REPORT

- Voice Over Internet Protocol -

Carlos Navarrete
Guillermo Rodríguez

What is VOIP?

VOIP is an acronym for Voice Over IP, or in more common terms phone service over the Internet. It's based on giving the same service as the traditional telephony system but using the Internet to send the voice as packets over the net. This new service implies big advances in communications systems such as cost reduction or videoconference service.

Nowadays VOIP is starting to get some followers that use VOIP in addition to their traditional phone service, since VOIP service providers usually offer lower rates than traditional phone companies. New software based on VOIP is being developed and currently there are many programs available to use for people with a reasonable quality Internet connection. VOIP programs don't offer all the destinations as the traditional telephony but they can reach between 80 and 95 % of them. VOIP also allows, in addition to reaching fixed and cell telephone, to speak with another person that uses a computer with speakers and microphone.

In the last years the evolution of VOIP is increasing exponentially and it is incorporating manufactures such as Cisco, Lucent, Intel, Quicknet, IBM and many other systems that support VOIP.

Data Networks versus Voice Networks

Traditional telephony networks are based on the concept of circuit commutation. This means that for making a communication there must be first a physic circuit establishment. A consequence of this is that during the time the communication lasts, it is not possible to share the resources that is using with another communication, even while nobody is speaking (but the communication is established).

Opposite to this are the data networks that are based on the package commutation concept. This means that for the same communication, different paths between origin and destiny can be followed by the different packages. The resources that are being used by one connection can be used by other connections at the same time. This makes that the data network use in a more efficient way the network resources.

One operator has less cost if it uses data networks instead of voice networks, because it can allow more clients speaking with a data network than with a voice network using the same investment on infrastructure. The disadvantages of the data network is that it has to send the information divided in packages so a connection is composed of several packages and some of this can be lost over the net without arriving the receiver. There is also no guarantee for the packages about the time that it will take them to arrive at the receiver. These problems imply a loss of quality of service. Fortunately, with the new technologies this problem will disappear little by little.

Classic Vocal telephony versus IP based telephony

The classic vocal telephony (***P****ublic* ***S****witched* ***T****elephone* ***N****etwork*) is composed of an access network, that includes cable and the necessary equipment between the house of the subscriber and the local telephone exchange, and a transport network that includes higher ranked telephone exchange and links between them. The communication is made thanks to the circuit commutation.

In the IP vocal telephony the fundamental change is on the transport network. This function is done by an IP protocol based network with package commutation (for example, the Internet). The access network can be the same used with the classic vocal telephony.
The elements needed for the IP vocal telephony depend a lot on what kind of terminals are used to communicate the users. If the terminals are "IP terminal" then elements such as multimedia computers or IP faxes are needed. If the terminals are not "IP terminal", elements as conventional telephones or faxes are needed. In the first case, the terminals send IP packages and in the second case the terminal need an intermediate device to transform the voice into IP packages.

Voice-over-IP Overview

A VOIP system consists of a number of different components: Gateway/Media Gateway, Gatekeeper, Call agent, Media Gateway Controller, Signalling Gateway and a Call manager.

The Gateway converts media provided in one type of network to the format required for another type of network. For example, a Gateway could terminate bearer channels from a switched circuit network and media streams from a packet network.

In VOIP, the digital signal processor (DSP) segments the voice signal into frames and stores them in voice packets. These voice packets are transported using IP in compliance with one of the specifications for transmitting multimedia (voice, video, fax and data) across a network: H.323 (ITU), MGCP (level 3,Bellcore, Cisco, Nortel), MEGACO/H.GCP (IETF), SIP (IETF), T.38 (ITU), SIGTRAN (IETF), Skinny (Cisco) etc.

Coders are used for efficient bandwidth utilization. Different coding techniques for telephony and voice packet are standardized by the ITU-T in its G-series recommendations: G.723.1, G.729, G.729A etc.

