Telephone Number Lookup

Robert S. Hanmer
Lucent Technologies
2000 Lucent Lane
P. O. Box 3033
Naperville, IL 60566-7033
hanmer@lucent.com

These patterns are part of a larger work of patterns that describe a telecommunications switching system in the form of patterns. NUMBERING PLAN describes the benefits of using a standardized numbering plan. TRANSLATIONS discusses a way to store a large database of telephone numbers for easy, and efficient access. SIGNALING and OUT OF BAND SIGNALING discuss two ways of getting information into the switching system to tell it who is being called.

Introduction

These patterns are part of a larger work that describes the architecture of a telephone switching system through patterns. In this larger work the patterns begin with a very general pattern to add a switching node to the telephone network (ADD A SWITCH) and end with detailed patterns such as these that describe how to solve the smaller architectural details.

We are all used to dialing a certain sequence of digits on a telephone to make a call. In North America we dial a 3-digit area code followed by a 7-digit telephone number. We can omit the 3-digit area code if our recipient is within the same area code that we are. In other countries the bunches of digits are different, but the general effect is the same, to divide the range of numbers into manageable chunks that humans and computers can remember and process easily. NUMBERING PLAN (1) deals with the chaos produced by random telephone numbers.

The North American Numbering Plan (NANP) defines the structure of telephone numbers in North America. Under it, a fully qualified telephone number is ten digits long. When a call is made from one telephone to any other telephone, the computer controlling the routing of the telephone call must determine how to process the call. Even with blocks of numbers in the dialing plan being illegal (for example, area codes that begin with "1"), a large amount of memory is required. The second pattern, TRANSLATIONS (2) describes an efficient solution to the large memory requirements.

SIGNALING and OUT OF BAND SIGNALING describe two ways of passing telephone number (address) information to a switching system. SIGNALING talks about the early development of in-band signaling where the signals are audible to the telephone caller. Advancements in the 1970's led to the development of out of band signaling that uses a separate network to exchange the address signals.

A glossary of some key words with specific meanings within the context of this paper is presented at the end.

1. NUMBERING PLAN

… Human operators work well with names for the network endpoints "*connect me with Betty please",* but computer systems (and predecessor electro-mechanical) work better with more numeric addresses. To address this, numbers are used to identify each endpoint in the network. Because there are many network endpoints, ADD A SWITCH [Copl] was employed to add multiple switching systems to the network.

Computers work well with regular, consistent patterns of input. They work less well with random, chaotic inputs. Every locality can create its own local way of numbering telephone endpoints. Every region can have its own way of numbering telephone endpoints. Different areas that have local telephone networks have different needs. A small town might only have fifty telephones, whereas a city might have hundreds of thousands. The small town might only require dialing three digits. In the city perhaps seven digits are needed. Addressing this variety of numbers in a regular way is a difficult problem for a computer.

❖ ❖ ❖

How can we make sure that each telephone endpoint can be uniquely identified? How can we make sure that they can all talk to each other?

A numbering system designed for people should be easy to use and to remember. When telephones were first made available, the telephone's name (and its address) was the name of its owner. The identifier should be consistent from one locality and region to another. It should be structured so that machines of differing capabilities can easily parse it. Computers can easily handle this kind of data.

If everyone were to be able to choose the number that they want, without any limitations, there would be total chaos. And this chaos would not support the ability of electromechanical and computers to process regular systems of numbers.

A system of numbering designed for a computer might not be easy for humans to use. An example of this is the IP numbering system (see sidebar). Its numbers are not easy to use so a Domain Name System was created to map the numbers into the commonly used names (such as hillside.net or google.com).

Therefore,

Create a standards body to define a standardized numbering plan across all the jurisdictions. The consistency of the standards will make the numbers easier to remember.

The numbers should be assigned by the telephone carrier in a way that makes the number's use and administration most efficient. This efficiency will benefit the telephone users by minimizing the difficulty and delay in making telephone calls and connections.

❖ ❖ ❖

The standardized numbering plans need to be defined. In the USA, in the mid 1950's the North American Numbering Plan (or NANP) was created and adopted for telephone numbers. In the data realm the common numbering plan is the IP address.

