

# **Making the world (of communications) a different place**

Report of a working session of the End-to-End Research Group

January, 2005

Version 4 3/24/05

This version for preliminary release

*David D. Clark, Craig Partridge, Robert T. Braden (chair), Bruce Davie  
Sally Floyd, Van Jacobson, Dina Katabi, Greg Minshall,  
K.K. Ramakrishnan, Timothy Roscoe, Ion Stoica,  
John Wroclawski and Lixia Zhang*

This report is the product of a discussion held at the January 2005 meeting of the End-to-End Research Group, which is part of the Internet Research Task Force. The challenge presented to the group for this discussion was the following:

*How might the computing and communications world be materially different in 10 to 15 years, and how might we define a research agenda that would get us to that world?*

There were a number of motivations for this discussion. The Internet itself arose because of a visionary answer to a question such as this one. Through an alignment of visionary leaders, the research community, and funding agencies, there was a coherent, long-term effort to build a running prototype of a major new communications system. That effort led to a number of new research results; results that substantially expanded and changed our understanding of the communications field.

The networking field does not have a shared vision of the future today. Perhaps as a result, much of the research we see today lacks a motivation to deepen or broaden our understanding of communications. Much of today's research is felt to be incremental (in the sense of "least publishable increment") and lacking a long-term motivation.

At the same time, the United States' National Science Foundation is interested in hearing about important focus areas that they might fund. While focus areas are some steps short of a shared vision, we thought that a discussion of visions of the future would help refine what the focus areas might be, and could even be a vehicle to bring the research community to a common objective.

In this context, the participants at the meeting speculated about possible visions of the future, and whether the time was right for a focused research push to move us toward that future. The next several sections talk about some of the visions. The report concludes with some thoughts about directions we might take.

## **1 First, What is a "Vision"**

Before talking about the different visions of the future, we would like to step back and talk a little about what constitutes a vision.

The older members of the data communications research community spent some of their formative years in the time when data communications was being revolutionized by the creation of a new paradigm: packet switching. While packet switching is now an accepted, indeed, lauded way to think about data communications, into the early 1980s it was still a radical idea and into the 1990s required periodic justification. The point of this observation is that the first “vision” that older researchers experienced was an all-transforming idea, that affected almost every element of data communications.

There are two aspects to an idea such as packet switching. First, there is the conception of how the future could be different and better. This aspect captures what is important about the idea. It is our story to the world, to excite people about our goals for the future. The other aspect is the technical approach to getting there. Both aspects are necessary for a vision to be a useful contribution to progress, but it is important to note that the technical approach may take on a very different form for different visions. It is rare to find a single technical innovation that has had the sweeping implications of packet switching. Rather, in mature fields, it may require a *suite* of technical innovations to accomplish an important objective. Sometimes, it requires that a number of ideas that already exist be put together in some novel and useful way.

In this document, we’re concerned entirely with the first of these aspects. The topics below are not characterized by their technical novelty and appeal. They are described in terms of the outcome—how the world might be different. We feel, for all of these topics, that the goal could be achieved—that there are valid technical approaches to solving these problems. But that was not elaborated in the discussion.

In fact, a unified set of technical research objectives might set a direction for the field that could accomplish several of these visions. NSF is contemplating a focus area to look at a new architecture for an Internet of the future. That work, carried through with the right breadth and mission, could move us toward a number of these objectives. It is important, if the field is to have a sense of where it is going and what its most important challenges are for the next several years, to have a thoughtful assessment both of approaches and objectives, and how these relate.

We feel that these objectives have the benefit that they open up new avenues for further research, and have great benefit in the larger context.

## **2 Technology at the edge—an access infrastructure for next generation devices.**

There is a natural tendency, when thinking about a new network architecture, to think about what sort of devices will make up the core or center of the network. For instance, the early 1990s saw the network community vigorously debating what technologies should sit in the core of the network, with concurrent work on SONET, Asynchronous Transfer Mode (ATM), and IP devices and much jostling in the research world, the standards community and the marketplace.

