6

VoIP AND UNIFIED COMMUNICATIONS DEFINE THE FUTURE

Knowing some history of telephony and understanding the new technologies should prepare you to face your specific business problems. The intent is to help evaluate, select, and deploy future communications systems for at least the next decade.

6.1 VOICE AS BEFORE, WITH ADDITIONS

The goal of UC is to improve communications, which means the mark of success will be increased usage per person. What new functions will a migration add? What effect will it have on network traffic? The direction is up, but where will you start?

Before that increase hits, you should calculate if the IP network is able to absorb voice and UC in addition to data. A PBX will provide some demand information; phone bills are another source.

When planning basic capacity, compare what you know to the assumptions that vendors make when recommending the size and number of servers. Some vendors openly state their basis for a calculation, for example, in terms of the

VoIP and Unified Communications: Internet Telephony and the Future Voice Network, First Edition. William A. Flanagan
© 2012 John Wiley & Sons, Inc. Published 2012 by John Wiley & Sons, Inc.

number of emails, phone calls, and IM messages each person makes in a day or hour. If you don't validate the assumptions against your own data, any prediction will be a wild guess and probably wrong.

The same applies to assumptions about costs when calculating a potential return on investment. More on that below.

RFC 5359, *Session Initiation Protocol Service Examples*, collects best practices for legacy services using SIP "methods" and protocols. Watch for similar documents to emerge in the future, not only from the IETF but also from vendors, associations, and industry forums.

6.2 LEGACY SERVICES TO KEEP AND IMPROVE WITH VoIP

You can still get directory service, but 411 isn't free any more. Cellular carriers will dial the number they look up for you, either for no additional charge or for a fee. The correct time is available in most areas for a local call (but is being discontinued by major carriers). The author's audible time reading was within 2 seconds of the computer display controlled by the Network Time Protocol (NTP). These you might want to keep.

Astrology readings, gambling hints, and recorded financial advice? Might not need to preserve them. With digital controls, it is possible to block connections in the same way that parental controls can block websites. SIP phone numbers and web addresses are all URLs or URIs.

The point here is that even analog phones provide more than a voice connection. Unified Communications will provide far more functionality, some new and some carried over. You will have to decide which services your system provides, with a little help from your friends in local, state, and federal government.

Once completely uncontrolled, VoIP has become too important to leave to end users and peer-to-peer desktop applications. The call volume carried on IP networks exceeded the volume on TDM networks for international calling around 2010 and continues to increase. VoIP service providers such as Skype and Vonage avoided regulation as telephone carriers for years but became too large to ignore by taxing authorities. In 2011 the requirements for E911 location reporting and certain taxes applied to almost everyone providing voice service. Certainly enterprise users need to comply with E911 laws, which are becoming more demanding in more states each year.

On the positive side, a VoIP provider classified as a local exchange carrier (LEC) gets to control public (E.164) telephone numbers, direct inward dial (DID) numbers, and can assign them to its customers. A LEC also is entitled to handle the phone number of a customer who chooses to take that number from another carrier—though this ability to port local numbers is advancing more slowly.

6.2.1 Flexible Call Routing and 800 Numbers

What exactly does a phone number stand for? How does the network find that phone? The introduction of packet switching for voice not only unifies communications but also changes the answers to those questions.

Legacy PSTN switches present a separate hardware interface for each analog phone or trunk connected to the switch. Some of these ports are in the central office, some at remote terminals. Originally the identification of that physical port, or connector, was the directory number (DN), the telephone number in the phone book that you dialed to reach that phone. This meant that a phone number corresponded to a specific local loop that extended to a specific location from a specific switch. The phone number represented the region (area code), the central office (the exchange), and the phone line (the last four digits).

To encourage people to adopt the telephone, the Bell System business model imposed billing on the caller—there was no incremental charge to receive calls (beyond the monthly connection fee). If a business wanted to save its customers toll charges, to encourage them to call from a distance, it would obtain a local number in another area served by a different switch. A business had to lease a private line from the switch that controlled the desired number the business wanted to make available to its customers. That foreign exchange (FX) line passed through the business's own telephone central office to a local loop. FX let a business have local numbers outside the service area of it own CO. The monthly FX line charge was by the mile, so it was expensive and economically practical only over a small region or metro area.

From early on in the Bell System it was possible to call and "reverse the charges" (have the called party pay), but that required the intervention of a live operator and acceptance by a person who answered the call. The process was a bother for the caller and also very expensive for the callee. But business wanted to encourage customers to call and were willing to pay. To fit the billing system, businesses that paid for incoming calls received virtual phone numbers, the 800 numbers, which were handled automatically as reversed charges. 800 numbers changed what the telephone number could represent. Now it could be any DN, anywhere in the country.

To map an 800 number to a DN that the switches understand requires a lookup in a database (Figure 3.1). This was not difficult when only 800 numbers required a translation. A software change enabled the switch to find a DN when it received an 800 call, then use the DN to make the connection.

VoIP numbers and SIP addresses likewise represent any location. The improvement is that the new addressing reaches anywhere in the world.

6.2.2 Call on Hold

Typically applied to voice sessions, the "hold" function cuts off the audio transmission while keeping the connection. The original hold button activated a mechanical switch to isolate the hand set. In VoIP, the initiator of the hold changes the state of the media session while keeping the control session open.

In SIP, a re-INV message from either end carries an SDP body in which one of the attributes (directionality) for the session is changed to:

- a = inactive if the initiator does not provide music on hold.
- a = sendonly if the end placing the hold will provide MOH.

6.2.3 Call Transfer

A UAS that accepts a call (UA = B) can transfer the caller (UA = A) to another number (UA = C) by using INV and re-INV messages. The first session/dialog (A−B, based on SD-1 and SD-2) is put on hold (see above). Then:

- UA-B sends an INV with no session description to the transfer target (C) to create a second dialog.
- The target C responds with an OK 200 message containing an SDP offer (SD-3), which contains its contact information and capabilities.
- UA-B puts SD-3 into a re-INV message on the first dialog to UA-A, who now has C's information.
- UA-A sends an OK on the first dialog, with a session description SD-4.
- UA-B puts SD-4 into the ACK to C on dialog 2.
- UA-B ACK's A.

If the user at UA-B hangs up before C answers, the system should complete a blind transfer. If B talks to C and then hangs up, it is a supervised or attended transfer, just like the old days.

6.2.4 Call Forwarding

A UA proxy may handle call forwarding for a phone registered with it. The server may apply a filter, forwarding only certain calling numbers or calls at certain times of day, to voice mail or another destination. The end user may also tell the proxy to forward calls to a temporary location by changing the primary registration.

To effect a forward, the proxy issues an INV message to the new target phone (or terminal or server; fax calls can be forwarded too, e.g., on detecting a fax modem tone), creating a second dialog. The message includes the SDP information from the caller and the caller's contact information. It is not required for responses to come back through the proxy if it is stateless.

6.2.5 Audio Conferencing

Conference call service used to be highly profitable. The audio bridge was large and expensive, not something most enterprises wanted to own. Governments lacked capital budgets, but they could pay by the minute, per participant line, to discuss projects. Travel savings justified the cost of conferencing.

More powerful computer chips and cheaper memory brought down the cost of a bridge. Even early digital PBXs offered bridging for a handful (or two) of participants as part of the feature set.

Digital signal processors (DSPs) brought the cost down further, enabling bridges to conference thousands of lines on a call. These large devices allowed rural local exchange carriers (LECs) to offer free conferencing service. Participants called, mostly from a distance. The LEC collected call termination fees paid by interexchange carriers (IXCs, long-distance network providers). Conference participants pay for the call to the free bridge. For a fee, typically paid by the host company, the bridge owner will let participants call on a toll-free number.

In the improved UC arena, conferencing has become standard at no additional cost in both hardware and software products. It may reside in a media gateway, drawing on DSPs in the hardware. Or a software bridge may run on a call control server or on a dedicated device such as the H.323 Multipoint Control Unit (MCU).

