Testimony of Gregory A. Jackson

Vice President & CIO, The University of Chicago
House Committee on Science and Technology, 5 June 2007

Mr. Chairman, representatives of the great state of Illinois, and members of the Committee, I am pleased to be here today. My name is Greg Jackson. I am Vice President and Chief Information Officer at the University of Chicago, where I have overseen central information technology and services for almost eleven years. Before that I was MIT's Director of Academic Computing, and before that a statistician and faculty member at Harvard and Stanford.

Two high-level policy questions frame our discussion today. The first is whether the copyright law that has grown up around industrially-organized publishing remains relevant and productive in today's widely distributed information economy. The second is to what degree network service providers should be responsible for illegal use of their networks.

I know that the Committee has engaged these larger questions in other hearings. Since I can claim no special expertise with regard to the larger questions, I will concentrate on the two topics I have been asked to address based on my experience at the University of Chicago: how we handle DMCA and related incidents, and the feasibility, in research universities like ours, of technologies one might use to reduce the illegal sharing of copyrighted materials.

My testimony emphasizes five key points:

- The University's business centers on intellectual property;
- Like most of its peers, the University deplores violations of copyright law;
- Market shortcomings are the principal drivers of infringement;
- Network-based anti-infringement technologies fail within high-performance networks, and eventually they will fail more generally; and
- Technological obstacles to behavior have only limited and transitory effects.

Let me begin with a few words about the University. We are a large private institution, one of the world's major research universities. We operate one of Chicago's principal medical centers and, through subsidiaries, two DOE research laboratories, Argonne and Fermi. We have a $2-billion operating budget, 13,000 students, 2,000 faculty, 5,000 staff, 150 buildings in five states and four foreign countries, 25,000 telephones, and – most important for today's topic – a high-performance network using about 2500 switches and routers to connect our 25,000 digital devices to each other, to research universities and labs worldwide, and to the Internet.

1. The University's business centers on intellectual property

Our research produces not only deeper understanding of how the world works, but also concrete products including many inventions and creative works. Our teaching instills in our students not only concrete knowledge and skills, but also insights into what's worth doing and what isn't, what's right and what's wrong. We protect our intellectual property: we patent inventions, copyright works, distribute online journals, value distinctive teaching, and so on. Yet research and teaching, the heart of higher education, also depend on access to intellectual property. This has implications for course materials, for our libraries, for publications, for the University of Chicago Press, for our relationships with outside entities, and in many other domains.

It is important to us that patents and copyrights be enforceable – even though in many cases we license our intellectual property, and especially our research, for free. But it is also important that we be able to do the best possible research and teaching, that technology advance rather than degrade our ability to do that, and therefore that technology promote rather than deter access to intellectual property.

A key challenge for all of us – copyright owners, publishers, transmitters, enforcers, and users alike – is to find the elusive right balance between mechanisms to protect intellectual property and mechanisms to make it accessible. Tradeoffs are inevitable. We should all be working together, across organizational and political lines, to find reasonable, manageable compromises among our diverse needs, rather than unilaterally and adversarially staking out fundamentally irreconcilable positions.

2. Like most of its peers, the University deplores violations of copyright law

The University of Chicago received 57 Digital Millennium Copyright Act (DMCA) complaints in 2006. At the current pace (33 complaints through April 30), we will receive about 130 complaints in 2007. Those complaints involve only about one half of one percent of our community. 58% of last year's complaints involved music, and most of the rest involved movies, TV shows, or software; this year the music percentage has dropped to 52%. (I should note that the MPAA "top 25" listing has an incorrect DMCA count for us – it's about ten times the right number. We have asked MPAA to clarify or correct this, but thus far have received no substantive response.)

The DMCA, as we understand it, requires the University, as a "network service provider," to end violations when we receive a valid, accurate DMCA complaint. (A valid complaint requires that data sufficiently detailed to locate the offending computer, plus various other elements including an affirmation and a signature, be sent to the University's "DMCA agent" – me, in our case.) We deal strongly with DMCA violations. When we receive a complaint, network-security officers first verify that the offending material remains available, or that our network logs confirm the access cited in the complaint (this is what makes a complaint accurate). If the complaint is valid and accurate, network-security officers immediately disable the network connection cited in the complaint, as DMCA requires. In addition, by University policy we identify who was using the connection at the time of the offense, and refer the offender to the appropriate disciplinary process.

For first offenders this means a formal hearing before a Dean (or an HR officer in the case of staff) and a file notation, after which we restore the network connection. (Very few offenders dispute the violation, although many assert – often with good reason – that the offense resulted from negligence rather than intent.) For second offenders we impose a fine of $1000, the proceeds of which become financial aid for others. Over the past five years we have had just six second offenses.