There are lots of numbers, and lots of routes through a network to connect the telephones that those numbers represent. Every system that is used to make a connection between these endpoints needs an efficient way to store and lookup where the destination number is, and how to get there in its SWITCH DATA STORE [Hanm2003]. TRANSLATIONS (2) addresses the efficient storage of this information. …

SIDEBAR: Example Numbering Plans

The North American Numbering Plan. [AT&T]

In 1950, the USA had a population of approximately 151,325,000 people [USCB]. Not every one of these people required their own telephone number. However the number must be increased by all the non-people entities (i.e. businesses) that use telephones. Some way of reaching at least this many people was required.

In the 1950's the North American Numbering Plan, or NANP was created. It is in use in 19 North American Countries.

In the NANP a telephone number is of the form NPA-NXX-XXXX, where N is a digit from 2-9, P is either 0 or 1, and both A and X represent any digit from 0-9.

Codes of the form NPA represent an "area code" or a "Numbering Plan Area". These are geographic regions that form the basis for geographic grouping of smaller telephone groupings.

312	Chicago Metropolitan
201	New Jersey

218	Northern Minnesota
425	Washington state
507	Southern Minnesota

Table 1 Example NPAs

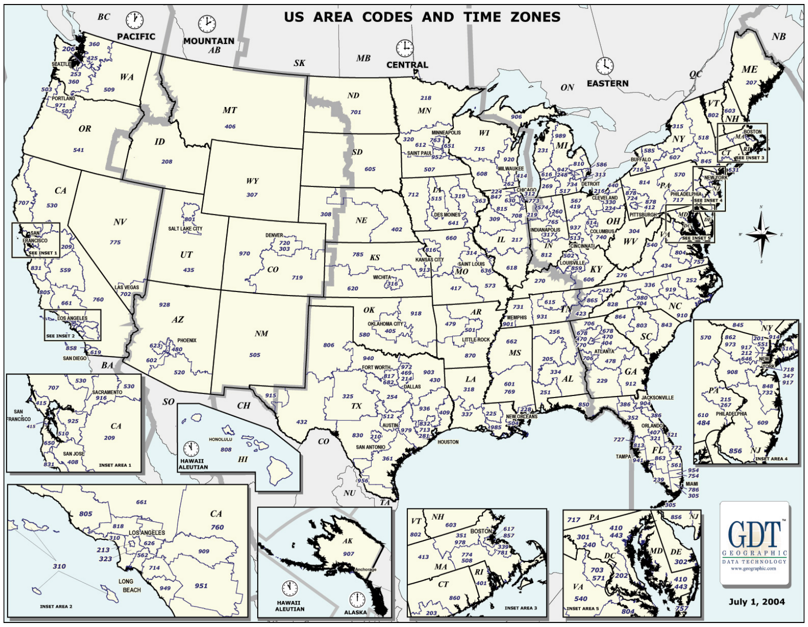


Figure 1 USA Numbering Plan Areas (from NANPA.com website)

Codes from the NXX represent office codes within a geographic region. Thus NPA-NXX refers to a specific unique telephone office within the country.

Within each office there are the provisions for 10,000 different telephone lines. Modern switches may have more than this number of lines, in which case they will have multiple office codes.

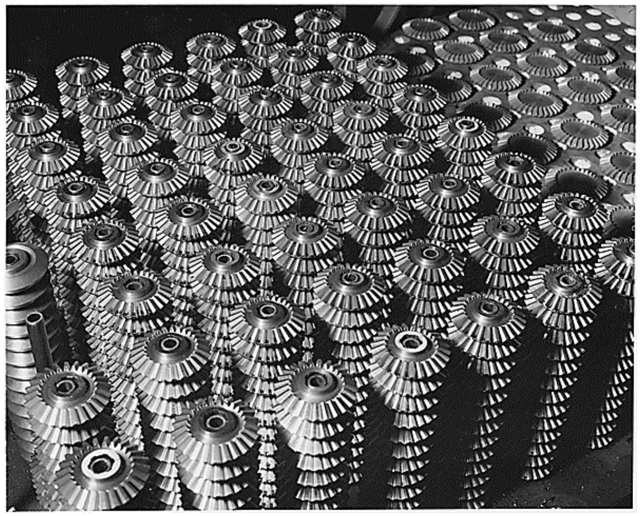
The NANP provides uniqueness. NPAs cannot have the same format as NXX. Prior to 1995 NPAs must have either a 0 or a 1 in the second position. For example, 312 or 202 but not 847 or 923. And the office code part of the number (NXX) cannot have 0 or 1 as either the first or second digit. For example, 724 or 979, but not 610 or 123. If the system looks at the first two digits it should be able to tell whether it is receiving an NXX or an NPA. If it determines that the number is the start of a ten-digit number of the form NPA-NXX-XXXX then it knows that it can expect seven more digits. If, on the other hand, it determines that it has received an NXX office code then it can expect only four more digits. And if it receives one of the special three digit codes, like "911", then it should route the call immediately.

