JR Richardson
Engineering for the Masses
JR.Richardson@cox.com

Leveraging Asterisk to Deliver Large Scale VoIP Services within a Carrier Environment

Preface

As the voice technology landscape pushes VoIP services closer to the edge and nearer to the customers, Carriers, ILEC's, CLEC's and MSO's struggle with the ever present battle of service delivery cost. When and how to deliver? What architectural model? Risk analysis regarding competition? What features to incorporate? Which network do we use, TDM or Packet based? Countless other questions arise when planning a large scale VoIP Services rollout.

The paramount question that is asked in the board rooms and the meeting halls is "How much will it cost and what returns can we expect?" The 10 year Return on Investment model went out in 1996. 2000 pushed out the 7year ROI in favor of the 5 year model and albeit today, smaller service providers favor the 3 year ROI model. Business is business, always has been and always will be, but timing is everything.

Deploying VoIP services 10 years ago arguably could have been the most important technological decision a Telecomm Company could have made. These days if you're asking that question, you're already behind. Asterisk can help launch VoIP Services provided by facilities based Telco's most inexpensively with arguably the best foundation and flexibility to build on with an ultimately scalable solution. Cost of capital, implementation, provisioning, billing, managing, repairing and other variables effect ROI.

Out with the old and in with the new is usually required when migrating to newer technologies. This paper investigates integrating new VoIP Services in with the old legacy VoTDM Services, with the least impact to variables effecting ROI. This enables decreasing Total Cost of Ownership and quickly deploying VoIP features and functionality to edge customers.

What is Asterisk?

"Officially, Asterisk is an Open Source hybrid TDM and packet voice PBX and IVR platform with ACD functionality. Unofficially, Asterisk is quite possibly the most powerful, flexible, and extensible piece of integrated telecommunications software available. Its name comes from the asterisk symbol, *, which in UNIX (including Linux) and DOS environments represents a wildcard, matching any filename. Similarly, Asterisk the PBX is designed to interface any piece of telephony hardware or software with any telephony application, seamlessly and consistently" *Reference "The Asterisk Handbook Version 2" Digium 2003*

Asterisk runs on general computing platforms. Linux is equipment agnostic and runs well on legacy and newer PC systems. Asterisk has a variety of telephony interface options, on the TDM side including FXO, FXS and T1. High Density DS3 interfaces are on the road map for Digium.

Asterisk supports SIP, H323, MGCP, SCCP and IAX2 (Inter-Asterisk Exchange Protocol). All of these protocol stacks have unique call handling capabilities and can be used to produce simple to elaborate solutions predicated on the environment of deployment and interoperability requirements. IAX2 is a robust protocol that works as a trunk between Asterisk servers and has features enabling firewall penetration for easy two-way communications establishment without the need to configure firewall access lists or port forwarding.

Dial plan scripting within Asterisk is by far the most flexible environment for configuring a soft switch. The administrator is only bound by his imagination when creating IVR and ACD solutions with Asterisk. With 80+ imbedded programs and utilities, Asterisk can handle the most complex tasks of a call center environment. Asterisk also has the ability to extend control outside of its applications and utilities and call on other programs within the Linux environment.

Asterisk can act as an application server such as Voicemail and Conferencing. Asterisk is a VoIP Registrar and User Agent. Asterisk is a trancoder for packet to TDM voice. Asterisk is a gateway between networks. Asterisk has a powerful API for programming new applications. Asterisk has a manager interface to interoperate with GUI programs. A more complete list of features can be found at www.asterisk.org under the features tab.

Hot Debate: Hardware or Software DSP?

Digital Signal Processing is the process of converting TDM or Analog Voice into Digitized or Packetized Voice and vice versa.

Digital Signal Processing has been a barrier to low cost deployment of VoIP Services since the technology hit the streets. Hardware DSP is expensive in that one dedicated chip or chip channel is required for every voice path given the same codec and possibly 2 DSP can be required if the voice codec's are different, i.e. GSM speaking to ULAW. Software DSP is handled by the CPU of the Soft Switch.

The use of software DSP technology has not proliferated in the carrier environment for several reasons, although the capability has been around for many years. Looking at the cost of hardware and processing power available in the mid 90's, software DSP was far from economical. Entering the 21st century, CPU speeds crested over 1GHz and general PC hardware cost declined rapidly. The software DSP function now warrants investigation.

