Peer Production of Survivable Critical Infrastructures

Yochai Benkler[*]

Abstract

The paper describes an approach to provisioning survivable critical communications and computation infrastructures by deploying radically distributed, peer-based systems for communication, computation, and data storage and retrieval. The paper focuses on system survivability as an approach to robustness, rather than on artifact-oriented hardness or impregnability. Based on a survey of experience with ad hoc wireless mesh networks, peer-to-peer storage and retrieval systems, and distributed computing, the paper identifies a class of solutions that mobilize redundant capacity resident in privately owned, extant unlicensed wireless devices, desktops, laptops, and handhelds around the periphery of the network to provide redundant, adaptive, self-healing systems. The paper explains how the disparate network-based practices we observe today can be understood as embodying a common production modality. It explains why individuals systematically invest in goods that have excess capacity relative to their needs, and why these goods might more effectively be harnessed through social relations-based sharing and exchange practices, rather than through either a price system or managerial/administrative direction. The paper concludes with some observations on how organizations or government agencies can act to implement such peer-based survivable systems using the excess capacity extant in the organization, or in society, respectively.

I.   Introduction

         Imagine a data storage and retrieval system that stores millions of discrete files, in a way that can be accessed, searched and retrieved by millions of users, who can access the system wherever they are connected to the Internet. Imagine that this system is under a multi-pronged attack. Its enemies have used a variety of techniques, ranging from shutting down the main search server under the threat of armed seizure, to inserting malicious files to corrupt the system, to capturing and threatening the operators of storage devices. Imagine that even through all these assaults, the system continues to operate, and continues to provide high quality storage, search, and retrieval functionality to millions of users worldwide. That would be a system worth studying as a model for cybersecurity, would it not?

         That system has in fact been in existence for five years, and it has indeed been under the kinds of attacks described over this entire period. It is the peer-to-peer music file sharing system. It is the epitome of a survivable system. Its primary design

---

[*] Professor of Law, Yale Law School.

characteristic is radically distributed redundant capacity. The primary economic puzzles in understanding whether it is a model that can be harnessed to design survivable systems more generally are: (1) why there is so much excess capacity for its core components—storage, processing, and communications capacity, in the hands of many widely distributed users; and (2) how one might replicate it for uses that are somewhat less controversial than sharing music files.

Peer-to-peer file sharing networks are but one, highly visible, example of a much broader phenomenon in present computation and communications systems. Shareable goods—goods that have excess capacity and are widely distributed throughout the economy in the hands of many users—are widespread in the digitally networked environment. Personal computers and handhelds; wireless transceivers; DVRs, microwave ovens, and many other devices have excess capacity of one or more of the following basic functionalities—computation, storage, and communications capacity. These goods are widely distributed and diversely located both geographically and in network topology. Their excess capacity and topological diversity can, with the right network design, be used to render their excess capacity available as redundant capacity that is highly survivable in case of attacks in either physical or network space.

The paper begins with a very brief definition of survivability and how it differs from hardening or impregnability as an approach to security. I then consider three areas of focus in which we already see sharing networks that provision the three primary functionalities of communications and computation infrastructures—computation, communications, and storage—on a radically distributed, peer-production model. All these systems share three characteristics that make them highly attractive from a survivability perspective. They all have (a) redundant capacity; that is (b) geographically and topologically diverse; and are (c) capable of self-organization and self-healing on a fully distributed model, without need for reference to any single point of organization that can, in turn, become a single point of failure. How do these networks of redundant capacity arise? Why is the excess capacity provisioned in the first place, and how is it then harnessed? The answer described in Part III is that the case studies represent specific instantiations of a general category of economic production—social sharing and exchange—that harnesses the excess capacity of a particular type of goods—shareable goods. The characteristics of shareable goods make them widely distributed in the population (at least of relatively wealthy nations), and hence available for the creation of survivable systems. The characteristics of shareable goods generally well-describe most, if not all, communication and computation devices presently deployed and planned. In other words, the universe of end user computation and communication equipment systematically includes the excess capacity necessary to produce the kinds of survivable systems that the case studies suggest. These can be harnessed in a variety of ways, ranging from managerially-mandated sharing of excess capacity in large-scale enterprises or networks for pooling redundant capacity among smaller enterprises, through secondary markets, and all the way to sharing models—like peer-to-peer filing sharing networks or ad hoc mesh wireless networks. But for reasons to do with transactions costs and the diversity of motivations of the owners of the resources, social sharing and exchange has

distinct advantages over secondary markets in pooling resources not already owned by a single enterprise.

It is the opportunity for social sharing that offers the most interesting policy implications. If indeed we live under technical market conditions that lead to many individuals in society owning large quantities of excess capacity of all three core functionalities in communications and information systems, what can policy do to facilitate their joining sharing networks that could eventually provide a valuable source of survivability to the digitally networked environment? One approach whose effects would mostly be felt in wireless carriage and storage and retrieval systems is adjusting the laws—like spectrum regulation—that have been hampering growth and adoption of sharing-capable systems. Another approach may be a *quid pro quo*, whereby the sharing networks receive some form of regulatory relief in exchange for making their capacity available for an emergency system. This might include, for example, permitting use of frequencies normally allocated to public use for unlicensed wireless devices that use ad hoc mesh architectures, in exchange for utilizing protocols that would recognize public use, on a "sirens and lights" model, and hand over their excess capacity to aid public safety in an emergency. It might take the form of a more explicit safe harbor for peer-to-peer sharing networks if their design makes them available to use incrementally for survivable data storage and retrieval systems. The last part of the paper explores these options.

The basic point of the paper is simple. We live in a technological context in which a tremendous amount of excess capacity of the basic building blocks of our information and communications infrastructure is widely deployed. The widely distributed and topologically diverse deployment of these resources make them ideally suited for building redundant, survivable backup systems for our basic computation and communications infrastructure. Harnessing this excess capacity to create such a survivable infrastructure will likely be done most effectively not through improving the ability to price these resources, but through improving the conditions for social sharing and exchange of the excess capacity users own. If we invest our policy efforts in hardening our systems to attack instead of rendering them survivable; if we ignore in our institutional design choices the effects of policy choices on social sharing and exchange, focusing solely on their effect on price-based markets and in enterprise organization, we will lose a significant opportunity to improve the survivability of our information systems at relatively low cost and with minimal bureaucratic intervention.

II. Survivable systems versus impregnable artifacts

In 69 A.D., facing a siege by the Roman legions led by Vespesian, two groups of Jews in Jerusalem took two radically different approaches towards defending Jewish life and praxis. The zealots, headed by Shimon Bar Giora and Yochanan of Gush Chalav fortified the citadel and the Temple. Rabban Yochanan Ben-Zakai, on the other hand, appealed to Vespesian to allow him, and other rabbis who wanted to join him, safe passage to Yavneh—an unfortified area of no strategic or spiritual consequence—to set up a center for rabbinic study. Vespesian agreed, hoping, no doubt, to win "hearts and

minds." The results of these choices are well known. The zealots burned down with their temple and citadel, bringing an end to ten centuries of Temple-centric Jewish life. Ben-Zakai began the rabbinic tradition that allowed Jewish belief and praxis to evolve and adapt to changing conditions in the face of repeated destruction of one major center after another—from Babylon to Alexandria, from Spain to Poland and Lithuania.

The story encapsulates two basic models of attaining security. One model is the survivable system model. The other is the hardened or impregnable artifact. System survivability as a concept for communications infrastructures has been around at least since Paul Baran's germinal 1964 memorandum, *On Distributed Communications*.[1] Its core parameters already then were understood to be distributed redundant capacity that can learn about system condition and adapt in real time to changing requirements and available resources. It presaged the core characteristics of the Internet as a survivable system. What it means for a system to be "survivable" is that it is capable of fulfilling its mission in a timely manner, in the presence of attacks, failures, or accidents.[2] The emphasis is not on repelling or preventing any attack or failure, the goal is to prevent such events from causing the system to fail to perform its mission. What it means for an asset or artifact to be impregnable, on the other hand, is that it is hard enough to break that it will not be broken or compromised for as long as it is relevant that it not be broken.

Redundancy, adaptability, and recombination are core properties of survivability defined in terms of a system. It is how cultures store knowledge. Durability, imperviousness, and retrieval are the core properties of an impregnable artifact. It is how buried time capsules preserve knowledge. Since survivability is in the system, not in any given artifact, it uses redundancy instead of durability to assure that any given functionality that the artifact could have is available from any artifact or relationship of artifacts usable for the functionality, rather than the durability of a particular functionality-producing artifact. It uses adaptability to respond to environmental changes and challenges to minimize stress on, or loss of, the system, rather than building imperviousness into the artifact to assure that it is not compromised by changes. And it uses recombination of capacities from the system, rather than retrieval of the uniquely durable and impervious artifact, to provide its functionality.

Whether a given asset requires survivability or impregnability depends on the user and the use that make the asset valuable. If the asset is valuable because of the use functionality it makes possible for its owner, then survivability is more important than impregnability. What is valuable to the user is that the asset continue to be usable by the user, even under attack, not that it be untouched by another. If the asset is valuable because it is unavailable for use by another, impregnability, rather than survivability, is genuinely necessary. Many assets have aspects of both, and the relative importance of survivability or impregnability will depend on the relative value of both—a comparison that would only be useful to engage in if one had to make tradeoffs between what survivability required and what impregnability required.

---

[1] RAND RM-3420-PR (AUGUST 1964), available http://www.rand.org/publications/RM/RM3420/.

[2] R.J. Ellison *et. al.*, Survivable Network Systems, An Emerging Approach (November 1997), CMU-SEI Technical Report CMU/SEI-97-TR-013, available http://www.cert.org/research/97tr013.pdf.

Consider a customer database in the hands of three different types of companies. Company A is a customer-oriented company. Its market position is widely understood as stemming from its responsiveness to customers, which combines excellent data on history of past purchases with a heavy emphasis on personal responsibility of each customer care agent for the entire pre- and post-purchase customer experience. Company B is a transactional data collection and sales company. Its business is built around an algorithm that, based on past purchase decisions allows it to predict within 75 microseconds of a user's browser identifying itself to the website what products, out of a lineup of products any given web-based merchant offers, are most likely to capture the attention and money of that particular customer, and then displaying those items to that user. Company C sells mailing lists to bulk mail advertisers. Clearly Company C needs strictly to exclude competitors from its customer database, because if they extracted the information, they could simply directly replicate its product. The same cannot be said about either Company A or B. Each company has another major input into the mix that gives it a market advantage. For each, if competitors got the contents of the customer information database, its operations would be little affected, and its competitiveness would also likely be little affected, unless the competitor successfully replicated either the customer-care staff or the algorithm, respectively. For each, on the other hand, denial of access or loss of, say, 30% of the data in the database would be entirely disruptive, because it would disable the leveraging of the human or algorithm advantage in 30% of cases, undermining the reputation of the firm. Companies A and B are, then, much more reliant upon the survivability of their database than upon its impregnability, while the inverse is true of Company C. Company B, in turn, is more dependant on the impregnability of its algorithm, because that, if taken by another, can lead to matched competition, whereas Company A's core advantage is outside any of its information systems exclusive features. As long as the system is up and running timely, that firm's advantage—human connection relying on the information flow from the system—is unaffected.

