POINT OF VIEW

Common VoIP Architecture

Executive Summary

This white paper describes the architecture of AT&T's common infrastructure for real-time communications services over Internet protocol, commonly referred to as VoIP. The infrastructure will be used to implement all planned and future real-time services that use IP as their main transport technology. Existing voice services will migrate to this architecture over time and when appropriate.

AT&T is migrating to a single Multiprotocol Label Switching (MPLS)-enabled IP core network in support of a major initiative, the Concept of One, which will ensure optimum use of resources by maximizing commonality. All services will eventually be carried over this IP network. The common infrastructure described in this paper is limited to the essential capabilities and components that will enable AT&T to develop and deploy real-time services, such as voice, interactive text and video, and various multi-media collaboration services.

The main goal of the AT&T VoIP architecture is to provide a single, common, and shared infrastructure that facilitates the development of real-time services with the highest quality and availability, the shortest possible time-to-market, and the lowest cost of operations and maintenance feasible. To accomplish this, the architecture is divided into separate independent layers. Each layer has a well-defined role and provides a set of capabilities for the layer immediately above it by utilizing the set of capabilities provided to it by the layer immediately below it.

The VoIP infrastructure is built as a virtual network on top of AT&T's Converged IP/MPLS Core Network and is called the VoIP Connectivity Layer. The Core Network is surrounded by a Multi-Service Access/Multi-Service Edge network that supports all popular access technologies including TDM, ATM, Frame Relay and Ethernet. Thus, the VoIP infrastructure can be reached via any of these access technologies. Moreover, the architecture provides capabilities to support various access protocols such as, H.323, MGCP, MEGACO, SIP, TDM/SS7, as well as any VoIP protocol may emerge in the future. This is achieved by surrounding the VoIP Connectivity Layer with Border Elements (BEs), which mark the trust boundaries of the VoIP Infrastructure and translate the specifics of various VoIP access protocols into Session Initiation Protocol (SIP) – the single common internal protocol used by all VoIP infrastructure components. BEs not only provide protocol conversion, but also enforce various policies including those needed for call admission control and VoIP-level security.

The Call Control Element (CCE) controls and manages the VoIP infrastructure and provides a single interface to application servers residing in the Applications Layer. Working with various BEs, the CCE creates and removes call legs and joins call legs to establish connectivity between end-points. The application servers residing in the Applications Layer provide the service logic capabilities to implement various types of services. The "plug-in" paradigm is used to allow easy addition of various services without impacting the VoIP Infrastructure. Vendor-provided application servers as well as internally developed service logic can both be used in a similar fashion.

Concept of Zero is followed to provide "zero-touch" provisioning and maintenance from a single virtual operations center.

Introduction to VoIP Layers

The VoIP architecture is devised in a series of Layers, as illustrated in Figure 1. The layers follow the principle of information hiding; that is, each layer has a well-defined role and provides a set of well-defined capabilities to the layer to which it is sending information, by using the capabilities of the layer it received information from.

Principles
AT&T's VoIP Common Architecture

- Use a common architecture for all real-time communications services
- Ensure that high reliability, availability, and performance is comparable to AT&T's circuit-switched network
- Support both existing and future services
- Provide for rapid deployment of new services
- Maximize usage on a single, shared set of resources
- Accommodate multiple popular access technologies
- Use SIP as the common internal signaling protocol for all signaling between components
- Leverage the latest, cost-reduced technologies
- Allow only Open Standard Protocols for all internal interfaces
- Ensure that components are interchangeable and interoperable
- Design performance and reliability into all components, services and processes
- Provide for the necessary security to protect AT&T and customer assets
- Support "users" in addition to "lines"
- Ensure that capacity closely tracks demand

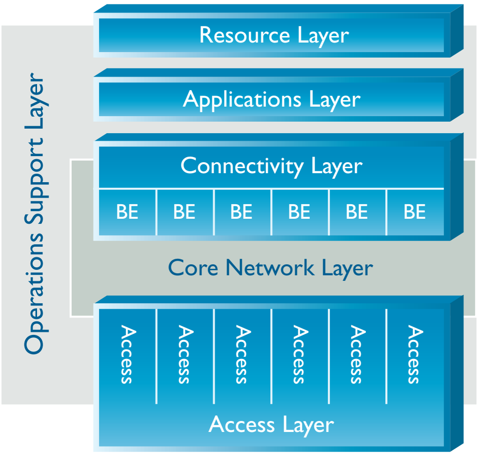

Figure 1: VoIP Layers

The Access Layer interfaces with customer equipment and provides customer connectivity using access-specific protocols.

The IP/MPLS Core Network Layer provides IP connectivity for all elements of the VoIP infrastructure.

The Connectivity Layer provides the VoIP infrastructure needed to process basic calls, support high performance network functions, send network primitives, provide media services, interact with Application Servers for more advanced calls, and support Call Detail Recording. Border Elements translate the access protocol to SIP.

The Applications Layer consists of several Application Servers, each providing one or more services.

The Resource Layer provides an environment for the creation of service logic and the management of services including customer record maintenance and billing plans.

The Operations Support Layer consists of multiple applications, databases and a supporting data communications network, which are used by internal and external personnel to manage the VoIP network and its elements.

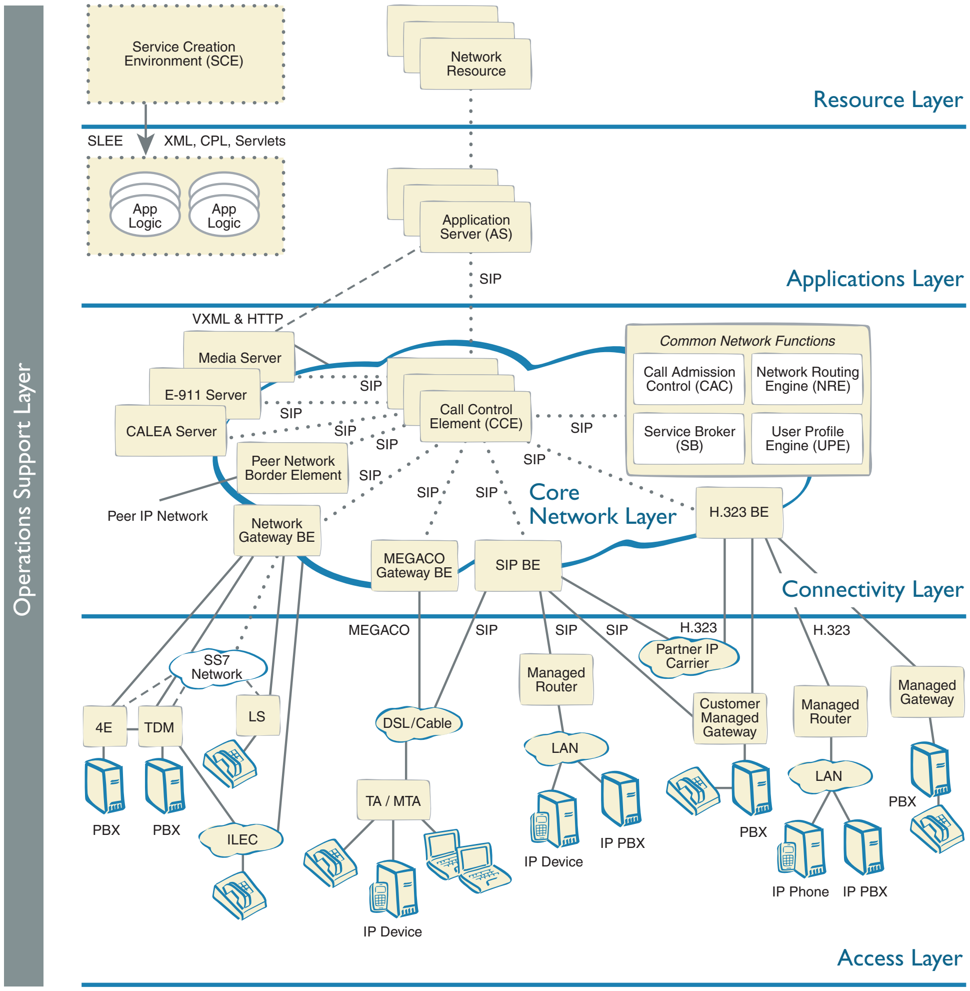

Figure 2: AT&T VoIP Functional Architecture

Functional Architecture

In Figure 2, layers are represented in a high-level AT&T VoIP functional architecture. This is a logical functional architecture and is not intended to represent the physical implementation. The architecture does not specify how functions will be distributed to devices. Instead, it allows the flexibility to package functions into various servers using cost-effective methods, as long as open interfaces to those functions are available for use by other functions.

