Happiness in Slavery: Trusted Computing, Law Enforcement, and Power Relationships

David "Novalis" Turner
novalis@fsf.org

Free Software Foundation
59 Temple Place – Suite 330,
Boston, MA 02111

Abstract. I describe the capabilities of Trusted Computing (TC) architectures as embodied in the Trusted Computing Group and Next Generation Secure Computing Base standards. TC advocates correctly claim that TC technologies provide only new capabilities, and do not themselves contain any constraints on user behavior. However, TC enables technologies that, should users accept them, will restrict their freedom. TC also fails to provide the expected user-level benefits, such as increased data privacy. Data privacy, like non-TC so-called "Digital Rights Management", is an area where technical constraints fail unless coupled with vigorous law enforcement – yet vigorous law enforcement works on its own. It may seem that having more capabilities is always better. In the case of TC, however, users may benefit from not having such capabilities. Voting provides an analogy. Technologists need to educate non-technical users about how life in a TC world will be different, so that they will oppose TC. If this education effort succeeds, TC will ultimately fail, because non-TC systems match users' desires better.

Keywords Trusted Computing, law enforcement, TCPA, TCG, NGSCB, privacy, power imbalances, remote attestation, DRM

1 How Reduced Choice Can Lead To Increased Freedom

Laws are generally aimed at forbidding behavior that others would object to. For example, laws forbidding theft are not enacted for the benefit of the thief, but for the benefit of his or her victims. Individuals usually want the widest possible legal freedoms for themselves. Similarly, users will be unlikely to reject a new technology which gives them new options, and, by default, comes with no constraints. However, paradoxically, sometimes constraining one's own behavior gives one more freedom.

Consider a proposal to make voting mandatory in Afghanistan (with "abstain" as an option on all ballots). This would reduce everyone's individual liberty. Anarchists (or other supporters of non-democratic systems) would not allowed to make a political statement by not voting. Busy, apolitical people would

be forced to waste their time by voting for something they don't care about. However, in Afghanistan, women are often discouraged or forbidden from voting by their husbands[1][1]. So, this proposal would increase the liberty of Afghani women, because it would make it significantly more unlikely for their husbands to prevent them from going to the polls (since it would result in the arrest of the women). So, Afghani women ought to support this proposal if they fear that they won't otherwise be allowed to vote – by reducing their choices, they increase their freedom.

Trusted Computing[4] (TC) is a technology that appears to increases personal choices by allowing users new techniques to improve their security and privacy. However, the ultimate effect will be to be to reduce liberty via network effects, lock-in, and power imbalances. The rest of this article will explore these effects.

2 Introduction to Trusted Computing Technologies

Trusted Computing is the common name for a family of technologies sharing common functionality. There are two major standards for TC: Trusted Computing Group (TCG, formerly Trusted Computing Platform Alliance), and Microsoft's Next-Generation Secure Computing Base (NGSCB, formerly Palladium). TCG 1.1 is implemented in some IBM laptops[5]; NGSCB has not yet been implemented. The fundamental operations of TCG are authenticated boot and remote attestation. TCG 1.2 also has some support for I/O authentication. To these, NGSCB adds curtained memory.

A brief outline of the TCG trusted boot and remote attestation process follows. Some unexciting details have been omitted for brevity. Proof of the security of the system against various attacks is provided in the literature. Although they are designed to with stand various physical attacks, TC systems are still vulnerable, given sufficient time and money. The following discussion will assume that hardware attacks are impossible; they will be discussed in a later section.

A TCG system contains a Trusted Platform Manager (TPM): a chip containing at least 16 Platform Configuration Registers (PCRs), one unique private key, and a fixed set of cryptographic functions. The TPM is not a generic cryptographic co-processor. It performs cryptographic operations internally for security, rather than speed. The term "measure" is used by TCG to mean "cryptographically hash." A TCG system also has a Trusted Software Stack (TSS), which consists of portions of the bootloader and operating system which are critical for security.

