The Architectures of Mandated Access Controls

Lawrence Lessig
Paul Resnick
Draft 1, to be presented at the Telecommunications Policy Research Conference*

* Professor of Law, Harvard Law School; Associate Professor, University of Michigan School of Information. Thanks to Lorrie Cranor for initially suggesting the symmetry between tagging speech and tagging people. Alexander Macgillivray provided valuable research assistance. Comments on this draft are most welcome. Please email them to lessig@law.harvard.edu and presnick@umich.edu. You can find information about the latest version of this paper at http://www.si.umich.edu/~presnick/papers/lessig98/.

Speech, it is said,[1] divides into three sorts — (1) speech that everyone has a right to (political speech, speech about public affairs); (2) speech that no one has a right to (obscene speech; child porn); and (3) speech that some have a right to but others do not (In the United States, *Ginsberg* speech, or speech that is "harmful to minors," to which adults have a right to but kids do not.) Speech protective regimes, on this view, are those where category (1) speech is dominant; speech repressive regimes are those where categories (2) and (3) dominate.

This divide may make sense when thinking about speech within a single jurisdiction. It makes less sense when thinking about speech across jurisdictions. For when considering speech across all jurisdictions, most controversial speech falls into category (3) — speech that is permitted to some in some places, but not to others in other places. What is "political speech" in the United States (Nazi speech) is banned in Germany; what is "obscene" speech in Tennessee is permitted in Holland; what is porn in Japan is child porn in the United States; what is "harmful to minors" in Bavaria is Disney in New York. And relatedly, while every jurisdiction will have some speech to which access is controlled, what that speech is will differ from jurisdiction to jurisdiction. In the general case, access is controlled, but the specifics of which type and for whom change.

This fact about the regulation of access to speech creates special problems when we consider speech in cyberspace. This is because of the "architecture" of that space. By architecture, we mean both the Internet's technical protocols (e.g., TCP/IP) and its entrenched structures of governance and social patterns of usage that themselves are not easily changeable, at least not without coordinated action by many parties. So understood, the initial architecture of the public Internet made it hard to impose jurisdiction-specific access controls, since within this initial architecture, the identity and jurisdiction of the speaker and receiver are not easily tracked, either by law enforcement or by the participants in conversation. Neither is the content of speech readily identifiable as it crosses jurisdictional boundaries.[2] As a result, real space legal rules are difficult to impose in cyberspace. Or put another way, the initial architecture of cyberspace in effect places all speech within category (1).

One possible response to the initial architecture would have been for governments simply to give up on access controls — to treat all speech in cyberspace as if it were speech of category (1). But the evidence suggests that this is unlikely. As the popularity of the net has grown, governments have shown an increasing interest in re-establishing mandated access controls over speech. In the

---

[1] *See* Lawrence Lessig, *What Things Regulate Speech: CDA 2.0 vs. Filtering*, Jurimetrics Summer 1998, at 10.

[2] The content is not easily identified primarily because content on the net is broken into packets, and not all packets will necessarily pass through the same channels. Even if they did, the content could be encrypted, which would further complicate identification.

United States, this speech is sex related;[3] in Germany, it is both sex and Nazi related;[4] in Asia, it is anything critical of Asian governments.[5] Across the world governments are moving to regulate access to speech in cyberspace, so as to reestablish local control.

We take as given this effort at re-regulation. We personally oppose the regulations that many governments seem eager to impose, although we are not in principle against regulability. In any case, the desire of governments to regulate is a feature of the current political reality surrounding cyberspace. This reality should push us to consider the options that regulators face — not because regulators need to be encouraged towards their regulatory end; but because we all should understand the consequences of any particular regulatory strategy.

Our aim in this paper is to suggest a way to think about the trade-offs that might exist among the various ways in which access control could be grafted onto cyberspace. If a government does not leave decisions about access to individuals alone, what options do governments have, and how effective and costly would these options be along various dimensions of value? Given that different jurisdictions will want different restrictions, and given that those restrictions would be differentially costly, our aim is to map how different architectures and different assignments of responsibility might effect these restrictions, and the trade-offs among these alternatives.

The approach is a type of sensitivity analysis. Regulation, in the view we take of it here, is a function of both law and the architecture of the Internet, itself subject to direct and indirect regulation by law. In this essay, we ask first how access can be controlled given the existing array of legal and architectural features. We then consider, second, how changes in the current array might yield a different mix of costs and benefits.

We evaluate the various outcomes of these different legal and architectural choices along four separate dimensions: first, the effectiveness of a particular mix at controlling access; second, the cost to participants in that access, whether sender, or receiver, or intermediary; third, costs to a system of "free speech"; and fourth, other second order effects, including in particular how different architectures might enable other regulation, beyond the specific access control that a given change was designed to enable.

---

[3] *See* Telecommunications Act of 1996, Pub. L. No. 104-104, Title V, 110 Stat. 56, 133-43 (1996) (Communications Decency Act).

[4] *See* Kim L. Rappaport, *In the Wake of Reno v. ACLU: The Continued Struggle in Western Constitutional Democracies with Internet Censorship and Freedom of Speech Online*, 13 Am. U. Int'l L. Rev. 765 (1998).

[5] *See* Geremie R. Barmé & Sang Ye, *The Great Firewall of China*, Wired, June 1997, at 138 and Philip Shenon, *2-Edged Sword: Asian Regimes On the Internet*, N.Y. Times, May 29, 1995, §1 at 1.

For concreteness, we focus on sexually explicit speech. We pick this type because in the American context at least, there are at least two levels of regulation with respect to such speech. Some sexually explicit speech is generally prohibited (obscene speech; child porn); some sexually explicit speech is prohibited only to minors (speech that is "harmful to minors"); and some sexually explicit speech is permitted to everyone. Thus, the range of regulations of sexual speech in the U.S. will be illustrative of the more general question of how access can be controlled when the desires to control differ among jurisdictions and among people within jurisdictions.

