GVP HSG Pages

Genesys Voice Platform 9.0.x

12/29/2021

Table of Contents

Sizing Genesys Voice Platform 9.0

9.x This version of the Sizing Guide applies to Genesys Voice Platform that is part of 9.0, starting with version 8.5. For version 8.1 of Genesys Voice Platform, see the Genesys Voice Platform home page.

How you can use this document

Read this chapter of the Genesys Hardware Sizing Guide to:

- Learn the recommended hardware and software needed to support Genesys Voice Platform (GVP) 8.5.
- Learn terminology, and about capacity, performance criteria, machine setup, application profile makeup, and typical call volumes.
- Review extensive performance data from testing numerous hardware and software configurations, to determine what will work best for your Genesys installation.

Intended Audience

Engineering, Sales and Marketing, Product and Program Management, Quality Assurance, Technical Publications, Production, Genesys Partners, and Genesys customers.

Content

Hardware and Software Tested	Traffic and Capacity Testing ...extensive test data	Performance Planning and Scalability ...graphs and interpretive tools
• Hardware and Operating Systems Tested ...minimum and optimal configurations • Hardware and Bandwidth Usage ...disk space and bandwidth specs	• Overview, VXML and CCXML Application Profiles • Traffic and Capacity Terms • VoIP Capacity Test Summary Tables • Component Capacity Test Case Tables • GIR-GVP Port Capacity Tests	• Performance and Scalability Comparisons • Application Test Cases • Component Test Cases • Single Server Test Cases • Multiple MCP Instances and Virtual Machines Test Cases • Call Setup Latency Test Cases

Hardware and Software Tested	Traffic and Capacity Testing ...extensive test data	Performance Planning and Scalability ...graphs and interpretive tools
		• Cachable VoiceXML Content Test Cases

Hardware and Operating Systems Tested

[+] HOW YOU CAN USE THIS TOPIC

Read our recommendations of the appropriate Hardware, Operating System, and Reporting Server for your GVP installation.

Hardware Tested

These configurations were tested on Windows and Linux.

Hardware	Tested
CPU	Dual Quad Core Xeon X5355 2.66 GHz (benchmark) or higher (For optimal performance, Genesys recommends Xeon with Core 2 technology.) Note: Other CPUs are also used for testing: • Dual Quad Core Xeon E5620 2.40GHz 16GB RAM • Dual Hex Core Xeon X5675 3.06GHz 32GB RAM • Dual Hex Core Xeon E5-2683 v4 2.10GHz 198GB RAM • Dual DoDeca Core Xeon E5-2695 v2 2.40GHz 128GB RAM • Single Hex Core Xeon X5675 3.06GHz 12GB RAM • Single Hex Core Xeon X5670 2.93GHz 12GB RAM • Single Dual Core Xeon X5160 3.0GHz 8GB RAM
Memory	4GB RAM minimum, 8GB recommended
Network	1 Gigabit or 100 Megabit Ethernet
Storage	RAID 1 HDD with at least 40GB with 15 K RPM

GVP-GIR Port Capacity Test Profiles 2014

Operating Systems Tested

- Microsoft Windows 2008, Enterprise Edition, SP2, x64
- Microsoft Windows 2008 R2, Enterprise Edition, SP1, x64
- Microsoft Windows 2012, Enterprise Edition, SP1, x64
- Microsoft Windows 2016, x64

- Red Hat Enterprise Linux 6.0, Update 4, x64
- Red Hat Enterprise Linux 7.0, x64

Important

GVP does not support the 32-bit OS versions any more:

- Microsoft Windows 2008, Enterprise Edition, SP2, x86
- Red Hat Enterprise Linux 5.0, Update 8, x64
- Red Hat Enterprise Linux 5.0, Update 8, x86

Operating Environments Supported

Reporting Server Recommendations

If you intend to deploy Reporting Server in partitioning mode to optimize a high-performance environment, you must ensure you are using a supported operating system. Before you deploy the Reporting Server, consider the information below.

Reporting Server Modes

All versions of Microsoft SQL Server are supported on Windows only. Reporting Server can be installed on Linux, however, the database must be installed off-board on a separate Windows host.

Reporting Server mode	Supported operating systems
partitioning mode	Oracle 10g or 11g Enterprise Edition Microsoft SQL Server 2008 Enterprise Edition
standard mode	Oracle 10g or 11g Standard Edition

Reporting Server mode	Supported operating systems
(no partitioning)	Oracle 10g or 11g Enterprise Edition Microsoft SQL Server 2005 or 2008 Standard Edition Microsoft SQL Server 2005 or 2008 Enterprise Edition

Reporting Server Performance in Different Modes

The Reporting Server will perform at optimal levels when in partitioning mode. In standard mode, Reporting Server 8.1.3 performance will be below optimal and comparable to Reporting Server 8.1.1 performance, which does not support partitioning. However, Reporting Server 8.1.2 and later releases in partitioning mode are much improved over standard mode Reporting Server 8.1.1 in terms of performance.

top | toc

GVP Hardware and Operating System Profiles for Genesys Interaction Recording (GIR)

- Hardware Profiles Used in These Tests
- Virtual Machine Profiles Used in These Tests

Hardware Profiles

Hardware Profile 1	Specifications & Recommendations	Comment
CPU	Single Hex Core Intel Xeon X5670@ 2.93GHz	
Memory	8 GB or more	4 GB is minimum and 8 GB is recommended
Network	GigaBit Ethernet	100MBit supported
Storage	15k rpm SAS HDD disk storage with at least 72 GB. RAID 0.	15k rpm recommended for maximum performance
OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

Hardware Profile 2	Specification & Recommendation	Comment
CPU	Single Hex Core Intel Xeon X5675@ 3.06GHz	
Memory	16 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP recording cache location. 15k rpm SAS HDD disk storage with at least 136 GB used for all other operations. RAID 0	SSD and 15k rpm SAS HDD are recommended for maximum performance
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 3	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	16 GB or more	8 GB is minimum and

Hardware Profile 3	Specification & Recommendation	Comment
		recommended
Network	GigaBit Ethernet	100MBit supported
Storage	15k rpm SAS HDD disk storage with at least 72 GB. RAID 0	15k rpm SAS HDD is recommended for maximum performance
OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

Hardware Profile 4	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	32 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP recording cache location. 15k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	SSD and 15k rpm SAS HDD are recommended for maximum performance
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 5	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	32 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	Multiple 15k rpm SAS HDDs disk storage with at least 360 GB used for all other operations. RAID 0.	Split VMs into multiple 15k rpm SAS HDDs.
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 6	Specification & Recommendation	Comment
CPU	Single Eight Core Xeon E5-2640 2.00 GHz	
Memory	64 GB or more	8 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP logs and recording cache location. 15k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	SSD and 15k rpm SAS HDD are recommended for maximum performance.
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 7	Specification & Recommendation	Comment
CPU	Dual 16 core Xeon E5-2683 v4 @ 2.10GHz	
Memory	32 GB or more	8 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	10k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	SSD and 15k rpm SAS HDD are recommended for maximum performance.
OS	VM vSphere or ESXi 6.x Windows Server 2016/RHEL 7 as Guest OS	VM vSphere 6.x as Host OS Windows Server 2016/RHEL 7 as Guest OS

Virtual Machine (VM) Profiles

VM Profile 1	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 2	1x X5675@3.06GHz, 16 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 2	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 4	2x X5675@3.06GHz, 32 GB RAM
CPU	4 x vCPU	
Memory	8 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location.At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance.
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 3	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 4	2x X5675@3.06GHz, 32 GB RAM
CPU	3 x vCPU	

VM Profile 3	Specifications & Recommendations	Comment
Memory	6 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance.
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 4	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 4	2x X5675@3.06GHz , 32 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance.
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 5	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 5	2x X5675@3.06GHz, 32 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	At least 36 GB 15k rpm SAS HDD disk storage.	
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 6	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 7	2x Intel® Xenon® CPU E5-2683 v4@2.10GHz
CPU	2 x vCPU	
Memory	4 GB RAM	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	At least 36 GB 10k rpm SAS HDD disk storage.	
Guest OS	Microsoft Windows Server 2016 or Red Hat Enterprise Linux 7.0	

Hardware and Bandwidth Usage

This section contains hardware / disk space usage and bandwidth estimates for the Reporting Server, and bandwidth usage estimates for the Media and Call Control Platforms.

