Governance of Trusted Computing

Clark Thomborson
Department of Computer Science
University of Auckland
cthombor@cs.auckland.ac.nz

Matt Barrett
Department of Computer Science
University of Auckland
mbarrett@auckland.ac.nz

Abstract

Trusted computing systems offer great promise in corporate and governmental applications. Their uptake has been very slow outside of the national security agencies for which they were developed, in part because they have been difficult and expensive to configure and use. Recent designs are easier to use, but some compliance and governance issues are unresolved.

Our analysis suggests that cryptographic systems, in order to be trustworthy in corporate environments, must support an audit of their most important operations. At minimum the audit record must reveal the number of keys that have been generated, as well as the creation times and authorities of these keys. This record of cryptographic activity must be tamper-evident, and must be open to inspection by the IT staff of the corporate owners as well as by their independent auditors.

Key Words: Trusted computing; Audit; Digital Rights Management.

1. INTRODUCTION

Security analysts use the word “trusted” in a way that may seem confusing to the non-specialist. Components are considered to be trusted only if their failure or misbehaviour might cause a violation of a security goal of a system. A security analysis is required to distinguish the components that must be trusted from the ones it is not necessary to trust. All subsequent analysis will focus on the trusted components. If any trusted component has a vulnerability which could be exploited in a foreseeable attack, then the security analyst will propose modifications to the component that make it more trustworthy.

The accuracy of the financial and operational records of many small corporations can be compromised by an attack on any one of their computers. Thus all of their computers are trusted components. Large corporations can afford more elaborate security mechanisms, minimizing reliance on individual computers. Even so, the theft of a chief executive’s laptop is likely to be a great cause for concern. A virus infection in a single desktop computer will threaten the availability of vital information systems, because connectivity may be impaired either as a direct result of this infection or indirectly by the preventive countermeasures taken to prevent its spread. We conclude that corporations are placing significant trust in most, if not all, of their servers, desktops, laptops, and handheld computing devices such as Blackberries.

Because of the diversity, complexity, and instability of corporate computers and their interconnections, it is infeasible to make an accurate assessment of the trustworthiness of the system. Due in part to this infeasibility, but also because of the large number of easily-addressed vulnerabilities that can be found in the information systems of a typical corporation, security consultants typically provide much simpler and more cost-effective services than a comprehensive security analysis. A common style of consultancy is to look for components with known vulnerabilities, such as unpatched servers. The typical analyst will then suggest mitigations which will improve the client's information security relative to the best practice for their industry segment.

The frequency and cost of corporate security incidents gives us strong incentive to move beyond a methodology of incremental improvement. To radically improve corporate security, we must greatly increase the trustworthiness of our information systems.

In this paper we explore the non-financial advantages and disadvantages, from a corporate governance standpoint, of the "trusted computing" features defined by an industry standards group (Trusted Computer Group 2004). A crucial component in this design, the Trusted Platform Module (TPM), is a hardware chip that is securely attached to the motherboard. Such TPMs are present on most corporate laptop and desktop computers, and they are specified for inclusion in other mobile computing devices such as cellphones, but these TPMs are not yet being used extensively.

One of the eventual benefits of shifting to a TPM-based design will be to provide secure digital rights management (DRM) for corporate data objects requiring either confidentiality (such as sensitive internal email), or integrity (such as contracts, invoices, and receipts), or both (such as confidential contracts with external suppliers). At present there are few technological controls over the creation, modification, and distribution of corporate data objects as soon as they are exported from the system in which they were created. The export operation may be as simple as an inclusion of a file in an email attachment. Everyone who handles such objects outside of their secure environment thus becomes a "trusted component" of the corporate security system. This wide distribution of trust precludes a robust security analysis, and more importantly it places a burden of diligence on employees whose time and attention could be more profitably spent on other matters.

The trusted computer architecture seems most likely to be embraced by both corporations and governments before it is accepted by consumers in their home computers. Our reasoning is as follows. Consumers expect to have complete control over the security policies in their own computers, whereas corporations and governments expect to control the security policies in all their employees' computers. Trusted computing allows external control of a computer's security policy, which is of direct interest to corporations and governments but not to home users. Certainly the secure kernel of a trusted computer architecture can not be under the complete control of its user-operator, otherwise the kernel could not provide the "mandatory access control" and remote-attestation features that are fundamental to its security model.