While the Internet as a whole functions on a distributed, redundant, self-configuring and self-healing model that results in enormous survivability, not all points or functionalities in the digital environment are similarly designed. Most processing that is directly relevant to users of any type occurs in discrete processors—personal computers, servers, etc. Most storage happens in discrete storage media—hard drives, CDs, tapes, etc. Most last/first mile physical links occur over single, non-redundant or at least low-redundancy links—such as a single cable into a home or small office, and sometimes no more than one or two major cables into a corporate campus. While a redundant survivable system approach is used to preserve the flow of packets once in the Internet cloud, then, computation, storage, and even physical links from the user into that cloud are not designed in the same way. They are, instead, usually designed on an impregnable artifact model—typified by the firewall, the password, and the locked office door. These are all artifact-oriented hardening strategies. They are not survivability strategies. Given enough time, and given interconnected, unbounded networks whose very purpose is to reach across organizational domains and entity-boundaries—individual or corporate—hardening to the point of perfect success is quite possibly unattainable. Given that survivability is the primary value to at least some substantial subset of users of

computation and communications networks, the remainder of this paper asks whether there are approaches to replicating the redundant, distributed, self-healing architecture of the Internet for functionalities other than long-distance packet forwarding—specifically, for physical link capacity in the first mile, for data storage and retrieval systems, and for raw computation.

III.   Examples of peer-produced capacity-generating systems

1.   Ad hoc mesh wireless networks

In the days following September 11, 2001, a group of volunteers delivered broadband connectivity to downtown Manhattan by setting up three WiFi hotspots.[3] The group, NYC Wireless, had been engaged in a number of experiments throughout the city of placing open WiFi gateways to allow users to connect freely to the Internet. When major infrastructure-based systems were down, connectivity was rapidly deployed, using unlicensed spectrum, not public safety spectrum, and off-the-shelf commercial equipment, not dedicated public safety equipment. The story provides a primitive motivator to look at other, more sophisticated systems that would not similarly require volunteers to come and set up connectivity. Two elements of currently developed systems would produce a similar WiFi based alternative connectivity system, but would do so automatically, without human intervention. These two elements are (a) mesh architecture and (b) ad hoc configurability.

WiFi systems as currently familiar to American users are deployed in a terminal-based station architecture, similar to a mobile phone system. We largely use WiFi devices to connect to a fixed gateway to the Internet. This is neither the only way, nor necessarily the most efficient way, to deploy an unlicensed wireless system. A number of companies, beginning at least with Rooftop Community Networks in the mid-1990s and developed most extensively in DARPA projects, have developed an approach called wireless mesh network architecture. In a mesh, every node, every device, represents both a user node and a component of the infrastructure. The purpose of mesh architecture is to create multiple possible paths between any two points in the network, such that for any given required end-to-end link, there are multiple possible paths at any given moment. As a result, if any one or a number of nodes in the system goes down because it lost power, or because its owner is a mobile unit that went out of range, or because a tree has fallen and blocked a transmission path given the propagation characteristics of the carrier frequency, nodes still active on the network can identify and compute new paths from any point, to any other point, in the network. A mesh network has no single point of failure. Redundancy of possible links replaces any single base station as a critical node of transmission, and replaces any efforts at making that single base station durable and impregnable to achieve a survivable communications system. Mesh networks can combine repeater networks—that is, each node helps the transmitting node to forward messages to their intended recipient, thereby reducing the power necessary to achieve effective communication and making the signal less subject to interference—and multi-

---

[3] Peter Myers, In A Pinch, Wi-Fi Fills Gap, New York Times (October 4, 2001).

user detection—allowing the receiver to cooperate with its neighbors' antennae better to filter out interference and correctly to decode a message from a transmitter.

An independent characteristic of these mesh networks is that many of them are also ad hoc networks. An ad hoc network is one that has no fixed infrastructure. Each node is an end user of the network, as well as a component of the network infrastructure. As a node is turned on, it searches for neighbors and announces itself to neighboring systems. It provides its own information, and in return receives information about its next neighbors and the network topology, or at least the near network topology. Nodes continue to inform each other at intervals about their presence and availability, roughly in the way that Internet routers do. Ad hoc architecture adds immense survivability to the network through its capacity to self-heal. Because each node can serve as infrastructure, and because the network is designed to be self-configuring as each node joins and leaves the network without requiring any dedicated infrastructure prior to the existence of usable nodes, an ad hoc network can deal with failures automatically and as part of the same process by which the network is formed in the normal course of its use.

In combination, ad hoc mesh networks are highly survivable in the event of an attack or other substantial disruption. Because of their mesh architecture, they are designed to create high degrees of redundant optional physical paths from any point in the network to any other point. Because of their ad hoc nature, they can adapt to radical changes in the environment, and recombine the components that survive into a communication system after an attack, using the very same characteristics that are built into them to allow them to form in the first place, when each node is first deployed.

While none of these technologies is already a major market presence, neither is any of them futuristic. The properties of repeater networks and multiuser detection in mesh networks and ad hoc architectures have become an area of burgeoning research.[4] A number of practical implementations and businesses are being developed and deployed in this space, including Mesh Networks and Packethop, both civilian commercializations of DARPA-funded projects aimed at providing the military with robust ad hoc mobile data networks, and Tropos. The first generation of ad hoc wireless mesh technologies, like Rooftop Networks (later Nokia Rooftop), preceded the explosion of WiFi, and therefore had to build the devices, the radio link protocols, and the network. This created a high entry barrier for adoption, because unless an entire neighborhood could be persuaded to join the network, buying one or two devices would provide little benefit to their users until a certain threshold adoption was in place, and a lack of economies of scale in production made the equipment unnecessarily expensive. The new generation of businesses focusing on ad hoc mesh networking can now rely on existing deployed wireless devices that are adopted by their users for mobile access that is "tethered" to an access point or cell site. What this new generation of products does is provide a software implementation that allows these already existing devices—most ubiquitously now WiFi enable devices—to form ad hoc mesh networks.

---

[4] For a brief bibliography and description of the research, see Yochai Benkler, Some Economics of Wireless Communications, 16 Harv. J. L. & Tech. 25,44-47 (2002).

Usefully for this paper, the first market into which these firms are selling is the public safety market. During the first quarter of 2004, a number of field deployments for public safety high speed mobile data networks have been undertaken. Garland, Texas and Medford, Oregon are deploying ad hoc mesh networks based on the Mesh Networks technology for use by their first responders. San Mateo California and North Miami Beach, Florida, have also begun to deploy an ad hoc wireless mesh network for first responder high speed data access, from Tropos. Both of these solutions rely on providing some fixed backbone access points, on the cellular style, instead of on a purely ad hoc infrastructure, but also enable each device to operate as a network node that is adaptable ad hoc to achieve survivability. The backbone is only necessary because of the relatively small number of nodes involved. If nodes were universally distributed throughout the relevant environment, there would be no need to seed the environment with fixed nodes. San Francisco, Marin County, and California State participants in the Golden Gate Safety Network ran a successful informal trial of PacketHop Inc.'s solution, based on a purely ad hoc mobile network with no preexisting backbone, in February 2004. The test reportedly enabled three different systems, whose traditional radio systems are not interoperable, to form an ad hoc, wireless local network on both sides of the bridge and in the water, and to bridge this ad hoc network to a command center—all enabling voice, text, and video to be exchanged among a number of fixed and mobile stations participating in the exercise.

While the technologies are reasonably well understood, their implications as a model of peer production of connectivity and survivability of networks is less emphasized. They are still packaged as either last mile solutions that competitive access providers could offer to compete with cable or DSL—this was the pitch of the first generation—or as approaches for a single entity—usually a municipality—to provision its own needs using its own network of devices and users as both the network and the users. This is so largely because mesh networks require some threshold density, derived from the power constraints regulations impose on unlicensed devices, to achieve the state where any given individual could purchase a single unit and be reasonably assured that when he or she turns that unit on, it will have neighbors with whom to form a metropolitan or wide area network. Last mile carriers or single entities can invest in strategically located fixed backbone nodes to assure connectivity under most conditions. While achieving threshold density is a genuine transition problem, as a matter of business rationale, a carrier or a fixed backbone are not necessary for the survivability of the architecture itself once a sufficient number of individual users have joined the network to create the redundancy. Indeed, a carrier is more a potential source of failure, because if configuration is dependent on a carrier, rather than attainable locally be each node and its neighbors, then carrier-operated nodes/base stations become potential points of failure.

Once mesh networking becomes a software overlay on top of WiFi devices, the path to a municipal level mesh network made of end user devices is fairly straightforward to map—even if not without some bumps. You start with the frustrations of any given household that there may be dead spots in the house, or in the yard. You add that the household has at least two WiFi-enabled devices. A utility that allows a users to place the desktop or laptop in such a location vis-à-vis the access point so that it bridges to the

dead spots, and also allows a visiting friend to check their email from their PDA in the yard becomes an obvious extension for off the shelf WiFi devices. Individual consumers, seeking to fulfill their own needs for wireless networking in the home or their business place, will provision devices that can cooperate to form a highly redundant, self-configuring and self-healing wireless network. Which leaves genuine policy questions regarding for example, how to make sure that people share their wireless capacity with each other to provide municipal coverage or how to coax people into giving public safety or mission critical data priority, at least in emergencies. These become the subject of the kind of policy interventions I outline in Part V.

For now, all that is necessary to realize is that we are already seeing the emergence of device-based wireless connectivity, that is built on top of an already-widely adopted infrastructure of devices—WiFi equipment. What is special and unusual about this connectivity infrastructure is that it requires no carrier, no fixed backbone, and no centralized source of investment. It could be built spontaneously, out of cooperative protocols implemented on top of end-user capitalized and owned equipment. If this were done, the metropolitan-level networks that could form would provide a distinct benefit to the owners of the equipment that participated in making the ad hoc mesh transport network—it would give them a last mile that they were all privileged to access, and would therefore enable tremendous competition in broadband Internet access—having circumvented the only major bottleneck to competition in that area, the last mile of physical connectivity to the Net. These networks would have highly redundant physical links, large degrees of excess wireless transport capacity by reusing spectrum and providing resources to the network whenever the device was not actively transmitting or receiving information—which in an Internet-model of communications is much of the time. This system of user-owned, cooperatively created wireless transport networks would then be available, with the right regulatory system or incentives, as a source of survivable carriage capacity in the face of emergency—irrespective whether its source were intentional attack, system failure, or some external stochastic event, like a hurricane that brings down major connection links. Understanding why it is that a user-based network of sharing can emerge, and the motivations and information characteristics of such sharing networks will provide the basis for the recommendations regarding how to optimize the adoption of such connectivity-producing cooperation, and how to coax its users-owners into making their connectivity available as a survivable backup system for mission critical communications.