- Reporting Server Hardware Usage
- Bandwidth Usage for MCP, CCP, RS

Reporting Server Hardware Usage

Factors affecting disk space requirements for Reporting Server:

- Retention period
- Call rate
- Number of IVR Profiles, Tenants, and DNs

Reporting Server Disk Space Estimates

This table provides information necessary to estimate the disk space required for Reporting Server data types. For more information about data retention and data types, see "Data Retention Policy Wizard" in "Chapter 6: Provisioning IVR Profiles" of the GVP 8.5 User's Guide.

Table: Reporting Server Disk Space Estimates

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Resource Manager				
CDR	Very High	600	Calls per day	`retention.cdr`
Calculation: `600 * number of calls per day * retention.cdr`				
Operational Reporting (5 minutes)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.5min`
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of`				

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
`RMs) * 2 * 1440 * retention.operations.5min`				
Operational Reporting (30 minutes)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.30min`
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * 48 * retention.operations.30min`				
Resource Manager				
Operational Reporting (hourly)	Medium	300	Number of: • DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.hourly`
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * 24 * retention.operations.hourly`				
Operational Reporting (daily)	Medium	300	• DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.daily`
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.daily`				
Operational Reporting (weekly)	Medium	300	• DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.weekly`

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.weekly/7`				
Operational Reporting (monthly)	Medium	300	• DNs • IVR Profiles • Tenants • RM, CTIC, PSTNC	`retention.operations.monthly`
Calculation: `300 * (number of DNs + number of IVR Profiles + number of tenants + number of CTIC, PSTNC +1) * (number of RMs) * 2 * retention.operations.monthly/30`				
Media Control Platform				
CDR	Very High	600	Calls per day	`retention.cdr`
Calculation: 600 * calls per day * retention.cdr				
Operational Reporting (5 minutes)	Medium	300	• IVR Profiles • MCPs	`retention.operations.5min`
Calculation: `300 * (number of IVR Profiles + 1) * (number of MCPs) * 1440 * retention.operations.5min + 100 * (number of MCPs) * 1440 * retention.operations.5min` **Note:** The first product is for the arrivals that are stored per IVR Profile for each MCP. The second product is for the peaks that are stored for each MCP.				
Operational Reporting (30 minutes)	Medium	300	• IVR Profiles • MCPs	`retention.operations.30min`
Calculation: `300 * (number of IVR Profiles + 1) * (number of MCPs) * 48 * retention.operations.30min + 300 * (number of MCPs) * 48 * retention.operations.30min`				
Media Control Platform				
Operational Reporting (hourly)	Medium	300	• IVR Profiles • MCPs	`retention.operations.hourly`
Calculation:				

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
`300 * (number of IVR Profiles + 1) * (number of MCPs) * 24 * retention.operations.hourly + 300 * (number of MCPs) * 24 * retention.operations.hourly`				
Operational Reporting (daily)	Medium	300	• IVR Profiles • MCPs	`retention.operations.daily`
Calculation: `300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.daily + 300 * (number of MCPs) * retention.operations.daily`				
Operational Reporting (weekly)	Medium	300	• IVR Profiles • MCPs	`retention.operations.weekly`
Calculation: `300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.weekly/7 + 300 * (number of MCPs) * retention.operations.weekly/7`				
Operational Reporting (monthly)	Medium	300	• IVR Profiles • MCPs	`retention.operations.monthly`
Calculation: `300 * (number of IVR Profiles + 1) * (number of MCPs) * retention.operations.monthly/30 + 300 * (number of MCPs) * retention.operations.monthly/30`				
Events	Very High	500	• events per call • calls per day	`retention.events`
Calculation: `500 * number of events per call * number of calls per day * retention.events`				
VAR CDR	Very High	200 per VAR CDR 150 per VAR custom variable	• calls per day • custom variables per call	`retention.cdr`
Calculation: `(200 +150 * number of custom variables per call) * number of calls per day * retention.cdr`				
Media Control Platform				

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
VAR Summary (5 minutes)	Medium	300	• IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	`retention.var.5min`
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 1440 * retention.var.5min`				
VAR Summary (30 minutes)	Medium	300	• IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	`retention.var.30min`
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 48 * retention.var.30min`				
VAR Summary (hourly)	Medium	300	• IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	`retention.var.hourly`
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * 24 * retention.var.hourly`				
Media Control Platform				
VAR Summary (daily)	Medium	300	• IVR Profiles • Tenants	`retention.var.daily`

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
			• MCPs • IVR Actions • unique call-end reasons	
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.hourly`				
VAR Summary (weekly)	Medium	300	• IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	`retention.var.weekly`
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.weekly/7`				
VAR Summary (monthly)	Medium	300	• IVR Profiles • Tenants • MCPs • IVR Actions • unique call-end reasons	`retention.var.monthly`
Calculation: `300 * (number of IVR Profile + number of tenants) * number of MCPs * (number of IVR Actions +1) * number of unique call-end reasons * retention.var.monthly/30`				
SQA Latency (hourly)	Medium	600	Number of components	`retention.latency.hourly`
Calculation: `600 * (number of components) * retention.latency.hourly * 24`				
Media Control Platform				
SQA Latency (daily)	Medium	600	Number of components	`retention.latency.daily`

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
Calculation: `600 * (number of components) * retention.latency.daily`				
SQA Latency (weekly)	Medium	600	Number of components	`retention.latency.weekly`
Calculation: `600 * (number of components) * retention.latency.weekly/7`				
SQA Latency (monthly)	Medium	600	Number of components	`retention.latency.monthly`
Calculation: `600 * (number of components) * retention.latency.monthly/30`				
SQA Failure Details	Medium	500	Calls per day Failure rate percentage	`retention.sq.failures`
Calculation: `500 * calls per day * failure rate percentage * retention.sq.failures`				
SQA Failure Summary (hourly)	Medium	200	• MCPs • IVR Profiles	`retention.sq.hourly`
Calculation: `200 * number of MCPs * number of IVR Profiles * retention.sq.hourly * 24`				
SQA Failure Summary (daily)	Medium	200	• MCPs • IVR Profiles	`retention.sq.daily`
Calculation: `200 * number of MCPs * number of IVR Profiles * retention.sq.daily`				
SQA Failure Summary (weekly)	Medium	200	• MCPs • IVR Profiles	`retention.sq.weekly`
Calculation: `200 * number of MCPs * number of IVR Profiles * retention.sq.weekly/7`				
Media Control Platform				
SQA Failure Summary (monthly)	Medium	200	• MCPs	`retention.sq.monthly`

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
			• IVR Profiles	
Calculation: `200 * number of MCPs * number of IVR Profiles * retention.sq. monthly/30`				
Call Control Platform				
CDR	Very High	600	Calls per day	`retention.cdr`
Calculation: `600 * calls per day * retention.cdr`				
Operational Reporting (5 minutes)	Medium	300	• CCPs • IVR Profiles	`retention.operations.5min`
Calculation: `300 * (number of IVR Profiles +1) * number of CCPs * 1440 * retention.operations.5min + 300 * number of CCPs * 1440 * retention.operations.5min` **Note:** The first product is for the arrivals that are stored per IVR Profile for each CCP. The second product is for the peaks that are stored for each CCP.				
Operational Reporting (30 minutes)	Medium	300	• CCPs • IVR Profiles	`retention.operations.30min`
Calculation: 300 * (number of IVR Profiles +1) * number of CCPs * 48 * retention.operations.30min + 300 * number of CCPs * 48 * retention.operations.30min				
Operational Reporting (hourly)	Medium	300	• CCPs • IVR Profiles	`retention.operations.hourly`
Calculation: `300 * (number of IVR Profiles +1) * number of CCPs * 24 * retention.operations.hourly + 300 * number of CCPs * 24 * retention.operations.hourly`				
Call Control Platform				
Operational Reporting (daily)	Medium	300	• CCPs • IVR Profiles	`retention.operations.daily`
Calculation:				

Data type	Usage	Estimated disk storage in bytes	Required estimates	Retention periods
`300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.daily + 300 * number of CCPs * retention.operations.hourly`				
Operational Reporting (weekly)	Medium	300	• CCPs • IVR Profiles	`retention.operations.weekly`
Calculation: `300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.weekly / 7+ 300 * number of CCPs * retention.operations.weekly / 7`				
Operational Reporting (monthly)	Medium	300	• CCPs • IVR Profiles	`retention.operations.monthly`
Calculation: `300 * (number of IVR Profiles +1) * number of CCPs * retention.operations.monthly / 30 + 300 * number of CCPs * retention.operations.monthly / 30`				
Events	Very High	500	• events per call • calls per day	`retention.events`
Calculation: `500 * number of events per call * number of calls per day * retention.cdr`				

top | toc

Bandwidth Usage

The following tables describe the bandwidth usage for the following components:

- Media Control Platform: Table: Media Control Platform Bandwidth Usage
- Call Control Platform: Table: Call Control Platform Bandwidth Usage
- Reporting Server: Table: Reporting Server Bandwidth Usage

Media Control Platform Bandwidth Usage

The table below describes the bandwidth usage when bi-directional traffic exists between the Media Control Platform and other servers.