A Model of Access Control: Elements

In our model of access control, we consider three relevant actors — a sender (S), a recipient (R), and an intermediary (I). The sender is the party who makes available the relevant speech; the recipient is the party who gets access to the relevant speech; and an intermediary is an entity that stands between the sender and the recipient. As these definitions suggest, nothing in our description hangs upon whether the sender actually *sends* material to the recipient. The model is agnostic about the mode with which the recipient gains access.

These actors, we will assume, know different things about the speech that is to be regulated. We assume that the sender knows about the item that is being "sent," where "knows about" means has knowledge about the character of the speech item at issue. We assume the recipient has information about who she is, and where she resides. And finally we assume that the intermediary has information neither about the content, nor about who the recipient is, or where she resides. Obviously, these assumptions are not necessary. A sender might not have knowledge about the speech she is making available; and a recipient may not know where or who she is. But we assume the general case.

Given this mix of knowledge, a government effects mandated access control through four separate steps. It first defines which transactions are illegal, where "transaction" means the exchange of speech of a certain kind between two kinds of individuals. Second, it assigns responsibility to one or more actors to effect that restriction. Third, it creates a regime to detect when assigned responsibilities are being violated. And fourth, it sets punishments when these responsibilities are violated.

In the balance of this part, we sketch issues relevant to each of these elements of a regulatory regime. In the next part we conduct the sensitivity analysis.

(1) Defining Blocked Exchanges

A regulatory regime first defines a set of illegal transactions, or "blocked exchanges." The criteria for deciding whether an exchange is blocked include: (1) the type of speech item exchanged; (2) the recipient, and (3) the rules of the recipient's jurisdiction. We can state this relation as follows:

(a) $B(I, R, J) = \{Y, N\}$

Where I = item type, R = recipient type, and J = jurisdiction type, and B() is a function determining whether exchange of the speech item is blocked.

Stated alternatively, a blocked exchange is access to a given item type by a given individual within a given jurisdiction which the law deems illegal.

Within this model, there may be "floor" recipients, and "floor" jurisdictions. In the specific context of sexually explicit speech within American jurisdictions, children are a floor recipient type (anything that is permitted to children is permitted to adults as well), and a Bible Belt small town may be a floor jurisdiction (anything that is not blocked there would be permissible everywhere). More formally, with j_f denoting a floor jurisdiction:

(b: floor type) For all I, J:
B(I, child, J) = N implies for all R, B(I, R, J) = N

(c: floor jurisdiction) For all I, R:
B(I, R, j_f) = N implies for all J B(I, R, J) = N

The two floors can be combined. Anything that is permitted to children in a floor jurisdiction is permitted to everyone in every jurisdiction:

(d: floor recipient and jurisdiction) For any I:
B(I, child, j_f) = N implied for all J B(I, child, J) = N.

In the general case, either the sender's or the recipient's jurisdiction may determine that an exchange is blocked. United States laws regulating cryptography, for example, restrict a sender's right to send certain encryption related material to another jurisdiction; French crypto laws regulate a receiver's right to receive such material.[6] For simplicity, however, we will focus on blocked exchanges in the recipient jurisdiction alone. This focus will be significant in the context of enforcement, since governments can more easily control their own populations than populations in other jurisdictions.

A jurisdiction, on this model of blocked transactions, may determine that a particular transaction is to be blocked in at least two different ways.

(1) A jurisdiction might publish criteria defining what is to be blocked, but require a judgment by the parties about how to apply that criteria. The jurisdiction may or may not then hold parties responsible for correctly making such judgments prior to a determination by the regulating jurisdiction.

(2) A jurisdiction may classify specific items as acceptable or blocked for particular recipient types. Such classifications could function as a form of pre-clearance of allowable speech, with the government promising not to prosecute parties for decisions made in good faith based on the pre-clearance, or the jurisdiction could define a list of prohibited speech. The determinations about acceptability may occur through a judicial or

[6] *See* Stewart A. Baker & Paul R. Hurst, The Limits of Trust 130 (1998).

administrative process, or the jurisdiction may delegate its authority to an independent rating service.[7] A jurisdiction might even rely on a computer program to provide an initial classification of the speech at issue, and publish that classification as a pre-clearance, perhaps with a stipulation that the initial classification might be changed in the future after human review.

In the American context, the ordinary procedure follows case (1). If a jurisdiction follows case (2), publishing a list of blocked items for a given recipient type, then the list of items must, ordinarily, be judicially specified.[8]

We believe that it is unclear, in the American context, whether a regime of voluntary pre-clearance would be permissible. On the surface, it would seem that any step that would reduce the uncertainty surrounding the distribution of speech would be speech enhancing. But some who have considered the matter believe that if this voluntary regime became, in effect, mandatory, with speech not appearing on a pre-clearance list being restricted, it would then become constitutionally suspect.[9] The net effect on speech, however, is not clear: Lower costs could lead to less chilling of speech (if it is clearer what is prohibited and what is not) but to more control on speech (if it results in greater prosecution of improper speech.)

(2) Assignments of Responsibility

The second step that the regulator must take is to define how best to allocate responsibility among actors to assure that access is controlled. Responsibility for controlling access could be assigned either exclusively to one actor among the three, or jointly, to any two, or all three. In this version of our analysis, we consider only exclusive assignments of responsibility for blocking, though we do consider requiring other parties to provide information to the blocking party.

By hypothesis, no party knows enough to determine whether a particular exchange should be blocked. (Again, the sender does not know the recipient; the recipient does not know the content of the item; the intermediary does not know either.) The law must therefore create an incentive for parties to produce the

---

[7] An example would be a third-party PICS-based rating service such as NetShepherd, or a proprietary rating service such as NetNanny. *See NetShepherd's Home Page* (last modified July 14, 1998) <http://www.netshepherd.com/> and *Net Nanny Main* (visited August 22, 1998) <http://www.netnanny.com/>.

[8] *See* Paris Adult Theatre I v. Slaton, 413 U.S. 49, 55 (1972) (Injunction could be used so long as adequate procedures to determine obscenity had been used).