However, for the Internet, much of the creative energy is at or near the edge of the network. It is at the edge that most applications are created. It is at the edge that most devices are connected. It is at the edge where we usually see the development of new networking technologies. It is at the edge of the network where the economic conditions most favor innovation, as the barrier to entry (for applications, devices, and networking technologies) is typically lower at the edge. And, at a fundamental level, the purpose of an Internet is to hook computers and similar “things” together, and we connect new “things” at the edge.

So if we want to think about where networking might be in 10 or 15 years, it behooves us to look at (r)evolution at the edge.

The current Internet was conceived in the era of main-frame computers and has matured in the era of workstations. The most common wireline edge technology is Ethernet, and the wireless technologies, such as WiFi, mimic Ethernet. The evolution of Ethernet is marked by improvements in speed at constant cost. As computers get faster, so does Ethernet.

In 10 years, we expect the most common devices on the network will be embedded processors, such as in sensors and actuators. An ambitious goal would be hundreds of billions of such devices capable of communication. In most cases, the market for these devices will be driven by dropping costs rather than increasing speeds. There is only so much communication required to perform certain types of measurements – and in many cases, improving battery life will be more important than improving performance.

Many of these devices will (inter)connect by radio, whether they communicate only locally or, at least intermittently, with the global network.

Concurrent with this evolution in devices may come a revolution in the use of wireless spectrum. Energy-aware radios capable of changing their transmission schemes to reflect the current state of local spectrum availability have the potential to dramatically improve spectrum utilization (which is currently estimated to be less than 10%<sup>1</sup>). We could see orders of magnitude improvement in wireless bandwidth as a result. Optimists dare to think of a world in which spectrum is plentiful rather than scarce.

The research community should conceive and design the communications infrastructure for these embedded devices at the edge :

*In 10 years, there should be a ubiquitous, low cost, open infrastructure suited for communication with low-cost computing devices such as sensors and controllers.*

---

<sup>1</sup> The typical measure is, standing in one spot, how many frequencies exhibit active transmissions per some unit of time. In fact, this measure underestimates available frequency as techniques such as underlaying may enable reuse of even some active frequencies.

There are many issues in the design of such an infrastructure. Just like the Internet, parts of the system will be purchased and operated by corporations, building owners and homeowners, and parts will be run as commercial services. How will devices move seamlessly between these various regions? What should the open interfaces be for this infrastructure? What are the issues around security, robustness, and ease of use? How do we dynamically program and manage the spectrum? How can we achieve sufficient ease of use in a context where there are not professional “network specialists” to manage these devices?

There has been research on pieces of this problem. It seems that the time is right to build on this research, to speculate on how far it might take us. From that perspective we can ask what a coherent infrastructure for broader class of embedded devices might be, and then build that infrastructure.

### **3 A universal system for location**

The Global Positioning System, or GPS, has revolutionized the practice of war; navigation for car, boat, plane and walking; emergency services; even accurate timekeeping. GPS technology is now small enough and cheap enough that it can be embedded in wristwatches and cell phones. However, GPS only works when the receiver is “in the open”, and able to receive signals from satellites. Imagine a world in which any properly equipped device can “know where it is” anywhere—outside or inside, in buildings, tunnels, and other shielded places. Imagine a world in which the precision of the location is variable—inside a building it might provide location to a room level, and inside a room it might provide even more precise location.

There are a huge range of opportunities and risks in this idea. In case of an emergency, responders can now be directed to exactly the right location. Missing objects can be tracked with precision. There is no longer any concept of “being lost”. Pets, children, or Alzheimer’s patients could be monitored and located if concerns arise.

Location information can be used to deliver an experience (access to information about restaurants, for example) tailored to where you are.

On the other hand, this system will raise serious concerns about privacy, surveillance and freedom of action. While providing location information can clearly be a one-way system (ala GPS) where the location providing tools do not track who is receiving, once the information is received by your phone, PDA, or other device, your location is potentially available to others. Who will have access to the knowledge of where anything is? The design of this system will require multi-disciplinary participation by technologists, social scientists, and societal observers.

Of course, it is not just people who can determine their location. Any properly equipped sensor can now know where it is, which makes installation and configuration easier and more automatic. One can now make queries about “the nearest  $x$ ” and find the object that

is closest physically, not in cyberspace. Devices that are in the same room can know they are in the same room, and organize accordingly.