There are two main causes for the evolution of the Voice over IP market:

- Lower Cost
- Increased functionality

Lower Cost

In general phone service via VOIP costs less than equivalent service from traditional sources. This is largely a function of traditional phone services either being monopolies or government entities. There are also some cost savings due to using a single network to carry voice and data. This is especially true when users have existing under-utilized network capacity that they can use for VOIP without any additional costs.

In the most extreme case, users see VOIP phone calls (even international) as free. While there is a cost for their Internet service, using VOIP over this service may not involve any extra charges, so the users view the calls as free. There are a number of services (such as Free World Dialup and Skype) that have sprung up to facilitate this type of "free" VOIP call.

Increased Functionality

VOIP makes easy some things that are difficult to impossible with traditional phone networks:

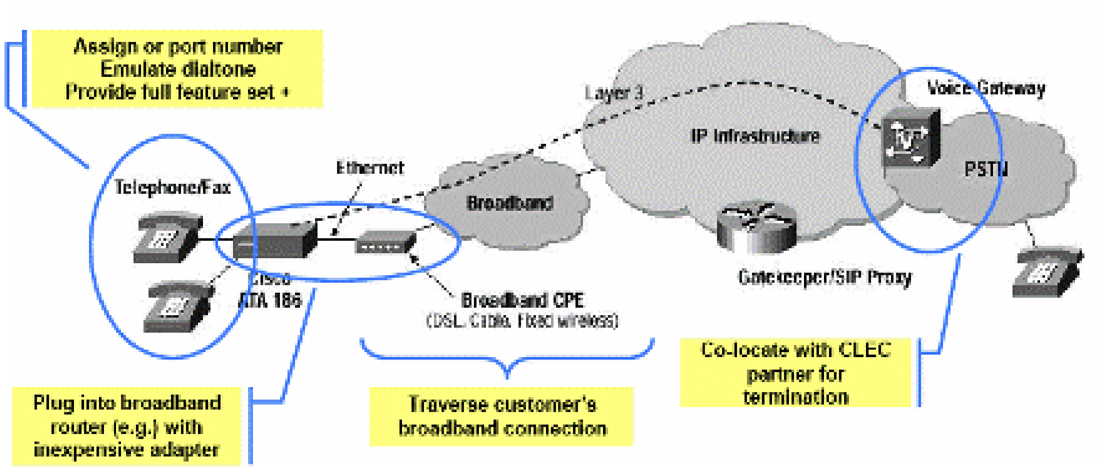
- Incoming phone calls are automatically routed to your VOIP phone where ever you plug it into the network. Take your VOIP phone with you on a trip, and anywhere you connect it to the Internet, you can receive your incoming calls.
- Call center agents using VOIP phones can easily work from anywhere with a good Internet connection.

Gateways and GateKeepers

The Gateway is the element in charge of connecting the classic telephone network with the IP network. It converts analogical signals into IP packages and in the opposite way.
A gateway can communicate one “no IP device” with an “IP device”. It works as a connection between a telephone exchange and an IP network. Sometimes other elements take the function of the gateway, for example Cisco has implemented a system with a router that has also the function of a gateway.
The Gatekeeper acts together with several Gateways and in charge of user’s authentication, bandwidth control, IP routing… is the brain of the IP telephony. One of the main problems for the IP telephony expansion has been that gateways and gatekeepers haven’t been compatible but this is being corrected and most systems work with H.323 protocol.

There are three main different possibilities in using VOIP depending of the type of terminal used:

- Phone calls between two telephones. In this case origin and destiny need to communicate with a gateway. The telephone (A) that is making the call will contact a gateway that will ask a gatekeeper how to get to the destiny telephone (B) and this gives back the destiny’s IP address to the gateway. So now the gateway from the origin transforms the voice from A into IP packages and sends them to the gateway of B, that transforms them into voice again that is what B receives.