[9] *See* Frederick Schauer, *Fear, Risk and the First Amendment: Unraveling the "Chilling Effect"*, 58 B.U.L. Rev. 685, 725-29 (1978). The closest case is perhaps Bantam Books v. Sullivan, 372 U.S. 58 (1963), where the Court invalidated a "blacklist" Commission. The pre-clearance idea is not quite a blacklist — the result of the submission would be a promise not to prosecute, not a determination that the material was "obscene." Again, however, we concede that the line is a difficult one to sustain.

information necessary to determine whether access should be blocked. The law can create this incentive by allocating liability to an actor for failing properly to block a transaction, or by setting a default rule for improperly blocked transactions.

We consider two such defaults.[10] Under the first rule, the sender is liable if she enters a transaction that is later determined to be illegal, without reliable indicators that in fact the transaction was legal. We call this the "prohibited unless permitted" rule, and as we have formulated it to turn on the steps taken to comply with the law, we believe it is distinct from a prior restraint.[11] Under the second rule, the sender is liable only if she enters a transaction that is later determined to be illegal, in the face of indicators that in fact the transaction was illegal. This is the "permitted unless prohibited" rule, and it is equivalent to a specific intent to violate the law.[12] One modification of this latter rule would hold the sender responsible if the sender should have known that the transaction is illegal. This would comport with a negligence standard, and we consider this change where relevant in the analysis below.

In cases of uncertainty, the "prohibited unless permitted" rule will be overbroad (it will block more speech than the state has a legitimate interest in blocking), and the "permitted unless prohibited" rule will be ineffective (since there will be insufficient incentive to discover the relevant information).[13] Our focus, therefore, will be on changes that might reduce the uncertainty that actors who are responsible for blocking exchanges face. Stated abstractly, these changes will be either changes that tag speech, or tag people. If speech is tagged, then it is

---

[10] We are not claiming at this point that either default would, for all types of speech, be constitutional under the American constitution. Nor are we speaking about the burdens of proof under a particular statute. We will assume throughout that the state bears the burden for all elements of the charge. *See* Smith v. California, 361 U.S. 147 (1959). Rather than a claim about what is constitutionally possible, our defaults will help clarify the relationship between the proscription and uncertainty. Like Schauer, our objective is to further explore this relationship, and the constitutional implications of uncertainty. *See* Schauer, *supra* note 9, at 725-27.

[11] It is distinct because there is no requirement of not sending, there is simply a punishment for sending without indication that the sending is legal. We concede this is a fine line, but the purpose of our defaults, as we have explained above, is not so much to limne the contours of constitutionalism, but to understand the relationship between these rules and uncertainty.

[12] Specific intent is equated to acting knowingly in the Model Penal Code. The relevant section reads:

> A person acts knowingly with respect to a material element of an offense when:
> (i) if the element involves the nature of his conduct or the attendant circumstances, he is aware that his conduct is of that nature or that such circumstances exists; and
> (ii) if the element involves a result of his conduct, he is aware that it is practically certain that his conduct will cause such a result.

Model Penal Code § 2.02(2)(b) (Official Draft 1985).

[13] The Restatement (Second) of Torts §282 defines negligence as: "conduct which falls below the standard established by law for the protection of others against unreasonable risk of harm."

easier for an intermediary or recipient to determine item types; if people are tagged, then it is easier for an intermediary or sender to identify recipient and jurisdiction types. Our aim here is to consider the various consequences of these different alternatives.

(3) Monitoring and (4) Enforcement

The final two regulatory steps are first, devising schemes for monitoring compliance, and second, implementing schemes for enforcing rules against non-compliant actors. In both cases, where the target of regulation sits, relative to the regulating regime, is an important factor in selecting among regulatory regimes. And in the case of monitoring, the technology used to effect the access control will significantly alter the costs of monitoring. Some technology, that is, would be open for an automated and random verification; others would not.

The major issues for enforcement all involve the question whether the target of enforcement can easily, or cheaply, be reached. We assume there are more receivers than senders, so one might believe targeting senders would be cheaper than targeting receivers. This, however, is complicated if the sender is outside the regulating jurisdiction, making the sender sometimes legally, and if not legally, then often practically, beyond the reach of the regulating jurisdiction. The cost of enforcement against aliens, there, may mean that it is cheaper to enforce a rule against receivers than senders.

Whether there are more receivers or listeners, however, there are certainly fewer intermediaries than either. Intermediaries, as we discuss below, may therefore be the optimal target of regulation, even though they have even less information than either the sender or receiver. Again, the savings of enforcing a rule against them may be greater than the cost of their obtaining the necessary information.

SENSITIVITY ANALYSIS

We consider now the consequences of various allocations of responsibility among our three actors, and within each, consider how changes in existing law and architecture might better achieve the aim of access control, with fewer free speech costs, or more access control effectiveness.

Sender responsible for blocking access

Our first rule would make the sender responsible for controlling access. To comply, the sender must determine both the law of the jurisdiction of the recipient, and depending upon that law, the character of the recipient. Under the present architecture, both determinations are costly. There is no simple way to determine the jurisdiction within which the recipient resides, and no way to be certain of characteristics of the individual.[14]

[14] A web server, for example, knows the IP address of the client computer that requests a web page, but usually knows little else about the recipient. An IP address does not readily identify a

Under the "prohibited unless permitted" rule, the cost of determining eligibility is likely to present a significant chill on the speaker's speech.[15] The sender would have to take steps independent of the architecture of the net to determine where a recipient is — by verifying an address, for example, or using an area-code on a telephone number as a proxy. And the sender would need to rely upon proxies from credentials (such as a credit card) to guess whether the individual is a proper age or not.

The United States Supreme Court has permitted this regime in the context of obscenity — where the sender must determine both the law of the jurisdiction and the jurisdiction relevant for the recipient.[16] It has not directly ruled on the same question in the context of speech "harmful to minors," where the sender must determine, in addition to the jurisdictional information, the age of the recipient. In the *Reno v. ACLU,* the Court did cite the burden of verification as one reason that the CDA's "indecency" provision was constitutionally suspect.[17] But *Reno* did not address the "harmful to minors" standard — or as described by some, the obscene-as-to-minors standard — and there is no clear indication by the Supreme Court that the test would be different.

