

16

The Ecology of Control: Filters, Digital Rights Management, and Trusted Computing

Joe Karaganis

The Internet was not designed as a distribution channel for media. Rather, it was a communications network, which gradually grew capable of communicating rich media content. Its architecture privileged the transmission of data, not the identification or control of what was conveyed. It was, in this respect, an open network, indifferent to the uses to which it was put. In the mid-1990s, as culture industries began to understand the Internet as a competing, uncontrolled distribution network for their goods, they had to confront this infrastructural fact. The underlying problem was not that tens of millions of file sharers seemed indifferent to copyright norms, but that the networked computing infrastructure itself could not guarantee the unique market conditions for cultural commodities—widespread distribution *and* controlled scarcity.

The complementarity between open networks and personal computers was not accidental, though in some respects its survival has been. The personal computer grew out of (and was sustained by) a broader culture of general-purpose computing which, from computing pioneer Alan Turing's day forward, treated computers as universal machines, capable of solving any problem that could be expressed in generalized programming languages. Software could be developed to almost any purpose, from atmospheric modeling to accounting to video compression. Hardware development also relied on relatively open frameworks that encouraged tinkering and innovation. Buses and ports, which govern input and output (I/O) within and between devices, favored minimum and open standards for connectivity. Their key feature was to impose few constraints on what was connected to the microprocessor or on how those "peripheral" devices were used. Over time, this allowed computer system components to diversify and evolve at different speeds, requiring less frequent revisions to the underlying architecture.

Developments in network technologies followed a similar path: The Internet embodied and to a large extent consolidated the notion of "end-to-end" networking, which held that any two parties should be able to communicate without the intervention of a third party, and—by extension—that the most widely shared protocols should be the least constraining.¹ Initially, these principles served the predominantly military goal of ensuring that communications networks could survive the destruction of individual nodes (e.g., by nuclear attack). In practice, this goal required solving a range of interoperability issues between the electronic networks that predated the Internet, and that subsequently composed it. Collectively, these efforts cre-

ated a network that supported not only survivability and interoperability but also a very wide scope for future innovation. The lowest-level internet protocols provided a platform for other networks and applications with more specific functionality. The World Wide Web, with its markup language (HTML) and hyperlink structure, is only one example. Instant messaging systems, peer-to-peer file sharing, and internet telephony are others.

Closed systems and centralized networks differ not in their absolute capacity for innovation, but in the control that system owners exercise over them. For cable television systems or gaming consoles, corporate owners retain the prerogative to decide how the system evolves. Like traditional content industries, they box out challenges to their infrastructure while profiting from—and providing third parties the opportunity to innovate around—fixed-format content. For some 70 years, AT&T's telephone system was the exemplar of a closed network, managing all communications and exercising control over what could be attached to the end of the phone line. In the United States, cellular phone networks operate on similar principles, with vendors dictating both the hardware available to users and the uses the network will support.

Centralized control offers a number of advantages. End-to-end networks and open architectures are poor at prioritizing among different objectives—"bad" uses compete for network resources with "good" ones. Innovation in end-to-end systems can be difficult to coordinate and aggregate; interoperability between systems becomes more challenging when there is no controlling authority to enforce compliance with standards. Innovation within a network, under these circumstances, acquires a strong social dimension, as each actor weighs the costs and benefits of change. Open source software projects, which sometimes manage contributions from hundreds or thousands of volunteers, typically address these issues by maintaining hierarchical control over the integration of new code. They rely on benevolent dictators or other carefully managed structures of authority to prioritize and evaluate innovation. For these reasons, "intelligent" networks that can discriminate among uses and users have been the norm when the ownership of the infrastructure and the functions of the network are clearly defined, as in ATM banking networks.

Content companies and technology companies have traditionally viewed these structures differently. Although end-to-end networks permit more sources of innovation, their lack of discrimination undercuts pricing opportunities for services carried over the network, which in turn limits the

market power of incumbents. Many companies—including technology companies—have been on both sides of this issue, defending openness in markets where they are minority players, while working to create chokepoints in areas of innovation where they dominate, which allow for “supernormal” profits. This is the structural complaint against Microsoft’s role in the operating system market, but it extends to many other market positions and technologies: I/O buses, media players, devices like videodiscs and DVDs, and so on.² Internet Service Providers (ISPs) have also been on both sides of this issue, acquiring immunity from liability for illegal content carried over their networks (in 1995), but also freedom from “common carrier” provisions that allowed competitors to use their networks (in 2005). The latter development promotes the growth of vertical monopolies that can link internet service provision and content production. Time Warner, News Corporation, and General Electric all have substantial, interlocking interests in both ISPs and media production. Although this consolidation has not yet produced overt discrimination of internet content in the United States—and may never if “network neutrality” principles are written into law—service providers practice subtle forms of regulation through such measures as “asymmetrical” upload and download speeds, which favor a consumption-oriented model of internet use. Home-operated servers, which shift the user from consumption toward production and distribution, are often penalized or contractually forbidden.

securing the distribution channel

Napster and newer file sharing networks are examples of the unpredictability and low barriers of entry associated with end-to-end innovation—the original Napster, like the first Mozilla browser, was written by a college student and freely disseminated on the Web. Because of the Internet’s open architecture, no technological constraints, controlling authorities, or market incumbents prevented their widespread adoption. As these and other innovations demonstrated, the Internet can support not only new content but also new and rapidly evolving distribution models.

The content industries have adopted a variety of tactics to combat this proliferation of distribution channels, from education campaigns, to efforts to co-opt file sharing in “high-risk” communities, such as universities,³ to continuous legal action against file sharing network owners (e.g., Napster, Sharman Networks, Aimster, Grokster, Limewire) and individual users of those services (some 18,000 by 2006).