Distinguishing differences among UC conference systems that can affect your business are:

- Number of ad hoc conferences at one time.
- Ability to schedule conferences to reserve resources for large calls.
- Total number of lines simultaneously in one or more conferences.
- Controls available to a conference moderator:
  1. Mute all lines or individuals.
  2. Choose video feed to distribute; from participants, speaker, or designated source.
- Authorization or authentication methods to control admission to a conference for privacy
- Facility to distribute documents or other files, via on-demand download or pushed by the host

6.2.6 Video Conferencing

Video conferencing scored high among end users when asked about the features they most desired in a Unified Communications environment. Video to date largely has been either one-way broadcast to a large audience (announcements, coverage of live events, speeches, etc.) or a meshed exchange of images among a relatively small number of users or sites. How you want to use video leads to another item on the previous list:

- Video procedures and image switching/combining during a conference or broadcast.

Three- to five-way video conferences allow all participants to view each other with at least a quarter of the screen per site in a static configuration.

Among larger groups, the best results come from switching the main feed to whichever site is producing the audio at the moment. The more sophisticated video bridges or Multipoint Control Units (MCUs) can tie the video feed to the current audio source.

Telepresence goes further by applying stereo sound imaging to place a speaker's voice with the image.

6.2.7 Local Number Portability

A later and much larger version of the 800 number data base now includes more area codes for toll-free numbers (866, 877, and additional free-call exchanges as needed). The same technology potentially applies to every phone number because of the US national policy called Local Number Portability (LNP). The Federal Communications Commission requires most carriers to support LNP by giving up or receiving the number of a customer who wants to change service providers but keep the phone number.

It's called local portability because the original scheme restricted the change to land lines in the same area code. Later cellular numbers were added. However, with the prevalence of unlimited long-distance calling, the importance of having local numbers or 800 numbers has declined.

Nevertheless, if you've grown attached to your phone number, or spent heavily to advertise it, you probably want to keep that number when moving from TDM service to VoIP. The FCC says you can do that. Routing calls from the PSTN to a VoIP system uses the same database that controls 800 calls. Either the called or calling party's carrier may operate a gateway that receives these calls and places them on an IP network. You may retain a gateway that receives calls on traditional TDM trunks, until replaced by SIP trunks.

6.2.8 Direct Inward Dialing, Dialed Number Indication

In the PSTN, dialed number indication (DNI) allows a call through a PBX (private branch exchange, the enterprise phone switch) to ring directly on an extension without requiring action by an attendant. Telcos call this feature direct inward dialing (DID) or dialed number indication service (DNIS). DNI lets the PBX route a call to a specific phone inside an organization.

DID depends on assigning a public (E.164) directory number (DN) to a phone. LECs charge a small monthly fee for DID numbers, which are assigned in blocks that may not be related to the primary phone number. Phones on a PBX that lack a DID number can call out by requesting a trunk (dial 9), but incoming calls must pass through the attendant. It is up to the customer (you) to have DID numbers for all extensions in addition to a general number that usually rings at an attendant station.

ISDN digital trunks from the CO always carry DNI in the call request in the format of the Q.931 packet signaling protocol. A PBX with a PRI (an ISDN trunk) interface understands the message. A media gateway that receives a

PSTN call extracts all the information from the Q.931 message and passes it in a SIP message to the call control server or media gateway controller (MGC).

If the central office switch sets up a DID call to a PBX on an analog trunk, the CO switch treats the PBX like another PSTN switch on a tandem trunk. The CO passes DNI to the PBX (or the MGW) during the 4 seconds after the first full ring (which is 2 seconds of 20 Hz a.c. imposed on the line to announce a call). COs use one of several signaling methods to send DNI on POTS lines:

- DTMF digits, TouchTone sounds.
- An asynchronous data protocol message of ASCII characters at 1200 bits/s. Frequency-shift keying, a V.23 modem signal, sends tones of 1200 Hz (1 or mark) and 2200 Hz (0 or space) to represent binary digits. There are detailed signal and timing requirements. The message includes the date, time, and a checksum to confirm accurate delivery.
- Multifrequency tone pulses, where each pair of six audible frequencies (odd 100s from 900 to 2100 Hz) represents a dialed digit (or a function, e.g., coin return). This older mode is as common as pay phones—not very.

Media gateways may be configured to receive DNI information from the PSTN and will pass the called number to the UA server that routes connections. The DNIS method on the MGW much match what the central office switch uses. In addition the call control signaling must also match. MGWs, in general, support FXO, FXS, loop or ground start, and possibly wink start and reverse battery. The carrier will supply this information about legacy trunks.

Reviewing data sheets for MGWs, it appears that not all will accept in-band DNI information (DTMF, MF). Those with only digital interfaces (T-1, E-1, ISDN) seem to prefer ISDN D-Channel signaling over channel-associated signaling (robbed-bit signaling) to receive DNI.

SIP trunking takes a view of DID and DNI similar to ISDN. All SIP calls are DID, targeted to a specific recipient. A SIP address, a URI of the form SIP: name@domain.tld or similar, identifies the called party so the recipient UAS always has a specific identity to locate (or group of phones). The attendant is just another specific DID address.

6.2.9 Call/Message Waiting

Media gateways in front of a PBX can recognize the signals for a waiting message:

- 100 V applied to the loop.
- Stutter dial tone.
- Proprietary signals to digital phones.

A SIP message to the IP phone turns on the light or triggers the display. For phones without a display the system can create a “stutter” dial tone.

Call-waiting tones and messages from a PBX are converted to a SIP Notify message. The recipient can see the information on the phone' display if it has one. If the target phone is on an FXS port, the gateway generates the message waiting indication (MWI), an audible tone heard only by the one party.

6.2.10 Call Recording

One reason to focus voice traffic at one point in the network is to record conversations—for "quality control or training purposes" as well as documenting financial transactions such as brokerage orders.

Several vendors offer customer premises equipment (CPE) for recording. Having your own system allows you to tag calls with date, time, employee number, customer phone number, internal extension number, and other data to facilitate retrieval on demand. Carriers and VoIP service providers who host call control may also offer call recording.

A choice between a service bureau and you own equipment will be influenced by cost but also by how often recordings need to be retrieved and the relative ease and speed to recover a specific conversation. Concerns about confidentiality may point to on-premises recording.

Many routers, media gateways, and session border controllers have the ability to duplicate media streams and send the copy to a recorder. There the packets are saved to disk. As with most collections of information, the disks will fill. Anticipate running out of disk storage with a policy to discard earlier recordings or to back them up to another device or medium.

An example is the Alcatel-Lucent RECORD suite of software. It will record not only the voice element but screen images of agents. It works with legacy phones and, through a MGW, IP phones.

6.2.11 Emergency Calling (E911)

When 911 appeared, the phone company programmed a translation into the CO switch for your phone. Dialing 911 sent the call to a local public service answering point (PSAP). That is, 911 was translated into a DN for a line to the PSAP, similar to the handling of calls to 800 numbers. PSAP lines are configured for centralized automatic message accounting (CAMA) to deliver the calling phone number, the calling line identification (CLID). An ISDN PRI trunk also will do the job in a different format. The idea was that the PSAP could call back if necessary.

At first the choice of PSAP reflected the physical location of the telephone line demark for the CLID. That is, the address where the telco terminated its local loop on inside wiring was reported to the PSAP. This is all the detail that a LEC is certain to know in every case. For a single-family home, the location is clear. A reverse directory lookup gives the address.

In a multitenant building, commercial or residential, the demark could be inside any office or apartment at the address. It's still reasonably specific

because the line carries a customer account name or apartment number. But things got messy.

The CLID itself no longer indicates an address without ambiguity when:

- Campus environments terminate phone lines in a PBX room, but the extensions cover many separate buildings, often far apart.
- Landlords operate phone services in large buildings, hiding customer names and exact locations behind a shared demark.
- Large corporations occupy dozens of floors in a skyscraper, all with the same address and general phone number.

People died because ambulances went to the main building in response to a call from another location that passed through the one PBX serving both locations.

As PSAPs evolved, the CLID became less certain to be specific. For example, an extension that lacks a DID number shows the general line as the CLID. A call to that rings at the attendant station, possibly in a different building. The need was to show a CLID that represented the location of the phone rather than the demark. PBX vendors modified their switches to provide a more flexible CLID, but much of the older equipment was not highly capable for this.