Hardware DSP Cost versus Software DSP Cost Exercise

One of the more critical planning events when developing a VoIP solution is to calculate PSTN access or, more specifically, TDM to Packet voice translation. This gives the basis on how many DSP functions are needed which allows planning to proceed to hardware costing.

Since we are investigating a large scale VoIP deployment, we'll be using a customer line side extensions count at 200K.

When building Switched Telephony networks, a line side (looking to the customer) to trunk side (looking to the PSTN) compression or over subscription ratio falls between 4:1 and 10:1, increasing with customer base. For this exercise we will use a 7:1 line to trunk ratio.

Let us look at the cost of dedicated hardware DSP boards for a soft switch. With a trunk ratio of 7:1 for 200K lines, we'll need 28,571 DSP chips imbedded in our hardware platform for trunk lines. A quick search for Dialogic boards on Google came up with several resellers offering new 2xT1 voice boards for ~$7,999. As a large carrier we have shrewd negotiation tactics and leverage economies of scale with volume purchasing. Working with a distributor or reseller 50% discounts should be in order. For the purpose of this exercise, l will set the price of a Dialogic 2xT1 card at $4,000. This roughly gives us $83 per port for dedicated hardware DSP functionality ($4,000/48 ports). Keep in mind, this cost is just for the DSP function and does not include server cost to house the card and make it work. So far we've spent $2,384,000 to provide dedicated hardware DSP for our 28,571 trunk lines.

Now let us look at the software DSP model. We will use the same interface requirements of 28,571 trunk lines needed. Currently Digium is selling 4xT1 cards for $1500 list. Again, as a Carrier we should be able to reduce cost by 50%, at least for this exercise we will. So at $750 per 4xT1 board, we have a cost of $7.81 ($750/96 ports). Off-loading the DSP function to the central processor makes a huge cost difference between dedicated DSP chips and software DSP functions, so far we have spent $1/10^{th}$ the cost at $223,500 for our 28,571 trunk lines.

Processor Loading

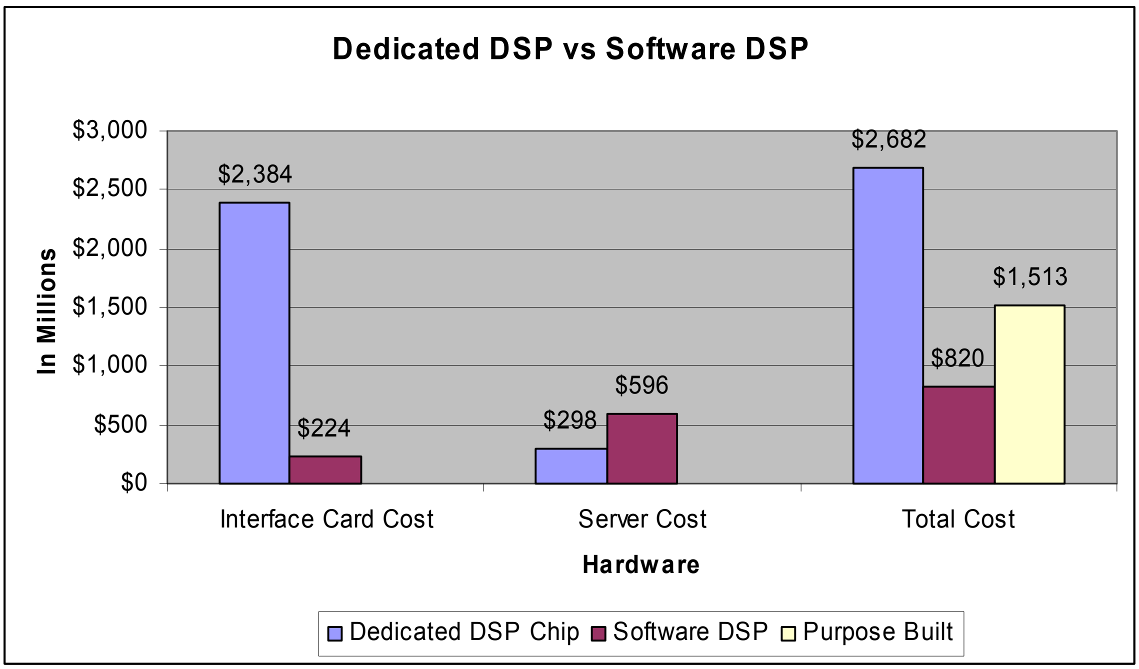
Loading the CPU with the DSP responsibility does have trade-offs when building soft switch platforms. To utilize the full channel capacity of a 4xT1 card (96 ports); the computer must be on the power side of computing by today's standards. Dual Xeon 2.8 GHz, 1 Gig RAM, 533 MHz front side bus, SCSI HD. This drives up the server cost to house the 4xT1 interface cards. Again, with strong Vendor relationships and volume purchasing server cost can easily be in the sub $2,000 range. For this exercise we'll use $2,000 as a base server cost and use 1 server per 4xT1 card. This requires us to have 298 servers. This adds up to $596,000 ($2,000 x 298 servers) needed in server hardware cost for our VoIP platform to access our PSTN through 28,571 trunk lines.

