The History of Multiple Access and the Future of Multiple Services through Wireless Communication

Address at Marconi Centennial and Rutgers WINLAB Decennial
September 30, 1999

Andrew J. Viterbi
QUALCOMM Incorporated

Between 1895 and 1901, Guglielmo Marconi first demonstrated the feasibility and tremendous potential of wireless communication, progressing during that time from small home grown experiments to transmission of the first transatlantic wireless messages. In the century that followed, our lifestyle and culture has been modified dramatically by a series of culture altering wireless applications, including broadcasting (both audio and video), radar, and mobile telephony. Yet, we find ourselves once again at a threshold of explosive growth, this time in wireless access for an ever widening user population. The six billion inhabitants of Earth are currently served by only about 600 million fixed wireline telephones, but already around 300 million wireless mobile phones are in operation and in the next decade the population served by wireless telephony is expected to well exceed that served by wireline. An even greater culture change may result from wireless data services which provide access from anywhere and at any time to a virtually unlimited base of knowledge. The key to unlocking these benefits is the implementation of efficient means for multiple access by numerous simultaneous users and the provision of multiple services to a population with diverse requirements and resources.

Multiple Access Technologies

The origins of multiple access date back to Marconi's early experiments. In 1900, he was awarded Patent No. 7777 for the "Tuned Circuit," which was the enabling technology for Frequency Division Multiple Access (FDMA). In fact however, the first experiments probably involved Frequency Division Multiplexing (FDM). The difference is that FDM refers to transmission of multiple sources from a single location by modulating each on a separate carrier separated sufficiently in frequency. Each receiver contains one, or several "tuned circuits" or more generally frequency filters, each of which isolates one of the received multiple sources and sends it to a demodulator to recover the desired source signal. FDMA operates on the same principle as FDM, except that the sources and their respective modulated carriers emanate from different transmitters, generally not co-located. Multiple Access is a term often used for both forms, but FDM is somewhat simpler because from a single transmitter the frequency separation of carriers is easily maintained, while FDMA with separate transmitters must carefully control carrier frequencies, particularly for moving users whose frequency is changing due to the doppler effect. In any case, FDM and FDMA are the only multiplexing and multiple access techniques which can handle both analog and digital transmissions.

For digital sources, two alternative technologies have evolved for multiplexing and multiple access: time division (TDM and TDMA) and code division (CDM and CDMA). With the beginnings of the computer industry in the 1950's, TDM evolved naturally since it is a way to multiplex several parallel data streams generated in a single location into one serial data stream. Thus time division multiplexing is synonomous with parallel-to-serial conversion. TDMA, on the other hand, had a less obvious initial motivation. In 1964, a U.S. Department of Defense study group compared the various multiple access technologies for geosynchronous satellite networks being designed for high speed (multi-megabit/sec.) trunked digital traffic among a few satellite earth terminals employing large dish antennas. In such cases, the limiting element is the satellite's transmitter power which is proportional to its launch weight and thus must be operated as efficiently as possible. To achieve this, it is necessary to drive the satellite's transmitting power amplifier into saturation, which results in nonlinear operation. When multiple transmitters' signals arrive simultaneously at the satellite, as in FDMA, and are amplified together by this nonlinear amplifier, cross-products are formed which both steal power from and interfere with the desired signals. On the other hand, TDMA guarantees that only one transmitter's signal is received at any instant which avoids the generation of the disturbing cross-products. Thus nearly all geosynchronous satellite digital networks in the 1970's employed TDMA in preference to FDMA or CDMA. As satellite dishes became smaller, while satellite transmitter powers increased, the emphasis on efficiency shifted to the earth terminal. For VSAT's (Very Small Aperture Terminals), the peak-to-average power ratio, which for TDMA is proportional to the number of terminals served, becomes a limitation, while satellite transmitters are no longer driven into saturation given the weaker uplink transmitter powers. Furthermore, the time synchronization of many small transmitters became a problem. The combination of these and other considerations led to the choice of FDMA for VSAT networks by the 1980's. In terrestrial communication, while the majority of digital cellular networks which were designed in the 1980's were designated as TDMA, they were in reality hybrid FDMA/TDMA systems. For in North American TDMA (D-AMPS) or in Japanese TDMA (PDC) only three users time share a single carrier, while multiple carriers 30KHz or 25KHz apart are transmitted using FDMA; the European GSM system time shares eight users but employs FDMA carriers 200KHz apart. In all cases, the total bandwidth used is 10 to 20 MHz wide so that between 50 and several hundred FDMA carriers are employed in such hybrid TDMA/FDMA networks.*