In the ideal scenario, during an emergency, or in the event of a major disruption, such as an attack on a major wired switching point, public safety officials would be able to rely on the widespread availability of wireless nodes to provide them with survivable communications, at the basic connectivity layer. If every apartment and house, or every other apartment or house, had a wireless node, connected in a mesh with some other set of neighbors, achieving coverage throughout a metropolitan area, it would be practically impossible to bring down communications capacity in that metropolitan area except by annihilating a very large portion of the homes in an area. The more the device manufacturers are permitted to build their equipment with dynamic power management, so as to increase or decrease power as hop distance requires, the more survivable the

network would be everywhere, because the lower the threshold density required for connectivity. Survivability and ubiquity of connectivity are achieved, in such a system, not through providing a particularly impervious access point, but by seeding a network of devices, installed by individuals throughout a metropolitan area for their own personal needs, and available to provide backup connectivity everywhere, anywhere, in a way that is highly robust to attack because redundant, adaptive, and self-healing or capable of recombining itself through dynamic reconfiguration.

One common objection would be the concern that wireless communications are less secure than wired communications? The answer is that there are two distinct meanings to the question. The first and more common meaning is: is the data secure? Is the data capable of being read by an enemy? The answer to that is like the answer to all Internet communications. Communications are as safe as the encryption they use. No more and no less. A well encrypted bit stream will take a lot of time and money to decrypt. Its content is secure even if the whole bitstream is acquired by an enemy. Moreover, because of the packetized, multi-path nature of communications in a mesh network, acquiring the entire bitstream is not a trivial task. The second meaning of "security" in this context concerns whether the link between a and b (friends) can be disrupted by enemies x, y. The answer to that question is the same answer to the question of survivability in general. If the most that an enemy can do is cause the desired communications to lose packets, or to fail altogether, the answer is in the survivability of the system, not in the imperviousness of the signal of a single link. To conclude, then, a wireless mesh network can provide a highly survivable network, deployed and owned by many users and manufactured by device manufacturers, rather than service providers who "own spectrum". Data that it carries is no less secure than any Internet communication, wired or wireless—its security dependant on the strength of data encryption, not any supposed physical imperviousness of the physical link channel.

2.   *Distributed data storage and retrieval*

The second piece of the puzzle for provisioning a survivable computation and communications infrastructure is data storage and retrieval. Here, as I mentioned in the beginning of this article, we have an extraordinarily robust example of such a system already in operation. It utilizes the storage capacity of millions of end-user computers to store data in redundant storage. It uses their distributed processing power to search for it. And it uses their excess bandwidth to retrieve it on a global scale. I am speaking, of course, of the peer-to-peer file sharing networks that have been used so successfully as a survivable, adaptive, self-healing system of data storage and retrieval. Despite a series of attacks on major points of failure, and in the case of Napster, a successful shutdown of the main point of data search and location, the data survived in user-owned storage media, new systems of search and retrieval were created, and these in turn proved substantially more distributed, and highly resistant to attack from a persistent enemy—to wit, the recording industry backed by the force of the United States and other sympathetic governments. While one may argue as to whether the particular use to which this system has been put is desirable or not, that is not the point. The point is that Napster introduced, and Gnutella, Kazaa, Grokster, Aimster, and the many other systems that

followed it developed, a highly survivable system for data storage and retrieval with tremendous resilience, redundancy, capacity to migrate data across the network to where it is most needed, and a high resilience to geographically or topologically focused attacks, as well as to diffuse injections of malicious data into the system.

The first major peer-to-peer file sharing network, Napster, relied on decentralized storage and retrieval, with one major deviation—a centralized server to store information about which peer had what music. While peers then communicated directly with each other and downloaded music one from the other's computer from one end of the network to another, no peer could communicate directly with the other without first identifying the relevant other by reference to a central listing. This made the communications protocol simple to design and execute, it made the system highly scalable, but also vulnerable to attack and failure. The particular attack on that server was legal, and the failure a court ordered closure. But the weakness was architectural—there was a single point of attack that could bring the system down. The response of technology developers was to treat the legal attack like any other failure, and to design a system that did not offer a single target. The result was Gnutella, and a class of approaches to peer-to-peer network design that avoided creating a center that would be vulnerable to attack. Each node in the network would tell some number of its neighbors what it had on its hard drive. Each node looking for a song would query its neighbors for the song. If both information about what everyone has, and about what every one wanted, propagated in this way throughout the network, then anyone who wanted anything that anyone else had could find it. The problem with this approach was that it required a lot of overhead, and the overhead increased as more users joined. This required some tradeoff between comprehensiveness of availability of material and the information cost required to recognize what people have and what people want.

Gnutella worked reasonably well up to a point, but was not perfect. FastTrack, with its best known clients KaZaa and Morpheus, was an improvement in adding structure to the network of peers, but structure that was not sufficiently centralized, stable, or single-point-dependent to suffer the failure of Napster. FastTrack introduced Supernodes—a dynamically determined class of nodes in the network that offer the functionality of the Napster server on an ad hoc, dynamically reconfigurable basis. Based on their material characteristics—underutilized processing power and bandwidth of their connection—the software assigns to users "Supernode" status. Instead of every node telling every other node everything it had and everything it wanted, as in the Gnutella model, every node sends the list of its available files and requests to its nearest Supernode. Supernodes tell each other—keeping network traffic and information overhead low, while retaining redundancy, flexibility in the distribution of the location of information, load balancing among nodes with more or less free capacity, and flexible rerouting around points of failure—or self-healing. The result is a system that is even more widely used than Napster was at its height, yet offers no single point of failure for attack. Like Napster, KaZaa was sued. Unlike Napster, KaZaa won in the Dutch courts. But even if KaZaa had lost, the Fastrack network architecture would not have been eliminated. If it had been shut down by litigation, there are other clients that could have taken its place effortlessly, or new network architectures on the same principle that could

have been written. The physical components of the network are individually owned by users. The connections among these individually-owned nodes are software-implemented network relations running on top of the Internet protocol. There is no "network" to shut down. There are only algorithms that tell user computers how to organize themselves into a network on an ad hoc, dynamically updated basis. As long as people have files on their hard drives that can be run on any generic computer, and as long as they want to share these, there is nothing that can be done to prevent them from using these architectural principles to do so. This is why the present attacks by the recording industry are aimed at hardware regulation—forcing the computer and consumer electronics industries to build equipment that will not run generic files, but only permission-controlled files[5]—and at raising the costs of sharing to users, by suing users civilly and by pushing criminal enforcement authorities to prosecute users criminally.[6]

From Napster to KaZaa, file sharing systems relied on complete redundant copies existing around the network. Other systems, however, have begun to increase both survivability and secrecy by breaking down individual documents and storing pieces of data on different computers, encrypted, so that no user who has a piece of a file on their hard drive can compromise the document, and they allow data "retrieval" to occur as a recombination of bits of files, rather than as retrieval of whole files, from many different points in the network, accompanied by what are essentially assembly instructions for recombining a replica of the document at the point of retrieval. These approaches include at the simplest level BitTorrent, which is an approach towards accelerating downloads in music sharing peer-to-peer networks that replicates portions of songs and allows a user to download segments of a song from different peers, load balancing for bandwidth and availability across nodes.

An earlier and more complete implementation of this approach, as well as a more readily recognizable system of data storage and retrieval that is even more resilient in its characteristics, is Ian Clark's Freenet.[7] Freenet was designed for censorship-resistant publication, not for easy music downloading. It trades off easy availability of files for music usage for a series of security measures that prevent the owners of the hard drives on which the data resides from knowing what is on their hard drive or controlling what is stored on their hard drive. This step is crucial in shifting from distributed storage and retrieval systems that are intended to give the owners of the contributed hardware access to the information stored, towards a more generally applicable facility for storing data in a survivable form by harnessing widely distributed excess storage capacity on the Net. If banks, hospitals, or national security bodies are to harness distributed storage approaches, either on individually owned unaffiliated computers or even only on employee computers harnessed for this purpose, then the function of providing data storage capacity must be separated from the capacity to read the data stored from the computer providing the storage capacity. This is precisely what Freenet does: segmenting documents among

---

[5] 107 S. 2048, Consumer Broadband and Digital Television Promotion Act (March 21, 2002).

[6] See, *e.g.*, Mark Landler, Fight Against Illegal File-Sharing is Moving Overseas, New York Times March 31, 2004 (describing transnational efforts, including criminal prosecutions in Italy).

[7] See Ian Clark, Oscar Sandberg, Brandon Wiley, and Theodore Hong, Freenet: A Distributed, Anonymous Information Storage and Retrieval System (2000), available http://citeseer.ist.psu.edu/clarke00freenet.html.

drives so that no single drive has the whole of the document, and preventing any single drive from becoming a significant point of failure or compromise. In order to improve access, the system automatically migrates the topological location of data based on demand patterns, so that, for example, documents critical of a repressive government that are often accessed by residents of that country will topologically migrate towards network locations that would increase the speed and availability of document retrieval.

A similar, but perhaps even more ambitious project, though not yet a deployed system, is OceanStore. OceanStore is a concept developed by John Kubiatowicz at UC Berkeley for what he describes as a global scale, self-repairing, secure and persistent storage system. It is built precisely on the concept of ubiquitous computing, which provides ubiquitous storage in units small and large (from desktops to handhelds, and from automobiles and microwave ovens to walls and shoes in the not too distant future). The particulars of the system are fascinating.[8] The overarching principle, however, is all that we need here. Storage exists almost wherever computation exists. And computation exists almost everywhere. Connectivity is also being brought in anywhere, mostly for purposes of controlling mechanical systems—Zensys, for example, is selling ad hoc mesh network enabled WiFi radios implemented in a $5 chip in anything from air conditioning units to refrigerators. When connectivity will in fact become ubiquitous, we have a global medium of storage as long as we solve a series of questions involving security, privacy, and authentication—all of which are done with encryption irrespective of whether the physical network is diversely owned or under single entity control— resilience over time, which Kubiatowicz proposes to do by migrating replicated data from older servers to newer ones automatically, so as to avoid sensitivity to units going off line and losing data, and adaptability to failure and attack. The survivability is addressed by wide fragmentation and distribution of redundant file fragments. Kubiatowicz's Tapestry retrieval, or rather recombination, system weaves the files out of fragments distributed across many computers, none of which is readable by the owner of the storage medium on which it resides. As we saw with mesh networks, peer-to-peer file sharing systems, or FreeNet, Kubiatowicz's system relies on a peer-based model, using the high degree of redundant storage/caching capacity owned by individuals distributed throughout the world and the network, but assuming that nodes are only mostly, not always, available or connected, and that they are at best partially reliable—either in technical reliability terms or in agent trust terms. The system as a whole uses redundancy of the basic desired capacity—storage—coupled with adaptive algorithms to achieve survivability through adaptation and recombination, and to manage load balancing and usability despite the unreliability of any given node. These same everyday design components that allow the system to operate efficiently under normal operating conditions are also germane to assuring the survivability of the storage and retrieval system as a whole.