Table: Media Control Platform Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Media Control Platform and SIP components			
SIP	• Simple inbound call: 5KB per call • Outbound with Supplementary Services Gateway: 10KB per call	Very high	SIP traffic can vary, depending on the call flow, the amount of user data, and number of treatments applied to the call.
Between Media Control Platform and MRCPv1			
RTSP MRCP RTP	• ASR: 8 KB per recognition, and 8 KB/sec of RTP traffic • TTS: 3 KB per prompt, and 8 KB/sec of RTP traffic	Very high	RTP traffic is uni-directional only.
Between Media Control Platform and MRCPv2			
SIP MRCP RTP	• ASR: 15 KB per recognition, and 10 KB/sec of RTP traffic • TTS: 6 KB per prompt, and 8 K/sec of RTP traffic	Very high	RTP traffic is uni-directional only.
Between Media Control Platform and RTP components			
RTP	• PCMU/PCMU/G.722: 20 KB/sec per call leg • G.729: 6 KB/sec per call leg • G.729d: 5.6 KB/sec per call leg • G.729e: 7 KB/sec per call leg • G.729-16: 8 KB/sec per call leg • G.726-24: 10 KB/sec per call leg • G.726-32: 12 KB/sec per call leg • G.726-40: 14 KB/sec per call leg • GSM: 7.3 KB/sec per call leg • AMR: 2-7.3 KB/sec per call leg • AMR-WB: 5-10 KB/sec per call leg (the rate varies, depending on the audio data) • H.263/H.264-1998: 10-70 KB/sec per call leg (the rate varies, depending on video data) • H.264: 20-90 KB/sec per call leg (the	Very high	Examples of RTP components are: • RTSP software • Soft phone • Media gateway

Protocol	Estimated bi-directional traffic	Criticality	Comments
	rate varies, depending on video data)		
Between Media Control Platform and HTTP Server/Proxy Server			
HTTP	1 KB per request and content size of the VoiceXML page or audio file in the HTTP request and response.	Very high	HTTP traffic can vary, based on the number of files that are used by the VoiceXML application, the maxage and `maxstale` settings of the VoiceXML application, and the expiry settings on the HTTP server.

Call Control Platform Bandwidth Usage

The table below describes the bandwidth usage when bi-directional traffic exists between the Call Control Platform and other servers.

Table: Call Control Platform Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Call Control Platform and SIP components			
SIP	Simple inbound call without join: ~7 KB per session Inbound call starting a simple dialog: ~20 KB per session	Very high	Significantly dependent on call flow and network conditions. If the network connection is poor, messages could be resent according to the SIP protocol.
Between Call Control Platform and HTTP Server/Proxy Server			
HTTP	1 KB per request and content size of the CCXML page in the HTTP request and response.	Very high	HTTP traffic can vary, based on the number of files that are used by the CCXML application, the maxage and maxstale settings of the CCXML application, and the expiry settings on the HTTP server.

For information about bandwidth usage for the Management Framework components, see the Management Framework chapter in this guide.

Reporting Server Bandwidth Usage

The table below describes the bandwidth usage when bi-directional traffic exists between the Reporting Server and other servers.

Table: Reporting Server Bandwidth Usage

Protocol	Estimated bi-directional traffic	Criticality	Comments
Between Reporting Server and Media Control Platform			
Proprietary (per call)	CDR: 1 KB per callEvents: 1 KB per call	Very high	CDR: 2 updates per call, 400 bytes per update. Events: 10 events per call, 100 bytes per event. **Note:** The number of updates per call depends on the application used.

Protocol	Estimated bi-directional traffic	Criticality	Comments
Proprietary (Operational Reporting)	OR: 100 bytes/min.OR: 100 bytes per IVR Profile per minute.	Low	One update per minute for peak (~50 bytes), and one update per minute for arrivals (~50 bytes).
Proprietary (SQA)	SQA: 50 KB per 15 min.SQA: 3 KB per IVR Profile per minute	Low	This depends on the frequency at which the SQA is configured to send data upstream to the Reporting Server. The default is 15 minutes. If the deployment is configured differently, the estimate must be adjusted.
Between Reporting Server and Resource Manager			
Proprietary (per call)	CDR: 3 KB per call	Very high	CDR: 7 updates per call, 400 bytes per update. **Note:** The number of updates per call depends on the application used.
Proprietary (OR)	OR:100 bytes per IVR Profile per minuteOR: 100 bytes per tenant per minute OR:100 bytes per DN per minuteOR: 100 bytes per CTI Connector or PSTN Connector component per minute **Note:** These data usage results are only for the IVR Profile, Tenant, Component, and DN that are invoked during each 5-minute period.	Medium	Two updates per minute per IVR Profiles, 50 bytes per update.Two updates per minute per tenant, 50 bytes per update.Two updates per minute per CTI Connector/PSTN Connector component, 5 bytes per update.Two updates per minute per DN, 50 bytes per update.
Between Reporting Server and Call Control Platform			
Proprietary (per call)	CDR: 1 KB per callEvents: 0.5 KB per call	Very high	CDR: 2 updates per call, 400 bytes per update.Events: 5 events per call, 100 bytes per event. **Note:** The number of updates per call depends on the application used.
Proprietary (OR)	OR: 100 bytes per minute OR: 100 bytes per IVR Profile per minute	Low	One update per minute for peak (~50 bytes), and one update per minute for arrivals (~50 bytes).
Between Reporting Server and an Off-board Reporting Database			
Proprietary (database vendor)	The sum of all estimates between the Reporting Server and all the Media Control Platform, Call Control Platform, and Resource Manager servers.	Very high	This bandwidth estimate applies when the database is off-board only (on a different server).

top | toc

Traffic and Capacity Testing

Overview

Use this section to determine the required capacity of your GVP servers, based on anticipated traffic characteristics or by running tests on an existing system.

When measuring peak capacity on a single GVP machine, CPU usage is usually the determining factor—memory has not been an issue in most test cases. Therefore, the sample test results in this section concentrate on CPU usage and other criteria.

In addition, the Media Resource Control Protocol (MRCP) server that supports Automatic Speech Recognition (ASR) applications, must not share a host with a GVP server. You can use multiple MRCP servers for a particular test, however, it is important that the MRCP resources do not cause a bottleneck during testing.

This section contains test summary tables to assist in the difficult task of sizing in the face of so much raw data contained by the tables in the following sections. Each table is prefaced with a description of its intent, with suggestions for interpreting and applying the data.

The complexity of VoiceXML and CCXML applications impacts capacity testing, therefore, the Genesys QA performance testing results in this section are derived from test cases using four different VoiceXML applications and two different CCXML applications.

VoiceXML Application Profiles

VoiceXML performance testing was conducted on four major application profiles. Their characteristics are outlined in the tables below. The call flow duration for each application profile is for a single call or CD1 (see Call Duration (CD) and Peak Capacity (PC)).

Profile: VoiceXML_App1	Profile: VoiceXML_App2
A simple DTMF-only application designed to refill calling cards. • Total number of digits (DTMF input only) = 52, including:	A complex application designed for insurance coverage inquiries. • Speech input, including: • Type of request

<table>
<tr><th>Profile: VoiceXML_App1</th><th>Profile: VoiceXML_App2</th></tr>
<tr><td>

Toll free number from the back of the card
Refill card PIN number
Refill dollar amount
Credit card number
Credit card expiration date
Zip Code of caller

Number of VoiceXML pages = 18
VoiceXML complexity = low
Number of audio prompts = 9
Number of audio files used in prompts (no TTS) = 107
ECMA script complexity = moderate
Number of VoiceXML pages = 6
Number of Java script help functions in each VoiceXML page = 13
Call flow duration:

74 seconds (pre GVP 8.1.2)
76 seconds (GVP 8.1.2 and higher)

</td><td>

ID card number
Confirmation
Relationship with insurance plan holder
Date of birth confirmation

Number of VoiceXML pages = 10
VoiceXML complexity (~ 1 MB of content) = High
Number of audio prompts = 7
Number of audio files used in prompts (3 with TTS) = 29
ECMA script complexity = high
Call flow duration:

70 seconds (ASR engine)
55 seconds (ASR engine simulator)

</td></tr>
</table>

<table>
<tr><th>Profile: VoiceXML_App3</th><th>Profile: VoiceXML_App4</th></tr>
<tr><td>
QA ASR/TTS load application.