Each PCR is a 160-bit, volatile storage location inside the TPM. PCRs have three major operations: Extend, Quote, and Seal/Unseal. Extend is the only way to write to PCRs (except by Resetting them to their initial values) and operates according to the algorithm:

$PCR_i \Leftarrow SHA1(PCR_i || input)$

[1] The number of women registered has increased since [2], but fraud is suspected, and voting may still be coerced[3]

Quote returns a copy of specified PCRs, signed by the TPM's private key (or some key authorized by that key). Seal encrypts data (usually itself a key) such that the TPM will only to decrypt (Unseal) the data when specified PCRs hold values identical to those specified at Sealing time[2]. Sealed data is also specific to a "platform" (individual TPM), although migration functionality is available.

In addition, many other TPM operations depend on the values of PCRs to determine whether the system is in a certain state. State is defined by the set of loaded software and configuration.

2.1 Trusted Boot

Trusted boot is a way that a system can ensure that it is a known state.

1. The PCRs are initialized to some known state. In addition, a number of other initialization functions, mostly related to key management, are performed.
2. The TPM measures the first boot stage and extends PCR0 with the result.
3. The BIOS runs first boot stage.
4. The first boot stage measures the second boot stage, and extends PCR1 with the result.
5. The first boot stage runs the second boot stage.
6. This process is repeated until the operating system is loaded.

At the end of this process, the system knows what state it is in by reading the PCRs. There is no Internet connection required for any part of the trusted boot.

Trusted boot provides security benefits on its own. One mode of attack is local (or remote) replacement of operating system or application files with modified versions with backdoors. The trusted boot process prevents this, by sealing files with PCRs matching the original operating system and application. A modified operating system, or a different application, could not unseal those files, because lower-level loaders would generate different PCRs. Of course, flaws in these systems themselves (for instance, a security hole which allows remote access to the data in memory) could allow confidential data to be revealed or altered.

2.2 Remote Authentication

On top of authenticated boot, TCG provides remote attestation of the system's state. So, a TCG system can prove what software it is running to other computers via the Internet. It can refuse to reveal this information, but it cannot lie. Here is the process, in brief:

1. A remote computer requests platform state, and sends a nonce (to show that the authentication is online).

[2] Note that the values specified for Unsealing need not be the same as those present at Sealing time.

2. TPM signs the concatenation of that nonce and whatever PCRs are requested, with some identity key which has been signed by the TPM's public key. That public key, in turn, is signed by a certificate authority (CA) which vouches for the state of the platform when the identity key was created, and for the TPM's compliance with the TCG standard.
3. The TPM sends this signed data, along with the signatures from various trusted third parties which verify that the implementation of the TPM matches the specification and is reasonably secure against physical attacks.

The remote computer now compares the PCR states with a list of states certified by software providers or independent certification agencies. If some of the PCRs are in an unlisted state, then the system is running software unknown to the requester. The requester may choose not to distribute information to systems which are in unknown states.

When the requester does know the platform's state, it has an expectation of how the platform will treat information it is given. The requester can then send information to this platform, which the platform will "seal". Sealed data is encrypted with a symmetric key, and this symmetric key, along with a list of current and target PCR values, is encrypted with some key which has been encrypted by keys in a continuous chain going back to the private key. The TPM will refuse to decrypt the symmetric key unless the target PCRs stored with it match the current PCRs.

This functionality has a few interesting details:

1. It's still vulnerable to software attacks against poorly-written software. So, if someone can convince the system to execute arbitrary code without updating the PCRs, they have full access to anything sealed to those PCRs.
2. Online attestations can restrict offline access to data.
3. Hardware breaks on one system only allow access to data sent to that system. The only "universal breaks"[3] are those which secretly compromise platform signing keys, or break one of the cryptographic algorithms that TCG depends on (mainly SHA1 and RSA).
4. The system does not directly restrict your actions. Contrary to popular belief, TC systems don't prevent alternative software from running on general purpose computers. The mechanism of influence is more subtle, and will be discussed further later.