If, on the other hand, the rule is "permitted unless prohibited," the existing architecture would make any access control ineffective. While in real space, certain facts about an individual are generally unavoidable, or self-authenticating (a 10 year old boy doesn't look much like a 20 year old man), in cyberspace, such facts are not self-authenticating. To determine either the jurisdiction or the age of the recipient requires affirmative steps by the sender. If no obligation to take such steps exists, or if no requirement exists to block unless such steps are taken, then the rule will not effect the intended access control.

The existing architecture therefore creates a great burden for the sender if the default is "prohibited unless permitted," and it defeats access control if the default is "permitted unless prohibited."

Sensitivity

---

geographic location, because the administrative practices surrounding IP address allocation have not been based solely on geography. By analogy with the telephone numbering system, IP addresses have been allocated more like 800-numbers than like the numbers in regular area codes. Moreover, there is currently no single up-to-date database indicating the location of the computer using each IP address. (In practice, to facilitate routing, address allocations do roughly follow geography, which means that such a database might not be too unwieldy if it were assembled). An IP address does not even uniquely identify a recipient computer, since dial-up connections through an Internet service provider typically are assigned a different address each time they dial.

[15] Though the use of the word has become quite general, we attempt in this essay to follow Schauer's definition of "chill," which refers "only to those examples of deterrence which result from the indirect governmental restriction of protected expression." Schauer, *supra note* 9, at 693.

[16] *See* Hamling v. United States, 418 U.S. 87, 104-06 (1974).

[17] *See* Reno v. ACLU, 117 S. Ct. 2329, 1997 U.S. LEXIS 4037, 24 (1997).

Some of the burden on the sender created by this rule could be reduced by certain architectural and legal changes. In this section we describe four, and consider the potential costs and benefits of each.

The first two changes involve ways more cheaply to identify facts about the recipient. The two facts unknown by the sender are the jurisdiction of the recipient, and characteristics of the recipient (that she is, for example, over 18.) The changes described here would facilitate the sender knowing both facts at a relatively cheap cost.

The first technique relies on digital certificates.[18] In the standard model of certificates, certificates identify who someone is. They are digital objects cryptographically signed by a certificate authority. The dominant use of such certificates today is to certify the identity of the holder. This is the model, for example, of the Verisign Digital ID, which Verisign describes as a "driver's license for the Internet."[19]

But there is no reason that the same technology couldn't be used to certify facts about the holder — or, more generally, to certify any assertion made by the signer. In our case, a signing certificate authority (CA) could then certify that X is from Massachusetts, and that X is over the age of 18, without identifying who X is.[20] Senders would then examine these certificates before granting access to regulable speech. Access would then be granted without a terribly cumbersome system of passwords, or ids.

We can call this a "credentialing" solution.[21] It requires that the sender make certain judgments about the speech at stake; but it allows the sender to rely upon representations about jurisdiction and the recipient that are necessary to determine whether an exchange is or is not blocked.

Under a "prohibited unless permitted" regime, access would be blocked except to those who could show that they carry the proper credentials. All else being equal, certificates would lower the cost of such a showing, and therefore reduce the burden, and hence chill, of the access control regime. Moreover, the burden on individuals under such a regime would be lower than under a regime where they must show a credit card, or other form of identification. The cost of a

---

[18] *See* A. Michael Froomkin, *The Essential Role of Trusted Third Parties in Electronic Commerce*, 75 Or. L. Rev. 49 (1996) <http://personal.law.miami.edu/~froomkin/articles/trustedno.htm>.

[19] *See Verisign* (visited August 22, 1998) <http://www.verisign.com/>.

[20] David Chaum was an early proponent of such characteristics certificates rather than identity certificates. *See* David Chaum, *Security Without Identification: Transaction Systems To Make Big Brother Obsolete*, 28 Comm. of the ACM 1030 (1985).

[21] Note that even though the technology for this solution is already in place, we refer to it as a possible architectural change, because a widespread change in social practices would be necessary for the technology to be used in this way.

certificate should be less than the cost of a card, and the possibilities for anonymity should be greater.

Alternatively, widespread use of digital certificates could also improve the effectiveness of a "permitted unless prohibited" regime, by providing senders with enough information correctly to block exchanges that would otherwise have been permitted by default. In this case, however, recipients have no natural incentive to provide credentials, since such credentials can only cause otherwise permitted exchanges to be blocked. Suppose, for example, that laws prohibited knowing transmission of certain materials to minors in some jurisdiction, but some other jurisdiction had no such restriction. Recipients in the restrictive jurisdiction would have no incentive to provide certificates identifying their jurisdiction, since withholding the credentials would enable the sender to assume that the recipient was in an unrestricted jurisdiction.

Thus, in order to achieve widespread use of jurisdiction and recipient type credentials under a permitted unless prohibited regime, it will be necessary to impose additional legal rules that require recipients, in certain situations, to provide relevant credentials. To minimize the burden of this rule, the requirement could be that the recipient provide the certificate only if the server asks, and the server asks only if the material is illegal in at least one jurisdiction. This regime would still burden recipients living in jurisdictions where the speech was wholly legal; its viability would rest then upon the significance of that burden.[22]

An alternative might be to impose on intermediaries the requirement that they assure users have valid certificates. If the state requires intermediaries to facilitate the supply of certificates then the cost of monitoring and compliance might be lower than if the same role was being performed by the state. The intermediary's advantage is not over the primary conduct — certainly senders and receivers are in a better position to certify than intermediaries — but in assuring that the primary conduct is properly regulated.

A second architectural change to help the sender identify the jurisdiction into which speech was to be sent would be an IP map — a table that would give a rough approximation of the location of the recipient's computer.[23] No doubt the map could not be perfect, and senders or recipients could use proxies to escape the

---

[22] Another possibility would be for the server to send a request of the form "if you are in jurisdiction X or Y, please provide a credential attesting to your age," which would further reduce the burden of the system.