In addition to the disciplinary process for offenders, we communicate broadly with the community on this topic. We deploy humorous but persuasive posters. We discuss the issue at student orientation. Faculty and instructors discuss it in class at relevant moments. It is covered by our acceptable-use policy. About once a year, I personally remind the entire community by email that the University takes DMCA offenses very seriously and that they can result in very negative consequences.

Many of our DMCA offenses, we believe, result not from intentional distribution of copyrighted material, but rather from how hard it is to disable the public-sharing features of peer-to-peer software. Because of this, we publish a Web page providing extensive guidance as to how a user can disable peer-to-peer sharing. Scores of other entities – including RIAA itself – have cited or linked to our materials.

Unfortunately, inaccurate DMCA complaints, discriminatory enforcement, and politically-structured "top 25" lists have proliferated lately. One movie company, for example, has an accuracy rate down around 20%, and even though commercial ISPs in some university towns serve precisely the same numbers and types of students who live in campus dormitories, the ISPs receive no DMCA complaints and never make top-25 lists even when the local university does. This is all becoming very problematic, since these problems waste resources, and the inconsistencies and discrimination cause offenders to dispute rather than accept our guidance.

3. Market shortcomings are the principal drivers of infringement

Media producers provide and protect their online wares inconsistently, incompatibly, inefficiently, inconveniently, and incompletely. For example, music purchased legally from Microsoft can't be used on Apple devices or *vice versa*, pricing seems high, managing keys and licenses is a major hassle, and no one offers Beatles tracks. So long as the right thing remains more daunting, awkward, and unsatisfying than the wrong thing, too many people will do the wrong thing.

Digital rights management (DRM), the principal mechanism vendors use to protect content sold online, involves packaging intellectual property so that it cannot be used without a special digital key. The digital key, in turn, is restricted to a particular customer or device with license to use the content. This is how iTunes, Zune, Ruckus, and Genuine Microsoft Validation work. Customers who want to use content protected by different DRM typically have to use different software – or even different devices – to gain access. Managing keys can be a major hassle, for example when one's device dies or is replaced. Moreover, poorly implemented DRM can disable customers' computers entirely, as one media company unfortunately demonstrated broadly with its CDs not too long ago.

DRM appears to be a good idea. However, it has been plagued by poor execution, and so has come to be a frustrating obstacle rather than a convenient enabler. Moreover, DRM has become a challenge to security specialists and hackers, who delight in showing how easily it can be subverted. This exemplifies the unwinnable arms race and has induced some vendors to begin selling unprotected content, points to which I will return.

4. Network-based anti-infringement technologies fail within high-performance networks, and eventually they will fail more generally

How Networks Transmit Files

Say that person A wants to send a file to person B. If A and B work at universities, the file might be a prepublication draft, a three-dimensional x-ray scatter image of a molecule, or the video of a procedure carried out within a containment facility, but the process would be exactly the same if A were sending a personal wedding video or an illegal copy of *Eleanor Rigby* to B. Here's what happens, in simplified form:

1. A's computer chops up the file (which may first be encrypted, for security) into many small chunks, much as I might cut up a large mounted photograph to make a jigsaw puzzle whose pieces would fit in regular envelopes. A "header" on each chunk contains limited information including as the address of B's computer and the kind of data being transmitted – by analogy, think of the addresses and "contains photo – do not bend" notations on an envelope. The encased chunk is now a "packet," in networking jargon.

2. One by one, A's computer sends packets to the network for transmission to B's computer. A's computer sends other files to other places at the same time. The packets from the other files get shuffled with the B-destined file's packets as they leave A's computer.

3. Once the packets reach the network, bucket brigades of routers and switches pass them along – again mixed with others, and again one by one – until each packet reaches its destination. Although packets headed for the same destination usually follow the same path, a great strength of the Internet is that they need not do so. Network equipment constantly monitors flows, and switches to alternate routes when particular paths get clogged.
4. As packets reach B's computer – some from A, some from other sources – B's computer sorts them and requests re-transmission for any missing packets. It then extracts the chunks of data from the packets and reassembles them into the original file.

I highlight four key attributes of this process. First, files move across the network in *discrete packets*, rather than as whole files. Second, packets are *intermingled* with other packets from other files as they leave the source, as they move across the network, and as they arrive at their destinations. Third, packets going from one source to one destination may follow *different paths* across the network. Fourth, this chopping and scattering is *intentionally designed* into the Internet to ensure reliability, speed, and robustness.

As particularly advanced users of networking, colleges and universities typically deploy networks comprising an array of main switches and routers interconnected in a ring or mesh with tentacles reaching out to smaller switches and routers, rather than connect everything to one telephone-like central switching point. Rings and meshes maximize the robustness and efficiency of networks. As a desirable consequence, they also make internal traffic on campus networks especially likely to follow multiple routes between points.

