Next Generation Internet and 4G Wireless

Lessons for Next Generation Internet we should learn from the wireless experience.

Peter Chapman, *Member, IET* and Mihaela Ulieru[1], Senior Member IEEE

***Abstract*—We believe the clamor to throw away our well managed stable and reliable Time Division Multiplexing (TDM) networks in favor of collision based networks, meaning code division multiplexing in the case of wireless and unmanaged packets in the case of wireline, is a serious mistake. Before we finally throw away a network that works so well, we are researching the use of TDM for wireline and wireless Internet traffic as means of providing Next Generation Internet. We examine the experience of convergence and migration as occurred from 2G to 3G wireless and identify some of the lessons learned with the hope of avoiding the associated pitfalls when defining the needs of next generation Internet. We also note that treating conversational voice as just another digital bit stream that can be mixed and communicated with others, disregards some of the human psychological requirements of conversational voice.**

We conclude that in defining the architecture for next generation Internet, a managed TDM system, may offer advantages over the unmanaged packet system we have today.

***Index Terms*—Next Generation Internet, 3G Wireless, TDM Internet.**

MOTTO:

"For a successful technology, reality must take precedence over public relations, for Nature cannot be fooled."

Richard Feynman

1. INTRODUCTION

People need to communicate. The ability to communicate in all its many forms is a key element in our rich society. Despite all the doom and gloom that the telecommunications industry has suffered in the last few years, communication is still one of the most basic human endeavours for which people are willing to pay. However, we must not lose sight of the fact that we need viable business models that satisfy users and providers of the services that enable communication. Equitable business models supported by viable technology are the key to our future.

As an industry we have allowed technological wizardry, public relations hyperbole and avoidance of rigour take communications technologies down a path which, though exciting in its promise, is in danger of being the victim of its own success. We are in danger of disregarding the scientific discipline required to understand the fundamentals.

How did we get here? What can we learn? Most importantly, how can we learn from the lessons of history and thus prevent repeating our mistakes?

1.1. Communicating Digital Information

Information that is to be communicated can

[1] Canada Research Chair in Adaptive Information Infrastructures for the eSociety

usually, easily be converted into bits. However this does not automatically mean that these bits can be equally handled in the same way or can comfortably share a communication medium. In this context communication includes:

- Speech
- Text
- Video
- File Transfer
- Picture transfer
- Music sourcing
- On line processing (software as a service)
- Remote management
- Remote monitoring
- Entertainment distribution
- Purchasing
- Browsing (window shopping)
- Booking
- News distribution

With today's technology, information that needs to be communicated or stored can and is typically converted to an ordered pattern of bits. Following from this is a view that information bit patterns can be mixed during transport and storage because to the storage or transport medium they appear indistinguishable from each other. However, depending on the purpose of the information being communicated, the requirements of the transport and storage medium differ. For example, speech can tolerate errors in the least significant bits of the speech information, a fact that has been very effectively exploited in the GSM standard. Similarly in video, some loss is acceptable in the low significant bits and this has been exploited in JPEG and MPEG compression technologies. Other information carried digitally typically cannot tolerate any errors. At the same time, conversational speech, (as distinct from voice mail leaving or retrieving), or video conferencing as examples cannot tolerate delay while other communications typically can. For this reason different communications with different characteristics cannot comfortably share an unmanaged communications medium. Characteristics beyond the fact that they are bit patterns need to be considered and they need to be managed accordingly by the communications medium. The human factors characteristics, such as delay and error sensitivity of the information the bit patterns are carrying is an important characteristic that the medium needs to consider in managing the information being communicated.

Digital traffic to be communicated comes in many forms, streams and blocks being two principal examples. Streams are typified by voice communication and some real time entertainment feeds. Blocks are typified by files, browser information, in fact any transaction in which the total amount of bits is known prior to it being presented to the network. Knowing the size of a block to be transmitted has the advantage that we can advise the network of the total number of bits in the transaction and it can allocate resources accordingly. This is a valuable attribute that is being disregarded by most network interactions particularly in the clamour to make every transaction packet based. Information that exists in a stored block can be presented to the network at a rate and in a form that suits the network, when and if the network can handle it. Today's thinking on collision based packet networks is to present the information in the hope that the network is sufficiently empty for it to be satisfactorily delivered.

Today's networks work well in carrying a mix of Internet and Time Division Multiplexed (TDM) traffic, for what they are being asked to do – and one can easily argue that they are only being asked to do what they can do well given that users abandon applications that are unsatisfactory. Which is cause and which is effect?

In the future the demand for critical control applications to communicate will dramatically increase [5]. Today's Internet does not provide the resilience, reliability, predictability, security and priority for critical applications, which include management of:

- power,
- water,
- food distribution,
- waste management,
- defence
- law enforcement
- coordinated air traffic control

1.2 Beyond packet switching

Packet switching has proved adequate for the current Internet which is still dominated by information entertainment and non critical activities. However packet switching is not suitable for critical activities due to the inability of the network operator to manage traffic based on its characteristics and to prioritize accordingly.