[23] Currently, the InterNIC maintains a database of which organization each IP address was allocated to. This database is public and a copy of it may be queried from any computer on the Internet. Unfortunately, some entries in the database are incomplete or out of date, and they do not necessarily identify the location of computers using the IP addresses. It has been suggested, however, that this database be used as a starting point for developing an IP to jurisdiction mapping. *See* Philip McCrea, Bob Smart, & Mark Andrews, *Blocking Content on the Internet: A Technical Perspective*, Appendix 5 (visited August 22, 1998) <http://www.noie.gov.au/reports/blocking/index.html>.

consequences of the map. But in the main, the map might suffice sufficiently to segregate restrictive jurisdictions from nonrestrictive.[24]

An IP map would provide benefits over a certificate system. Under the "prohibited unless permitted" regime, an IP map may burden speech even less than the certificate regime, since the cost to the recipient of this form of identification is zero, and the processing costs to the server would be lower than processing a certificate. The "permitted unless prohibited" regime becomes more effective as well, since now the sender has an assured way of knowing the jurisdiction into which the material is being sent, though not information about the recipient's age or other characteristics.

But there are important social costs associated with this IP-to-geography mapping. These flow from its generality. Since jurisdiction identification would be determinable with any IP transaction, the regime would effect jurisdiction identification independent of the kind of speech being accessed. This raises obvious privacy concerns, which might be mitigated by structures that would limit the use of the list for specific purposes. But for obvious reasons, it would be difficult to limit the use of this information.

The final two architectural changes would aid senders in classifying their speech according to the categories of various jurisdictions. While we presume that the sender knows about its speech, it may not understand the classification scheme of every legal jurisdiction. As discussed on page 4, pre-clearance of the sender's materials can eliminate the sender's uncertainty. If the pre-clearance is judicial, the cost of the clearance would still be high. It may be possible to use automated regimes to facilitate this certification, so long as the government had the right, prospectively, to change its mind about a certification.

A second way to reduce uncertainty about how to classify items according to particular jurisdictions' categories would be a thesaurus that relates the categories of different jurisdictions. Thus, if the sender is able to classify an item according to one jurisdiction's categories, it could infer the classification in some other jurisdictions. For example, it may be that anything classified as child pornography in jurisdiction A would be classified as obscene in jurisdiction B, though the converse inference might not hold. The thesaurus functions as a more complex version of the base jurisdiction model that we described in Equation C on page 4.

Recipient responsible for not taking access

Our second rule would make the recipient responsible for illegal transactions — targeting the buyer, that is, rather than the seller. This rule has some advantages over the sender-responsible rule — the recipient, for example, may be

---

[24] We note that already, companies such as Microsoft are using IP addresses to assure themselves that the user is within the United States, so that Microsoft does not become an "exporter" of high grade encryption technology. *See Internet Explorer Products Download: Microsoft Strong Encryption Products (US and Canada Only)* (visited August 23, 1998) <http://mssecure.www.conxion.com/cgi-bin/ieitar.pl>.

in a better position to know about the law of its jurisdiction, and about its own recipient type.

But the recipient is in a worse position, relative to the sender, to know about the kind of information that the sender is making available. A sender may find it burdensome to classify its speech according to a particular jurisdiction's categories, but at least the sender begins with knowledge about the content of the speech at issue.[25] The receiver does not. This means that a recipient cannot determine the legality of an exchange until after the exchange has occurred. Thus, under a "prohibited unless permitted" rule, the receiver risks liability[26] in the very act of determining whether a particular exchange complies with the law. And if the rule is "permitted unless prohibited," then restriction is likely to be completely ineffective.

A second problem with placing liability on the receiver is that if the receiver is classifying speech for someone else (her child, for example) then the costs of classification might be so high as to create an incentive for a highly conservative classification.

Finally, putting the responsibility on the receiver may increase the costs of enforcement. Receivers are ordinarily individuals, and a regime that regulates only individuals would either be costly to monitor, or dangerous because of selective enforcement.

Of course these burdens are contingent, and it could be that they in fact indicate something very different. It may be that given the distribution of recipients and senders, and the likelihood that more recipients will be within the control of the regulating jurisdiction than senders, punishing recipients will be cheaper than punishing senders. While there may be more recipients than senders, their compliance could be checked randomly; assuming the probability of detection were high enough, their compliance may be easier to secure.

Sensitivity

A recipient-responsible rule could be made less costly if there were cheaper ways to identify item types. An obvious solution here is a kind of labeling or rating of items. Two sorts of labeling are possible. One, we have already described — pre-screening — and here the same techniques for reducing the costs of pre-clearance would apply, including the use of automatic text classification and delegation of the pre-clearance powers to an independent third-party rater. As we

---

[25] It would be different, of course, if the sender were considered as a bookstore, without knowledge, or any simple way to get knowledge, about the content of its books. *See* Cubby v. Compuserve, 776 F. Supp. 135 (S.D.N.Y. 1991). We would consider such a "sender" to be an intermediary in our analysis.

[26] This depends upon the level of knowledge required for someone to be guilty under such a provision. If the statute were criminal, the knowledge requirement would be quite strong, so inadvertent liability would not be possible. But for a lesser prohibition, the knowledge requirement may be less.

mentioned before, however, there remains a concern about the constitutionality of even a voluntary pre-clearance regime. In the American context, despite the reduction in uncertainty, this might be a prohibited regulatory change.

The other solution is to create the incentive for senders to label, by giving recipients immunity if they in good faith rely upon a sender's labeling structure. This solution simply inverts the certificate solution — since here it is the sender that is offering a "certificate" and the receiver who is relying, while in the case above, it was the recipient providing the certificate, and the sender who was relying. The analysis is also analogous.

A "prohibited unless permitted" regime would give senders an incentive to provide rating labels, since they would allow more recipients to receive the speech. There would of course be a transition period, during which only a small percentage of materials would carry self-rating labels, rendering most of the net blocked under a strict "prohibited unless permitted" rule.