Adding up the PSTN gateway cost for 28,571 trunks using a software DSP model, we come up with $819,500 ($596,000 for servers + $223,500 for 4xT1 boards).

With the dedicated DSP hardware, the boards still need to be housed in servers. The servers do not have the responsibility for DSP function in this model so hardware requirements would not need to be as robust thereby reducing the cost of the server per card or the same server would be able to house more than one card. For this exercise we'll maintain our server cost at $2,000 per server but install 4-2xT1 boards in each server. This requires us to have 149 servers at a cost of $298,000 ($2,000 x 149 servers) for the server hardware cost for our VoIP platform to access our PSTN through 28,571 trunks lines.

Adding up the PSTN gateway cost for 28,571 trunks using a dedicated hardware DSP model, we come up with $2,682,000 ($298,000 for servers + $2,384,000 for 2xT1 boards). This is 3 times the cost needed for the software DSP model in this exercise.

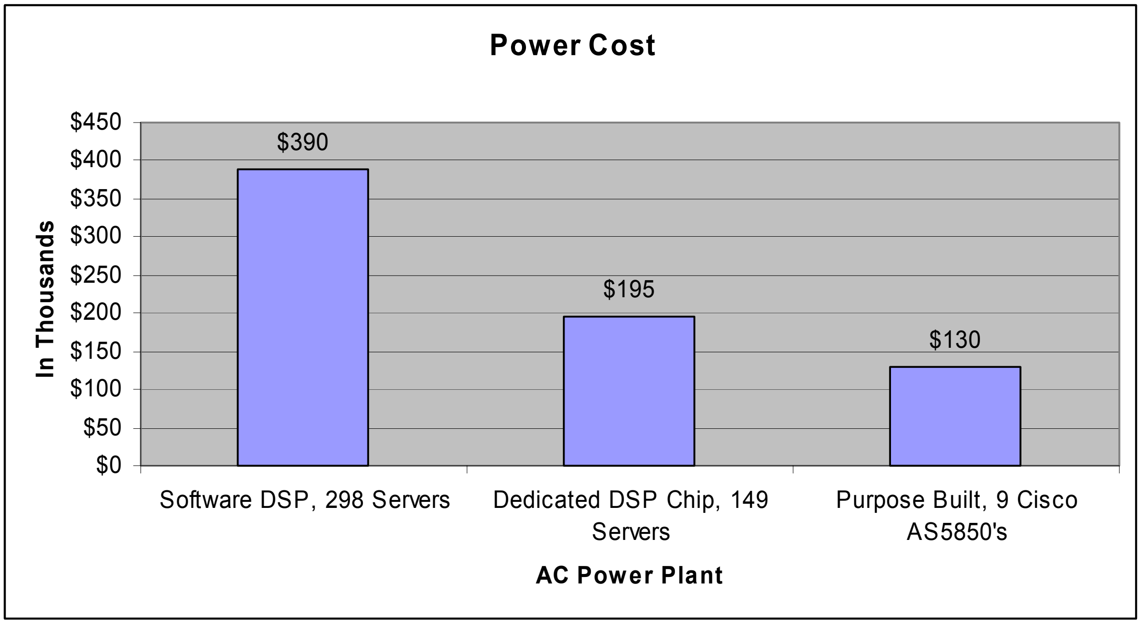
Also, let us look at a dedicated platform for this PSTN gateway purpose, the Cisco AS5850. This device uses dedicated DSP chip based technology to convert TDM analog voice to packetized voice. A fully loaded chassis can support 3360 channels (5 CT3's). With this in mind, 28,572 trunks required, we would need 8 1/2 AS5850's. After searching Google, I found several resellers offering AS5850 fully loaded for sale. The lowest price I found was for $175,000. This comes out to be $52 per port ($175,000/3360 ports). Total cost for this device to perform the PSTN gateway function for our VoIP platform would be ~$1,513,000. This is still 1.8 times the cost of using Asterisk servers and Quad T1 cards.



