbwire, and lookouts that protected against a frontal assault on France. It was through this break that the Germans passed on their way to Paris.

The French might have learned from successfully repulsed sieges of the Middle Ages that a perimeter has to be complete to work. But even multi-walled castles with layers of protections — moats, drawbridges, narrow entry points under commanding fields of fire, interior water sources, stored food, weaponry, and thousands of defenders — have been known to fall. Hideyori, daimyo of Osaka, failed to hold such a structure, Osaka-jo castle, against Tokugawa Ieyasu in 1615 (Figure 1).

Figure 1: Osaka-jo Castle, a Permeable Perimeter



Secure Connection

During World War II, an interesting and vital technique practiced by American intelligence relied on secure connection. The technique depended on Navajos, people indigenous to the American Southwest, who spoke a doubly-coated dialect of their native language plus symbolic representation (e.g., a platoon was "has-clish-nih," which translates to "mud"; Germany was "besh-be-cha-he," which translates as "iron hat"). They spoke over channels that the Allies understood to be monitored by Axis intelligence, but the trick was that only the sender and receiver could decode the message (Figure 2). Although this technique has been superseded by advanced encryption for data protection, it was good enough at the time.

Figure 2: Code Talkers in the Jungle

Of the two schemes, secure connection is more desirable theoretically. A perimeter is made up of all kinds of elements thrown together (walls, rivers, wires), and the interstices represent potential vulnerabilities. A secure connection can be made between any two elements in a given group of elements that might want to talk to each other. Once such a connection is established, the two parties can chatter away freely without worrying about who might listen in. In the real world of rising reliance on the Internet, which exposes ever increasing amounts of value to potential depredations, the pure secure connection model is unlikely to take hold in the near future. For now, security will be made up of perimeter defenses and a degree of secure, bilateral connections.

How the Industry is Addressing Security

Although well established, the perimeter approach continues to suffer from breaches by ever more creative assailants. Threats have only increased as connectivity and network complexity have risen. To address this problem, the industry created a special interest group (SIG), called the Trusted Computing Group (TCG). The TCG's mandate was to specify, with input from all sectors of the computing industry, a comprehensive approach to enterprise security based on compatible technology building blocks. Members and non-members alike can use these technologies, which take the form of open specifications, royalty free to create products. The SIG format has proved to be a highly successful way to introduce new technologies to the entire industry on an open basis. Past successes include the PCI and USB specifications.

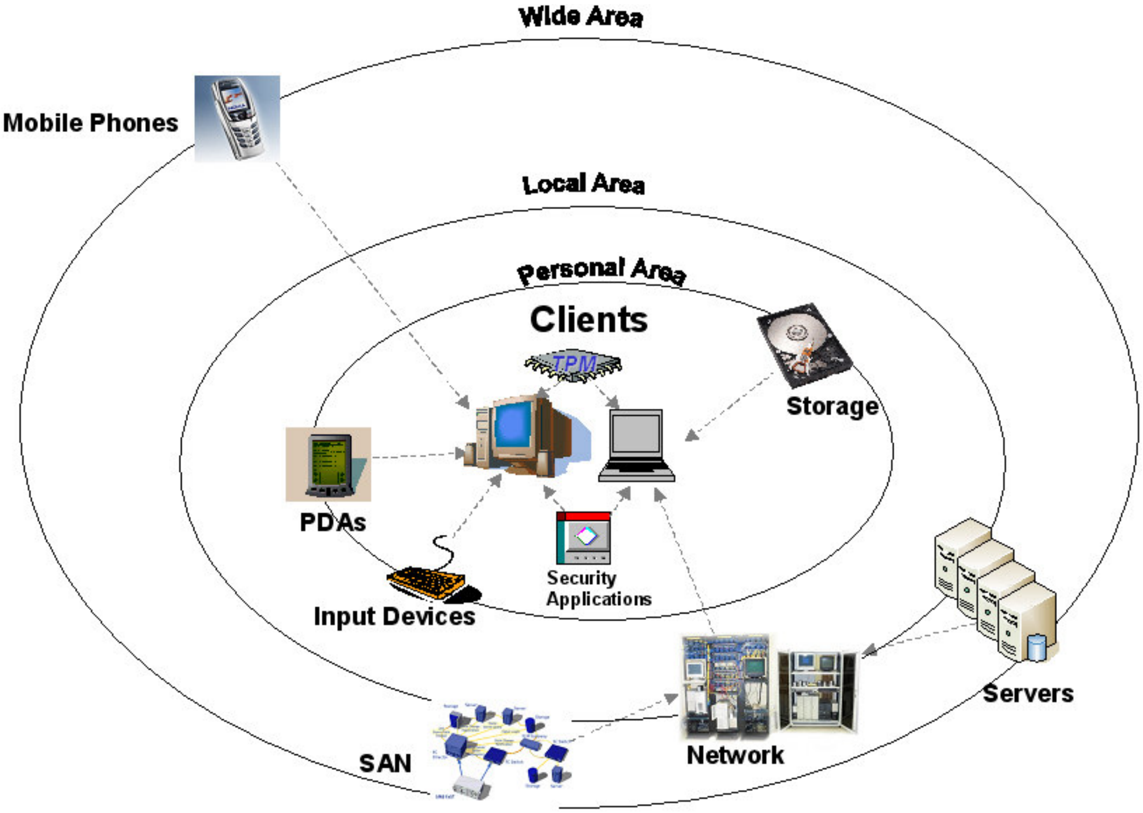
The tone of this group was set by IBM, which created the first dedicated security chip in the late 1990s. However, the nature of a universal secure-connection scheme requires that all nodes speak the same language. To gain widespread acceptance, IBM contributed its development work on the chip, now called the Trusted Platform Module (TPM), to the group.

