# Service Architecture and Content Provision: The Network Provider as Editor*

J. MacKie-Mason
Univ. of Michigan
jmm@umich.edu

S. Shenker
Xerox PARC
shenker@parc.xerox.com

H. R. Varian
Univ. of Calif., Berkeley
hal@sims.berkeley.edu

December 26, 1995

### Abstract

There are at least two competing visions for the future National Information Infrastructure. One model is based on the application-blind architecture of the Internet; the other is based on the application-aware architecture of cable TV systems and online services. Among application-aware architectures, some are content-aware and some are content-blind. In this paper we examine some consequences of these different network architectures for content provision.

---

*This work was supported by National Science Foundation grant SES-93-20481.

# 1 Introduction

The formation of the National Information Infrastructure is often depicted as a battle between various economic entities such as the cable TV companies and the regional Bell operating companies. More recently, the Internet – with its enigmatic origins, anarchic organization, and startling growth – has gained widespread visibility and is now seen as a fundamental component of the NII. In addition, the various online services (AOL, Prodigy, CompuServe, MSN, etc.), direct broadcast TV, wireless telephony, and other emerging technologies will also likely be important parts of the NII.

These various delivery systems display a wide range of characteristics. The physical transmission medium runs the gamut from satellite transmission to coax cables to fiber optics. Some components are broadcast in nature (everyone in the system receives every signal) and others are switched (signals are directed to specific receivers). The corporate entities deploying these component technologies are diverse — from startups to established concerns, and from local monopolies to international competitors — and they face disparate regulatory restraints. These physical, corporate, and regulatory issues have been thoroughly discussed in the literature, and we do not address them here. We focus instead on another issue that, to date, has received considerably less attention.

The purpose of these systems is to transport information *content*. We are interested in how the differences between systems affect the offering of content to consumers. We begin with the observation that these systems have radically different *service architectures*. By architecture we are not referring to the actual physical implementation of the network. Rather, we are referring to the nature of the transport service offered. In this paper, we concentrate on one crucial feature of service architecture: *transparency*. We call the entity that transports the bits to consumers the *network provider*. In some architectures, the network provider is aware of the content of transported bits (e.g., movie, point-to-point video, music, etc.), and in others the content is completely opaque to the network provider.[1] See (Shenker, Clark, & Zhang, 1994) for a brief and speculative discussion of the implications of this distinction.

We then ask the question: how does this difference in architecture af-

---

[1] In Section 2 we clarify that there are gradations of awareness, from awareness of the applications (e.g., teleconference vs. file transfer) to awareness of the actual content (e.g., the specific movie being transmitted).

fect the content provided to consumers? Does the difference in the network provider's awareness affect the selection of existing content that is made available to consumers? These are questions that are getting increasing attention in the press concerned with the strategic developments in competition between the Internet, telephony and cable networks. For example, George Gilder writes that:

> Networks promote choice, choice enhances quality and quality favors morality. Television is culturally erosive because its small range of offerings requires a broad, lowest-common-denominator appeal. Linking to millions of cultural sources, global networks provide a cornucopia of choices, like a Library of Congress at your fingertips. On the Net, as at a giant bookstore, you always get your first choice rather than a lowest-common-denominator choice. A culture of first choices creates a bias toward excellence and virtue. ("Angst and Awe on the Internet", *Forbes ASAP*, 4 December 1995, p. 132.)

Also important is whether architecture choice affects the incentives to create *new* content. We address this question in another paper ((MacKie-Mason, Shenker, & Varian, 1995)).

In this paper we explore one way in which architecture may affect content provision: through the extent to which the network provider can play an editorial role in selecting the content made available to consumers. In an aware architecture the provider can offer an editorial service; in a blind architecture it cannot. We characterize the different architectures in Section 2. Then, in Section 3 we consider the effects of architecture on the selection of already created content to be offered on the network.

There are other important ways in which service architecture can affect content provision. In a related paper (MacKie-Mason et al., 1995) we examine two: through technological and institutional delivery costs that vary across architectures, and through the extent to which architecture permits the network to differentiate transport prices for different goods. We discuss briefly these effects — and some effects on the incentives to create new content — in Section 4.

The relationship between architecture and content provision is too complex to yield a simple and definitive answer. Our purpose here is to provide some initial intuition about the effect of architecture on content provision. To that end, we analyze this question in the context of some very simple models;

we use these to identify some of the major issues, and illustrate them through examples. In particular, throughout our discussion we restrict ourselves to the case of a single network provider that serves customers who choose to connect; our present analysis does not address the more complicated, and realistic, case of multiple coexisting and competing architectures.

# 2 Architectures

The architectural distinctions we make here concern the extent to which the network provider distinguishes between the bits it conveys. The Internet, telephony, and cable TV occupy very different places along this spectrum; they are representatives of the three basic architectural choices that we describe below.

## 2.1 Architectural Choices

### 2.1.1 Application-blind networks

One of the Internet's central design principles is that the network provides only bit transportation; it is up to the end hosts to construct higher-level applications on top of this raw transport mechanism. This architecture has the feature that it need not be modified as new applications arise, because applications are implemented entirely at the end hosts and no centralized authority need approve such applications.

We will use the term *application-blind* to refer to architectures where a general interface is made available to end users, who then implement their applications on top of this interface. For purposes of our analysis we assume that application-blind networks operate as common carriers. They offer a single, nondiscriminatory price for transport and accept any and all traffic at that price.

Network blindness has both advantages and disadvantages. The ability for end-users to develop and implement new applications without network intervention may encourage creative and experimentation. For example, the Internet has seen a proliferation of applications, such as electronic mail and the WWW, that were not envisioned when the IP protocol was originally designed. However, a blind network provider is unable to provide a gateway or editorial service which selects which applications and goods to make available to users. For example, the explosion of content on the Internet has led to an increasing amount of low-quality and irrelevant material that users must search through to find valuable content. A network cum editor can filter and certify content, much as the editor of a newspaper does.[2]

### 2.1.2 Application-aware networks

Telecommunication infrastructures developed by private enterprise such as cable TV and telephony are more tightly coupled to specific applications. The underlying transport function deep within these networks may be general but the interface presented to users is highly restricted. In this type of network the architecture is aware of the type of application that is being used.

