Next Generation Peer-to-Peer Engineering

Next Generation Peer-to-Peer Engineering:

Mediated Computing on the Edge

By

William J. Yeager and Rita Yu Chen

Cambridge Scholars Publishing

Next Generation Peer-to-Peer Engineering:
Mediated Computing on the Edge

By William J. Yeager and Rita Yu Chen

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-9479-3
ISBN (13): 978-1-5275-9479-1

To the memory of Gene Kan,
a dear friend and colleague.
His contributions to Gnutella
helped to bootstrap P2P in
the new millennium.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Although the goal of this book is technical in nature as well as instructional, that is to say, if one wishes to know how to build a P2P network, reading this book provides multiple recipes for both the student and the practitioner. Yet, its arrival is not independent of existing economic trends and technical market forces. Thus, up front, before diving headlong into the engineering chapters, let's take a closer look at the significant events that have motivated the necessity to write a book on "P2P Engineering and Computing on the Edge" at this time.

1.1 The .COM Supernova

The year 2000 was amazing. The beginning of the new millennium began with fireworks displays broadcasted worldwide as successive capital cities greeted the stroke of midnight. The high-tech economy responded likewise. Over the next year it experienced worldwide growth to reach new limits as the .COM frenzy took hold. This folie arrived at the point where a clever new domain name could generate a few million dollars in venture capital to fund startups, and early stage, emerging companies. This explosive economic growth had to self-destruct sooner or later. It did, and sooner, when in March, 2000 the stock market bubble burst like a sudden .COM supernova that sent hundreds of these Internet companies into bankruptcy. Since that time most of those .COMs that drove the NASDAQ to new heights have practically all vanished.

It is historically important to note that the .COM stock market "bubble" bursting is not a new phenomenon. The same bubble effect occurred concurrently with the inventions of electricity in 1902, the radio in 1929, and the promise of the mainframe computer in 1966. Initial expectations drove stocks' P/E ratios to unrealistic heights only to see them abruptly come tumbling down again. The new technology in each case was sound and produced long-term revenue, the investors' enthusiasm and the market's response were not.

One can ask first, why did the .COMs come into being, were they and/or their supporting technologies legitimate? And from this point of view, one can try to understand the above events. Secondly, given the void that their disappearance had left, what, if anything, could be done to put in their place the products and services that would result in long-term, stable economic growth in the high-tech sector? This collapse has shown this sector to be the heart of the cardio-vascular system of the new millennium economy. There are those who denied the latter statement and looked to the old gold-standard industries in times of economic decline; but the new economy was and is as dependent on the high-tech Internet technologies as the automobile industry is on our highway system.

The birthrate of .COMs was directly tied to the ease with which one can deploy on the Internet client/server web-based architectures, and this drove the .COM expansion further than it should have. While this ease of deployment was certainly a big plus for the underlying technologies, the resulting digital marketplace could not provide the revenue streams necessary to either support the already existing .COMs or sustain the growth being experienced. One wonders about their business plans since office space was rented, employees hired, software written, and hardware purchased and connected to the "Internet Information Highway," and all with speculative, venture capital investments. Venture capital support was supposed to be based on a sound business plan as well as new ideas based on Internet technology. Here it is extremely important to note that the technology is not at fault. The technology was and is fantastic. Rather, it was the desire to make a quick buck today without worrying about tomorrow's paycheck that brought the economy to its knees. One senses that a sound business plan was irrelevant, and that the hopes for a quick and highly inflated IPO were the goal. When the .COMs went supernova, and their hardware capital was reabsorbed into the economy with bankruptcy sales, this brutalized companies, like Sun Microsystems, that depended on hardware revenue. Similarly, tens of thousands of very talented employees were then jobless. This can only be viewed as a course in "business 101," as "lessons learned" for the young entrepreneurs and their technical counterparts. For those who profited from this bubble implosion-explosion, and there are many, these young entrepreneurs came back, and in force with their technical teams; but also wiser, smarter; they came back and as well made the necessary changes to the system that exploited their talent, energy and dreams.

On the technical side of things, again, the ease of deployment of the hundreds of .COMs was a proof of concept of not only the web based, distributed computing model but also the high-tech design, manufacturing

and delivery cycle. High-tech was and would continue to be responsive across the board, and clearly, the Internet did not go away either as a way of doing e-business, or as a means of enhancing both personal communication and one's day-to-day life. With respect to personal communication, a beautiful thing happened during this time. Strong partnerships and alliances were created between nearly all aspects of the high-tech and telecommunications sectors because personal communication in all its forms can be enhanced with Internet connectivity.

Yes, there was a rush to add features to mobile phones before the technology was mature, but the i-Mode, Java/MIDP experiment of the early 21st century alone proved the business and technical models are sound. And at the same time, the WAP GSM and 2.5G experiments proved that without uniform standards across this space, as well as responsive data networks, these services will be abandoned. The former partnerships continued to flourish, and the latter problem was realized mid-stream at the WAP Forum; WAP and its WAP gateways as proxies to the Internet were abandoned; and end-to-end Internet protocols were adopted for mobile devices. This is discussed in detail in Chapter 6.