The impact of these efforts is unclear. Some studies have suggested a correlation between the legal threats against individuals and apparent dips in file sharing numbers, but by most accounts the numbers have fluctuated since 2003 and measurement difficulties abound (Rainie & Madden, 2005; Hindo, 2004). Nearly all the major actors in the content industry understand these legal efforts as stopgap measures—bad for public relations and of dubious value in slowing the growth of informal distribution, but of potentially greater value in fostering a political climate in which more effective legal and technical solutions can be enacted. As the late Motion Picture Association of America (MPAA) president Jack Valenti made clear, the longer term and more consequential goal is to transform the principles of openness that make computing culture such a dangerous environment for media companies, and to erect legal protections around this new technical and cultural infrastructure. Nothing less will “build the barricades tight and strong so that creative works are shielded and made safe in order that investments in more and more programming can be planned and made without fear of being burglarized by others” (Valenti, 2000). Nothing less will ensure that the distribution channel is successfully remediated—that is, returned to the old intermediaries.

Jack Valenti’s retirement from the MPA in 2004 marked the departure of one of the dominant voices in political debates about digital culture. For Valenti, culture was synonymous with the major culture industries, and an uncontrolled distribution system was an unequivocal and growing threat. Valenti was less open about the extent to which this “crisis” of the content industries also represented a vast new business opportunity. The same technologies that expanded the power to copy and distribute digitized cultural goods could also tilt the other way, and dramatically enhance the power of rights holders to control distribution and consumption. The technologies that permitted new forms of access to content also enabled new forms of audience surveillance and marketing, which might diminish the unpredictability of the consumption channel by allowing more precise matching of products with targeted demographic groups. In Chapters 14 and 15 of this volume, we explored the cultural and market logic of this vision. In this chapter, we explore the technological methods.

filter

Napster, the first peer-to-peer (P2P) file sharing network, was launched in 1999 and developed quickly into a 70-million user community. Napster pro-

vided fresh answers to two problems associated with distributed resources in digital networks: the search problem, or how to identify resources scattered across millions of individual machines; and the bandwidth or continuity problem, or how to ensure that large files could be reliably transferred across a network characterized by intermittent connections and/or uneven bandwidth. Napster answered the first question by providing a centralized database for listing files distributed across a large number of individual machines. This database brokered relationships between those seeking resources and those holding them. Napster addressed the second challenge, continuity, by providing a means of tracking the presence of users on the network. This enabled file transfers to be interrupted and resumed over time, thus greatly reducing the penalty of file sharing over low-speed or erratic connections.

With these relatively modest innovations, Napster succeeded in leveraging the contemporaneous development of several other technologies: a widely available audio compression format (MP3, patented in 1996), personal computers powerful enough to run MP3 “codecs” and reproduce high-fidelity audio, and the growth of commercial broadband services, which allowed individual “super-sharers” to emerge as key distribution points on the network.⁴ Together, these technologies combined to make an effective extracommercial distribution channel (and also de facto archival technology) for music. Because music had relatively modest technical requirements for digitization and high-quality reproduction—compared, for example, with the high-bandwidth requirements of film or the cumbersome screen technologies that continue to limit e-books—it was an optimal good for early peer-to-peer networks.

By all accounts, Napster thrived on the exchange of copyrighted music. Although the number of infringing files and successful transfers is hard to determine, there were 2.79 billion *initiated* file transfers at Napster’s peak in February 2001. In 1999 the Record Industry Association of America (RIAA) brought a copyright infringement lawsuit against the company. It argued that Napster bore responsibility for the infringing behavior of its users, even though Napster had no control over those uses, and even though the network also supported a wide range of legal behavior, such as the sharing of public domain works, pictures, and—in large quantities—amateur pornography. Napster argued that, as a neutral service provider rather than a content provider, it benefited from “safe harbor” provisions in the Digital Millennium Copyright Act—the same provision that had been successfully

invoked by ISPs in 1998. Unlike the ISPs, Napster lost this argument: The presiding judge ordered Napster to eliminate infringing files from its network.

The scale of this task necessitated a technological fix. Napster began by filtering files that bore the titles of copyrighted songs—initially some 250,000 titles supplied by the record industries. This reduced the sharing of infringing files, but the permutations of titles made the system unreliable: Infringing files slipped through, and legitimate files were sometimes stopped. Music sharers took advantage of this inflexibility by renaming songs. During Napster's appeal in 2000, the RIAA successfully argued for a stricter standard: File sharing would be permitted only if copyright protection was perfect. Napster had prepared for this outcome and implemented, in June 2001, a system designed to identify songs by their sound, to be checked against a database of copyrighted songs. This filter proved wildly indiscriminate. Overnight, it eliminated nearly all traffic from Napster's network. Usage plummeted, and Napster shut down definitively in July 2001.

Content filtering is an appealing mode of control for regulators and other actors who are uncomfortable with the lack of traditional intermediaries in the end-to-end world of the Internet. Filtering addresses the asymmetry between the offline world and the Internet, reinserting intermediaries with the capacity to discriminate content before it reaches the end user. When "good" file transfers, page visits, or other uses of the Internet can be distinguished from "bad" ones, the Internet can become a much more powerful tool for social regulation. It can replicate and extend the kinds of social, institutional, and material barriers that regulate conduct and access to information goods in the offline world, separating children from pornography or citizens from unauthorized news, or reducing the effective scale of copyright infringement. In this respect, filters try to reproduce nondigital sociotechnical arrangements and the power relationships that define them. Unlike those arrangements, however, filters have fewer points at which power is visible or negotiable. They can deny not just access to resources but also to *knowledge of resources* more completely than in other kinds of social space. In the offline world, the exercise of such power tends to be thickly mediated by laws and norms, by the materiality of technologies, and by the friction of social relations. In the digital environment, this social and institutional context is much thinner, making slippery slopes on issues of control or surveillance easier to descend, and transgressions of privacy more difficult to specify. In networked computing environments, there are few practical dis-

tinctions between a little control and a lot of control—enabling one often enables the other. The borders between public and private are less clearly drawn and the means of establishing them are often technically demanding.