For example, telephone networks are designed around voice traffic and cable TV protocols are designed for video traffic. Similarly, online service providers such as CompuServe and AOL generally know what kinds of applications (voice, data, images, etc.) are being transmitted.

By an *application-aware* architecture we mean that the network service provider can identify the general type of application being invoked (e.g., e-mail, audio playback, interactive video). This permits some degree of editorial service. For example, AOL permits users to access Time (r) magazine, but for its first several years did not provide access to content on the Internet.

### 2.1.3 Content-aware networks

Besides being able to determine the general applications customers use, some networks can also monitor and even control the content that is transported over some applications. For example, cable TV can distinguish basic from premium channels and video-on-demand systems know what movie has been requested. Online services also sometimes know what sort of content is being requested: airline prices, bibliographic references, cartoons, etc. We define a *content-aware* architecture as one in which the network provider can identify the network content.

---

[2] One might argue that indexing, rating, information filtering and editorial services are emerging on the Internet. But these services merely shift the problem to another level: now the network user must sift through a growing number of filter services to select a preferred one. We are implicitly assuming that the network provider has some advantage in providing an editorial service if the architecture permits it. This advantage might follow from a more credible ability to signal quality, perhaps due to the large fixed investment the network makes in infrastructure.

It is not the case that all *application*-aware networks are also *content*-aware. For example, telephony is content-blind, as is electronic mail or WWW browsing.[3] When a network is content aware, the provider can play a more discriminating editorial role.

## 2.2 Architectural Implications for Content Provision

Our fundamental point is that network architecture can have important implications for the nature of information goods made available. Anecdotal evidence is certainly consistent with this view. For instance, the application-blind Internet supports a diverse and rapidly growing set of applications such as electronic mail, file transfer, teleconferencing and WWW. The application-aware telephone system supports only a narrow range of services (basically FAX, low-speed modems and telephony), but provides access to varied content in the form of 900 numbers. Cable television, which is content-aware, offers only one application and rather limited content. Are these differences related to architectural differences? Since the network provider is the entity actually delivering the goods to consumers, the policies of the network provider will affect which goods actually are available to consumers; we investigate the role of architecture in shaping those access decisions.

To focus on the role of architecture, we simplify or eliminate most other relevant factors. For instance, we assume the network provider is a monopoly. This provider is free to maximize profits without competitive (or regulatory) pressures. We also assume that content provision is competitive, with a large number of content providers, and incurs no marginal cost (we can easily incorporate a finite marginal cost into the formalism at the expense of notational simplicity).[4] Therefore, the monopoly network provider can set prices without negotiating with the competitive upstream content providers.

Our modeling of goods and consumers is also quite simple. In most of what follows, we label separate goods by an index $i$, although in some ex-

---

[3] The line between application and content awareness is not always clearly defined. For example, an online service might know that a user is accessing Time (r) magazine, but may not be able to track which articles the user is viewing. Is this application or content awareness? In pat due to this blurriness, we do not distinguish between the two types of awareness in our present analysis.

[4] With competitive content provision and no marginal content cost, there is nothing to be gained from vertical integration between the network provider and content providers. Nor could *bundling* be used to raise profits. We intentionally sidestep these interesting and important strategy questions in this paper.

amples it is more convenient to consider a continuum of goods labeled by $x \in [0,1]$. These goods all have equivalent bandwidth requirements. We model the consumers as consuming at most one unit of each good, with a reservation value $v_i^\alpha$ for consumer $\alpha$. Initially, we assume that marginal willingness-to-pay is independent of the consumption of other goods (later we will modify this assumption slightly in Section 3.2), so a consumer's satisfaction is merely the sum of the consumed $v_i^\alpha$ less the price paid for the goods, and any other non-marginal costs imposed by the network.

In a content-blind but application-aware architecture, different goods denote different applications. In a content-aware architecture, different goods can refer to different content as well as different applications. Given this different definition of good in the content-aware and application-aware architectures, our analysis need only distinguish between aware and blind architectures. The blind architecture cannot distinguish between goods, and the aware architecture can. This key architectural distinction has several different implications for content selection from already created goods, which we now discuss.

## 3 Content Provision

How do different service architectures may affect the provision of already-created information content? Our focus in this paper is the opportunity for an aware network to provide an editorial service by controlling the content available. We explore two possible motivations for limiting content: *clutter costs* and *attention costs*. Clutter cost is an increase in the difficulty of finding or processing information that results from the total number of information goods *available* on the network. Such costs arise at the network level — they are not attributable directly to individual consumer decisions. Attention costs, on the other hand, are a reduction in the value of information goods as the number of goods purchased by a users increases. These costs can be traced to decisions made by individuals.

We believe that costs of clutter and attention begin to distinguish the economics of information goods from other more traditional fields in economics. Of course, there is nothing novel about externality costs or interdependent demands, which are the formal characterizations we use to specify clutter and attention. However, in the context of information networks, much of our attention in this paper is directed to whether there are too few or too many goods offered. This is fairly unconventional for an economic problem; more choice over available goods is routinely assumed to be unambiguously desirable. Questions of this sort do appear in the economics of advertising, but research in that area has also been somewhat unconventional, and plausibly is an early example of the economics of information.[5]

The effect of different types of cost on content provision depends on the network architecture. An aware network — with its greater control — is advantaged when there are significant *network costs*. That is, the selection of which goods to provide is more efficient. There is no such advantage with *user costs* since these can be allocated efficiently by the actions of individual users without network intervention.