The CLID became an index to Public Service–Automatic Location Information (PS-ALI). This is a database that holds actual location information on a location information server (LIS). The local exchange carrier (LEC) or its contractor maintains the LIS for each region.

For each call received, the PSAP uses the CLID to retrieve the caller's address. The screen pop requires a dip into the ALI database. At a minimum, the ALI contains a street address. A perceived need for more specific location information grew into legislation in several states, and federal initiatives for Enhanced 911. E911 makes space for at least an additional 20 characters that can further identify the location. The enhanced database record can be more specific by adding a building number, floor quadrant, office number, or a combination of pointers to deliver a more specific emergency response location (ERL).

An ERL should represent no more than a zone on a specific floor in a building. In the most demanding jurisdictions a call to 911 must generate an ERL to identify the location of the calling phone within 100 feet. Other jurisdictions' requirements can be less strict. Almost all state laws have several points in common:

- 911 calls must not be blocked, either deliberately (to prevent prank calls) or accidentally (by not planning for them in a new phone system).
- ANI (caller ID, CLID) must be sent with a 911 call.
- Specific location data complying with local law is required; more is better, send as much as possible, including the calling extension number.

The oldest PSAPs remain tied to the ALI and CAMA trunks. The large number of jurisdictions with public service answering points (PSAPs) and a push to update them to handle more modes of communications means that the state of 911 communications is in flux. Within a decade expect development of "Emergency Context Resolution with Internet Technology" (ECRIT) to change the PSAP's connections. Work proceeds in the IETF, with several drafts undergoing work. Eventually, the Emergency Services IP Network (NSInet) in each region will tie together the PSAP(s), first responders, medical facilities, and emergency offices.

Large organizations that have a security staff will want notification of 911 calls in progress sent to the watch desk, with the location. Security staff can speed medical responses by alerting check-in points to admit ambulance crews, calling and holding elevators, and so forth.

Next Generation E911 will enable a PSAP to accept text messages, SMS messages, GPS location information from a caller's cell phone, and VoIP calls via SIP with the location information in a newly defined field of the INVITE message. A client-server architecture will be more adaptable to future technologies.

A more flexible stand-in for the CLID is the emergency location identification number (ELIN). This continues to be the phone number that the PSAP can use to call back, and the index to the location database, but it no longer need refer to the caller's specific phone. The ELIN may be a CLID, but it can also refer to a location zone that includes many phones, one of which the PSAP can reach by DID. An updated PBX reports the caller ID on a 911 call as the ELIN, not the main number or that extension's DID number. A good ELIN might ring at the desk of an employee designated as assistant emergency coordinator for the zone.

An organization obligated to provide detailed location information for the ALI has several ways to conform. The choices are roughly the same for a legacy PBX or a VoIP system.

- Install POTS lines at each location, that is, in each location zone. Route 911 calls to them, and let the LEC maintain the ALI. Changes can be slow to make. The LEC updates records once per day typically.
- Purchase a "gateway account" from the LEC to gain access to the Public Safety–ALI database. Each organization manages the records for its own lines and numbers. The gateway device costs several thousand dollars. The charge to store records is about $1 per year each. Updates take effect quickly when delivered to the gateway.
- Hire a contractor to process orders for phone moves/adds/changes into updates to ALI records. This process requires the customer to buy an ALI gateway, on-site server software, and the support service.
- Adopt E911 as a service. Small and medium businesses (SMBs) can put everything on the contractor, "in the cloud." Larger firms may need to stand up one or more servers to track all of its phone locations and help handle 911 calls.

Dealing directly with the PS-ALI database works best for firms with no more than a few hundred phones and not much movement of those phones. Larger firms may want the help of specialized server software, such as a 911 manager, a LIS, and voice positioning center (VPC). Firms with wide geographic presence or many remote offices may want a contractor that tracks the more than 5000 PSAP areas by city, county, and so forth. Political jurisdictions often share a PSAP so the mapping of PSAP coverage requires detailed local knowledge down to the street address.

Alternatively, an enterprise may engage a contractor to build and maintain a database of phone locations and zones. The "full-service" E911 contractor may process all 911 calls for an enterprise. Commercial services that provide E911 location information with calls to a PSAP offer both a local database of locations, within a campus, for example, and an interface to the public LIS. Either may satisfy the requirements for E911 calls from a corporate office, depending on the extent of the private network and the costs to build and maintain lists in the two environments.

The architecture for one vendor involves:

- A 911 server on customer premises, the location information server (LIS), that maintains a database of phone locations both inside and outside the private network.
- An IP-based call control server to forward 911 calls to the contractor's data center.
- A dedicated IP connection between the enterprise call control server and the contractor's voice positioning center (VPC), that matches locations to the PSAPs that cover them.
- A SIP-based IP channel to carry 911 calls from the enterprise call server to the contractor.
- Notification to the enterprise's security office.

All 911 calls pass through the enterprise call server, which picks up the phone's location form its own VPC. The call becomes a SIP INVITE message (containing the calling number and its location) to the contractor's softswitch. That switch uses the location to complete the call:

- Pick the appropriate PSAP and its regional PS-ALI database.
- Create an ALI record for the current location of the phone and an ELIN.
- Insert that ALI record in the regional ALI database.
- Find the telephone routing number for the PSAP from the master list.
- Forward the 911 call to an Emergency Services Gateway (ESGW) in the PSTN using the national Transport Number (TN). The 911 transport network then delivers the call to the selected PSAP.

An E911 feature is built into the firmware of routers intended for small or branch office locations that normally rely on the IP WAN for telephone service.

The 911 calls don't go to the VoIP call controller but are routed immediately to a local POTS line so that they reach the appropriate PSAP. For a small site there's no effort on the customer's part to maintain the PS-ALI database.

If the call passes through a media gateway to a circuit switched line such as a PRI, the gateway inserts the ELIN as the calling phone number in the signaling message. Calls other than to 911 carry the phone's DID number for the standard operation of caller ID.

6.2.12 Tracking IP Phone Locations for E911

VoIP phones can plug in anywhere. A user can move a phone to a different floor, take it home, or pack it on a business trip. When the phone or user authenticates to the call server, it has the same phone number but a different IP address. When an enterprise tracks migrated phones, it will also update the public service ALI database (the LIS) so that the PSAP finds the location for a phone number. In a future IP environment, the mobile or nomadic device will discover its own location, for example, via GPS or a location service of a carrier.

One common way to resolve the location of an IP phone on the enterprise network relies on its IP address. The network operator assigns IP addresses in the enterprise's Dynamic Host Configuration Protocol (DHCP) server so that every subnet address range falls into a known and controlled range of physical cabling within a defined floor area no larger than that legally required for an ELI. The location radius varies widely by political jurisdiction.

Configuration of the DHCP server restricts IP address assignments available to each zone. Any device that plugs into the network in that zone will receive an IP address, via DHCP, from the IP range assigned to that zone. When the phone registers with the call server, an important parameter is the phone's IP address. A 911 management server receives notifications of new registrations from the call processor, including the IP address in each new registration. This interface may require some system integration, often performed by the two software vendors in a partnership that enhances the value of both products.

The 911 server examines the database of IP address assignments to identify the location zone, and tells the call server the telephone number to use as the emergency location identification number (ELIN) when that phone calls 911. The call processor handles 911 calls by associating the calling phone with the corresponding ELIN, not the phone's DID number or DN. Thus the PSAP dips into the ALI database for the ELIN, which returns the correct location of the phone's zone.

Because the PSAP needs the caller ID, the SIP Forum's SIP Connect recommendation for carrier trunking requires behavior that conflicts with privacy requests. When a SIP-PBX places an emergency call it must not withhold caller ID even if the caller asserts a right to privacy on that information by including a P-Asserted-Identity header.

TABLE 6.1 Two methods to identify an internal location of an IP device

	IP Address/vLAN	Switch Port
IP address assignment	By DHCP	By DHCP
Extent of smallest zone	Subnet or vLAN	Cable drop or switch port
How 911 server resolves location	Configured table of IP's (same as in DHCP server)	SNMP search for switch port with registered IP and MAC addresses
How 911 server selects ELIN	Table matches IP address to ELIN zone	Table matches switch port to ELIN zone

Another way to locate an IP phone depends on mapping physical switch ports that terminate LAN drops to phones. Each port or group of ports is assigned to a location zone in the 911 management server. When an IP phone registers with the call control server, it notifies the 911 server, which searches the network for that IP address using Simple Network Management Protocol (SNMP). When found, the port is identified and with it the location zone. This approach resembles the earliest assumption that a number represents a location.