*In 10 years, there should be a ubiquitous, open infrastructure that allows any properly equipped device to determine its location, both inside and outside, to a suitable level of precision.*

As this discussion indicates, location alone is an exciting service with the potential to transform certain applications. It is also likely a component (and, importantly, an enabling component) of the larger vision of the network edge offered in the previous section. There has been a great deal of research in this area, and a number of proposals for suitable technology. Is now the time to define an architecture for location, and to drive for deployment?

#### **4 A new design for secure, robust operation.**

The Internet was designed in simpler times, when the user community was smaller. It was reasonable to trust most of the users, and it was possible to track down and deal with misbehavior. It is clear today that the Internet, like the real world, includes a population of malicious users. Policing is difficult since the Internet crossed jurisdictional boundaries. War in cyberspace is a possibility, and terrorist manipulation of the Internet is likely. Greed is leading to a range of anti-social behavior, including spam, spyware and adware, and phishing.

A great deal of effort has gone into improving the security and robustness of the Internet, but this effort involves patches and bandages on the original architecture. If there is any single motivation for a redesign of the Internet architecture, it may well be to give us a coherent framework for a more secure, robust and trustworthy infrastructure.

An important aspect of secure, robust operation is to design mechanisms that survive and adapt to attacks, including major attacks that aim to disrupt the whole network. In times of crisis, parts of the network may fail, parts may become isolated, and abnormal traffic patterns and overloads may occur. Critical web sites and sources of information may be disabled. The commercial world is hard-pressed to invest seriously in mitigating this problem, because they are disciplined by competition to invest primarily in features that return a profit under normal circumstances. The research community should take it as a mission to make sure our communications infrastructure is robust enough to resist attack.

*In 10 years, our communications infrastructure should be based on an architecture that provides a coherent framework for security, robust operation in the face of attack, and a trustworthy environment for services and applications.*

There has been a great deal of research concerning security. It is now the time to pull the community together in pursuit of a new architecture for security. While security is often

seen as a technical problem, many of the hardest of the issues in this space arise from a mix of technical, economic and social causes. So breakthrough research in this area will call for a multi-disciplinary approach.

## 5 Operation in times of crisis

Not only should the Internet be robust to attack, it should be designed to support those special needs that arise in times of crisis, both big and small. Today, the Internet has no mechanisms equivalent to an Emergency Broadcast Service, no ability to dedicate resources to first responders, or to provide something similar to 911 access. To the extent that the Internet is becoming the communications infrastructure of the future, we should make sure that it meets the needs of the individual, the region and the nation in times of emergency.

*In 10 years, the network itself, and critical applications that run on it, should address the special needs that arise in times of crisis.*

## 6 Anti-Scale: small networks

A major theme of networking research over the past twenty-five years has been on network architectures and protocols that scale seamlessly over ever-larger distances and ever-larger numbers of attached devices. This theme represents both a positive desire to create a global, interconnected, network and a backlash against work on localized networks that proved completely unable to scale.

But there are signs that it may be productive to revisit localized networks, with an eye not towards using them to enable a global network, but rather to dramatically rethink the local computing environment. A few straws in the wind that illustrate the potential:

- ❑ The rise of Bluetooth. Bluetooth is a very primitive networking technology. Bandwidth is low and any device can have only a handful of concurrent peers, yet the technology's success is proving there's a demand to interconnect devices in a local area. If the number of devices goes up sharply, a new technology will be required.
- ❑ The demise of the backplane. There has always been a blurred line between a small local area network, a cross-connect fabric and a backplane. There's a line of thinking that the computer bus or backplane is now a nuisance rather than help – that we're better off treating the components of a computer as individual devices and networking them together. This is a technology argument that could change the way we compose devices out of parts, including our personal computers, our entertainment systems and our laboratory equipment.