Consequently, business opportunities with personal communication, media-sharing and life-enhancing applications were speculated to be a large part of the P2P economic model. We will point out throughout this book how P2P can better address areas like these, and more easily solve the multiple device problems for which client/server web-based solutions have not been adequate. In this book we describe how in some of these cases a marriage of these latter technologies with P2P networks will be suitable collaborative P2P applications in the enterprise, where the documents produced are proprietary and need to be highly secured, regularly checkpointed, with all transactions audited; and in others, pure, ad-hoc P2P networks will be the best solution. Here, one might have the exchange of content, like family trip photos, or video sessions between neighbors connected either on a shared Wi-Fi network, or with cable or ADSL. In all such cases, content will be protected from ISP snooping. And above, in all such cases using P2P to provide computation on the edge, that is, nearer to the users, will significantly improve user-to-service bandwidth. It will as well lessen the bandwidth load on the long haul, backbone networks data must traverse to reach these massive Cloud Services. Also, by distributing such services in this manner, users and providers escape all your services in one location, syndrome. Thus, the distributed servers on which these services run are less vulnerable to attacks.

As the .COM rollout proceeded in the first decade of the new millennium, limitations of the underlying network, its administration and

services were also revealed; SPAM continued to be an uncontrolled problem; bandwidth was not sufficient in the "last mile;" Internet service to China, India and Africa was minimal to non-existent (Allen, 2001, 76-83); and denial-of-service attacks, and security problems were then and still are a constant concern. These are discussed in section 1.3.1.1, and P2P engineering solutions, where applicable, are found throughout the book. For example, security, distributed-denial-of-service (DDoS) attacks, and SPAM are covered in Chapter 5.

In the final analysis, the .COM supernova was not technology gone wrong, but rather a business failure. Yes, we did have those unemployed, creative engineers, managers, and marketeers. Happily, creative beings tend not to be idle beings. Creative energy is restless energy. From this perspective, the "supernova" was a positive event: Just like its astronomical counterpart which must obey the laws of thermodynamics, where the released energy and matter self reorganizes itself to form new stars, one can view the laid off, post .COM supernova employees as a huge pool of intellect, of creative energy that did not remain dormant like a no longer active volcano, but rather, regathered in small meetings in bars, cafes, homes and garages, to create business plans based on surprising and disruptive technologies, some which appeared to come out of "left field," and others from a direct analysis of what is required to advance Internet technology. And the post .COM entrepreneurs were not much older but were much wiser. As a result, the technologies they will, and did introduce were and are based on sound computer science; an examination of the underlying reasons why failure of such huge proportions happened so abruptly; and thus, yielded products with long term revenue as a goal rather than a short-term doubling of one's stock portfolio. Life continued for these individuals with these dreams in place, is here to stay, reshaping itself as necessary, and the .COM supernova was the natural progression of things: A necessary reorganization of a huge amount of energy that cannot be destroyed.

1.2 The Changing Internet Marketplace

Why does a marketplace change? How do we get into online shopping in large farmer-like markets? Fundamental to these two questions is access to the distribution and aggregation of the commodities being sold. The Internet, digital or e-Commerce marketplace is not different. As we accelerated through the events discussed in the previous section, rules for accessing, distributing, and delivering digitally purchased items were put in place. But, right after the .COM bubble burst, most of which was purchased

on the Internet could not be delivered in the same manner. Many examples come to mind. Two typical ones at that time were eBay and WebVan. One requires revenue streams to deliver. Thus, the supply chain of purchased, e-Commerce products to the consumer stalled. It is interesting to note that there were two salient exceptions: Amazon and Netflix. They did not overexpand. Their business models were steady as you go. How such budding companies can build a P2P, Mediated, Edge-Computing as a cost saving infrastructure is one of the several goals of this book.

With respect to digital music, Napster's P2P network was an initial effective as well as inexpensive delivery mechanism. It was brought down by copyright infringement problems and is discussed below. Fortuitously, as a side-effect, it helped bootstrap P2P technology: Even if Nutella, Freenet, LimeWire and BitTorrent were noteworthy early efforts that were also drawn into the fray of copyright infringement, their P2P technologies were an excellent motivation for the P2P research and the diverse systems that followed.

As an example, one of the more successful, early P2P efforts was Sun Microsystems open-source Project JXTA (Project JXTA, 2011). It was initiated in 2001. This open-source project provided a P2P platform on which anyone could build P2P applications and services. This technology was globally downloadable at no cost. And its more than 20 United States patents permitted unlimited derivative works and were royalty free. The latter permitted its CTO to work with the IETF's Internet Architecture Board to create the first P2P Internet Research Task Force. Sun Microsystems as a company had as one of its goals to create Internet standards.

How can P2P ease the access to products and services that are for the most part achieved through a web-based, browser interface or a mobile app. If one needs to search for a hotel at a particular price in each region, then Internet access to this digital information can be extremely tedious and time consuming. While the results are often gratifying, the means of access can certainly be improved. If one wishes to purchase an orange, then it should not be necessary to go to Florida. For the http GET mode of access, one almost always returns to the home site for fresh data.

Napster showed us that one can much more effectively deliver digital content using hybrid P2P technology. Yes, Napster no longer exists but the technology it used is only getting better. There were legal issues with respect to copyright protection, and so digital rights management software has been written to protect digital ownership. Why? Because for those who stopped using Napster, the recording industry realized the huge potential of unloading the infrastructure cost for delivering mpeg to the users' own systems and networks. P2P is sure to eventually become an essential part of

the digital marketplace because safeguards can be put in place to both guarantee the payment for purchased, copyrighted material as well as its P2P distribution. Dr. Stefan Saroiu pointed out that about 40-60% of the peers shared only 5-20% of the total of shared Napster files (Saroiu, 2002), i.e., Napster was used as a highly centralized, client/server system with very little incentive for sharing the wealth of music in the system. One can speculate that the resistance to sharing was directly correlated with the fear of getting caught for copyright theft, and that a system of sharing incentives such as digital coupons leading towards free or low-cost content will be easy to put in place in a marketplace that discourages digital content theft.