Accuracy in placing IP phones into zones depends on keeping the list of assigned IP addresses or switch ports up to date. Precise recording of the as-built configuration tends to get neglected. An outside contractor that specializes in network audits may prove more reliable. Automated discovery software is ofter a help in IP address management.

A record associating each ELIN with a location zone must be added to the SP-ALI database. This process may be manual, but automation in the 911 server, through a gateway, speeds the work for large networks and increases accuracy.

When the enterprise's VoIP call processor sees a 911 call, the ELIN is inserted in the call request or INVITE message as the caller's phone number.

Nomadic users of IP softphones pose a particular problem. They will acquire IP addresses from DHCP servers in hotels and airports that may not have a strictly defined geographical area that maps cleanly to a PSAP. It is possible to include these IP phones if the user will input a location when on trips.

There's at least one application for that, which runs on the same notebook computer running the softphone and links to the 911 management server. This application intercepts the registration process to interpose an additional step. Before the call control server will accept a remote softphone, the user must enter an address and any additional location details into the 911 server. That server then validates the address in the national Master Street Address Guide (MSAG) before allowing the phone to operate.

Upon verifying the location, the 911 server creates an ALI record and stores it in a temporary ALI. It is only when that phone makes a 911 call that the temporary ALI record is inserted into the proper regional SP-ALI database where the PSAP can retrieve the location.

Note that the phones do not need to be IP based. The location database for legacy PBXs has been populated manually but automated procedures are also used. For example, the human resources record typically includes a phone number and a workstation address for delivering postal mail. Those data can fulfill the E911 requirement.

A PSAP may convert to VoIP, on the way to UC, before the NG911 network becomes available. The site would insert a special purpose MGW's function to terminate the CAMA trunks from the CO switch (Figure 6.1). The gateway converts a call from the PSTN in the legacy format to a SIP INV addressed to the PSAP's proxy.

The E911 call that originates on the PSTN follows the legacy procedure as far as routing by the serving office to a 911 tandem switch. The call appears on a trunk to the PSAP, shown in the example as a loop-start analog trunk.

The E911 tandem switch seizes the line, in this case by closing loop and drawing current from the gateway. The FXS device acknowledges the line seizure with a wink signal (250 ms of reverse battery). Seeing the wink, the tandem sends the MF "Spill," the digits with ANI. The gateway creates a SIP INVITE message to the PSAP. All the MF spill information goes in the SIP From: header in one string.

The PSAP operator may transfer the call by sending a hookflash (a SIP INFO message with a "hookflash" body). The gateway converts that to a wink toward the tandem, preparing it to receive DTMF digits for the phone where the call should be transferred, sent in a separate message. A SIP REFER message from the PSAP, addressed to the gateway's IP in the Refer-To URI, will transfer the call to another extension at the PSAP whose number is in the "userpart" of the Refer-To URI.

When the VoIP call in the PSAP ends, the IP phone sends a BYE to the gateway. It then reverses the polarity toward the tandem switch, returning to idle.

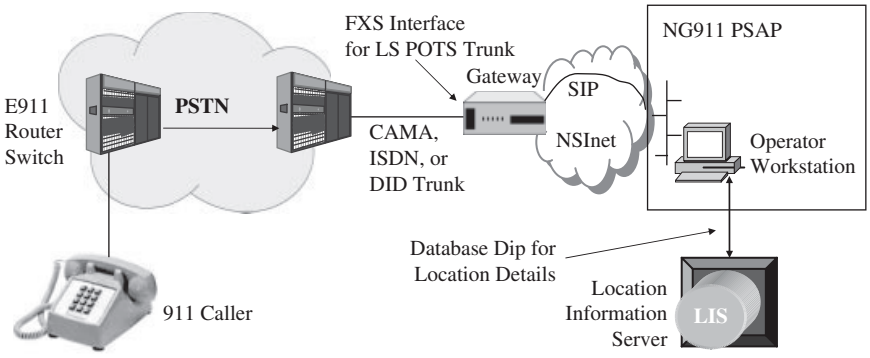

FIGURE 6.1 Gateway to receive a 911 call from the PSTN switch presents an FXS interface to appear as another PSTN switch.

6.3 FACSIMILE TRANSMISSION

Faxing remains important as a way to deliver legally recognized signed documents. Faxes are difficult to alter. By law they should be marked with the sender's name, phone number, and date stamp. A UC application can deliver faxes to individual email boxes and track in- and out-bound faxes for auditing purposes.

Unfortunately, every fax over IP implementation seems to differ in some way from every other. Hosted PBX and SIP trunking providers may handle fax calls the same as voice, using PCM encoding of the modem tones. Or they may decode the modem into a digital stream, carried in packets. Either way, they usually charge about $10/month for a fax DID number, which may be a dedicated MGW or the second port on a voice MGW.

Fortunately, the SIP Forum spent most of 2010 analyzing facsimile connections and came up with recommendations for changes to T.38. ITU is working on a draft of a revised T.38 at the same time. With the problems more clearly understood, it is highly likely that SIP interoperability testing will sort out a solution in the near future.

6.3.1 Facsimile on the PSTN

ITU's Recommendation T.30 describes real-time faxing over the public switched telephone network (PSTN) between two fax machines with analog (modem) interfaces. It's a universal format, the primary standard for almost all fax machines. The procedure is a dial-up connection originated by one machine and answered by the other. They recognize each other as fax machines by the tones they play toward each other. T-30 is digital at its core but most often travels as analog modem tones. A modem converts the digital fax format to an audible sound that passes like voice through a dial-up connection on the PSTN. Modems imitate voice in a way that PCM can reproduce.

Recall that the PSTN has practically no jitter and creates no interruptions to the audio path. Modem designs assume those conditions, which are not true for VoIP or IP traffic. Fax images, in particular, assume a dedicated voice channel for its format of a constant stream of pixel data.

T.30 persisted when communications technology converted to digital because modems were designed when channel banks had already digitized parts of the PSTN—fax machines were designed for the digital network. Almost always on the PSTN the voice channel remains uncompressed, using PCM/G.711 end to end, in part to ensure success of fax transmissions.

Carriers that compress voice can ensure success in faxing with fax detectors that recognize the modem tones and prevent compression. In a SIP environment the media gateway that connects a fax machine to an IP network should recognize the fax modem tones and re-INV to change the session from voice to T38fax. Problems exist because T.38 isn't precise on which end should invoke the change.

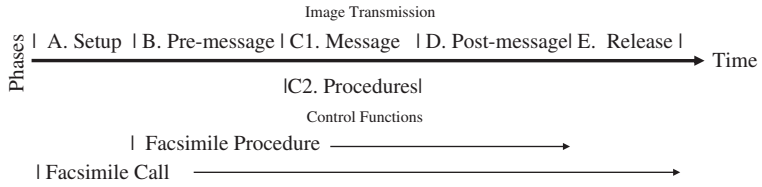

FIGURE 6.2 Phases of a fax transmission.

Fax-aware gateways respond to the difficulties with modems over IP by extracting the scan data (demodulating the modem tones) and transmitting digital information. The T.30 process assumes that the initial scanning to a digital signal follows ITU Recommendation T.4, which also includes run-length compression of blocks containing consecutive white or black pixels on a scan line.

On the WAN connection, T.30 puts the control information in HDLC frames marked by the 0x7E flag character (01111110) that is not used elsewhere. The T.4 data stream from a scan is not framed but runs continuously. Each scan line ends with a special EOL symbol (000000000001) that cannot appear anywhere else in a message. Between EOLs is a series of run-length codes, alternating between black and white, indicating a number of contiguous pixels of the same color. Here lies the sensitivity to jitter and lost packets. If the transition from black to white is confused, the received page can be illegible. On a voice channel the transmission "resets" after each line of pixels. When packetized, catching an error such as a lost EOL symbol is harder.