To filter is to monitor and parse communication. There are formidable challenges to discriminating content beyond the grossest level, even for works as distinctive as music or video files. Simple workarounds, such as encryption, make the identification of files problematic (see, e.g., Brief *Amicus Curaie*, 2005). Content filtering also faces a practical and politically charged difficulty regarding the point of application: Where should filters do their work? The Napster case revealed the limitations of targeting individual services. As Napster's filter diminished its value to users, music sharers simply moved on to other file sharing networks—Grokster, Morpheus, KaZaA, eDonkey, and others. By most accounts, the global file sharing community is now larger than during Napster's 100 million-user heyday (OECD, 2004). With the Napster case, the music industry proved that it could crush institutional competitors, but not control the Internet itself. The new P2P services learned from Napster's vulnerability and adopted fully decentralized network models that distanced network owners from responsibility for—or even knowledge of—the content circulating on the network. Unlike Napster, the new services had no central directory. In addition to complicating legal efforts against network owners, this structure made filtering more difficult at a technical—not merely contextual—level. Without a centralized database to govern resources, there is no obvious point of application for filtering technology. Advances in filtering technologies, such as new tools for identifying “audio fingerprints,” haven't escaped this architectural tradeoff.

Although Internet Service Providers have been mostly successful in avoiding liability for the activities carried over their wires, their central role in internet access makes them a logical and attractive target for content filtering.⁵ Renewed RIAA attention to ISP filtering in the United States has focused on universities, which typically act as ISPs for their student bodies and have a potentially higher degree of liability for student activity. If RIAA succeeds in establishing this liability, universities will begin to bear the costs of an “arms race” with file sharing networks and other related technologies (Fisher, 2007).

ISP-level filtering of this kind remains technically daunting: Parsing heterogeneous ISP traffic is much more difficult than scanning self-selecting file sharers, and the possibilities of defeating such filters (e.g., via encryption) are numerous. General filtering of internet use also runs up against power-

ful free speech claims and more diffuse expectations of online freedom. This former issue, especially, has blocked efforts to force commercial ISPs to play a filtering role. The RIAA hopes that the university–student relationship will allow for a different arrangement of rights and expectations—one that, in all likelihood, will not look like the current Internet. Content filtering on a large scale will inevitably sweep out protected forms of speech and expression along with infringing material: file sharing networks, like the Internet itself, support both. The Communications Decency Act of 1995, which required intermediaries to filter or otherwise control access to “indecent” Internet material, was declared unconstitutional on these grounds.

Filters operate, in many respects, on the presumption of a heterogeneous system—they operate in a context of diverse and changing services and network flows. This imposes a high complexity cost in distinguishing different kinds of content. To date, machine evaluation of context and meaning is poor, and choices made by human editors inevitably reflect a range of conscious and unconscious biases. Filters work best when the values they check are themselves binary—copyrighted/not-copyrighted, on/off. The problem suggests its possible solutions. What if copyrighted works could be distinguished by the presence of a binary tag, built into digital files? What if the network could be made less heterogeneous?

digital rights management

Content filtering is not the only form of cultural regulation applicable to digital networks. Digital rights management (DRM) systems take a different approach to cultural regulation: rather than sort through heterogeneous materials, DRM systems tag and encrypt cultural goods at the outset and provide authorized end users the tools for unlocking that content. Authorized users and uses are determined by the content provider and mediated through end-user contracts—notably, the “click-through” agreements that precede the installation of most software. Increasingly, these contracts provide for the ongoing monitoring and updating of software by the provider over the Internet, enabling forms of control that extend beyond the sale (von Lohmann, 2002; Godwin, 2003; May, 2006)

DRM efforts, to date, have been disorganized and mostly unsuccessful. Many narrow and incompatible systems have been pushed into the marketplace, reflecting corporate anxiety about control but much less attention to consumers’ expectations about the permissible or convenient uses of

digital goods. A 2003 report on e-books sponsored by the American Association of Publishers and the American Library Association notes a number of areas where existing DRM strategies have undermined the acceptance of digital formats. Chief among these were the lack of platform neutrality (e.g., across Mac, PC, and Linux systems), the lack of portability (of files between devices or machines), the lack of transferability (of files to another person), and inadequate excerpting, highlighting, or print features. Many of the same complaints can be applied to other digitally distributed media, such as music and, increasingly, video.

Deirdre Mulligan, John Han, and Aaron Burstein (2003) raise additional questions about the continuous and often surreptitious monitoring that accompanies many DRM systems. Because these systems cannot be legally disassembled and analyzed, due to provisions of the 1998 Digital Millennium Copyright Act (DMCA), it is difficult to know what they report back or how company claims about privacy are to be verified. Competition among the major content holders creates a different order of problem, as companies vie for the control of media delivery platforms. There are, for example, multiple, competing DRM formats for the commercial sale of digital music, each vertically integrated across hardware and software systems.

The proliferation of systems reflects not only different corporate judgments about consumer expectations but also the ubiquitous use of DRM as a tool for market segmentation and competition. An industry snapshot of digital music from summer 2004 provides an idea of the pattern: Apple's FairPlay DRM is part of a vertically integrated commodity chain that runs from the company's licensing deals with record companies to its online music store to its media management software on individual computers to its dedicated player device, the iPod. The iPod supports Apple's chosen digital audio format, AAC, but not Microsoft's WMA format. Microsoft, for its part, has declined to make an Apple-compatible version of its media player software, with the consequence that music services that use WMA (like the relaunched, industry-sponsored Napster or Musicmatch store), can't sell to either Apple computer or iPod users. Real Networks tried to build its own vertically integrated commodity chain around an online music store, its Helix DRM architecture, and RealPlayer media player, but it had only limited success, and has since devoted considerable energy to breaking open the closed DRM commodity chains of Apple and Microsoft. Until 2006 Sony insisted on its own audio compression format, called ATRAC. The deliberate incompat-

ibilities among these systems meant that all players supported the de facto standard, MP3, which has no DRM strings attached. To date, music DRM operates at the margins of a larger “open” MP3 ecology. With the recent decision by EMI to offer its catalog DRM-free, through iTunes, it is possible that this costly phase of DRM experiments may be drawing to a close.