The TCG's work is quite far along. The SIG has developed specifications for second-generation TPMs as well as specifications for accompanying software, client PCs, storage devices, mobile phones, servers, and secure network access. Multiple suppliers of the various security elements already have products on the market. TPMs are now built into many notebook computers sold to enterprises, and other implementations of security technology such as servers, network devices, and secure mobile systems are in the works.

Elements of a Trusted Solution

The following paragraphs lay out both the existing and planned elements of a complete security solution (Figure 3).

Figure 3: Secure Clients in Context



Existing Elements

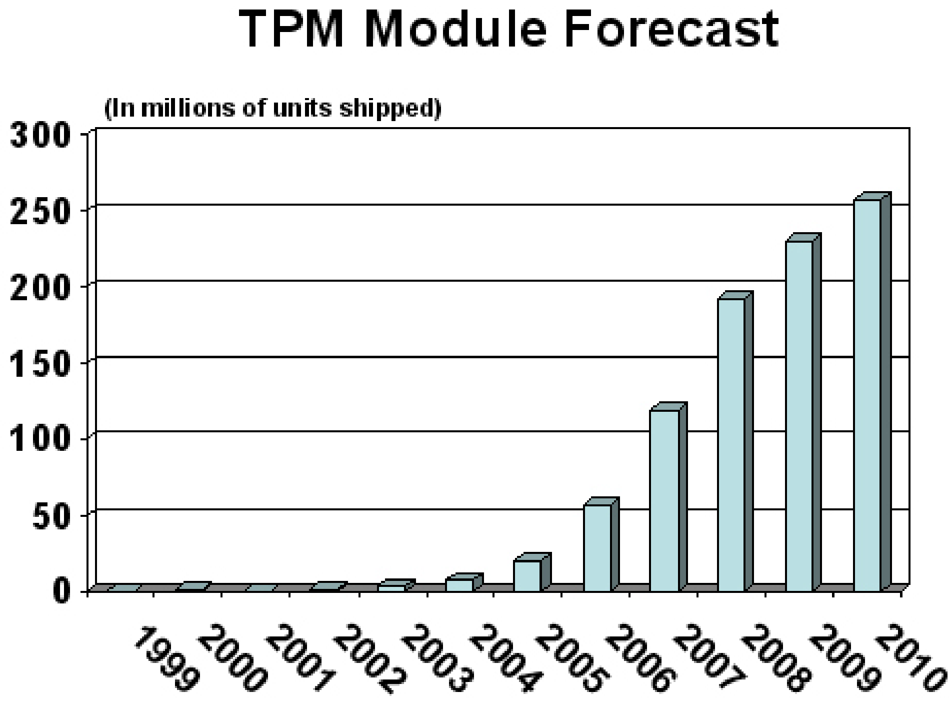
The following security technologies have already been specified.