Space and Power

Space requirements for this exercise are as follows: 298 servers for the software DSP model (8 racks), 149 servers for the dedicated DSP chip model (4 racks) and 9 Cisco AS5850 (3 racks). Rack kits for blade servers and telecom equipment range from $500 to $1,500 depending on shape and purpose. Averaging the rack cost to $1,000 we see that space will probably be more critical than cost. Added into the overall cost of each solution, the rack hardware is negligible. On the other hand space availability has to be considered on a case-by-case or space-by-space basis.

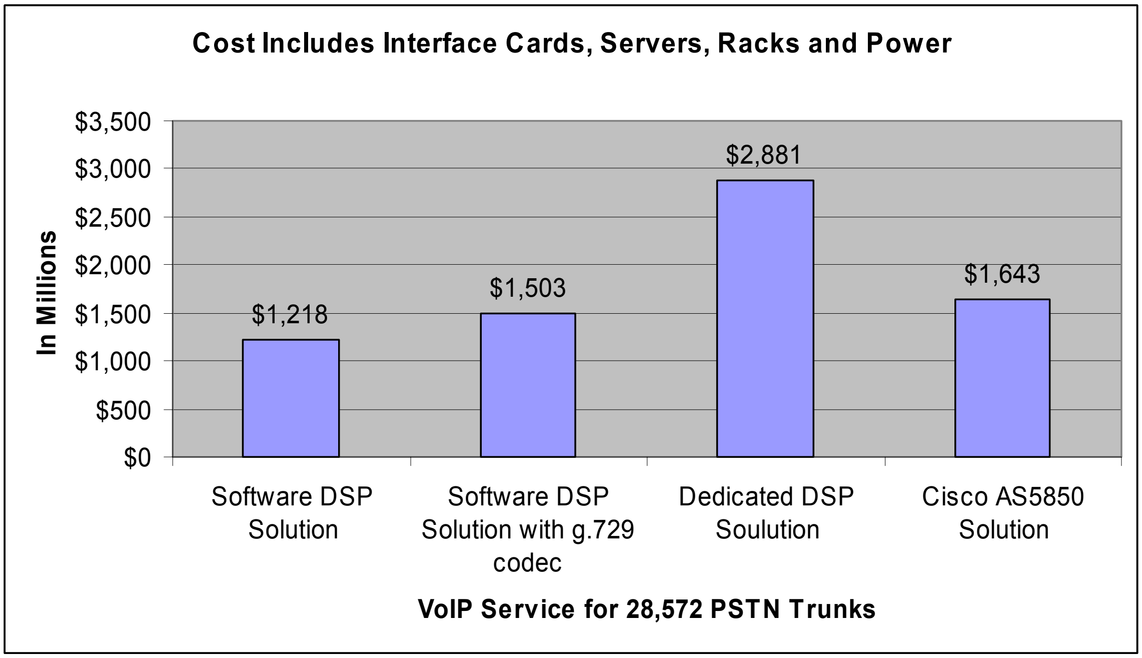
Power requirements break down as follows: 298 servers @ 300 watts per server equal 89.4KW, 149 servers @ 300 watts equals 44.7KW and 9 Cisco AS5850's @ 2.4KW equals 21.6KW. 2 years ago we augmented our collocation facility power plant and added 30KW of AC power. This cost us $130,000 but we had many extenuating circumstances that drove up the cost. Using this figure as a baseline, I'll estimate the cost to power 298 servers @ $390,000, 148 servers @ 195,000 and 9 Cisco AS5850's @ $130,000.



Codec Cost

Depending on the codec used can drive up cost per line when using the software DSP model. Currently the highest compression (lowest bandwidth utilization), best audio quality codec is G.729. Digium licenses this codec for $10 per channel. To use this codec in this exercise with 28,572 trunks, this would add an additional cost of $285,572 to the software DSP model. Ultimately Network architecture and available upstream bandwidth from the client will determine codec usage and will drive the decision to use a high quality, high bandwidth free codec or one that cost a bit and takes advantage of lower bandwidth.