The different locus of control also implies a difference in the selection order for content. An aware network controls content selection, and thus orders choices by the profits each good generates. (We make our notion of "ordering" precise below.) In a blind network, with its single transport price $p$, customers control content selection. Any good with positive demand at price $p$ will be purchased by some consumer, so goods are generally ordered by maximal willingness-to-pay.

This contrast between ordering by profitability versus maximal willingness-to-pay has interesting implications for the content diversity on various networks. Consider, for example, two very different kinds of goods: low-value *mass-market* goods, which have low maximal willingness-to-pay but high total revenue, and high-value *niche* goods, which generate relatively little revenue but have high maximal willingness-to-pay. Aware architectures will tend to favor low-value mass-market goods, while blind architectures will favor high-value niche goods. We discuss this more fully in (MacKie-Mason et al., 1995). The bias toward mass-market goods is consistent with what we observe when comparing the offerings of the Internet with cable television.[6]

We now make these observations more concrete by considering how clutter and attention costs affect the provision of goods.

---

[5] There is quite a bit of research on the optimal variety of goods when products are differentiated, but that literature is concerned with closely related substitutes. We are investigating markets with very different information goods; indeed, when we model clutter costs we assume that demands for different goods are strictly independent.

[6] This is not to imply a causal link; there are many other differences between these two systems that could explain this observation. For instance, the broadcast nature of cable television would favor mass-market goods while the switched nature of the Internet does not, although the advent of multicast is changing that somewhat.

## 3.1 Clutter Effects

When there are many information goods or applications available, the *clutter* that results can decrease the value to consumers. It becomes harder (and slower) to locate the desired content, or the interface becomes more difficult to use.

Blind networks cannot directly control the variety or quality of content available. They only set a single, uniform transport price which acts as a cutoff. Any good or application worth more to some user than the uniform cutoff price will be offered, regardless of the clutter costs imposed on other users. The concern with the resulting clutter in blind architectures is evidenced by the huge popularity of Web indexing services (`http://www.yahoo.com/` receives about 1.5 million visits per week) and the demand for them from commercial providers (for example, AOL recently purchased WebCrawler).[7] Readers of Usenet newsgroups and Internet mailing lists are also familiar with clutter costs. The current Usenet feed is about 5MB of new, mostly unmoderated text material per day.

Content-aware architectures can control the content available to users and play an editorial role; in fact, this is already standard practice in the moderated discussion groups of many online services. Online services also restrict the variety and quantity of content available.[8]

Clutter costs are *network costs* – they depend on the total number of applications or goods available, not the choices made by particular individuals. However, these costs arise not as expenses directly incurred by the network in providing a service, but in reduced user utility as more goods are offered. We model clutter effects by decreasing an individual's utility as more goods are offered (see (Shiman, 1995) for a discussion of similar congestion effects for electronic mail). For instance, if the network offers a set of

---

[7] The field of Web indexing services is itself getting quite cluttered now: there are at least a dozen competing services available. To help users choose an index service — so that they may then use the index to help them choose content — we have already seen several comparative studies of index services; see for example `http://www.zdnet.com/ pccomp/features/internet/search/index.html`. And there are several meta-search services (that search multiple search services), such as the All-in-One Search Page, CUSI, Internet Exploration Page, and SavvySearch. That is, we are already seeing editorial meta-services that recommend editorial services!

[8] Interestingly, the online services have recently engaged in a mixed strategy, providing both an editorially-controlled service and a gateway to unrestricted Internet content. We discuss this evolution below in Section 5.

goods $G$ and a consumer $\alpha$ has purchased a set $G_\alpha$ of goods, then the total valuation by that consumer is $\sum_{i \in G_\alpha} v_i^\alpha - F_c(|G|)$ where $F_c$ is some nondecreasing function. If $\sum_{i \in G_\alpha} v_i^\alpha - F_c(|G|) < 0$ we assume the consumer leaves the network. Let $\mathcal{C}$ denote the set of connected consumers, those for whom $0 \leq (\sum_{i \in G}(v_i^\alpha - p_i)_+ - F_c(|G|))$.

A network with a blind architecture sets a uniform transport price $p$. Let the revenue to a content provider for good $i$ be $R_i(p)$. Then provider $i$ will offer its good if and only if $R_i(p) > 0$. However, adding a good increases clutter costs for all network users: some marginal customers may disconnect. Thus revenue must be computed over those consumers still attached to the network, $\alpha \in \mathcal{C}$. From this we can make the following observation about content provision in a blind network with clutter costs: the selection of goods is ordered by maximal willingness-to-pay. That is, any good $j$ for which $\max_{\alpha \in \mathcal{C}}[v_j^\alpha] > p$ will be offered. Thus, *the users* select the content available, rather than delegating the selection decision to the network.

In the presence of clutter, a network provider with an aware architecture can decide which goods to make available. The problem is to choose this set $G$ and the associated $p_i$ to maximize the total revenue, which is given by the sum $\sum_{\alpha \in \mathcal{C}} \sum_i p_i \delta(v_i^\alpha \geq p_i)$.[9] This maximization problem depends on the details of the distribution of consumer preferences. The striking difference from the blind network is that the goods are ordered by maximal revenue rather than by maximal willingness-to-pay.[10] This is the natural consequence of the locus of editorial control: the aware net makes a profit-maximizing choice for the network of users as a whole; in a blind network each individual user self-selects the desired content.

The different ordering of content selection can be interpreted as a bias towards mass market goods in an aware network. Offering a new good has an opportunity cost: the revenues lost from customers who detach due to the

---

[9] The delta function is defined $\delta(z) = 1$ if $z$ is true, 0 otherwise.