For each access type, a corresponding Border Element (BE) is deployed to manage the access-specific requirements and interfaces. BEs are the demarcation points for the Connectivity Layer, which is a common and shared layer on top of the converged IP/MPLS network, or Core Network Layer. The BEs translate access protocols to SIP, then provide call details to the Call Control Element (CCE). The CCE manages the VoIP infrastructure and creates, removes and joins call legs. To provide a service, the CCE invokes an Application Server (AS), using SIP to communicate. By sharing a common architecture, new access technologies can use all existing and future ASs, and new ASs can support all existing and future access technologies.

This architecture supports all real-time communications scenarios, like prepaid card, click-to-chat, and teleconferencing. For a simplified basic call flow, the following steps are performed:

1. The caller's phone connects to the BE using the appropriate access-specific signaling protocol.

2. The BE sends a SIP INVITE to the CCE with Request-URI for the destination's phone number. The CCE consults with the Service Broker, which indicates that this call has no features.

3. The CCE sends an INVITE to the destination's BE, which communicates with the destination's device using the appropriate signaling protocol.

4. The call is set up between the two end-points, and the caller and destination talk.

5. The destination hangs up. The BE sends a SIP BYE to the CCE.

6. The CCE sends a BYE to caller's BE, which disconnects the caller.

Access Layer

The Access Layer provides connectivity between the Customer Premises Equipment (CPE) and the BEs. The Access Layer must support the endusers' service requirements such as Quality of Service (QoS), Security, and Availability. Current technologies will be supported allowing AT&T to continue to offer existing profitable services at a low cost.

The VoIP architecture is designed to provide customers with menus of services that are independent of access technologies. In the past, new access technologies were managed differently. With this architecture, new access technologies will be deployed quickly and efficiently without disturbing the existing VoIP infrastructure.

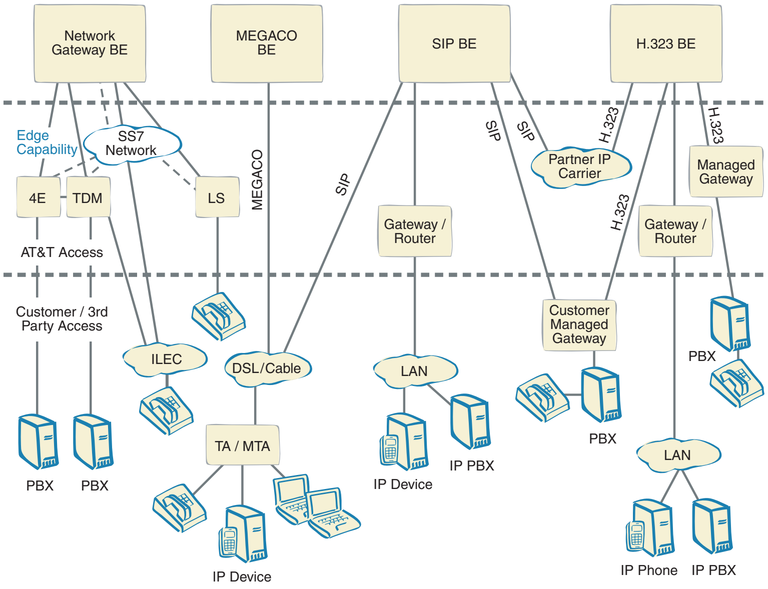

Figure 3: Access Layer

Access Methods

Ubiquitous access to customers is achieved with an Access Layer based on a combination of AT&T and 3rd party segments. While customers in large enterprise locations can be economically served with dedicated access (Leased Line, FR, ATM, etc), customers in home offices and small branch offices need to be served by shared access methods (ILEC Local Service, Cable Modem, DSL, etc). To meet these widely varied requirements, the Access Layer is composed of a variety of access methods:

- **TDM via Edge Switching –** Customer access occurs via existing connections to AT&T Edge switches. This includes shared access from the ILEC and dedicated access to 4E or 5E switches. Bearer and SS7 signaling for Edge switches interface with the Network Gateway BE. In the future, the Edge switches could support an IP interface, eliminating the need for TDM to IP conversion in the BE.
- **TDM Direct Access –** Customer PBX or ILEC switches could interface directly to a Network Gateway BE. Full range of SS7, ISDN, CAS signaling needs to be supported.
- **TDM via IP Direct Access (Managed Router/Gateway) –** A direct-access line shared with an AT&T IP Data Service, (e.g. MIS, MDNS, MRS, EVPN) in conjunction with a VoIP Gateway on the CPE Managed Router, is used to provide a TDM service demarcation on the customer's premises. QoS implementation is at Layer 3 for direct IP Access, and at both Layers 2 and 3 for FR/ATM access.
- **IP MIS (Managed Router) –** A direct-access line shared with an AT&T IP Data Service, (e.g. MIS, MDNS, MRS, EVPN) in conjunction with a CPE Managed Router, is used to provide an IP service demarcation for VoIP service on the customer premises. VoIP protocols supported include SIP for SIP Phones or SIP Proxies and H.323 for IP PBXs. QoS implementation is at Layer 3 for direct IP Access, with possible additional Layer 2 QoS features for FR/ATM access.
- **IP Broadband Agnostic –** Customer-purchased Broadband Access, (e.g. Cable, DSL) in conjunction with an AT&T-provided Terminal Adapter, is used to provide Remote Office or second-line service to individual end users. End user devices can be SIP Phones, SIP Clients, Black Phones or other IP devices. For Black Phones, the AT&T Terminal Adapter will convert to VoIP. QoS is achieved by queuing mechanisms implemented in the Terminal Adapter. Privacy is achieved by IPSec sessions between the Terminal Adapter and BEs.
- **IP Peering –** AT&T provides or purchases VoIP minutes over an IP Interconnect with another VoIP carrier. Currently, the VoIP signaling protocol is H.323. Potential relationships with cable operators may introduce CMSS, a PacketCable variant of SIP. Encryption for Privacy is an option.

Customer Premises Equipment

Customer Premises Equipment (CPE) are end-point VoIP devices at customer locations. Examples include gateways, IP PBXs, and IP phones. AT&T selects and certifies CPE to work with AT&T's VoIP architecture. CPE interfaces will comply with standard protocols such as SIP, MEGACO, MGCP, and H.323.

CPE Functional Components

Signaling – CPE implement supported access protocols (H.323, MGCP, MEGACO, SIP). CPE are typically the end point devices for originating or terminating signaling and/or media streams.

Media Control – CPE terminate the media streams going in and out of the Connectivity Layer and provide for final media format conversion.

Core Network Layer

CPE Types	
H.323	Implements a gateway or terminal function for access.
MEGACO/MGCP	Implements a MEGACO/MGCP end-point for access.
SIP	Implements a SIP User Agent for access.
TDM	Connects through an NG BE to access VoIP service.