Much as it's easy to attain perfect network security by detaching computers from networks, it's easy to protect intellectual property by locking it in a strongbox where no one can retrieve it, or by disabling networks that might transmit it. The value of intellectual property depends largely on circulation, however, so using a strongbox or disabling networks reduces the value of the property. Implementing the strongbox or complicating the network diverts resources from more productive pursuits. And so the challenge we are discussing today: Can anti-infringement technologies work without degrading the efficiency and productivity of the campus networks critical to research and teaching?

There are two principal network-based technologies for forestalling, detecting, or reducing illegal network file sharing: traffic shaping and signature matching.

Traffic Shaping

Traffic shaping involves handling packets differently depending on information in their headers. Thus, for example, we might assign Web packets higher priority than email, or Berkeley-bound packets higher priority than Emory-bound ones, or locally-originated packets higher priority than others. Higher priority translates into faster transfers, so varying priorities in this way "shapes" traffic according to policy.

The most common shaping tools are firewalls, which block traffic according to source, destination, or other header attributes. Packeteer, cGrid, Clouseau boxes, or other more sophisticated shapers can also speed or slow traffic according to packet headers. Traffic shaping can be quite effective when offending traffic (a) has stable or predictable header attributes and (b) those header attributes clearly, reliably, and accurately distinguish illegal from legal traffic.

Unfortunately, much illegal file sharing fails these tests. Newer peer-to-peer software routinely switches addresses and ports in increasingly complex ways. It often mixes infringing transmissions with legitimate ones, for example by disguising transmissions as Web traffic or legal transfers. Moreover, a great deal of illegal file sharing no longer uses distinctive peer-to-peer software or protocols. Traffic shaping has thus become rather ineffective against illegal network file sharing, although it remains an important mechanism for network management.

Signature Matching

Signature-matching technologies compare a file's content to a database of abstracted "signatures," and then take specified action when they find a match. The most typical examples are virus or spam checkers, which perform the matching exercise when a computer opens a file or message and block the file if it matches the checker's database. The comparisons necessary for signature matching can be slow, since accuracy requires detailed comparison. However, virus and spam screening appears not to slow things down, mostly because personal computers and email servers operate so much faster than people use files or read email.

Signature matching for network traffic is much more challenging. In order to do high-quality comparison on network traffic, an entire digital file must be available for comparison to the signature database. Accurate signature matching thus entails three requirements: that all packets travel through one network point where they can be gathered and reconstituted, that reconstitution and comparison be as fast as network transmission, and that matching methods and databases identify only illegally transferred files – that is, there can be no false positives. These are the challenges for Audible Magic and similar products.

The requirements for satisfactory signature matching appear unattainable within the typical campus network. (The network border is a separate issue, to which I will return.) First, as I pointed out earlier, a file's packets are mixed in with others, and may travel different routes across the network. This makes gathering and reconstitution *en route* difficult at best, and often impossible. Second, networks are equipped and optimized to transmit packets without decoding anything but headers, and only the headers are standardized and optimized for this purpose. Since campus and research networks carry traffic at very high speeds, there is no practical way to do full-file comparison without seriously degrading network performance. Third, legal and illegal copies of files sometimes are identical. This will become more common as Apple, Amazon, and other companies sell more copyrighted content without DRM.

What about partial signature matching using data from individual packets? In general, even this cannot be done at campus-network speeds, since reading headers does not suffice, and reading anything else slows the network. The larger problem is accuracy: the smaller the basis for comparison, the greater the likelihood of errors, both positive and negative. Compounding the problem, newer peer-to-peer software and other file-sharing mechanisms use strong, increasingly sophisticated encryption to protect or disguise files, and therefore to defeat signature matching.

Border and Host-Based Approaches

Two signature-matching strategies might make technical sense. One is more promising than the other, but neither will work for long.

The less promising strategy involves signature matching on users' computers, rather than the network. Over the past few years, colleges, universities, and other networking providers have very successfully persuaded their users to install anti-virus and anti-spam software on their personal computers. Since signature-matching software works analogously, and the target files are already intact, installing anti-infringement signature-matching software might not degrade the performance of personal computers.

Users like and are happy to use anti-virus and anti-spam software because it reduces problems without constraining or suppressing benefits. Unfortunately, much as we might wish otherwise, experience has shown that many users likely would perceive anti-infringement software in precisely the opposite way. If installation of such software were to be required, compliance and technical workarounds would become major problems. We already see this problem with copy-protected DVDs: users easily and inexpensively replace software that complies with copy protection with software that doesn't.

The requirement might also have serious indirect negative consequences. Users resisting anti-infringement software, for example, might become suspicious of anti-spam and anti-virus software. If this caused a backlash and led users to remove, disable, or bypass those protections, requiring anti-infringement software might not only have failed to achieve its own objectives, but it would also have reversed the Internet-wide security and privacy gains anti-virus and anti-spam software has yielded over the past few years.