Finally, given the vast extent of the now primarily In the Cloud based access by the means of an ISP, the user computing on The Edge will be able to access proximity Cloudlets Services. This will optimize the Cloud's bandwidth, thus providing faster Internet service with the local, high-speed distribution and streaming of media in all its forms. This includes voice recognition without delays; enhanced eCommerce; optimized virtual reality services and apps; and significantly increases AI algorithm's response times. In fact, AI at the edge can potentially yield near-instantaneous decision-making.

For P2P to become a fundamental building block in the existing digital marketplace, the digital content needs to be aggregated closer to home. While powerful servers are essential for the financial management and initial distribution of this data, always "going to Florida" is no longer a sensible solution. Just like the freeways are filled with semi-trucks during off hours to deliver the oranges to local markets, the same use of the information highway makes good "bandwidth sense." As the Internet Cloud continues to expand like the Universe, the consumer experience will require local-proximity-based, Cloudlet data access. This is thoroughly discussed in Chapter 2.

As mentioned above, even with Google search, the web-based process of finding what you wish to purchase is tedious, time consuming and can be streamlined if, for example, the user wishes to purchase from a nearby merchant. One might like to have a digital aide with a template exactly describing a desired purchase, and a P2P app that does the shopping by querying one's local, on the Edge, marketplace Cloudlets and returning **only** the best matches. Scanning, for example, Google lists is really a waste of time. Cloudlets with digital yellow pages for locally available purchases will create a very attractive e-Commerce marketplace. We discuss this thoroughly in Chapter 7 under the topic of Java Mobile Agents on P2P networks.

It is also important to understand that commercial P2P does not exclude user P2P networks for user communication as well as exchanging their personal media with the usual legal requirements for copyright protection. P2P applications using Computing on the Edge technology can enable such small communities in Virtual Private P2P networks.

1.3 What Is a P2P Network and Computing on Edge?

What is P2P? That is the real question. Most people still believe that P2P means the ability to violate copyright and exchange what amounts to billions of copyrighted mpeg or jpeg files free of charge. The music and motion picture industries are strongly against P2P for this reason. We will discuss the historical origins of P2P in Chapter 2, and this history makes no reference to the "Napster-like" definitions of P2P. Rather, it will discuss the foundations of distributed, decentralized computing. One finds also that the non-technically initiated have also begun to equate decentralized computing with a dark force of the computing psyche. It almost comes down to the old battle of totally centralized versus decentralized governments. Note also that Cloud Services provide centralized computing in The Cloud. Amusingly enough, for example, The United States and European Union are organized somewhat like hybrid P2P networks in the definition we will propose below. And capitalism was born out of such an organization of political and legal balance of power.

Where does that leave the cabalistic opponents of P2P? We leave that to the reader to decide. But this opposition is at least partially neutralized by a lack of understanding of P2P, and thus a goal of this book is to help shed some light on this P2P issue. Decentralized, distributed computing is a powerful tool for organizing large collections of nodes on a network into cooperative communities, which in turn can cooperate with one another. Yes, the one possible instance of this paradigm is anarchy where each member node of such an organization of computational resources can do what it wishes without concern for others. At the opposite extreme is a dictatorship. An organization of network nodes that leads to either anarchy or a dictatorship does not have a place in P2P networks as viewed from this book's perspective. Nearly everything in between does. And, certainly, we need to establish some rules to prevent the violation of others' rights whether they are members of society or nodes in a network. Napster from this point of view was not P2P. Rather, it was a centralized system that encouraged non-cooperation among the member nodes or a subtle form of anarchy which is almost a self-contradiction. Centralized because all mpeg distribution began with centralized registration and access to copyright-

protected mpeg files from about 160 servers (Saroiu, 2002). And anarchy-like behavior among the nodes because once a node possessed a copy of a mpeg file the tendency was not to redistribute it, and thus, ignore the underlying "share" the content model which is at the roots of P2P.

Clay Shirky (Winer, 2002) gives us a litmus test for P2P in:

1. Does it treat variable connectivity and temporary network addresses as the norm, and
2. Does it give the nodes at the edges of the network significant autonomy?

While this is a test for P2P, there will be instances of P2P networks from this book's point of view that will treat fixed network addresses and 24x7 connectivity as the norm. Our goal is not to have a purist litmus test that excludes a major segment of the computational network, but rather a test that is less exclusive. To this end a collection of nodes forms a P2P overlay network or P2P network if the following hold:

1. A preponderance of the nodes can communicate with one another; can run app-services enabling them to each play the role of both a client and a server; and exhibit a willingness to participate in the sharing of resources,
2. Nodes may be completely ad-hoc and autonomous, or use traditional, centralized, client/server technology as necessary.

Here one notes the term overlay network. From this book's point of view, P2P networks are overlay networks on top of the real network transports and physical networks themselves, as shown in Figure 1-1.

P2P means mutually supportive nodes on the one hand and being able to use as much of the available technology as is necessary on the other, and thus, to arrive at a network that behaves optimally.

A P2P network in an enterprise will be different from a P2P network in a neighborhood, and the two may or may not communicate with one another. The former will in all probability be stable, and the latter most likely ad-hoc and unstable.