- Phone calls between a PC and a telephone. In this case, only one side needs to contact a Gateway. The PC needs an application capable of establishing and maintaining a phone call. We suppose that a computer A is calling a phone B. A needs first the IP address of B's Gateway, so A asks his Gatekeeper. Then the phone application on A establishes data connection through the IP network with B's Gateway that will transform the incoming IP packages into analogical voice for B.
- Phone calls between two PC's. This case is quite different from the other two. Both computers only need to have the same phone application installed and to be connected to an IP network (generally Internet) to make an IP phone call. It's just the same as any other Internet application (for example a chat).

We can see that, in the three cases, the most or the total of the connection across sender and receiver works with data network instead of voice network.
Sometimes the VOIP gateways are referred as ATA (Analogue Telephone Adaptor) or just as TA (Telephone Adaptor) and they usually have an Ethernet jack and a RJ-11 phone jack. They have a little management and administration system and they have an IP address too.
Sometimes Gateways are not necessary in case the phone is an "IP phone" because these include the gateway functionality in their own design.

Bandwidth

The bandwidth needed for sending voice and/or video in real time over the net was quite big some years ago but now, the voice that a Gateway receives is digitalized and compressed with several (only one at a time) algorithms (GSM, G.723.1, G.711, G.729) that manage to obtain bigger compression ratios (instead it takes more time to compress and uncompress the data) so the bandwidth require for real time communications is not so big.

The best compression algorithms manage to compress voice packages in approximately 8 Kbps. The IP protocol adds to the digitalize voice package some headers to transport it over the network so finally the bandwidth necessary to send voice is 16 Kbps.

But it is possible to reduce this bandwidth with the “silence suppression” that avoids sending voice packages when no agent is speaking. Considering this, it is possible to affirm that the average package size for a conversation is 8 kbps.

The VOIP Standard

Internet telephony or VOIP applications from different vendors have been incompatible for some time, due to fundamental differences in voice coding, silence suppression, addressing and dialling plans, call management, and other related functions due to lack of clearly defined VOIP Specification. To address these problems a number of non-profit working groups of industry players have coalesced in the last several years and sets of standards, VOIP Specification or "interoperability guidelines", have been born. These include the International Multimedia Teleconferencing Consortium (IMTC), the Voice over Internet Protocol Forum (VOIP), and the SCSA standard set.

The VOIP Standard was defined by the ITU (International Telecommunications Union) in 1996 and it provides a group of norms to the different manufactures so they can all evolve together as a group.

One of the main advantages of the Standard is that it allows controlling the network traffic so it is less possible that failures happen. Because it works in the same level as the IP it has the following advantages:

- Independence of the physical network.
- Independence of the hardware used.
- It can be implemented in software and/or in hardware.
- It can integrate video and TPV.

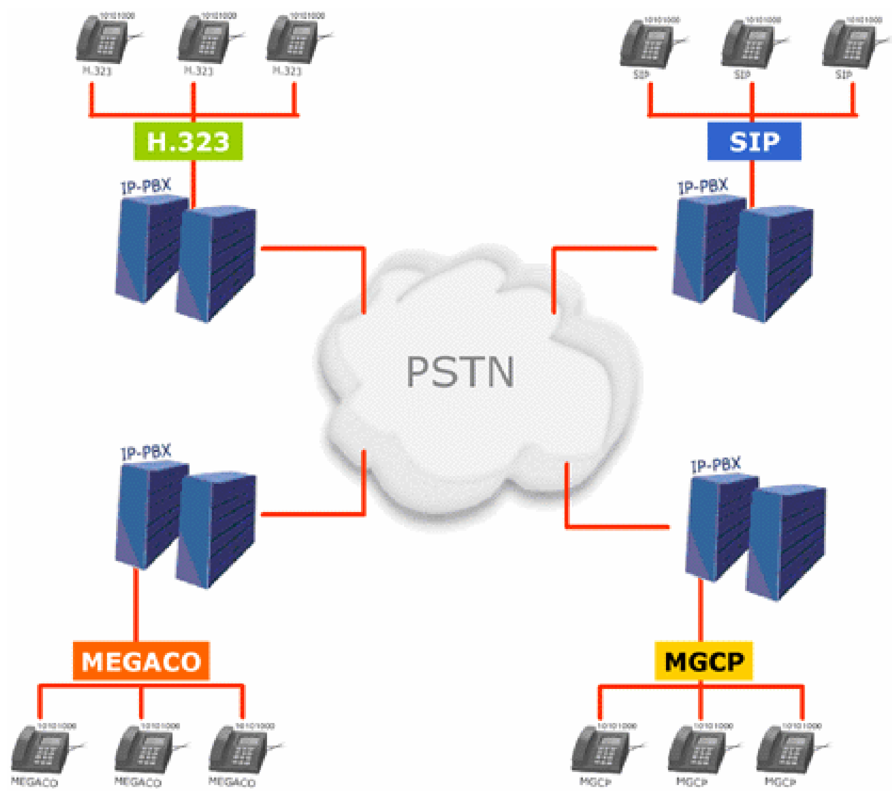
The net architecture is based on three elements: Terminals, Gatekeepers and Gateways (that have been explained before).