Governments expect to exert sovereign power over their internal operations, and over some aspects of their citizens' computer usage. Trusted computing is a double-edged sword from the governmental perspective, because some residual and surreptitious control might conceivably be exerted by the manufacturer of the trusted platform module or by the provider of the kernel of the operating system. Indeed, two governments have already developed principles about the usage of trusted computing and digital rights management technologies (BSI 2004, NZ SSC 2006).

Current designs for trusted computing could, we argue in this paper, be modified to provide sufficient assurance to large corporations (and to governments) that all security policies are in fact under their complete control. Small corporations do not expect complete control over security policies. Consumer adoption will likely be initially motivated by a "killer app" that requires a trusted computing architecture. We believe that a well-informed consumer would willingly cede some of their control to a trusted third party if there is sufficient watchdog oversight by a regulatory agency or a consumer-advocacy group. A suitable application would also drive initial adoption amongst comparatively uninformed consumers, unaware they are ceding control without any oversight. Thus we see scope for a modified version of trusted computing being acceptable to a wide range of corporations, governments, and home users.

Our argument above has reduced the problematic acceptability of trusted computing to a question of proper governance. Who should have ultimate control over the security policy of a trusted computer, and how can we place effective limits on this control?

2. TRUSTWORTHY SYSTEMS

A trustworthy system has a small number of trusted components, each of which is simple enough to be completely analyzed for its vulnerabilities. As a case in point, the Windows XP operating system can not be considered trustworthy, even after the "hardening" features introduced in its second major service patch, SP2. Because this operating system offers no highly-secure "sandboxes" in which it is completely safe to run possibly-hostile code, there are approximately 40 million lines of operating system code which must be trusted. Additionally, all web-browsers, email clients, and other common applications must be trusted unless they are disabled from receiving data from uncertified sources.

Users can lessen their trusting reliance on their operating system and communication applications by running virus and spyware scanners. These scanners provide important defences against known threats, but no security scanner can provide a strong assurance of security. Signature-based scanners are unable to keep up with the latest threats (Naraine 2006). Anomaly-based scanners can not distinguish, with sufficiently high accuracy (Axelsson 2000), between beneficent but strange-looking code and maleficent but normal-looking code (Cohen 1993).

The operating system Linux is somewhat better designed than Windows XP from a security perspective, because it has a well-defined kernel. Even so, the Linux kernel is far too large to be thoroughly analyzed and pronounced trustworthy. This line of reasoning leads us to a principal objective of most designs for secure operating systems, which is to implement all trusted functions of an operating system in a kernel of minimum possible complexity and size. Recent examples are Linux Kernel 2.6.0 and the not-yet-released Microsoft Vista. The integrity of the kernel can be assured by comparing its hash signature to a fixed value stored in some secure location. A special computer chip, called a "trusted platform module" or TPM, has offered this hash-signature kernel-checking function on corporate server, desktop, and laptop computers for the past few years. It's a curious situation. Even though this form of hardware support for more trustworthy computing is widely available, it is not yet in widespread use.

One reason for the slow adoption of more secure operating systems is that many device drivers, utility routines, and user applications must be redesigned to conform to the more restrictive security model of a secure operating system. This redesign was especially difficult in the highly-secure Unix variants of the 1980s and 1990s, such as LOCKix (O'Brien 1991). The redesign is also likely to be problematic in Windows Vista, because third-party software installations in earlier versions of Windows have been allowed to modify device drivers and even the kernel (Field 2006, Fisher 2006). The malleability of the Windows kernel, and its near-total lack of internal security boundaries, has in this author's view been one of its greatest assets in the past. Legitimate third parties have used these features to speed their development and installation of novel applications for Windows. Regrettably, but not surprisingly, illegitimate third parties have also taken great advantage of this openness. A lock-down process is painful and costly, but necessary for security. The lock-down of Windows has been ongoing since its earliest versions. Each major release, and most of the service packs, has closed down some of the design flexibility that was previously available to legitimate (and illegitimate) third-party software.

3. REQUIREMENTS ANALYSIS

In this section we develop a set of security features for trusted computing that would make it very attractive to corporations. Section 3.1 outlines static data security requirements, Section 3.2 outlines dynamic security requirements, and Section 3.3 outlines governance requirements. We will analyse Microsoft's IRM v1.0 along these three dimensions in Section 4.