Real Network’s effort to reverse engineer Apple’s FairPlay is an example of how firms try to enter new markets by cloning products—better, cheaper, or differently—that can provide access to larger system ecologies. Reverse engineering is a protected right within international patent law and a basic principle of technology transfer. The history of the personal computer is largely a history of this practice—of competition among component makers made possible by stable (and often open) architectures and standards. Self-updating DRM systems, continuously in contact with the corporate licensor, permit incumbents to break this practice by continually changing the details of system architecture. When hackers developed FairUse, a software utility designed to allow the legal owners of iTunes music files to remove the constraints of FairPlay DRM, Apple quickly updated iTunes to refuse these files. Apple also retaliated against Real Networks with an update that refused Real’s FairPlay hack. The persistent connection to the vendor after sale, in the DRM world, means that the vendor has much greater power to refuse outside innovation and block competition. The vendor also exercises discretion over the portability and transferability of content, as with the arbitrary device limits that encumber iTunes and other music services. Other digital media have their own strategies of control: Adobe’s e-book Reader software is integrated into its document management products, and is incompatible with Microsoft Reader e-books; Microsoft Reader, for its part, won’t work on Apple or Linux systems. All the readers are incompatible with the now defunct Gemstar Rocketbook system, which was one of several efforts to bind e-books to dedicated reading devices. In practice, any socket, connection, or exchange can be DRM-protected—Lexmark, Hewlett-Packard, and Xerox printers all contain embedded logic that can refuse generic cartridges or, in Hewlett-Packard’s case, set an expiration date regardless of whether the cartridge is empty.⁶ The value chain can be secured from competitors and users can be locked in.

These examples share ground with other struggles for the control of IT platforms and standards, and for the supernormal profits that come with such control. The web browser wars, highlighted by the long antitrust action

against Microsoft for embedding Internet Explorer into the Windows operating system, have been repeated in the media player arena. Microsoft's integration of its media player into Windows was recently the subject of a successful antitrust action brought by the European Union (2005). The fact that Microsoft continued to fight this issue in the face of hundreds of millions of dollars in fines speaks to the perceived value of platform dominance. Far more than in the browser case, which Microsoft ultimately won but could not commercially exploit, the control of media formats brings with it the prospect of incorporating the vast distribution chains of the content industries into a single revenue stream. It holds the prospect of Microsoft or Apple control of the new fee and service models that DRM and closed architectures make possible.

The *structure of participation* associated with DRM has numerous implications for freedom of expression and other core principles of an open and participatory digital public sphere. Because use of DRM'd commodities and services is circumscribed by voluntary contractual agreements, DRM systems are far less vulnerable to the claim that they undermine free speech rights. The content provider can retain complete control of the good, ensuring that it never circulates in ways detrimental to the owner's interests. This raises serious political concerns as media and economic interests consolidate. Increasingly, the capacity for censorship becomes a precondition of use. Intellectual property law, and in particular the fair use provisions to copyright law that support access to knowledge goods, are made largely irrelevant in this context. The contractual framework is a private one, with no obligation to balance public and private interests and with diminished forms of accountability. DRM signals a postcopyright regime of control over circulation and use, in which the terms of use are set not by public policy but by private interests.

an open ecology

At present, DRM systems operate within an "ecology" of computing and network technologies that is both intensively and extensively diverse. Many different systems and architectures coexist within the same information space—the average U.S. household, according to Novell's estimate from 2002, contains some 145 microprocessors. The vast majority of these systems are not PCs but devices with embedded chips—cell phones, DVD players, home appliances, and so on. Whereas some 150 million PCs were shipped worldwide in 2002, embedded systems accounted for 5.3 billion shipments.⁷

Despite the rollout of wireless services and trends toward “ubiquitous computing,” very few of these devices are currently networked. Among other things, this means that the vast majority of vendors exercise no effective control on after-sale use of these systems. They can be hacked, reverse engineered, and cloned; closed commodity chains can be opened by competitors. This familiar characteristic of commodity chains has an important political dimension: In such contexts, it is very difficult to exercise control over the global information ecology. System heterogeneity is reinforced by political heterogeneity, reflecting different national and regional needs and capacities. It is also reinforced by the facts of uneven development, which underwrite the “recycled” or “pirate” modernity described by Ravi Sundaram (Chapter 4) and Brian Larkin (Chapter 5) in this volume, with its patchwork of systems, illicit software, and low-cost computing initiatives.⁸ In such environments, complicated vendor–user relationships, structured by ongoing fees, contractual relationships, and postsale monitoring, are unfeasible. In such circumstances, computing infrastructures favor untethered resources—pirated software with broken keys, recycled PCs, and, increasingly, noncommercial products such as Linux, and other open source software.

First-world DRM provides ways of securing particular product channels, but has little impact on what people do with their computers outside those channels or elsewhere on the network. Open system architectures make it very difficult to monopolize formats; heterogeneous computing ecologies make it very difficult to universalize surveillance or compliance with contracts. DRM’s potential for controlling content is greatly undermined by this heterogeneity and openness, with its easy-entry/easy-exit conditions of innovation and use. Those who don’t like the subscription structure of the reborn Napster, or who become too frustrated with Apple’s device certification requirements for songs purchased on iTunes, can fall back on the “uncontrolled” MP3 audio format. At present, the digital cultural economy is mixed—not all sources are known or trusted; not all uses are regulated.

legal closure

Content industries have supported a range of legal innovations in the last decade to reduce this heterogeneity—initially through the World Intellectual Property Organization’s Internet treaties (WIPO; 1996), and later in national and regional implementations of WIPO treaty provisions, such as the U.S. DMCA (1998) and the EU Copyright Directive (2001). The most significant

of these measures is the criminalization of attempts to circumvent encryption or other protections attached to digital goods. These anticircumvention clauses extend legal protection beyond the copyrighted work itself into the penumbra of technological methods for limiting access. They have equated circumvention with theft, and in the process have shifted the power to make decisions about what constitutes permissible use from consumers (and when contested, the courts) to the copyright holder. Such measures effectively close off much of the latitude available for the unanticipated uses of digital goods, whether or not those uses infringe the copyright. They also affect a number of basic practices of technology research and development—most directly in the area of encryption, which advances through continuous testing and breaking of encryption algorithms (Litman, 2001; Gillespie, 2004; Brief, 2005).