Exercise Conclusion



This cost exercise was intended to show the potential cost differential between the two methods and gain interest in the software DSP function as a cost effective method of converting TDM and Analog voice to packetized voice. Ultimately I will show Asterisk as viable software to accomplish this task.

IMHO, I feel there are no true arguments concerning voice quality between the two methods, when both are implemented properly. I will say, in my experience that dedicated hardware DSP VoIP can be more forgiving where latency and jitter are concerned. When the network is owned and managed end-to-end, latency and jitter can be easily controlled and reduced.

Architecture and Integration

In a carrier environment, segmenting the functionality of service is preferable to facilitate a growing customer base. This is seen is all large scale deployments of many service offerings such as voicemail, e-mail, web hosting, authentication, conferencing, IVR and so forth. In a small enterprise, putting all services onto one server works well but ultimately is not scalable in the most efficient manner.

The VoIP platform in a large scale carrier environment is no exception. The PSTN Gateway will perform just that function. The voicemail system will be on dedicated servers with tape backups. The conferencing services could reside on separate servers with dedicated trunk access. The customer aggregation servers will handle registration, dial plan lookups and some advanced features, call hold, call transfer, etc. The Internet Gatekeeper will be fire-walled coming and going. We will look closely at the PSTN Gateway function and briefly discuss the application and registration servers.

PSTN Gateway

This is where we get the most bang for our buck. In a green field environment, the VoIP PSTN Gateway would back into the access network through SS7 links. In this environment of an existing Class 5 Switching network, SS7 links are presumably in place and have capacity to grow the line side requirements needed for our soft switch implementation.

The line side links from the Class 5 Switch (we will call it the TDM Switch), will be the trunk side links into the soft switch PSTN Gateway. See figure 1. There are three methods we will discuss to egress the TDM switch with line side trunks and the effects each will have on the call handling for the VoIP platform. PRI's are a good way to over subscribe the channels with more DID (Direct Inward Dialing) telephone numbers than the actual T1 can handle. Also multiple PRI's can be grouped together to produce greater efficiencies in relation to DID TN's. GR303 is another method of over subscribing physical T1 channels and can also be grouped together for fault tolerance. Asterisk supports GR303 trunks as of Jun 1, 2004. If the carrier already has a GR303 aggregator, this can be placed between the TDM Switch and the VoIP PSTN Gateway. The last method is a channelized T1, one two-way TN dedicated to a single channel. This is our most inefficient manner but is within the scope of this paper.

SS7 is the standard for connecting telephone service providers to the PSTN. Currently Asterisk does not support SS7 trunking but the OpenSS7 community and code writers within the Asterisk community are engaged in a project to implement this signaling set into Asterisk. The evolution of Asterisk continues so stay tuned.