3.1 Static Requirements

A system can be said to have static data security, if it maintains all three of the "CIA" properties (Lampson 2004):

- Confidentiality (data is only read by authorised parties),
- Integrity (data is only written by authorised parties), and
- Availability (data can always be read by, and written to, authorised parties)

In this paper we discuss security only in its broad generalities, although we realise that the "devil is in the details" of any security analysis. Every user of every system has somewhat different security goals. This diversity must be taken into consideration by a security analyst. At the generic level, however, the CIA taxonomy allows us to identify three categories of internal documents and three categories of externally-generated documents.

Corporate documents, by default, should have a primary requirement of integrity. This is an appropriate requirement for documents intended for external parties, such as official price sheets, external correspondence, web "presence", and advertising copy. Some internal documents, such as signed contracts and financial accounts, also have integrity as a primary requirement. Undetected and malicious modification of any of these documents could have significant repercussions including short term financial losses, legal difficulties in the medium term, or long term damage to a corporation's reputation. Any data that is accessed routinely in an operational setting has a primary requirement of availability, with integrity running a close second or possibly first-equal. A malicious "denial of service" attack on a mission-critical information resource can be extremely costly and disruptive, even if the downtime is brief.

A small percentage of corporate documents, such as strategic plans and trade secrets, have confidentiality as a primary requirement. Correspondence with external parties should, ideally, only be readable by the intended recipient, but the maintenance of this confidentiality requirement is only rarely of concern to the sender after a successful delivery has been made. Our analysis above suggests that most internally generated documents fall into our first two categories (of "integrity first" or "availability and integrity first"), in which confidentiality is a secondary requirement. These less stringent confidentiality requirements could be enforced sufficiently by privacy filters, by encrypting all correspondence under the intended recipient's public key, and by deterrents such as well-publicised random audits of employees who claim an operational requirement to access corporate databases and confidential documents.

Externally-generated documents form three additional categories. Documents signed by external parties, such as contracts and receipts, must retain their integrity and availability so that the signature remains non-repudiable. Unsigned external documents form a second case. Examples are email received from outsiders and data downloaded from the web. All of the documents in the second category must be open to inspection by appropriate authorities in the corporation, for at least the following reasons. Employees are not generally allowed to use corporate equipment to keep secrets from their management, and outsiders should never be provided with a secret storage area on a corporate computer unless carefully managed controls are in place. This availability requirement on externally-generated documents is one of four principles for trusted computing and DRM technologies recently established for e-government in New Zealand:

For as long as it has any business or statutory requirements to do so, government must be able to use the information it owns/holds [and] provide access to its information to others, when they are entitled to access it (NZ SSC 2006).

The supporting policies for this principle allow for an exceptional case, which is our third category of externally-generated documents. Non-owned documents may be held on governmental computers with access limited by a digital rights management system, but only if these access controls are well-notified and acceptable to the government. One condition on acceptability is an ability to detect and reject harmful communications:

Agencies will reject the use of TC/DRM mechanisms, and information encumbered with externally imposed digital restrictions, unless they are able to satisfy themselves that the communications and information are free of harmful content, such as worms and viruses (NZ SSC 2006; see also Garden 2003).

To summarise our analysis, the default category for internally-generated corporate data and services is "integrity first", with exceptions for operational documents ("availability and integrity first") and secret documents ("confidentiality first"). Externally-generated documents also fall into three categories, the default being an unsigned document ("availability first"), with exceptions for externally-signed documents requiring both integrity and availability ("integrity and availability first"), and for non-owned documents in cases where the corporation has agreed to accept the licensing restrictions ("externally controlled confidentiality").

Requirements will change whenever business conditions change. This implies that the static view of information security as developed above is insufficient, except for the most moribund of corporations. In Section 3.3 below we consider the procedures for changing a corporation's security arrangements, but first we discuss the dynamic processes which maintain the corporation's current security policy.

3.2 Dynamic Requirements

From a dynamic viewpoint, security is a process of maintaining a set of static requirements in the face of a changing environment. The required processes in a secure system are listed below. These are sometimes called the "gold standard", because *Au* is the Latin abbreviation for gold (Lampson 2004).