It is important to note here that the diversity of the nodes—their topological and geographic dispersion, the diversity of users and uses that are their primary tasks— immensely increase the complexity of the network they form and in the task of assuring a steady flow of functionality from a network formed from such nodes. You cannot store a unique copy of a document that is needed on average every 30 minutes on a laptop that

---

[8] See Sean Rhea *et. al*., Maintenance Free Global Data Storage, IEEE Internet Computing, (Sep. Oct 2001).

sometimes goes on a plane and is unavailable for six or twelve hours. But this same characteristic also adds to the survivability of the system as a whole. The fact that not all nodes are always connected increases the survivability of the system, because any node not connected to the network at a moment of attack is safer from the attack. It can later become available as a source of backup storage to recreate the data if all nodes connected to the network at the time of attack are affected. Having a storage and retrieval system resident on many devices, owned and used by many different users, in many locations, with more or less stable network locations and availability and with highly diverse usage patterns makes for a complex, but highly survivable, system.

3.   Distributed computing

Distributed computing is a particular implementation of the broader phenomenon of parallel computing. Parallel computing is driven by the economic fact that it is cheaper to lash together large numbers of commodity processors and divide a task among them than it is to build a single fast computer capable of running the same number of operations in a given time frame on a single processor. Imagine that a computation requires 100 operations to run in a time unit t. If the cost of a processor that runs 50 operations per second is c, the cost of a processor that runs 100 operations per second is $4c$, and the task of coordinating among more than one processor imposes a 30% loss of computation efficiency, then at a price of $3c$ a computer designer could build a processor capable of running [3*50*0.7=]105 operations a second. This would be more efficient than investing in fabricating the more expensive ($4c$) single processor system that would perform 100 operations per second. This simple description explains why parallel computing came to dominate more exotic-materials-based approaches in supercomputing over the course of the 1990s. Parallelization became over that period a major area in computer engineering—concerned with how to design problems so that they were amenable to computation in many discrete computations that can be run in parallel, without relying to heavily one on each other for inputs.

Distributed computing involves the same calculus, but instead of being built around systems that lash together thousands of processors in a single box, that are all connected very closely to minimize communication time among the processors and from them to the system memory, they are built around tens to hundreds of thousands, and in one case millions, of processors, which are donated by Internet users during their downtime.[9] Problems for such systems get parallelized on the assumption that vast computation is available, vastly more than in single proprietary supercomputers, but that no single processor is reliably available, and communications among processors and to system memory are orders of magnitude slower than those in a supercomputer. Using this approach, SETI@Home has become the fastest supercomputer in the world, performing, as of the end of 2003, calculations at a speed *seven times* greater than the

---

[9] See, *e.g.*, Michael Shirts and Vijay Pande, Screen Savers of the World Unite!, 290 Science 1903-04 (2000).

fastest supercomputer that IBM was able to build at that point in time,[10] 60% faster than the NEC Earth Simulator, formally the fastest supercomputer in the world at the same time, and four times as fast as the next fastest supercomputer.[11] SETI@Home harnesses idle processor cycles of about 4,500,000 users around the world. The users download a small screen saver. When the users are not using their computer, the screen saver starts up, downloads problems for calculation—in the case of SETI@Home, radio astronomy signals to be analyzed for regularities as part of the search for extraterrestrial intelligence—and calculates the problem it has downloaded. When a solution is calculated, the program will automatically send its results to the main site. The cycle will continue for as long as, and every time that, the computer is idle from its user's perspective so that the screensaver is activated. I have elsewhere described SETI@Home and other distributed computing projects on the Internet today in more detail.[12]

Distributed computing provides a readily available model for producing a system of survivable computation capacity, when combined with the insights one gets from the model of ad hoc mesh networks and distributed storage systems. While the focus in distributed computing today is on its cost/performance characteristics, another critical feature of distributed computing is its topological diversity. No attack or failure in a particular geographic location, or in a particular portion of the network, will disable the entire array of processors participating in a distributed computing exercise. Because it relies on parallelization that explicitly assumes nodes that are highly variable and unreliable in terms of availability, and have extremely high delays in communications among processes and from processors to memory, distributed computing can tolerate many more points of failure and radicals shifts in availability of processors due to attack. Like a mesh network, a distributed computing system need not have any single point of failure, may run redundant processes, may reassign discrete pieces of computation to available nodes when nodes otherwise tasked with running a process do not deliver, and can load balance among the available processors according to their particular conditions of availability of processing and communications capacity. In other words, such systems have the same characteristics of redundancy, adaptability, and a capacity for recombination as a basic design feature of their normal operation. These make distributed computing system highly survivable.

Distributed computing also introduces us to another potential point of interest, which is less readily available (though still possible) for connectivity, but is shared with

---

[10] IBM ASCI White, at Lawrence Livermore Laboratories, could perform 7.3 Teraflops, See Top 500 Supercomputer Sites, at http://www.top500.org/list/2003/06/, as compared to 54 Teraflops calculated by SETI@Home.

[11] SETI@Home statistics showed a speed of over 54 Teraflops per second. The more traditional supercomputers,—where a single firm or organization builds on computation cluster from processors under its ownership or control achieved 35.8 Teraflops for the NEC Earth simulator, 13.9 Teraflops for the HP ASCI-Q at Los Alamos National Laboratories, the Linux Network with 7.6 Teraflops, and the IBM ASCI White mentioned in the preceding note, both at Lawrence Livermore Laboratories. See Top 500 Supercomputer Sites, at http://www.top500.org/list/2003/06/. These numbers and relationships offer, of necessity, a snapshot. When looked at again five months later, the numbers for the mainstream supercomputers remained similar, but the number for SETI@Home had increased to over 64 Teraflops.

[12] Yochai Benkler, Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production, *forthcoming* 114 Yale L. J. __ (Nov. 2004).

distributed storage. That is, that it can be implemented not only in reliance on the participation of individual users/equipment owners, but can also be implemented within a single enterprise, or within a coalition of enterprises. Any enterprise that owns and places into operation in the hands of its employees thousands or tens of thousands of desktops, laptops, and handhelds, owned either by the firm or by its employees as work tools, or any alliance of enterprises with similar characteristics, can self-provision a substantial back-up emergency computation system solely from within enterprise or alliance resources.

IV.   <u>Some economics of shareable goods</u>

1. What are shareable goods?

The systems described in Part III share the following characteristics:

1.   They rely on widely distributed physical goods;
2.   That have excess capacity of some desideratum (transmission, storage, processing);
3.   They are privately owned by many distinct users
4.   Who pool their excess capacity without relying on either prices or managerial commands to motivate and coordinate the pooling

The wide distribution of physical capital goods capable of provisioning some level of capacity of the desideratum—be it storage, processing, or communications carriage capacity—is a direct requirement of survivability. It is the wide distribution, and the diversity of types of nodes, and their geographic and topological diversity that makes the systems they combine to form survivable. This design requirement—that the system integrate the capacity of widely distributed, diverse physical goods that generate capacity—raises an unremarkable economic question—who will invest in these goods, and why will they pool their capacity?

The second, third, and fourth characteristics shared by the systems described offer an unorthodox answer to this fairly straightforward economic question. The system is provisioned by many different users, who are not otherwise affiliated, and who coordinate their behavior without reference to a price mechanism or a managerial hierarchy. These users pool their privately owned, standard economic goods—the CPUs, hard drives, and wireless transceivers are all privately owned rival goods, they are not public goods, and at least the CPUs (for processing) and hard drives do not have network externalities/demand side economies of scale. The owners of units that participate in these practices appear to do so for a variety of motivations, ranging from some that look like altruism—in the case of distributed computing—to others that look like in-kind exchange that is clearly instrumental, but not quite barter—because it does not engage in accounting of the values given or received, and seems to tolerate high variability and lack of correlation among contributions to and calls on the system—or free riding, if you will—as in the case of file-sharing. My point in this part will be to (a) provide a set of sufficiency conditions for this kind of non-market-based (in the sense of an explicit price

mechanism), non-firm-based social provisioning of a good or service that depends on material inputs; and (b) provide a transactions costs-based and motivation-base explanation of why agents might choose to act through a social-relations transactional framework rather than through a price-based or firm-based transactional framework. I have provided the more complete statement of the argument elsewhere[13] and here will only restate the central elements of the argument.

There is a class of goods that have the following two characteristics. First, they are lumpy, or indivisible—at a given technological state, they cannot be produced in quantities that precisely match demand for them. The minimum size of a CPU available in markets today, for example, exceeds the processing needs of many home users; CPU capacities, hard drive capacities, wireless transmission capabilities increase in coarsely-grained intervals, such that the probability that any population of users would be able to find a distribution of technically available units that precisely matches their demand for the capacity of these units is low. Second, the package size or granularity of the indivisible good is relatively small—unlike an airplane or a steam engine—and their cost, relative to the degree and distribution of wealth in a society, is sufficiently low such that a substantial portion of the population could justify putting them into service to satisfy their own individual or household needs. I called these "mid-grained" goods because they are neither so coarse-grained that putting them into service necessarily requires pooling the demand of many users—through a market, a firm, or a club, for example— nor so fine-grained that an individual could purchase precisely the amount of capacity the individual requires and no more.

Where these two conditions hold, we will see a distribution in the society of capacity-generating goods that are widely owned by many users for their own personal uses, but have excess capacity relative to the needs of their owners. This describes the state of CPUs, storage media, and wireless transmission devices today. It also describes the carriage capacity of automobiles in the United States today, and underlies the phenomenon of carpooling, which accounts for roughly 17% of work related commuting trips—four times more than all forms of public transit combined.[14] It could one day, depending on the costs of home solar power electricity generators on the one hand and of fossil fuels on the other hand, describe electricity generation.

I call goods with these characteristics "shareable goods," because when they hold, individual owners of capital goods can, if they choose, share their excess capacity in social sharing or exchange networks—relying on social and psychological cues and motivations, rather than on money or commands, to inform and motivate behavior. The conditions are not necessary for sharing to occur—people share goods that have no excess capacity in acts of pure altruism where they in fact sacrifice something for the good of another, or exchange them in complex instrumental gift and social exchange systems. Conversely, people sell excess capacity of their "shareable" goods as I define

---

[13] *Id.*

[14] See John Pucher & John L. Renne, *Socioeconomics of Urban Travel: Evidence from the 2001 NHTS*, 57 TRANSPORTATION Q., at *? tbl. 3 (2003); *see also* Sharing Nicely, *supra* note 12 (surveying carpooling literature and available data).

them here. They may, for example, sell a used book in a second hand book market rather than lend it to a friend—indeed there are two online distributed computation projects to which people can sell their computer cycles. My claim is much narrower. These conditions make social relations-based clearance of the excess capacity feasible. Whether it is desirable or not—which may include the question of under what conditions it would be more efficient to share or exchange socially than to sell—is a separate consideration. Whether social sharing or exchange will or will not occur—even if it is feasible—will also depends on cultural conditions—most importantly, are these shareable goods deployed in a society with a well- or poorly-developed social transaction framework on the one hand, or market transactional framework, on the other hand.