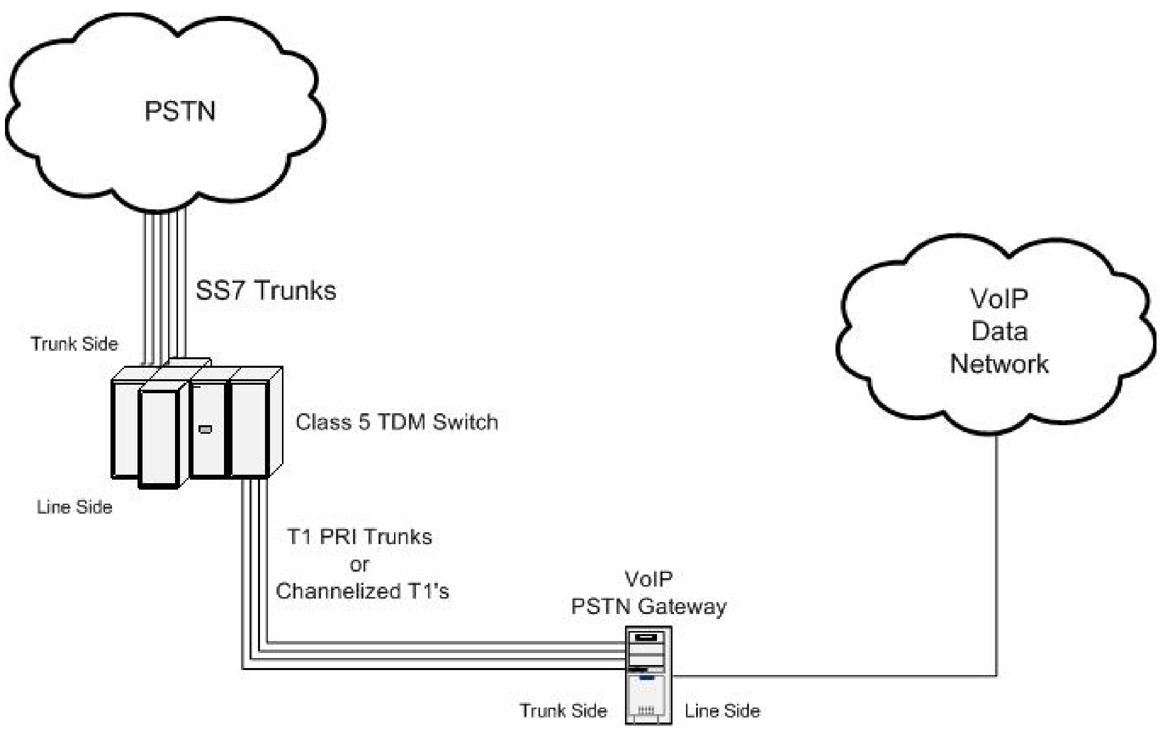


Figure 1

PSTN Gateway PRI's

Bringing in PRI's into the VoIP PSTN Gateway is a great way to efficiently use trunk capacity within the soft switch but can make for a complex dial scheme in the VoIP environment. Also, bundling multiple PRI's into a single gateway server help for any single T1 failure, traffic will still pass on the remaining PRI links, just capacity to the server is reduced until repairs are made.

Provisioning Asterisk to route calls in and out of a PRI can get complex as there are no dedicated channels per TN. In a one box PBX this is not a problem but PRI's being dynamic in nature can wreak havoc in a large scale environment with respect to record keeping and provisioning new service. This increases TCO to manage and maintain but is a trade-off for lower port density. In a large scale build-out, lower port density is good but constant provisioning of adding dynamic trunk groups to the soft switch can quickly eat up any cost savings on equipment.

Using PRI's into soft switches is a good way to manage line side trunk volume, redundancy and restoration during failure.

PSTN Gateway GR303

GR303 is another trunk protocol that takes advantage of over subscribing physical channel capacity per T1 or T1 group. This is good for the TDM switch as it reduces the number of physical T1's needed and offers a great scaling model. This does not require a GR303 aggregator in between the TDM switch and the soft switch as Asterisk does support this trunk protocol. An aggregator in between the TDM switch and Asterisk can enhance the ability to produce a static VoIP gateway provisioning model. See figure 2. Speaking to the cost of this device, the technology has been around for many years and a multitude of Venders manufacture competing equipment. Price for up to ~ 600 T1 aggregations can be in the sub $20,000 range. In the over all scheme of using GR303, this expense is worth it as it simplifies the provisioning task of the soft switch platform.