The protocols used in VOIP for connecting are very important because they will define the communication efficacy and complexity. The most important protocols are (in chronological order):

- H.323, defined by the ITU-T.
- SIP, defined by the IETF.
- Megaco (also known as H.248) and MGCP that are control protocols.
- Skinny Client Control Protocol, Cisco proprietary protocol.
- MiNet, Mitel proprietary protocol
- CorNet-IP, Siemens proprietary protocol.
- IAX.
- Skype, peer-to-peer proprietary protocol used in the Skype application.

- Jajah peer-to-peer proprietary protocol used in the web-phones Jajah SIP, IAX.
- IAX2.

So different telephones can communicate between themselves using different protocols as the next image shows:



The second version of the H.323 set of ITU-T specifications is currently the predominant and is the implemented Voice over IP standard. According to the latest news, this protocol leads by a 2:1 margin over the second-place Voice over IP protocol, MGCP (the Media Gateway Control Protocol).
But H.323 is fast falling from the top. When asked what Voice over IP standards their new products will support, the vendors prefer H.248/Megaco, MGCP, SIP and SIP+. H.323 support is losing mainly because of its many different, and generally incompatible, versions. While most companies do not view SIP (the Session Initiation Protocol) as directly competitive with MGCP, many regard the new and untested H.248/Megaco specification as directly competitive with H.323 for VOIP equipment interoperability.

Because H.323 is the most used protocol with VOIP, some of its characteristics will be explained. H.323 is a family of standards defined by the ITU for using multimedia communications over LAN networks. It’s defined specially for those technologies that don’t guarantee quality of service. Some examples are Ethernet, Fast Ethernet and Token Ring. The network technology more common to work with H.323 is IP. Actually, H.323 is the first complete specification that manages to work over IP.
The VOIP working with H.323 uses a set of other protocols for different functions:

- Routing and Forwarding:
  - RAS (Registration, Admision and Status). Communication protocol that allows a H.323 station to locate other H.323 station using a Gatekeeper.
  - DNS (Domain Name Service). It does the same as RAS but using a DNS Server instead.
- Signalling:
  - Q.931. Initial signalling of the phone call.
  - H.225. Call control such as admission, signalling, registering and packaging and synchronising the voice stream.
  - H.245. Control protocol to specify opening and closing channels for voice streams.
- Voice compression:
  - Required: G.711 and G.723.
  - Optional: G.728, G.729 and G.722

The G.723.1 standard is emerging as one of several coders for Internet telephony, mostly because Intel and Microsoft are already shipping products that support it. In spite of some drawbacks, such as large latency, VOIP voted to recommend G.723.1 as the default voice coder the IMTC voted to accept the recommendation.