Speech input, including:

Words
Digits
Hotkey (NGI)
Yes or no confirmation

Number of VoiceXML pages = 1
VoiceXML complexity = low
Number of audio prompts = 7 prompts involve 7 audio files and 7 TTS
ECMA script complexity = low
Call flow duration = 62 seconds

</td><td>
Composer-generated application designed for IVR-assisted banking.

Input a total of 20 digits (DTMF only):

Input current customer number
Confirm contact ID
Input debit menu option
Input debit banking menu
Input personal option
Input 6 digit secure code

Number of VoiceXML pages = 20
VoiceXML complexity = medium (~ 400 KB of content)
Number of audio prompts = 6 (no TTS, 12 audio files)

</td></tr>
</table>

Profile: VoiceXML_App3	Profile: VoiceXML_App4
	• ECMA script complexity = moderate (4 general JavaScript function files) • Call duration = 85 seconds

Profile: VoiceXML_App5	Profile: VoiceXML_App6
VoiceXML_App1 with IVR recording function. In addition to running the VoiceXML_App1 application, IVR recording was also started when the VoiceXML_App1 began and the call was recorded until the end. **Recording details** • No of channels = 2 • Recording type = mp3 • Bit rate = 16 kbps • Recording destination = http • Recording metadata = enabled	Simple IVR recording application with continuous speech input from the caller. • Number of VoiceXML pages = 1 • VoiceXML complexity = low • Number of audio prompts = 2 (2 audio files) • Call flow duration = 75 seconds (NGI) **Recording details** • No of channels = 2 • Recording type = mp3 • Bit rate = 16 kbps • Recording destination = http • Recording metadata = enabled

Profile: VoiceXML_App7	Profile: VoiceXML_App8

A simple voice input application designed to get transcript from Google ASR directly from MCP (NativeGSR).
• Number of VoiceXML pages = 1
• VoiceXML complexity = low
• Number of audio prompts = 5
• Call flow duration:
  • ~ 5.5 seconds
	A simple voice input application designed to use Text to Speech service from Google directly from MCP (NativeGTTS).
	• Number of VoiceXML pages = 1
	• VoiceXML complexity = low
	• Number of TTS prompts = 1
	• Number of characters in TTS prompt = 344
	• Call flow duration:
	  • ~ 22.5 seconds

CCXML Application Profiles

CallControlXML (CCXML) performance testing was conducted on two major application profiles. Their characteristics are outlined below. The call flow duration for each application profile is for a single call or CD1 (see Call Duration (CD) and Peak Capacity (PC)).

Profile: CCXML_App1	Profile: CCXML_App2

An outbound application that joins multiple call legs, dialogs, and conferences.
• Includes the following steps:
  ◦ Call customer and connect to a dialog
  ◦ Call agent and connect to dialog
  ◦ Exit agent dialog
  ◦ Exit customer dialog
  ◦ Create conference
  ◦ Join customer and agent to conference
  ◦ Disconnect agent
  ◦ Disconnect customer
  ◦ Destroy conference
• Number of CCXML (JSP) pages = 2
• CCXML complexity = medium
• Customer call duration = 8.7 seconds
• Agent call duration = 8.6 seconds
• Conference call duration = 6 seconds
	Simple conference recording call.
	• Includes the following steps:
	  ◦ Create a call to agent
	  ◦ Agent receives an invite and a dialog is created for agent to ring back
	  ◦ Agent answers the call and a conference is created to join caller and agent
	  ◦ Conference is established and dialog is created for recording
	  ◦ Call is disconnected from caller after 15 seconds of recording
	• Number of CCXML pages = 1
	• Number of VoiceXML pages = 2
	• CCXML complexity = medium
	• Call duration = 21 seconds

top | toc

Traffic and Capacity Terms

Capacity Metrics and Formulas

- Call Arrivals Per Second (CAPS)
- Port Density (PD), also known as Peak Ports
- System Capacity

Performance Metrics and Formulas

- Call Duration (CD) and Peak Capacity (PC)
- Call Setup Latency (CSL)
- Caller Perceived Latency (CPL), also known as Response Time Latency
- Call Passrate (CP)

Capacity Metrics and Formulas

Two units of measure are used for capacity planning: *Call Arrivals Per Second* and *Port Density*. This section also provides the formulas used to calculate capacity and performance.

Call Arrivals Per Second (CAPS)

CAPS measures traffic within the system. For example, 10 CAPS means that GVP is receiving 10 calls every second, which is considered busy traffic. CAPS is similar to Busy Hour Call Attempts (BHCA) or Centum Call Seconds (CCS), which is the legacy engineering term for telephony traffic.

Use the following formula to calculate CAPS in terms of CCS:
`CAPS = CCS/36`

CAPS measures can be applied to components which handle messages or data associated with a call. For example, the reporting server will have a CAPS value based on the number of call records written to it, which will often relate one-to-one with a completed call.

Throughout this chapter, including the tables, the capacity of a function and/or component is defined by its Peak CAPS (the maximum number of calls per second that the system can handle for that function without suffering from latency).

Port Density (PD) *and Average Port Density*

PD is the maximum number of simultaneous calls that can be served by GVP at any given time. In the tables of this document, PD is called “Peak Ports” because it also specifies the number of ports that are required to handle the call traffic.

Use the following formula to calculate Port Density:
`PD = CAPS x Avg(CD)`
...where Avg = Average.

Average Port Density (APD) is the average number of simultaneous calls that are being served by GVP at any given time. The formula to calculate APD is:
`APD = CAPS x Average(Call Duration)`

Due to PD being a random variable, it can be higher than its average value APD. Use the following formula to calculate Port Density:
`PD = APD + 3*SQRT(APD)`
...where SQRT(x) is the square root of x.

System Capacity is a function of the maximum number of ports (PD) or maximum call-arrival rate (CAPS) at which GVP can maximize its use of hardware resources, while maintaining all of the criteria within a predefined threshold.

Performance Metrics and Formulas

Four units of measure are used to assess performance—*Call Duration*, *Call Setup Latency*, *Caller Perceived Latency*, and *Call Passrate*.

Call Duration (CD) *and Peak Capacity (PC)*

CD is the length of time that a call stays in the GVP system. Use CAPS and CD to calculate the port density required for handling such traffic.

Instead of measuring individual Caller Perceived Latencies within an application under test, you can use data collected by GVP to measure the increase in the total call duration to determine system loading.

For a single call, the assumption is that the normal call duration (where the hang-up is performed by the application) is CD1. When the load increases on the system, the call duration is expected to increase due to an increase of latencies within the application. Assume that for x simultaneous calls in the system, the measured average call duration is Avg(CDx) and the measured 95th percentile call duration is 95%tile(CDx). The capacity measurement goal is

```
Avg(CDx) / CD1 <= 110%
95%tile(CDx) / CD1 <= 120%
```

When the 110% (and 120%) threshold is reached, the call volume x is considered to be the Peak Capacity (**PC**) for this criterion.

Call Setup Latency (CSL)

CSL is the delay between the initial SIP INVITE message and the first audible RTP packet sent from GVP.
CSL consists of the following requests and responses:

- User SIP INVITE request received > SIP 200 OK response sent.
- SIP 200 OK response sent > User SIP ACK request sent.
- User SIP ACK request sent > First audible media response sent.

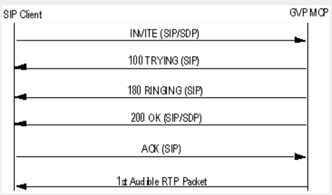


Figure: Typical SIP Call Flow #1

One example is the dialog in a typical call flow shown in **Figure: Typical SIP Call Flow #1**.