- Sandwicing optical or wireless devices on CMOS. It is now possible to put useful optical or wireless transceivers, with high bandwidth, on top of a silicon chip. Beyond enabling the vision of the network as a bus, these innovations allow us to imagine that information between chips on a card is carried optically or wirelessly. Another way to think of this situation is that the network termination is moving from the edge of the computer enclosure into the middle of each chip. What should an into-the-chip network look like? Does it need to defend against DDOS attacks? Does it run IP? TCP? If my name server goes down, can I still talk to my disk drive? Can I hijack the processor(s) of the computer being used by the guy next to me in the Internet café?

These observations lead to the following challenge:

*In 10 years, we should have local communications architecture that allows the local interconnection of dozens or hundreds of small (e.g. chip size or slightly larger) devices, with price-performance ranging from very low cost to very high bandwidth.*

## **7 Assume Quantum Computers Work**

At this point, quantum computers are a little bit of working logic, and a gleam in many inventors' eyes. But given that there is some working logic, let's assume that the various physical challenges of building a quantum computer are solved and that we have working quantum computers in the next decade. What does this innovation do to networking?

Potentially quantum computers could have multiple profound effects on networking. We sketch two possible effects here, with the understanding that more effects are likely.

The first is simply how do we network quantum computers? Quantum computers don't work on binary data. They use quantum bits (qubits), which are very different from binary. It is an open question whether we will want to network quantum computers, and if networked, whether the quantum computers will wish to exchange qubits. But suppose we do wish to exchange qubits. How would we transmit packets of qubits? Clearly we cannot use optical-electrical regeneration of qubits, as the electrical logic is binary (and would destroy the multi-state information in a qubit). Do we need qubit routers? Does optical amplification work, or does it also damage qubits?

A second effect is on network security. One of the algorithms that quantum computers are particularly well suited to solve is prime factorization. Much of network security today is based on keying systems that assume prime factorization is difficult. We presumably will need new keying systems. We might also need more keys – as one might imagine that quantum computers are also good at the kinds of multiple state searches to crack a key and thus keys will “wear out” faster.

One possible solution to new keying systems and having more keys are *quantum key distribution networks* now being experimentally deployed. While their name suggests a close affiliation with quantum computers, quantum key distribution networks are very different technology. Their important feature is they encode bits in individual photons and then use the physical properties of individual photons to construct key distribution channels that cannot be eavesdropped upon. But much as transmitting a qubit is challenging, building networks to forward individual photons (which cannot be “read” or converted into an electrical signal by anyone other than the recipient without destroying its encoded bit) requires us to discard much of our current optical transmission technology.

*In 10 years, we should have a network ready for the existence of quantum computers. Such a network would allow quantum computers to communicate, and would also have a security architecture that protects the privacy of data, even if quantum computers are available to crack keys.*

## **8 Rethinking the Control/Data Plane Dichotomy**

One of the features of the current Internet is that it is designed to provide a general transport service capable of supporting many different applications. This “data plane” is not designed or optimized for any single application, but is designed for generality and evolvability. The core of the network just forwards packets; knowledge of the application is localized to the edges, where the attached hosts sit.

In parallel to the data plane sits a control plane, which manages the network infrastructure and ensures that data can continue to flow. But the control plane is often equally oblivious to the applications being used.

This is a powerful division of function, because it facilitates innovation and the deployment of new applications, but it has one drawback. Since the core of the network is oblivious to what the user is trying to accomplish, it cannot reliably detect when the user is having problems or experiencing a failure. While the network is designed to detect and correct its internal failures (e.g. rerouting after a link failure), it cannot detect misconfigurations or inconsistencies (e.g. errors in a DNS server) that can keep an application from running.

We are also reaching certain limits in control plane architecture. A central feature of the current control plane is that a device manages its own information. That is, if you want to know what is happening at a particular router in the network, you ask the router (or perhaps a device that is acting as a proxy for the router). The limitations of that model are increasingly being exposed. For instance, if a key router begins to under perform, the last thing we want is every affected user to query the router. Yet attempts to simply push data about devices out in the network have hit the problem that the data any device keeps about itself is quite large and pushing large amounts of unstructured data around the

network does not, in truth, sharply improve network understanding (and harms performance by creating a new, ever present, flow of mostly-unused traffic).