As in other areas, new tools for controlling the commodity chain also prove to be powerful tools for excluding competition: Manufacturers and content producers can simply encrypt functions or input/output protocols to block the development of compatible devices. The DMCA, in particular, effectively criminalizes reverse engineering except for the narrow purpose of improving interoperability—and then on terms highly favorable to the owner of the encrypted system. Although some of these efforts have been defeated (such as Lexmark’s 2004 bid to exclude generic printer cartridges), manufacturers must now be taken to court to secure such freedom to operate, raising the cost of entry for innovators. The DMCA failed, moreover, to provide any specification of personal fair use—giving a nod only to the institutional purposes of libraries and archives, through guidelines subject to periodic reexamination by the Library of Congress.

In practice, the DMCA has been used by the content industries to restrict independent research on existing DRM technologies—including for the purposes of academic research. There have been several high-profile cases of such censorship, such as the 2001 lawsuit brought against Professor Ed Felton of Princeton University, when he sought to document his decryption of the Secure Digital Music Initiative’s encryption algorithm *as part of an SDMI-sponsored contest* (see Felton, 2003); or the charges brought against Russian programmer Dmitri Sklyarov when he presented his work on Adobe Reader encryption to an American conference audience; or the charges brought against 16-year-old Jon Johanson for writing a seven-line program to strip DVDs of their CSS encryption, so that they could be used

on Linux systems, which have no CSS-licensed media players. In practice, the DMCA works to shield content industry value chains from competition and the innovation processes characteristic of open system ecologies.

trusted computing

Trusted computing (TC) describes a different model of network architecture under development by a range of major hardware and software companies, including Microsoft, Intel, IBM, Hewlett-Packard, and AMD. Although the initiative has had a variety of names and stated goals, most accounts highlight the need to “give individuals and groups of users greater data security, personal privacy and system integrity.”⁹

Although project details are often difficult to pin down, TC’s core functions are relatively clear: where the PC platform has been open to user modification, a trusted computer will have its status continuously certified by a controlling authority (such as the Trusted Computing Group, comprised of the major players). Certification will depend on the presence of system components that comply with industry requirements for the secure delivery of services. These include all input and output functions at all points of the system, from the graphics and sound card to hard drives and CD burners. Although this will make it more difficult for viruses and other malware to run on a local computer, much of the initiative has focused on preventing the user from making unauthorized use of media content: Trusted status means that no unapproved device or software can capture digital output or even analog output that can be redigitized (the so-called “analog hole” at stake in debates about high-definition television signals and recording devices). All components and peripherals will have to be able to adhere to rules dictated by content providers. Storage of all kinds—hard drives, CD burners, and random-access memory—will be subject to those rules.

The public face of trusted computing usually involves the claim that individuals can opt out—that users will be at liberty to turn off the TC framework and retain the current “freedom to tinker,” to use Ed Felton’s phrase. In practice, the incentives to work within the trusted environment are likely to be high: Trusted status will be the condition of the delivery of a wide range of services, and could easily be made the condition of more basic forms of communication, such as the receipt of Microsoft Word documents, or Outlook-generated emails. If this seems unlikely, it is worth noting that DRM for both MS Word documents and Outlook email was integrated into Office 2003,

allowing such features as expiration dates, encryption readable only by specific persons, or blocks on forwarding documents or emails outside the home institution. At present, these features need to be supported by a Microsoft network environment. In a trusted system ecology, permissions can be so extensive and interconnected that documents could be restricted wherever the trusted system was in force. Ross Anderson notes the authoritarian undercurrent of these capabilities, which not only enhance the potential for secrecy, but also the possibility of retroactively altering the documentary record. Citing one possible use, “Word documents created on civil servants’ PCs [can be] ‘born classified’ and can’t be leaked electronically to journalists” (Anderson, 2003). When documents have trackable identification numbers, a trusted system could erase them, alter them, or simply render them unusable wherever they had circulated. By stemming the viral capacity of the Internet to spread information, document DRM could bring about a dramatic decline in the accountability of private and public institutions. It signals a change, in Geoffrey C. Bowker’s (Chapter 2, this volume) terms, in the power relations that define a society’s “memory practices.”

It should come as no surprise that software piracy is an explicit target of this initiative.¹⁰ The blacklisting of known, pirated software serial numbers is already common practice. Usually, blacklisting doesn’t prevent the installation and use of software—pirate hacks often work around the initial validation—but it does restrict the steady flow of patches and updates provided by the vendor.¹¹ Vendors, in turn, must weigh whether their paying customers are better served by blocking access to security patches for unauthorized users, which in a highly connected environment can make all users less secure. The resulting system leaks on all sides, and more so in the developing world, where the majority (and sometimes the overwhelming majority) of the installed software base is pirated.¹² TC-enhanced control raises the possibility of a broader transformation of this global IT ecology, and in particular of a very different kind of pressure brought to bear on developing countries. Anderson (2003) describes how the developing-world IT infrastructure, characterized by low ability to pay and a high capacity for piracy, could be made to play by Microsoft’s rules:

For years, Bill Gates has dreamed of finding a way to make the Chinese pay for software: TC looks like being the answer to his prayer. . . . The proposed use for this is that if everyone in China uses the same copy of Office, you do not just stop this copy running on any machine that is TC-compliant; that would just motivate the Chinese

to use normal PCs instead of TC PCs. You also cause every TC-compliant PC in the world to refuse to read files that have been created using this pirate program.

The dependency relationship between developed and developing countries would make this a very powerful incentive to adopt trusted computing, whether or not it came with new media services. Adoption would occur for different reasons based on economic status: for developed countries, the incentive would be better security and richer media services; for developing countries, it could become access to the communication networks of the developed world.