So, the characteristics of shareable goods define conditions which, if present, are sufficient to make clearance of a set of physical capital goods through a social sharing or exchange system feasible. The question we turn to now is what might be reasons for the excess capacity of these goods to be cleared through a social transactional framework, as opposed to a market, or price-based, transactional framework.

2. Differences in information costs

Social transactional frameworks and market transactional frameworks each have substantial and different setup costs. Markets require the definition of property rights and contracting arrangements, legal enforcement systems, often physical exchange locations and information flow mechanisms, etc. Social arrangements require norms to be developed, social networks to be formed, cultural values of cooperation inculcated, etc. Assuming, however, that a society has invested in both types of transactional frameworks, individual marginal transactions within each system also have a marginal cost. We have long understood that these marginal transaction costs can lead resources to be allocated through a managerial firm-based transactional framework, rather than through a price-based transactional framework. What we add here is that a similar calculus could lead resources to be most efficiently cleared through a social sharing and exchange system rather than through the price system.

For purposes of goods that meet the focused definition of shareable goods, there are two discrete differences between the information characteristics of market, as opposed to a social, transactions, the first more important than the second. A market transaction, in order to be efficient by its own measures, must be clearly demarcated as to what it includes, so that it can be priced efficiently. That price must then be paid in equally crisply delineated currency. Even if initially a transaction may be declared to involve sale of "an amount reasonably required to produce the required output", for a price "ranging from *x* to *y*," at some point what was provided and what is owed must be crystallized and fixed for a formal exchange. The crispness, or completeness of the information regarding all aspects of the transaction, is a functional requirement of the price system, and derives from the precision and formality of the medium of exchange—currency—and the ambition to provide refined representations of the comparative value of marginal decisions through denomination in the exchange medium that represents these incremental value differences. Social exchange, on the other hand, does not require

the same degree of crispness. As Maurice Godelier put it, "the mark of the gift between close friends and relatives … is not the absence of obligations, it is the absence of 'calculation.'"[15] The point is that social exchange does not require defining: "I will lend you my car and help you move these five boxes on Monday, and in exchange you will feed my fish next July," as it would require "I will move five boxes on Tuesday for $100, six boxes for $120," etc. Instead, actions enter into a cloud of good will or membership, out of which each agent can understand him or herself as being entitled to a certain flow of dependencies or benefits in exchange for continued cooperative behavior. This may be an ongoing relationship between two people, a small group like a family or group of friends, and up to a general level of generosity among strangers that makes for a decent society.

Because the difference in cost is at the margin per-transaction, it increases linearly with the number of discrete transactions necessary to obtain a sufficient quantum of capacity to achieve the goals of a person relying on flows of capacity from the owners of the capital goods that provide the capacity. Imagine that in order to run one computation one needs the excess capacity of only one computer, and the difference in transaction costs between using market based clearance and social exchange is x. If, instead, in order to run one computation the person seeking to use excess capacity of others needs to pool the excess processing power of two idle computers, then the cost difference is $2x$ *per computation*, and so forth. This pattern indicates that when slack capacity is located in small amounts distributed among many owners, it becomes increasingly more costly to harness that excess capacity through markets than through social exchange systems. Given that the interest of each individual owner to buy as little excess capacity as technically feasible places a downward pressure on the expected amount of each unit's excess capacity, shareable goods are likely to have this characteristic—widespread distribution of excess capacity in smallish dollops. This is precisely the domain in which shareable goods become very interesting as objects of social sharing and exchange, rather than market exchange, and in which they describe well distributed computing, file sharing, and ad hoc mesh wireless networks.

The second information difference is that markets sacrifice texture for computability and comparability. In order to allow the comparison of different courses of action in a clearly computable form, market systems must abstract from the particulars of any given sought transaction to a limited set of standardized data—price, quantity, more-or-less standardized quality measures. Social systems, on the other hand, tend to use more "analog" or textured modalities of information rendering, which make decisions less clearly computable, but likely provide more richly textured information about the relevant transaction. The answer to an agent's question "should I sell my excess processor cycles at all, and if so should I sell them to a or b," is informed in different ways when the relevant data are: "a will pay me 0.2 cents per processing hour, and b will pay me 0.25 cents per processing hour," as compared to "a is dedicated to finding extraterrestrial life, b is dedicated to finding promising compounds for AIDS drugs, and I was so close to John who died in the early days of the pandemic…."

---

[15] Maurice Godelier, The Enigma of the Gift 5 (1997).

The tradeoff between formal clarity and computability, on the one hand, and texture on the other hand, suggests that social systems will be relatively weaker in organizing actions for which there are clear, computable, but fine differences between alternative courses of action. Conversely, such systems will be particularly valuable as information processing systems where the context, precise nature of the alternative possible actions, and the range of possible outcomes are persistently vague or difficult to specify formally. To the extent that information about production opportunities, cooperative actions, and motivational inputs can be represented effectively through social communications systems, it would represent a more complete statement of the factors relevant to agent's decisions than could information available in systems—like the price system—that require formalization of the data so that they can be represented adequately for the particular process of computation defined as necessary to a decision in those systems. This is particularly important where uncertainty cannot be eliminated at an acceptable cost.

3. Differences in motivation structures

Markets and social production systems appeal to different motivations. Actual behavior is diversely motivated in either type of transactional setting, and agents may act for composite motivations in either setting. But the basic motivation in market transaction is to improve material welfare. The motivations in social settings are more varied. They range from material instrumental motivations—such as when *a* helps take care of *b*'s child today, expecting to be able to call on *b* some time in the future when *a* needs help with his children—intrinsic psychological motivations—such as acting in a way that makes one understand oneself to be a self-respecting, well adjusted member of society—and social relations building—such as spending time with friends—to pure altruism like donating blood to an unknown recipient. This diversity of motivations, and the differences between market and social relations in harnessing them, would not matter if we thought that (a) all agents were identical in their motivational structure; and (b) the two types of motivation were purely cumulative. In that case, working through the market or through social relations would motivate all agents identically, and the number and mix of agents would depend purely on the magnitude of the reward, not on its form (material welfare, social-psychological). It also would not matter whether the total value of the reward was comprised purely of price, purely of social-psychological rewards, or in some mix—adding more of one or the other would always increase the level of activity.

But agents are diverse in their motivational preferences, and money and social-psychological rewards are not cumulative. The former is not a highly controversial claim; the latter more so. The *locus classicus* of this latter question is the Titmuss-Arrow debate. In his classic critique of the U.S. then-market-based blood system,[16] Richard Titmuss found that the U.K., which had an all volunteer system, had lower hepatitis-infected blood. He claimed that when money was offered for blood, donors were driven away from donating blood, leaving money-seeking blood-sellers who did not care about the unobservable (at that time there was no relevant test for hepatitis) quality of what they

---

[16] Richard M. Titmuss, The Gift Relationship: From Human Blood to Social Policy (1971).

were selling. Kenneth Arrow criticized Titumss interpretation—though not his basic finding of the differences in infection rates. Arrow's claim was that eliminating the market in blood supply could in fact remove incentives of the "bad blood" suppliers to give blood, thereby improving the overall quality of the blood supply.[17] In other words, Arrow claimed that people are in fact diversely motivated, and that some people who will be moved for money will not similarly be moved for social reasons. What he rejected, however, was the claim that by introducing a market in blood, by commodifying blood, we would crowd out the volunteer donors. He saw the two systems as independent of each others—those moved to donate would donate, those moved to sell would sell. Adding money would therefore likely increase the activity, because it would bring sellers into the cycle of blood suppliers without crowding out volunteer donors.

Even this minimal divergence from an assumption that all agents are similarly motivated suggests that a system that is seeking to harness resources from widely divergent individuals needs to know something about the population it is seeking to harness. Imagine that there are two types of agents, B, who are driven to do good in society, and C, who are driven to make money. Imagine that putting a motivation scheme in place is costly. The designer of a project that wishes to tap the resources under the control of these agents would want to choose the higher-yield strategy, and must study whether the population has more Bs or Cs: if the former, then the project organizer should invest in social-psychological motivations and vice versa. Of course, if Arrow is right in his second critique of Titmuss, and if investing in motivation were costless, then the ideal strategy would be to include both motivations.

Subsequent theoretical and empirical literature has substantially bolstered Titumss's claim that introducing markets for an activity crowds out volunteer efforts towards that same activity.[18] There are now quite extensive studies that show that adding money to an activity may decrease the level of activity from levels previously attained when the activity was carried on purely for social and psychological motivations.[19] One theoretical framework—put forward in economics mostly by Bruno Frey—for explaining these findings is based in social psychology, and focuses on the introduction of monetary interventions being understood by agents as impairing their sense of self-determination

---

[17] Kenneth Arrow *Gifts and Exchanges,* 1 Phil. & Pub. Aff. 343, 351-55 (1971).

[18] For a more complete statement, see Sharing Nicely, *supra* note 12.

[19] See Bruno S. Frey and Reto Jege, Motivation Crowding Theory: A Survey of Empirical Evidence, 15(5) J. Economic Surveys 589 (2001) (surveying the literature); See Bruno S. Frey, Not Just for Money (1997); Brnuo S. Frey, Inspiring Economics 52-72 (2001). T.F. Bewley, A Depressed Labor Market as Explained by Participants, 85 Am Econ. Rev. 250 (1995) (providing survey data about managers' beliefs about the effects of incentive contracts); Margit Osterloh and Bruno S. Frey, Motivation, Knowledge Transfer, and Organizational Form, 11 Organization Science 538 (2000) (describing effects of money-based intervention on willingness of workers to impart implicit knowledge to team mates); Bruno S. Frey and Felix Oberholzer-Gee, The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out, 87 *Am Econ. Rev*. 746 (1997) (NIMBY phenomena increase, rather than decrease, when money is added in addition to arguments about helping the common weal); H. Kunreuther and D. Easterling, Are Risk-Benefit Tradeoffs Possible in Siting Hazardous Facilities? 80 Am. Econ. Rev. 252-256 (1990) (same). Uri Gneezy and Aldo Rustichini, A Fine is a Price 29 J. Legal. Stud. 1 (2000) (finding that introducing a fine for tardy pickup increased, rather than decreased, tardiness by parents).

and self-respect.[20] An alternative causal explanation is formalized by Benabou and Tirole, who claim that the person receiving the monetary incentives infers that the person offering the compensation does not trust the offeree to do the right thing, or to do it well of their own accord, and the offeree's self-confidence and intrinsic motivation to succeed are reduced to the extent that the offeree believes that the offeror—a manager or parent, for example—is better situated to judge the offeree's abilities.[21] These causal explanations may well be an important part of the story, but they tend to underplay the instrumental importance of social interactions to their participants, which has been so central to the claims of the social capital literature.[22] For social capital to be not perfectly fungible with financial capital, as social capital is usually described, actions within a social framework must be non-fungible with actions in a market—otherwise market-based behavior could accumulate social capital seamlessly. It is fairly plausible to assume that ready availability of any given action or service from a market would debase the currency of that action as a mode of building social networks of reciprocity and social exchange. The personal sacrifice and connection between the actor and reciprocator loses its value if the agent's action is perfectly fungible with that of total strangers in the market. This instrumental, social-capital based hypothesis for crowding out would then explain the empirical literature that shows that the introduction of money decreases reciprocating cooperation.[23]

The most likely implication of the empirical findings and the theoretical interpretations is that people are indeed diversely motivated. They seek material welfare. They seek psychological coherence. They seek social relations of affirmation and companionship, as well as relations of reciprocity and instrumental advantage. The relationship between these various social-psychological motivations and the presence of money in the relationship is at least ambiguous. It may be cumulative, with one adding to the value of the other. It may, however, be negative—in particular, where the presence of money could undermine the quality of the social-psychological relationship. Paying a friend at the end of a dinner party or paying one's sexual partner are ready examples of stark cases where money would completely crowd out and social-psychological rewards

---

[20] See Frey, Not Just for Money, *supra*, relying on the line of psychology literature he follows to Edward L. Deci, Effects of Externally Mediated Rewards on Intrinsic Motivation, 18 J. Personality and Social Psychology, 105-15 (1971).