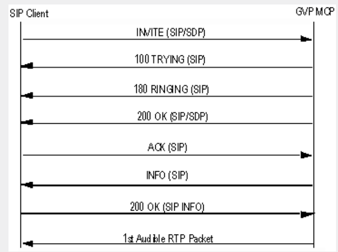


Figure: Typical SIP Call Flow #2

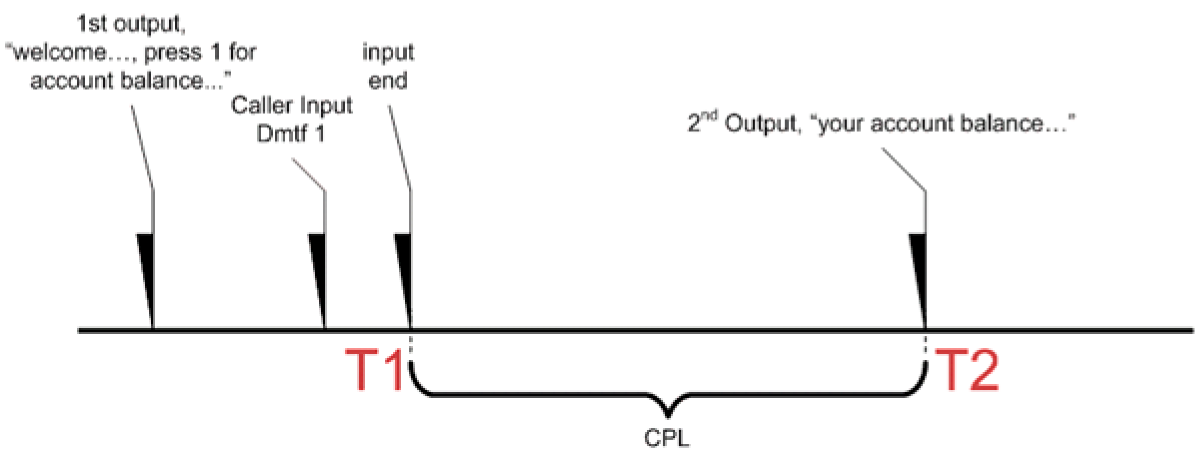
CSL is the delay between the time that the initial SIP INVITE (top line from the typical call flows shown in both diagrams above)
is received to the time that the first audible packet (bottom line from **Figure: Typical SIP Call Flow #2**) is sent out by GVP.

Other than call setup latency, SIP INFO response latency (from SIP INFO with MSML embedded to 200 OK response from MCP for SIP INFO request) is also an important factor. It should be measured and reported separately, although this duration is already part of call setup latency.

Caller Perceived Latency (CPL)

CPL, also known as Response Time Latency (RTL), is the time between the last user input (speech or DTMF) and the next prompt. As illustrated below, the time between T1 and T2 is the period of CPL.

Use the following formula to calculate CPL:
`CPL = T2 - T1`



Caller Perceived Latency

CPL is impacted by the following factors:

- Recognition engines.

- End of speech or a DTMF time out.
- Application pages and prompts.
- Grammars caching and fetching mechanisms.
- The size of application pages.
- Call traffic, including call arrival rate and call duration.
- Speech density: during a call, the percentage of time that speech or DTMF detection is on and the caller can provide input.
- The size of speech recognition grammars and how often they are used in an application.
- Backend operation: the length of time required to obtain information from the business layer (such as, database queries and CRM findings) and return the results to the caller.

Call Passrate (CP)

CP is the measurement that indicates how many calls finished the pre-defined call flow successfully during the performance load test. For example, if a total of 1000 calls were placed, and five calls did not finish according to the call flow for any reasons, then the pass rate is:
`(1000-5)/1000 = 99.5%`

The capacity measurement goal is:

ASR/TTS dependent application	`Passrate >= 99.95%` or `Error rate <= 0.05%`
DTMF only application	`Passrate >= 99.99%` or `Error rate <= 0.01%`

CP is similar to call duration, in that when the threshold is reached at call volume x, then x is considered to be the peak capacity for this criterion.

VoIP Capacity Test Summary Tables

Some capacity test summaries in this section were performed on systems with hardware specifications other than those in Hardware and Operating Systems Tested. Major differences in test results can occur, depending on the CPU model and the number of CPUs that are used.

Certain tests may not be conducted with the hardware specified in Hardware and Operating Systems Tested; the major difference is the CPU model and the number of CPUs being used. The Hardware column in the tables below describes the CPU setup that was used in each test and the observed capacity. The results are based on Next Generation Interpreter (NGi) configured in Media Control Platform (MCP), unless it is stated as GVPi.

VoiceXML_App3 was used for both single server testing and PSTNC testing. See Table: Single Server All-In-One Capacity Testing and Table: PSTN Connector and SSG Capacity Testing.

Click a link in the list below for specific details about intent and use above each table:

- Table: GVP VOIP VXML/CCXML Capacity Testing
- Table: Multiple VMs Versus Multiple MCP Capacity Testing
- Table: Single Server All-In-One Capacity Testing
- Table: Standalone VM with Single MCP Instance Capacity Testing

GVP VOIP VXML/CCXML Capacity Testing

This table shows the fundamental performance of a single physical server process in terms of peak throughput and peak port capacity; either VoiceXML applications for MCP or CCXML for CCP. You can use this table as the first basis of your assessment.

Table 2: GVP VOIP VXML/CCXML Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Windows				
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	17 (preferred)	1300 (preferred)	*Preferred* means the highest capacity that the system can sustain while maintaining optimal user experience.
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	17	1300	Using TCP and TLS.

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	23.6 (peak)	1800 (peak)	Ignore call setup latency threshold on Window 2003 and 2008 R2, x64. *Peak* means the highest capacity that the system can sustain regardless of the user experience.
VoiceXML_App1	1x HexCore Xeon x5770 2.66GHz	26 (peak)	2000 (peak)	Ignore call setup latency threshold, Windows 2008 R2 x64 SP1
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	10 (preferred)	800 (preferred)	Using GVPi.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	7.2	400	MCP on a physical server. Tested with offboard NSS engine MRCP v1.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	7.2	400	MCP on a physical server. Tested with offboard NSS engine MRCP v2 (NSS 6.2.x + NR 10.2.x + NV 5.7.x) with session XML enabled. GVP 8.1.7 or later.
VoiceXML_App2	1x HexCore Xeon X5670 2.93GHz	8	450	MCP on a physical server. Tested with offboard NSS engine MRCP v2 (NSS 6.2.x + NR 10.2.x + NV 5.7.x) with session XML disabled. GVP 8.1.7 or later.
VoiceXML_App2	2x Core 2 Dual Xeon x5160 3.00 GHz	4.5	250	MCP on a physical server. Tested with simulated speech server.
VoiceXML_App2	2x Core 2 Quad Xeon x5355 2.66 GHz	1	60 (GVPi)	MCP on a physical server. Tested with Nuance Speech Server.
VoiceXML_App4	2x Core 2	9.4	800	

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
	Quad Xeon x5355 2.66 GHz			
CCXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	30	N/A	
CCXML_App2	2x Core 2 Quad Xeon x5355 2.66 GHz	20	420	
Linux				
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	17 (preferred)	1300 (preferred)	
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	23.6 (peak)	1800 (peak)	Peak, ignoring call setup and tear-down latency threshold.
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	23.6	1800	Using TCP and TLS.
VoiceXML_App1	2x Core 2 Quad Xeon x5355 2.66 GHz	14.5	1100	Inband DTMF.
VoiceXML_App2	2x Core 2 Quad Xeon x5355 2.66 GHz	7.2	400	MCP on a physical server. Tested with simulated speech server.

Multiple VMs vs. Multiple MCP Capacity Testing

This table provides a comparison of capacity testing results when multiple virtual machines (VMs) are used versus multiple Media Control Platform instances.

The table below shows the effect of stacking server processes on the same hardware server where there is one MCP associated with a VM instance on the same hardware server. The effect is the increased total port capacity that you can achieve using stacked processes provides a comparison of capacity testing results when multiple virtual machines (VMs) are used versus multiple Media Control Platform instances.