Although elements of open source software practice could probably survive in a TC environment, the certification requirement would gut one of its core values: the right of anyone to run modified software. Although a TC-compliant Linux is almost certainly technically possible and is reportedly under development by IBM, any significant modifications to such a system would require recertification by the TC authority. User control of the direction of open source projects would be greatly diminished and possibly subject to industry veto.¹³ The user-centered development model would be severely constrained, as it would be too costly for a central authority to test and certify thousands of small-scale changes.

Although Microsoft and others can implement some of the TC agenda in software, the real security lies in hardware integration—initially in the form of a chip soldered to the motherboard (sometimes called the “Fritz chip,” after retired U.S. Senator and DRM enthusiast Fritz Hollings), which would conduct the authentication and certification process for the machine during the boot process. Versions of hardware authentication have been in use for some time in gaming consoles such as Microsoft’s Xbox and Sony’s Playstation, ensuring that—for all but skilled hackers—only licensed games can be played (Xbox launched in fall 2001; “mod chips” were available by spring 2002; the security protocol itself was cracked in 2003). In the PC arena this functionality will be complemented by and likely subsumed into the microprocessor itself, which will make modification far more difficult.¹⁴

To date, TC has flown almost entirely under the public’s radar. Some of this reflects difficulties in implementation—the TC agenda has advanced in bits and pieces, rather than as part of a concerted master plan. The low profile also reflects industry concerns about public outcry. Intel learned a lesson in 1999 when it announced that Pentium III microprocessors would henceforth contain a unique identification number, enabling hardware-based

authentication of users. Privacy and civil liberty groups saw this as a direct threat to user anonymity on the Internet, and defended the conception of the Internet as an extension of private space except when explicitly relinquished by the user (as in commercial transactions). The strength of public opposition turned the issue into a significant embarrassment for Intel, which ultimately backed down.

Trusted computing reconceptualizes privacy in ways that make analogies to “real-world” expectations of privacy—and abridgements of it—difficult to assert. Notably, TC implements parts of the wish list of digital privacy advocates, who welcome the added barriers to unauthorized access to data on personal computers. The difficulty is that TC embeds effective power over the definition of privacy into the commercial relationship between individuals and vendors. TC equates security not with anonymity, but with constantly authenticated identity. It thereby runs counter to the strong libertarian strain within computer culture that privileged anonymity is a protection against both malicious individuals and the state. Trusted computing is the fusion of security discourse with property discourse. It builds a transition path away from open networks and system architectures toward an environment capable of supporting much more comprehensive control over information. And it shares with other forms of security discourse recurrent blindness to the implications of certain kinds of security for public discourse and for the forms of accountability necessary to an open society. As the basis for a digitally mediated democratic society, TC represents a very large step into the unknown.

Like the current battles over DRM, trusted computing will probably have a bumpy future. Elements of the architecture are extremely ambitious. Earlier generalized architectures for content control have met with both technological and political difficulties, such as the Secure Digital Music Initiative or recent legislative efforts to reengineer hard drives. It is not clear who the certifying authority will be. But large parts of the TC architecture are already in place. Lagrande was built into Pentium IV chips (2003–2004), although not yet activated. Microsoft’s trusted document environment exists on its office network servers. Vista, Microsoft’s new version of Windows, makes extensive use of hardware and software monitoring. Congressional action, shaped by content industry lobbying, has been looming for several years, and could result in legislation that mandates steps toward trusted computing on all computers and media-capable devices.¹⁵ Regardless of the short-

term outcomes, TC is a bet on the long term—on the need to secure content delivery channels once and for all. Although there are many proposed flavors of trusted computing, they all share, with other digital technologies of control, the lack of a *necessary* middle ground. The conditions of a little control are also the conditions of a lot. However useful some of the features might be—in developing a fair-use friendly digital circulation model for authors or musicians, for instance—TC places controlling authority in the hands of the corporate intermediaries, who, facing diminished competition, will face less pressure to make generous arrangements for the secondary cultural lives of their goods.

core common infrastructure and digital freedom

One way to strengthen public cultural agency, in this context, is to support a “core common infrastructure” for digital culture that provides low-cost access to basic services and maintains a high degree of openness to different kinds of secondary activities (Benkler, 2001; 2006). In Yochai Benkler’s account, core common infrastructures exhibit neutrality toward different users *and* uses: They do not “discriminate” by raising costs for particular kinds of use. Such neutrality is inevitably circumscribed by social and legal definitions of harmful use, but the general principle holds true for many basic forms of infrastructure: The public highways, for example, support many different kinds of traffic at equal cost to the traveler. “Common carrier” regulations have traditionally required private network owners, such as railroads and telecommunications providers, to serve anyone—including competitors—willing to pay a standard price. Competitive markets for services can also respect neutrality toward uses and users—a fact that becomes increasingly important when infrastructure is privately owned. In these diverse but related senses, core common infrastructures are bound up with, and often conceived as prerequisites of, the exercise of multiple kinds of freedom—free speech, freedom of movement, free markets, and, in Lawrence Lessig’s addition to this lexicon, free culture—one in which individuals enjoy wide latitude to create and share (Lessig, 2004). In turn, the substantive meaning of freedom, at any particular moment, is inseparable from the characteristics of these networks: Free speech is exercised through the dominant technologies of communication; freedom of movement through the dominant means of transportation; freedom to buy and sell within a marketplace that discriminates neither for nor against certain buyers.

Digital technologies bring their own native characteristics that shape the available forms of cultural participation and notions of freedom. The almost limitless fungibility of digital representation is the most essential of these, with arguably the largest impact on the objects and forms of circulation of public culture. As digital convergence becomes the normal condition of cultural expression, it is easy to lose sight of the underlying shift implied in digital representation. For the first time, the content of culture is easily dissociable from its storage medium—text from paper, moving images from film, sound from records and CDs. Nothing *necessary* anchors a digital representation to a particular device or object. Instead, the characteristics that matter most are those that define the technologies themselves: operations upon data; blurred lines between storing, accessing, and copying; and increasingly fast and cost-free transmission.