There are many parallels between the Public Switched Telecommunication Network (PSTN), based on Time Division Multiplexed (TDM) technology and the Global System for Mobile (GSM) cellular network, which we will examine. We use the term TDM to describe the traditional digital PSTN, which is often, somewhat misleadingly, described as "circuit switched". We make this distinction because it is circuit switching, meaning requirement for end to end allocation of actual or virtual circuits that created one of the limitations on the network. It is not the fundamental Time Division Multiplexing and associated Time Slot Interchange switching technology that limited the use of the network and therefore a network implementing TDM transport does not suffer the limitations of a circuit switched network.

It is instructive to examine the GSM and PSTN as successful networks and the reasons for their success. GSM uses a combination of Frequency Division Multiplex (FDM) and Time Division Multiplex (TDM). An examination of the PSTN network and the GSM cellular network provides the following parallels:

- Network designed for one primary service, real time speech
- Technology optimized for user satisfaction (delay, distortion, ring back)
- Technology choice optimized for revenue generation

In both cases, attempts to carry non voice traffic over the network have proved unsatisfactory [1]. Both Internet Protocol and Third Generation (3G) wireless have adopted technologies which we describe as unmanaged collision systems, displacing highly managed admission systems, GSM and TDM.

1.3 Unmanaged collision systems

We define unmanaged collision systems as systems to which traffic is presented without the network to which it is presented, having any prior allocation or knowledge of the traffic and having no knowledge of the characteristics of the traffic. Ethernet is an example of an unmanaged collision system as are most variants of packet switching, notably Internet Protocol. Code Division Multiple Access (CDMA) and its wireless version - WCDMA can also be regarded as unmanaged collision systems. A characteristic of collision systems is that all users occupy the same communication medium, Fig 1. Every user affects the performance of all other users using the medium. Managed systems on the other hand allocate a separate channel for each user, which controls each user's performance independently and creates an inherently stable system, Fig. 2.

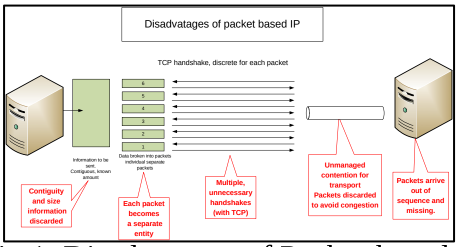


Fig 1: Disadvantage of Packet-based IP

As stated earlier, the reasons for the success of the Internet are not, inherently due to the use of packet switching, but, rather, due to the removal of restriction on, what can be communicated, the definition of a single interface standard and on the separation in real time of the two ends of the communication. In Section 2 we make the argument that an Internet not only can be created on a TDM network, but that such an Internet would provide a number of advantages over today's Internet for the types of services and uses being envisaged for the future.

1.4 2G to 3G: Managed admission to unmanaged collision

CDMA systems, including WCDMA, can be regarded as unmanaged collision systems in that they admit traffic up to a level when error rates become acceptable and then disable further new traffic. Although operators typically set a maximum number of connections based on experience, it is still an

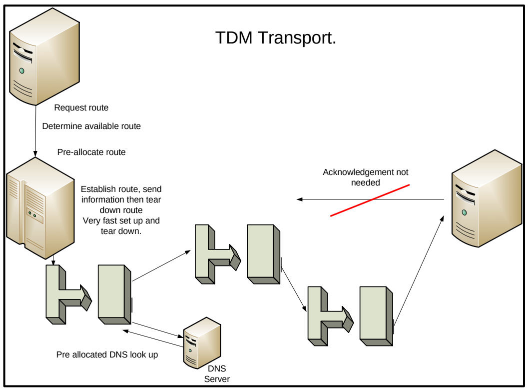


Fig 2: TDM Transport

unmanaged system and depends for its capacity on the statistical pauses in conversation when handsets are not transmitting to achieve the intracell interference performance. This means that statistically, the interference threshold can be exceeded for short periods, which has disastrous effects on non voice communication.

In the case of the move from 2G to 3G, in moving from FDM/TDM to WCDMA we have again a move from a managed admission system to an unmanaged collision system. It became necessary to overlay on the unmanaged collision system (WCDMA) a managed TDM system in High Speed Downlink Packet Access (HSDPA) and its planned derivatives - High Speed Packet Access (HSPA) - in order to achieve satisfactory results.

1.5 CDMA and spread spectrum.

WCDMA and CDMA should not be confused with spread spectrum. Whereas it is possible to use a code to spread a signal so that a signal to noise advantage can be gained at the expense of occupied bandwidth, the theory and research on this technique assumes that the signal is resident alongside Additive White Gaussian Noise (AWGN). Spread spectrum uses a pseudo random code to spread the signal and the benefits only apply within uncorrelated noise such as AWGN.