Code division multiple access (CDMA) has a far different pedigree, also dating back to the 1950's. As its name implies, users' signals are isolated not by separate time or frequency slots, which are occupied in common by all users, but rather by unique underlying codes, which when decoded restore the original desired signal while (ideally) totally removing the effect of the other users' coded signals. For this ideal case the codes must be time-synchronized and orthogonal , meaning that any two users' codes must differ in half their symbols and agree in the other half. This characteristic

---

* Actually in a sense FDMA underlies all other multiple access techniques because spectrum regulators assign

and particularly synchronization in time is easily achieved for code division multiplexing, where all sources destined for all users are transmitted from the same location, such as a base station. For multiple access, on the other hand, time synchronization is generally not practical since users are separated in distance which will change with motion; additionally, multipath may produce different time shifted replicas of each user's transmitted signal and code. Thus for CDMA, users' codes are generally chosen to be non-repetitive over a very long period, which does not guarantee orthogonality over the shorter period of each user's transmission , but does ensure a small effect on the demodulators of other users.

An important side effect of code division is that each user's transmitted bandwidth is greatly enlarged by making the coded signal's symbol rate, or clock, run much faster than the digital data rate of the source. For example, if the data rate is 10K bits / sec, the code clock symbol rate may be 1Mbit / sec or 100 times as fast. The result is an occupied bandwidth approximately equal to the coded rate; hence the term "Spread Spectrum" is often used interchangeably with CDMA. This, in fact, better describes the origins of CDMA. As early as World War II but with greater intensity and sophistication beginning in the 1950's , spread spectrum was employed in military communications to protect against hostile interception and interference or jamming. For if the enemy does not know the communicator's code, the latter's signal will appear merely as noise. More significantly, if the enemy tries to jam the transmission with any form of radio

---

licenses in frequency increments for various different purposes.

signal, the intended friendly receiver's demodulator in the process of decoding the desired signal will transform the hostile signal into a spread spectrum form approximating wideband noise. The effect is to reduce the hostile jammer's effectiveness by a factor known as the "processing gain" or "spreading factor" which is the ratio of the code rate to the original source's bit rate (100 for the example just given previously). Essentially, spread spectrum or CDMA is the "best" signaling modulation for even the "worst" form of jamming signal. (This is sometimes called the mini-max solution of a game between communicator and jammer.)