3 Escalation

"Information wants to be free." -Stewart Brand[6]

Remote attestation is a building block for so-called Digital Rights Management (DRM) technology. But how did we get to the point where computers have a special chip just to keep people from copying data?

[3] "universal breaks" are those which affect all platforms of a certain type, rather than just one. Of course, all breaks which expose confidential data are universal with respect to that data, since once exposed, it can be copied freely.

The history of software from about 1976[7] is a narrative of copying and attempts to prevent copying. Non-possessors of software wish to possess it. So, they copy it from possessors. This violates copyright law, but is undetectable, since it's all on floppy disks and local dial up bulletin boards. Software publishers then try to restrict copying technologically. The process is escalatory; for every improved copying technology, there is an improved anti-copying technology or law. Bitstreams which can be accessed can be copied; most attempts to restrict copying restrict access to bitstreams.

Generally, restrictions on anything depend either on the scarcity of some resource, or on police powers. The reason there is so much spam is that there is effectively no limit on the resources (electricity, computers, bandwidth) needed to send it, and that there are no international laws against it. The IETF's Sender Policy Framework [8] for e-mail relies on the scarcity of domain names (a scarcity enforced by the financial cost of domain names) to reduce spam. Other proposals, such as Reusable Proofs Of Work (RPOW) [9] depend upon the scarcity of computational cycles.

Trusted Computing creates scarcity by cutting off TPMs which have been broken. The limitation is that people who wish to share media in spite of its copyright holders wishes must constantly acquire new TPMs (which, with properly designed TPMs, entails new motherboards at the least).

Given the failure of anti-copying technology, proprietary software companies come up with four ways to stop copying:

1. Software-based prevention techniques
2. "Lite" versions
3. Essential network services
4. Disgruntled employees

Software-based techniques are doomed to failure. A program's only means of knowledge about the world is that data it receives from the computer itself. A program's beliefs about the world can be altered by altering the data that the program reads, or by altering the program to read different data, or to ignore the data it reads. Included in this category are dongles, hardware tokens which are difficult to duplicate. The software which verifies their presence has the same fundamental flaw as pure-software techniques. Despite this, there is a constant escalatory war of technology between would-be-copiers and software publishers.

"Information wants to be expensive." -Stewart Brand[6]

Since it was nearly impossible to reduce the supply of copied software, software companies attempted to reduce demand. Some companies released less functional versions of their software for reduced prices or for no cost at all. Examples include Microsoft Works, Autocad Lite, and the shareware version of Doom.

In the games market, software publishers have also turned to essential network services. Some games are exclusively or mainly played via a centralized network. Examples include Battle.net, and The Sims Online. This network can

easily restrict access to people who have purchased the software by giving each purchaser a unique id and cutting off ids which are used by two people at the same time.

In parallel with this effort, many software publishers who supply mainly corporate markets solicited reports of illegal copying from disgruntled employees. Home copying remains undetectable, but the larger corporate market now must weigh the risks of being caught against the savings from not paying for software[4].

In software, increased enforcement has been significantly more successful in ending illegal copying of software than technical measures ever were. According to the Business Software Alliance, it's not skill at breaking access control systems which causes illegal copying, but lack of law enforcement [10]. This isn't surprising, since in the pre-Trusted Computing world, all breaks are "universal breaks".

It is only recently that music and movies ("media") have been copyable digitally and losslessly. Unlike software, illegal copying of media cannot be found by soliciting reports, since while many people are willing to report on their employers, few are willing to report on their friends. Media producers have tried the same things that software producers tried in the 1980s: that of ever-increasing technological barriers to copying. The music industry has tried a number of methods[5] to prevent CDs from being copied. However, since the data can still be read, it can be copied. The movie industry has also tried a few methods, such as the original Macrovision and the Content Scrambling System. All of these methods except CSS are based on somehow mangling the media's table of contents or other metadata, so that readers will play them correctly, but will not copy them correctly. Early software publishers tried similar techniques, and they didn't work then, either.