In the case of CDMA, a similar spreading code technique is used with the code used to separate the signal at the receiver. However, this is done for many signals within the same spectrum, broadcasting at the same frequency and with identical pulse shape characteristics. The codes are chosen to be orthogonal to ensure that each code does not create an identifiable signal for receivers other than the intended recipient. This means that each spread signal does not derive the signal to noise benefits equivalent to those buried in AWGN. Instead they are all competing with each other for the attention of the receiving station which has to separate them from each other. This is a difficult task and is one of the reasons for the poorer than theoretical performance of WCDMA. Multi user detection is one of the biggest challenges for CDMA technology. It is particularly difficult where there is no synchronizing signal between senders and recipient. Many of the popular texts relating to CDMA technology trivialize the difficulty of multi user detection and analyze a received clean digital signal, disregarding the difficulty of achieving such a clean digital signal, as per e.g. [6].

"All the WCDMA properties come rather "through the back door" by the wideband properties of the signals when examined at the system level, rather than the level on an individual radio link."

1.6 Stability in shared systems

Shared systems are those in which a communication medium is shared among a number of independent users. The PSTN is a shared system but is highly organized and managed. The only way users affect each other is when the system reaches its maximum capacity and further users are denied access. The same is true of the GSM system which uses a finite number of discrete channels. The Internet and CDMA both come into the category of collision systems because traffic from one user actively interferes with traffic from another. The capacity of the network is limited by performance and suffers degradation as the number of users or the traffic generated by each user increases. It is widely believed that allowing the transport mechanism to expand until the user experience degrades, enables more capacity to be extracted from the system. This is not necessarily true due to the unmanaged nature of the traffic. It also means that error performance increases dramatically as the system moves into its capacity limited stage. Not only does user experience suffer much slower response, due to re-sends, but stability of the system also suffers. It works like this: as capacity reaches the point where queue lengths become unacceptable, packets are dropped. This causes a TCP time out which generates a packet re-send. This increase in traffic contributes to the queue length which results in more dropped packets and more re-sends. TCP has mechanisms for detecting this effect and backing off or reducing the rate at which it presents packets to the network, but it is a retroactive activity, meaning it is initiated after a problem is detected.

1.7 The problem of competing error recovery between radio links and TCP/IP.

In the event of errors occurring across the radio link, the radio link arranges a resend and corrects the errors. The delay created in the network by this radio link resend causes TCP to detect the absence of an acknowledgement within the specified time and initiates a re-send of the unacknowledged packet. As a result network capacity, and, in particular, valuable radio resource, is wasted by these duplicated resends. There are other more subtle and complex problems due to the different time cycles of the resends which we have investigated as part of a study of an equipment manufacturer that was experiencing performance issues due to this problem. The solution that was found was to modify certain TCP parameters for exchanges carried out across radio links.

1.8 CDMA and WCDMA cell stability

Voice communication is inherently stable, even in a CDMA cell. Real time human communication, typified by conversational voice, is inherently stable because the communicators clear the call and redial in the event of degrading voice performance. Non-voice (machine to machine) communication is inherently unstable due to the simple process of resends of corrupted data creating more intracellular traffic which contributes to the error rate of all other users in the cell, which requests more requests for re-sends. A simple unstable positive feedback loop is readily established.

Given the difficulty of mixing voice and non voice communication due to the very different requirements, it is difficult to know why it was considered necessary to combine them in the same standard, same interface and handset. By providing a managed interface, where the handset and base station negotiate capacity before the transaction commences, a far superior level of service can be achieved. For real time services, this will almost always be true.

2. On TDM, Circuit Switching and Packet Switching.

2.1. TDM

Today's PSTN is a TDM system as distinct from the "circuit switched" network it supersedes. The term 'circuit switched' was originally coined to describe an end to end electrical connection established through electromechanical switches, in such a manner as a connection was established to enable real time communication between two ends with no buffering of traffic along the way. With the introduction of digital switching, and in particular, Time Slot Interchange switching, the term continued to be used even though actual circuits were not established. In fact, a factor in converting the analogue network to digital was to provide digital switching using Time Slot Interchange, which required buffering of traffic by small amounts to enable exchange between incoming and outgoing time slots. TDM requires a disconnection of the two ends, albeit for, by today's standards, very short periods.

End to end connection is not a prerequisite of TDM. All that is required in order to achieve Time Division Multiplexing is for the recipient address to be determined by the position of a, usually, fixed size information element within a TDM stream. TDM elements are interlaced with each other, this way very predictable real time performance and very low delay is achieved. (We avoid the term latency to avoid confusion. The term latency was originally introduced to mean the time taken to fill an information, packet or time slot, before that element could be accepted by the network. Today it is all too often used to mean delay).

2.2. Packet Switching

Packet switching is the method by which the destination of a sequence of information is defined by the initial sequence of information in the stream. Compare this with TDM, where the destination is pre assigned and then defined by the position of a finite sized sequence within the stream. Packet switching by its nature does not easily support interlacing, whereas TDM does. Packet switching does not provide predictable real time performance, whereas TDM does.