[21] Roland Bénabou, and Jean Tirole, Self-Confidence and Social Interactions. NBER Working Paper W7585 (March, 2000).

[22] See, *e.g.* Nan Lin, Social Capital: A Theory of Social Structure and Action at 150-51 (2001) (Making the claim that "there are two ultimate (or primitive) rewards for human beings in a social structure: economic standing and social standing," and elaborating a thesis where in both cases these represent relational standing, in terms of capacity to mobilize resources, some that can be mobilized by money, others those that can be mobilized by social relations); James S. Coleman, Social Capital in the Creation of Human Capital, 94 Supp. Am. J. Soc. S95, S106 (1988) (similarly is focused on the functional characteristics of social networks); Mark Granovetter, The Strength of Weak Ties, 78 Am. J. Sociology, 1360-80 (1973); Mark Granovetter, Getting a Job (1974); Yoram Ben-Porath, The F-Connection, Families, Friends and Firms and the Organization of Exchange, 6 Population and Development Rev. 1 (1980).

[23] In particular, studies that focus on crowding-out of reciprocity, are supportive of a social causal theory. See, *e.g.*, Ernst Fehr and Simon Gechter, Do Incentive Contracts Undermine Voluntary Cooperation? IERE Zurich Working Paper, No. 34, (April 2002) (describing laboratory experimental results for the effects of incentives on cooperation).

from the activity.[24] But the relationship of money to social-psychological rewards is not a fixed or uniform. At one time, being a "professional" paid athlete or performer detracted from the social respect one could earn in these fields. That has now changed drastically. In any given period, money can be culturally framed in ways that are congruent with esteem and social motivations, rather than opposed to them. For example, an academic offered a million dollars to make a scientific claim would not gain any respect in academia—though may well decide to forgo the respect for the money and write the paper nonetheless. The same academic who writes the paper in the pursuit of, and eventually receives, the Nobel Prize, gains respect rather than losing it.

If the relationship between money and social psychological motivations described here is correct, this has substantial implications for how one is to structure one's activities in attempting to harness human behavior—be it labor or the contribution of excess resources like processor cycles. If we return to the little problem of the actor trying to harness efforts or resources offered by a population of Bs and Cs, for example, then one has to assume the possibility of an even stronger loss of Bs from the introduction of money. In other words, use of a cumulative activity—offering both money and social relations—will not lead to an aggregation of Bs and Cs. It will affirmatively crowd out Bs, unless the person seeking to harness the agents finds a way of culturally framing the money so that it does not crowd out the socially-driven contributors.

It is possible, indeed likely, that there are different people with systematically different preferences on social psychological as opposed to monetary motivations. But it is equally plausible that any given individual has different preferences along these dimensions at different stages of the day, week, month, and lifetime. Once an agent has received enough money to maintain whatever material standard of living she aspires to in the relevant timeframe (today, this week, this lifetime—ready for retirement), her monetary appetite is at least partially satiated. At that point, her relative appetite for companionship or psychological affirmation may increase, and she would divert her actions towards activities that would satisfy those motivations. Anyone who looks forward to going home at the end of a workday or workweek and spending time with family and friends, rather than staying to work overtime, is experiencing such a motivation model. In designing a system to harness contributions, then, one of the considerations must be how well the platform for harnessing the effort can be calibrated to allow people to participate only when they are at moment of the day or week when their preference for money is relatively satiated, and vice versa when one believes that money is likely to be the primary motivator for agents who are likely to participate in his project.

4. Information and motivation—cumulative considerations

Anyone considering how to structure a transactional framework for harnessing resources—like CPUs, or storage media, or wireless capacity—is faced, then, with a choice among three transactional frameworks. One can choose to try to structure the transaction through one of the two more widely used models—either a market-based

---

[24] See Yochai Benkler, Coase's Penguin, or Linux and the Nature of the Firm, 112 Yale L.J. 369 (2002).

transactional framework or internally, through an enterprise. But one also could consider using a social sharing or exchange system. Such a system is likely to have lower marginal information costs, provide its users more textured information, and will have a greater capacity to harness social-psychological motivations.

In the case of shareable goods whose excess capacity is very widely distributed in relatively small incremental quanta—that is, in precisely the case that could be most valuable for constructing survivable backup systems for communications and computation infrastructures—the costs of using a market-based systems may be high, and the crowding out effect strong. This is because in order to achieve a usable robust system, contributions from a very large number of discrete individuals, in many discrete instances, will have to be harnessed. As the number of transactions necessary to achieve a working system increases, the relative per-transaction cost advantage of social sharing system will become more salient. Furthermore, if each contributor is likely to provide a relatively small incremental contribution to the overall system, the amount of money that can be paid per user or per transaction is quite small. Crowding out theory predicts that it is precisely where the negative effects of commodifying an activity are made to kick in by the introduction of payment for the activity, but the positive effects of adding money are low because the compensation offered is small, that the crowding out effect is likely to dominate and the activity level decline. This does not necessarily mean that social sharing is the only feasible approach towards harnessing the large quantities of excess capacity that users connected to the Internet have for purposes of building survivable infrastructures. But it does suggest that for those organizations or in those settings where a social transactional framework can be used, that framework will be more effective than using a market-based approach.

Part V. Some practical proposals for organization and policy

The basic point of the paper to this point has been to describe a particular class of survivable infrastructures for computation and communication—those that rely on widely distributed components that are also widely owned and shared by their owners cooperatively. I suggested that the cooperative practices that underlie these behaviors are varied and diverse, and that they range from quite directly instrumental relations, as in the case of peer-to-peer file sharing systems, to a quite generally altruistic form, like SETI@Home. In this last part I will briefly outline the approaches that organizations and individuals seeking to harness excess capacity in the traditional modalities—of markets and firms—might take, and then provide a more detailed analysis of what sort of policy interventions would be appropriate to foster the adoption of such cooperative practices for purposes of survivability more generally.

1.   *Pooling resources through firms*

The simplest avenue for adopting some of the insights of the systems described in Part III is to reorganize the use of computation and communications systems within firms. The personal computer revolution got a great lift when firms shifted from

mainframes to PC-based information systems. Their use of PCs has remained unabated. The consequence is that any large firm has employees using large numbers of computers, usually all or mostly networked for purposes of information exchange, as well as backup from the personal units to a centralized backup facility. Firms that have employees in more than one location, or that furnish or subsidize at least some of their employees' home computers/laptops, handhelds, etc., can solve the cooperation problem trivially, by fiat.

A storage system like OceanStore, a processing load-balancing utility like SETI@Home, or a campus-wide high speed data network like mesh networks could be implemented by a single firm if it has a sufficient number of employees, with a sufficient number of devices. The more diverse the locations of the users, and the more diverse the devices on which the systems are implemented, the greater the internal survivability that the firm is able to capture. A single firm implementing any such system need not face any of the motivation or coordination problems that may, or may not, create difficulties for the other alternative means of pooling these resources because of its ability to manage, rather than motivate and inform, action by independent agents.

One degree removed from within-firm cooperation can be platforms for pooling of firm resources. Here, again, instead of efforts aimed at pooling individually-owned resources, mid-sized firms can decide to form alliances or networks of pooled resources. The transaction costs would, of course, be much higher than those of a single large firm, but still the number of independent decision makers who must reach agreement is smaller, per-computer, than fully decentralized systems. One option for such a network to emerge would be through a market actor that would provide the technology and know-how to pool resources, and would cut its own costs by using the physical facilities of the businesses it organizes. The advantages that a cross-enterprise effort would have is in the relative diversity of its constituent components, which would give a successful cross-enterprise implementation greater diversity and resistance to viruses and other system failures, in addition to intentional attacks, because of the diversity of systems involved relative to the system of a single large firm.

2. Secondary markets in excess capacity

As one moves away from enterprise-owned resources towards attempting to harness the excess capacity of individually-owned resources, the standard economists' market-based response would take aim at pricing capacity. This approach, despite its higher transaction costs and the possibility of lower participation than sharing-based systems, may nonetheless be the only approach available to certain types of enterprises. For example, a motivational message that sounded roughly like: "contribute a little bit of hard drive capacity to help assure that our bank records do not get deleted in the event of a terrorist attack on our main databanks" is not likely to have the same motivational effect as, say, "contribute your unused CPU cycles to help fight AIDS." While there may be some degree of goodwill towards helping prepare for terrorism, it is still likely that helping commercial entities will not easily harness social-psychological mechanisms.

This does not mean, however, that it is impossible to achieve similar results through markets.

In the distributed computing world we have the examples for Gomez Performance Networks and Capacity Calibration Networks as enterprises that harness user capacity through a payment system. Gomez Performance Network specializes in analyzing the actual speed and availability of corporate websites for access by users in different locations and with different connection speeds. Gomez appears to use end-user installed clients not for their aggregate computational capacity but for their location in network topology and their connection speed. They harness end users to run a client by promising payment based on a rate schedule per computation, per minute on the Net, and per-user referred. But the users cannot set how often they participate or where they will go. Rather, the client, running testing tasks downloaded from the company's site, checks the speed and availability of websites when accessed from the user's desktop. Payment is capped at $45 a month. What is interesting about the site is that, while it uses the term "peer to peer" often, it constantly reinforces the message of "making money while you sleep", a message particularly clear from the client interface, which constantly shows the user's session, monthly, and lifetime balance of returns to running the program. It is difficult to measure the comparative efficacy of this network's use of money to the science-oriented volunteer distributed computation projects, however, because Gomez is not really harnessing raw distributed computation power, but rather the participants' topological diversity. It limits the number of users connected based on the information sought, not on raw processing capacity needed. Therefore, neither the processing cycles harnessed nor the number of users recruited can offer a quantitative measure of comparative efficiency. Capacity Calibration Network (CapCal), like Gomez, primarily harnesses the topological diversity and bandwidth of users, rather than their processing power. It pays users $0.30 per hour that CapCal uses the client. The user has no control over when the client is used.