SIDEBAR:

IP Numbering scheme

There are two Internet Protocol (IP) numbering plans in current use. IPv4 and IPv6. In IPv4 the address is a 32-bit number that is usually expressed as four numbers separated by periods, for example 36.120.120.3. In IPv6 the address space uses 128 bits expressed as eight 4-digit hexadecimal number separated by colons. For example, 1080:0:0:0:800:0:417A.

2. TRANSLATIONS[1]



… The system has a small SWITCH DATA STORE [Hanm2003] that contains the assignments of telephone numbers to specific communication channels, i.e. to lines or trunks. NUMBERING PLAN (1) has been used to define a plan that provides a consistent method of numbering locations and telephones. This numbering plan allows for a varying number of digits. It includes some specialized codes for some special telephone numbers, such as "911" to reach local emergency services and "411" to reach directory information. The North American Numbering Plan (NANP) provides for 1×10^{10} different telephone numbers (even though not all are used) and it provides the flexibility to intermix both short (seven-digit) numbers and long (ten-digit) numbers. This flexibility brings with it many unused numbers. This in turn makes storage of the numbers problematic.

❖ ❖ ❖

With a large variety of information that must be associated with different telephone numbers, efficient storage and efficient retrieval of telephone numbers is essential.

Ideally when a telephone number is received from a customer the switch can immediately know how to route the call and can make the connection directly. The NUMBERING PLAN (1) helps this happen by creating a regular pattern of telephone numbers that have well-defined rules how to parse. But it also complicates the issue by potentially forcing the definition of large blocks of unused numbers.

By allowing a variety of numbering formats in the numbering plan (the NANP is assumed throughout this pattern, refer to the sidebar in NUMBERING PLAN (1) for more information) the problem of translating from the dialed digits to a destination is made

[1] Strategy alluded to in [GHRS], pp. 1206-7.

more difficult. SIGNALING (3) provides for serial input of address digits. If the caller dials three digits, do they represent an office code, that is the NXX part of a NXX-XXXX telephone number, or does it represent a special 3 digit code such as "911"? In some locations an initial "1" digit is used to signal that the call is leaving an area code. This simplifies the deciphering because it indicates that the following three digits represent a Numbering Plan Area, the NPA part of NPA-NXX-XXXX and not an office code.

In many scenarios the digits are received serially and collected one at a time as they are received. This is not the case when OUT OF BAND SIGNALING (4) is employed. If a human is entering the digits there might be pauses between the digits sent. These can be misleading. Did *#-#-#-pause-#*… indicate NXX and then the start of the remaining digits, or was it a special code and then a mistake? Should that following X be discarded or used? The digits must be examined to determine this

Sometimes calls begin with special codes like "1" or "011". These special digits are usually used to select the type of treatment that the call receives and are not used during the lookup.

One way that the SWITCH DATA STORE could be arranged is for it to store every combination of digits. Then when it determines that it has collected all of the digits in a number it would look up the value in the Data Store's tables and know how to route the call. This direct access based on the telephone number is very fast. But it won't work in practice because of the number of possible telephone numbers in the NANP; the system probably doesn't have enough storage to handle this number of values. Another drawback is that large blocks of memory might be empty because they correspond to invalid telephone numbers.

Consider the system indexed on the full ten-digit telephone number, and the special codes were just ordinary entries in this table. This would lead to $O(1x10^{10})$ indices into the table. Even if only a very small amount of data were stored in the table, this still requires more memory than a great many systems have available.