The apparently more promising strategy involves the border between campus or dormitory networks and the commodity Internet (that's the regular Internet, as opposed to special high-performance research networks such as Internet2 or National LambdaRail). Commodity connections are expensive, and so colleges and universities typically buy no more capacity or speed than they need. Moreover, all traffic destined for the commodity Internet flows through one or two connections at the typical campus border, so gathering packets seems more feasible than it does within campus networks. As the Committee has heard today, sufficiently fast signature matching therefore might be possible at commodity border points.

But even perfect border screening can succeed only partially and temporarily. For example, it cannot detect or act on file sharing within campus networks. As peer-to-peer encryption becomes more common and powerful, it will become increasingly difficult to identify files. As some vendors begin selling music and movies without DRM, it will become impossible to differentiate legal from illegal transmissions using signatures. False positives and false negatives will increase, thus rendering even border screening ineffective and counterproductive.

5. Technological obstacles to behavior have only limited and transitory effects

I have confined my remarks thus far to technical feasibility. Let me conclude with a broader observation.

Unexpected problems arise in networked environments. In large part this is because fast, extensive networks enable people to do foolish things much faster – and at much greater scale – than they could otherwise. Since colleges and universities started providing high-performance networking to entire communities earlier than anyone else, we have lots of experience assessing and solving problems that arise in intensively networked environments.

An important lesson we have learned is this: When the problems that arise are about personal and organizational behavior, about the rights and responsibilities of community members and citizens, the only successful, robust way to address them is with social rather than technical tools. We must educate people to understand why certain behaviors are counterproductive for their own community or economy. If we do that together – by which I mean owners, publishers, transmitters, and users – collective good will trump individual malfeasance. When we instead restrict behavior technologically, we get nothing but an arms race we can't win.

I hope that this Committee can translate this lesson into effective policy and collaborative practice, and I appreciate the opportunity to provide whatever help I can.

Biography

Gregory A. Jackson is Vice President and Chief Information Officer at the University of Chicago. In this capacity he reports to the President, and manages the University's central computing facilities, telephones, communications, network services, administrative computing, academic computing, computer store, and related entities.

The umbrella organization for these activities, Networking Services and Information Technologies, spends about $70-million annually overall. It employs about 350 individuals (not counting students). Jackson also works closely with the University's widely diverse academic and administrative units to frame and guide more distributed information-technology activities, and to make sure the University makes optimal use of information technology in its education, research, and administration. He serves on University-wide committees, councils, and boards including Budget, Computing Activities and Services, Patents and Licensing, Research Infrastructure, Intellectual Property, Provost Staff, Executive Staff, President’s Council, and various others.

Jackson has served on the Boards for EDUCAUSE, National LambdaRail, and Internet2. He has served as a member of the EDUCAUSE Recognitions Committee, chaired the Internet2 National Planning and Policy Council, and is an active participant in the Common Solutions Group and the Ivy+ and CIC CIO groups. He also has served on the higher-education advisory boards for Dell, Sun, Apple, Microsoft, and Gateway.

From 1991 to 1996 Jackson was Director of Academic Computing for the Massachusetts Institute of Technology. From 1989 through 1991 he was Director of Educational Studies and Special Projects in the Provost's Office at MIT. Concurrently with his administrative work at MIT, Jackson was Adjunct Lecturer in Harvard's Kennedy School of Government, and Lecturer in the Harvard University Extension. From 1981 through 1990 Jackson was Associate Professor of Education at Harvard University (Assistant Professor 1979-81), teaching in the University's doctoral and management programs in higher education. He served as one of the founding Directors of Harvard University's Educational Technology Center, which studied the use of technology to advance educational practice. He also served as Assistant Director of the Joint Center for Urban Studies of MIT and Harvard University, a multidisciplinary research organization then operated by the two universities. Jackson was Assistant Professor of Education at Stanford University from 1977 through 1979.

Trained as a statistician, Jackson has taught analytic methods for clarifying decision making, including statistical and qualitative research methods; policy analysis and evaluation, especially in higher education; and computer programming. At MIT Jackson also taught an MIT freshman seminar on the scientific integrity of murder mysteries.

Jackson has worked extensively on evaluation and planning methods in higher education; on research, instructional, and library computing in universities; and on admissions and college-choice issues including the differential impact of financial aid on minority and majority college applicants. He is co-author of two books – *Who Gets Ahead?* and *Future Boston* – and of numerous articles, reports, and teaching cases.

Born in Los Angeles and raised in Mexico City, Jackson earned his bachelor's degree from MIT and his doctorate from Harvard.

Contact:

Gregory A Jackson, Vice President & CIO
The University of Chicago
5802 South Ellis Avenue, Chicago IL 60637
773-702-2828
gjackson@uchicago.edu