- Authentication is the process of verifying the identity of a user, a data object, or an executable object. For example, if a user is identified by login name, then the authenticator may be the corresponding password.
- Authorisation is the process of deciding whether or not a user or executable has the right (under the current CIA policy) to make a requested access to an object.
- Audit is the process of maintaining the history of a trusted system, so that subsequent analysis can reveal the frequency and intensity of security breaches, and to support a forensic investigation in cases where a severe breach is suspected.

In a traditional "gold standard" design, physical access to secure computer systems is guarded by a person who is trusted to authenticate everyone who seeks access. The guard's physical-entry logbook is an important component of the audit record. Surreptitious oversight of the guard's activity allows the veracity of their audit record to be assured, giving us confidence in the trustworthiness of the guard, as well as in the trustworthiness of computer system under guard. In some early targets for office automation, for example in accountancy systems, the cost of auditing was not seen as problematic for there was a pre-existing operational requirement for maintaining a complete record of transactions in the journal. Adequate physical security could be maintained for a small number of data-entry terminals, and for the “big iron” in the back room. In some other highly structured information systems, such as those developed for intelligence agencies, all data accesses could be authorised against a well-defined and stable set of security policies. Thus it was not inconceivable to meet the gold standard for information security in the corporate computing of the 1960s or 1970s.

Modern corporate information systems are rarely designed to meet the gold standard for security. Security generally runs a distant second to usability and feature-set in overall system objectives. Any insistence on 100% gold-standard security, in addition to these usability and feature set requirements, would result in an overly expensive “gold-plated” system. The required feature-set of the information systems in a modern corporation make it increasingly difficult to discover a defensible security perimeter. Removable storage devices can rapidly and surreptitiously retrieve, or inject, massive amounts of data through an unsecured USB2 port. Wireless communication devices such as laptops, handheld computers, and cellphones provide multiple modes of access, each of which must be secured. All of our portable computing devices, and the mobility that these devices enable, greatly limit the prospects for physical security, and greatly increase the risks of eavesdropping on data and theft of authentication credentials. Globally distributed corporate operations, and rapidly changing corporate partnerships and consultancies, make it difficult or impossible to maintain complete and up-to-date authorisation policies in traditional, centralised security architectures. In response to these difficulties, security chiefs at some major corporations such as Rolls-Royce, Boeing, Qantas, Procter & Gamble, and Standard Chartered Bank, have recently begun developing standards and concepts for de-perimeterized and micro-perimeterized security in the Jericho Forum. More trustworthy computing platforms are a key requirement for the realisation of the Jericho vision of enabling corporations

> *… to embrace the Internet and to securely exploit public infrastructure and services directly within the organization’s technology and business boundaries. Such a model would connect an organization and its business processes to all external stakeholders, seamlessly and securely, enabling employees, suppliers, and customers to collaborate anytime, anywhere, and at the lowest cost to all (Open Group 2006).*

The metaphoric responses of a security engineer to these security challenges are to apply an authenticating gold veneer at the security perimeter, to sprinkle auditing gold-dust uniformly but very sparingly over the most important security areas, and to place an authorising golden seal on all of the most important accesses.

3.3 Governance Requirements

Most information systems are in a state of flux, because of the instability of our technological and business environment. In poorly managed corporations, the changes are reactive and ill-directed. Pro-active change is well-directed, when corporate managers continually seek better answers to three questions of security governance (Lampson 2004):

- Specification, or Policy (answering the question of what the system is supposed to do),
- Implementation (answering the question of how to make the system do what it is supposed to do), and
- Assurance (answering the question of whether the system is meeting its specifications).

We would say that IT security is properly governed if, and only if, its managers set broad goals for each of these three activities, if they oversee progress toward the goals, if they arrange appropriate levels of financial and other resources, and if they intervene appropriately whenever progress is insufficient to meet the broader objectives of the corporation.

Our thesis, to be explored in the next section, is that current products for trusted computing and digital rights management do not allow proper governance. These systems do not fully support the static and dynamic security policies outlined in the first half of this section, and they do not fully support assurance procedures. Assurance is the only way a corporation's governors can assess, and then lessen, their reliance on required but untrustworthy systems.

4. ASSESSMENT OF CURRENT SYSTEMS

In this section we make a tentative assessment of Microsoft's upcoming Vista operating system and the latest official release, in Outlook 2003, of its Information Rights Management product IRM v1.0. Our assessment is tentative because we are working from incomplete information about the features to be included in Vista. We are assuming that a new version of IRM would take advantage of the trusted computing features of Vista. In future, we hope to gain Microsoft's assistance in conducting a more careful assessment. We also plan to investigate the DRM products which are designed for corporate use on a trusted Unix platform such as SELinux.