The primary function of the Core Network Layer, also known as the IP/MPLS Converged Network, is to provide high performance intersite routing and transport of IP packets. In support of the Concept of One, the Core Network Layer is a single network that provides IP connectivity for the VoIP infrastructure. The Core Network also has the added value of architecturally isolating the BEs and the associated Access Networks they connect with, from the CCE, and associated service intelligence. The design facilitates inter-working among end-points in different access domains. That is, a multi-location customer could have a single voice-oriented service that works in the same way at all of the customer's locations regardless of how network connection is achieved. (e.g.,TDM, IP, ATM, Frame Relay, or remotely via a service partner like a cable company.)

IP Quality of Service Capabilities

In order to address the performance needs of each of the typical traffic streams associated with the VoIP Architecture (bearer channels, signaling, management traffic, etc.), the Converged IP/MPLS Network supports tuning of per-hop behaviors within the network. These behaviors are consistent with the performance needs of four classes of traffic flows, including Real-Time, High Performance Data, Medium Performance Data, and Best-Effort. This includes setting up priority queues, where appropriate, to support traffic that has real-time characteristics (i.e., bearer channel traffic). It also includes allocating the remainder of the queue available on those links to three distinct classes of service (high, medium and best-effort data traffic), with servicing characteristics existing so that the lower classes cannot overwhelm capacity reserved for the higher classes, when contention occurs. The equipment signals to which of the four traffic classes to map into by marking IP Precedence settings within the IP Type of Services (TOS) header (also known as Packet Marking). In the future, the Converged IP/MPLS Network will support explicit Admission Control mechanisms.

Security

By design, the VoIP Architecture is intended to be resilient to known forms of security attacks. This includes both simple attempts to steal service, as well as brute force distributed denial of service (DDOS) attacks intended to dramatically reduce the Quality of Service (QoS) provided to customers.

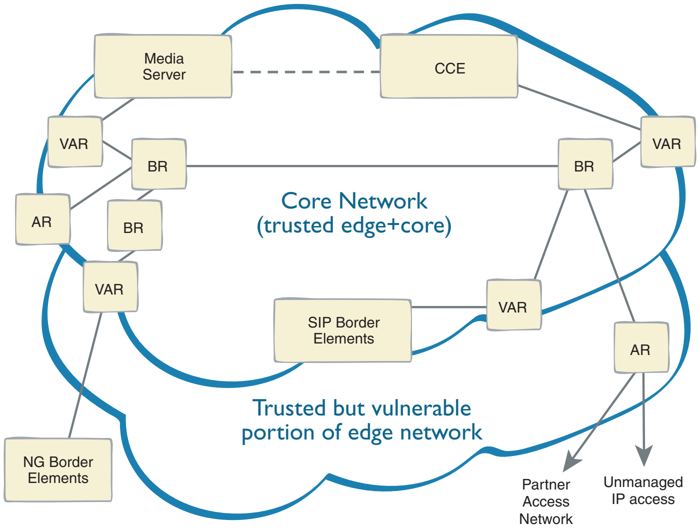

Figure 4: VoIP Security Framework

The definition of a core trust domain where VoIP Architectural Elements participate is fundamental to this design. Also, the security design assumes that the BE will act as a gateway into that trust domain, from customers who are natively attached to networks outside of that trust domain (e.g., partner networks, customer premises networks, etc.).

In the early stages of this architecture, these trust domains may need to be implemented by classifying all of the traffic on a given interface to be part of a particular type of VPN. In the target implementation, MPLS VPN will be used, which will both improve the flexibility of VPN management, and allow the components in the other layers of the VoIP Architecture to mark different traffic streams (via labels) that belong in different VPNs on the same physical interface.

Connectivity Layer

The VoIP Connectivity Layer provides all network primitives needed for applications to implement services. This includes establishing simple connectivity between end-points by providing capabilities to create, join, remove (tear down) and report the status of call legs. The following services are also enabled in this layer: Basic media services, E-911, CALEA and Call Detail Recording.

The Connectivity Layer provides a unified, shared environment that supports the addition of new services and access technologies without changing the basic infrastructure.

This layer is comprised of a number of zones and a collection of network functions. A zone consists of a CCE managing one or more BEs. Network functions and resources, as described in this section, are shared by CCEs across many zones.

Border Element

The Border Element (BE) is the point of demarcation for the Connectivity Layer. It identifies the Boundary of Trust for the AT&T VoIP Network and provides an entry point into the VoIP infrastructure. For each customer site, BEs will ensure reliability and availability.

Connectivity Layer Definitions	
Border Element (BE)	Translates an access protocol to and from SIP, enforces security, enforces admission pollicies in coordination with CAC, detects DTMF and other types of in-band user requests and redirects media when instructed by the CCE.
CALEA Server	Responsible for the interception of communications for law enforcement and other purposes.
Call Admission Control (CAC)	Responsible for the enforcement of network-wide admission policies.
Call Control Element (CCE)	Manages and interacts with all BEs in a zone, sets up and tears down call legs and interacts with ASs to implement their requests.
E-911 Server	Routes emergency calls to the appropriate Public Safety Answering Points, (PSAPs) based on the caller's location.
Media Server (MS)	Handles and terminates media streams, providing functions such as announcement, DTMF, text-to-speech (TTS), automatic speech recognition (ASR), bridging and transcoding.
Networking Routing Engine	Provides route information and translates a network address or logical address to an IP address.
Service Broker (SB)	Provides the CCE, upon request, the address of the appropriate AS to treat the call.
User Profile Engine (UPE)	Contains user data, including registered devices, presence information and geographical location.
Zone	Consists of a set of BEs managed by one CCE.

As the interface between AT&T's internal address space and the customer site's address space, the BE is responsible for recognizing and authorizing end-points (IP-PBXs, SIP phones, etc.). Functionally, the BE can be logically decomposed or distributed, if needed. A configuration is supported in which the media-related processing of the BE takes place at a trusted CPE. The signaling function of the BE always occurs in an AT&T network location.

BE Functional Components

A BE has four main functions: Signaling, Media Control, Security and Call Admission Control.

Signaling – A BE proxies both the caller and the called end-points, thereby providing a point of signaling control at the edge. It translates the access protocol (H.323, MGCP, MEGACO, SS7, CAS, ISDN, etc.) to and from SIP.

Media Control – A BE examines all media streams going in and out of the Connectivity Layer for security, media format conversion, and media transfer. The BE also redirects media streams upon request from the CCE without impacting the actual caller and called party, and provides the means for the CCE to define, detect and report DTMF strings during the call.

Some services require the ability for the AS to detect a signal that does not need to be on the media path (e.g. DTMF, flash hook). For example, a prepaid card application may permit the caller to enter a sequence of digits (e.g. "**9"). This forces a hang-up of the destination and provides the opportunity for another call to be placed. To enable efficient utilization of network resources, the BE will allow an AS to register event triggers via the CCE, and the BE will signal the AS when the event occurs. While there is not currently a standard mechanism for accomplishing this task, AT&T supports the IETF proposal, using HTTP to send DTMF signals.

Security – A BE provides all necessary security and screening for the customer sites it interacts with. It authenticates subscribers, customers, and partners, and provides NAT and firewall functions as appropriate.

Call Admission Control – A BE enforces admission policies over the access pipe such as:

- Call gapping – limiting the call set up rate
- Call limiting – limiting the number and type of calls
- Bandwidth management – ensuring media bandwidth being sent matches the profile negotiated through signaling

Furthermore, the BE uploads local policy information via policy protocol or Operations Support, and keeps track of resources for access networks.

BE Types	
Network Gateway (NG)	Provides all required PSTN interface functions, including SS7 signaling and media conversion. The NG BE also operates as a gateway between circuit-switching and packet-switching networks. From the CCE, it acts as a SIP User Agent (UA).
H.323	Implements gatekeeper-routed signaling for H.323 CPE gateways, from the access side. From the CCE, it acts as a SIP UA.
MEGACO/MGCP	Implements Call Agent (or Media Gateway Controller) signaling for MEGACO/MGCP devices, from the access side. From the CCE, it functions as a SIP UA.
SIP	Implements SIP Proxy and back-to-back user agent (B2BUA) functions for SIP devices.