Under a "permitted unless prohibited" regime, there would be little incentive to label by the sender, since that could only reduce legal access. However, if a significant population adopted a filtering solution voluntarily, then the market demand for labels might be a sufficient incentive on its own.[27]

In either case, one difficulty with labels might be addressed by requiring senders to provide labels. This may raise a constitutional question in the United States if labels were considered compelled speech. Some have argued that they would not,[28] but this is a question nonetheless. To reduce the cost to senders of labeling, a government might subsidize third party rating or itself produce suggested ratings. Its rating could not be definitive[29] in such a system, but may be an aid to senders in self-labeling.

The burden of labels might be minimized by simply requiring labels only where speech is potentially regulable (comparable to requiring that people up to the age of 25 carry IDs to purchase cigarettes, even though the prohibition reaches only those 18 and under). Even here, however, the requirement raises difficult questions, since it is requiring speech by the sender in the form of a label even if the underlying speech is clearly legal in the jurisdiction into which it is

---

[27] For example, consumers might turn on the facilities in Microsoft's Internet Explorer (version 3 and higher) or Netscape Navigator (version 4.5) to voluntarily block access based on senders' PICS-formatted self-labels. In order for this to create an incentive for senders to label, however, the voluntary filters that consumers installed would have to follow a prohibited unless permitted rule.

[28] *See* R. Polk Wagner, *Filters and the First Amendment* (visited August 22, 1998) <http://www.pobox.com/~polk/filters.pdf>.

[29] In no case could the government's own ratings be determinative of whether speech were delivered or not, absent a judicial finding. *See* Rowan v. U.S. Post Office, 397 U.S. 728 (1970).

being sent. Thus the most restrictive jurisdiction is determining whether the speaker must label.

We note one interesting constitutional asymmetry. It seems plain that there could be no law that required receivers not to receive any speech unless it were plain that the speech was legal. (A rule, that is, that would punish the receiver if she received speech without a clear indication that it was not illegal.) Even with respect to obscenity, that restriction would be too overbroad, or fail a minimal mens rea analysis. It is at least arguable, however, that a law that required senders to check every digital ID before sending material was not constitutionally invalid. Analytically these two regimes are quite similar: In both, the transaction is conditioned upon verification of its legality. But the burden on the sender is likely less constitutionally troubling than the burden on the receiver.

Beyond the constitutional issues, there is a practical enforcement problem with mandating that senders provide labels. Just as it may be difficult to enforce blocking requirements across jurisdictional boundaries, it may be difficult for authorities in one jurisdiction to enforce a labeling requirement in another.

Intermediary responsible for blocking

We have assumed that the intermediary has neither information about the recipient, nor about the item the sender would send. It might therefore seem odd to consider the intermediary as a possibly responsible actor.

But this intuition is misleading. Intermediaries are a cheap target of regulation. There are fewer of them than receivers or senders, and they are typically more stable, or harder to move. Just as it is easier for the government to regulate telephone companies than it is to regulate telephone users, it would be easier for the government to set requirements on intermediaries which intermediaries could then enforce upon their customers. More importantly, because intermediaries have an interest in reducing the cost of compliance, regulating intermediaries is more likely to get innovation in the methods of compliance.

In addition to a lack of information, intermediaries may have limited capabilities for implementing blocks. Blocking can either be implemented at the application layer (e.g., web page requests) or at the network layer (i.e., individual packets). Network layer blocks are of necessity much cruder: only the sender's and receiver's IP addresses and the port number (a rough indicator of whether the connection is being used for a web transfer, email, or something else) are available. Thus, a network layer block can either block all web requests to a particular IP address, or none of them.[30]

We consider two types of intermediaries. One type is an Internet access or service provider or an employer or a school (for simplicity, we will refer generically

[30] For a more complete description of application layer and network layer blocking, *see* McCrea, *supra* note 23.

to any of these as an IAP). Many but not all IAPs run proxy servers (and other application layer gateways) which intercept some kinds of Internet traffic. Most commonly, a web proxy at an IAP will keep copies in a cache of frequently accessed web pages; when a customer requests a cached page, the proxy sends it to the customer, without fetching it again from the sender's web server. Proxy servers permit application layer blocking: requests for certain URLs can be blocked. Moreover, an IAP may configure a firewall that forces all requests to use the proxy server. This is done most frequently to enhance corporate security, by restricting the Internet traffic entering and leaving a corporation to only that which passes through proxies. In those cases where an IAP does not employ proxy servers, however, only cruder network layer blocking is possible.

The second type of intermediary is a cross-jurisdiction transit point — for example, the first router on a path inside a jurisdiction. Such transit points do not normally employ proxy servers or other application layer gateways. Thus, only the cruder network layer blocking is possible at cross-jurisdiction transit points, given the current Internet architecture.

One final difficulty with blocking by intermediaries is that recipients may find ways to bypass the blocks, especially if the senders cooperate. For example, the same prohibited document may be available from several different URLs, so that a recipient can access one even if the others are blocked. A technique known as tunneling, where the contents of one packet are wrapped inside another packet, may bypass a network layer block.[31]

Sensitivity

Given how fundamental the architectural features are that yield the conclusion that intermediaries are not in a position to control access, one might well conclude that it would be unadvisable to make any changes to increase their ability to control. However, because intermediaries are also practically easier for a jurisdiction to regulate, we will consider what changes might makes this control possible.

A combination of the architectural changes discussed in previous sections could provide intermediaries with enough information to decide which exchanges to block. That is, information about item types could come either from senders' labels or from pre-clearance lists provided by jurisdictions. Information about recipient type could come from certificates and information about recipient jurisdiction could come either from certificates or from a database lookup on the IP address.