The TPM

At the core of the trusted solution is the TPM itself. Currently embodied in a discrete or integrated component, the concept represents the clear superiority of security algorithms executed in hardware. The security itself is still software: keys used to encrypt data for storage or transmission. But the operations themselves are done in a closed hardware environment. nCipher, a security consultancy in Cambridge, England, proved definitively that even very good encryption was vulnerable to attack if it took place in the usual locations (i.e., in main memory, on the hard drive) because a good sniffer could detect the key by its highly entropic character; that is, a program that could distinguish a good random number from its surroundings could find the key and break the code. Thus, with all operations taking place inside the chip, no location exists where the key can be tested by intruders.

About 20 million TPM chips shipped in 2005, most of them in notebook PCs. By 2010, worldwide shipments of TPM modules in PC client systems will reach more than 250 million (Figure 4). However, as categories beyond PCs — e.g., mobile phones, storage systems, embedded applications, and peripherals — adopt TPMs, the total number of chips shipped could rise dramatically, even exponentially. Should TPMs become common in phones, for example, annual unit shipments could rise by hundreds of millions.

Figure 4: TPMs in PCs



Endpoints

The first specification for a device into which a TPM could be embedded was the endpoint or client — the desktop and notebook PCs at the edge of the network. Too long, these endpoints had been neglected as security efforts were focused on the network and the servers that contain an organization's most valuable information. But the whole network could be jeopardized by a single compromised client. An intruder could simply pose as a legitimate user and gain access to whatever rights and privileges that user had. Thus, one of the first standards set by the TCG was the endpoint or client security specification for PCs.

Although IBM originally targeted stationary systems (desktops), it rapidly became apparent that mobile systems (notebooks) represented a larger vulnerability. Desktop PCs could be locked to the desk and locked in the building after working hours, and generally only employees had physical access to them. Notebooks were much more vulnerable to theft or other breaches. Thus, implementation turned toward mobile clients fairly early on.

Implementation of endpoint security is fairly straightforward. Many commercial notebooks (and some desktops) come with the TPM chip embedded in them today. These systems also have client software, such as IBM's Client Security Software, that allows the end user to set up and use the system for authentication (matching a user to a machine at log-on), password management, and data and file encryption. Many of these systems now have fingerprint readers as well, and these readers are tied directly to the TPM via software for user authentication and convenient password management.

Applications

Software, too, is critical to the security environment. Security programs represent a special class of applications, which must not only be secure themselves, but also safeguard the integrity of the perimeter. Particular security applications that support the capabilities of the TPM include:

- Data protection solutions
- Data and file protection
- Secure document management and electronic signature products
- User managed credentials and auto-log-in products
- Identity protection
- Backup, restore, and migration administration
- TPM management tools and capabilities
- Network protection
- Enhanced email security
- Enhanced Web client authentication

Trusted Network Connect

Trusted Network Connect (TNC) is a specification that covers the dynamic relationship between clients and servers in the enterprise network. The essence of the TNC specification governs how clients request access to the network, and how a server grants that access. The protocol involves the packaging of a network request with information that positively verifies the user and the hardware as well as the state of the requesting device. A user can be verified by way of a password, fingerprint, or other method, including multi-factor authentication. A device can be verified by an algorithm that interacts with its TPM, producing a unique identity. The state of a device is measured by a number of factors, including whether or not is has up-to-date patches and virus definition files. If a request measurement falls below some threshold set by policy, access can be denied. If the measurement falls within a certain range, the device can be given provisional access, shunted to a quarantine area, and remediated before being let on the network. If the measurement exceed some threshold, access can be granted with no further delay.

TCG's specification for TNC is available and a number of companies are developing products to support it; some of these products are available now. TNC takes an open architectural approach so that users can mix and match products in their organizations.

Planned Elements

The TCG is working specifications for the following security technologies, and products based on these specifications will hit the market in 2006 and 2007.

Servers

In some sense, server security has lagged behind that of clients because, the common wisdom goes, servers are behind the perimeter defenses, and anything that gets to them has already been vetted. Also, server security protocols can affect performance, and any performance penalty at the server level is unacceptable.

However, servers need to be part of the security story, if for no other reason than that they should not be left vulnerable to a malicious intruder successfully masquerading as a legitimate client. Also, with newer systems, the effect of the performance penalty can be limited. Most servers have multiple processors, many have dual-threaded processors, and the latest have dual-core processors. These developments mean that security tasks can be executed by a single virtual processor while the rest are occupied with the main task.