Call Control Element

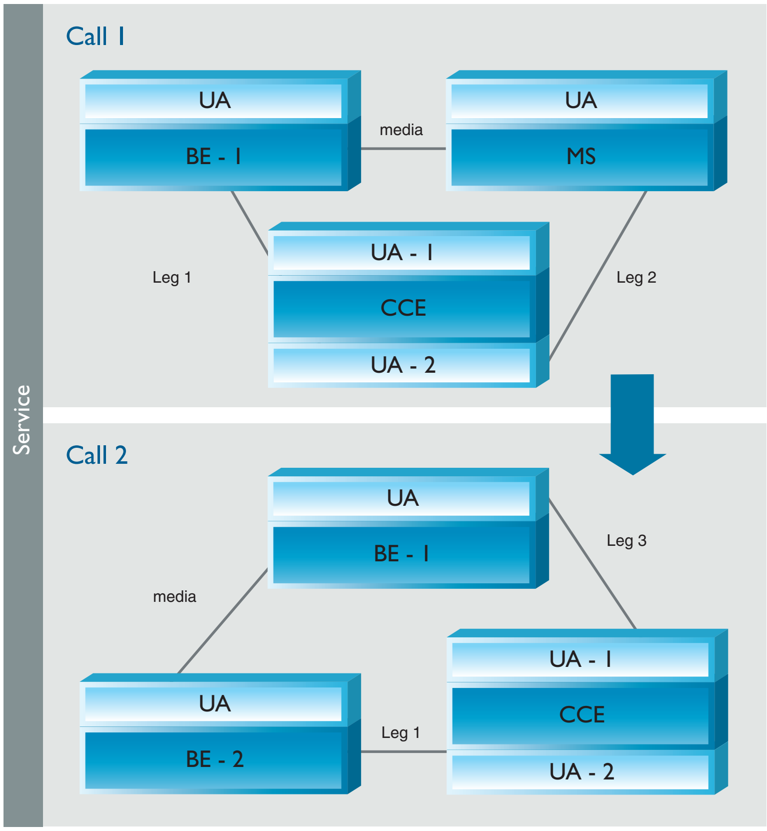

Figure 5: Call Leg Relationships

The Call Control Element (CCE) is responsible for providing a call leg view of the Connectivity Layer to the ASs. Working with the BEs, the CCE completely hides the specifics of the Access and Connectivity Layers from the ASs; allowing them to deal with logical end-points (users), call legs, and calls terminating at logical end-points. An AS can ask the CCE to establish a call leg to or from a logical end-point (user), to join call legs to form calls, and to tear down call legs and calls. The CCE does not need to interact with an AS to serve basic calls with no service features.

The CCE functions as a SIP Back-to-Back User Agent (B2BUA), and is a signaling end-point for all call legs with all BEs. Media paths, however, are established directly between BEs without going through the CCE.

The CCE may receive a "call invocation" (SIP INVITE) from any BE on behalf of a subscriber, or from any server inside the network. An invocation has information about the caller, the called party, and characteristics of the call (class of service, media type, bandwidth, compression algorithm, etc.). The CCE may instruct a BE to redirect the media channels associated with a call to a different destination.

A CCE manages all BEs in its zone. It is aware of the status of each BE it manages, including whether the BE is operational or congested. Moreover, the CCE enforces various routing policies, such as deciding which BE to use to set up a call leg.

The CCE interacts with the Service Broker (SB), the Network Routing Engine, (NRE), the User Profile Engine (UPE), and the Call Admission Control (CAC). In addition, the CCE also communicates with the resource servers residing in the Connectivity Layer, such as MS, E-911, and CALEA.

The CCE interacts with the SB to determine the services associated with a call leg (either in-bound or out-bound). For each call, the CCE provides relevant information and requests the AS to execute the corresponding service logic.

The CCE interacts with the NRE to translate a network address (such as an E.164 voice network address) or logical address to an IP address. The CCE uses the physical address of an end-point when considering BE availability, load, etc. A physical address consists of the IP address of a BE and all other information needed by the BE to locate the end-point in the address space it manages. Physical address characteristics include:

- A local IP address, when a BE manages an IP address space
- Identification tokens, when a peer network is involved
- The end user's telephone number for a PSTN

The CCE may also consult with the NRE to learn the terminating BE's address. If the terminating BE is not in the zone of the originating CCE, the NRE returns the address of the CCE in charge of the terminating BE. Then, the originating CCE consults with the terminating CCE, which functions as a SIP redirect proxy, to learn the address of the terminating BE.

The CCE interacts with the UPE to determine if a user is registered or logged-on, and to determine which address or number should be used to route the call.

The CCE interacts with the CAC to authorize users and admit calls at the time of call setup, leveraging network-wide conditions and policies. The CCE formulates the QoS/Bandwidth/SLA policy for network-wide resources in conjunction with the local policy and customer-specific requirements provided by the BE.

In the future, policy information will be obtained from the lower IP network layer entity, like Bandwidth Broker/Router, to verify bandwidth is reserved, thus ensuring QoS before acceptance of the call (e.g., ringing will not be provided unless resources have been reserved).

Common Network Functions

Common Network Functions used to provide services across the Connectivity Layer are the Service Broker (SB), the Network Routing Engine (NRE), the Call Admission Control (CAC), and the User Profile Engine (UPE). These functions can be used by more than one CCE.

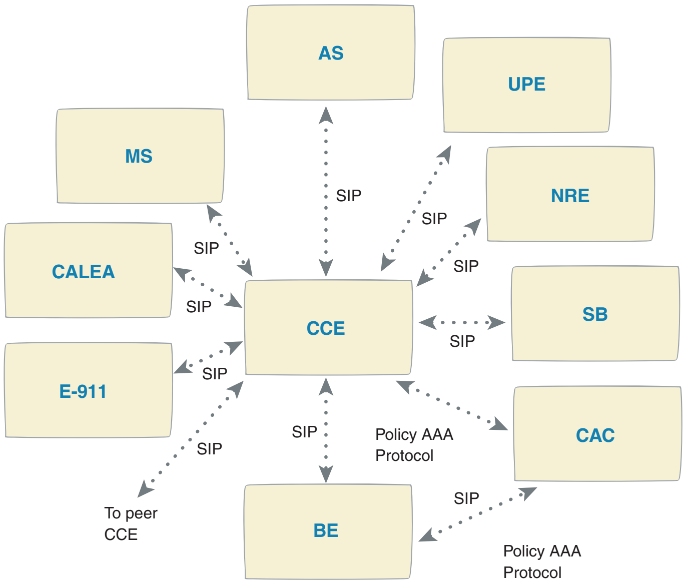

Figure 6: Architeture of Common Functional Entities in the Connectivity Layer

Service Broker

The Service Broker (SB) maintains subscribers' service information from a database of installed and activated services. This database can be shared with other functional entities, such as the NRE and the UPE. The SB acts as the SIP Redirect server and provides tables that define the services subscribed by each individual user. These services, along with the associated ASs, are usually provisioned in the SB using Operations Support provisioning capabilities. In the future, it is expected that any new services and associated ASs will be provisioned with the SB dynamically, using SIP.

The SB also performs service precedence and identifies to the CCE the primary applicable AS and a list of other services applicable to the call.

Network Routing Engine

The Network Routing Engine (NRE) provides the route information, upon finding the destination BE. This route information is required to set up the call leg between the source and destination BEs. SIP is used between the CCE and the NRE, with the NRE acting as the SIP Redirect server.