These characteristics inform the substantive meaning of a core common infrastructure for digital culture. In particular, they provide an important and, in the current environment, very demanding condition for nondiscrimination: minimal constraints on the use of the basic capabilities of the technologies themselves. This condition is central to several contemporary information and technology policy debates, from the debate about discrimination among services on the Internet signaled by the term “net neutrality,” to debates over the availability of unlicensed radio spectrum, which has supported, among other things, the growth of WiFi. The open source/free-software movement has also demonstrated how commitments to open, nondiscriminatory infrastructures can underpin more formal concepts of digital freedom, expressed through rights to examine, modify, and distribute software code, and embedded in participatory values and practices. The industry campaign against file sharers has different objectives, but similarly tests whether file sharing generates its own *informal but substantive* concept of digital cultural freedom, rooted in greatly expanded lower-cost access to media.

Because digital convergence brings together once-distinct modes of expression (textual and audiovisual), as well as different modes of communication—one-to-one (telephony), one-to-many (broadcasting), simultaneous, and asynchronous—the larger cultural field comes to rely increasingly on the constitutive choices made for the underlying infrastructure. One of the main achievements of the current wave of law and technology scholarship has been to articulate the interplay between these technical choices and other more familiar forms of cultural regulation. For Lessig (2002, 2004), law

and technical architectures operate as complementary modalities for shaping our freedoms and notions of cultural agency (together with norms and markets). Benkler (2001), for his part, builds on a traditional three-layer model of network architecture to identify regulatory opportunities for defending social, political, and cultural freedoms: the physical layer, involving the organization of material infrastructures such as fiber optic networks and radio spectrum; the logical layer that structures traffic (such as TCP/IP protocols for the Internet); and the content layer structured by norms of ownership and by intellectual property law.

These schemas map the politics of openness and closure proper to network technologies. They also identify a new context for the construction of democratic cultural values, described partly by the reconfiguration of major political values such as privacy or free speech within emerging digital institutions, and partly by the ways that digital culture has focused political attention on once-minor values, such as sharing, collaboration, and creativity. The digital environment provides a range of native contexts for the growth of these forms of subjectivity and agency. It also creates the conditions for a more extensive lockdown of cultural production than was possible in the analog era, marked by pervasive monitoring, the steering of behavior through design choices, and the strengthening of commodity chains.

As Lessig has observed throughout his work, code is a form of *de facto* legislation of the digital environment. Part of our problem is that we have done little to subject code to the same public processes or safeguards that the democratic tradition has thought essential to the formulation of law. The outpouring of serious attention that 17th- and 18th-century thinkers gave to the question of constitutions—in recognition of the opening of the political “design space” as the divinely justified social order lost legitimacy—has not been reproduced as we move into an era of new sociotechnical systems. Civil society for the digital age is radically underdeveloped, and the important struggles over values are no longer limited to defining the boundaries between the individual and the state.

Although older versions of public life were no less dependent on technological infrastructures and pathways of information, they were far more constrained by the limits of human cognition, communication, and agency. They cohered in part because of the comparatively slow evolution of those information technologies—print, mail, the telegraph, the telephone—although each of these catalyzed important changes in the organization of

politics and society. The new spaces of digital culture are no less social and political than the old ones they displace, but they are far less visible to the classic citizen of democracy—the person without expert knowledge. A democratic digital culture will be one in which the inevitable conflicts of values in this technical sphere are recognized and adjudicated through public processes. A democratic digital infrastructure will be one that supports participation in those processes.

references

- Adar, E., & Huberman, B. (2000, September). Free riding on Gnutella. *First Monday*, 5(10). Retrieved August 2005 from http://www.firstmonday.org/issues/issue5_10/adar/
- Anderson, R. (2003). *Trusted computing frequently asked questions, Version 1.1*. Retrieved August 2005 from <http://www.cl.cam.ac.uk/~rja14/tcpa-faq.html>
- Benkler, Y. (2001). *Property, commons, and the First Amendment: Toward a core common infrastructure*. White paper for the Brennan Center for Justice, NYU Law. Retrieved August 2005 from <http://www.benkler.org/WhitePaper.pdf>
- Benkler, Y. (2006) *The Wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.
- Brief *Amicus Curiae* of the Computer Science Professors Harold Abelson, Thomas Anderson, Andrew W. Appel, Steven M. Bellovin, Dan Boneh, David Clark, David J. Farber, Joan Feigenbaum, Edward W. Felton, Robert Harper, M. Frans Kaashoek, Brian Kernighan, Jennifer Rexford, John C. Reynolds, Aviel D. Rubin, Eugene S. Spafford and David S. Touretsky Suggesting Affirmance of the Judgment. Metro-Goldwyn-Mayer Studios, Inc. et al. vs. Grokster, Ltd. (2005). Retrieved August 2006 from http://www.copyright.gov/docs/mgm/computersci_profcs.pdf
- Borland, J. (2003, November 6). Napster to give students music. CNET News. Retrieved August 2005 from http://www.pewinternet.org/pdfs/PIP_Filesharing_Marcho5.pdf
- Erikson, J. (2003, April) Fair use, DRM, and trusted computing. *Communications of the ACM*, 46(4), 34–39.
- Felten, E. (2003). A skeptical view of DRM and fair use. *Communications of the ACM*, 46(4), 57–59.
- Fisher, K. (2007). Congress, RIAA and universities prepare for P2P arms race. *Ars Technica*, June 7.
- Gillespie, T. (2004). Copyright and commerce: The DMCA, trusted systems, and the stabilization of distribution. *The Information Society*, 20(4), 239–254.
- Godwin, M. (2003) *What every citizen should know about DRM, a.k.a. 'digital rights management'*. Washington, DC: Public Knowledge, America Foundation. Retrieved August 2006 from <http://www.publicknowledge.org/content/overviews/citizens-guide-to-drm/attachment>
- Hindo, B. (2004, January 16). Did big music really sink the pirates? *Business Week*. Retrieved August 2005 from www.businessweek.com/technology/content/jan2004/tc20040116_9177_tc024.htm
- Lessig, L. (2002). *The future of ideas: The fate of the commons in a connected world*. New York: Random House.
- Lessig, L. (2004). *Free culture: how big media uses technology and the law to lock down culture and control creativity*. New York: Penguin.
- Litman, J. (2001). *Digital copyright: Protecting intellectual property on the internet*. Amherst, NY: Prometheus Books.
- May, C. (2006). *Digital rights management: The problem of expanding ownership rights*. Oxford, UK: Chandos.
- Mulligan, D., Han, J., & Burstein, A. (2003, October). *How DRM-based content delivery systems disrupt expectations of "personal use."* Paper presented at the ACM Digital Rights Management Work-