One proposition to mitigate these problems is to add a new construct to the Internet—a distributed layer or plane in the architecture where knowledge about what the user is *intending* to do can be recorded, and can be compared to what is actually happening. Application code, running on edge nodes, can make assertions into this layer about what *should* be happening, the system can monitor what *is* happening, and discrepancies can trigger diagnosis and correction in ways that cannot be automated today.

*In 10 years, the Internet should be augmented with a new set of mechanisms for diagnosis and configuration, which can improve the usability of the Internet, reduce the need for manual intervention, and provide a linkage between application intentions and network behavior.*

Users today find the Internet very frustrating when something goes wrong because it is often not clear where the problem is or who should be notified to fix it. The purpose of this system is to improve the response of the system when things go wrong.

One hypothesis is that techniques from AI, including knowledge representation, fault modeling and machine learning may be useful in building this layer. Whether this approach is the most fruitful one, we should set a goal of mitigating the current problem.

## **9 Giving everything a presence in cyberspace**

Any physical object has an appearance--a manifestation in physical space. In many cases, these are public. Buildings present a facade. Products are wrapped in packages with attractive marketing. People have an appearance, and may choose to offer you a business card. We can imagine a world where everyone has a scanner and a viewer, and all objects manifest some sort of cyber-identity, so that a person can not only "look at" an object in the real world, but "look at" its manifestation in the cyberworld. Products can link to information and marketing. Buildings can link to a list of tenants, or other information about what is inside. Vending machines link to a means of online payment. And so on. This vision raises issues about privacy (under what circumstances should people be linked to information in cyberspace), but would transform a wide range of activities today.

*In 10 years, any physical object should be able to tag itself in a way that links it to relevant information and functions in cyberspace. A context of scanners and online viewers will allow users to see this information in a convenient and interactive manner.*

## **10 Reduce the energy required for communications**

Given current trends in energy-consumption by computing chips (falling sharply per unit of computation thanks to Moore's Law) and radios (largely determined by the laws of

physics and thus going down only if we think hard about how to drive them down), it seems likely that in 10 to 20 years, the wireless interface will be the primary consumer of energy in any device that combines computation and radios (be it a sensor, laptop, or handheld device). For battery-operated devices, the energy consumption will become the single most important figure of merit.

Current research suggests that the community could set itself a bold goal here—not a fractional reduction but orders of magnitude:

*In 10 years, in a wireless network, the energy cost per bit of data transferred should be 1/1000<sup>th</sup> the cost it is today.*

If we can reduce energy requirements by three orders of magnitude, we can transform our concept of what a portable wireless device is. We may be able to abandon batteries all together, and use scavenged energy, like the calculator powered by a solar cell that can work anywhere it is bright enough to see the display. While this goal may sound ambitious, it is certainly within reach. Aggressive research funding by DARPA in the past few years has shown the potential to reduce energy by over two orders of magnitude. With sufficient effort, it seems possible that we could achieve the third order of magnitude in ten years.

This research vision has the virtue that it can clearly drive innovation in a number of directions.

First, it gives a crisp metric for measuring efficiency. For instance, we've known for years that exchanging more state information in reliable protocols usually gives a performance improvement – but quantifying whether the improvement was worthwhile was often hard. The energy standard gives us a clear metric. (Though we should not mistake it for the only metric of merit).

Second, it can further drive the creative development of new radio technologies. A range of radio technologies is believed to help reduce energy consumption, including dynamic transmission energy control, directional antennas, and improved routing protocols.

Third, the advent of software radios makes it far easier (e.g. less expensive) to perform the wide range of experiments required to meet this goal. Expressing this point from the other perspective, reducing energy per bit is an ideal challenge for the emerging software radio community to address.

A final point: we could, of course, commit to a similar goal for wireline networks. But it isn't clear that's an interesting goal. Typically, a wireline network connection is physically close to a source of electricity – and so wireline devices are not, except in emergencies (e.g. battery backup to fiber optic terminators), reliant on battery lifetimes.