Table 3: Multiple VMs vs. Multiple MCP Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Using VMWare				
VoiceXML_App1 1 VM	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	17	1300	One VM image is configured and enabled with only one MCP installed in the image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 2 VMs	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	21	1600	Two VM images are configured and enabled with only one MCP installed in each image. Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 2 VMs	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	21	1600	VM images (using VMWare ESXi 5.0) are configured and enabled with 4 Media Control Platform instances—2 installed in each image.Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 4 VMs	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	29	2200	VM images (using VMWare ESXi 5.0) are configured and enabled with 1 Media Control Platform instance only installed in each image.Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 4 VMs	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	26	2000	VM images (using VMWare ESXi 5.0) are configured and enabled with 8 Media Control Platform instances—2 installed in each image.Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1	2x Core 2	34	2600	VM images (using VMWare ESXi 5.0) are

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
8 VMs	Quad Xeon x5355 2.66 GHz 12GB RAM			configured and enabled with 1 Media Control Platform instance only installed in each image.Guest OS is Windows 2008 Server SP2 x86.
VoiceXML_App1 1 VM	2x Core 2 Quad Xeon x5355 2.66 GHz 4GB RAM	8 (tested)	600 (tested)	VM image (using VMWare ESXi) is configured and enabled with all GVP components (except Reporting Server) together with SIP server.Guest OS is Windows 2003 Server.
VoiceXML_App1 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad-Core Xeon E5620 2.40GHz 16GB RAM	39	3000	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1.
VoiceXML_App2 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad-Core Xeon E5620 2.40GHz 16GB RAM	8.6	600	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1. Tested with Nuance Speech Servers which run on another 4VMs of same hardware spec as MCP.
VoiceXML_App4 (4 VMs, 4 MCPs, 1 MCP per VM)	2x Quad Core Xeon E5620 2.40GHz 16GB RAM	21	1800	4 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is Windows 2008 Server R2 x64 SP1.
VXML_App1 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	52	4000	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 5.8 x64.
VXML_App1 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	3.9	300	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 6.4 x64. GVP 8.1.7 or later.
VXML_App2 (6 VMs, 6 MCPs, 1 MCP per VM)	2x Hex-Core Xeon X5675 3.06GHz 32GB RAM	8.6	600	6 VMs under EXSi 5.0 are configured and enabled with only one MCP installed in each VM. Guest OS on each VM is RHEL 5.8 x64. Tested with Nuance Speech Servers which run on another 4VMs of a host of 2x Quad Core Xeon E5620.
VMXL_App1	VM Profile	14.5	1100	One VM image is configured and enabled with

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
(1 VM)	6			only one MCP installed in the image. Guest OS is Windows 2016 x64.
VMXL_App1 (1 VM)	VM Profile 6	15.8	1200	One VM image is configured and enabled with only one MCP installed in the image. Guest OS is RHEL 7.0.
Not Using VMWare				
VoiceXML_App1 1 Media Control Platform instance	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	17	1300	All Media Control Platform instances are configured on one server. Windows 2008 Server, SP2, x86.
VoiceXML_App1 2 Media Control Platform instances	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	27.5	2100	
VoiceXML_App1 4 Media Control Platform instances	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	30	2300	
VoiceXML_App1 8 Media Control Platform instances	2x Core 2 Quad Xeon x5355 2.66 GHz 12GB RAM	27.5	2100	
VoiceXML_App1 2 Media Control Platform instances	2x Core 2 Quad Xeon x5355 2.66 GHz 4GB RAM	39.4 (peak)	3000 (peak)	All Media Control Platform instances are configured on one server. Squid is bypassed and call setup latency threshold is ignored. Windows 2003 Server only.

Single Server All-In-One Capacity Testing

This table describes the capacity testing for a single server with multiple components installed (see Comments column). Tests were performed using a single instance of the Media Control Platform on Windows and Linux systems with 1 Core 2 Dual Xeon x5160, 3.0 GHz CPUs with 8GB RAM. This table shows the effect of having many GVP processes, including Nuance speech components, on just one physical server, which Genesys calls "the single server solution."

Table 4: Single Server All-In-One Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Windows 2008, SP2, x86 and Windows 2008 R2				
VoiceXML_App1	1x Core 2 Dual Xeon x5160 3.0 GHz 8GB RAM	7.9	600	A single server hosting Management Framework, Media Control Platform, Resource Manager, Reporting Server, Web Application Server (WAS), and SIP Server.
VoiceXML_App2 MRCP v1	1x Core 2 Dual Xeon x5160 3.0 GHz 8GB RAM	1.2	100	
VoiceXML_App3 MRCP v1	1x Core 2 Dual Xeon x5160 3.0 GHz 8GB RAM	2.5	160	
VoiceXML_App3 MRCP v2	1x Core 2 Dual Xeon x5160 3.0 GHz 8GB RAM	1.9	120	
Red Hat Enterprise Linux 4				
VoiceXML_App1	1x Core 2 Dual Xeon x5160 3.0 GHz 8GB RAM	4 (maximum CAPS tested)	300 (ports tested)	A single server hosting an Oracle DB Server, Management Framework, Reporting Server, Media Control Platform, Resource Manager, SIP Server, Web Application Server, and Linux.

Standalone VM with Single MCP Instance Capacity Testing Table

This table describes the capacity testing results performed on a standalone VM with single MCP instance (see Comments column). Tests were performed using a single instance of the MCP on VMs running Windows and Linux systems with 2 Virtual Cores, Xeon E5-2683 V4, 2.099 GHz CPUs.

Table 5: Standalone VM with Single MCP Instance Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
VMWare				
VoiceXML_App5	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	6.6	500	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS RHEL 7.0, x64.
VoiceXML_App6	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	5.2	350	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS RHEL 7.0, x64
VoiceXML_App5	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	6.6	500	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS Windows 2012, x64
VoiceXML_App6	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	5.2	350	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS Windows 2012, x64
VoiceXML_App7	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	18.18	100	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS Windows 2016, x64
VoiceXML_App7	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	18.18	100	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS RHEL 7, x64
VoiceXML_App8	2 Virtual Cores, Xeon E5-2683 V4	17.3	400	VM images (using VMWare ESXi 6.0)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
	2.099 GHz			are configured and enabled with one MCP instance installed in Guest OS Windows 2016, x64
VoiceXML_App8	2 Virtual Cores, Xeon E5-2683 V4 2.099 GHz	30.2	700	VM images (using VMWare ESXi 6.0) are configured and enabled with one MCP instance installed in Guest OS RHEL 7, x64

Important

- The limitation factor for *Voicexml_app7* came when the network usage was at ~20 Mbps (including uplink & downlink) for both Windows & Linux.
- The limitation factor for *Voicexml_app8* came when the network usage was at ~90 Mbps and ~180 Mbps (including uplink & downlink) for Windows and Linux respectively.

The Mbps network usage numbers (20, 90, and 180) are to indicate the network usage in the MCP hosts at max ports and a minimum requirement for achieving the claimed max ports. Higher bandwidth doesn't necessarily mean it can scale higher since MCP is limited by the number of max network threads and buffer size it can utilize.

Component Capacity Test Case Tables

Capacity test case results for the GVP components appear in the following tables:

- Media Control Platform Capacity Testing (Physical Servers Running Windows)
- Media Control Platform Capacity Testing (Physical Servers Running Linux)
- Media Control Platform / Media Server Capacity (Virtual Servers)
- Resource Manager and MRCP Proxy Capacity Testing
- Reporting Server Capacity Testing
- CTI Connector and CTI Connector with ICM Capacity Testing
- PSTN Connector and SSG Capacity Testing

For additional sizing information for Genesys Media Server with SIP Server, see the chapter “Genesys Administrator” in the Genesys Hardware Sizing Guide. The capacity testing results for the Media Control Platform are described in the next three topics. Tests were performed using a single instance of the Media Control Platform on Windows and Linux systems with 2x Core 2 Quad, Xeon x5355, 2.66 GHz CPUs.

Tip

Media Services Only: If your deployment is limited to Media Services, then see critical information for sizing the MCP in Media Control Platform Capacity Testing and the section “Genesys Media Server Sizing with SIP Server” in the Genesys Hardware Sizing Guide.

Media Services plus VoiceXML Applications: If you have both types of services on the same GVP system, then the actual performance will be a roughly proportional combination of media service performance and VoiceXML performance. As this is difficult to determine, Genesys recommends that you default to the media performance metrics if transcoding is prevalent or media services are significant.

Media Control Platform Capacity Testing (Physical Servers Running Windows)

This table does not focus on GVP as a whole, but rather shows the impact of media services (announcements, call parking, bridging, conferencing, transcoding and video) on the performance of the Media Control Platform (MCP).