- shop. Retrieved June 23, 2004, from http://www.sims.berkeley.edu/~john_han/docs/p029-mulligan.pdf
- OECD. (2004). OECD information technology outlook 2004: Peer to peer networks in OECD countries. chapter 5. Retrieved October 2006 from <http://www.oecd.org/dataoecd/55/57/32927686.pdf>
- Rainie, L., & Madden, M. (2005, March). Music and video downloading moves beyond P2P. Pew Internet Project Data Memo. March, 2005. Retrieved August 2005 from http://www.pewinternet.org/pdfs/PIP_Filesharing_March05.pdf
- Saltzer, J., Reed, P., & Clark, D. (1984). End-to-end arguments in system design. *ACM Transactions on Computer Systems* 2(4), 277–288.
- Stam, N. (2003, September 22). Inside Intel’s secretive “LaGrande” project. *Extremetech*. Retrieved August 2005 from <http://www.extremetech.com/article2/0%2C1558%2C1274119%2C00.asp>
- Stefik, M. (1997). Shifting the possible: How trusted systems and digital property rights challenge us to rethink digital publishing. *Berkeley Technology Law Journal*, 12(1), 137–59.
- Turkey, J. (2002, November 11) Embedded Processors, Part 1 & 2. *ExtremeTech*. Retrieved August 2005 from <http://www.extremetech.com/article2/0,1558,18917,00.asp>
- Valenti, J. (2000, June 15). *Intellectual property: Why it must be guarded and preserved in the long-term best interest of this nation and all who live in it*. Presentation to the House Judiciary Committee’s Subcommittee on Intellectual Property.
- von Lohmann, F. (2002, April 16). *Fair use and digital rights management: Preliminary Thoughts on the (irreconcilable?) tension between them*. Presented at the Computers, Freedom, and Privacy conference. Retrieved August 2005 from http://www.eff.org/IP/DRM/fair_use_and_drm.html

notes

- 1 The key principles of end-to-end networking and its consequences for innovation were developed by Saltzer, Reed, and Clark (1984). A number of scholars have used the end-to-end analogy to describe the organization of a democratic public sphere (e.g., Lessig, 2002; Benkler, 2001, 2006). See also Sack’s discussion of the network metaphor in Chapter 11, this volume.
- 2 IBM’s patented MCA bus for PCs (1987) illustrated this tension. IBM’s effort to charge licensing fees for access to the bus resulted in the rapid development of an open alternative, the EISA bus, by other hardware manufacturers.
- 3 Industry-friendly digital music services such as the “new” Napster have struck deals with Pennsylvania State, Yale University, Wake Forest University, Vanderbilt University, and others. Under most of these arrangements, universities privilege or subsidize the music service on campus in return for diminished liability for copyright infringement on campus networks (see, e.g., Borland, 2003).
- 4 This “super-sharer” structure of the most popular P2P networks is documented and discussed in Free Riding on Gnutella (Adar & Huberman, 2000). They concluded that on the P2P network Gnutella, in 2000, 5% of users were sources for 70% of the transfers. The recent music industry

approach to suing individual users of file-sharing networks is premised not just on the general deterrent effect of highly publicized fines, but on the potential cascading effect of removing the super-sharers.

- 5 ISP-level filtering is the basis of the “Great Firewall” of China maintained by the Chinese government, which restricts both individual websites and types of speech on the Internet.
- 6 Lexmark has invoked the Digital Millennium Copyright act against generic cartridge manufacturers who have reverse-engineered its cartridge authentication protocol; on the HP cartridge expiration date, see <http://www.ddjembedded.com/resources/articles/2002/0209k/0209k.htm>
- 7 Roughly half of these run neither Windows nor Linux, but a Japanese open source operating system called iTron. See Turkey (2002).
- 8 From the Indian Simputer, to Brazil’s \$300 Popular PC, to the more recent One Laptop Per Child initiative, which has targeted a \$100 price point.
- 9 See Microsoft “Palladium” business overview: <http://www.microsoft.com/presspass/features/2002/jul02/0724palladiumwp.asp#challenge>
- 10 See TC pioneer Mark Stefik’s account of the initiative as grounded in the presumption that “the consumer is dishonest” (Stefik, 1997); also Erikson (2003)
- 11 The Windows Genuine Advantage program, distributed through Microsoft’s automatic security updates to Windows XP, implemented an early version of this surveillance in 2006.
- 12 Rates of business software piracy, especially, are said to hover around 70-80% in most developing countries—although the industry data and methods have been disputed. For annual summaries of industry country studies, see the International Intellectual Property Alliance site: <http://www.iipa.com/countryreports.html>
- 13 For Free Software Foundation Richard Stallman’s views on “treacherous computing,” see <http://www.newsforge.com/business/02/10/21/1449250.shtml?tid=19>
- 14 See, for example, Intel’s Lagrande effort, which brings curtailed memory, protected input/output, and sealed storage features into the microprocessor (Stam, 2003).
- 15 See, for example, 2003’s CBTBDA act and 2004’s wildly broad INDUCE act, which would ban any device that could contribute to copyright infringement.