In the following table the codecs and the bandwidth that they require are shown:

VoCodecs	Bandwidth
G.711 PCM	64 kbps
G.729 CS-ACELP	8 kbps
G.728 LD-CELP	16 kbps
G.723.1 CELP	6.3 / 5.3 kbps

- Voice transmission:
  - UDP. It doesn't offer data integrity but it exploits better the bandwidth than with TCP. This is because UDP header is smaller than TCP header. TCP also sends again a package when it doesn't arrive to the receiver on a first attempt. This has no sense in VOIP because the package is needed just in real time, not after.
  - RTP (Real Time Protocol). Is in charge of the temporization aspects. It adds the UDP packages the required information so that the package arrives at the destiny on the correct time.
- Transmission Control:
  - RTCP (Real Time Control Protocol). It detects network congestion situations and makes some actions to correct them.

MGCP

Media Gateway Control Protocol (MGCP) is used for controlling telephony gateways from external call control elements called media gateway controllers or call agents. MGCP assumes a call control architecture where the call control intelligence is outside the gateways and handled by external call control elements. The MGCP assumes that these call control elements, or Call Agents, will synchronize with each other to send coherent commands to the gateways under their control. MGCP is, in essence, a master/slave protocol, where the gateways are expected to execute commands sent by the Call Agents. The MGCP implements the media gateway control interface as a set of transactions. The transactions are composed of a command and a mandatory response.

SIP

Session Initiation Protocol has been developed by the IETF as a standard for the initialization, modification and finalization of interactive user sessions with multimedia elements such as video, voice, instants messaging, online games and virtual reality. Is one of the possible signalling protocols of the VOIP.

H.248

Also known as Megaco, is the Standard that allows a media gateway controller (MGC) control a media gateway (MG). This protocol substitutes MGCP and MDCP and is complementary to the H.323 and SIP protocols.

Quality of services

The main problems in the VOIP are:

Packet Loss
Packets that contain voice data can be lost for several reasons:

*Insufficient bandwidth due to poor capacity planning
*Packets arrive at their destination too late
*Network outages

The system designer has to ensure that a network has the capacity to meet the bandwidth requirements of voice traffic. In doing so, the designer should consider the aggregate bandwidth requirements of all the communications that are taking place on a network. If the underlying network does not have sufficient bandwidth, it is said to be “oversubscribed.” In some cases, it may be necessary to dedicate a LAN to VOIP to avoid this condition.

Oversubscription can lead to lost packets. The probability of CSMA/CD detecting a channel that is "free" for transmission decreases as more devices use the network, but if the network is never free, it will discard the packet, causing packet loss.

Jitter

RTP is essentially a continuous stream of voice data that is transmitted in discrete packets. In an ideal VOIP scenario, packets traveling between points A and B would never be lost, always arrive in the same order as transmitted, always follow the same path, and always take the same amount of time to cross the network.

Jitter Buffers

Unfortunately, the ideal scenario rarely happens in real-world VOIP networks. Packets leave their source in order, but they are very likely to use different paths as they travel over the network to their destination. This is especially true when packets must travel across a WAN. Because packets almost never follow the same route from point A to point B, each packet experiences different amounts of delay. When packets arrive irregularly at their destination, they are said to be experiencing jitter.

In some cases, arrival delays might cause packets to arrive out of order. If the voice packet transmissions shown in Figures 6 and 7 were played immediately upon their arrival, the transmission quality would be very poor. To compensate for jitter and out-of-order packets, a jitter buffer is used to eliminate jitter.

Jitter and Lost Packets

Unfortunately, a jitter buffer adds further delay, and if the delay caused by the jitter buffer is too long, it will degrade voice quality further by injecting short periods of silence during playback. VOIP systems must keep the jitter buffer short enough to maintain the natural pace of the conversation. However, if the jitter buffer is too short, some packets may arrive after playback, and those packets are effectively lost. A packet which arrived after the buffering time had expired maybe with very serious consequences. Receivers have algorithms that interpolate and compensate for packet loss. However, the algorithm may not work perfectly, and the listener may perceive degraded voice quality.