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
Audio bridge transfer G711u <-> G711u (baseline ~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	6.8	800	Bi-directional audio streams. Tested on Windows 2003.
Transcoding with bridge transfer G711u <-> AMR (~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	2.6	300	Bi-directional transcoding. Tested on Windows 2003.
Transcoding with bridge transfer G711u <-> AMR-WB (~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	2.0	230	Bi-directional transcoding. Tested on Windows 2003.
Transcoding with bridge transfer G711u <-> G722 (~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	3.0	350	Bi-directional transcoding. Tested on Windows 2003.
Transcoding with bridge transfer G711u <-> G726 (~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	2.6	300	Bi-directional transcoding. Tested on Windows 2003.
Transcoding with bridge transfer G711u <-> G729 (~117 seconds duration)	Quad-Core Xenon 5355 2.66GHz	3.0	350	Bi-directional transcoding. Tested on Windows 2003.
SRTP with bridge transfer – G.711u (~67 seconds duration)	Quad-Core Xenon 5355 2.66GHz	18	1200	The capacity is the same for RTP and SRTP of both encryption and decryption, one direction only of audio stream. Tested on Windows 2003.
MSML CPD + VXML	Quad-Core Xenon	30	n/a	CPD enabled

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
dialog (helloworld) (8 seconds overall call duration which includes 2.5 seconds CPD time)	5355 2.66GHz			within MSML which also invoke a VXML dialog using default helloworld page. VXML dialog will start after CPD result returned the result of human successfully.
Netann announcement – 3 seconds audio	Quad-Core Xenon 5355 2.66GHz	120 (preferred) 200 (peak)	500 (preferred) 1100 (peak)	Preferred – with call setup + call tear down latency < 1sec (500ms each) Peak – ignore call setup/tear down delay
Netann announcement – 10 seconds audio	Quad-Core Xenon 5355 2.66GHz	90 (preferred) 150 (peak)	900 (preferred) 1500 (peak)	Preferred – with call setup + call tear down latency < 1sec (500ms each) Peak – ignore call setup/tear down delay
Netann Play Treatment - G.711u, G.729, GSM (~60 seconds audio)	Quad-Core Xenon 5355 2.66GHz	30	1800	No transcoding. The capacity is the same for G.711u, G.729, or GSM. Tested on Windows 2003.
Netann 2 party Call Recording - G.711u (~60 seconds duration)	Quad-Core Xenon 5355 2.66GHz	12	720 call legs (360 recording sessions)	Tested on Windows 2003.
Netann 2 party Call Recording - G.729 (~60 seconds duration)	Quad-Core Xenon 5355 2.66GHz	9	540 call legs (270 recording sessions)	Tested on Windows 2003.
Netann 2 party Call Recording - GSM (~60 seconds duration)	Quad-Core Xenon 5355 2.66GHz	8	480 call legs (240 recording sessions)	Tested on Windows 2003.
MSML Conference (all participants using the same codec) – G711u, G.729, GSM (3-party; ~60 seconds duration)	Quad-Core Xenon 5355 2.66GHz	6	360 participants (120 conference sessions)	The capacity is the same for G.711u, G.729, or GSM. Tested on Windows 2003.
MSML Conference (different codecs	Quad-Core Xenon 5355 2.66GHz	6	360 participants (120 conference	Tested on Windows 2003.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
between participants) – G711 and G.729 (3-party; ~60 seconds duration)			sessions)	
MSML Conference (different codecs between participants) – G711 and GSM (3-party; ~60 seconds duration)	Quad-Core Xenon 5355 2.66GHz	6	360 participants (120 conference sessions)	Tested on Windows 2003.
MSML Conference (3-party conference; all participants using the same code – G711, HR Timer disabled ~60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	6	360 participants (120 conference sessions)	Tested on Windows 2008 R2 x64 SP1 with HR Timer disabled in 8.1.6.
MSML Conference (3-party conference; all participants using the same code – G711, HR Timer enabled ~60 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	5	300 participants (100 conference sessions)	Tested on Windows 2008 R2 x64 SP1 with HR Timer enabled in 8.1.6.
MSML Conference (One giant conference with 3 speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711)	1x Six-Core Xeon X5675 3.06GHz	0.72	1300 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
MSML Conference (One giant conference with 3 speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec H263 + G.711)	1x Six-Core Xeon X5675 3.06GHz	0.61	1100 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
MSML Conference (One giant conference with 3	1x Six-Core Xeon X5670 2.93GHz	0.72	1300 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711, HR Timer disabled & gain control enabled)				one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
MSML Conference (One giant conference with 3 speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711, HR Timer enabled & gain control enabled)	1x Six-Core Xeon X5670 2.93GHz	0.56	1000 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
MSML Conference (One giant conference with 3 speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711, HR Timer enabled & gain control disabled)	1x Six-Core Xeon X5670 2.93GHz	0.78	1400 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
MSML Conference (One giant conference with 3 speakers; all other participants are listeners. Each participant stays 1800 secs (30 mins) in the conference. Codec G.711, HR Timer disabled & gain control disabled)	1x Six-Core Xeon X5670 2.93GHz	0.99	1800 participants (1 conference session)	Tested on Windows 2008 Server R2 x64 SP1 with only one MCP instance. Threaded outputs enabled (`conference.threadedoutputs = true`).
Note:				

Application Type	Hardware	Peak CAPS	Peak Ports	Comment

- *Preferred* means the highest capacity that the system can sustain while maintaining optimal user experience.
- *Peak* means the highest capacity that the system can sustain regardless of the user experience.

Media Control Platform Capacity Testing (Physical Servers Running Linux)

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comments
Audio bridge transfer G711u <-> G711u (baseline)(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	9.4	1100	Bi-directional audio streams. Tested on Linux RH EL5.
Transcoding with bridge transfer – G.711u <-> G.722(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	3	350	Bi-directional transcoding. Tested on Linux RH EL5.
Transcoding with bridge transfer – G.711u <-> G.726(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	2	240	Bi-directional transcoding. Tested on Linux RH EL5.
Transcoding with bridge transfer – G.711u <-> G.729(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	2	240	Bi-directional transcoding. Tested on Linux RH EL5.
Transcoding with bridge transfer – G.711u <-> AMR-WB(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	2	240	Bi-directional transcoding. Tested on Linux RH EL5.
Transcoding with bridge transfer – G.711u <-> AMR(~117 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	1.7	200	Bi-directional transcoding. Tested on Linux RH EL5.
SRTP with bridge transfer – G.711u(~67 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	22.4	1500	The capacity is the same for RTP and SRTP of both encryption and decryption. One direction audio stream. Tested on RH EL5.
SRTP with bridge transfer – 3gp of H.264 video (352x288) + AMR audio (~125	Quad-Core Xeon x5355 2.66 GHz	3.2	400	The capacity is the same for RTP and SRTP of both encryption and decryption. One

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comments
seconds duration)				direction RTP stream. Tested on RH EL5 x64.
MSML CPD + VXML dialog (helloworld)(8 seconds overall call duration which includes 2.5 seconds CPD time)	Quad-Core Xeon x5355 2.66 GHz	40	n/a	CPD enabled within MSML which also invoke a VXML dialog using default helloworld page. VXML dialog will start after CPD result returned the result of human successfully.
Netann announcement – 3 seconds audio	Quad-Core Xeon x5355 2.66 GHz	120 (preferred) 200 (peak)	500 (preferred) 1100 (peak)	Preferred – with call setup + call tear down latency < 1sec (500ms each) Peak – ignore call setup/tear down delay
Netann Play Treatment – G.711u(~60 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	30	1800	No transcoding.
Netann Play Treatment – video h263(+)(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	10	1200	No transcoding.
Netann Play Treatment – video 3gp/avi (h263)(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	8.3	1000	No transcoding.
Netann Recording Single Call – G.711u (raw, au & wav), G.722, G.726(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	8.3	1000	The capacity is the same for G.711u, G.722 & G.726.
Netann Recording Single Call – G.729, AMR(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	5.8	700	The capacity is the same for G.729 & AMR.
Netann Recording Single Call – AMR-WB(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	6.6	800	
Netann Recording Single Call – video	Quad-Core Xeon x5355 2.66 GHz	4.2	500	