## **11 Embrace the software radio revolution**

The advent of software defined radios clearly portends a potential revolution in wireless communication. Software radios allow us to view every aspect of wireless transmission and reception as programmable: how data is encoded, what frequencies the data is sent on, what energy is used to transmit, what media access layer is used, and so forth. All these characteristics are “soft” instead of hard-wired. Fielding new coding techniques or new media access rules is a matter of a software change.

Furthermore, we can use the software radio’s receiving logic to examine the spectrum and discover how the spectrum is being used. With this information, we can then begin to allow opportunistic use of the spectrum, where a portion of the spectrum is allocated but unused, or more daringly a portion of the spectrum which is in use, but in a fashion that permits underlaying (using a signal whose coding or energy level or both are such that the signal does not interfere with the incumbent signal), is exploited by the software radio’s ability to dynamically reprogram itself to transmit in a way that does not interfere with existing use.

*In 10 years, we should have working software radio systems that demonstrate that spectrum can be used and managed in revolutionary ways. We should demonstrate highly efficient use (and reuse) of spectrum, and establish a regulatory regime that permits these modes of operation.*

It is possible that the concept of software defined radios may have an impact similar to that of packet switching. The degree of flexibility, efficiency and adaptability this technology might provide could be as revolutionary and transforming as packets were to circuits.

Software radio is a technology-rich research area. Converting this potential revolution into a real revolution requires a coherent program of research. First, we need to create programming infrastructure for software radios. Currently the available software, while promising, is fragmentary. Second, we need to encourage experimentation. Software radios need to be widely distributed, with appropriate wireless interfaces, and appropriate licenses from applicable regulators to permit experimentation. Third, we need to examine how to enable software radios to co-exist in an environment where much of the spectrum is allocated and has incumbents whose rights must be respected. At the same time, we would like to have a world where incumbents have an incentive to encourage the use of software radios. (This area is one where DARPA has taken a research lead, but there’s always room for more smart ideas).

## **12 Getting there—research and experiment**

This paper describes how the world might be different. It does not describe the research agenda to get there. A research plan is a necessary next step, but this meeting did not provide the time to discuss that. Nor does this document provide citations to the body of work that would sustain these visions. As we noted above, the participants felt that there were valid approaches to achieving all of these objectives, but the articulation of the

research agenda must await a next meeting.

To validate and demonstrate any of these visions, it will be necessary to build some sort of prototype, testbed, or experimental infrastructure. So part of the challenge in achieving these visions is to agree, as a community, what sort of infrastructure would best serve us in our experiments. Some of the objectives in this list might be met as part of a fundamental redesign of the Internet itself, and this research objective would call for a testbed that can demonstrate a new network architecture. Some of our visions would require a new sort of infrastructure—a location infrastructure or a quantum infrastructure. Still others simply require the presence of a support infrastructure, such as a foundry for experimental optical chipsets or a software team maintaining a reference implementation for software radios.

So we offer two challenges to the research community: first, to set itself some long-range visions and work to achieve them, and second to agree as a community on the test infrastructure necessary to support those visions.

### **13 Some missed opportunities**

It might be argued that having a unified vision of the future is not necessary these days. One might ask whether increased commercial interest in computing and communications will provide the driver to translate research results into running systems. Optimistically, this might happen, but it is hard for industry to set a long-term direction or to arrive at a cross-sector unified architecture.

The research community paid little attention to an architecture for home networking. Home networking today is happening, but slowly and incoherently. There were several competing standards for wireline and wireless options (differing more in commercial implications than fundamental capabilities), and different sectors of the consumer market (e.g. computing and entertainment) have not yet converged on common approaches. While there is beginning to be a technology shakeout in this space, there are still issues around ease of use, security, management and debugging, and changing standards. Might this world have happened sooner, and happened more coherently, if there had been some leadership and architecture from a research community?

One could ask about other cases where things may not have matured or converged as well as they might, such as a networking framework for automobiles (where perhaps it is yet not too late). We also acknowledge that all not visions established by the research community are successful (e.g. body-area networks, micropayments and various sorts of middleware and naming architectures). In many cases, it could be argued that the “failure” is not the failure to have a good idea, but a failure to keep pushing until the idea takes root. Part of a long term vision is that it justifies a long-term attention to fulfilling the potential of our good ideas.