TCG has released a specification for secure servers that defines how trusted servers are created, managed, and maintained. Trusted servers can store and protect digital keys, passwords, and certificates to handle:

- Asset management
- Configuration management
- Data migration and backup
- Distributed trusted computing

- Document management
- Financial transactions
- User and platform authentication

Servers are also integrated to Trusted Network Computing (TNC), TCG’s initiative for network access. When the client side asks for access, the server side, after assessing the client's trustworthiness, grants access or not, depending on the policy.

Storage

Storage represents a special class of devices with distinctive security issues and technology. For example, even if a PC is password protected, a thief can remove the hard drive and insert it in another PC as a slave device. Unless they are encrypted, all the files can then be opened.

Hard drives can be brought inside the perimeter by equipping them with an unchangeable hardware partition not visible to the operating system. This partition can contain the necessary keys to ensure that the drive is able to talk only to an authorized host over a secure communications link. Drive security applications include:

- Full-disk encryption
- Disk-erase enhancement
- Drive locking
- Forensic logging

In general, buyers purchase storage with the CPU, and so it is fairly easy to determine from documentation whether both main unit and storage subsystem have the required embedded security. TCG is in the process of finalizing an open specification for development of trusted storage for the enterprise. Prototype products have demonstrated that a trusted storage device simply won’t work if separated from its authenticated client. In the final specification, TCG will address all types of storage (direct attach, SAN, NAS).

Input and Output Devices

Other devices that need to be part of the chain of trust include the input and output devices normally associated with a PC. These devices include the mouse, keyboard, and potentially other devices on the input side, and displays, printers, and other devices on the output side. Data traveling to and from these devices can be intercepted if they are not inside the perimeter. Future devices will have built-in trust elements that interact with host-based TPMs.

Mobile Phones

Another obvious client platform is the mobile phone. With capabilities increasing such that they could begin to rival PCs, phones have a clear need for security. More value — in the form of both important data and actual financial transactions — is passing through phones. As data devices, phones now have not only a phone number, but a dynamically allocated IP address as well. Phones are becoming nodes on the network, yet to date there has been no organized effort to provide comprehensive security for them.

Just like other clients, phones (and their users) need to be authenticated and their integrity or "digital health" ascertained. Other applications of trusted mobile devices include:

- Protection of user data
- Protection of content and services
- Secure software download
- A secure communication channel
- Mobile ticketing
- Mobile payments
- Secure software usage

These operations can be achieved via the TPM capabilities and associated applications.

Get Started Now

Now is the time for the IT community to get started on trusted computing implementation, whether the first moves involve seeding the environment with TPM-enabled clients or rolling out a full-blown trusted-computing-based security solution.

How to Approach Trusted Computing

It is not necessary for an organization to implement comprehensive security all at once. Such an effort could be disruptive to the business and is also likely to be outside the budget in any given year. The good thing about TCG-specified products is that they can be layered on over time, each new layer adding a greater degree of security. The IT department can rest assured that the TCG has envisioned the entire solution from scratch and that the last element implemented will work as planned with the first one. Thus, the rollout can be gradual, implemented one step at a time. Already, many TCG-compliant hardware and software products are commercially available (see "Resources" for product availability).

The job of securing an enterprise can be overwhelming if viewed as a monolithic task. However, broken into parts, the undertaking is not so daunting. The order of implementation is important because some things depend on others being in place, and some are more critical than others.

We recommend the following steps in the order shown:

1. Authentication
2. Data protection
3. Network attestation and platform measurement
4. Application protection
5. Content protection

Authentication

This level of security is relatively straightforward. Authentication involves ensuring that a user is who he or she says he or she is, and that the machine requesting access is the machine that that user is supposed to be operating. Endpoints (desktops and notebooks) can be definitively authenticated if they are equipped with a TPM and related authentication software. We strongly recommend multi-factor authentication involving at least two factors, typically biometric (i.e., fingerprint) and password.

These days, an increasing number of commercial clients are offered with embedded authentication features, including a TPM chip, security software, and a fingerprint reader right in the bezel. These units can be set up for authentication right out of the box. It is also possible to retrofit existing clients with fingerprint or smart card readers, but, to be effective, the TPM must be embedded.