When a user registers or subscribes to a service, the database used by the NRE may be provisioned with the addresses of the end-points, residing in the access network, that are reachable through a particular BE. Also, the database is populated with a list of BEs through which a user can be reached.

The Location Server is part of the NRE and uses the same database. It collects the telephony route information (E.164) from the PSTN/TDM-IP gateway when users are reachable through those gateways in the PSTN/TDM network. The PSTN-IP gateway may be a part of the BE while the TDM-IP gateway may be a part of the CPE. The following protocols are used:

- TRIP-GW registration protocol is used by the gateways to register and transfer the route information from the gateway to the Location Server (LS)
- I-TRIP protocol is used between the NRE and the LS, and between the LSs within the AT&T network
- TRIP protocol is used between the LSs of the AT&T Network and AT&T partner or other service provider network

Call Admission Control

The Call Admission Control (CAC), engaged at the time of call setup, considers network-wide conditions and policies. The CAC manages capacity, controls congestion, observes firewall restrictions, and interprets SLA, QoS, Network Address Translation (NAT), and security policies.

The CAC is envisioned as a centralized function controlled by the CCE in collaboration with the BEs. While a BE implements local and customer-specific policy and provides authentication for entities that access the AT&T network, CAC functional entities help the CCE determine whether to admit or reject a call request. These functional entities include the Policy Server and the Authentication, Authorization and Accounting (AAA) server.

The Policy Server is used primarily for SLA and QoS policies. This server may also be used for security, accounting, billing, firewall, and NATs. The policy protocol used between the Policy Server and other functional entities (e.g., CCE, AAA Server, or BE) must be based on an open standard protocol that can interwork with SIP, once appropriate extensions are added to the protocol.

The Authentication, Authorization and Accounting (AAA) Server receives authentication from the BE and uploads security policy. The AAA protocol used between the AAA server and other functional entities (e.g., CCE or BE) must be an open standard protocol that can interwork with SIP once appropriate extensions are added to the protocol.

User Profile Engine

The User Profile Engine (UPE) is a functional entity that keeps both static and dynamic user profiles. The UPE consists of a SIP Registrar and a Presence Server (future implementation), which, along with the NRE, share the same logical database. Access to the UPE is controlled by the CCE for registration and presence information, using SIP signaling messages.

The Registration Server receives SIP registration requests under the control of the CCE, via BEs. The user location information obtained from the registration messages is updated into the user profile database. Users who subscribe to AT&T services will register with this server under the control of the CCE, via BEs.

The Presence Server (future implementation) stores and conveys the user's location and status (e.g., offline, busy, other). Users will register with the Presence Server (PS) under the control of the CCE, via BEs. The PS will subscribe to the users' presence information. When a status changes, (e.g., online, offline, busy) notifications will be sent by users' presence agents and immediately update the PS. User's availability should map to presence information so that the call can be expediently routed to the address where the user is currently available.

Media Servers

Media Servers (MS) typically operate with ASs to handle and terminate media streams, and to provide services such as announcements, bridges, transcoding, and Interactive Voice Response (IVR) messages. Using SIP to communicate, the AS sends an invite to the MS, via the CCE, setting up the call and specifying the script the MS executes or the function it performs. The MS returns the status and results to the AS via HTTP posts. With the exception of network-busy announcements provided by NG BEs, components of this architecture will not provide their own MS functions.

Special Network Servers

The following are special network servers used in the VoIP network.

CALEA Servers provide the ability to identify and collect content of voice telephone calls traversing the VoIP network, as mandated by the Communications Assistance for Law Enforcement Act (CALEA). When a CCE detects that a call needs to be monitored for legal reasons, it sets up the call via the CALEA server.

E-911 Servers route calls to the appropriate Public Safety Answering Points (PSAPs) based on the caller's location. When the CCE receives an emergency 911 call, it sends the call to the E-911 server and ensures that call waiting is disabled for the duration of the call.

Call Detail Recording

Call Detail Recording is required to provide data for billing and operations. The target is to have all BEs, CCEs, and ASs generate Call Detail Records (CDRs) for Operations Support, and to have the CCE generate an integrated CDR for billing purposes. The BEs and ASs should send the required information for billing to the CCE, via call signaling. There are several key requirements that will enable state-of-the-art network management and billing:

- Near-real time delivery to traffic monitoring, fraud detection, and billing systems
- Vendor-independent formats based on industry or standards-based formats
- Inclusion of quality of call and routing information
- Capturing PSTN interconnection and termination release codes
- Capturing a Unique Call ID from the originating BE
- Clearly-defined audit methods
- Record storage and redelivery in the case of lost connectivity

Applications Layer

Application Servers (ASs) implement features that leverage the VoIP Architecture, using a common interface to the VoIP Connectivity Layer. The "plug-in" paradigm facilitates fast and easy addition of new features, and SIP creates a unified interface to the Connectivity Layer.

"Plug-In" Features

The CCE, as directed by the SB, uses SIP to invoke an AS. The AS may also initiate communication sessions via the CCE (third party call control). While each AS is responsible for implementing its own service logic, all ASs implement a common interface to the CCE, by which the CCE may invoke an AS, and by which an AS may ask the CCE to set up, tear down, modify, or join call legs. This interface is implemented using SIP, thus providing common SIP signaling throughout the VoIP Architecture. ASs are free to implement only the subset of this interface that they require in order to perform their functions. While this interface may evolve over time, it will maintain backwards compatibility so existing ASs will continue to function, as the interface evolves.

While SIP provides the flexibility to allow ASs to set up call legs without leveraging the CCE, in the AT&T VoIP Architecture, ASs must use the CCE to set up, tear down, modify, and join call legs. Thus, from the AS's point of view, each AS has a CCE acting as its SIP Proxy for all inbound and outbound SIP messages.

While there is a well-defined SIP-based interface between the CCE and all ASs, each AS may invoke other Network Resources, using the appropriate protocol for that resource. Thus, an AS implements an abstract that enables new feature "plug-ins" without having to introduce new capabilities or protocols to the core VoIP Architecture.

For example, a voice mail application may leverage an IMAP message store and an LDAP directory, both of which are examples of Network Resources. Network Resources need not be operated or even hosted by the VoIP service provider, as in the case of a "Find Me" application leveraging a user's calendar that is available over the Internet as an XML Web Service.

As shown in the diagram, the AS may also invoke the MS to play prompts, collect user input, etc., by asking the CCE to create a call leg to the MS via SIP. The CCE will use SIP to create the leg to the MS. The MS will then request VoiceXML scripts and audio files, and post back user input via HTTP, as defined by the VoiceXML 2.0 standard. For efficient use of network resources, an AS may ask the CCE to instruct a BE to notify the AS when it detects mid-call DTMF.

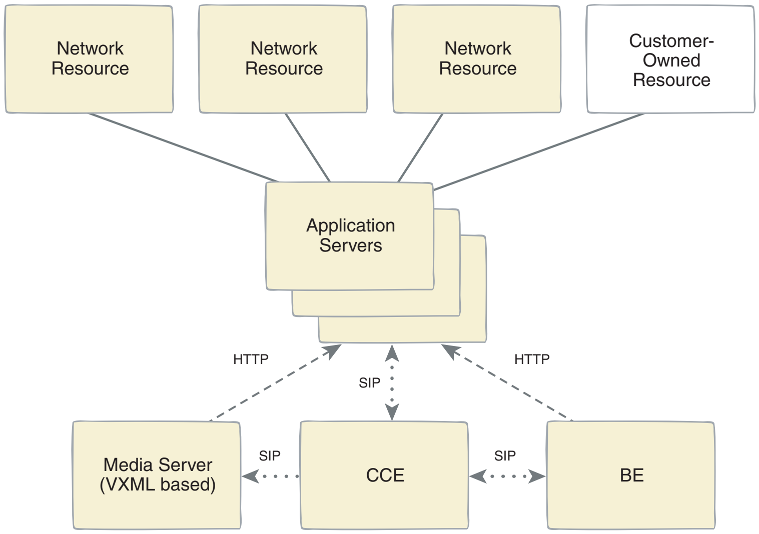

Figure 7: Application Server Interface

Service Creation

While many ASs implement specific service logic as defined by the vendor, certain ASs will instead serve as software platforms for implementing new services. These ASs have two elements: Service Creation Environment (SCE) and Service Logic Execution Environment (SLEE).