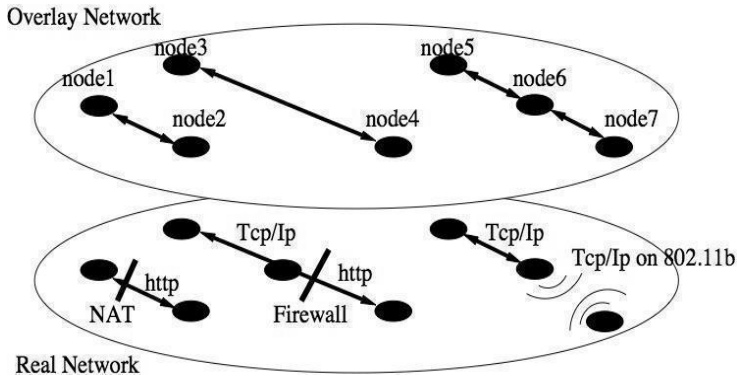

Figure 1-1. P2P Overlay Network

The lifetimes of network addresses and connectivity, as well as an autonomous node's symbolic "Edge" position in the Internet topology lay at the far end of a very broad P2P spectrum of possibilities offered by the above definition. If one wishes P2P to be a success, then the engineering principles to which it adheres as well as its domain, must be able to encompass, and find ways to both interact with and improve current Internet, centralized client/server, based app-services. In fact, the appropriate vision is to view the ultimate Internet as a single network of nodes for which P2P provides an underlying fabric to help assure optimal, and thus, maximum service to each device limited only by the device's inherent shortcomings, and not by its symbolic position in the network. Yes, an ad-hoc, autonomous, self-organizing, network of unreliable nodes is inherently P2P. Yet, a highly available cluster of database systems supporting a brokerage firm can also be configured as a P2P network as can these systems' clients. The members of such a cluster can be peers in a small P2P network using P2P communication for the exchange of availability and fail-over information; the clients can also be members of the same network to help both mediate network wide load balancing, and data checkpointing; as well as a member of a client P2P network to share content and suspend and resume tasks on one another's systems.

In such a configured P2P network there may be centralized client/server relationships to, for example, insure authenticated, authorized access, and this P2P network as well as the pure, ad-hoc P2P network both satisfy the definition, both being points in the P2P spectrum. The application of the fundamentals in this book will enable one to create such networks. But, standard, distributed client/server email and database systems are not P2P even if the clients may keep data locally and can act as auto-servers either

to improve performance or at times of disconnection. These later client/server systems do not communicate with one another as peers and adhere strictly to their roles as clients and servers. This book does not discuss the fundamentals of these latter systems but will describe methods for morphing them towards the P2P paradigm for better performance. Such morphed systems will certainly be hybrid rather than pure P2P, and an extremely important step in the evolution of P2P technology.

Figure 1-2. The P2P Spectrum

The symbolic "Edge" of the network is really better viewed as pre-Columbian network terminology in the sense that before Columbus, the western world believed the world was flat, and had an edge. When Columbus looked for the edge of the world, he never found it, this fictional point of view was dropped, and the possibilities for travel have been limitless ever since that time. If one is at any node in the Internet, then there is not a sense of, "I am at the network's Edge." One's services may be limited because of a slow or poor connection, and this can happen anywhere on the Internet. It is much better to view each node as located at the "Center" of the network, and then do what is possible to make all such "Centers'' equal. This is closer to the task P2P has set out for itself in this book.

1.4 Why P2P Now?

Internet e-Commerce will be as successful as the acceptance of the willingness to both use on a regular basis and pay for the applications and services (app-services) that the digital marketplace offers. One of the reasons we had a .COM supernova was the consumers did not accept the app-services offered by the .COM marketplace in the above sense. Some of the app-services were used some of the time, and very few were used all of the time. Those very few are those that survived.

The acceptance one envisions here means much more than, "Hmm... This is an interesting URL, maybe I'll try it someday." Acceptance is expressed by users saying things like, "This app is so cool I can't get along without it"; "This service is so compelling that I feel like I am under charged for its use"; and "this app is as necessary as my car, my roller blades, and my hiking boots, and, hey, this is fun!" The app-services must offer a break

from the tedium of day-to-day living. People spend enough time waiting in lines, sitting in traffic, and being overloaded with junk postal mail, spam and those unwanted, popup advertisements. Each of the above produces revenue but why perpetuate pain when it is not necessary? In the last three cases the advertisers are neither thinking of, nor care about the recipient of the advertisements, rather they use any technique they can to increase sales revenue. How many times have you, the reader, been frustrated by these advertising methods? As its primary goal, the kind of acceptance envisioned here requires maximal service with minimal user hassles. Furthermore, optimal app-service response time is essential to enable the creation of real sources of revenue rather than using bogus nuisances for this purpose.

To achieve maximal service with minimal user hassles we must look beyond the client/server/Cloud-Service-mode of distributed computing that drives the Internet. The era has arrived where billions of devices will be interconnected. Although the client-server/Cloud-service structured network is straightforward, and has served us well up to now, even software applications looking up the right server is a painful job for both clients and servers as our every-day directory service, Domain Naming Service (DNS) has the potential to become a bottleneck with the sustained growth of the Internet. In fact, a Distributed Denial of Service (DDoS) attack in October of 2016 on Dyn, a company that controls much of the DNS infrastructure, brought down most of the United States' Internet for an extended period of time.