[10] The revenue again must be evaluated with respect to those consumers still connected to the network, given the other goods offered. We can formalize the notion of profit ordering by using a necessary condition for offering good $j$, conditional on the set of other goods that are offered at an optimum. The incremental good $j$ is sold to some subset of connected users, yielding revenue $R_j(p_j)$. However, some users detach from the network as a result of the extra clutter. Let $\mathcal{D}_j$ denote the set of users who detach, that is, those for whom $F_c(|G_{-j}|) - F_c(|G|) - v_j^\alpha \delta(v_j^\alpha > p_j) > S_{-j}^\alpha$, where $G_{-j}$ is the set of offered goods excluding good $j$, and $S_{-j}^\alpha$ is the surplus obtained by user $\alpha$ from goods $G_{-j}$. Then, we can say that goods are ordered by revenue in the sense that all offered goods satisfy $R_j(p_j) > \sum_{\alpha \in \mathcal{D}_j} \sum_{i \in G_\alpha} p_i$, and all other goods do not.

increased clutter plus the revenue lost from lowering some prices to keep other customers attached. This opportunity cost must be overcome for a new good to be offered. Consider two new goods with identical value to all users, but which appeal to different fractions of the user base. The mass market good is purchased by $f_m$ users; the niche good by $f_n$ users; and $f_m > f_n$. Then the revenue from the mass market good with $f_m$ customers will be greater than the niche good revenue, and it will more likely exceed the opportunity cost. Indeed, even if the niche good was more highly valued by users, the aware network favors the mass market good if the difference in market size is large enough: $f_m p_m > f_n p_n$.[11]

When there are clutter costs a network with an aware architecture offers different goods than does a blind network, but does it handle clutter better? Clutter costs are an *externality*: offering a new good benefits some users but hurts all others. In a blind network, content providers will decide to enter the market as long as there is a single users with $v_j^\alpha > p$; the resulting clutter will tend to reduce total consumer surplus. This is the typical tragedy-of-the-commons phenomena present in many congestion models; see (MacKie-Mason & Varian, 1995) for an analysis of congestible resources.[12] In general, an aware network provider will not only select different goods to offer, but will also limit the number of goods offered in order to reduce total clutter costs.

This is easy to see in the case when all users who purchase a good value it the same: then the aware network provider charges a price equal to this value for each good. In order to induce the consumer to connect, the provider pays a subscription rebate equal to the user's clutter costs, $F_c(|G|)$. In this way, the network provider extracts exactly $\sum_{i \in G_\alpha} v_i^\alpha - F_c(|G|)$ from each consumer $\alpha$, or the total consumer surplus. Since the aware provider is maximizing

---

[11] Although surely not due to clutter alone, the reports that about one-quarter of users of online services like AOL detach within a year suggests the number of marginally attached customers is sufficiently large that this bias in selecting new content could be important. See (MacKie-Mason et al., 1995), op. cit., for a more complete discussion on the mass market bias in aware networks.

[12] Another aspect, which we have not explored here (in fact, it is precluded by our particular modeling of the clutter effect), is that if different goods experience different amounts of clutter (i.e., are devalued differently), then goods that are less susceptible to clutter can displace those that are more susceptible to clutter, even if their intrinsic value is less. This is akin to the effect that in a computer network congestion tolerant traffic can squeeze out congestion intolerant traffic, even though the congestion intolerant traffic may intrinsically be much more valuable.

total surplus, it selects the first-best set of goods — the clutter externality is internalized. More generally, the network will not achieve the first best. However, reductions in consumer's surplus due to clutter do reduce the profit it can extract, so the aware provider acts as an editor by limiting the number of goods offered, in order to reduce total clutter costs.

The clutter cost externality has important implications for interpreting the different content orderings we discussed above. Recall that the aware architecture will favor mass market over niche goods (appropriately defined). This is not necessarily a bad thing. A niche good, by definition, benefits a small subset of users while imposing additional clutter cost on all others. On net, total welfare may be lessened by introducing a niche good. In the Appendix we show that, indeed, too many niche goods will be available on a blind network.

We have characterized two effects that service architecture has on content provision when there are clutter costs. First, an aware network offers content based on a profit ordering; content is offered on a blind network according to maximal willingness-to-pay. Second, the clutter cost is an externality that can be at least partially ameliorated by an aware network provider that acts as an editor or gateway. We now illustrate these aspects of the clutter effect through two numerical examples. A more general treatment is presented in the Appendix.

### 3.1.1 Example 1: No Blind Equilibrium

There is a continuum of goods labeled by $x$, with $x \in [0, 1]$. Each individual consumer $\alpha$ values each of these goods an amount $v^{\alpha}$, and these values $v^{\alpha}$ are uniformly distributed between 0 and 1 throughout the population.

Let $p$ be the transport price of these goods (which are priced the same even in the aware architecture since they are essentially identical). Suppose a network offers $x$ goods; if $v^{\alpha} > p$ then user $\alpha$ connects and buys all $x$ of the goods, to receive utility

$$U_{\alpha} = \int_{1-x}^{1} (v^{\alpha} - p) ds - F_c(x)$$

We assume that the clutter effect function $F_c$ is of the form: $F_c(z) = \lambda z^2$. By solving for $U_{\alpha} = 0$ we find that all users with $v^{\alpha} > p + \lambda x$ are connected, and thus, the fraction of connected users and the total demand for each offered good is $1 - p - \lambda x$. Total profit is calculated as demand times price times the number of goods, or $\pi = (1 - p - \lambda x)px$.

An aware network can set both $p$ and $x$ to maximize profit; the maximizing values are $p = \frac{1}{3}$ and $x = \frac{1}{3\lambda}$, yielding a demand of $\frac{1}{3}$, a revenue of $\frac{1}{9}$ and a consumer surplus of $\frac{1}{6}$ per unit good. Goods are offered as long as they yield revenue at least $\frac{1}{9}$; at $x = \frac{1}{3\lambda}$, any additional goods would yield less revenue. Although the example is stylized, in that each good is essentially the same, it illustrates the result that goods are selected based on a profit condition in an aware network.

