

CloudWare Deployment on Data Networks – VoIP Best Practice

This White Paper sets out to consolidate ‘Best Practice’ information drawn from various sources and established VoIP network engineering principles.

Engineers deploying CloudWare software (or ANY VoIP Softswitch product) will need to emphasise the ‘real time’ nature of VoIP when encountering customer resistance to implementation of ‘VoIP Best Practice’ - especially when it is compared to other applications running on their network. While a 1 second delay in delivery of an email or in the transfer of a file from one PC to another is tolerated, it is disruptive and unacceptable in a carrier/business grade phone call. This means that by definition critical network elements must be optimised for efficient delivery of VoIP/SIP data packets in a converged solution.

Key Pre-Deployment Questions:

- Has a VoIP specific network audit been carried out?
- Have the audit findings been implemented?

It’s an obvious statement but following general VoIP best practices will also produce a successful project which, with a satisfied customer will become a good reference site.

The principles of general layout and design models for integrated VoIP and data networks provide a strong base for VoIP deployments so as a start point, always separate and handle voice traffic through standard, vendor-neutral protocols and design practices. A dedicated VLAN, setting QoS and enough bandwidth to take account of the VoIP traffic allowing approximately 90Kbps each way per call (plus a 50% margin for any additional proprietary signalling for SoftPhones) will each help build a successful, converged VoIP solution atop an existing network.

VoIP Models

Regardless of a VoIP network’s specific design or the particular system being deployed, follow the standard guidelines appropriate for the overall VoIP network model. This document examines two widespread VoIP network models – a centralised design and a distributed design. Centralised models provide call setup and teardown through a VoIP core involving one or more Servers. A distributed model performs call setup and teardown at multiple locations. The locations are normally arranged geographically with links for inter-site connectivity. The centralised model has the advantage of hardware and PSTN consolidation. The distributed model is best used with multiple sites of equivalent size where most traffic remains local.

Redundancy is a critical success factor for a centralised model and is usually accomplished with two or more CloudWare Servers that can each handle the total load. These servers should be dual-homed with two NICs connected to separate switches (*NOTE: The two NICs should be in hardware redundant mode presenting to the Operating System as one NIC/controller or where two individual NICs are fitted ONLY ONE NIC in each server should be enabled). Also, to **avoid duplex/half-duplex conflicts, NICs must be ‘locked down’ to full duplex with no rate limiting and avoid server bottlenecks by removing any bandwidth throttling on the switch’s server port.

Backup power should be provided via a UPS with enough available load to outlast multiple brownouts as well as multi-hour blackouts. The entire phone system depends on these servers retaining connectivity.

Key Points in configuring a resilient VoIP network and its associated hardware

- All switches and routers ‘SIP/VoIP compliant’ and QoS capable
- One VLAN dedicated to VoIP with all endpoint users, the CloudWare Server and its NIC port as members
- *One NIC active only per CloudWare Server connected to the VLAN and the SIP trunk via a common switch – duplicate in redundant server, **no rate limiting
- No bandwidth throttling on the CloudWare Server’s switch port
- No other software application running on the VoIP server
- Use IP addresses only in endpoint device configurations rather than DNS server addresses, hostnames, and domain names
- Use long DHCP IP address lease times e.g. 8 days or more
- Use bandwidth techniques such as ‘Weighted Round Robin’ (WRR) or ‘Modified Deficit Round Robin’ (MDRR) **only** when the highest priority for VoIP RTP & control packets can be ensured

VoIP call processing requires time-critical access to Operating System functions, memory allocation and to the NIC therefore the servers should have no other software applications installed other than CloudWare Server.

Note that if redundant links are in place between the server and switches, and spanning tree protocol (STP) is used to mitigate data loops, STP may introduce packet delays which will adversely affect voice quality. However data loops must also be prevented as this can result in 'packet storms', and in particular SIP Real Time Control Protocol (RTCP) storms.

While a centralised design will need to be highly resilient, it usually involves less hardware overall. PSTN/SIP trunk connectivity can be consolidated at one location or in a dedicated CloudWare Transit Server. Organizations typically use a centralised model when they have a large central site, such as a Carrier Platform or corporate headquarters, and several remote locations. WAN links connect sites into the main location and provide administration of voice communications. Call setup and teardown will be handled over the WAN links, but local site calling will remain local. The centralised model's disadvantage is the requirement to maintain connectivity to the central site from each location. WAN link redundancy is typically expensive, but the cost can be mitigated through local failover options. CloudWare Server offers full redundancy and failover capability.

In a distributed model, multiple large sites usually require localised CloudWare Servers. These sites will pass inter-site voice calls through VPNs that link the sites and sometimes local trunks. The benefit here is a decreased need for redundant connectivity. The distributed model costs may increase however, due to the overall hardware requirements and design complexity.

Network considerations

Design consideration will greatly influence the overall success when integrating VoIP with an existing network. Network equipment will need to support virtual LANs (VLANs), quality of service (QoS), and compression. Use VLANs to separate voice and data traffic but have them co-exist on the same medium. VLANs offer a simple way to provide a distinct level of service as well as allowing the individual traffic types to be monitored.