4.1 Static Assessment

We start our assessment at the lowest, static, level of data security. In Section 3.1, we had discussed the CIA taxonomy of static data security, identifying the key properties required for six categories of corporate documents. Confidentiality of documents is a strong point of Microsoft's IRM v1.0. Controlled documents are encrypted under a symmetric key which, after being encrypted under a rights-server's public key, is stored with the document itself as part of its rights management metadata (Garden 2003, Microsoft 2005). The server's private key is required to decrypt the document key, and traditional server-side security is (at least arguably) sufficient to maintain the confidentiality of the server's private key.

The trusted computer architecture provides secure hardware support for the asymmetric encryption requirements at the client, and at the server for that matter. Additionally, the symmetric encryption routines of IRM could be protected by the memory and diskfile protection features of a trusted operating system such as Vista. The client's connection to the server, and vice versa, could be secured by the remote attestation features of Vista, whereby each platform can be assured of the other's identity and that it was securely booted i.e. not compromised. Of course there may be defects in the cryptographic protocols of IRM, or in the key-generation routines (Garden 2003), so some third-party assurance is required. However procedures for a third-party assurance of the TPM are already being developed at Bochum University. Overall we see little cause for concern with the confidentiality requirements for documents that are in formats supported by IRM.

The integrity of documents is apparently controlled, under IRM v1.0, only by symmetric encryption (Microsoft 2005) and not by the usual technique of a digital signature or cryptographic hash. Encryption is arguably sufficient to ensure the integrity of a document, because any change to a well-encrypted file is overwhelmingly likely to result in a substantial (and therefore easily recognizable) amount of unreadable gibberish. Furthermore, it is a simple matter to introduce a CRC or other error-detection code with the plaintext of the document, so that any unauthorised modification of the encrypted document can be very quickly and safely detected. We therefore see little cause for concern with the integrity requirements on corporate computing, under the assumption that at least one of the formats supported by IRM/Vista is appropriate for long-term archival storage, *e.g.* a compressed bitmap, PDF/A, or other well-documented and stable format.

The availability of documents is problematic under IRM v1.0, because all controlled documents are stored in encrypted form (Garden 2003). Master keys are held at the rights management server. This server could be a single point of failure, in a denial of service attack on document availability. The loss, or temporary unavailability, of a server master key would make it very difficult to read any of its slaved documents, except in cases where the document has been accessed recently. Slaved symmetric keys on recently accessed documents might be retrieved from the temporary keystore on a client workstation, but this would not be an acceptable defence against a sustained denial of service attack. A full key-escrow service, or a fully redundant rights management server, might be an appropriate (but expensive) mitigation to the risk of a denial-of-service attack on the server.

A more problematic availability requirement is that all protected documents must be secured against malware threats, as noted in Section 3.2. We cannot see any reasonable mitigation, other to require that all documents be made available to an offline malware scanner (Garden 2003). The threat scenario is that encryption applied by IRM might have occurred at a time when some malware in a source document was not recognised as a threat. A subsequent decryption may occur at any time, possibly many years later. In addition, malware scanners can only look for common threats. Any online scan that is invoked immediately after an IRM-controlled document is decrypted for reading must be extremely rapid, to avoid inconveniencing users. In the exceptional case of a document whose confidentiality restrictions are controlled externally, a corporation might be willing to accept a relaxed condition on availability at the cost of a great decrease in usability (due to the requirement for a stringent online scan and/or a carefully sandboxed execution) and a concomitant increase in system complexity and cost.

In summary, our static assessment suggests that IRM on a Vista platform could meet a typical corporation's confidentiality and integrity requirements, but not its availability requirements. Most corporate documents have a strong, possibly overriding, availability requirement. For this reason it seems to us that IRM/Vista is unlikely to find much market acceptance until its implementation is modified to allow stronger guarantees of availability. We believe that the level of availability, integrity and confidentiality required for most corporate operations could best be obtained by encrypting most documents under symmetric keys that are shared widely through the company. This is in contrast to the existing IRM v1.0 design, which stores all symmetric keys on a central rights management server. In a trusted computing-enabled environment, it is reasonable to assume that all computational platforms are trustworthy enough to retain adequate key control, as well as to generate an auditable record of document access. Our suggested approach prevents any temporary or permanent loss of availability of the rights management server from affecting the availability of most corporate documents, although it does weaken guarantees on their confidentiality, We are thus confident that DRM systems and their trusted-computing hosts could be designed to meet the static requirements of corporate document management.