A network with a blind architecture can set $p$ but not $x$; content providers will enter the market as long as some customers are willing to buy. For any $p$, as long as the demand per good $1 - p - \lambda x$ is positive new goods will enter, ultimately driving the demand to zero. Thus, in the blind architecture no stable equilibrium can have positive demand, revenue, or consumer surplus. Given the set of offered goods, a new good will be purchased (and profit will be positive, so it will be offered) as long as at least one customer values it more than the transport price; that is, if $\max_\alpha v^\alpha > p$. Since the optimal price is $p = \frac{1-\lambda x}{2} < 1$ and $\max_\alpha v^\alpha = 1$, all goods are offered, until congestion crowds out all users. This illustrates both that goods are selected according to maximal user value in a blind architecture, and that clutter creates a congestion externality.

### 3.1.2 Example 2: An Inefficient Blind Equilibrium

Consider an example with a continuum of goods labeled by $x$, with $x \in [0, 1]$. Each individual consumer values each of these goods an amount $x$ (i.e., the $x$'th good is valued an amount $x$). The clutter effect function $F_c$ is: $F_c(z) = \frac{1}{2}(z - \frac{1}{2} + \epsilon)_+$. In the blind architecture, all goods have the same price $p$ and all goods with $x \geq p$ are offered. Since $R = p(1 - p)$, revenue is maximized at $p = \frac{1}{2}$ and the resulting revenue is $R = \frac{1}{4}$. The consumer surplus is $S = \frac{1}{2}(1 - p)^2 - \frac{1}{2}(\frac{1}{2} - p + \epsilon)_+$, which takes on the value $\frac{4\sqrt{\epsilon}}{8}$ when $p = \frac{1}{2}$. Then, for sufficiently small $\epsilon > 0$, the total welfare and the consumer surplus are increasing in the region $\frac{1}{2} < p < \frac{1}{2} + \epsilon$, even though the blind revenue is maximized at $p = \frac{1}{2}$. This last statement can be verified by noting that the price derivative of the total welfare is $D'(p)(p - F'(D(p))$ where $D(p)$ is the demand function. Thus, the total welfare increases with $p$ if and only if $F'(D(p)) > p$.

In this example the clutter externality in a blind network can be resolved with a traditional Pigovian tax to raise the price by $\epsilon$. As an alternative, an aware network fully internalizes the clutter cost and achieves the first best. Of course, consumers would prefer the blind network because they get to share in the surplus, all of which is extracted by the network provider before the architecture is aware.

### 3.1.3 Summary

In the Appendix, we present a more general model of content provision with a clutter externality. In that model, a Pigovian tax may lower welfare in a blind architecture, in contrast to Example 2, above. However, the editorial service provided by the aware network against internalizes the clutter problem. We formally show that a blind network will offer *too many* goods, and that the first-best involves eliminating some *niche* goods.

In our analysis of content selection with clutter costs we have uncovered two principles:

- **Externality**: Content value is reduced as the menu of offerings becomes more cluttered. An aware network provider can serve a beneficial editorial function, increasing the value of the net by limiting the offerings.

- **Content Ordering**: Since the aware net provider controls the offerings, content selection will be ordered by profit under an aware architecture, and by willingness-to-pay under a blind architecture.

Suppose that network providers or others develop ways to control clutter cost even in a blind architecture. Even so, an individual user has limited time and attention to devote to different information goals. We now explore how network architecture affects content provision when users experience attention costs.

## 3.2 Attention Effects

In the previous subsection we explored the fact that when many goods are purchased by a consumer, her satisfaction may be significantly less than the sum of her $v_i^\alpha$'s. We assumed that the whole may be less than the sum of the parts due to clutter: the more goods or applications available, the harder it is for the user to find what she wants. Another possibility is that users have limited attention, and their enjoyment of a given good is decreased when they are also consuming other goods. There has been considerable recent discussion of the extent to which we are moving from a service economy to an *attention* economy (see, e.g., (Lanham, 1993)). For example, a subscription to HBO has reduced value if the user also subscribes to additional movie channels. This is a special case of the goods being (imperfect) substitutes for each other. Attention depends on the goods *consumed*, whereas clutter results from the goods *offered*.

It might appear that attention effects are a modest variant on clutter effects, and not worth the bother of separate analysis. Indeed, we initially conjectured that the same would hold: we thought content selection would be ordered differently in aware and blind networks, and that the aware architecture would have an advantage because it could limit the attention cost imposed on users.

In fact, both of these conjectures turn out to be wrong. With a little thought, it may be obvious why they are wrong. However, it is worthwhile to explore how attention effects differ from clutter effects. For one thing, demand interactions of this sort surely *are* relevant for an aware net provider selecting content (though not in the way we originally guessed). In addition, the economics of service architecture is a novel topic, and it is instructive to understand how very slight differences in modeling assumptions can lead to quite different results.

### 3.2.1 No Externality

The first thing to notice is that the attention effect *does not* create an externality. Attention cost is not imposed on users — they impose it on themselves. Suppose we write individual utility as $\sum_{i \in G_\alpha} v_i^\alpha - F_a(|G_\alpha|)$ for a consumer who purchases a set $G_\alpha$ of goods, where $F_a(\cdot)$ is the nondecreasing attention cost of consuming those goods. The $F_a(\cdot)$ term represents interdependence in a single user's utility function; it is unaffected by what goods other users consume. Thus, there is nothing intrinsic about the attention affect that favors the aware architecture over the blind.

We use a simple example with homogeneous users and heterogeneous goods to illustrate the absence of an attention externality in a blind network. Consider a continuum of goods labeled $x \in [0, 1]$, where all consumers agree that $x$ is the value of the good. In a blind architecture, a user will purchase all goods that have value $x > q \geq p$. The cutoff $q$ will generally be greater than the price $p$ because for each good consumed the user pays both $p$ and an incremental attention cost, $F_a'$. Therefore, surplus is

$$U = \int_q^1 (v - p)dv - F_a(1 - q).$$

We suppose that the attention cost function is given by $F_a(z) = \frac{1}{2}(z - \frac{1}{8})_+$.