Collisions and Local Area Network Load

Compensating for the occasional lost packet is relatively easy, but compensating for the loss of several consecutive packets is much more difficult. Because multi-packet loss typically occurs during peak network traffic in "bursts," paying careful attention to network load conditions is very important. A lightly loaded shared Ethernet network is normally defined as one utilized between 0 and 50%. On such a network, few collisions generally occur, and it is very probable that the packets involved in the collisions will be successfully transmitted during a second attempt. As network utilization increases beyond 50% and up to 80%, the chance of collisions is much greater, and the overall throughput of the network decreases. Fewer packets containing information arrive at their destinations, and such conditions can have a severe negative impact on voice quality. Little can be done to compensate for the loss of five or more consecutive voice packets in a stream. The situation worsens as a network approaches saturation with utilization greater than 80%. Under such conditions, latency can exceed a full second.

The amount of utilization that can be tolerated by VOIP applications depends on the type of network in use.

In Shared Ethernet VOIP applications should work best in shared Ethernet networks with an average of 30% utilization in Half-Duplex Switched Ethernet VOIP applications should work well with an average utilization of 70% and in Full-Duplex Switched Ethernet should work well with an average utilization of 80%.

Subjective Effects

High jitter and lost packets can cause several negative effects.
* Voice transmissions can sound unnatural or robotic.
* Extended periods of silence or gaps in the signal can occur.
* Excessive clicking or popping sounds may be heard.

Clock Drift and Synchronization

Clock drift can be another source of voice quality problems. Excessive clock drift can cause overrun of a jitter buffer or introduce additional delays.

* If the clocks of the receiver and transmitter are too far out of synchronization, the transmitter may fill the receive buffer faster than the receiver can process the data (overrun).
* The receiver may have to wait for a slow transmitter, causing unnecessary delay.

Clock drift can cause the same type of problems as packet loss and delay..

Echo

The presence of echo can have a positive or negative effect on telephone conversations. Side tone is a positive type of echo, and helps the user feel that a telephone is working correctly. Echo becomes a problem when delay is greater than 20 ms, and the echo is loud enough to be heard distinctly. Generally the human ear will not perceive a received sound as a troublesome echo if it occurs within 20 ms of the time when it is spoken and is attenuated by 25 dB relative to the level of the speaker’s voice.
Leakage that causes echo always takes place in analog components because digital components cannot leak any portion of the transmitted audio into the receive path. Two types of echo introduced by analog components are acoustic and hybrid echo.

Acoustic Echo

Acoustic echo is produced in a handset due to poor voice coupling between the microphone and the earpiece. Hands-free phone operation can also generate acoustic echo

Hybrid Echo

Hybrid echo is produced in analog phone networks at the hybrid device that converts 4-wire trunk circuits to 2-wire local loops. The device is a transformer that can reflect a portion of the speaker’s transmitted voice back into the speaker’s receive path.

Echo Cancellation and Suppression

In order to minimize echo and improve perceived voice quality, engineers have developed echo cancellation and suppression algorithms. The adaptive filter estimates and predicts echo in order to eliminate it from the Send-In signal. Several metrics are used to evaluate the performance of an echo canceller.

■ **Echo Return Loss (ERL)** — Ratio of Receive-Out and Send-In power, expressed in dB. ERL measures the Receive-Out signal loss when it is reflected back as echo in the Send-In signal.

■ **Echo Return Loss Enhancement (ERLE)** — Ratio of Send-In power and the power of the residual error signal immediately after cancellation.

Improvement of VOIP with IPv6

For a large enterprise, implementing a VOIPv6 solution will reduce the load on the external gateway needed for making external IP phone connections. This is because external IPv6 lines do
not go over the external gateway but through a firewall. Thus IPv6 telephony traffic will have to be assigned a port to reach the Internet. This reduces the load on the head office external gateway and lowers costs.

The implementation of VOIPv6 will simplify Internet VPN traffic paths for small nodes. Then the firewall of each node will provide a port to the Internet only for IPv6 telephony Traffic. The advantage of this arrangement is that no new IPv4 addresses will be needed which reduces costs (a similar configuration under IPv4 will require additional global addresses). Voice traffic paths via Internet VPN are also simplified.