This historical military application, took on added importance with the proliferation of military geosynchronous communication satellites in the 1970's and 1980's, which are particularly vulnerable to jamming from almost anywhere. The first commercial applications of CDMA were also to communication satellites because as the geosynchronous orbit space became more crowded and earth antennas became much smaller with consequently wider apertures, transmission to and from satellites began to interfere more severely with one another. Hence the interference suppression properties of CDMA made this the multiple access technology of choice; and this has remained true as well for the new generations of low earth-orbit satellite networks. Terrestrial mobile cellular telephony became the overwhelmingly pervasive multiple access application of the 1990's with approximately 300 million subscribers at the end of the century. The industry moved from the analog modulation of the 1980's, which can employ only FDMA, to digital modulation in the 1990's to take advantage of voice compression and advanced modulation and coding techniques in order to serve more subscribers per base station in their allotted frequency spectrum. The impetus which began in European standards organizations in the early 1980's was towards TDMA, or more precisely to the hybrid TDMA/FDMA approaches mentioned previously. CDMA mobile cellular telephony was not standardized until 1993, nearly a decade later, yet today it supports nearly 50 million subscribers, about a quarter of the digital cellular population, and is the fastest growing technology. From a market perspective this is primarily due to its ability to serve more subscribers per base station in the allotted frequency band than any other multiple access technology. And the technical reasons are several, but they are summarized by three attributes in which CDMA excels; channel measurement, control and interference suppression. Briefly, measurement accuracy increases with the bandwidth occupied, and spread spectrum helps particularly in identifying, isolating and mitigating multipath propagation. Control is facilitated by the sharing of bandwidth by all users, with an important corollary being the ability to perform soft handover between base stations. Interference suppression is the hallmark of CDMA, which not only permits more users per base station, but also avoids the efficiency reduction in FDMA and TDMA necessitated by the requirement to assign different frequency allocations to neighboring base stations to avoid mutual interference, and even to antenna sectors within the same base station.

While this last viewpoint is not yet appreciated universally, the majority of manufacturers and service providers worldwide are in basic agreement. The first supporters were the North American based manufacturers and the service providers for a majority of the cellular and PCS market. These were followed soon thereafter by the South Korean cellular industry, whose government was the first to standardize CDMA exclusively for digital service. Eventually all the major Asia-Pacific and South American nations also adopted CDMA as an accepted standard; in fact the Asian subscriber count, led by Korea and Japan, currently exceeds the number in the Americas. Western Europe is the only regional holdout, driven by political and economic considerations to approve a single standard (GSM based on TDMA) through its regional standards organization, ETSI. Yet for future so called "Third Generation" services, requiring higher speed wireless data networks, new standards based on CDMA are being universally endorsed by Asian, European and American standards bodies. While the European and Asian versions differ from the North American, mostly in insignificant details, although in one important characteristic (base synchronization), they all benefit from the CDMA advantages described previously.

The Future of Multiple Wireless Services

The wireless industry is preparing for a plethora of new services ranging from e-mail and fax delivery to high speed document distribution, video telephony and even game playing. But to achieve all this, the industry is faced with numerous new challenges, both in marketing and in technology development and acquisition. The impetus for both are the expansion of personal computer ownership and the success of the Internet.

Early wireless data service is simply imitating the evolution of wireline data in the 1980's and 1990's, but unlike the situation in wireless telephony, wireless data has a long way to go. The combination of fax, e-mail and Internet access has propelled wireline data to the usage level of voice and beyond, while wireless data accounts for less than one percent of voice usage, which is not altogether surprising considering that the introduction of digital service to any major degree occurred only within the last couple of years, while the transmission of data on analog wireless modems was far less successful than the experience with wireline. Only very recently have wireless digital data services been available in the U.S. and these at only 14K bits/sec or lower. Korea and Japan are about a year ahead with such offerings and are currently upgrading to service at 64 Kbits/sec or slightly above, a service which will probably not be available in North America for several months. But what is really desired is service at several hundred Kbits/sec or even above 1 Mbit/sec for high speed data transfers and for downloading from the Internet and the World Wide Web. These are services which are just becoming available in major metropolitan areas of the U.S. through cable and wireline digital subscriber line (DSL) network connections.

What then does wireless have to offer for such Internet and Web-based services? First, obviously it uniquely serves the nomadic user, including the "road warrior" with anytime, anywhere data requirements accessed through his or her portable laptop personal computer. But the label of nomadic user now applies to an ever larger percentage of the population of developed countries. As personal digital assistants

(PDA's) or palm-size terminals acquire many of the capabilities of PC's, ever larger numbers will access their e-mail, directories, sports scores and stock quotes, and even surf the Web with wireless data communicators built into their PDA's. Another impetus for wireless is its avoidance of installation requirements, in contrast with cable and DSL which until now have required several person-hours per location to install. And while the wireless bandwidth resource may be more scarce and costly, it represents only one segment of the overall network costs.