The consumer chooses to purchase all goods with value greater than $q = \frac{1}{2} + p$, where the $\frac{1}{2}$ is simply the marginal attention cost. Substituting back into the surplus function and taking the derivative with respect to $p$ we find that

$$\frac{\partial U}{\partial p} = \begin{cases} -\frac{1}{2} + p, & \text{if } p \leq \frac{3}{8} \\ -1 + p, & \text{if } \frac{3}{8} < p \leq 1 \end{cases}.$$

For all feasible values of the transport price, surplus is decreasing in price. There is no externality, and reducing the set of offered goods (by raising price) can never make consumers better off.[13]

### 3.2.2 Content selection by value or profit?

The attention effect is a function of consumed goods, not offered goods, and users choose the goods consumed. As a result, it turns out that in both architectures the goods are generally ordered by maximal willingness-to-pay rather than total revenue. Why is this result different from the ordering by revenue for the aware architecture with clutter effects? The clutter cost is a per good cost borne by the network as a whole: it lowers the total consumer surplus available for extraction. This network-wide cost from offering an incremental good must be recovered. Attention cost, on the other hand, is a per good cost borne by individual users. Since there is no network-wide cost to be recovered, the network offers goods based on individual consumer valuations, not total revenue.

As before, the result is trivial for a blind network: the net provider merely sets the uniform transport price, $p$, and all goods that some consumer values more than $p$ are offered (consumer valuation is net of the marginal attention cost induced by consuming the good). For an aware network, the result

---

[13] It would never make sense to eliminate a different set of goods, because all consumers agree on the value ordering of the goods. Thus, raising price eliminates those goods that are valued least by all consumers.

is also straightforward. The net provider should offer all goods for which $\max_\alpha[v_i^\alpha - F_a'(|G_\alpha|)] > 0$.[14] To see why this is so, suppose not for good $j$: if the network adds $j$ to its offerings, then at least one user would experience higher utility by purchasing it, and there are no external effects on other users, so no one detaches. Thus the net could charge a $p_j > 0$ and increase its profits.

It may be worth noting that ordering by willingness-to-pay is equivalent to ordering by profit in this case. The point is not that the network does not order by profit — of course it must if it is maximizing profits — but rather that with only attention costs the two orderings are identical.

## 3.3 Architecture Choice and Social Welfare

We have characterized the effects that choice of service architecture have on content provision. What are the welfare consequences of architecture choice? Given our focus on the selection of goods offered, it is natural to first compare how many goods are made available under different architectures. However, since different goods will be valued differently, we are also interested in how content selection affects the total surplus obtained: that is, the sum of profits and consumers' surplus.

Consider first a network with attention costs. The aware provider can charge a different price for each good or application; the blind provider chooses only one price. As a result, profits will be higher and more goods will be offered in an aware network. The aware network number of goods is greater because all goods are offered for which $\max_\alpha[v_i^\alpha - F_a'] > 0$, whereas in the blind net only goods for which $\max_\alpha[v_i^\alpha - F_a'] > p > 0$ are offered.

How do the number of goods provided by aware or blind monopoly network providers compare to the first-best welfare optimum? It is easy to determine the welfare-maximizing outcome when there are only attention costs. Each user orders all of the goods from highest $v_i^\alpha$ to lowest. The user then adds the goods to her consumed set in order until $v_j^\alpha < F_a'(j)$. The union of the goods desired by all users is the optimal offered set. That is, all goods for which $\max_\alpha \left[v_j^\alpha - F_a'(|G_\alpha|)\right] > 0$ should be offered. This is the set of goods offered by the aware network, so the aware provider makes the optimal selection, whereas a blind network offers too few goods.

---

[14] We abuse the notation slightly by referring to the incremental change in attention costs as the derivative $F_a'$. In fact, the argument changes by integer values, and the incremental change is a first difference.

Even though the aware architecture offers the optimal number of goods when there are attention costs, both architectures achieve less than the optimal level of social welfare. For the blind architecture, the result follows directly from offering too few goods. In an aware network, the provider chooses prices $p_i > 0$. Now consumers order the goods by $v_i^\alpha - p_i$. Thus, the order in which the consumer selects goods is distorted. Further, the consumer purchases goods only until $v_j^\alpha - p < F_a'$, so an individual consumer takes fewer goods than the social optimum ($v_j^\alpha < F_a'$). That is, the right set of goods is offered, but too few consumers buy them. This result contrasts strongly with the result for a network with clutter costs: in that case the provider used its editorial opportunity to offer *fewer* goods under an aware architecture.

Although network pricing lowers social welfare when there are attention costs, there is no market failure as there was with the clutter effect. The result is simply the usual monopoly result: the net provider restricts the output in order to raise prices above marginal cost and earn supracompetitive profits. The editorial role in an aware network is not an intrinsic advantage when there are attention costs: private consumers internalize the problem and solve it themselves.

Both architectures offer suboptimal social welfare when there are attention effects. Can we say which does better? Unfortunately, no. Though more different goods are offered in an aware net, $p_i < p$ for some $i$, $p_j > p$ for some $j$, and the net effect on social welfare is ambiguous.[15]

What happens to the number of goods offered and social surplus when there are clutter costs rather than attention costs? We showed in Section 3.1 that the blind architecture offers fewer goods than an aware network. This provides another sharp contrast between clutter and attention effects, since a blind network offers more goods in the latter case.

How does an aware network compare to the social optimum when there are clutter costs? We showed earlier that when all purchasers of good $i$ have the same valuation for that good, $v_i^\alpha = v_i^\beta$, the aware provider could achieve the first best by charging $p_i = v_i$ for the goods and then giving a subscription rebate to users to cover their clutter costs (assuming that everyone experiences the same clutter costs). Now suppose there is some variation across users in the clutter cost function. Then some users will get positive surplus from the original subscription rebate, while others will

---

[15] That is, more different goods are offered, but less is consumed of some of the goods.

want to detach. To retain some of the surplus from those users who want to detach, the network will exclude some low profit goods to reduce clutter costs. Thus, we expect to see too few goods on an aware network with clutter costs, consistent with the usual monopoly result. We cannot be sure whether the blind net offers too few or too many goods relative to the optimum because the monopoly profit incentive and the clutter externality effect work in opposite directions.