For authentication and authorization, centralized, server-based services such as Kerberos are in use. Internet security protocols like Secure Socket Layer (SSL) and Transport Layer Security (TLS) currently require centralized Public Key Infrastructures (PKI) and well known, centralized Certificate Authorities (CA) to manage X509 certificate creation and distribution. These systems are also vulnerable to disruptions like DDoS attacks and are required to do secure Internet transactions.

Finally, Cloud Services such as AWS, Azure, and Google Cloud that Build, Deploy, and Manage Websites, Apps or Processes certainly save money since the private data centers given the costs of the latter. But, attacking any one of these Cloud services can disrupt thousands of businesses and millions of users. This is an "all of your eggs in one basket" mode of operation. It is in fact one of the problems that inspires distributed computing on the edge. There is no reason why one cannot have edge-based Cloud services, which we call Cloudlets, where such a service is equivalent to peer in a P2P network of a regional or local user base. These Cloudlets will store the only data that is in use by their P2P network. This will not only significantly improve the achievable bandwidth and thus user

satisfaction but also make these more finely distributed services more difficult to attack. These Cloudlets are not Internet cache storage. Rather, they are complete, user proximity services. The latter can abandon media that is no longer in use, and just as easily refresh it from the corresponding Cloud service as required. It is maximal resiliency.

Beyond these computational infrastructure limitations in the client/server/Cloud model, we are also faced with the Mobility paradigm. One travels with her/his mobile device(s), or a laptop and would like to communicate with another such system. There is no reliable solution for the problem of ad-hoc mobility where a node can appear in a network, joins it, and begins to communicate with other nodes. We now have millions of mobile devices with disposable IP addresses and no home address identifiers. We need solutions for these mobile devices to discover and communicate with one another with end-to-end, secured connections. Facebook, Twitter, and Instagram don't cut it with their viral behaviors.

One can also envision P2P sensor networks as well as P2P drone networks. The former sensors might monitor a bridge for material failures; shopping centers for misbehavior; aid firefighters in locating hotspots; locate poachers; monitor traffic and alert congestion; go to areas of high radioactivity, and, etc. These can be viewed as inter-P2P networks, AI enhanced swarms with well-known home bases for real-time data communication as well as guidance. With the proper computation power and enhanced cameras becoming mobile, mini-neural networks executing machine learning algorithms, they can send their thus preprocessed images to their home bases for more intensive machine learning. They will save lives of humans as well as animals as well as protect both public and private property.

To build a reliable computing powerhouse to serve billions of clients and applications, during the past few decades, companies, institutes, and governments have viewed Moore's Law as a monarch to follow, as well as a limit to challenge. Sooner or later, the limit will be reached at the level of each individual machine, and scientists have already begun to investigate the possibility of building more powerful computing engines by using more advanced technologies from optical to quantum that will no longer be subjects of the Moore monarchy. We are excited about the future, at the same time, we are worried about the present: idled computers; Internet services wedged like the 5 p.m. traffic on a Los Angeles freeway; devices no longer able to communicate with one another; the impossibility of secure communication between any two devices; and wasted manpower and energy outages. Are we solving the right problem? Are there better solutions already available?

We, more than ever, need P2P now because with the duplication of information at various locations closer to the systems accessing it, even when one or more these sites are disabled, the information can be retrieved more efficiently since both the load on, and access bandwidth to a centralized resource are minimized; with the direct connection between any two devices identified by unique IDs virtually achievable, the centralized directory lookup will no longer be either a "must-have" component or a source of bottlenecks in the network; multiple device, drones, and sensors types of ad-hoc mobility can be achieved; and with P2P network based, mobile agent services, objects, applications and clients can communicate and share information with each other in such a way as to minimize the users' involvement in tedious search tasks and thus make systems that are more user responsive. There are many more possibilities brought by P2P technology and any one of them can lead us toward the next wave. With respect to timing and the state of current technology, these possibilities are much closer to realization, and preferable to us sitting here and waiting for the next technical revolution.

So, why will P2P now help optimize Internet access and blow away this illusion of the user isolated at its edge? A short answer to this question is that the current Internet technology without P2P cannot support the sustained, optimized growth of multidimensional, multiple device app-services, and the network topology which P2P implicitly creates will be location independent, and hot with localized activity everywhere. Let's look at why this is true.

As mentioned above, one of the first requirements is user-based, app-services that fully support multimedia data. This means that among other things, at least music, and video must be delivered on demand. The evidence is already here that centralized services cannot support the current demand for domain name lookup (Cheriton, 1988, 314-333), and the massive exchange of multimedia content is a huge problem in comparison. The bandwidth is not there, and the centralized, Cloud based web-services solution always adds more bandwidth capacity, and larger server farms. This is historically viewed as keeping up with the increasing demand by providing the same quality of service. This is neither acceptable nor successful at providing user satisfaction. The analogy is adding more lanes to the freeways to "improve" traffic flow, rather than seriously looking at alternative solutions that will either be convenient or compelling enough for drivers to consider using them.

The build-out of Wireless LANs (WLANs) based on 802.11a/b networks anticipated in the first decade of the 21st century has arrived. As will be discussed in Chapter 6, P2P is a natural fit for optimal content

distribution in WLANs. In section 3.4 it is pointed out how P2P will encourage an evolution of the organization of a mixture of network devices again leading to an optimal use of bandwidth and services to eliminate the centralized bottlenecks reinforced by the pre-Columbian illusion of where the center of the Internet is located. The world is not flat and the Internet's edges as well as the sky itself with air born, mobile P2P networks can all be hot spots of computation and communication.