Transmission via modem takes place in phases (Figure 6.2) of a T.30 connection:

1. Call setup/clearing, such as dialing across the PSTN.
2. Facsimile procedures or commands by which the fax machines recognize each other and select formats.
3. Message transmission, an image encoded per T.4 or other data, sent continuously.
4. Confirmation of receipt.
5. Call termination.

Moreover fast modems have internal echo canceling, which can't train if there is even a small amount of jitter. A dynamic receive jitter buffer still allows enough jitter to pass through to spoil the connection. Network managers should turn off the dynamic adjustment feature for jitter buffers handling fax channels. Turn off VAD as well; some VAD implementations may ignore a modem tone and emit silence instead. That means end of page to the receiving fax machine, and end of transmission.

The slower modems don't cancel echo themselves, so they may be able to communicate as VoIP without resorting to T.38 gateways. However, the time to send a page increases significantly.

Greater problems for fax emerge when the voice carrier transcodes the originating PCM format to compress it into something that requires less bandwidth. ADPCM at 32 or 16 kbit/s has been used for international connections for decades. Two conversions in the path (slower, then faster) distort the reproduction of full-speed modem tones and usually make them unintelligible to the receiving fax machine. Most machines can shift to simpler modulation schemes, at slower bit rates, which may allow a transmission to succeed, but slowly.

The delay in a satellite hop is another problem, which some machines can overcome. Adding the delay of a packet network to a satellite hop make the case more difficult. Latency over 3 s (the so called T4 timer) may trigger retransmissions that overlap with delayed packets. Good spoofing of the T.30 protocol allows a fax relay system to operate with 5 s total latency.

If this is confusing, don' feel bad. Fax machines use multiple modem modulation schemes, from V.21 at 300 bit/s to V.17 to V.34bis at up to 33.6 kbit/s—on each call. The procedures and tone frequencies changed when the Recommendation was updated in 1996. Machines built to that earlier standard should be out of service by now.

Faxing over a packet network (via VoIP) offers three major scenarios under current standards:

- Real-time transmission or fax relay (T.38).
- Store and forward fax transmission (T.37).
- Revert to T.30 for the WAN portion of a connection.

6.3.2 Real-Time Fax over IP: Fax Relay or T.38

ITU Recommendation T.38 addresses the problem of how to fax in real time over a packet network by defining gateways and a protocol between them. This network replaces the audio tones in a voice channel with the digital run-length codes of the scanned pages.

Real-time transmission most closely emulates the analog operation over the PSTN. The users know they have concluded their transaction. The communications carrier never "holds" the information, so never has liability for its security under laws about health information (e.g., HIPPA) or privacy of personally identifiable information.

On the hardware, T.38 describes an interface to the packet network that carries the digital information from the scan output, defined in Recommendation T.4, in packets. A very few fax machines have a digital T.38 interface.

Almost all ordinary fax machines connect in analog (T.30 mode) on a modem interface designed for the PSTN. Rather than send that audible signal over a PSTN connection via modem, a T.38 process takes the modem tones into a device (a gateway) that demodulates the T.30 modem traffic into digital image information. The T.30 data stream then breaks into data blocks that are

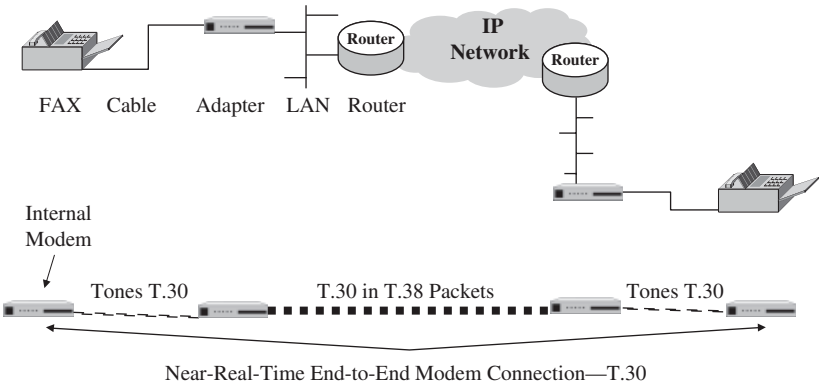

FIGURE 6.3 Real-time transmission of facsimile over and IP network (T.38).

encapsulated in a special version of UDP (with error correction), then in IP packets. A digital network such as a LAN or WAN (Figure 6.3) delivers the fax stream to another fax gateway.

A modulation function in the gateway at the delivery end recreates the modem tones from the transmitted data for the receiving fax machine.

The PSTN offers minimal latency and introduces practically no jitter on an audio channel. To isolate the fax machines from the variable delays and possible packet loss of a packet network like the Internet, a fax gateway will spoof at least part of the T.30 protocol. That is, the gateway terminates or generates the calling and answering tones of the fax call setup procedure (Phase A in Figure 6.2) and accepts DTMF dial strings, emits ringing voltage, and so forth. The packetized page scan content may be protected by optional end-to-end error correction.

Lacking T.38 gateways, fax over VoIP may be treated as voice at "normal" speed (pages per minute) if using the G.711 (PCM) codec. This encoding generates 64 kbit/s plus packet headers to carry a 9.6 kbit/s modem signal—ugly, and still not guaranteed reliable. Even an all-PCM fax connection (no compressions) may fail if any part of the connection is packetized, as in VoIP, without good QoS. Nominally able to carry the modem signal, G.711 over IP often suffers enough from packet loss and jitter to make fax transmission uncertain. This is a concern because the majority of international calls in 2010 were carried as VoIP.

Codecs that compress voice channels cannot carry high-speed modem tones accurately, so the fax machines back down to simpler (and slower) modulation schemes. They too may fail, for the same reasons.

Because the fax machines see only the analog modem signals, a T.38 gateway to the LAN needs to operate in real time, or very near it. That is, the gateway creates packets at short intervals when connected to a sending machine and reproduces a constant analog signal when delivering to the receiving fax. Silence causes the receiving fax machine to hang up.

When T.38 first appeared, RTP wasn't established. A form of UDP called UDPTL (UDP Transport Layer) was standardized and widely implemented—at this time it is still the dominant protocol for T.38 implementations. Version 3 of T.38 emphasizes RTP as the better protocol.

Each UDPTL protocol data unit (PDU) contains a sequence number followed by the current or primary data block—an Internet Facsimile Protocol (IFP) message—from the fax scanner. An IFP message contains an internal sequence number. The two numbers must be the same.

If error correction is configured, there are two options:

- In redundant mode, copies of one or more earlier IFP messages follow the current IFP message in the packet. With the current and the previous interval in each packet, two consecutive IP packets must be lost to lose any scan data. With two earlier IFP messages in each packet, a data loss requires three missing packets, and so forth.
- In forward error correction (FEC), each UDPTL packet contains the primary IFP message followed by a field that indicates the number of earlier messages that are included in the FEC process, then a parity-encoded representation of those messages.

IFP messages processed for FEC must have contiguous sequence numbers, starting with the sequence number just before the primary IFP message. That is, if the primary IFP message is number S, the sequence number of the first redundant message is S − 1 of the second, S − 2 and so forth.

The clever FEC process creates one composite message equal in length to the longest message included in the group. The previous messages are stacked vertically, the shorter ones are padded with 0's to the length of the longest, then each bit-wide column is "parity checked": if the number of "1" bits in a column is odd, the new message has a "1" in that position; otherwise, the result is "0" for that bit position.

With this parity information from *N* packets of FEC fields (sent in *N* packets), the receiver can recreate one missing packet (one primary message) in *N*. There is also a more complex FEC process that generates multiple FEC messages per IP packet. This scheme protects against some error bursts that lose multiple consecutive IP packets but requires more processing power.

FEC is optional. A gateway receiving redundant IFP messages may ignore them.

To identify a fax connection, the media type in an SDP block may read:

- m = audio/t38, indicating the transmission emulates fax-modem connectivity.
- m = image/t38 when the data is a TIFF file containing an image of the faxed page.