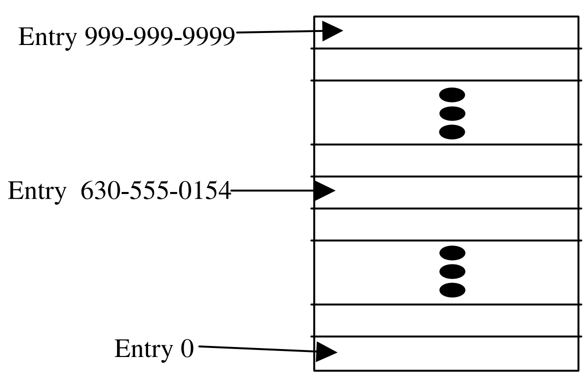


Figure 2

If the numbers that aren't possible are eliminated, for example all the five digit numbers since they aren't valid under the NANP the table would still need to be addressable by $O(1x10^{10})$. That's 1000 three-digit numbers, $1x10^7$ seven-digit numbers and 1×10^{10} ten-digit numbers unless something can be done to eliminate the impossible numbers. Some of the numbers are not valid, for example the seven digit numbers that begin with “1”, or the ten-digit numbers that have a digit between “2” and “9” in the second position. This is impossible to do in a simple table.

The numbers could be used as the hash table key to point to the correct bucket. This is efficient for memory but takes considerable processing power. Converting a ten-digit number to a hash key was a non-trivial task when this problem was first addressed. It required processing resources that were not always present. Modern computers can perform these operations much more quickly, but today’s larger number of subscribers mitigates this benefit.

Different communities have different needs for telephone digits. A small town might not require that seven digits be used to route a call since there are only several hundred telephones in the town. When there was a human operator this was not a problem, the telephone numbers could just be the short number. The first switching systems that these communities received were not computer controlled telephone systems. Their systems were hardcoded to map to this shorter telephone number. Other locations are very large and actually have more than one NPA assigned to them. For example, Chicago has two NPAs and its suburbs an additional five NPAs (in 2004). The telephone system should be flexible enough to support these varying requirements, allowing a small community to continue to route its calls based on seven-digit numbers, and the metropolitan areas to route their calls based on a full ten-digit number.

The SWITCH DATA STORE should store only those records that make sense. As seen above, it doesn’t make sense to store the invalid numbers. But there is no easy way to eliminate them that doesn’t require processing time (as in hashing the number). Another way of storing these records is to use a sequence of tables.

A 1000 entry table will handle all of the three-digit numbers that represent the special codes, all of the NPA’s and all of the NXX value, refer to Figure 3. Each entry stores two data items. One indicates now many digits are expected; this is called the Acceptable Digit Count (ADC). The other data item is a pointer to the table with information about what to do next. When a number is looked up, an ADC value of zero will indicate that no more digits should be accepted and that routing can take place. For a special code the ADC value will be zero (i.e. 911 needs no more digits), and the pointer item will be null.

3 Digit Code	*Acceptable Digit Count*	*Next Table*
999	invalid	invalid
…		
911	0	Route now
910	7	NPA 910 Office code table
…		
000	Invalid	Invalid

Figure 3

When ADC is not zero it will indicate how many additional digits should be accepted. For an NXX entry, the ADC will be four, which is the final XXXX part of the address; for an NPA entry, the ADC will be seven, indicating that an entire local number is still expected. When ADC is not zero the pointer points to the next data table that should be consulted, as additional digits are received.

When the NXX is entered it will point to a 10,000-entry table that contains specific information about how to process each NXX-XXXX number that is possible. When an NPA is entered, it will point to another 1000-entry table that contains each of the possible office codes for that area code.

There are two cases to consider if a ten-digit number is being received. The number can either be inside the current NPA, or it can be in a different NPA. In the case where the number is inside the current NPA, the NPA can be ignored (discarded) in the processing and the system can just start looking for the seven-digit local number. In the case where the NPA part of the number points to a different NPA then processing must wait until the final four digits have been collected, so that they may be passed along to the destination. They will not be used directly for routing at this time, but will must be preserved and passed along. The switch in the destination NPA will route based on the office code and line number part of the number.

The SWITCH DATA STORE will contain multiple 1000-entry tables and also multiple 10,000-entry tables. Sometimes tables can be reused, for example the 10,000-entry table for two NPAs might be identical (for example in the case when the office code is unused). Even without any reuse the storage required is $O(1 \times 10^4)$, which is a considerable savings from the $O(1 \times 10^{10})$ required to store all possible digits. Table 2 looks at the memory required for the worst-case table sizes.