However the most significant difference between packet switching and TDM is that in a packet switching system every user is in conflict for resources with every other user while in a TDM system, dedicated resources are assigned to each user.

2.3 Success of the Internet

No one will deny the outstanding success of the Internet. From a system for transferring files between research labs it has grown to the worldwide communication medium that has become indispensable to our daily lives. As mentioned earlier the success of the Internet is often credited to the choice of packet switching as the transport methodology. This is not valid. The Internet could have succeeded with other transport architectures.

Rather than owing its success to packet switching, the Internet was successful because of the fixed interface and store and forward capability provided by the network. While it is interesting to note that packet switching forced the introduction of store and forward and it was this store and forward that was the major contributor to its success, in fact a TDM based store and forward system would have provided all of these benefits in addition to superior performance.

2.4 Statistical multiplexing.

Statistical multiplexing is the means by which more virtual circuits or capacity is allocated than the physical structure is designed to carry. It depends on the fact that during typical transactions there are times when no information is being sent. In particular, during a voice call, most of the time the conversation is in one direction only, the capacity of the allocated link in the opposite direction can then be occupied by other traffic. It is important to realize that such technique, although generally associated with packet switching, is not unique to packet switching. Statistical multiplexing was originally introduced on expensive long haul voice circuits, such as submarine cables, within a TDM environment. Statistical multiplexing introduces variable delay, and, unless there are good reasons for implementing it, it is best avoided in real time communications. In the case of non voice or non real time communication, there are few benefits as the sending equipment typically has a store of information waiting to be sent and the rate at which it is sent is controlled by the network. Managed scheduling on a “fair share” basis is preferred to the “free for all” collision philosophy.

It is widely believed that collision systems provide some form of statistical multiplexing and that this is desirable and needed. In an unmanaged network the traffic level is allowed to grow until performance degradation either limits it by users clearing and re-establishing the communication or because the operator has put certain limits on it. This prompts two questions: firstly, is the capacity of collision systems greater and secondly, are today’s networks capacity limited? If the first of these, capacity of collision systems exceeding that of managed systems, is not true, then we should not be introducing collision systems. If the second, today’s networks are not capacity limited, is true, we should not be doing it because there is no need. Either way, introducing collision systems in the hope of increasing capacity would seem to be a mistake. One reason to use a collision system is to obviate the need to create mechanisms to manage traffic.

3. Real time requirements of conversational voice

Human communication is most often carried out using conversational (two way) voice. In a two way conversation, a certain level of unintelligibility is allowed because the other party in the conversation can ask for an immediate repeat. In non conversational voice, for example when creating or retrieving a voice mail message, there is no opportunity for such a repeat, so intelligibility must be very good, though delay is unimportant. On the other hand, conversational voice has very exacting requirements for delay. If the delay exceeds certain limits conversation becomes impossible and the communication becomes one way, akin to “push to talk”. This manifests itself as, in one respect, people have short duration conversations. It also has the effect of one person sensing that they are being “talked over”, when in fact they are not. This tends to make the conversation more hostile than intended and is particularly noticeable in heated discussion or arguments. Voice traffic can therefore be categorized into two very different types with divergent characteristics:

- Conversational voice: low delay, but tolerates errors
- One way voice: delay immaterial, but intolerant of errors

Increased delay meets the business needs of cellular operators, where often cell capacity is a limiting factor, so short duration calls are encouraged. However, for long distance operators, who generate revenue based on the duration of a long distance call, it is far from ideal. In fact one of the drivers towards a low delay digital network was to provide a speech experience that would encourage long duration calls as a means of increasing revenue. Designing wireline communication systems based on experience of acceptable delay on a cellular network is a mistake. Many communication systems are built around intelligibility and not around delay, resulting in a poor user experience. Designers of systems have been known to remark "it sounds good enough to me" having listened to a one way voice transmission, with little appreciation of the subtlety of human psychology in voice communication.

4. The Wrong Mutation (Why did Modems and ISDN fail and the Internet succeed?)

4.1 Modems

In order to communicate data over voice lines, the original technology was to make the data emulate the characteristics of voice, in terms of limited bandwidth, and so the modem was created. The problem with early modem technology was two fold. Firstly there was no universal standard, though many attempts were made to establish a number of standards. These standards resulted in an unfriendly user interface. The user had to know the settings of the recipient modem, such as parity settings etc. There are 36 commands in the basic Hayes command set alone. There are numerous standards. In addition the recipient modem needed to be connected, powered and ready, prior to the exchange of information. The result was marginal success combined with a lot of user frustration. It remains a mystery to this day why some of these issues did not become resolved in the standards, or even why there were ever user options on the technical settings needed. The modems seemed to have been specified for the engineers, not end users. The second problem was that, because of the way the network was created for voice traffic, it was necessary for two modems to both be connected in real time and to establish a handshake before communication could commence. Compare this with today's Internet, where the communication is managed through the standardized TCP/IP settings, of which the user generally requires no knowledge.