A second way P2P can optimize the Internet now is by taking advantage of the processing power of quiescent CPUs at no cost. It was projected in 1966 that mainframe computers would revolutionize the world. Neither the arrival of the now-extinct mini-computer nor the microprocessor was anticipated. A mobile phone's processor is more powerful than a typical 1980's mainframes! Mobile devices included, there are more than a billion computers out there all connected to the Internet, and most of the world's population is not yet connected. The existing available processing power is massive. Using P2P one can create coordinated, loosely coupled, distributed nodes as members of a P2P Network. SETI@Home (SETI@home, 2022) is successful as an administratively centralized, computing grid where the responsibility for decisions is administered by software at SETI@Home's University of California's laboratory. With the addition of P2P capabilities SETI@Home like P2P networks will be able to offload some of these administrative tasks by adding server capabilities to each SETI@Home client. This will help to both lessen the bandwidth used to and from their laboratory to the volunteered computers, and speed up the overall grid computation by, for example, permitting checkpointed jobs to be off-loaded to another client directly.

Right now, one's home can become a fully connected P2P edge network. This network in turn can be connected to either a laptop, PDA, mobile phone, automobile, or workstation in one's office behind a firewall giving each family their personal peer community. This book presents the fundamentals sufficient to initiate the process.

1.5 Basic P2P Rights

P2P Networks are organized overlays on top of and independent of the underlying physical networks. We have a definition that is general enough to permit almost any device with network connectivity to be a node in some P2P network. And our definition permits these networks to be both ad-hoc and autonomous, and their nodes to be anonymous. It also admits the possibility of strongly administered well authenticated, secure networks. And, in either case, both openness and secrecy can and will exist. This

paradigm is frightening to some because on the one extreme it is antagonistic to George Orwell's state in his book, "1984." "Big brother" will not know that you exist and therefore cannot be "watching you." It is frightening to others because it also permits Orwellian P2P networks where there is no secrecy, all communication is both monitored and audited, and all data is in clear-text and is either monitored in real-time or saved for surveillance. The former and latter will then use contemporary Big Data Analytics. The latter, among other things, can take advantage of both Natural Language Processing and Machine Learning algorithms. What's important is the freedom of choice P2P offers. This choice has concomitant responsibilities, a P2P credo if you like: Respect one another's rights, data, CPU usage, and share the bandwidth fairly; spread neither viruses nor worms; be tolerant of one another's resource differences; be a good network neighbor; and do no harm to others. The nature of credos is to be violated. That is understood and part of human nature. The goal is to minimize non-altruistic P2P behavior by either localizing it to those P2P networks where it is acceptable, or appropriately punishing it when it is discovered. The usual punishment will be expulsion from the P2P network.

Rightly enough, in the sanctity of one's home can be a P2P network whose members are all the devices and appliances that have network presence. As you will learn in this book, a P2P overlay network is independent of the multiple possible "underlying" real networks and their protocols, for example, bluetooth, Wi-Fi and ethernet networks, some of which may or may not use the Internet protocols on the local area network in the home. All in-the-home communication and data can be private, and only search warrants will permit entry.

The United States Bill of Rights can be viewed as a P2P supporting document since freedom of assembly, speech and the avoidance of unreasonable search and seizure are at the heart of P2P. And certainly, one can imagine a well-organized "network militia" bearing its software arms to secure the network and the data resident therein. Freedom of access for network devices and their associated data are at the heart of P2P networks. The rules for the network and data use are decided by the member nodes and are member nodes' policies. One should be able to then purchase several devices and appliances as well as applications using the appropriate, hopefully open source, standard based P2P protocols[1], to enable a P2P

[1] This book teaches the reader the fundamentals of how to build a spectrum of secure P2P Networks that support Edge Computing. Ultimately accepted standards created by a recognized standard body may vary from this book. For without standardized P2P protocols the ever present, dark side of the Internet prevents practitioners from guaranteeing secure P2P networks.

network, and permit the various nodes to join the network.

1.5.1 "Freedom of Assembly" for Right to Connect Everything

The first decade of the new millennium saw an exponential growth of network aware devices capable of sending and receiving data. The list is ever increasing, and the combinatorics defy one's imagination. Along with computers we have home routers and modems, PDA's, mobile phones, automobiles, TV's, surveillance systems, door bells, temperature controls, light switches and light bulb receptors, fans, refrigerators, alarm systems, wrist watches, stoves, dishwashers, ovens, home theaters, electricity and gas meters, pet licenses, eye glasses, rings, necklaces, bracelets, etc. Any combination of these can be interconnected to form ad-hoc P2P networks. One might ask, "To what end?"

It is easy to place oneself in a scenario having just left home and worrying if the oven or stove was left on. Rather than turn back, a simple control panel on these devices which are peers in a home P2P network and this home edge network can be accessible with either a wireless device in one's automobile or a mobile phone, both as peers in one's private home network, is sufficient to make a quick check. In fact, one could launch a mobile agent securely within the home LAN to do a full integrity check of the home and report back. A few seconds later one will receive an "all is well," or "you left the stove on, and it's turned off?" Before and after photos can be sent.

Another scenario is ad-hoc networks of people in coffee houses, railroad stations, sitting in the park, or in an outdoor cafe. Here, jeweled bracelets, or necklaces might light up in similar colors for all of those who are members of the same ad-hoc, P2P network community. In the evening when viewed from the distance, one can imagine small islands of similar colored lights moving about, towards and away from one another, in a beautifully illustrated, ad-hoc social contract as "Smart Mobs (Rheingold 2005)".