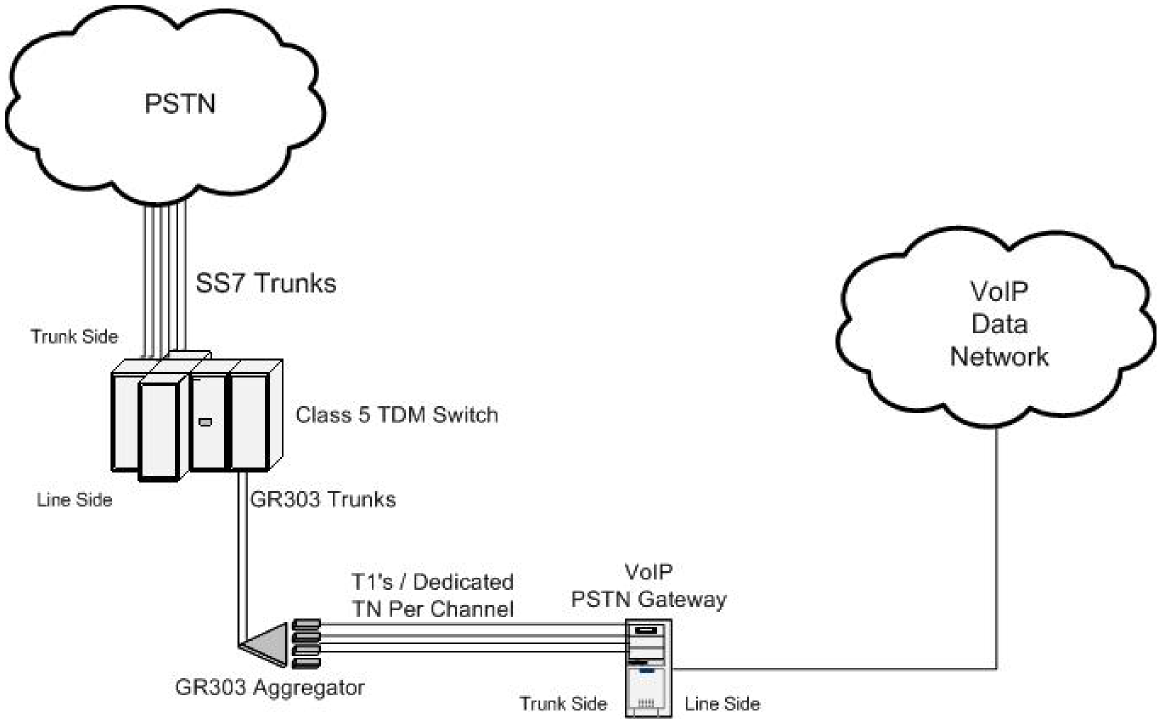


Figure 2

Soft Switch Provisioning Model

Bringing new service offerings into a company requires provisioning of those services. How automated the provisioning is will determine how much human or machine interaction is involved with actually turning up a new customer. In an existing Telco, many devices and systems already have provisioning routines for daily service turn up and disconnect. Adding new VoIP services in the middle can be complex and daunting setting up the right picks in and out of data bases, adding extensions, configuring features, creating dial plan call routing, maintaining security, etc. If we use a 1 TDM channel to 1 VoIP channel provisioning model, our VoIP platform can be brought into our existing infrastructure with minimal impact to daily provisioning tasks. When you turn-up a customer TN in the Class 5 switch, that extension or TN is bonded to the VoIP channel associated with its location on the ingress T1 on the VoIP PSTN gateway platform.

With this concept in mind we can imagine the provisioning of the VoIP environment as a static event in relation to our existing network. The VoIP provisioning scheme can be setup prior to actually turning up customers. Then when a new customer needs to be turned-up, the normal Telco provisioning events take place and pre-defined channel association with the VoIP platform dictate the CPE equipment registration and channel assignment, at least through the VoIP PSTN gateway. Pre-defined and provisioned accounts with varying applications can be setup ahead of time. Since the physical port to channel relationship is static, the configuration files can be backed up and restored quickly if a gateway element fails.

Application Servers

Asterisk follows modular software design as does Linux. Applications within Asterisk can be turned on and off, started and stopped, loaded and unloaded without impact to the overall system or other applications. Asterisk can easily be setup as a monolithic application server performing specific tasks. As a standalone PBX, Asterisk loads and runs most modules and applications to perform many tasks. As a task specific server, if a module, application, codec or protocol is not needed then Asterisk can be configured not to load modules that will not be uses, thereby increasing platform security, stability and reliability. When new voicemail, conferencing or other applications are written, code changes will not affect other modules on that server due to other modules not loaded or used.

In a server farm arrangement, specific applications servers can grow and scale as they are needed. If the majority of customers do not require voicemail hosting, then the voicemail application servers would not grow proportionate to customer aggregation servers or PSTN gateway servers. This gives the overall VoIP platform greater flexibility when scaling applications and controlling equipment cost. Also, different applications have different CPU loading properties. With low processor loading and or volume of customer usage, applications can be concentrated or expanded matching available hardware resources. For instance, a dual 2.8 Gig Xeon processor machine may handle 96 simultaneous analog to digital voice channel conversions in the role as a PSTN gateway server, this services 96 customers. The same server may handle 500 voicemail customers in the role as a voicemail application server.

Asterisk has gotten some notoriety as voicemail and conferencing servers backed into other Vendor solutions such as Cisco Call Manager allowing cost savings over a Unity implementation. Also Asterisk is being used by some Internet Telephony companies (unnamed) for IAX registration gateways.