QoS is a method of distinguishing IP packets so they can be treated by distinct policies. The VoIP implementation must prioritise traffic to prevent existing data traffic from undermining voice communication integrity. This process requires matching specific fields within either the Layer 2 or Layer 3 headers inside the data units. Layer 2 QoS, for example 802.1p, allows for multiple levels of prioritization by tagging frames with certain fields. At the switch port level, QoS can then be provided to traffic entering the port. Therefore, as data and voice traffic simultaneously enter the device, voice traffic can be sent first and the data traffic queued for delivery. CloudWare Server adopts the Differentiated Services (DiffServ) model: Precedence '5', LowDelay & High Throughput. (Hex B8). Other values for the outgoing RTP TOSByte can be implemented by adding a new setting in the CloudWare 'server.ini' file under section [Default] 'TOSValue=nnn' where 'nnn' is the decimal equivalent of the hexadecimal value required. Proprietary (e.g. non-SIP) VoIP control and voice packets (such as for CloudWare SoftPhones) are normally handled via direct IP address to IP address port mapping.

On Layer 3 links, such as WAN links between sites, QoS can also be used to match on fields within the IP header. Again, this provides queues for data traffic for delivery while sending voice traffic with priority. As these QoS processes are fairly detailed, consider how much bandwidth will need to be allotted to individual traffic types. For example, 75 percent of a WAN circuit could be set to carry voice traffic, leaving 5 percent for call setup and teardown activity, and reserving 20 percent exclusively for data traffic. It may also be useful to identify highly-critical data traffic that can be placed somewhere below voice, but above the general data queue. When dealing with limited bandwidth, allow ample time to create an effective QoS policy. QoS policy will ultimately determine inter-site voice quality as otherwise the network sends traffic on a first-come, first-served basis.

To avoid packet loss that will disrupt voice quality and control signalling, only use bandwidth techniques [such as 'Weighted Round Robin' (WRR) or 'Modified Deficit Round Robin' (MDRR)] where delay-sensitive VoIP traffic (RTP and control packets) can be prioritised at the highest level.

After QoS, the codec is the next consideration. Compression will create smaller packets by shrinking duplicated information however this is at the expense of overall voice quality. The specific VoIP codec CloudWare Server uses to ensure carrier/business grade voice quality is G.711 – this also directly impacts bandwidth. G.711 requires approximately 90Kbps

VoIP Deployment Notes

- Centralised VoIP networks help consolidate hardware and PSTN connectivity and require a highly resilient central site
- Distributed VoIP networks require more hardware than centralised networks and don't require as much redundant connectivity as centralised networks
- QoS and VLANs enable separate voice and data traffic treating each with distinct transmission policies
- The VoIP network should include redundant hardware, redundant links, and adequate power backups
- Use standard network monitoring techniques and VoIP-specific tools to continuously monitor the network's health and identify problems

each way for voice communication. The overall goal is to transmit the maximum number of voice calls while maintaining acceptable call quality.

Network resilience

End-users expect reliable telephone service, and a successful VoIP solution requires a resilient and redundant network. Where possible, the network should contain redundant hardware, redundant links, and adequate power backups. To meet the needs of a VoIP network, core network and VoIP hardware should run off UPS backups. If the VoIP solution uses power over Ethernet (PoE), the local network switches will also need power backups. These switches transmit power to the individual telephones, which will need power and connectivity during an outage.

A strong network design will translate to a good VoIP design. Medium and large VoIP deployments require a close examination of the redundancy at the network's core. Redundancy measures may include dual Layer 3 switches that share default gateway duties through a standards-based or proprietary protocol. The network may also have redundant links from the access layer into the distribution and from the distribution into the core. Not all networks are built to such standards, but an audit should be carried out by a specialist VoIP network engineer to identify and mitigate any potential points of failure.

While DHCP and DNS standards apply in most data-only networks, for VoIP there is a high element of risk that communications will be disrupted by IP address changes and delayed name resolution. For example Cisco recommends against the use of DNS server addresses, hostnames, and domain names in VoIP deployments, opting for static IP addresses for both the server and all end points. As for DHCP Avaya recommends lease times of 6 weeks or more, while Cisco recommends 8 days or more to mitigate the risk of disruption to calls in progress, polling and registration of VoIP end points.

Maintaining a healthy VoIP network

Once installed it is imperative to regularly monitor and test the VoIP network to ensure sustained integrity. Although a properly designed voice network will initially respond well, quality can quickly degrade if problems are not detected and resolved. Regularly monitor bandwidth usage and each hardware device's CPU and memory use. Use VoIP-specific management tools to monitor the voice network and catch problems basic IT management techniques will miss (e.g. ClearSight Analyzer, WildPackets OmniPeek).

An effective patch management process is also critical for VoIP network health. Consistently update VoIP hardware and software as with any other system. CloudWare regularly releases code revisions and updates for CloudWare Server and SoftPhone software that should be installed as soon as possible to maintain application integrity. Similar policies should apply to all network elements such as SIP phones, switches and routers.

The bottom line

The overall design and implementation process will greatly affect a VoIP installation and make the CloudWare or any other VoIP Softswitch deployment a success. Networking design functions such as VLANs and QoS are definite requirements and can easily make or break VoIP call quality. Make the network strong enough to handle both day-to-day problems and occasional catastrophes. Finally, monitor and maintain the VoIP network to increase its reliability and effectiveness. When voice is sent over the data network, quickly detecting and fixing problems is critical for success. Once deployed, VoIP traffic will likely become the most important information a network transmits.

This document draws from various public sources and from consolidated research by others of 'VoIP Best Practice' network engineering principles. Many site-specific environmental factors may influence the success or otherwise of the practical implementation or application of any of these principles in the field. As a result CloudWare Pty Ltd accepts no liability whatsoever for consequential damages arising from any such application or implementation including but not limited to loss of anticipated profits or other economic loss.