One potential change in the architecture to facilitate the implementation of blocking would be to require an application-layer gateway at IAPs or cross-jurisdiction transit points, and require that all customer traffic use these gateways (perhaps enforced through a firewall). This would have high costs for Internet

---

[31] McCrea et. al. detail these and other ways that senders and recipients might bypass intermediaries' blocks. *See* McCrea, *supra* note 23.

flexibility and operation. First, it would be computationally expensive to assemble all packets into messages at cross-jurisdictional transit points, especially for traffic where there is no counter-acting performance gain from caching. Second, messages may be encrypted for privacy or security purposes (e.g., in SSL connections) so that even at the application layer only crude blocks based on sender and receiver address are possible. Third, innovations that introduce new applications would be stifled, since the application layer gateways would not initially know about the new applications and hence would block them.[32] The Internet's current architecture has enabled experimentation and rapid deployment of new applications (examples of applications that blossomed in part as a result of this flexibility include the world wide web, push services, and ICQ[33]). One final cost might come in the form of reliability. It is relatively easy for a service provider to provide multiple routers, so that network layer service is not interrupted if a router is temporarily disabled. It may be more costly to arrange for continued service if an application layer gateway is temporarily disabled.[34]

CONSEQUENCES

Our focus so far has been on how to effect mandated access controls. We have considered three different techniques — tagging the sender, tagging the recipient, or regulating the intermediary to help effect either of the two taggings.

The aim in this section is to consider side effects of each strategy. All three strategies envision a general infrastructure that can be used for purposes beyond those initially intended. We consider these other purposes "side-effects" of the initial regulatory objective. These side effects, we believe, should be accounted for in selecting among strategies for regulating access.

IDs and Regulability

To effect sender or intermediary control, we envisioned the development of identity certificates designed to facilitate the credentialing of certain facts about a recipient — how old that person is, where she is from, etc. We also proposed the development of a database that maps IP addresses to jurisdictions. But it should be clear that, if enabled for this purpose, these certificates and databases would have

---

[32] Many corporate firewalls do prevent employees from using experimental applications that the corporate proxy or gateway is not configured to handle. *See* William R. Cheswick & Steven M. Bellovin, Firewalls and Internet Security: Repelling the Wily Hacker 75 (1994).

[33] ICQ ("I Seek You") maintains a worldwide registry of users and their status (online, busy, away, etc.) allowing users an easy way to keep track of friends and acquaintances. The ICQ client software interacts with the registry updating a user's information and receiving information about others on that user's "contact list". The ICQ client also acts as a platform for chat and other message exchange between any two registered ICQ users. *See What Is ICQ?* (visited August 23, 1998) <http://www.icq.com/whatisicq.html> and *How To Use ICQ*(visited August 23, 1998) <http://www.icq.com/icqtour/new-quicktour.html>.

[34] *See* McCrea, *supra* note 23.

significance well beyond this purpose alone. They might facilitate other jurisdiction-based regulation or voluntary access controls imposed by senders.

Certificates or IP databases would facilitate a more general structure of jurisdiction-based control, including taxation and privacy regulations. The reason is fairly straight forward. Local jurisdictions at least have the legal authority to regulate their own citizens, both while the citizens are at home, and while they are away. A certificate rich Internet would facilitate the identification, then, of who could be regulated by whom, or what standards could be imposed on whom. And this, in turn, could facilitate a more general regulation of behavior in cyberspace.

We might imagine the model to look something like this: States would enter a compact, whereby they, as a home jurisdiction, agree to require senders, or intermediaries, within their own jurisdiction, to respect the rules of other jurisdictions, in exchange for senders, or intermediaries in other jurisdictions doing the same for the home jurisdiction. These rules would specify the restrictions imposed on citizens from a given jurisdiction, and the range of citizens for whom the restriction applies. For example, a jurisdiction might specify that citizens from it may not engage in Internet gambling; the jurisdiction within which a gambling server sits, then, would require the server to check for a person's citizenship, and condition access based upon whether they held the proper credential. And presumably the jurisdiction would do this only if there were restrictions that it wanted imposed in other places, and which it needed other jurisdictions to respect.

If a jurisdiction database or a credential-rich Internet were in place, we might expect voluntary uses of that infrastructure to proliferate. Some voluntarily imposed restrictions might seem reasonable. For example, recording companies might refuse access to their web sites from countries where pirated copies of intellectual property were rampant. Other voluntary uses might not have such sanguine effects. For example, some Serbs and Croats might refuse to allow each other access to their web pages. In both cases, a form of discrimination is being enabled by the certificate infrastructure.

Labels and Improper Control

The alternative solution that we have identified for effecting mandated access control works to label content as a way to facilitate filtering by recipients or intermediaries. The labels might be provided by senders or by governments in the form of pre-clearance lists. An inexpensive and widely used labeling infrastructure would have its own secondary impacts, including both the possibility of more widespread speech regulation and voluntary or collective uses of labels for blocking beyond the state's legitimate censorship interest.

First, if available speech labels describe categories beyond those that a jurisdiction would normally regulate, the mere availability may tempt regulation within these new categories. Thus, the widespread use of a general labeling infrastructure may start governments on a slippery slope toward regulating all sorts

of speech, even if the initial impetus for labeling is limited to only a few kinds of speech.[35]

Second, labels might be used for voluntary access controls as well as mandated access controls.[36] That is, recipients or intermediaries might choose to block more exchanges than governments require. Parents in the United States, for example, may choose to block young children's access to hate speech or speech about sex education, even though such speech is legal for children in the United States. Alternately, a search engine may provide a filtered search service that, when queried for "toys", returns links to pages describing children's toys rather than sex toys, without necessarily reporting that certain sites have been blocked.[37]

The availability of voluntary access controls by parents and teachers is widely viewed as socially beneficial, since it give control to people who can tailor controls to individual and local needs. In a world of perfect transparency and competition, such control imposed by IAPs or search engines as well may be unproblematic.

But in practice, there are a number of reasons why these access controls might be less than ideal.