The Service Creation Environment (SCE) allows a development team to develop and test new service logic scripts. These scripts integrate various building blocks into a service, where the building blocks may be precompiled, implemented by other SIP-based ASs, or implemented as Network Resources.

The Service Logic Execution Environment (SLEE) provides the run-time environment for executing service logic scripts. While the SLEE appears to the rest of the VoIP Architecture to be just another AS, it plays an important role in the ability to deploy new services quickly, leveraging the capabilities of the SCE, as well as web-based management of subscribers and web-based service customization.

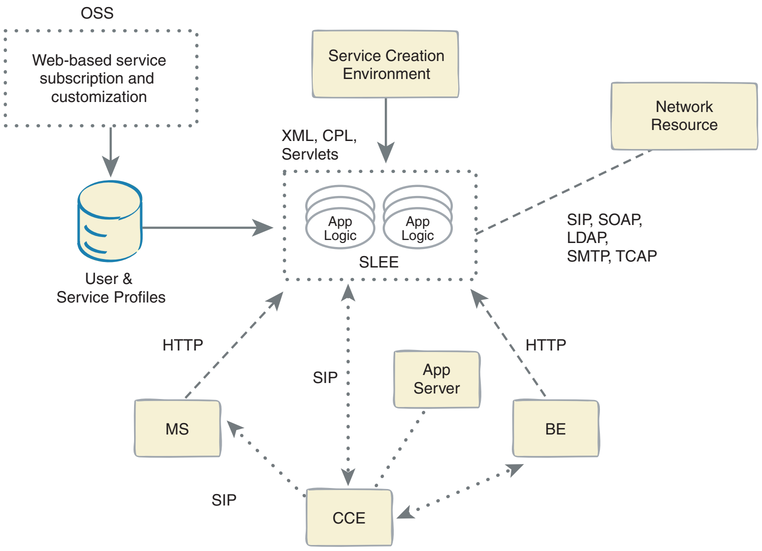

Figure 8: Service creation and Execution Environment

Resource Layer

The Resource Layer contains all systems that various ASs may use to provide their services. Most of these resources are owned and operated by AT&T, but a few may be owned and operated by AT&T customers. Protocols are not defined as a part of this architecture.

Operations Support Layer

The Operations Support Layer consists of multiple applications, databases, and a supporting data communications network used by internal and external personnel to manage the VoIP network and its elements. The target Operations architecture supporting the VoIP Network relies on the following design principles:

- **Concept of One** – Evolve to one Virtual Operations center from which all network elements can be monitored and managed. Accordingly, consolidations toward one organization, one network, one operations support architecture and one process result in reduced costs and increased efficiencies.
- **Zero Touch** – Minimize manual operator touch of the network elements for all network management functions. Requires an accurate network inventory to support process automation.
- **e-Enablement** – Provide access to network operations support systems via administrated permission to web interfaces, in order to facilitate the Concept of One and Zero-touch principles. The native Element Management System (EMS) provides access for the work centers into the VoIP infrastructure components via a single step cut-through.
- **Open Standards** – Allow multiple vendors to interoperate, and AT&T to adopt off-the-shelf solutions and processes to meet business objectives. In general, suppliers are expected to meet open interface standards with their VoIP network elements.

The following Operation Support systems and processes are described in this section: Fault Management, Configuration Management, Network and Service Provisioning, Capacity Management, and Performance and Incident Management.

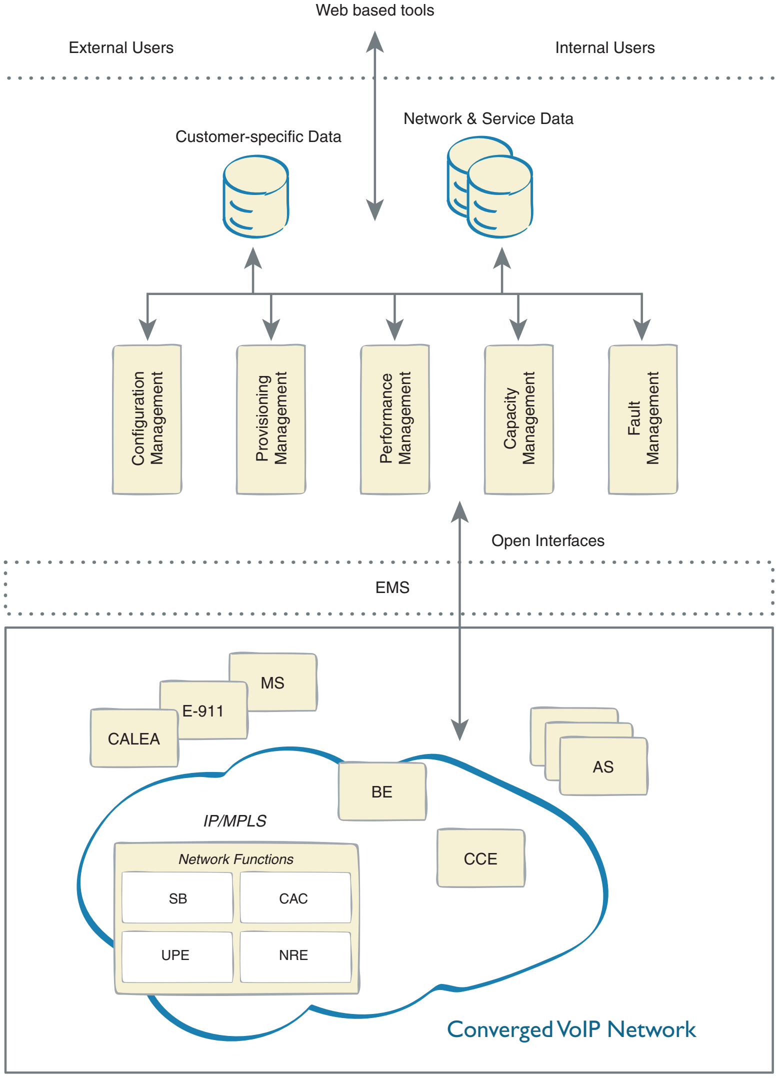

Figure 9: VoIP Operations Architecture

Fault Management

All VoIP network elements are supported through a single Fault Management process. This includes both end-customers and VoIP infrastructure components (BEs, Servers, Engines, Databases, CCEs, CACs, etc.).

Faults from VoIP infrastructure components are autonomously reported to the Fault Detection and Rules Management Systems through an Open Standard Interface (e.g., SNMP). Fault indications provided by the VoIP infrastructure components will support unique trap identifications and cause codes.

The VoIP infrastructure components support end-to-end testing across the converged networks from a common test platform.