Table Size	Worst Case Number of Tables	Worst Case Table Sizes (assuming each entry of ADC & pointer takes 1 memory unit)	Table Use
1000	1	1000	NPA or NXX first table

1000	# NPAs valid = 8*2*10 = 160	160,000	NXX subsequent tables, one for each valid NPA
10,000	# NXX valid = 8*8*10 = 640	6,400,000	Line numbers for all offices within the current switch's NPA.
		6,561,000	*Worst case memory units*

Table 2

This table solution is optimal in both processing time and memory usage. Processing requirements are minimized because the digits are used directly, without any computations of hashes or other indices. Populating only those tables that contain unique information optimizes memory.

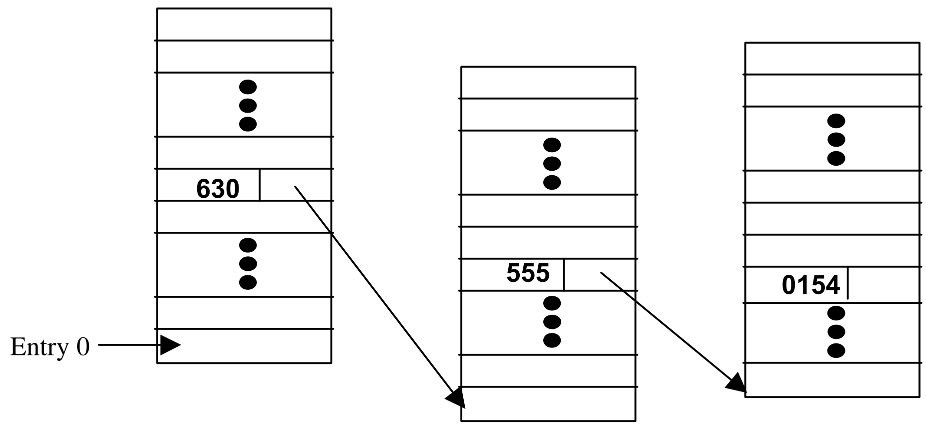


Figure 4
Example tables for 630-555-0154

The discussion above assumes that SIGNALING digits are being received in a serial fashion, i.e. one at a time. This approach with tables works even if all of the digits are received in a bunch, as will happen with OUT OF BAND SIGNALING (4). The parsing of digits will proceed through the tables from left to right, processing each group of numbers in turn. The groups of numbers will be those that represent the table structure, and ultimately the NUMBERING PLAN (1).

Therefore,

Use a sequence of tables to store information about how to process received telephone numbers. Process received digits in groups of digits by looking them up in the table. Store either information about what to do with all the digits processed so far or a pointer to the subsequent table that should be consulted.

❖ ❖ ❖

This pattern has talked about the NANP, but the solution will work equally well for other NUMBERING PLANS (1). The key is to choose the initial table sizes appropriate for the numbering plan. For example the key sizes in the French numbering plan would all be three digits long. In the UK they would a mixture of three and four digits long. The structure of the Numbering Plan leads to the structure of the tables and the grouping of digits that are processed.

There is still the issue of keeping all of this information synchronized between systems. A PROVISIONING SYSTEM [Hanm2003] comes into play by providing an external database of all the information and a means to get consistent views of the data into each switching system. …

3. SIGNALING

… One switch needs a way to talk to another switch. Within a switch intraprocessor communication methods support the two parts of a HALF CALL [Hanm2001] communicating and setting up a telephone call. The connection has evolved over time from simple beginnings from a single switch. When more than one switch is needed to connect a telephone call then inevitably one of them has information that the other needs.

❖ ❖ ❖

As the telephone network grows to contain more than one switch, a way of communicating requests from a HALF CALL **in one switch to another half call in a different switch is needed. How can this be done?**

History plays a part in the solution of this problem. The first telephones were paired in direct connects from one specific party to another party. Switchboards with human operators were then placed in between to connect the parties. The signaling that was used consisted of verbal instructions for the telephone operator.

When automatic switches were introduced, signaling as described in the context and problem were needed. Since the first automatic switches replaced the human operator the same wires between the party's telephone and the operator were used. The signaling had to be something that could go over the same wires.

The means of signaling has to be simple because of evolutionary nature of the simple machines that recognized and processed the signals. It must be basic because of the limited bandwidth available on a voice telephone line.

Any signal information uses the same line as the speaker's voice will use. While the signaling is going on nothing else can be happening, i.e. the speakers can't be speaking. The switch needs to receive the address uninterrupted, but this is okay because the parties are not yet connected so they won't hear the signals.