These scenarios are endless, practical, and part of our future. P2P engineering for wireless devices and networks is discussed in Chapter 6.

1.5.2 "Freedom of Assembly" for Survival of the Species

Ecosystems are self-organizing, self-maintaining and in case of reasonable injury, self-healing. There is life and death, and the eco-system evolves using natural selection. This process continues and new life forms arrive over time as the result of mutation. Eco-systems are great for trial-and-error testing. The same must be said for P2P overlay networks. Peers

come and go, crash during data transfers, lose their visibility, and are rediscovered. New devices are accepted on face value and permitted to participate in the network's activities. P2P networks are spawning grounds, playgrounds for creative thinkers. In this manner, a P2P network can continue to gather new members, make them as known as policy permits, and behave much like ecosystems where diversity leads to survival of the fittest. Peers are free to assemble with others for the interchange of content. Peers like mobile-agents are free to traverse the P2P network, visit those peers where entry is permitted, gather information, and return it to their originators.

As such, "Freedom of Assembly" is the ultimate P2P right. As "what is P2P" defines, although each single device is part of a cooperative whole, it is a node in a P2P network and makes its own decisions and acts independently. A device's presence is neither required nor denied. Hence, the failure of a device should not be harmful to other devices. If two devices are connected, and one abruptly crashes, this should be a little hiccup in the P2P network, and there ought to be a redundant device to take its place. Still, everything has two sides, this independence also means that there might not initially be anyone who will help this poor, temporarily stranded guy. As for a highly available, client-server system, there always are servers behind each client, and back-up servers configured for each server, but they are subject to bottlenecks, resource depletion and denial-of-service attacks. So, these self-maintaining, self-healing and self-adaptive features cannot always reduce the burdens on client/server, centralized systems. On the other hand, for a device in a P2P network they are not only essential but rather they are inherent in the network ecology. Thus, the "poor guy" who was sharing content and abruptly lost the connection can expect to resume the operation with another node although this recovery may not be instantaneous. During its apparent isolation it might receive the welcome visit of a wandering mobile-agent that is validating the network topology and can redirect the peer to a new point of contact. Similarly, it is difficult for hackers to use denial-of-service attacks because, like an ecosystem, there is no center to attack. P2P networks have built-in redundancy.

From a software engineer's perspective, ideally, P2P software must be written to reside in a self-healing network topology. Typically, any device uses software to monitor its tasks, schedule its resources to maximize its performance, set pointers and re-flush memory for resuming operations efficiently after a failure. At the higher level, the P2P software should be able to adjust to the new environment for both recovery and better performance. For example, a device might have dedicated caches or tables to store its favorite peer neighbors to be used for fast-tracking connections

or network topology sanity checks. When either a neighboring device fails or one of its buddies is not so "truthful" for an intolerable period, the P2P software on the device should be able to dynamically modify its records. In this way, at least the same connectivity can be guaranteed. This is just one of the most straightforward cases showing that P2P software needs to be self-healing and self-adaptive if the network is to behave in the same manner. The engineering dynamics of these scenarios is discussed in detail in later chapters.

Unfortunately, not all devices are capable of self-management, for example, the small, wireless sensors. Such small devices don't have enough computing power and memory to have such sophisticated software or hardware embedded. So, they must rely on other machines for the above purposes. Although the P2P purists hate to use the "server" word, it is true that the small devices require such "server-like," proxy or surrogate machines, and this fits perfectly with the definition of P2P overlay networks defined above.

As mentioned just above, "Freedom of Assembly'' in P2P networks is supportive of a multiplicity of devices and software "organisms." They arrive, try to succeed, to find a niche, and either continue to flourish or die out.

Since the early 1990's mobile-agent technology has been looking for the appropriate execution environment. They can be invasive, pervasive, informative, or directed, and come in all shapes and sizes. For the authors research, they work best when implemented in JAVA because the byte code is hardware independent. Mobile-agents can be written to adapt to self-healing, ad-hoc network behavior and, in fact thrive in such an environment. The very fact that they are mobile by nature, and can have self-adapting itineraries and agendas, can be signed and thus secured, and are opportunistic as they travel, they have always required a network eco-system for their survival and evolution in mainstream technology. The authors of this book are advocates of mobile-agent technology as applied to P2P overlay networks. The engineering details are discussed in Chapter 4.

1.5.3 "Freedom of Speech" for the Right to Publish Data and Meta-data

As previously mentioned, the data or information published and transferred on the Internet is multi-dimensional, and enormous in volume. Thus, brute force pattern matching techniques for searching for data are no longer durable and become dinosaur-like relics from the era when only textual data was available. A file sharing system which depends on such

simple techniques can be easily hacked since it only requires data corruption by a virus to destroy data. Now, a description of data, data about data, meta-data, is an essential component of the organization of data on the Internet to make tasks like search achievable. With meta-data, for example, one can keep signed hashes of the data in meta-data that permit one to detect modification attacks. Similar techniques can be used to detect data that has been modified while being transferred on the Internet. Nodes on a P2P overlay network have the absolute right to exchange data or meta-data or both.