A number of regularities in these sites are useful to note. First, all these payment-based systems are sensitive to transaction costs, and use PayPal—a system designed specifically to be a cheap payment mechanism for micropayments on the Internet—to pay their participants. Another characteristic shared by Gomez and CapCal is that they need topological diversity, and hence contributors are non-fungible to them. It is insufficient for the purpose of these projects simply to harness as many users as possible. They must harness only some users, and reject others, in order to create a good topological distribution of users. The specificity of the requirements, which generates the need to reject many putative contributors, likely contributes to the difficulty of attracting volunteers. Who, after all, wants to volunteer for a project that rejects many of its applicants for no apparent reason other than that they happen to be instrumentally useless? This likely adds to the fact that the motivational call to "help organizations solve their business problems" does not sound quite as inspiring as "help fight AIDS". Money then steps in as a reasonably well understood motivational source for a project ill-suited to mobilize contributors socially. Measuring the comparative efficiency of these projects to the non-market projects is difficult. On the one hand, none of these systems has grown to a size that seriously competes with the volunteer projects as mechanisms for

mobilizing the vast untapped resources connected to the Internet. On the other hand they have been in existence for a while, and do seem to generate the resources they need by paying for users' excess capacity.

A central design requirement for expanding the use of price mechanisms will be structuring the flow of funds so as to minimize the "crowding out" phenomenon. One approach to solving this problem may be to build resource-harnessing strategies into pre-existing consumer relations. For example, one could imagine banks harnessing consumers who use online banking services as components in a financial data backup system. One—simple but relatively weak— version would be to make a tamper-proof, time-stamped cached copy of the users' own account information, that gets automatically updated whenever the user logs on. Another would be to make a more explicit trade, where customers are given some discount in exchange for agreeing to offer backup facilities. This would then require an architectural approach more similar to Freenet and OpenStore, to make sure that no participating customer could tamper with any document. The general point, however, is a simple one. Organizations that have pre-existing relations with consumers, and can integrate the survivability-creating cooperation into their service, may have opportunities to harness distributed capacity through various add-on or bonus features at rates that would not have drawn into a pool of resources users without a prior relationship. Such low rates will, of course, be crucial given the amount of resources needed in relatively fine-grained quanta, can be paid to participants.

3. *Social production of survivable infrastructures*

a. instrumental exchange and non-instrumental giving

Observation of social production systems suggests that they capture a wide range of phenomena, which draw on a wide range of motivations. File-sharing systems, for example, are systems of instrumental reciprocity. Participants do not explicitly barter with one another directly (I will give you these three popular songs in exchange for that rare live recording), but rather enter a vast transactional framework where they provide their own songs to some unspecified number and identity of users, in exchange for access to the songs stored on the hard drives of other unspecified and anonymous or pseudonymous others. An application like Skype—the voice over Internet application provided by the makers of KaZaa that offers fully end-to-end encrypted telephony service from PC to PC with no dedicated infrastructure, using the FastTrack architecture— similarly seems most readily explained as an instrumental exchange system. Users who wish to have the benefits of using Skype and being available so that their friends can call them will keep their applications running. Whenever they are not themselves speaking, the computer on which they run the application ready to receive calls is also available to relay the calls of others or to provide SuperNode functionality. Although there are no test reports on the quality of Skype, anecdotally the application delivers quite impressive quality of service despite its lack of dedicated infrastructure, and its reliance purely on peer-to-peer provisioned excess bandwidth, processing, and storage.

Distributed computing systems, on the other hand, are not helpfully described as exchange relationships. The users do not have either direct or possible downstream access to resources or products they need. The contribution must be explained in terms other than reciprocal instrumentalism. Observation of the practices and interfaces of these systems suggests that they mix-and-match a variety of assumptions about the nature of the social giving that they pool. Sites describe the scientific purpose of the models and the specific scientific output, including posting articles that have used the calculations, as well as, in some cases, offering interfaces that allow—with greater or lesser degrees of legibility—contributors to see the scientific contribution their computation is making.[25] In these components, the project organizers seem to assume some degree of taste for generalized altruism and the pursuit of meaning in contributing to a common goal. They also implement a variety of mechanisms to reinforce the sense of purpose, so there are aggregate statistics about total computations performed—again, indicating a sense that users want to belong to a project that is successful and larger than themselves. Alongside these more solidaristic practices, the sites run components that are much more indicative of the mainstream anthropology of gift literature interpretation of giving as agonistic, on a "Big Man" model, that is, as a form of expressing social position. Here, we see most sites not only allowing individuals to track their own contributions, but also providing "user of the month" type rankings—displaying who contributed the most cycles and similar statistics. Nonetheless, in an informal survey that SETI@Home conducts among its own users, only a little more than 5% focused on answers that tended to suggest a search for fame or explicit recognition, like getting one's name on the top 100 list of the site. The majority of users self-describe as contributing "for the good of humanity" (58%) or simply to keep their computer productive (17%).[26] Another interesting feature of these sites, which pool capacity from a range of the most advanced economies around the globe, is the appeal to ready-made identities, encouraging users to compete not as individuals, but as teams sharing national or ethnic bonds, (*e.g.*, Overclockers Australia or Alliance Francophone), technical minority status (*e.g.*, Linux or MacAddict4Life), or organizational affiliation (University of Tennessee or of Alabama).

b. *Policy and design issues for harnessing instrumental exchange*

These observations suggest that the organizers of distributed computing systems do not themselves have a clear model of why it is that some or most of their contributors contribute. Nor do we have empirical or theoretical basis to state whether different modalities of motivating social sharing and exchange crowd each other out, so that implementing one approach (e.g., user of the month) would have a similar crowding out effect of introducing money relative to other, less individually-oriented or agonistic forms of social motivations—like "help find a cure for AIDS". The little we have to go on suggests that the most successful projects have indeed been using mixed strategies. The applicability of crowding out theory to different social motivational structures would seem to be a fruitful and important focus of future research.

---

[25] Climateprediction.net in particular provides each user with a discrete run of a simulated world with certain assumptions about CO_2 levels, which shows the user how his/her "world" is doing relative to a baseline run.

[26] http://setiathome.ssl.berkeley.edu/polls.html.

As a practical matter, ad hoc mesh networks to provide survivable last/first mile connectivity are readily obtainable on an instrumental exchange model, and can most readily be achieved or seeded by specific policy moves. Adoption of WiFi for in-home or office wireless LANs, on the one hand, and the emergence of WiFi-based protocols for making transceivers self-organize into ad hoc mesh networks that use autonomic formation and self-heal provide the building blocks of such a system. The primary remaining stumbling block to simple market-based emergence of neighborhood-level and municipal last mile redundant wireless transport networks is the timing of reaching a critical density of adoption of these devices and networking protocols in any given area. Once devices are deployed in geographic space with sufficient density, the advantages of buying a device and implementing a mesh networking protocol as a means of reaching a competitive broadband Internet Point of Presence that can circumvent the wired pipelines to the home or small office will be available to any individual making an individual decision about purchasing a device with a networking capability installed. Until that density of deployment is reached, however, the added cost, whatever it may be, of adding mesh networking capability, will do the individual no good because she cannot rely on there being a sufficient number of others to allow her to reach the neighbor she wants, the local public library on a municipal WAN, or, most importantly, the competitive broadband POP that could save her $10 a month on her broadband access fee. Because of the path of development of WiFi, this problem may eventually solve itself, because users are deploying WiFi networking devices for their own home networking needs. Nonetheless, the move to implement wider-range networking may be delayed, or even stalled, by a lack of sufficient adoption of networking capabilities.

There are two distinct types of interventions that different levels of government can make that would speed up adoption and solve the problem of threshold density. Each could work independently of the other, but they would be mutually reinforcing.

At the federal level, the FCC could orient its unlicensed devices spectrum use rules with an eye to making mesh networking simpler. One type of intervention would be to provide greater flexibility in the limitations on the design of unlicensed devices. Currently, constraints on unlicensed devices are defined in terms of a maximum peak power spectral density permitted for each device. What that means, as a practical matter, is that whether a device is deployed in Manhattan or in a ranch in Montana, it may only transmit a certain level of energy in a given bandwidth—which in turn limits the distance that a signal can travel. In Manhattan, the limitation is plausibly related to preventing interference with other nearby devices, and given the density of neighbors, may impose no functional constraint on the ability of one transceiver to reach the neighbors it needs to form a mesh. In Montana, on the other hand, the device could be cranked up to much higher power without interfering with any neighbors, and if it is not cranked up, will fail to reach the next door neighbor in the mesh. While implementing a "fix" for this problem is far from trivial, the basic idea of *in situ* measurement proposed by the FCC's own Spectrum Taskforce (SPTF) Report may provide the correct way of thinking about a more locally-adaptable regulatory requirement.[27] While the specific implementation in

---

[27] Federal Communications Commission, Spectrum Policy Taskforce Report 25-33 (Nov. 2002).

that report—the "interference temperature" model—may or may not work, the basic insight that devices that are context aware and can measure whether they are in an energy-soaked or energy-sparse environment should be permitted to adjust their transmit power accordingly is promising.

The second type of FCC intervention is more of a "carrot" than regulatory relief, and was also suggested by the SPTF Report. Here the idea is that device manufacturers get the "carrot" of being permitted to operate their devices in public safety spectrum, as well as in the general unlicensed spectrum, in exchange for designing their devices to be aware and compliant with the highly bursty pattern of public safety use. The basic model is a "lights and sirens" model, where the highway is used by all vehicles, rather than having a reserved fire lane or emergency vehicle lane, but all drivers know they must pull over when an ambulance, fire truck, or police car are flashing their lights and sounding their sirens. Device manufacturers who build devices that will recognize a siren and shift the device from normal user mode to an "infrastructure only" mode available to help in public safety communications but not using those frequencies for the users' own communications will be able to take advantage of wider and lower frequencies most of the time—improving their performance relative to competitors who do not take advantage of the same framework.

At the state and even more so the local level, government agencies can help to seed mesh networks by adopting the kinds of public safety systems described in Part III, and, importantly, making those facilities available to non-public safety users under non-emergency conditions. The idea parallels the public safety spectrum use described in the preceding paragraph, but implemented using resources under the control of local and state governments—fixed-infrastructure points on fire houses, police stations, public libraries, schools, or utility poles to provide threshold density—not by changing the operating parameters of the equipment, but by locating a larger number of off-the-shelf devices that are mesh enabled throughout their communities. This would increase the probability for any given individual that he will be able to connect to a network if he buys a device, even without coordinating with anyone else. The sirens and lights effect would then have to be implemented in the network access protocol, rather than in the end user device, but since the mesh architecture is now becoming separate from the off-the-shelf device, and since the local government will be the primary purchaser of the mesh protocol, it would be possible to specify respect for sirens in that protocol instead of in the device itself.