4.2 The Integrated Services Digital Network - ISDN

The Integrated Services Digital Network (ISDN) achieved little success, primarily because of the requirement for the receiving device to be online and ready. Had file exchange and browsers been available at the time of the introduction of ISDN, it is likely it would have been far more successful than it actually was. The creation of a store and forward ISDN network would potentially have produced a successful Internet. If the requirement had been recognized earlier and the then current TDM system been converted to a store and forward system for non real time traffic, and remained as is for real time traffic - we would today have an Internet that supports both real time and store and forward services.

Instead we have an Internet that only provides non real time services and we are attempting to carry real time services, such as voice, over the Internet. We believe that voice over Internet Protocol is today only viable because most of the transport links have far more capacity than traffic they are being asked to carry. Attempts to prioritize traffic and manage flows separately and thereby create a needed QoS are unwieldy and unlikely to be successful because the basic fabric of the Internet has never been designed for them. In order to create Internet architecture for the future we need to revisit the conflicting demands of real time and non real time traffic rather than try to carry real time traffic over an infrastructure that cannot support it.

We need to assess what we are asking the next generation Internet to do. What today's packet switched Internet has given us is a perspective of what can be done with a network. We need to create the next generation to do it.

5. An examination of GSM (2G) to UMTS (3G) as a salient lesson in how to NOT migrate the Internet

In migrating from 2G (GSM and others) to 3G (UMTS and others), a number of mistakes were made which resulted in disappointment in 3G both for the end user and for the service providers.

Second Generation wireless (2G) - characterized in its most popular form as GSM, was designed for and provided a superb voice service. It met all of its design aims, primarily:

- Good capacity
- Good voice performance
- Worldwide roaming (subject to frequency allocations)
- Good handset performance (weight, price, battery life)

The GSM radio system provides an excellent compromise between error correction for the most significant bits and the disregarding of errors for the least significant bits. This works well for voice but does not translate to non voice applications. As a result, despite the modest success of GPRS on the GSM network, the system was not really expandable to non voice applications, for which it was never intended.

As a result the International Telecommunications Union (ITU) created the requirement for a third generation service. Third generation was intended to overcome some of the drawbacks of GSM in applying non voice applications. It failed. Let us now examine the reasons for the failure. Firstly the original requirements:

- Universal roaming based on a worldwide standard and frequencies.
- A mix of voice and non voice services
- Availability of wide bandwidth services

Whereas these were worthy intents, it was not appreciated how difficult it was to provide a single radio interface to meet these needs.

Let us now examine these in detail together with the rationale for each of them.

5.1 Universal roaming

This was an excellent and obvious objective which failed for two reasons. Different standards and licensing authorities could not agree on either the radio standard or the frequencies to be used ITU: IMT-2000 specifies six incompatible radio interfaces:

- W-CDMA,
- CDMA2000,
- TD-CDMA / TD-SCDMA,
- UWC/EDGE,
- DECT,
- Mobile WiMAX,

The reasons for these is beyond the scope of this paper, but a contributing factor was that there was no agreement on a radio standard that could be considered a clear winning candidate either from a technical or a business perspective.

The inability to agree on frequencies was largely a political and administrative issue.

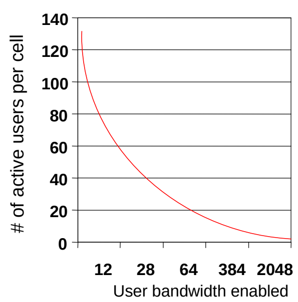
5.2 Mix of voice and non voice services.

It is not clear why it was chosen to mix voice and non voice in the same radio interface and spectrum given the very different requirements and difficulty of achieving this. A lack of understanding of the very different requirements and the view that all digital signals can be considered to have similar requirements is a possible reason. This view permeates much of the thinking today and results in inappropriate use of resource.

As was noted earlier, conversational voice has very exacting requirements in terms of delay, but less so in terms of errors in the less significant bits. This is quite different from most other digital information, which can generally tolerate no errors but some delay.

To mix these two types of traffic, can be seen in retrospect, to introduce unnecessary difficulties.

It is worth examining the impact of offering wideband services in contention for capacity with voice. If the cell is providing wideband to only a few users, then it impacts many low bandwidth users, requiring voice and other services. The service provider can generate more revenue from lots of low bandwidth services than for a few high bandwidth services, so a few high bandwidth services creates technical difficulties and occupies an excessive amount of available spectrum. This impact of wideband users on narrow band users is frequently underestimated. Despite the standards defining many virtual channels, we must remember that in a CDMA cell there is only one radio channel and virtual channels are all digitally multiplexed onto the one channel and every handset has to unscramble the channels and the Base Transceiver Station (BTS) has to unscramble all the signals from all the handsets simultaneously in order to retrieve the multi user signals. Frequently, diagrams of clean digital patterns with well behaved orthogonal signals are used to explain the separation of the signals. Unfortunately, the real world is much more complex.



Estimated single cell capacity in a WCDMA/UMTS network.