Unfortunately, the effect of architecture choice on total surplus is ambiguous when there are clutter costs, as it is for attention costs. In general, the welfare will be below the optimum with both architectures when the network provider is a monopolist. However, either the blind or the aware architecture could surpass the other, depending on consumer preferences.

## 4 Price Differentiation and Provider Costs

In another paper ((MacKie-Mason et al., 1995)) we have explored other ways in which the difference in service architecture can affect both content provision and the incentives to create new content. We summarize some of the main results here to flesh out a richer view of our ideas on the economics of service architecture.

Suppose there are no clutter or attention costs, or other consumer disutilities. Then the main feature of the difference in architecture is the fact that an aware network can charge a different price for transporting different goods or applications, while a blind network is limited to a single price. We also add one element of provider cost ignored in the present paper: there is a fixed cost per good offered that a network with an aware architecture must pay. We refer to this as a gateway or liability cost. For example, the aware architecture might have to reprogram to add a gateway to deliver a new application. Alternatively, under current and emerging U.S. law (at least), an aware network provider may be liable for certain types of content (e.g., libel or obscenity) that it transports.

When we compare the aware and blind architectures under these conditions, we find that potentially severe inefficiencies and political economy conflicts can arise. For example, if gateway costs are relatively low, the monopolist net provider will generally prefer an aware architecture (since it can imitate blind price as one option, or do something better). But consumers will often (not always) prefer the blind architecture, because they can retain more surplus. Total welfare could be better under either architecture, depending on the specifics.

For example, suppose most consumer variation in $v_i^\alpha$ is in $i$ (that is, "within" variation with most users having the same preferences but each having widely varying valuations across goods). Then the aware architecture has higher total welfare (and network profit), but consumers prefer the blind architecture. If, on the other hand, most variation is "between" — different valuations for the same good by different consumers — then both consumer surplus and total welfare can be higher under a blind architecture.

When we examine the creation of content, we find an interesting problem of commitment. The aware network can extract all of the surplus from a new information good, while the blind architecture leaves more surplus as a reward — and thus incentive — for the creator. Therefore, we might expect the blind architecture to favor creation. However, as above, we note that the aware network can mimic the pricing policies of the blind network. Therefore it might seem an aware network should be able to induce at least as much creative effort as a blind network. However, the mimicry strategy is not compelling here, because the problem is dynamic: the aware network cannot credibly commit that it will not expropriate the surplus in a future period, after the investment in creation activity is sunk. By choosing a blind architecture, then, a network may be making a credible commitment to leave incremental surplus in the hands of creators and thus it may induce a higher steady-state level of creative activity.

There is another obvious difference that is likely to affect content *creation*: in an aware network, a single firm (the network provider) effectively decides whether to invest in creating new content, while multiple firms each make independent decisions in a blind network. We expect that in a blind network there will be more experimentation with content creation, and collectively more risk-taking (not because of differences in risk aversion, but because of differences in beliefs about the likely success of projects).

# 5 Discussion

One key distinction between various competing visions for the NII is the extent to which the network provider is aware of the content of the bits it is conveying to consumers. Aware architectures are aware of the content, while blind architectures are not. Our focus, in this paper, is the impact this architectural distinction has on the provision of content.

The most striking difference between networks with aware and blind architectures is that the selection of offered goods proceeds by maximal profit in aware networks, whereas goods are ordered by the maximal willingness-to-pay among users in a blind network. We noted that in an aware network this ordering favors mass market over niche goods. These results apply, formally, when there are either clutter or attention costs. However, when the costs take the *attention* form (and if there are no other network wide costs from offering an incremental good), then the profit and willingness-to-pay orderings coincide.

Even though the selection orderings coincide when there are only attention costs, the *number of goods* offered under the different architectures do not: an aware network will offer more goods when there are attention costs. This occurs because the aware network provider can differentiate prices between goods and thus will offer some low-value goods that do not pass the uniform price threshold on a blind network.

In striking contrast, when the costs are experienced as clutter, a network provider with an aware architecture offers *fewer* goods than a blind network. This is the consequence of our third major finding on content provision: clutter costs are an *externality*, but they can be internalized if the architecture is aware. Therefore, in a blind network with clutter, too many goods are offered; the aware network provider exercises its editorial capability to reduce the number of goods offered.

We have studied a very simplified and stylized model of network service architecture and content provision. We plan to explore the effects of service architecture in richer settings. For example, we would expect more than one network provider to compete. Then it is easy to imagine that some customers who find clutter very costly will choose to subscribe to an aware network that controls clutter, at the cost of receiving mostly mass market goods. Other networks might cater to customers with niche tastes but sufficiently high values for these niche goods to overcome the resulting clutter costs.

However, once we begin to think about multiple networks with multiple architectures and different menus of available content, we have to take seriously another feature of information economies: positive network externalities. That is, the value of belonging to a network for many consumers tends to increase the more other users who are connected. For example, email is much more valuable if it can be exchanged with anyone, not just to subscribers on one of several competing networks. Therefore, we expect to see a growing demand for interoperability. Interoperability between multiple proprietary networks with multiple incompatible architectures is costly and difficult: interoperability seems to favor the blind architecture. Thus, rather than multiple competing network "clubs", we might eventually see a unified internetwork with a blind architecture, on which customer types who suffer high clutter costs pay for editorial services provided by competing sellers over the blind network. Such editorial services may not be as cost-effective or able to build reputation as editorial control provided by a network provider, but they may be the second-best compromise that results from the value of interoperability. Our casual observation suggests this vision is consistent with the recent movement by proprietary on-line services towards providing a gateway to blind Internet services, while attempting to differentiate their products through the quality of their competing editorial services.