Authentication can also be applied at the level of the virtual private network (VPN), such that only an authenticated user-machine combination gains access to a corporate VPN. This capability is useful for mobile executives who must stay in close touch with home base while passing through territory with varying degrees of insecurity. VPN software can be tied to TPM software through the settings interface.

Data Protection

Once users are clearly identified and married to their machines, the next level, data protection, becomes possible. Data protection takes many forms, but in essence it involves ensuring that user (or customer) data remains inviolate. Protected data cannot be changed without an authorized user's knowledge. It cannot be lost. Unauthorized people cannot access it.

Data protection is more than just a matter of keeping the wrong people out of places they shouldn't be and not having valuable records disappear or morph. Data protection is driven by a host of new legal requirements that protect customer privacy. Organizations that fail to protect their data in the face of these new mandates become subject to legal action and potential severe financial loss.

The newly formed TCG Storage Specification covers secure storage on fixed magnetic media. Critical to data protection will be the secure linking of host CPU and hard drives. Compliant products are designed so that a hard drive cannot be accessed if it is removed from its host. In some cases, the data can be destroyed deliberately if it a hostile access attempt is made. With the CPU tied to the user, who has to authenticate to log in, the stored data is secure because of the trust link set up between the central unit and the storage device.

Since as much as half the intellectual property of a given organization resides on PC hard drives, battening down this particular vulnerability is critical. Data and file encryption are key technologies on which data protection is based.

Today, the flaw of data protection software is that its secret key sits in the registry, whether obfuscated or not. The key can be extracted from the registry with readily available software tools. However, several solutions now on the market use a TPM to protect the private key, a monumental improvement in data protection.

Network Attestation and Platform Measurement

The TCG has created an open specification, TNC, for network attestation and platform measurement. This level of security can be added onto the platform-level steps laid out in the previous two sections. Essentially, when a client requests network access, it is required to go through a handshaking sequence with the access-granting server, which makes use of data gathered from the client's TPM, the user's authentication method, and a "health check" status composed of information about the status of various client platform measurements. These measurements include the version of the virus scan engine and DAT file, firewall and other settings, and patch status.

Implementation of TNC moves the organization closer to a virtually secure environment in which the perimeter becomes less important, since any two nodes in the system can authenticate to each other and then conduct interchange over a secure link no matter how hostile the environment..

Application Protection

In Microsoft's soon-to-be-released operating system, Vista, applications will run under a new model in which users will operate under the least privileged mode possible. This schema will allow applications like Outlook and Office to run in protected areas of the system without the potential threat of invasion by malefactors. The code will be protected from intrusion by partitioning the system and having applications run in an area accessible only by a user with high level privileges (i.e., the administrator).

This system is viable because it rests on a foundation of hardware-based security and TCG protocols. The virtual partitioning is hardened by authentication, data protection, and network attestation layers. Thus, the applications are physically secure as well as logically secure.

Content Protection

Finally, new schema for digital rights management (DRM) will be able to operate on top of a highly secure platform. Content owners will be able to release their intellectual property with confidence, knowing that they will be paid for value rendered. Although the TCG has not specified a particular DRM protocol, its security elements are designed to be open and work with whatever DRM software is specified.

Because the "root of trust," the machine-level keys, are hidden in the TPM and are used to create a "chain of trust," higher-level keys that are each in turn encrypted with the keys below, many different content protection schemes are possible, each relying on its own key pair. The content owner can rest assured that only the authorized user will be able to open the content because the public and private keys, which are interrelated, are both required to complete the transaction. Content "signed" with a particular user's public key can only be opened with that user's specific private key. Thus, the direct relationship between the owner and user (or buyer and seller) is preserved.

Conclusion

IT managers should view security in terms of a computing ecosystem, a complete environment, additive in security and performance as TCG solutions are implemented. The TCG array of products allows IT departments to buy a solid security platform now and yet be able institute future security technologies as they become available.

The benefits of trusted computing are clear once the broad outlines of the technology are perceived: simpler, more robust, and more convenient security. These technologies represent the building blocks for a secure computing environment, but also one in which procedures are easy to understand and user productivity can be increased.

Resources

For more information on TCG platforms, go to www.trustedcomputinggroup.org.

Roger L. Kay is founder and president of Endpoint Technologies Associates (www.ndpta.com).