Revenue from high bandwidth users does not compensate for lost revenue from voice users.

Operators revenue is optimised at many low bandwidth users, not a few high bandwidth users.

Revenue /MB

Cable TV	$0.0001
Wired Phone	$0.0800
Mobile Phone	$3.0000
SMS	$3000.0000

Figure 3: WCDMA single cell capacity

Figure 3, derived from a study of capacity carried out by the first author, while employed at a company producing UMTS wireless equipment, shows the single cell capacity of a WCDMA cell. As can be seen, the capacity is severely limited as soon as a few wideband users are admitted into the cell. What is equally significant is the chargeable revenue per bit based on wideband or narrow band use. The price per bit, representing the value to the end user per bit, is considerably greater in the case of low bandwidth services than high bandwidth services. For the operator, the result of admitting high bandwidth users is undesirable for generating revenue from the use of the spectrum.

CDMA and WCDMA are regarded as having the advantage that if there is no activity to or from the handset, then it does not occupy any capacity. By automatically detecting activity, in effect statistical multiplexing occurs thereby allowing a greater number of users in the same spectrum. In practise this is not as advantageous as it may seem when mixing voice and data traffic. Real time (conversational) voice does exhibit some of the indeterminacy and gaps that render it applicable for statistical multiplexing. As well voice tends to be one directional only, again facilitating statistical multiplexing by not sending in the other direction. However, in order to render voice acceptable to users it is necessary to send a low data rate background noise signal known as “comfort noise”. If this is not done the user believes the connection has been lost and typically questions the other party as to whether they are still listening. In the case of non voice activity, generally information is buffered in the sending device and as such will be sent as fast as the network allows, so the benefits of statistical multiplexing do not apply. In the case of transactions such as web browsing which is highly interactive, the average throughput is very low, typically around 10kbps and it is buffered. In fact, the average throughput is about the same as that of compressed voice communication but it has completely different characteristics.

6. The myth of spectral efficiency

Claude Shannon [4] in his seminal paper derived the formula by which the information carrying capacity of a communication channel is related to the signal to noise performance of that channel. Shannon’s formula applies to all communication channels to this day and techniques promising one technology improved spectral efficiency over another technique are largely illusionary. Shannon's theory is a mathematical theory and is independent of any particular implementation and is fundamental to all real world communications,

The information carrying capacity of a channel is dictated by its noise performance, whichever way it is used to share among users. Decisions on which multiplexing technique to use should therefore be made on other characteristics and not on dubious claims of spectral efficiency. Characteristics to be considered should include:

- End user performance
- Usability by the service provider
- Battery performance of the handset
- Price of the handset
- Acceptability in all territories

A number of researchers have compared the capacity of CDMA and WCDMA cells with TDMA cells with what are considered surprising results [2], [3].

> *…for moderate to high user densities, the BER performance degradation due to interuser interference is substantially higher in CDMA, even when error correction coding is used. For small user densities, the interference power is much less than the thermal noise power, in which case TDMA and CDMA have a similar BER performance. These observations were confirmed by our numerical results computed for AWGN channels and flat Rayleigh fading channels.*

Further, [7] states:

> *…analyzing cellular TDMA and CDMA with the same maximum capacity and bandwidth occupancy, the present authors found that in a cellular network with hexagonal cells, assuming that signal attenuation is proportional to the fourth power of the propagation distance, multiple access interference is approximately 10 dB higher in CDMA than in TDMA when the cells are fully loaded. In turn, this means that CDMA must virtually use 10 times more bandwidth than TDMA to achieve the same interference level, which may seem surprising given the common perception that CDMA has substantially higher capacity than TDMA*

As well as spectral efficiency there are other problems with WCDMA. In principle, making the Base Transceiver Station BTS less complex while making the handset more complex is not a good idea. Each BTS serves potentially several thousand handsets. A small complexity increase in the handset multiplies accordingly. The best engineering for cost is to make the handsets as cheap as possible at the expense of the BTS. In addition handsets are very sensitive to battery power consumption and forcing the handset to continuously unscramble CDMA code in the event that a message might be for them is not good engineering. It is better to use a TDM system so that the handset only needs to listen in its allocated time slot, or to a periodic pilot, to determine when it needs to listen for information addressed to it.

A parallel with the creation of television is as follows. In the early days of television, the standards were written to maintain as much simplicity at the TV receiver at the expense of complication at the transmitter, it being well understood that one transmitter could serve potentially millions of consumers. The equation made a lot of sense, the result of which is that we now have a "gamma" non linearity written into the standard, which was incorporated in order to make monochrome TV receivers with Cathode Ray Tube (CRT) displays as simple as possible. Interlace was introduced for similar reasons.

Applying the same logic to the wireless industry we should be designing our standards to simplify handset design, not just for cost, but for low battery consumption, small size, weight etc. WCDMA, by requiring a single transmitter and receiver for multiple users, enables a lower cost BTS, but requires a more costly handset than a frequency division or time division system. Given that a BTS typically services several thousand handsets, this is not a cost effective trade off.