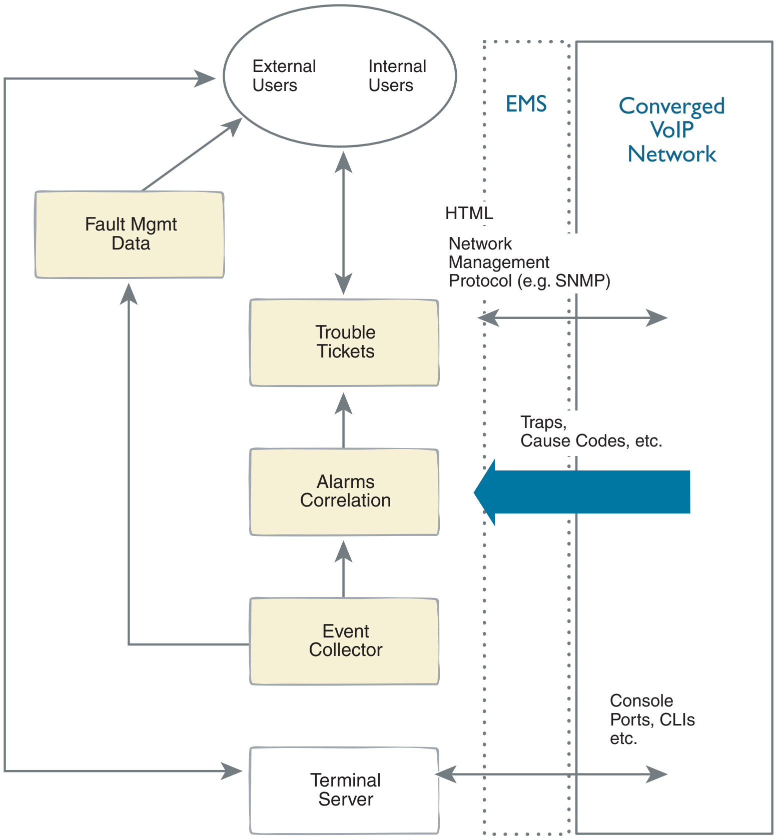

Figure 10: Fault Management Architecture

Configuration Management

Configuration Management supports the initial loading of element parameters, changes in element configuration, and retention of data for recovery for all network elements (BE, Server, Engine, Database, CCE, CAC) that are physically connected to the VoIP Converged Network.

Configuration Management refers to the core operations process of loading the following types of data:

- Data elements needed to uniquely identify the component and to establish connectivity with other IP network elements and operations systems
- Data elements that are static and populated at the time of installation
- Ongoing loading of data triggered from facility, trunk (group), link, or routing orders

Configuration Management must support the following:

- Retention of golden copy configurations and a configuration validation process (where configuration data is stored in the network element)
- An interface with the Network Inventory Database to maintain growth changes in the overall VoIP Network topology, in concordance with the Capacity Management process.

- The test & turn-up process at equipment deployment
- Showing the configuration and provisioning architecture for the VoIP systems. Appropriate open interfaces must be supported.

Network Provisioning

Network Provisioning is the process that assigns previously installed and configured VoIP network element capacity (e.g., NG BE) to support a level of service. Network and Service Provisioning are interdependent, as Service Provisioning provides the element's services.

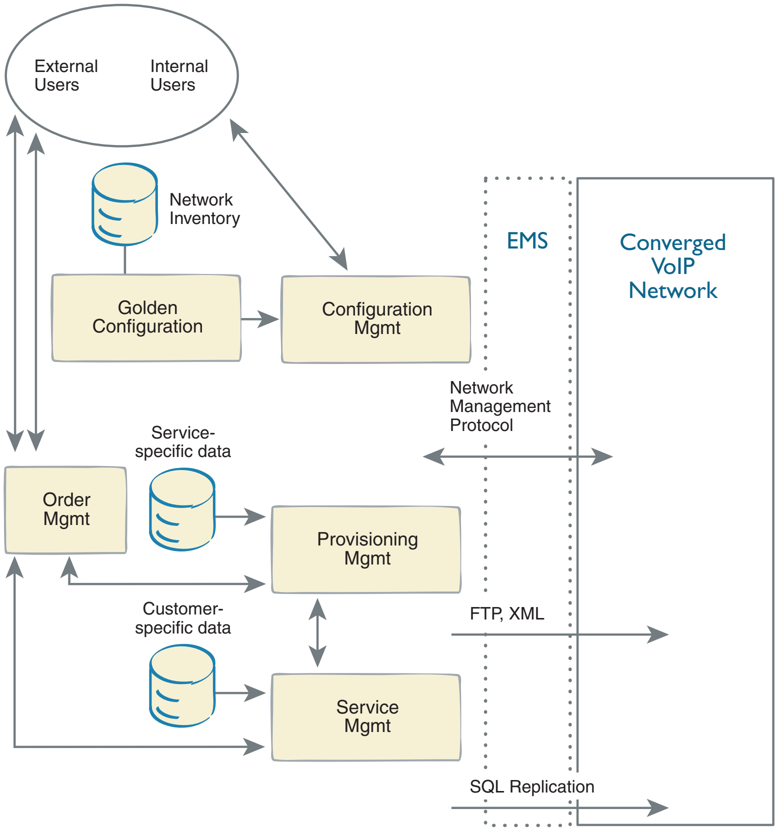

Figure 11: Configuration and Provisioning Architecture

Network Provisioning of the VoIP infrastructure components requires standards-based machine-to-machine interfaces to both a transaction initiator, (e.g., legacy OSS) and the actual infrastructure element being provisioned. Infrastructure elements must be able to support provisioning requests from multiple clients for changes to the state of a customer (e.g. in-service, out-of-service). In general, VoIP infrastructure components should support direct or indirect (e.g., via EMS) access for automated provisioning via SNMP, FTP and/or database replication tools as necessary. If an EMS is used, it must be able to acknowledge receipt of transactions and report successful completion of transactions, or reason for failure, to upstream provisioning systems.

Service Provisioning

VoIP Service Provisioning requires a flow-through mechanism to provision customer-specific orders within the Converged VoIP network, as well as orders to provision facility, trunking, and routing associated with access to the VoIP converged Network.

In adherence to Concept of Zero, the VoIP infrastructure elements will support flow-through of customer service requests by allowing the following:

- Real-time system interfaces to SB, allowing update and query of the services or applications (8YY, SDN, etc.) that a customer has purchased

- Real-time system interface to service ASs, allowing update capabilities and query of the features for a given application
- Real-time system interface to the VoIP Service Provisioning Server for network provisioning
- Real-time update of the UPE
- Real-time system interface to billing systems

Capacity Management

Capacity Management is the primary tool used to maintain adequate in-service inventories of all VoIP Network Elements, (BEs, Servers, Engines, Databases, CCEs, CACs, etc.) in order to meet customer demands of the network. AT&T uses extensive performance measurement data on the network elements, including average and peak buffer utilization, as well as packet loss. Using these inputs, along with service forecasts, the system will be used to forecast the network's physical growth needs.

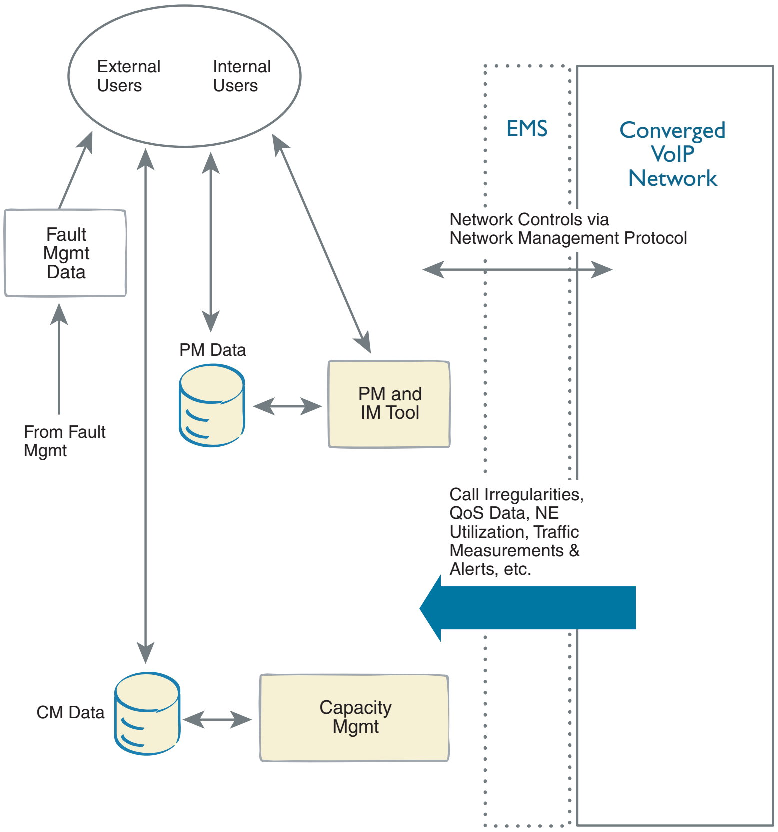
External Users
Internal Users
EMS
Converged VoIP Network
Fault Mgmt Data
Network Controls via Network Management Protocol
PM Data
PM and IM Tool
From Fault Mgmt
Call Irregularities, QoS Data, NE Utilization, Traffic Measurements & Alerts, etc.
CM Data
Capacity Mgmt