Because of the wide use of UDP and UDPTL, a vendor contribution to the IETF draft of proposed revisions to T.38 offers an example of what a fax-only INV might look like between T.38 gateways (Figure 6.4).

```
INVITE sip:+1-212-555-1234@bell-tel.com SIP/2.0
Via: SIP/2.0/UDP kton.bell-tel.com
From: A. Bell <sip:+1-519-555-1234@bell-tel.com>
To: T. Watson <sip:+1-212-555-1234@bell-tel.com>
Call-ID: 3298420296@kton.bell-tel.com
CSeq: 1 INVITE
Subject: Mr. Watson, here is a fax
Content-Type: application/sdp
Content-Length: ...
v=0
o=faxgw1 2890844526 2890842807 IN IP4 128.59.19.68
e=+1-212-555-1234@bell-tel.com
t=2873397496 0
c=IN IP4 128.59.19.68
m=image 49170 udptl t38
a=T38FaxRateManagement:transferredTCF
a=T38FaxUdpEC:t38UDPFEC
m=image 49172 tcp t38
a=T38FaxRateManagement:localTCF
```

FIGURE 6.4 INVITE message for fax connection.

The first "m=" line indicates a preference for UDPTL over TCP, which appears in the second "m=" line. The response will indicate port=0 for the rejected option and a valid port number for the accepted transport. The UA software takes care of creating the message, including the Content-Length field, which was not done for this hypothetical message.

For an INV asking for RTP/UDP, the message could resemble that shown in Figure 6.5. The first "m=" line now contains RTP with the Audio Visual Profile.

After establishing an audio connection between two UACs, one of them will re-INVITE with T.38 as an attribute, moving the call to fax mode based on PCM encoding (G.711) and no silence suppression (no VAD). The SDP message in the body of the re-INV could look like Figure 6.6.

The language of Recommendation T.38 leaves some ambiguity, which allowed different implementations of fax gateways. In some circumstances they don't interoperate. For example, T.38 isn't perfectly clear on which end of a sip connection should issue a re-INV to change from voice mode to fax mode. While gateways can detect fax modem signals of various kinds, only the HDLC flag characters in the initial V.21 modem modulation will distinguish a fax machine from a data modem.

Some T.38 implementations default to a behavior where the sending MGW waits for a media packet from the receiver before starting to send page images. It's like a wink-start telephone trunk, which can be good but raises a problem when working through a firewall with NAT. The media port won't be open to that first packet from the receiver until the sender opens it with a packet from the inside. To work in this case, the sender should be configured to start sending

```
INVITE sip:+1-212-555-1234@bell-tel.com SIP/2.0
Via: SIP/2.0/UDP kton.bell-tel.com
From: A. Bell <sip:+1-519-555-1234@bell-tel.com>
To: T. Watson <sip:+1-212-555-1234@bell-tel.com>
Call-ID: 3298420296@kton.bell-tel.com
CSeq: 1 INVITE
Subject: Mr. Watson, here is a fax
Content-Type: application/sdp
Content-Length: ...
v=0
o=faxgw1 2890844526 2890842807 IN IP4 128.59.19.68
e=+1-212-555-1234@bell-tel.com
t=2873397496 0
c=IN IP4 128.59.19.68
m=image 49170 RTP/AVP 100 101
a=rtpmap:100 t38/8000
a=rtpmap:101 parityfec/8000
a=fmtp:101 49173 IN IP4 128.59.19.68
a=T38FaxRateManagement:transferredTCF
m=image 49172 tcp t38
a=T38FaxRateManagement:localTCF
```

FIGURE 6.5 Connection request for RTP/UDP in a fax connection.

```
v=0
o=faxgw1 2890844526 2890842807 IN IP4 128.59.19.68
s=FAX message
e=faxsupport@company.com
t=2873397496 0
c=IN IP4 128.59.19.68
m=application 49170 udp t38
a=t38errctl:parFEC
```

FIGURE 6.6 SDP message in the body of a re-INVITE.

blank or idle packets (also call "no-op") immediately on the port of the image session. That will open the port on the firewall for packets from the sender.

The SIP Forum has a working group on the task of improving fax over IP. A key goal is to define repeatable tests to verify interoperability and capacity to handle multiple FoIP sessions, which usually require more CPU cycles than a voice session.

Version 3 of T.38 was nearing completion in 2011. It increases the maximum speed of transmission (lowers the time per page) by standardizing on a V.34 modem, described in the ITU Recommendation of that number, which is faster (up to 33.6 kbit/s). Older machines are built on V.17 modems for page images, 14.4 kbit/s maximum. When selecting a fax machine, fax server, or media gateway with fax capability, look for the T.38 Ver. 3 capability but be sure it can fall back gracefully to V.17 modulation.

6.3.3 Store-and-Forward Fax Handling

To automate handling of faxes, the receiving process can be emulated in software on a server equipped with a fax modem. Calls from the PSTN arrive over analog or ISDN lines. The modem demodulates received fax tones into a digital form that the server converts to a file format (typically TIFF) for delivery, storage, and viewing.

If the fax server is owned by the recipient, the server is considered the end point and the arrangement is nearly real time. If a fax server is in the middle of the network—outsourced to the carrier or a third party—the message transfer may become a "store-and-forward" process described by T.37 (Figure 6.7). It is more like email for faxes. The sending fax machine transmits to a server, in either T.30 or T.38 format. The server puts the digital information on disk, where it may sit until the recipient asks for it. Or the server may deliver the message to an outbound gateway that dials a destination fax machine over the PSTN and sends the fax via the T.30 procedures.

A server holding faxes may be subject to compliance and security requirements. Like emails, stored faxes may be archived for auditing purposes and to satisfy requirements for legal discovery. The server can offer additional security by authenticating users who try to retrieve faxes, encrypting fax files, and logging when and by whom faxes are originated and retrieved.

The server may deliver the fax at a scheduled time, immediately after receiving it, or on demand. With the fax in a stored file, it is available as part of a file system or the server can forward it as an attachment to an email.

When fax servers email the image files to the user, there's a MIME type for that (AUDIO/T.38) so that the form of the attachment is understood. The recipient can then pick up the fax like any email. Other delivery options include converting the fax text to spoken words when the recipient calls in, and delivery to a specified printer.

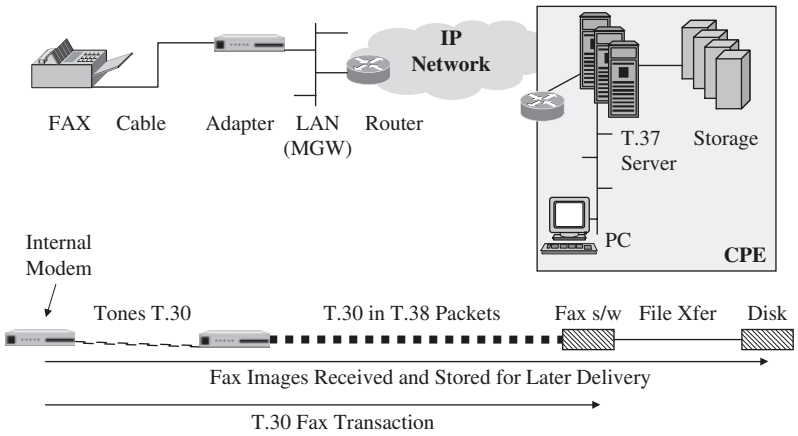

FIGURE 6.7 Fax via a store-and-forward server on a packet network (T.37).

Digital local loops such as ISDN lines will identify the called number with each inbound call. Caller ID and DID services are also offered on analog POTS lines. The fax server can associate a fax with the intended recipient through those numbers for delivery to an email address or a voice mail box.

Hosted fax services boast of large capacity to receive simultaneous faxes to the same phone number. An enterprise need not dedicate a large number of lines or machines to receiving faxes in response to a temporary rush such as a closing date, special product sale.

6.3.4 IP Faxing over the PSTN

So far, fax over IP as part of Unified Communications sounds fairly straight forward. However, the standards are not always interpreted identically by all vendors of equipment and software. There have been incompatibilities, though they are diminishing. Even the most popular brands of Fax to IP adapters don't work reliably enough for some critical applications.

The most common problems arise from the sensitivity of fax machines to delay and lost information. Too much of either and a fax machine can hang up and issue an error report.

When connected over a dialup connection on the PSTN, a voice channel has lower latency and practically no jitter compared to a packet network. IP transmission almost always increases latency (from buffering packets at multiple routers and switches across the network). If the packet network is congested, latency can vary (variation in latency is jitter).