6.3 Unsuitable technology choices – what we have learned from the 3G experience.

The choice of WCDMA for UMTS can be considered one of the poorest technology choices ever made [6]. The simultaneous channel sharing capacity of WCDMA has largely been swept aside by the introduction of HSDPA, which can be regarded as a time sharing (TDM) technique. Why this happened makes an interesting study. Contributing factors to the poor choice are optimism on WCDMA capability, poorly researched engineering, commercial or vested interests, with little of the interest of the consumer, the effective use of valuable spectrum, or even the business interests of the service providers sufficiently well researched or considered. It is interesting to note that licensing authorities, when issuing spectrum licenses, included a clause allowing the purchasers of those licenses to apply for use of a different technology if they could demonstrate that they could not create a viable business using the specified technology. This is an indication of the recognition of the risks associated with the uncertainty around the then untested technology.

The current system of HSDPA overlaid on WCDMA is similarly wrong. Given that often the whole base station capacity is given to each handset in time slots, it is difficult to understand the benefit of WCDMA. It seems an unnecessarily complex addressing mode, requiring all handsets to decode all signals continuously. The same result can be achieved by scheduling time slots, based on radio link quality or readiness of the information and indicating this to the handset through a management channel rather than require every handset to continuously descramble all information. The use of WCDMA as an addressing mode is now largely the result of legacy rather than good design. WCDMA scrambling and descrambling can be viewed as achieving no more than addressing the recipient handset while avoiding the need to communicate the scheduling of the wireless resource to the handset.

There are a number of lessons to be learned from the 3G wireless disaster, which will avoid making the same mistakes with next generation Internet. These mistakes can be summarized as follows:

- Choice of technology must be able to support services based on user needs.
- Choice of technology must be able to service lots of users rather than enable a few high bandwidth users at the expense of the majority.

The literature around 3G wireless is littered with phrases like “speeds of up to x Mbps” instead of “x users can simultaneously achieve consistent speeds of xx in each cell”, or even “each cell can simultaneously provide service to xx users” which would be far more meaningful for the service provider and the user.

There was a lack of understanding of two of the fundamentals that our industry needs to remember:

- What does the end user want to do that they will be willing to pay for?
- Can the operator make money?

The choice of WCDMA would seem to have more to do with providing high bandwidth to a few users with scant disregard for the traditional services that wireless operators have been providing. It has been replaced with HSDPA. Had the requirement been to create a burst mode high speed packet service, with the cell allocated to a few high speed users in very short bursts, as happens with HSDPA, then it is unlikely WCDMA would have been chosen as the addressing methodology.

8. TDM as a better alternative for next generation Internet.

As discussed earlier, the benefits of the current Internet came not from the perceived benefits of packet switching, but from disengaging the two ends of a transaction, thereby forcing the network to become a large store and forward facilitator. In addition, the defining of an interface, IP, that is completely invisible to the end user, created a level of user friendliness unseen in previous attempts.

Figure 1 describes some of the challenges of conflicting traffic while figure 2 describes specific problems associated with IP as well as problems introduced by TCP - which was created to overcome some of the basic IP difficulties. Figure 2 explains in very simplified terms how a TDM system will address the challenge.

Given that we now have a good indication of what makes the Internet succeed, in looking at next generation services, we can use this knowledge to create an Internet architecture that avoids some of the limitations of the present packet switched network. Such as:

- No real time capability
- No network-created delivery guarantee
- Difficult for the network to provide originator authentication
- No broadcast and multicast capability
- No network awareness of traffic requirements.

8.1 Intelligence in the networks or in the terminals?

It has been incorrectly argued that the Internet is a dumb network with all the smarts in the terminals whereas the TDM network utilized dumb terminals with all the smarts in the network. A closer look at the network revels that the present Internet requires a great deal of intelligence in the network, which is contained in routing tables and a sophisticated system for updating routing tables. In addition, the network needs to make judgments as to fragmenting packets, discarding packets, rerouting etc. Compare this with TDM systems whereby a relatively simple out of band protocol establishes a route, while confirming status of the recipient terminal before any information is presented to the network. All that is required to create a viable Internet is to segment the network into buffered TDM zones, thereby removing the requirement for simultaneous connection of both ends and providing network store and forward.

7.2 Proposed solution

Based on these ideas we are proposing a next generation TDM system for transport of all information across the network. Such an architecture will benefit from the advantages of disconnecting end to end, creating an "as needed" network store and forward capability, yet providing real time capability. Giving more responsibility to the network for acceptance of delivery of packets, renders the task of network based authentication much simpler, thereby creating a safer Internet. In proposing network-based authentication we are not proposing any sense of restriction of who can contact whom, but simply that the recipient has the right to know who is "knocking on their door" before admitting them. The reason network-based authentication can be more readily provided on a managed TDM system is that, essentially the sender requests a route to the network, the network then admits it or denies it prior to scheduling it. The authentication is provided prior to admitting the traffic. The requirement to authenticate separate packets is removed, as would be needed with the present packet architecture.