4.2 Dynamic Assessment

In Section 3.2 we summarised the dynamic requirements on corporate IT security with a metaphor. The system must have "an authenticating gold veneer at the security perimeter, … auditing gold-dust [sprinkled] uniformly but very sparingly over the most important security areas, and … an authorising golden seal on all of the most important accesses."

One of the design goals of any trusted computing system is to enforce an authenticating perimeter with a secure kernel login. This goal is typically addressed by designing a trusted path from one or more user input devices such as a keyboard or a biometric reader to the secure kernel, and another trusted path from the kernel to some user-visible output device such as a reserved area on a display terminal. If users are trained to recognise spoofing attacks, so that they will very rarely (if ever) attempt to login using untrusted I/O paths, then we believe that the identity of users can be acceptably authenticated on a trusted computing platform such as Microsoft's Vista.

Other portions of the security perimeter are somewhat more problematic, especially when one is trying to limit outward information flows as is required to support a confidentiality requirement. However our static analysis has suggested that confidentiality requirements are of less importance in corporate information systems than are the integrity and availability requirements. In this respect the trusted computing architecture is suited to corporate requirements. Its emphasis on an authenticated boot, and on maintaining the integrity of the operating system, should greatly improve the integrity and availability of corporate documents even though it can do little to mitigate many types of intentional breaches of its security perimeter by an authorised user. For example, someone who is authorised to read but not to copy a document could use a cellphone's camera to take a picture of a computer display showing this copy-restricted document. However an authenticated boot is a very important security measure, for it greatly reduces the risk of automated and externally controlled breaches. Covert modifications of the operating system and associated libraries may be prevented entirely. In combination with an appropriately secured desktop environment, trusted computing technology should greatly reduce system vulnerability to external attackers, and should make it much more difficult for an insider to mount an automated attack.

The security perimeter of a trusted platform includes its network interfaces and its removable storage devices. Any trustworthy implementation of a DRM system on a trusted platform would require cryptographic authentications across such interfaces. Incoming data can be given meaningful integrity characteristics if is cryptographically signed by its source provider, by the receiving device, or by both. Confidential data must not be released on an outbound channel unless there is a trusted path to a trusted recipient, so that the data can be encrypted with the recipient's public key (or with some mutually-agreed symmetric key). All of these requirements can be implemented on a trusted computing platform, so we see little cause for concern with authentication in a well-designed DRM system.

We also see little cause for concern with authorisation processes. Even though we have very little knowledge of how these are implemented in Vista or in IRM, we imagine that the storage mechanisms for access-rights information of an IRM port to Vista will have sufficient integrity to preclude adversarial modification by any but the most skilled of outsiders. Insider attacks are always more difficult to counter, but a digital rights management system for corporate use should make some provision for double-signature controls on access rights to very important documents. These provisions should rely on some trustworthy double-signature authorisation process which must be followed whenever the security posture of any Vista platform is loosened. We have no knowledge of any double-signature authorisation processes on current designs for IRM or Vista but it would not require "rocket science" to develop them. Our main concern is with the audit provisions in IRM and Vista. Our preliminary analysis suggests poor audit support is a critical weakness. We understand that current-generation TPM chips can create a secure, cryptographically-signed audit record of their activity. We see no indication that Microsoft, or any other developer of trusted computing systems, has as yet attempted to use this audit facility to construct a trustworthy audit record of all the key-generation activity of a TPM. A trustworthy audit record is the only way for corporate governors to be assured their security goals are being met and enforced by the trusted computing architecture. The opacity of the security features provided by a TPM necessarily prevents inspection during a security audit of the system in a production environment.

We believe that few corporations would wish to permit surreptitious cryptographic activity to occur on any platform under their control. A complete record of key-generation by the TPM, as well as a random sampling of its key use, would thus seem to be an appropriate and technically-feasible audit objective. Additionally, we would expect the audit log of each client in a DRM system to record a random sample or, depending on volume or document importance, a complete log of its document accesses. The rights server would also maintain a log. Although

IRM v1.0 seems to have no provision for creating such an audit record, it is not unreasonable to presume that this facility would be included in some future IRM version on Vista.