Other advantage is the growths in IP internal telephony will double the need for addresses. For example, in an office with 100 employees, servers, routers, printers and wireless applications will require 50 addresses; PCs (DHCP assignment) will require another 150. That is a total of 200 which can be handled by a /24 prefix. If 120 IP internal phones are added to this, a total of 320 addresses will be needed. This is more than a /24 prefix can handle. In IPv4, this would require changing the subnet mask or adding a separate segment (defining IP phones as a separate segment), in fact, making it necessary to redesign the subnet mask and thereby increasing design costs. In IPv6, these redesign costs (caused by adding terminals) are unnecessary.

SKYPE and the differences with VOIP

Each Skype user must have the Skype software running on his/her computer. This software is currently available free of charge and can be downloaded from the company website, but the software is proprietary.

The main difference between Skype and other VOIP clients is that it operates on a peer-to-peer model rather than the more traditional server-client model. The Skype user directory is entirely decentralized and distributed among the nodes in the network, which means the network can scale very easily to large sizes (currently just over 70 million users) without a complex and costly centralized infrastructure.

Skype also routes calls through other Skype peers on the network, which allows it to traverse Symmetric NATs and firewalls, unlike most other VOIP programs (The two most common VOIP protocols, SIP and H323 are usually UDP and point-to-point, making NAT traversal problematic. This, however, puts an extra burden on those who connect to the Internet without NAT, as their computers and network bandwidth may be used to route the calls of other users. The selection of intermediary computers is fully automatic, with individual users having no option to disable such use of their resources. This fact is not clearly communicated, however, and seems to contradict the license agreement which would limit Skype's utilization of the user's "processor and bandwidth to the purpose of facilitating the communication between the user and other Skype Software users".

The Skype code is closed source and the protocol is proprietary which has raised suspicion and drawn broad criticism from software developers and the VOIP user communities.

The Skype client's application programming interface (API) exposes the network to software developers. The Skype API allows other programs to use the Skype network to get "white pages" information and manage calls.

Skype doesn't not allow calling to anyone outside their service over the Internet. For example, a Skype user cannot call a SIPPhone, IPTEL.ORG, or FWD user, but IPTEL.ORG, SIPPHONE, and FWD users can call everyone *except* a Skype user. Skype and other non-standards-based services are essentially VOIP islands, excluded from the open SIP VOIP community.

In the past, VOIP services where characterized as PC-to-PC or PC-to-Phone. In those days, PC-to-PC meant users on VOIP service A could call other users on VOIP service A. If another user was on VOIP service B, there was no way for the two to talk, unless they both *upgraded* to a PC-to-Phone service, essentially using the PSTN as the interconnect. Today, using the SIP standard, users on VOIP provider A can communicate with users on VOIP provider B, directly over the Internet. The media has not yet grasped the significance of this and they still report about VOIP in the same old terms, still considering the traditional PSTN as the means to interconnect.

Because of standards, users on Internet service A can communicate via email with users on Internet service B. Likewise, now that interoperable standards exists for VOIP, users on VOIP provider A can place calls to users on VOIP provider B, if both provider A and B operate interoperable SIP-based VOIP services. This is as it should be. That's how we expect the Internet to work. We don't need the PSTN to communicate with each other if we are both using VOIP services. Would we sign up for an Internet service provider that only allowed us to exchange email with other users on the same service, or charged extra for sending such email? I don't think so. And there will come a day when we will not use VOIP services that only let us call other users on the same service, or charge extra to call other users on other VOIP services.

But the media doesn't get it yet. They should be telling us which services are based on SIP-standards, which ones support standard SIP hardware/software as opposed to requiring propriety hardware/software, which ones support inter-provider calling, which ones support future standards like ENUM, and, just as importantly, which ones are closed systems with no inter-provider connectivity like the services of days gone by.

Using VOIP services, allows you to call anyone on the same service as well as people on any of the other VOIP services, and even anyone on the Internet with a SIP address/number, which includes a rapidly growing number of corporations, universities, and smaller service providers.