For the purposes of multi cast and broadcast, because the network is managing streams not packets, an individual stream can be stored and redistributed as a stream not as individual packets:

- TDM manages streams not fragments of streams, (packets).
- Information does not come in packets: it comes in streams.

7.3 Self Managing Network

Networks need to be self managing. One of the ways of achieving this is for the source of information to communicate to the network the characteristics of the transaction and the network to allocate and adapt accordingly. This is a challenge for next generation networks, and details of how to implement it need to be researched. The current Internet collision systems have the capability of reacting to changed conditions, not of anticipating and managing for them in advance.

7.4 Proposed approach

At the start of the transaction (Fig. 1), the sending device indicates to the network the type of transaction. Note that, except in the case of conversational voice, video telephony and similar applications, where an indefinite time period two way path is required, for almost all other transactions a finite time, known total capacity one way path is required. Knowing in advance how much information is to be transferred, allocating the path, which will take only a few milliseconds if there is no congestion, and submitting the traffic will generate a far more efficient Internet transport than the current system which sends a packet and then sends another only after it has received an acknowledgement that the first has been received. The actual TCP protocol is more sophisticated than this, in that it anticipates receiving a successful acknowledgement and only backs off the rate at which it submits information after it determines that congestion is occurring. Nevertheless the current TCP principle being used is to submit information assuming it will be delivered. It only works for an empty network which is what we have today.

Our current research is based on the following principles:

- A pre allocated path for DNS look up, which is created and reallocated within a few milliseconds.
- Time slot based hot links.
- A pre allocated path to the next hop time space switch with fast set up and automatic tear down.
- Pre defined paths between all networked time space switches.
- Controlled delay at each time space switch
- Buffering at each time space switch in order to disconnect network regions
- A system of request of path with parameters such as total permitted delay
- Originator authentication of stream and flow.

8 Summary

The unexpected success of the Internet to become a flexible, ubiquitous communications medium has led many to conclude that the current Internet architecture is the only one possible. We disagree with this position. The purpose of this discussion is to address the human needs of the network and ask people to move away from the “it has to be packet” mindset that has become entrenched in our industry. [1], [3]. One argument for this is the perceived inherent statistical multiplexing capability of packet switched networks.

With this in mind we decided to examine the reasons for the success of the packet based Internet and the applications that have succeeded and those that have failed. We concluded that the success of the Internet was not due to the packet based architecture, but rather the freeing of constraints imposed by the hardware implementation and policies by the service provider.

We are proposing TDM Internet architecture because we believe that the network is far from achieving its potential as a communication medium. We also believe that the current trend to remove all time space switching and replace it with packet switching is based on flawed thinking and will result in the replacement of an excellent network with an inferior network.

Arguments in favour of packet switches (routers) being less expensive to create are based on a comparison between the time space technology which is often thirty or forty years old and packet technology which is current. A time space architecture built around current technology can potentially be as inexpensive as current router technology. Any proposed transport architecture must be compatible with and coexist with the current architecture.

Among the services that need to be accommodated in one universal network are:

- Speech
- Text
- Video
- File Transfer
- Picture transfer
- Music sourcing
- On line processing (software as a service)
- Remote management
- Remote monitoring
- Real time control
- Entertainment distribution
- Purchasing
- Browsing (window shopping)
- Booking

- News distribution

Many of these require a level of performance, reliability, security and availability that the current network cannot guarantee to provide. We need to migrate the current network to architecture to provide all of these, with the characteristics that users will value and therefore will generate revenue for the service provider. The current Internet has given a us a glimpse of the potential. It is now up to us as an industry to provide the technology, management and services to make it happen.

REFERENCES

[1] David D. Clark, Karen Sollins, John Wroclawski, Ted Faber, Addressing Reality: An Architectural Response to Real-World Demands on the Evolving Internet, Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture, Karlsruhe, Germany 2003, Pages: 247 - 257

[2] HIKMET SARI Heidi Steendam, Marc Moeneclsey, On the Capacity of Cellular CDMA and TDMA over Nondispersive Channels*

[3] Pablo Molinero-Fernandez and Nick McKeown and Hui Zhang, Is IP going to take over the world (of communications)? ACM SIGCOMM Computer Communication Review, Volume 33 , Issue 1 (January 2003), Pages: 113 - 118 Pages: 113 - 118

[4] C. E. Shannon A Mathematical Theory of communication, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

[5] [SMC 2009] – Special Issue on Engineering Cyber-Physical Ecosystems, IEEE SMC Part A – to appear in 2008.

[6] WCDMA for UMTS, Harri Holma and Antti Toskala, Wiley, 2002.

[7] Hikmet Sari,.Frederik Vanhaverbeke and Marc Moeneclaey, Extending the Capacity of Multiple Access Channels Communications Magazine, IEEE Publication Date: Jan 2000 Volume: 38, Issue: 1, p 74-82 ISSN: 0163-6804