A prudent response to the availability issues caused by encryption is some form of key-escrow service. Key-escrow is the practice of off-server or off-site backup of all encryption keys in use by the corporation. This reduces the risk of key loss resulting in the subsequent loss of all encrypted documents. All keys generated by the corporation must be stored by the key-escrow service for the practice to be effective. A performance audit of the key-escrow service would rely on the key generation record of the trusted computing architecture for assurance that all keys are being escrowed.

One complication in the audit requirement is that corporate governors may require that their PCs be capable of booting more than one trusted operating system during the lifetime of the hardware. Absent from this requirement, governors may feel unduly constrained by "vendor lock-in". Switching between operating systems on a TPM would invalidate its audit record, unless the record is maintained in a standardized way under both operating systems. This is, we believe, a novel requirement on the software portion of the trusted computing base.

It is possible that some corporate governors will not be very worried about "vendor lock-in" on their computing hardware, so Microsoft may be able to market a locked trusted platform which would maintain an adequate audit record of its key-generation activity. However we believe the purchasers of DRM systems would benefit if the Trusted Computing Group were to lend its support to the development of an audit standard for recording key-generation activity of the trusted computing base. Corporate governors might then reasonably require all platforms to adhere to this standard. We are currently considering starting a new forum in the Open Group, or broadening our current work with the Jericho Forum, with the goal of developing system-level requirements for this audit standard. These requirements will become implementable, we believe, if we receive support from the Trusted Computing Group, because some modification to the software component of the trusted computing base will be required.

In summary, our preliminary dynamic assessment has revealed little cause for concern with authentication and authorisation. Audit seems problematic for open systems until changes are made to the design of the trusted computing base, but we are optimistic about the prospects of developing a set of standards that would meet the requirements of corporate users, as well as addressing the concerns of governmental users and regulators (BSI 2004, NZ SSC 2006).

4.3 Governance Assessment

At the governance level, our concerns are with specification, implementation, and assurance. Corporate governors rarely, if ever, find themselves in the enviable position of having an implemented system which meets all desirable requirements. Furthermore governors rarely have an accurate idea of which requirements are not being met, and how often these failures occur. A typical approach to governance is one of constant amelioration. Governors ask their staff to develop a range of feasible possibilities for change, with an estimate of costs and benefits. Then the governors are in a position to make changes that are likely to be more desirable than the status quo.

Periodically, governors require assurance that requirements are being met so that they can prudently certify compliance with any relevant governmental regulations in their jurisdiction. A statement on the security position of a corporation making use of a trusted computer architecture must be derived in part from an audit of the activity of the individual TPMs in the corporation. If such a statement is required to show compliance with a specific regulation, these assurance procedures rely heavily on the presence of security audit records discussed above. Our preliminary analysis of Vista/IRM has identified auditability and availability as the primary areas of tension between the desirable requirements and the available implementations. Of these two areas of tension, auditability seems the most problematic, because a lack of auditability implies that the system is ungovernable. By contrast, a failure to meet an availability requirement might be regrettable, but acceptable, in many non mission-critical applications.

5. CONCLUSIONS

We have sketched a desirable set of requirements for corporate use of trusted computing and digital rights management technology. Our preliminary analysis of Microsoft's recent and likely forthcoming products strongly suggests that meeting our proposed requirements will become feasible, after some changes are made to product design. Developing trustworthy computing platforms and digital rights management systems are monumental undertakings. For this reason we do not expect to see more than a few viable product offerings. Already we have seen some massive but instructive failures, for example by InterTrust and MediaSnap (Stamp 2006).

We have argued that it would be in the best interests of large corporations to join forces with the few governmental agencies who have started to develop standards in this arena (BSI 2004, CEC 2004, NZ SSC 2006). We consider a standards-led approach to be more important in trusted computing than in many other areas of computer technology. Networking failures due to differences between vendor implementations of a specific protocol, for example, will not render data permanently inaccessible. However interactions between trusted computing architectures from different vendors may give rise to behaviour that renders encrypted data permanently unavailable. In addition, it should be clear that while an intra-corporation trusted computing architecture could be proprietary or unique, inter-corporation architectures must be standards-based so that usage-rights restrictions and agreements are universally understood.