CSS represents a move towards cryptographic techniques which has been mirrored in other recent technology, such as Apple's iTunes. Unsurprisingly to people who have watched the software copying battles, none of this technology has done anything to stop or even slow illegal copying. As a result of these failures of technology, media companies have launched a three prong counterattack: additional legislation, legal attacks on the mode of distribution, and legal attacks on individuals.

In the late 1990s, the Internet exploded in popularity. Computers became sufficiently powerful to store and play compressed music. At colleges and other venues, local networks of computers allowed people to illegally copy music on a larger scale than the merely local CD copying. With the creation of Napster in 1998, the term "file sharing" came to refer to illegal copying via the Internet.

---

[4] An economist examining this situation might propose a black market in insurance against illegal copying lawsuits.

[5] Examples include SafeAudio, Cactus Data Shield, SunnComm MediaMax, and Key2Audio.

The first shot in the war on file sharing in the United States[6] was the NET Act of 1997[11], which provided stiffer penalties and federal enforcement for illegal copying. The NET Act's penalties are based on the price of the infringed data, rather than the profits of the infringer. This opened the door to prosecution of non-commercial copiers. However, few cases have been brought under the NET Act[12].

In 1998, the notorious Digital Millennium Copyright Act (DMCA) was passed. This gave technological access control measures the force of law. Media companies will encrypt their media. Every break is still a universal break, but now companies can sue those who break their copy prevention schemes. The intent of the law is to stop the release of content by deterring people from breaking access controls.

From the perspective of preventing illegal copying, this has been stunningly ineffective. A few cases have been brought against producers of software for accessing DVD content, but this has not slowed copying of DVD content[7]. The software needed to copy DVDs is widely available, and is likely to remain so.

Since the DMCA passed media companies have lobbied for an alphabet soup of new federal laws along the same lines: SSSCA, CBDTPA, Pirate, and INDUCE, PDEA. None of these have passed. On the state level, six states have passed[13] "Super-DMCAs", which create DMCA-like rules on the state level.

Since attacking the initial source of illegally copied code has proven ineffective, media companies have attacked the modes of distribution. In 1999, Napster was released. The design of this first filesharing network was primitive – clients submitted lists of files to a central server, which fielded search requests.

Despite its many weaknesses, Napster exploded in popularity. By the end of its year-and-a-half of life, it indexed multiple terabytes of data. A few months after it went online, major record companies sued Napster for contributory infringement. Napster itself didn't infringe copyrights (it stored no music). But it knew about, aided (via the search functionality), and made no effort to control, users' primary infringement. This was not a novel legal theory – contributory infringement in copyright cases, though not detailed in statute, has a long history[8].

All post-Napster systems[9] reject control, the central aspect of contributory infringement. This has had mixed legal success. Madster (formerly Aimster) lost at the appeals court level in the 7th Circuit[14], while Grokster won in the 9th[15]. The latter case is expected to go to the Supreme Court due to the disagreement between circuits. From a practical perspective, however, it has had

[6] This article focuses on United States laws here, but similar laws have been proposed internationally.

[7] It is harder to measure personal DVD-to-DVD copying, as opposed to filesharing.

[8] It was accepted by the Supreme Court in *Sony v. Universal City Studios*

[9] Gnutella, the first such system, was released in March of 2000, by a rogue group inside AOL. It was pulled a few days later, but was quickly reverse engineered and cloned.

a huge effect. No modern P2P system can be shut down by lawsuits against its creators, because there's no central server.

Since shutting down filesharing services is time-consuming and uncertain process, media companies have begun suing individual users of these services. So far, the number of suits is small compared to the number of filesharing users, but the deterrent effect is larger[16]. The likely future effect of this will be to return filesharing to smaller groups([17]). Given the vast storage capacities of modern hardware, individual libraries can easily contain thousands of movies or episodes of television shows.

This brings the situation back to its status as of 1998. Since attacks on the networks are running out, anti-copying strategies will likely shift back to attempts to keep the initial content from being copied. Unlike previous attempts, media companies today have a new tool: Trusted Computing. I'll present a protocol by which Trusted Computing can prevent copying, then I'll discuss the philosophical implications of this.