Fax/IP systems that face regular fax machines need to look like another regular fax machine connected on the PSTN, a form of "spoofing." That is, a fax gateway needs to respond quickly and consistently to the signals from a fax machine, while dealing with a packet network that can introduce jitter and information loss.

As of this writing, the bullet-proof fax system is based on keeping the Fax over IP transmission on a LAN where the quality of service is controlled to a high level. Off-premises faxing reverts to the PSTN.

Hardware consists of fax modem cards in a server on each LAN. The POTS interfaces on the modem cards connect to local trunks. Workstations at both sites may run fax software, obtaining the benefits of message logging, easily viewable files, and flexible delivery. But each site uses the gateway to the PSTN for off-site faxing (Figure 6.8). The sending side dials up a voice connection and sends modem tones (T.30). The receiver may be another fax server or a standard fax machine.

This architecture certainly isn't elegant, or all-digital, and it doesn't rely on the Internet for inter-site transport. Many designers will reject it as inappropriate to include in a "UC" solution. However, the facsimile protocol in this case operates only over the PSTN, as it was designed. Users get all the UC benefits of routing and delivery to and sending from the desktop.

The optimal choice for you depends on how the business and the people use fax transmissions.

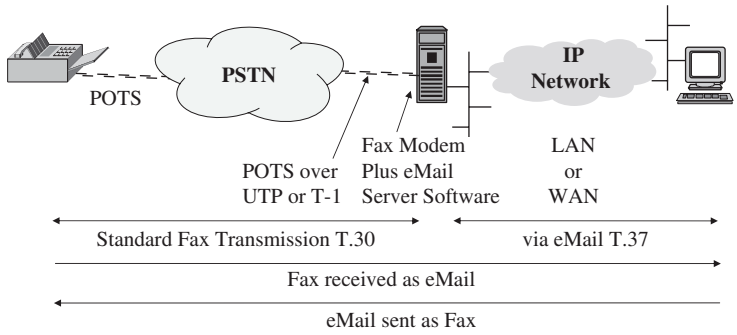

FIGURE 6.8 Sure-fire fax handling connects directly to the PSTN for off-site communications.

6.4 PHONE FEATURES ADDED WITH VoIP/UC

More than 50 projects in the IETF, additional ones in the SIP Forum, and proprietary efforts reminiscent of the PBX feature wars keep adding to the functions available on new telephone systems. Unified Communications integrates many additional features such as video, presence, and instant messaging. There are more than enough to distract the potential buyer.

The focus, then, must be on what's essential to your way of doing business. Review of sales literature will say more about emerging features. For starters, here are some things that should be available from most sources.

- **Coupling the PC and the phone:** manage the phone and calls on a PC. See who's calling, place calls, record calls, set up conferences from a web browser window or calling application.
- **Complete call lists:** not just completed calls, but missed or abandoned inbound calls—return the call with a mouse click.
- **Company and personal directories:** extensive information about fellow employees, provided from the HR database. Your contacts entered with space for as much information as you want from virtual business cards.
- **Privacy from integrated encryption:** IP phone participates in a virtual private network (VPN), IPsec, or Secure RTP—perhaps all three plus other methods—and can be configured and invoked easily from the phone or browser.
- **Choice of codecs:** for voice and video. Provides a range of trade-offs for bit rate versus voice quality.
- **Facsimile:** with a public telephone number (E.164), receive faxes in your private email box. Retrieve them from any browser or an application on a mobile phone or desktop computer. Send documents and images to any fax machine. For more details, see Section 6.3, Facsimile Transmission.

6.4.1 Presence

A survey (by Rad) revealed that end users consider presence very important in the business environment. This is not just for instant messaging (are you logged in?), but based on any information available about a user. At this writing, systems can display the busy status of the phone and meetings scheduled in calendaring servers. One vendor claimed to update presence if the keyboard is used at all.

In VoIP, status comes from the server where a user agent registers. This could be a hosted service or an "owned-and-operated" server. Many companies operate their own IM servers, for confidentiality, message archiving, or other reasons. The IM server can also track presence.

6.4.2 Forking

Find me/Follow me service in legacy switches would ring a series of telephone numbers until one answered or the call reached voice mail. Because numbers rang several times at each phone before the "ring no answer" count tripped the move to the next number, it could take significant time to run through the list.

VoIP systems improve on this feature by "forking" a call to ring multiple phones at the same time. The first to pick up gets the call. The phones that don't answer quickly return to idle status so that they can send or receive other calls—they are not tied up until the answered phone goes back to the on-hook state.

Forking proved handy in early VoIP deployments and became a favorite of people who weren't hiding behind voicemail. The caller got through whether at an office desk phone, cell phone, home phone, car phone (remember them?), or the girl friend's phone.

However, a proxy that forks an INVITE to multiple UASs will not do so unless the calling UAC provides acceptable authentication credentials for all of the UASs that ask for them. Smart proxy servers can still apply filters of various kinds to block or send to voicemail calls from specific sources, at defined times, or on designated days of the week.

6.4.3 Voicemail = eMail

Some organizations live in voicemail, some in email. UC promises to merge the two.

The technology to convert voice mail to text is speech recognition. It is highly developed in over 40 languages. When a caller leaves a voicemail, a UC system can convert it to text and deliver it via email. With cell phone displays now big enough to read the text, this can be a good way to retrieve messages in noisy areas such as airports or sports arenas.

Conversely, an email or text message converts to a voicemail via the text-to-speech feature.

The original message remains available in its first format so the recipient can check words that didn't translate or seem suspect from the context. Vendor

offerings vary, so check for convenience, reliability of conversions, and costs in terms of delivery latency and required servers.

6.4.4 SMS Integration

Short Message Service (SMS) was part of the GSM cell phone technology from the start. SMS is better known as "texting." Teens live by SMS, but big business has found applications such as alerting customer to delayed air flights, low bank balance, or items on sale. An interesting possibility suggested by an SMS vendor is to update an employee's presence status, including call forwarding, through a text message to a directory server.

The background on SMS involves Signaling System 7, which has a "part" for mobile users to send text messages. The destination typically is another mobile phone. Like most telco protocols, SMS was designed to be compact—the payload is limited to 140 octets on the GSM control plane. If using 7-bit ASCII codes, the maximum number of characters rises to 180. Complex character sets (Chinese, Japanese and Korean, Cyrillic) require 16-bit Unicode so only 70 characters fit in the SMS packet. Yes, that includes the title and space characters.

Extensions and services added by carriers and entrepreneurs has grown into a huge volume of traffic. In 2010 the number of text messages exceeded the number of phone calls on cellular system.

Twitter is based on SMS messaging. Carriers offer delivery of email via SMS, where up to the first 1600 octets of the email message is split over multiple SMS packets. Sprint will deliver an SMS message to a PSTN phone via automated text-to-voice conversion. If not on a plan with unlimited text messages, the charge of $.15 each can add up quickly.

The specialized nature of SMS has been a barrier to integrating the service in business processes. As mentioned, carriers will convert and deliver a text message in other forms. Specialized service bureaus or data brokers also offer ways to connect computers (e.g., of call center agents) to SMS. Unified Communications controllers can treat SMS as just another medium. Incoming texts can be queued at a call center with email and voice calls, to answered in order.

A common interface to send and receive SMS texting is as an email. The conversion requires a gateway server, these days connected on an IP network. Formerly a modem link was part of the infrastructure, but it's no longer is needed when HTML is the transport.

Only a few of the more sophisticated call control software platforms integrate SMS. At this time applications called Twilio and Voxeo Prophecy have the capability to send and receive SMS. FreeSwitch can receive but not send SMS. Asterisk, one of the VoIP software packages that's been around the longest, doesn't handle SMS at this time.

Because SMS is essentially a service of cellular carriers, it is harder to set up private services as is possible with instant messaging.

6.4.5 Instant Messaging

IM educated users about the concept of presence. The IM application on your PC displays your contacts who are on line at the time, as you are shown to them. Almost 6% of respondents to a survey in 2011 indicated this function is important to their businesses.

To distinguish IM from SMS, the difference is in the transmission method and the kind of server that mediates the service. An enterprise may host the IM server if it wants to control its internal communications.