# Reference

Lanham, R. A. (1993). *The Electronic Word: Democracy, Technology, and the Arts.* University of Chicago Press, Chicago.

MacKie-Mason, J. K., Shenker, S., & Varian, H. R. (1995). Network architecture and content provision: An economic analysis. Tech. rep., University of Michigan.

MacKie-Mason, J. K., & Varian, H. R. (1995). Pricing congestible network resources. *IEEE Journal of Selected Areas in Communications*, *13*(7). Available at URL:[16].

Shenker, S., Clark, D. D., & Zhang, L. (1994). Services or infrastructure: Why we need a network service model. In *Proceedings of Workshop on Community Networking*, pp. 145–149.

Shiman, D. (1995). When e-mail becomes junk mail: The welfare implications of the advancement of communications technology. Tech. rep., State University of New York-Oswego.

---

[16] ftp: //gopher. econ. lsa.umich.edu/pub/Papers/pricing-congestible.ps.Z

# A Clutter as an Externality

One key aspect of the clutter effect is that the offering of a good affects all consumers negatively (actually, it only affects those consumers who remain connected), but may only offer some users benefit. This negative effect can be controlled by the aware architecture, but not in the blind architecture.

Suppose there are many possible goods, and that each good appeals to an entirely separate group of like customers. That is, for each good $i$, there are $f_i$ potential customers, each of whom value the good at $v_i$, and no consumer wants more than one good. Order the index numbers for the goods so that $v_i > v_{i+1}$. Everyone bears the same clutter costs, which depend solely on the number of goods offered: $F(N)$.

In a blind architecture, the network provider can set a single transport price. Customers will participate in the network only if their surplus is positive, so we obtain an individual rationality (IR) constraint for each group of the form $v_i - F(N) - p > 0$, given $N$. Given the ordering of the goods, $N = \arg\min_i[v_i - F(N)]$, so the IR constraint can be binding for the least valued good, at most. A profit-maximizing network will thus set $p = v_N - F(N)$, conditional on its choice of $N$. That is, the price will be set to extract all of the surplus of the group with the lowest surplus. The network then chooses $N$ to maximize its profits where the profit function is $\pi(N) = [v_N - F(N)]\sum_{i=1}^{N} f_i$.

Now we can consider some welfare consequences of the clutter externality. In general, we would expect that because clutter imposes an externality, there will be *too many* goods in equilibrium. The first way to investigate this conjecture is to ask whether welfare increases if the price were increased. Consumer surplus at price $p(N-j)$ is

$$\begin{aligned} S(N-j) &= \sum_{i=1}^{N-j} f_i[v_i - F(N-j) - p(N-j)] \\ &= \sum_{i=1}^{N-j-1} f_i[v_i - v_{N-j}] \end{aligned}$$

where the simplification results from substituting in $p(N-j) = v_{N-j} - F(N-j)$. The change in consumer's surplus from raising price from the profit-maximizing level $p(N)$ to any higher price that is profit-maximizing for the smaller number of offered goods is:

$$\begin{aligned} S(N-j)-S(N) &= \sum_{i=1}^{N-j-1} f_i[v_i - v_{N-j}] - \sum_{i=1}^{N-1} f_i[v_i - v_N] \\ &= -\sum_{i=1}^{N-j-1} f_i[v_{N-j} - v_N] - \sum_{i=N-j}^{N-1} f_i[v_i - v_N] < 0. \end{aligned}$$

Thus, forcing the network provider to charge a higher price (say, by imposing a tax) discouraging some low-value goods from being offered will in fact lower consumer surplus, despite the clutter externality. Since we already know that a higher price will also lower profits, it follows immediately that a higher price will also lower total welfare.

This does not mean that the optimal set of goods is being offered, however. Suppose it were possible to pick and choose which goods would be offered, which is precisely what an aware network can do. How would total welfare be changed by eliminating a single good? Let $G_{-j}$ refer to the set of goods when the $j$th good is removed from the profit-maximizing set. The change in total welfare from removing the good is

$$\begin{aligned} W(|G_{-j}|)-W(N) &= \sum_{i\in G_{-j}} f_i[v_i - F(N-1)] - \sum_{i=1}^{N} f_i[v_i - F(N)] \\ &= -f_j[v_j - F(N)] + [F(N) - F(N-1)] \sum_{i\in G_{-j}} f_i \end{aligned}$$

The first term reflects the loss of surplus from excluding the $j$th group of customers; the summation reflects the reduction in clutter costs that offering the $j$th good imposes on all other customers. When the clutter savings are large relative to the surplus from the $j$th good, welfare would increase by excluding the good. This will tend to be the case for niche goods (small $f_j$ small relative to $\sum_{i\in\tilde{G}} f_i$), and, of course, when the marginal clutter cost $F(N)-F(N-1)$ is large.

Does an aware network necessarily deal better with the congestion problem? Yes: aware profits are maximized by extracting the full consumer's surplus with $p_j = v_j - F(N)$, so the first term in the expression is the negative revenue from good $j$, and the aware network drops goods with the lowest revenue until this foregone revenue exceeds the additional revenue (surplus) that can be extracted from other users as clutter decreases. Thus, the aware network completely solves the externality problem, and the welfare-maximizing set of goods is offered.

As we have seen, when there are clutter effects an aware network will order content selection by maximal profit, while a blind network will order by maximal willingness-to-pay, just as with the liability/gateway effect. However, with a clutter externality, too many goods will be offered in a blind architecture, particularly too many niche goods. Forcing the network provider to raise its transport price may not solve the clutter problem, and in fact may reduce both consumer surplus and total welfare. Adopting an aware architecture will internalize the clutter externality, and in the special case in which the network can extract all surplus, will even result in the socially optimal set of goods.