4 A Protocol for Preventing Copying

To obtain media:

1. The client computer boots into a known state. Note that it's not the case that every program on the system needs to be known – only the operating system, audio and video drivers, and any debuggers or other programs which can access the memory of another process.
2. The operating system measures and runs a media management application ("media manager"). A modern example of a media manager, which relies on traditional anti-copying technology rather than TC, is Apple's iTunes.
3. The media manager uses remote attestation to prove that it is a version known to the media distributor. The distributor sends encrypted media to the media manager, which has the decryption key. Because no other application can read the media manager's memory, the decryption key remains secret while the manager is running.
4. The media manager directs the TPM to seal the media to the machine's current configuration.

To play media:

1. The media manager directs the TPM to unseal the media the user requests. It does not allow the user to copy the media, merely to play it.
2. To prevent analog recording from the system's audio and video outputs, TCG 1.2 has secure I/O channels, which can authenticate a/v components (speakers, monitors) via a trivial public key protocol.

5 Power

Hal Finney's RPOW[9] uses partial hash collisions to ensure that users' computers have done a certain amount of work, and then issues proof-of-work tokens. The RPOW server redeems tokens exactly once, or redeems and exchanges them for tokens of equal value. One goal is to fight spam by allowing users to reject or mark e-mail without RPOWs attached. Ordinary users' computers will have to do very little work (since e-mails they receive will include RPOWs that they can reuse). Large senders of e-mail, who are usually spammers[10], will be unable to afford the computation time needed to produce the RPOW.

RPOW uses the TC capabilities of the IBM 4758 to prove to users that it is operating correctly. It is critical that the RPOW server operates as users expect, since it is effectively a mint for a new form of digital currency[11]. RPOW cannot turn off this attestation functionality without making users suspicious. Because RPOW's aim is to gain adoption, and because currently, nobody needs to rely in RPOW to meet specific goals, the operators of RPOW will keep attestation on.

This seems like an excellent case for Trusted Computing – users know that they can trust the people responsible for minting their currency to limit the supply to the amount of computing power available. Many countries formerly had a similar, though less transparent, system for currency valuation, the Gold Standard. This was used worldwide until World War I, when the system began a slow decline ending in 1971[18]. Modern currency is "fiat currency", backed by trust in its tradability and in its governmental issuer's respect for property rights. If RPOW were ever established as a genuine digital currency, and the TC feature were switched off, the value of the RPOW wouldn't collapse, just as the value of the dollar didn't collapse when the United States moved away from the gold and silver standard. Internet users would continue to use RPOW just as they now use fiat currency in a variety of online games[19].

So, the RPOW server's use of TC is based on its current need for acceptance – that is, its lack of power relative to its potential users. If the power relationship changes, the RPOW server could turn off TC. Greenstadt and Raymond [20] have suggested that TC capabilities could be used for medical privacy. However, as they point out that the major problem in medical privacy is not the release of detailed source data, but of short conclusions. No TC system[12] can control users' minds, so laws are still needed to protect medical privacy. In fact, such laws already exist in the United States and Europe. TC constraints cannot help enforce these laws, because the problem is human actions unrelated to technology.

In addition, while TC may be used internally to the medical community, its remote attestation features won't be available to those outside the community –

[10] As in all proof-of-work proposals, legitimate mailing lists will need to be whitelisted by users.

[11] Given the history of modern computing, the currency's inflation rate is roughly 30% due to hardware improvements alone

[12] Yet[21]

the people whose privacy depends on these features. There is a power imbalance between medical customers and healthcare providers – if a customer chooses not to get medical care, they may die. But if a provider chooses not to serve the desires of a small set of customers, they will likely not go out of business. If the number of users who value privacy were large enough, this might not be the case. This power imbalance makes TC useless to protect privacy.

Media companies are collectively and individually vastly richer than single individuals. Due to copyright law, each has a monopoly on certain content. So, the power is in the hands of the media companies. They may set terms for users' distribution of copyrighted content, and users do not have the ability to legally go to other suppliers. They don't even have the ability to require that media companies don't sell whatever personal information they are given – one thing which could easily be ensured with TC. This is true of vendors in general.