Implementing a parallel policy intervention for distributed storage and retrieval systems will likely be much more controversial. As the opening paragraph of this paper suggested, we already have a highly survivable storage and retrieval system in the shape of peer-to-peer file sharing networks. We have working distributed storage systems like Freenet, and working models like OceanStore, that are already available as models to achieve highly distributed, redundant, survivable storage and retrieval systems. Because of the powerful lobby of Hollywood and the recording industry, however, peer-to-peer systems are under tremendous pressure, and deploying such systems presents a substantial business risk. While some of the most recent decisions, particularly in the

Grokster case, suggest that systems with diverse uses may enjoy the benefit of the *Sony* case's privilege for manufacturers of devices/systems that have noninfringing, as well as infringing uses, the fate of Napster and Aimster must weigh heavily on any U.S. domestic firm, that is a stable and reachable defendant, and that intends to offer a distributed storage and retrieval system. As Larry Lessig put it so well in *Free Culture*, however, the theoretical availability of various defenses, like fair use, may be enough for lawyers to claim that a law is "balanced;" it does not, however actually provide a comfort zone for businesses given the realities of liability insurance and litigation risk.[28] In order to overcome this risk, one would need to amend the Digital Millennium Copyright Act to make explicit a safe harbor for distributed storage and retrieval systems that have non-infringing uses.

The primary question with respect to using approaches oriented toward instrumental social exchange in the context of storage and retrieval would be the extent to which one could structure the network so that its benefits were globally reciprocal. Consider OceanStore. Its designer, concerned with designing the technical attributes of the system, as opposed to its social design, simply assumed that the capacity will be harnessed and distributed through some price mechanism.[29] This assumption was not, however, necessary to the design. One could instead design the system such that it is available as a back up system for all of its contributors on a reciprocal basis. In that form of system, the reason I would contribute my hard drive to the system would be that I could then use it for my own data backup. If I want to use the system, I must contribute some capacity of my hard drive. The potential problem of free riding could be solved by tying the amount of resources that the system uses on one's own machine to the amount of storage that the user requires of the system, harnessing some ratio greater than 1 of the resources of the user relative to the resources the user uses.

This problem of how to resolve free riding in peer-to-peer file sharing networks arose first in a critique of the Gnutella network. Adar and Huberman showed that the distribution of files shared was highly skew—that 70% of Gnutella users did not share their own files, and that over 50% of the files shared came from the top 1% of nodes.[30] There was one interesting theoretical response, and one valuable set of practical responses to this concern. The theoretical response was Clay Shirky's.[31] Shirky focused on the nonrivalry of music files, the renewability of bandwidth, and the high costs of micropayment systems—like Mojo Nation—that might replace social sharing.   His basic point was that systems that share nonrival or perfectly renewable resources can withstand a large degree of free riding, because the free riders, while not contributing to the quality of the enterprise, take relatively little from it. The practical reality is that (a) efforts like Mojo Nation, which attempted to formalize the barter system in file sharing by giving anyone who gave resources "Mojo"–system-specific money that could be used

---

[28] Lawrence Lessig, Free Culture 97-99 (2004).

[29] Rhea *et.* al., Ocean Store *supra*, note 8, at 41.

[30] Eytan Adar and Bernard Huberman, Free Riding on Gnutella, First Monday, (October 2000) available http://www.firstmonday.org/issues/issue5_10/adar/index.html

[31] Clay Shirky, In Praise of Free-Loading, OReilly Open P2P, 12/01/2000, available http://www.openp2p.com/pub/a/p2p/2000/12/01/shirky_freeloading.html

to "purchase" system resources—failed; and (b) clients of the Gnutella and later FastTrack systems—like Limewire, BearShare, and later KaZaa, implemented much coarser mechanisms to improve the performance of a user's system, depending on the extent to which the user contributed resources for the benefit of the system as a whole.

KaZaa, for example, has created a system that motivates its users to (a) share more content; and (b) review the quality of the files they share so that they have both technical and content integrity.[32] The system assigns to each user a "participation level" that is coarsely defined as low, medium, or high. A user begins with a level of medium. A user who systematically downloads more megabytes than he or she allows others to download, or whose files are often corrupt or incomplete, will eventually be downgraded to "low." Users whose files are downloaded more often than they download other users' files, and/or who assure that the files they make available have high technical and content integrity, are upgraded to high participation level. Participation level is translated into a somewhat improved search capacity. In order not to put new users at too great a disadvantage, the system does not use participation level to affect initial search results. But the system offers "deeper" searches under a "search more" function, and the availability of the "search more" function is coarsely rationed based on the user's participation level ranking.

Without suggesting that KaZaa's solution is the only or best way of achieving ideal levels of participation in peer-based survivable storage and retrieval systems, it does offer the flavor of the design solution. If one is designing a system like OceanStore, and one is concerned about improving participation levels, one could (a) make the system available for its users to store and back up their own data—and as unique files like digital photographs that users would want to store and view from anywhere, and whose loss would be irretrievable increase, this function will become ever more valuable; and (b) tie the availability of system resources to the degree of contribution that each participant makes. The latter component should not, however, consistent with the claims about the relative transaction costs advantages of social exchange systems relative to market exchange systems, and the potential crowding out effect, seek to provide too fine an accounting system. Relatively coarse measures of participation seem to comport more with the relatively informal ways in which social sharing and exchange systems often work, and certainly have been the more successful approach in the peer-to-peer file sharing space.

c. Non-instrumental sharing and giving

The description of the motivational structure of distributed computing projects suggests that these represent the least studied and understood set of social production—that is, non-price, non-enterprise/firm based activities. They seem to be instances of cooperation that are not exchange systems, but rather operate on a model of pure redistribution with no exchange expected.[33] At most, if they do involve exchange, they

---

[32] http://www.kazaa.com/us/help/glossary/participation_ratio.htm.
[33] See, *e.g*., James Woodburn, Sharing is Not A Form of Exchange, in Property Relations (C. Hahn, ed. 1998) at 48-63.

are not a form of coarsely-accounted in-kind barter, but are rather a form of producing social relations and social hierarchy, or some other symbolic meaning. This latter form has been the dominant focus of the anthropology of gift literature.[34]

The practices of the distributed computing models outline the elements that would be available either to individuals or organizations seeking to harness excess capacity for computation, communications, and storage in ways that were not easily attainable through instrumental exchange systems. To generate the signaling effects associated with public giving one would have to assure that individuals can track their own contributions and that there would be a facility to identify and recognize high contributors. To provide the context for social connections, communications systems like discussion boards for participants could be maintained, so as to produce activity-based social interaction. In the alternative, one could attempt to harness pre-existing identity categories, such as nationality, or university or other organizational identity, by fostering extramural competition, as SETI@Home and Folding@Home do in reliance on national and other affiliation identities.

These are approaches familiar from the philanthropy world, and to a great extent this aspect of sharing systems is parallel to the phenomenon of in-kind giving. Systems like distributed computing have the advantage, however, that they are harnessing privately owned resources in ways that make it unusually "cheap" for the donor to give. First, because things are less fungible than money, the opportunity cost of giving a thing is lower than the opportunity cost of giving money. Giving up $600 in cash imposes a higher opportunity cost on the donor than donating $600-worth of an old computer, because the mechanisms for liquidating those $600 and reusing them for some other activity are more costly than they are to do so with the retained cash. Second, as long as the systems are tailored to harness genuinely excess resources—like computation, or bandwidth, or storage that is genuinely renewable or non-scarce—the "contribution" given by the individual is costless to the individual. The only "costs" are (a) the forgone market opportunity; and (b) the transaction costs of participation in the social giving system. Because of the relative advantage of social production systems in terms of marginal transaction costs, secondary markets may not be a feasible alternative for widely distributed excess capacity. Because the resources at stake—electrons and photons in one form or another—flow easily over a network in ways that need not be transparent to the contributor, the decision and implementation costs are lower than, say, donating a physical computer to a school. A system that reduces participation costs sufficiently should be able to motivate participation as long as it provides any meaningful social-psychological rewards.

---

[34] The anthropological literature on sharing and the gift has been vast, starting with Bronislaw Malinowski's Argonauts of the Western Pacific (1922) and Marcel Mauss, The gift; forms and functions of exchange in archaic societies (1925) (trans. 1954). A combination of a broad intellectual history and a major contemporary contribution to this line is Maurice Godelier, The Enigma of the Gift (1997); see also James G. Carrier, Property and Social Relations in Melanesian Anthropology, in Property Relations, Renewing the Anthropological Tradition, C. M. Hann, *ed.* (1998) at 85-103 (providing brief intellectual history of the literature); C.M. Hann, Introduction, in Hann, *supra*, note 2, at 23-34 (same).

VI.   Conclusion

In an open network whose value inheres in global connectivity, achieving security through impregnability is difficult. For large classes of system requirements, survivability is more important than impregnability. In particular, where the value of a system—what makes it a critical system—is its availability for reliable use by its intended users in the face of attack, failure, or accident, rather than its denial to an enemy or competitor, survivability, not impregnability, is the primary security desideratum of a system.

Since the early 1960s, and consistent with the design of the Internet itself, survivability has been understood to require distributed resources and redundant capacity, organized so that the system can learn of failures and self-heal by reconfiguring its redundant resources to provide the functionalities threatened by the attack or failure. The greater the degree of distribution or diversity in the loci and types of sources of capacity, the greater the resilience of the system to an attack on any particular resource or class of resources.

In the past half decade to a decade a series of developments in distributed computing, ad hoc mesh networking, and peer-to-peer file sharing systems have outlined an approach towards replicating the survivability characteristics of the Internet as a data transmission network with regard to the three central attributes of any computation and communications system: basic physical-layer local connectivity (through wireless); data storage and retrieval (p2p); and processing (distributed computing). These systems all rely on pooling of individually owned private computation and communications resources—PCs and laptops—into networks that share the excess capacity—processing, storage, and bandwidth—productively. They provide us with a template for how we could deploy similar strategies to develop peer-based survivable infrastructures for the core requirements of a computation and communications-dependant economy and society.

These systems are instances of the general category of social production, or sharing and exchange of shareable goods. These are goods whose technical characteristics make it highly likely that there will be a relatively large amount of excess capacity distributed in a society, distributed in relatively small dollops owned by individuals. Under these conditions, the transaction costs associated with harnessing the excess capacity of these resources through a secondary market are likely high, and the crowding out effect of trying to do so through markets, are likely to decrease, rather than increase, total activity, because they will likely drive off more social contributors than they will lure in market vendors of excess capacity. Resources distributed in this particular pattern may instead be more amenable to being harnessed through social production systems, rather than through price-based markets and firms. The paper concludes by offering a variety of specific policy and organizational actions that could be used to improve the degree to which we harness these excess resources towards providing our society with redundant, survivable backup communication and computation systems.

While my focus here has been on computation and communications, nothing in the analysis limits my observations to these resources alone. Any type of capital good or source of capacity that is sufficiently fine-grained to be widely owned in society, yet sufficiently coarse-grained or lumpy so that it usually has more capacity than its owner can use is a candidate for mobilization through such social production systems. Learning how to build sharing systems, rather than focusing purely on ways to perfect price-based markets in capacity or other types of goods, and focusing instead on facilitating the conditions of social sharing and exchange, could yield greater improvements in productivity, and in the case of goods that can be harnessed towards providing survivable critical infrastructures, security.