Figure 12: Performance and Capacity Management Architecture

Performance Management and Incident Management

Operational support for Incident and Performance management is crucial to ensuring AT&T's VoIP network maintains the highest performance levels. Direct and indirect control of people, process and network is coordinated in these functions, with the goal of increasing service quality. The functions of Performance and Incident Management can be divided into four functions: Performance Surveillance, Incident Management, Traffic Management and Performance Reporting.

Performance Surveillance functions require near real-time availability of network behavior measurements (e.g., faults, call processing irregularities, traffic measurements, equipment alarms, quality of service information (delay, loss, jitter) and configuration changes. Measurements are analyzed against configurable and perhaps dynamic thresholds. Resultant performance surveillance information is used to detect and locate problems that degrade service. Measurements also provide feedback on the effectiveness of restoration and traffic management activities.

Incident Management requires information about the customer impact resulting from abnormal network behavior during and following service-affecting events. This requires the ability to relate service degradation with customer impact in terms of service, to identify which customers were impacted, and to determine the length of the event.

Traffic Management is the practice of real-time traffic control to protect network resources from overload, or to redirect traffic to make better use of under-utilized network resources. Either goal serves to maximize the effectiveness of the network. Traffic management is most often associated with mass calling events and national crises. This discipline allows for prioritization of restorative activities to reflect business imperatives.

Performance Reporting functions are supported by data extracted from stored aggregate performance measurements. This data needs to be reported by time series (trending), by event, by customer, by service, by network device or by any ad-hoc extract, for the purposes of quality assurance, process improvement, equipment maintenance, network configuration, network planning, and business evaluation or customer inquiry.

Performance and Incident management are interdisciplinary, multi-service and centralized functions. OS solutions from multiple vendors for multiple services need to interoperate. OS support requires centralization and high levels of integration. Open APIs and standardized system interfaces are needed for network elements, EMSs and Network Management Systems (NMSs).

Performance and Reliability

The network platforms that support the VoIP Operations Support Layer, which include Operation Support Systems and their Data Communication Networks, must deliver appropriate contributions to performance and reliability to ensure that the objectives of all services and operations functions can be met.

AT&T has employed various layers of protection to support survivability by preventing and, when required, restoring the network/service with minimal customer impact. These layers of protection include outage prevention, equipment protection switching, facility restoration, routing resiliency, network management and disaster recovery.

Conclusion

This document has illustrated the high-level concepts of AT&T's Common VoIP Architecture. The architectural platform will enable AT&T to develop and deploy existing and new real-time communication services that use IP as their transport mechanism. By sharing a single, common VoIP infrastructure with interoperable network elements that support open standard interfaces and protocols, AT&T will be able to develop new services and applications, and interconnect with various access networks.

AT&T is committed to creating a high-capacity VoIP network that ensures reliability, availability and performance. In adherence to the Concept of One, a single infrastructure will be developed to support existing and new services. And, in adherence to the Concept of Zero, intelligence will be built into AT&T's core network, providing little or no need for human intervention. The AT&T Common VoIP Architecture creates a foundation for the complex and intricate process of migrating voice and other real-time communications services from a circuit-switched to an IP packet-switched network.

Acronyms

AAA – Authentication, Authorization and Accounting

API – Application Program Interface

AR – Access Router

AS – Application Server

ASN – AT&T Switch Network

ATM – Asynchronous Transfer Mode

B2BUA – Back-to-Back User Agent

BE – Border Element

BR – Backbone Router

CAC – Call Admission Control

CCE – Call Control Element

CDR – Call Detail Record

CLI – Command Line Interface

CMS – Call Management Server

CMSS – CMS Signaling

CPE – Customer Premises Equipment

DDOS – Distributed Denial of Service

DSL – Digital Subscriber Line

DTMF – Dual Tone Multi Frequency

EMS – Element Management System

ES – Edge Switch

EVPN – Enhanced Virtual Private Network

FR – Frame Relay

FTP – File Transfer Protocol

GW – Gateway

HFC – Hybrid Fiber Coaxial

IETF – Internet Engineering Task Force

ILEC – Incumbent Local Exchange Carrier

IMAP – Internet Message Access Protocol

IP – Internet Protocol

OSPF – Open Short Path First

ISP – Internet Service Provider

I-TRIP – Intra-domain TRIP

IVR – Interactive Voice Response Message

LDAP – Lightweight Directory Access Protocol

LS – Location Server

MDNS – Managed Data Network Service

MEGACO – Media Gateway Control

MGCP – Media Gateway Control Protocol

MIS – Managed Internet Services

MPLS – Multiprotocol Label Switching

MRS – Managed Router Service

MS – Media Server

MTA – Media Terminal Adapter

NAT – Network Address Translation

NG – Network Gateway

NMS – Network Management System

NRE – Network Routing Engine

OS – Operating System

OSS – Operational Support System

PBX – Private Branch Exchange

PS – Presence Server

PSTN – Public Switched Telephone Network

QoS – Quality of Service

SB – Service Broker

SCE – Service Creation Environment

SIP – Session Initiation Protocol

SLA – Service Level Agreement

SLEE – Service Logic Execution Environment

SNMP – Simple Network Management Protocol

TA – Terminal Adapter

TDM – Time Division Multiplexing

ToS – Type of Service

TRIP – Telephony Routing over IP

UA – User Agent

UPE – User Profile Engine

VoIP – Voice over Internet Protocol

VPN – Virtual Private Network.

POINT OF VIEW

DISCLAIMER:

This document presents information about the current plans and considerations for AT&T's VoIP Infrastructure. This information is subject to change without notice to you. Neither the furnishing of this document to you nor any information contained in this document should be interpreted by you as a legal representation, agreement, or commitment by AT&T concerning (1) any information or subject contained in or referenced by this document, or (2) the furnishing of any products or services by AT&T to you, or (3) the purchase of any products or services by AT&T from you, or (4) any other topic or subject. AT&T may own intellectual property which relates to the information contained in this document. Notwithstanding anything in this document to the contrary, no rights or licenses in or to this intellectual property are granted, either expressly or impliedly, by this document. Rights to AT&T intellectual property may be obtained only by express written agreement with AT&T, signed by AT&T's Chief Technology Officer (CTO) or the CTO's authorized designate.

The information contained in this document represents current considerations of the AT&T Common VoIP Infrastructure as of the date of publication. The contents of this document should not be interpreted to be a commitment or legal representation on the part of AT&T to you regarding AT&T's VoIP Infrastructure, and are subject to change without notice. AT&T MAKES NO WARRANTIES, WHETHER EXPRESS OR IMPLIED, IN THIS DOCUMENT. BY PUBLISHING THIS PAPER AT&T IS NOT GRANTING TO YOU, EITHER EXPRESSLY OR IMPLIEDLY, ANY RIGHTS UNDER ANY OF ITS INTELLECTUAL PROPERTY.

For more information, contact your AT&T Representative, or visit www.att.com/business.

AT&T