This meta-data can be public or private, freely published and subscribed to by everyone, or secret and viewable by a select few. Meta-data can be stored as clear text and posted to a public domain site for wide distribution of the data described by this meta-data. One of the immediate uses of these sites is to share research publications among institutes. On the other hand, P2P applications have the choice of not hiding or hiding meta-data. They can have strong encryption or use secure IP (IPsec) so that data or meta-data that is being exchanged can be impossible to monitor because well written security systems can assure the end-to-end privacy of these "conversations." Thus, encrypted meta-data can and will be impossible to detect on peers' systems. Also, access to a system's data directories, i.e., the meta-data describing the files on that system, can require authentication with appropriate credentials for access, and this meta-data can be transmitted as clear text or encrypted descriptions of these directories. Thus, again it may only be visible to the possessor of these credentials. Thus, given the current encryption key sizes and algorithms, viewing the clear text is impossible. Processing speed is so fast that encrypting or decrypting a megabyte of data is negligible. Thus, the processing time required to keep both local and remote user data and meta-data secret is a given. "Freedom" of Internet privacy protection has almost no obstacles because the cryptography code which implements the required algorithms is freely available on the Internet (Free Code 2003) (The Legion of the Bouncy Castle. 2022) (OpenSSL 2015).

Noting that thirty three percent of all internet traffic is directed towards pornographic sites (Buchholtz 2019), will P2P networks be any different with respect to data and meta-data that is published? The answer is probably not. The "Freedom of Speech" gives one the right to publish data and meta-data within certain legal limitations, and the Internet knows fewer and fewer boundaries to data exchange. But do note with caution that the first amendment to the United States Constitution being applied world-wide does inspire significant resistance from governments that wish to control the information that crosses their borders.

When P2P networks are pervasive, the publication of content may or may not reside on individuals' systems as a function of the networks placed on the P2P spectrum. As such, some of these systems will be much more difficult to locate on the Internet because their visible IP network addresses are NAT endpoints. Still, the permitted and accepted private exchange of data and metadata must be no different than a VoIP telephone conversation moving from base station to base station. As such, the problem does not reside with the system that is used to enable the conversation to take place, but rather with the endpoints of the conversation. Endpoint nodes on a P2P overlay network are assigned permanent identifiers as addresses that are globally independent of their real network addresses. And thus, the P2P protocols and routing we define on the overlay network adapt to the arbitrariness of IP addresses in real-time to give nodes true IP address independence. Why? The overlay network binds to the Internet layer and recognizes and changes to the device's IP address. It then updates the overlay routing of {endpoint node ID, IP address} and this propagates across the overlay network as necessary. The details are explained in Chapter 4.

New technology forges new pathways across all systems including legal ones. This always has and always will be the side-effect of crossing new frontiers. It's exciting for some and frightening for others, and one generation's laws may become the next generation's blue laws, i.e., outdated relics of historical interest which will always have some diminishing in time support. Undeniably, all users will be subject to the current "laws of the land," and open to arrest and prosecution for their violation. But P2P technology will create new markets for honest digital commerce with enormous economic incentives and will also permit private network conversations between consenting adults along with the expected criminal activity. The latter is well understood human behavior that P2P neither encourages or discourages much like a highway neither encourages or discourages a driver to exceed the posted speed limit. The solution to these problems is neither to abolish driving nor stop innovative technological progress because it can be misused. Clearly such reactions are neither well thought out, nor well founded, and will not lead to long term, fair and meaningful solutions. The latter will come from the technologists themselves working with lawmakers and media firms, and always respecting an individual's Basic P2P Rights. The engineering aspects of data and meta-data are discussed in Chapter 3.

In the next section we give an overview of the book as well as a brief description of a P2P Pseudo-Programming language that we invented.This enables a more technical description of our protocol design.

1.6 Contents of This Book and P2P PSEUDO-Programming Language

This book is organized as follows:

- Chapter 2 gives an historical perspective on both the evolution of the Internet in two phases, pre-WWW and post WWW, and the roots of P2P.
- Chapter 3 covers the essential engineering components of the generic P2P model. These include the descriptive language and resulting document space, unique peer identity, the overlay network, and communication components. The chapter concludes by showing how these components can be assembled to permit communication on the overlay network given the limitations of the real, underlying network and physical layers.
- Chapter 4 gives life to the components, and describes protocols used to create an active P2P network. Here connecting, routing, load balancing, and querying are discussed.
- Chapter 5 we present the details of how one can implement standards-based security in such a system. We conclude this chapter by applying these security fundamentals to demonstrate how to create secure Java mobile agent P2P services.
- Chapter 6 is a thorough discussion of wireless networks followed by showing how P2P can enable exciting new applications that are device and bearer network independent, and thus be a long needed, unifying force between wired and wireless networks. We also describe what is required to build a Java P2P application for mobile handsets.
- Chapter 7 explores some possible P2P applications starting with the familiar email, and chat programs and ending up with less familiar and innovative, almost science fiction-like possibilities.

To be able to explicitly express the engineering principals in this book, a P2P Pseudo-Programming Language, 4PL has been devised. The syntax and semantics of 4PL are defined in Appendix I. 4PL permits one to both programmatically define nodes on the P2P overlay network, as well as describe their interaction by defining each P2P component we introduce in Chapter 3 as a 4PL data type and creating a set of associated 4PL commands to which one can pass and return typed variables.

As mentioned above, in Chapter 4 we define several overlay network communication protocols. We will use 4PL here to create flow charts to

describe the precise protocol behavior. Thus, 4PL will give a solid logical foundation to the engineering and protocol definitions and eliminate the possibility of inconsistent behavior barring any 4PL programming bugs. It is suggested that the reader uses Appendix I as a reference when reading Chapters 3 and 4.