Customer Aggregation Servers

Aggregating VoIP customers can be accomplished in several ways with different protocols and configuration methods. The flexibility that comes with using Asterisk allows active or static provisioning of new customers. When using dedicated and secured CPE devices, Asterisk aggregation or registration servers can be pre-provisioned statically that contains entries pre-mapped to each CPE device. Those static maps can flow through to specific voicemail boxes and PSTN Gateway ports. With a static configuration scheme across the entire VoIP platform, each CPE device can be pre-mapped to a specific DS0 at the PSTN Gateway and ultimately on the TDM network. Realistically a provisioning plan can be put in place that will accommodate turning up new customers without the need to provision the VoIP platform. When each CPE port corresponds to a switch port or DS0, existing provisioning methods can be maintained and the new TN assigned to that DS0. With mild data-basing functions cross-referencing TDM switch DS0 mapping to VoIP CPE ports, minimal impact to overall provisioning routines can be realized.

Asterisk also has the ability to segment or pool customer types or quantities of users into their own dial plan configuration such as an ACD or IVR. Asterisk can handle multiple customers aggregated on the same server, all with different dial plans and call handling routines.

The aggregations servers perform the task of registering customers and must hold the call in session when individual CPE devises are used. This is so the advanced features are usable on the customer side of the call. When registering through IAX with an Asterisk server at the customers premise, that server can handle the advanced features.

Aggregation servers can be setup to accommodate specific protocols such as SIP, IAX, MGCP or H323. Also this allows for securing each protocol in the best manner using firewall specific protection.

Dial Plan Servers

The dial plan servers are the core switching unit for the VoIP Platform. With Asterisk PSTN Gateways, Customer Aggregations Servers and Application servers, IAX can be utilized within the VoIP core to keep call setup and transfer to a minimum. The servers themselves should be configured in a hardware and software redundant arrangement with active heartbeat monitoring and self diagnosis for real-time fault condition switching of load.

Within the dial plan server, the IAX protocol natively releases the call after forwarding the call to another Asterisk servers running IAX. This allows the dial plan server to maintain the database, route the incoming call to and from the appropriate Aggregation, PSTN and Application servers, then drop out of the path to handle more calls.

Asterisk can actively reload dial plan configuration files without the need to restart the service. When all dial plan configuration and call routing is handled within the dial plan servers, this precludes the need to configure and reload Customer Aggregation or PSTN Gateway servers for add, move and changes.

The Asterisk dial plan server can be optimized for housing databases locally or pull dial plans remotely from network storage. GUI front ends can be developed to assist the process of adding new call control statements or making changes to the call routing environment.

Operation and Maintenance

Operating within this environment would demand two, maybe three VoIP specialists given the quantity of servers, applications and integration with existing infrastructures. One of the first tasks is to create a dial plan scheme, predicated by best practices of provisioning in the host environment and integration techniques used to connect the existing infrastructure.

Maintaining the VoIP platform would consist of keeping track of software development that adds features and or fixes bugs. Keep an eye on log files and monitor health of the VoIP servers. Track usage on application servers and resource loading across the platform. This allows planning for increasing capacity when needed.

Software and Hardware Improvements

Future plans for Asterisk include increased port density on TDM cards, creating a DS3 TDM interface or working with an equipment manufacturer to build a high density, low cost unit that integrates seamlessly using the IAX protocol. SS7 protocol integration would allow Asterisk to back into direct PSTN trunks from an ILEC. Integration with presence systems such as Jabber, AIM, MSN or ICQ would greatly enhance when and where to receive calls. There are already software clients that integrate the PC desktop with the IP Hard Phone registered to Asterisk.

Because Asterisk is written in the C programming language, there are many individuals capable of developing new applications and management interfaces to integrate with Asterisk.

Paper Conclusion

The state of technology within the telecommunication field has never been static and shows no signs of ever being static. Relying on equipment Vendors to increase the product portfolios of services provided to customers has been the norm within this industry. If the ILEC, CLEC or MSO is large enough, demanding features to be built into equipment can usually be accommodated by equipment Vendors, but the bottom line is they still own the technology and the features. Historically, Service Providers have touted themselves as technology companies, the reality is they provide services built and delivered by other Vendor's technologies.

Asterisk is a project that delivers into the hands of the novice and experienced alike, a platform for developing telecommunication applications that bridge the gap between the existing TDM technology and the future VoIP technology. Asterisk is at a price point agreeable to a small business and quite possible robust enough to deploy in a carrier environment to deliver VoIP services to a very, very large customer base.

Pictorial Overview of Integration Model