A standards-led approach, especially if it is controlled by the purchasers rather than the vendors, is likely to maximize interoperability, minimize the appearance of incorrect implementations, and minimize costs to the purchaser. Our analysis suggests that large corporations would find it in their long-term interest to finance the development of an audit standard for TC/DRM implementations. An appropriately drawn standard would provide assurance for internal governance, and define a "safe harbour" for compliance with the regulatory requirements of external governors. Our initial analysis suggests that governmental agencies have many similarities to large corporations, in their desired uses and security requirements for TC and DRM. For example, both corporations and governments will want strong guarantees on availability and integrity on important documents they receive from external sources. Although there is a fundamental tension between governmental regulators and those who are regulated by government, the confluence of interest in their internal uses of these technologies should make it possible for large corporations and governments to work together very productively, to define a unified set of purchaser's requirements for these technologies.

REFERENCES

Axelsson, S. (2000). The Base-rate Fallacy and the Difficulty of Intrusion Detection, *ACM Transactions on Information and System Security*, 3(3): 186-205.

Bundesamt für Sicherheit in der Informationstechnik (BSI 2004), Federal Government's Comments on the TCG and NGSCB in the Field of Trusted Computing, 8 pp., March 2004. Retrieved from: http://www.bsi.bund.de/sichere_plattformen/trustcomp/stellung/_2a_e.pdf, 19 October 2006.

Cohen, F. B. (1993). Operating System Protection through Program Evolution, *Computers & Security* 12(6): 565-584.

Commission of the European Communities (CEC 2004). Management of Copyright and Related Rights in the Internal Market, COM(2004) 261 final, 19 pp., 16 April 2004. Retrieved from: http://eur-lex.europa.eu/LexUriServ/site/en/com/2004/com2004_0261en01.pdf, 19 October 2006.

Field, S. (2006). An Introduction to Kernel Patch Protection. In *windowsvistasecurity*, 11 August 2006. Retrieved from: http://blogs.msdn.com/windowsvistasecurity/archive/2006/08//695993.aspx, 19 October 2006.

Fisher, D. (2006). Vista Kernel Limits Have Security Vendors on Edge. *SearchSecurity.com*, 11 August 2006. Retrieved from: http://searchsecurity.techtarget.com/originalContent/0,289142,sid14_gci1210002,00.html, 19 October 2006.

Garden, J. (2003). Review of Microsoft Information Rights Management V1.0: Report to the E-government Unit of the State Services Commission, 87 pp., December 2003. Retrieved from: http://www.e.govt.nz/policy/trust-security/irm-200202/irm-report.pdf, 19 October 2006.

Lampson, B. W. (2004). Computer Security in the Real World. *IEEE Computer* 37(6): 37–46.

Microsoft (2005). Technical Overview of Windows Rights Management Services for Windows Server 2003, 29 pp., updated April 2005. Retrieved from: http://www.microsoft.com/techinfo/overview/rmenterprisewp.mspx, 19 October 2006.

Naraine, R. (2006). Alert Raised for MS Word Zero-Day Attack, *e-week.com*, 19 May 2006. Retrieved from: http://www.eweek.com/article2/0,1895,1965042,00.asp, 19 October 2006.

New Zealand State Services Commission (NZ SSC 2006). Trusted Computing and Digital Rights Management Principles and Policies, Version 1.0, 32 pp., 5 September 2006. Retrieved from: http://www.e.govt.nz/policy/tc-and-drm/principles-policies-06/tc-drm-0906.pdf, 19 October 2006.

O'Brien, R., Rogers, C. (1991). Developing Applications on LOCK. In *Proc. 14th Nat'l Security Conf.*, Washington DC, pp. 147-156.

Stamp, M. (2006). *Information Security: Principles and Practice*. Wiley: New Jersey.

The Open Group (2006). Jericho Forum: An Overview and How to Get Involved, 8 pp. Retrieved from: http://www.opengroup.org/tech/JF0602.pdf, 19 October 2006.

Trusted Computer Group (2004). TCG Specification Architecture, Overview Specification, Revision 1.2, 54 pp., 28 April 2004. Retrieved from: https://www.trustedcomputinggroup.org//IWG/TCG_1_0_Architecture_Overview.pdf, 19 October 2006.