Network effects also affect the balance of power – if nearly everyone uses TC, it will be harder for individuals to refuse. Today, around 90% of web users use Flash. So, many web sites depend on it. To refuse to use Flash is to refuse many of the most popular web sites. Those who aim to do business exclusively to TC users need TC to attain the same level of ubiquity. This suggests that if some areas of the world or segments of the market refuse to use TC at all, it will not be widely used for DRM. Specialized applications, like RPOW, might continue to use TC even if it is not in general use.

TC will be subject to hardware breaks, just as the anti-copying technology in console games has been since the early days of console gaming[22]. These hardware breaks will be semi-universal – they will be limited to a single model of TPM, and to users who can purchase and install a hardware solution. As a result, media companies will propose increasingly Draconian law enforcement measures (as they have been during the past seven years). While technology will continue to be defeated, law enforcement will grow increasingly effective in destroying larger copying networks. Efforts to use Trusted Computing to protect privacy will fail because of the power imbalances described above.

DVDs still use CSS and region coding, even though they have been thoroughly and repeatedly broken. If Trusted Computing is widely implemented, it is likely to remain a part of the computing landscape even though it fails to accomplish most of its goals. So, it's important to educate users that sometimes giving up choices can lead to increased freedom.

References

1. http://www.irinnews.org/report.asp?ReportID=39193
2. http://news.bbc.co.uk/2/hi/south_asia/3600742.stm
3. http://www.canoe.ca/NewsStand/LondonFreePress/News/2004/03/09/375477.html
4. Pearson, ed. *Trusted Computing Platforms*, New Jersey: Prentice Hall, 2003
5. http://www.pc.ibm.com/europe/think/en/security.html
6. Clarke, Roger. "Information Wants to be Free..." http://www.anu.edu.au/people/Roger.Clarke/II/IWtbF.html

7. Gates, William "An Open Letter To Computer Hobbyists", February 3, 1976 http://www.blinkenlights.com/classiccmp/gateswhine.html
8. Wong, Meng. "Sender Policy Framework" http://spf.pobox.com/
9. Finney, Hal. "Reusable Proofs Of Work" http://rpow.net
10. Business Software Alliance "Global Software Piracy Study" http://www.bsa.org/globalstudy/
11. "The No Electronic Theft ("NET") Act" http://www.usdoj.gov/criminal/cybercrime/17-18red.htm
12. Goldman Eric, and Gladstone, Julia Alpert, ""No Electronic Theft Act" Proves a Partial Success", National Law Journal http://eric_goldman.tripod.com/articles/nljnetact.htm
13. http://www.eff.org/IP/DMCA/states/
14. http://www.riaa.com/news/newsletter/pdf/aimster20030630.pdf
15. http://www.eff.org/IP/P2P/MGM_v_Grokster/20040819_mgm_v_grokster_decision.pdf
16. Rainie, Lee, et al. "The state of music downloading and file-sharing online", Pew Internet & American Life Project http://www.pewinternet.org/pdfs/PIP_Filesharing_April_04.pdf
17. http://crypto.stanford.edu/DRM2002/darknet5.doc
18. Wikipedia, "Gold Standard" http://en.wikipedia.org/wiki/Gold_standard
19. Castronova, Edward. "An Introduction to Virtual Item Trading", http://mypage.iu.edu/%7Ecastro/VirtualItemTrading.html
20. Greenstadt, Rachel and Raymond, Jean-Francois. "Applications of Trusted Computing for Medical Privacy", 2004, http://www.eecs.harvard.edu/ greenie/portia.pdf
21. Gibson, William and Swanwick, Michael. "Dogfight", *Omni* 7.10 (July 1985)
22. Sega v. Accolade 977 F2d 1510 (9th Cir. 1992), http://www.eff.org/legal/cases/sega_v_accolade_977f2d1510_decision.html