I believe the challenge in wireless service provision is to understand the nature of the traffic and the technological capabilities and tradeoffs available. There are currently two important misconceptions in the industry:

A) that with digital voice, the co-existence of telephony and data services in the same spectrum is a natural consequence;

B) that wider is better in the sense that efficiency grows significantly with increased bandwidth.

The co-existence fallacy is in not recognizing that all bits are not created equal. Voice bits must be delivered with a common quality of service and with minimum latency for all users, no matter where they are located (near or far from a base station, in a shadowed or in a clear location). This requires allocation of disproportionally more resources for disadvantaged users, both in the base station and in the portable handset.

With data, on the other hand, variable latencies, requirements and resources are the norm. Being able to offer variable levels of service by advanced network measurement and allocation techniques, without overly penalizing the weakest users, we can provide overall data throughputs which are more than triple the total throughput for voice. But co-mingling the voice and data service prevents us from employing such techniques and thus from achieving such improvement multiples.

The second belief, that increased bandwidth makes for increased efficiency, is also overestimated. The current standard bandwidth occupancy for CDMA telephony (IS-95) is about 1.5MHz, based on a 9.6 Kbit/sec voice bit rate and a spreading factor of 128, with a resulting coded clock rate of 1.2288 MHz. The proposed “Third Generation” clock rates are either three times this number, 3.6864MHz, or alternatively 3.84MHz, both of which fit in a 5MHz allocation. We refer to this as tripling the bandwidth. The only advantage in efficiency can be that which is due to increased trunking efficiency of a larger accessing population. This may account for 5% to 10% more than the tripling afforded naturally by the bandwidth expansion. This small advantage may be more than offset by the loss in flexibility when having to allocate bandwidth in 5MHz rather than 1.5MHz segments. Furthermore, if this wider bandwidth is shared between voice and higher speed data, even this small advantage is illusory. Data service, as just noted, will require a variety of data rates depending on the user’s needs, location and resources, so trunking efficiency for such a diverse population is hardly a meaningful measure, especially if, as noted, throughput can be tripled through alternate service assignment protocols.

Which leads us finally to consideration of the means for enhancing throughput of data-only services. We need to recognize, first of all, that the most demanding service application is Web browsing and downloading data from the Internet. Thus the heavily loaded direction is the forward one, to the user terminal from the base station. The reverse direction, from the user to the base station, consists primarily of point-and-click commands. But even the uploading of long files, such as images or long reports, will be far less sensitive to latency than for downloading data to the user. Thus the need for the much higher throughput is not as significant for the reverse direction. In this scenario, by judicious management of latency, resources and requirements, along with some technological improvement, forward throughput can be increased by a factor of three to four times, resulting in average throughput in excess of 600 Kbits/sec and peak data rates of 2.4 Mbits/sec, all within the current CDMA bandwidth. An obvious question then is why triple the bandwidth if we can more than triple the downloading throughput in the current bandwidth allocation?

In summary, just as a decade ago CDMA was proposed against mainstream opinion, and it has since prevailed, so today the common wisdom of voice-and-data coexistence coupled with wider spreading requirements is being questioned. In contrast, we propose that in tripling the bandwidth allocation, rather than supporting co-mingled voice and data in the entire band, we employ only the current bandwidth spreading for a data-only service, which can be tripled in throughput by techniques applicable only to data, while saving the other two-thirds of the tripled bandwidth for voice-only service which is likely to remain a major requirement for some time to come. When data becomes dominant, the voice service can flexibly be reduced back to make room for the increased data requirements. This concept which has been labeled High Data Rate CDMA (HDR/CDMA) is a major R&D program of my company with a pilot mobile data network to be publicly demonstrated this fall on the tenth anniversary of our first demonstration of CDMA technology, which happens to nearly coincide with the tenth anniversary of WINLAB.