· First, consumers may have a hard time determining which blocks are in their own best interest, as the criteria for selection may not be transparent, or readily understandable.[38]

· Second, even if the criteria were transparent, the present architecture would still allow filtering "upstream" (for example, by a search engine) without the consumer knowing (thus a nontransparency not about the rating, but about who is effecting the filter.)[39]

· Third, individuals may face a social dilemma about whether to adopt filters. Individuals may themselves prefer to have filtered content (to perfect their own choice), but not want society to have filtered content (to preserve social

---

[35] Obviously, the most significant concern here would be jurisdictions outside of the United States, or outside of places where a strong free speech right exists. The norms that the United States sets for the net, however, would certainly spill over into those places, however; and our view is that this spill over ought to be reckoned in any regulatory regime.

[36] In fact, voluntary access controls were the main motivation for the creation of PICS.

[37] *See* Jonathan Weinberg, *Rating the Net*, 19 Hastings Comm. & Ent. L. J. 453 n.108 (1997).

[38] *See* Rikki McGinty, *Safety Online: Will It Impede Free Speech?* Media Daily, December 5, 1997.

[39] *See* Weinberg, *supra* note 37, at n. 108.

diversity).[40] If everyone can easily satisfy their individual preference for filtering, the collective preference for social diversity may be ignored.

· Finally, if IAPs bundle filters with service, then the choice among filters might be less robust than ideal. Put another way, in practice, the competition among filters may not be sufficiently diverse. This could yield very broad filters, which if common, could create secondary impacts on the variety of speech available on the Internet — since senders may tailor their speech to what will pass the filters.[41]

These secondary impacts — a slippery slope of regulation and potentially chilling voluntary uses of labels — have led one of the authors previously to describe PICS, which provides the technical infrastructure for labeling, as "the devil."[42] The other author (one of the developers of PICS) believes that the net impact of a widespread labeling infrastructure would be positive, because of the many positive voluntary uses.[43]

CONCLUSION

This article has proposed an abstract model of mandated access controls. It includes three types of actors: senders, intermediaries and recipients. Control decisions are based on three types of information: the item, the recipient's jurisdiction, and the recipient's type.

With the architecture of today's Internet, senders are ignorant of the recipient's jurisdiction and type, recipients are ignorant of an item's type, and intermediaries are ignorant of both. It is easy to see, then, why, with today's Internet architecture, governments are having a hard time mandating access controls. Any party on whom responsibility might be placed has insufficient information to carry out that responsibility.

While the Internet's architecture is relatively entrenched, it is not absolutely immutable. Our abstract model suggests the types of changes that could enhance regulability. Senders could be given more information about recipient jurisdiction and type, either through recipients providing certificates, or through a database mapping IP addresses to jurisdictions. Recipients could be given more information about item types, either through senders providing labels or through government pre-clearance lists of permitted or prohibited items.

---

[40] *See* Cass R. Sunstein, Democracy and the Problem of Free Speech (1995).

[41] *See* Weinberg, *supra* note 37, at 477.

[42] *See* Lawrence Lessig, *Tyranny in the Infrastructure*, Wired, July, 1997, at 96.

[43] *See* Paul Resnick, *Filtering Information on the Internet*, Scientific American 62 (March 1997) and *PICS, Censorship, & Intellectual Freedom FAQ*, (Paul Resnick, ed.) (last modified January 26, 1998) <http://www.w3.org/PICS/PICS-FAQ-980126.html>.

Table 1 summarizes this sensitivity analysis. Since the two interventions are analogous, the analyses of their costs and effectiveness are analogous as well. In either case, there will be a natural incentive to provide information if the default action of the responsible party is to block access unless the information is provided (a prohibited unless permitted regime). Otherwise, there will be no natural incentive, and the government will have to require the provision of that information.

The secondary effects of these two infrastructures are also analogous, but quite different. The by-product of a certificate regime is a general ability to regulate based on jurisdiction and recipient characteristics, even for issues beyond content control, such as taxation and privacy. Such a regime also enables senders voluntarily to exclude recipients based on jurisdiction or type, a facility which might be used for negative as well as positive purposes. The by-product of a widely used labeling infrastructure is a general ability to regulate based on item characteristics, even characteristics that governments have no legitimate reason to regulate. Such a regime also enables intermediaries and recipients voluntarily to exclude some item types, a facility that may empower parents and teachers but may also be overused if it is poorly understood or difficult to configure.

If intermediaries are to be responsible for blocking, they will need both types of information. In addition, architectural changes will be necessary to enable application layer blocking of individual items rather than cruder network layer blocking of all traffic from or to an IP address. A requirement of application layer blocking, however, introduces significant costs in terms of openness to innovation and vulnerability to hardware and software failures. Intermediaries, then, are the most costly place to impose responsibility. On the other hand, they are the most easily regulated, since there are fewer of them, they are more stable, have assets and their governing jurisdictions are clear.

While our sensitivity analysis does suggest consequences that might not have been readily seen, our ultimate conclusion is one others have reached as well. It will be difficult for governments to mandate access controls for the Internet. Given today's architecture, any such mandates would of necessity be draconian or ineffective. Changes to the technical infrastructure or social practices could enhance regulability, although such changes would entail both direct costs and would create secondary by-products whose value is debatable. Given that the costs of any such architectural change would be significant, it is important for governments to answer the fundamental question of how important such changes are: perhaps a lessening of governments' traditional power to control the distribution of harmful information would be preferable.

SENSITIVITY TABLE (TABLE 1)

	Sender	**Intermediary**	**Recipient**
Missing information	· jurisdiction; · recipient type	· jurisdiction; · recipient type; · content of item	· content of item
Possible architectural and legal changes	· IP to geography mapping, jurisdiction certificates · Recipient type certificates · pre-clearance, thesauri	as for sender and recipient, plus: · responsibility to assure sender /recipient compliance · use of proxies and application gateways	· pre-clearance · sender's self-rating · third-party rating
Consequences	Enables general regulability of behavior on the net based on recipient type and jurisdiction	IAPs as the state	Enables greater control of speech content on the net beyond that initially required by governments
Notes	Enforcement problems significant, if sender outside the jurisdiction	Enforcement is easier, since ISPs are not mobile, there are few players, and they have commercial assets	Enforcement problem: number of recipients leads to selective enforcement, though a greater portion of the regulable public is within a given jurisdiction