As technology improved ways of sending the address information over the telephone line also improved. Touch-tone was a method invented in the 1960's to encode a pair of tones onto a telephone line to represent the digits. Each digit of a telephone number needed to be signaled separately still, but each digit was represented by a single tone instead of a varying number of off/on's of the telephone line. With Touch-Tone the equipment that receives the tones can listen during a voice call. This can provide much greater flexibility than the earlier (dial pulse) methods and allows for the development of future features that use these tones.

Therefore,

Create a class of special tones that should be placed onto the telephone line to "signal" how to handle the call. These can be pulses of on/off or real tones.

❖ ❖ ❖

This doesn't resolve the force that the customer doesn't want to hear the signal information. As additional features and services for the telephone network were devised then sometimes-additional information was needed between the switches to handle it. And people began to find ways to defraud the telephone companies by spoofing the signals. All of these led to the need for OUT OF BAND SIGNALING (4).

The protocol that should be used between the switches needs to be defined. The pattern PROTOCOL FOLLOWS APPLICATION DESIGN [Marq] discusses a way to generate a protocol. When applying this pattern we have decided on the application, telephone communications. …

SIDEBAR:

Dial pulse and Touch-tone specifics.

Both are in-band signaling methods.

Dial Pulse

Early signaling over the same telephone wires as the voice would use consisted of pulses created by opening and closing a contact. This is the "pulse" mode found on many telephones today. The information conveyed in these signals contains the address of the destination telephone.

The equipment to receive dial pulses and interpret them at the telephone switch was specialized and could only be connected to the telephone lines when the receipt of digits was expected. The cost of the equipment to monitor every call for received digits is prohibitive, and once the address is received the equipment is reused on different calls.

Touch Tone

Touch-tone signaling works by combining two tones of different frequencies, a high-frequency tone from the column of the keypad, and a low frequency group from the rows. The following table shows the frequency assignments [Rey, p 276]:

	1209 Hz	1336 Hz	1477 Hz
697 Hz	1	ABC 2	DEF 3
770 Hz	GHI 4	JKL 5	MNO 6
852 Hz	PQRS 7	TUV 8	WXY 9
941 Hz	*	Oper 0	#

Every line at the telephone switch can usually recognize touch-tone signals. This allows the tones to be detected during a voice conversation, which allows for the use of tones during a call.

4. OUT OF BAND SIGNALING[2]

… SIGNALING (3) passes the tones over the same channel that the voice conversation will ultimately use. This has several limitations. The first is that the parties can hear the tones during the call. Another is that only a limited amount of information can be conveyed given that the customer might hear the tones and become impatient that their call is not being established. And yet another limitation is that since the parties can hear the tones they can spoof them to defraud the carrier.

❖ ❖ ❖

Low bandwidth audible tone signaling doesn't allow sufficient information to be conveyed between systems.

Telephone lines were optimized for the maximum of 8000 Hz that the human voice uses. This is a very low bandwidth that doesn't allow much flexibility to add other information to the conversation.

Sending signals over the voice channel results in a long signal latency, because information is sent serially over the line. Taken with the low bandwidth available then only a small amount of information can be conveyed.

A greater set of possible messages, as well as higher bandwidth/lower latency could be achieved if there were a way to communicate between switches that does not use the voice path. The additional messages that are possible can allow new and enhanced features that the carrier can charge for usage. This parallel channel could take advantage of advances in data communications technologies and transmit packets of information more quickly and easily than encoding the message within the limitations of the voice channel.

---

[2] Out of band signaling is described in [Rey, pp 280-284].

The majority of the signals that need to be sent between systems are related to are related to call-setup, specifically telephone number addresses. For the rest of the call they there is little or no information. If the signals do not use the voice channel then several can be multiplexed onto the same signaling channel.

Therefore,

Remove the signal from the voice path and put it on a separate out of band channel.

❖ ❖ ❖

This sets the stage for digital communications historically by demonstrating that call control can be passed outside the normal bandwidth and communications lines used for voice.

Using a separate channel removes the delays and overlaps between voice communications and the signaling communication. This makes expanding the range of signaling contents possible. This in turns makes complex interactions possible, where the systems working to create a call can communicate more than the basics. One of the first features that took advantage of this capability is toll-free (800) calling. This is because toll-free calls have at least twice as much addressing information as other calls; they have the toll-free number that was dialed as well as the address of the phone to which the call should actually be sent. Many other features and capabilities are possible because of Out of Band Signaling. A discussion through patterns of Advanced Intelligent Networking, which is one system of using out of band signals for complex call control, can be found in [HanKoc].