- IM relies on a server where users register to obtain an account, then log in to establish their presence. The server may reside at AOL, Yahoo, another third-party data center, or on premises.
- SMS is a carrier function, since it relies on the signaling system of the cellular networks. SMS "brokers" may transfer messages between the cellular side and email or another form of message transfer. This function is integrated, for example, into comprehensive call center software.

IM can attach files, and messages can be longer than on SMS. SMS doesn't necessarily report on presence for your contacts.

The protocol of IM, XMPP, *Extensible Messaging and Presence Protocol* (RFC 6120), originated as the Jabber protocol about 1999. XMPP uses streams of XML (eXtensible Markup Language) to convey information, request responses, and maintain presence. Streams may be encrypted as TLS or authenticated (using a Simple Authentication and Security Layer, SASL) mechanism within this protocol.

IM servers authenticate themselves to each other, but each server has the responsibility to authenticate its own registered users. An IM server therefore cannot vouch for the name or address of a user registered elsewhere. Address imitation and forgery are possible.

XMPP addresses can be in any properly coded character set. Beyond the simple faking possible by substituting a numeral one (1) for a lowercase letter L (l), a user can switch character sets to be more creative with "confusable characters" in imitating a trusted address or name.

Unfortunately, there seems to be no technical way to prevent deliberate confusion. An IM server can make fakes harder by limiting all names to one font and character set (see Figure 6.9).

IM relies on XML over TCP. Signing on to an IM server starts preferably with a TCP connection negotiated with Transport Layer Security (TLS) for channel encryption; other transport is possible.

On that connection the client and server open an XML stream. In a sense the procedure is the same as starting to transfer an XML document. The two ends bind their specific resources to an XML stream. However, rather than closing the connection after a file transfer, the XML "document" remains "partly transmitted" for as along as the user is logged in.

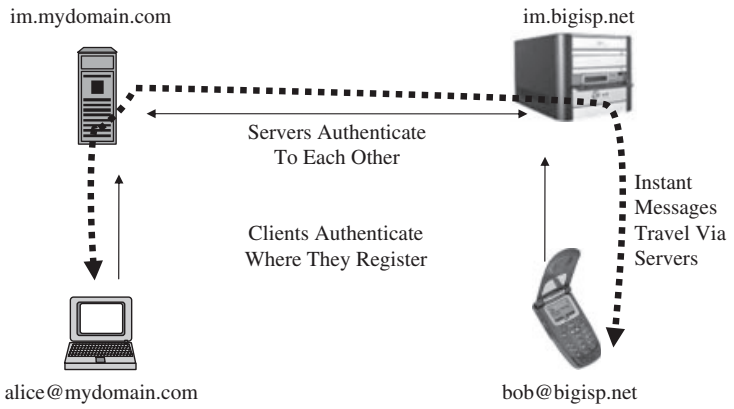

FIGURE 6.9 Instant messaging servers authenticate locally and between servers when accepting and delivering messages.

Adding an XML session fits easily with TCP as both are stream oriented. Each end of the XML connection binds a resource, like an application, to the stream. TCP and XML connections remain open indefinitely, allowing either party to push information to the other at any time.

When connected to the IP address and port of the receiver, the initiator opens a stream by sending the "initial stream header" to the receiving entity:

```
<?xml version='1.0'?>
<stream:stream
    from='juliet@im.example.com'
    to='im.example.com'
    version='1.0'
    xml:lang='en'
    xmlns='jabber:client'
    xmlns:stream='http://etherx.jabber.org/streams'>
```

The receiver replies with the "response stream header":

```
<?xml version='1.0'?>
<stream:stream
    from='im.example.com'
    id='++TR84Sm6A3hnt3Q065SnAbbk3Y='
    to='juliet@im.example.com'
    version='1.0'
    xml:lang='en'
    xmlns='jabber:client'
    xmlns:stream='http://etherx.jabber.org/streams'>
```

The entities then finish the stream negotiation process, which must include authentication via credentials using SASL. Servers may require encryption

(TLS). The order is TCP, TLS, SASL, then XMPP. For enhanced security, the devices can "forget" some of the information learned during the negotiation: the client and server may flush memory by exchanging new stream headers.

Headers may include "features/" elements, listing what is required or must be negotiated. Certain features require a restart when negotiated, for example, TLS, which is done by sending new stream headers on the same TCP connection. The XML stream gets a new ID number but is not closed.

Communications takes the form of blocks of information called XML stanzas. Any number of stanzas may be exchanged over the XML stream. There are three kinds of stanzas:

- **Message:** a push capability to send data,
```
<message to='foo'>
   <body/>
</message>
```
- **Presence:** publish-subscribe mechanism to advertise availability on the network,
```
<presence>
   <show>
   </presence>
```
- **iq** (**Info/Query**): request−response mechanism similar to HTTP.
  - Request:
```
<iq to='bar'
    type='get'>
  <query/>
</iq>
```
  - Response:
```
<iq from='bar'
    type='result'>
  <query/>
</iq>
```

These three elements are allowed only after completion of the stream negotiation. If either end attempts to send on earlier, the receiver must reject it and close the stream.

Users can send and receive any number of XML stanzas with any other users on the network. When a user signs off, the IM client closes the XML stream first, by sending a "stream/" tag, and then closes the TCP connection after confirming the XML closure. This procedure prevents opening a security vulnerability which arises if the TCP connection closes first.

The terminals on a long-lived TCP connections may not notice that it has failed during a period of inactivity. Therefore a stanza may be lost.

The "xmpp-client" uses port of 5222 for client-to-server connections. Port 5269 is standardized for server-to-server connections. These are the default ports registered with IANA).

6.4.6 Webinar Broadcasts

Lots of webinars in late 2010 replaced the media/analyst/consultant tours that vendors previously had to make when announcing a new product or company. The time and travel budgets saved certainly helped in the Great Recession. In many ways the web meeting is better for the recipient of information, too. For one, the recorded meeting can be replayed later to confirm details. Vendors use replays for internal training and promotional pieces.

With just two parties on a web conference, they can collaborate on writing or editing a document, create slides, or examine and discuss a complex chart or diagram in detail. It's more than a video conference.

6.4.7 Telepresence

Think video conferencing, but with feeling. The concept is to duplicate a meeting "across the table" by applying high technology and psychology. The images are hi-definition, the audio is hi-fi (7,000 Hz at least), and the lighting, camera angles, and seating are designed to contribute to the feeling of being there. Stereo sound places the speaker's voice near his image.

Broadband connectivity and H.264 video compression make the technology practical; only a half Mbit/s is needed. Dramatic drops in pricing for the components make it practicable.

Standardization proved a boon to telepresence. Different vendors' equipment interoperates. There is no need to keep the participant list to those inside one organization.

Global networks and global business leverage the savings from time and travel expense. One report claims a complete payback on a video conferencing investment in seven days of training sessions for people across the world.

A major push for the use of telepresence in the medical field shows promise of making doctors, particularly specialists, available to rural and remote areas. Thinly populated regions can't support all the specialties all the time. Doctor on demand quickly responds to a need without incurring the expense of full-time staffing at all locations.

6.4.8 More UC Features to Consider

Each UC vendor wants a unique selling proposition to convince prospects to buy its version. Look for the differentiators among features that can help your business, which could include:

- Real-time collaboration at a distance, between two users or across a group on a conference call (multicast connections). Meetings may be ad hoc, or configured administratively for a fixed team. Avaya features a drag-and-drop ability to pick participants from a directory listing to add people to a meeting quickly.

- In addition to audio conferencing on demand, UC that offers white board sharing, co-editing documents, photo distribution, file sharing, and prepared video footage as well as the speaker's image.
- Mobility whereby several vendors offer complete VoIP/UC phone features on portable devices such as netbooks and tablets as well as laptops and smartphones.
- Automated directories that combine all employees with contractors, vendors, and other contacts such as press and analysts.
- Instant Messaging integrated with system services, allowing side messaging during a conference call to any or all participants.
- Calendaring and scheduling playing an important role as at some firms where every meeting request comes through an invitation for a specific slot on the calendar. This can be too formal for some users, but it does make it clear when people are not available.
- Dashboard display of system status including usage, presence, and the modern busy lamp field.