An Out of Band Signaling network sets the framework for moving the packets of voice onto the signaling path. The address signals are sent as packets. The change from including only signaling information to signaling and also voice took time to develop, but was a conceptually simple advancement.

SIDEBAR:

Common Channel Interoffice Signaling (CCIS)

The first implementation of Out of Band Signaling in the telephone network was Common Channel Interoffice Signaling. It was introduced in 1976 on the 4ESS Switching System using 2.4 kilobit per second data links. [Rey, pp 283,284]

…

Acknowledgements

I would like to thank Paul Adamczyk for his feedback and encouragement during previous PLoPs and by shepherding these patterns. Federico Balaguer also provided very useful feedback. The photographs used to introduce each pattern are used courtesy of the US National Archives and Records Administration (*www.nara.gov*).

I would also like to thank my PLoP 2006 Writer's Workshop group for their comments.

Pattern Thumbnails

PATTERN	*Reference*	*Pattern Intent*
ADD A SWITCH	[Copl]	Add nodes to a network to reduce the number of direction connections.
HALF CALL	[Hanm2001]	Use a 2 part (half call) model for call processing.
NUMBERING PLAN	1	Use a standard numbering plan.
OUT OF BAND SIGNALING	4	Put the Signaling information on a separate network.
PROVISIONING SYSTEM	[Hanm2003]	Use an adjunct system to download the Switch Data Store.
SIGNALING	3	Use special tones on the voice path to communicate between the switches.
SWITCH DATA STORE	[Hanm2003]	Install a small database to route the call.
TRANSLATIONS	2	Use a sequence of tables to efficiently store the cross-reference between numbers and processing instructions.

Glossary

Line	The communication path and the wiring necessary to communicate between a switching system and an end-user. The end-user may have either a telephone or some device such as a modem, fax machine, or an answering machine.
Office	A telephone switch is located in a "central office", frequently abbreviated as "office". This refers to the building as well as the switching system in the building.
Provisioning	The act of providing the information that customizes generic switching systems into a telephone system. Provisioning associates the specific generic capabilities and components of a system with the details of a specific installation.
Routing	As a noun, Routing refers to information about the path that a telephone connection will take across the network.

	As a verb, Routing refers to the act of advancing a telephone call from switching system to switching system across a network.
Translation	The operation of converting information from one form to another. In Switching Systems, the process of interpreting all or part of a destination code to determine the routing of a call. [Rey]
Treatment	The type of handling that a telephone call should receive. For example, whether the call is ready to be routed, or if more digits should be collected, and so on.
Trunk	The communication channel between two switching systems. [Rey]

References

[AT&T]—. **Notes on the Network**. AT&T, 1980.

[Copl] Coplien, J. O., "Worth a Thousand Words", **C++ Report,** vol. 10, no. 5, May/June 1998.

[GHRS] Giunta, J. A., S. F. Heath III, J. T. Raleigh, M. T. Smith, Jr., 1977, "No. 4 ESS: Data/Trunk Administration and Maintenance." **Bell System Technical Journal**, vol. 56, no. 7, Sept., 1977: 1203-1237.

[HanKoc] Hanmer, R. S. and K. F. Kocan,, 2005. "Documenting Architectures with Patterns." **Bell Labs Technical Journal**, vol. 9, no. 1, Sept., 2004: 143-163.

[Hanm2001] Hanmer, R. 2001. "Call Processing" in Proceedings of PLoP 2001 Conference.

[Hanm2003] Hanmer, R. 2003. "OAM-2," in Proceedings of 3rd Latin American Conference on Pattern Languages of Programs, Porto de Galinhas, Pernambuco, Brazil, August, 2003.

[Marq] Marquardt, K. 2000. "How to Define a Protocol for Object Transportation" in Proceedings of EuroPLoP 2000 conference.

[Rey] Rey, R. F. **Engineering and Operations in the Bell System, 2d Edition.** Murray Hill, NJ: AT&T Bell Laboratories, 1984.

[USCB] US Census Bureau, statistical abstract – http://www.census.gov/prod/www/statistical-abstract-04.html.