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comments
raw h263(+)(~120 seconds duration)				
Netann Recording Single Call – video avi (h263+G.711u)(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	4	480	
Netann Recording Single Call – video 3gp (h263+amr)(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	2	240	
Netann Recording Single Call – video raw h264(~120 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	2	250	
Netann 2 party Call Recording - G.711u (~60 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	11	660 call legs (330 recording sessions)	
MSML Play announcement – one prompt (SIP INFO), one audio file - 3 seconds	Quad-Core Xeon x5355 2.66 GHz	80	260	Call duration 3.13 seconds and GVP precheck is on.
MSML Play announcement – one prompt (SIP INFO), one audio file - 10 seconds	Quad-Core Xeon x5355 2.66 GHz	200	2000	Call duration 10.34s and GVP precheck is off.
MSML Play announcement – one prompt (SIP INFO), two audio files – 4 + 6seconds	Quad-Core Xeon x5355 2.66 GHz	200	2000	Call duration 10.34s and GVP precheck is off.
MSML Play announcement – two prompts (SIP INFO), two audio file – 4 + 6 seconds, one file per prompt.	Quad-Core Xeon x5355 2.66 GHz	130	1400	Call duration 10.46s and GVP precheck is off.
MSML Play announcement – one prompt (SIP INFO), one audio file - 20 seconds	Quad-Core Xeon x5355 2.66 GHz	150	3000	Call duration 20.34s and GVP precheck is off.
MSML Play announcement –	Quad-Core Xeon x5355 2.66 GHz	130	2600	Call duration 20.35s and GVP

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comments
one prompt (SIP INFO), three audio files – 4+6+10 seconds				precheck is off.
MSML Play announcement – three prompts (SIP INFO), three audio filed – 4+6+10 seconds, one file per prompt	Quad-Core Xeon x5355 2.66 GHz	100	2000	Call duration 20.60s and GVP precheck is off.
MSML Conference (all participants using the same codec – G711u) 3-party; ~60 seconds duration)	Quad-Core Xeon x5355 2.66 GHz	6	360 participants (120 conference sessions)	The capacity is the same for G.711u, G.729, or GSM.

Note:

- *Preferred* means the highest capacity that the system can sustain while maintaining optimal user experience.
- *Peak* means the highest capacity that the system can sustain regardless of the user experience.

Media Control Platform / Media Server Capacity Testing (Virtual Servers)

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
Windows				
Video bridge transfer (H264 + AMR, 720P, 30fps, 1Mbps, level 3.1, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6	400 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.1, 720P -> CIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.11	8 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.1, 720P -> QCIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.21	16 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 30fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	7.2	500 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 1Mbps, level 3.0, VGA -> CIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.43	30 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 +	1x Six-Core Xeon X5675 3.06GHz	0.72	50 calls	Unidirectional down scale transcoding.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
AMR, 30fps, 1Mbps, level 3.0, VGA -> QCIF, 70 seconds duration)				Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, CIF, 30fps, 256Kbps, level 2.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	11.3	800 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, 30fps, 256Kbps, level 2.0, CIF -> QCIF, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	1.43	100 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 60fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	4.3	300 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video bridge transfer (H264 + AMR, VGA, 60fps, 1Mbps, level 3.1, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6.43	450 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA 1Mbps, level 3.1, 60fps -> 30fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.43	30 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA, 1Mbps, level 3.1, 60fps -> 15fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.13	9 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
Video bridge transfer (H264 + AMR, VGA, 30fps, 1Mbps, level 3.0, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	6	420 calls	Unidirectional rtp (video + audio) stream. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, VGA 1Mbps, level 3.0, 30fps -> 15fps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.34	24 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 1.5Mbps -> 500Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	0.5	35 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 1Mbps -> 192Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	1.12	80 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
Video transcoding with bridge transfer (H264 + AMR, CIF, 30fps, 500Kbps -> 192Kbps, 70 seconds duration)	1x Six-Core Xeon X5675 3.06GHz	2	140 calls	Unidirectional down scale transcoding. Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Answer Machine (~12.8 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	150	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Busy Machine (~7.7 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	120	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
MSML CPA Fax Machine (6.3 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	140	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA Human Machine (7.8 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	170	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML CPA SIT VC Machine (2.3 seconds duration)	2x Six-Core Xeon X5675 3.06GHz	130	n/a	MSML CPA only. Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec G711, 120 seconds duration)	2x Quad-Core Xeon E5620 2.40GHz	50	6000 calls	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (MP3, Any KHz, Any Kb, Cache enabled, Negotiated codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (MP3, 320Kbit, 44.1KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	18	2160 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	25	3000 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz Negotiated codec: G.711, 120	1x Six-Core Xeon X5670 2.93GHz	13	1560 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
seconds duration)				
MSML Play Announcement (MP3, 320Kbit, 44.1KHz, Negotiated codec: G.711, 120 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	12	1440 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec AMR and H263 CIF, 512Kbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	23	1380 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H263 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8	480 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 128Kbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	25	1500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 256Kbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	17	1000 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, CIF, 512Kbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	8.5	500 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play	1x Six-Core Xeon	22	1300 calls	Tested on 3 VMs of

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
Announcement (Codec H264 and AMR, 4CIF, 512Kbps 10fps, 60 seconds duration)	X5670 2.93GHz			EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 1Mbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	16	960 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	7.5	450 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 720P, 1Mbps 10fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	19	1100 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 2Mbps 15fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	9	540 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement (Codec H264 and AMR, 4CIF, 4Mbps 30fps, 60 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	4	240 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec H264 and AMR, 720P, 4Mbps 30fps, 60 seconds duration, no transcoding)	2x Hex-Core Xeon X5675 3.06GHz	2.5	150 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec H264 and AMR, 3gp file 720P, 4Mbps 30fps, high profile level 3, transcoding, cache disabled, 60 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	0.2	12 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
MSML Play Announcement, (Codec H264 and AMR, 3gp file 720P, 4Mbps 30fps, high profile level 3, transcoding to main profile level 2 CIF, cache enabled, 60 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	16.6	1000 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, VGA, 30fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	20	2400 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, CIF, 30fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	30	3600 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play Announcement, (Codec VP8 and G.711, avi file, QCIF, 20fps 60 seconds duration, non-transcoding)	2x Hex-Core Xeon X5675 3.06GHz	40	4800 calls	Tested on 6 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Play and Digit Connect (Codec G711 and SIP INFO Digit, 34 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	50	1700 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Record (MP3, 96Kbit, 32KHz, 120 seconds duration)	1x Six-Core Xeon X5670 2.93GHz	3	360 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Record (MP3, 320Kbit, 48KHz, 120 seconds duration)	1x Hex-Core Xeon X5670 2.93GHz	2	240 calls	Tested on 3 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1. One MCP per VM.
MSML Conference (32 participants per conference, all speakers. Each	2x Quad-Core Xeon E5620 2.40GHz	2.6	768 participants (24 conference sessions)	Tested on 4 VMs of EXSi 5.0, Guest OS Windows 2008 Server R2 x64 SP1.

Application Type	Hardware	Peak CAPS	Peak Ports	Comment
participant stays and speaks (300 secs in the conference. Codec G.711)				One MCP per VM.
Application Type	**Hardware**	**Peak CAPS**	**Peak Ports**	**Comment**
		Linux		
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	24	2880 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 6.4 x64. One MCP per VM. Play cache enabled as default. GVP 8.1.7 or later.
MSML Play Announcement, (Codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	42	5040 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 6.4 x64. One MCP per VM. Play cache disabled. GVP 8.1.7 or later.
MSML Play Announcement, (MP3, Any KHz, Any Kb, Cache enabled, Negotiated codec G711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	60	7200 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.
MSML Play Announcement, (MP3, 320Kbit, 44.1KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	16	1920 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.
MSML Play Announcement (MP3, 92Kbit, 32KHz, Cache disabled, Negotiated codec: G.711, 120 seconds duration)	2x Hex-Core Xeon X5675 3.06GHz	23	2760 calls	Tested on 6 VMs of EXSi 5.0, Guest OS RHEL 5.8 x64. One MCP per VM.

Resource Manager and MRCP Proxy Capacity Testing

Table: Resource Manager and MRCP Proxy Capacity Testing describes the capacity testing for overall system performance when the Resource Manager and MRCP Proxy (Windows only) are tested with multiple MCP instances.

Table: Resource Manager and MRCP Proxy Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Resource Manager (Windows)				
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	Using both TCP and UDP. Results occur regardless of the port density or the type of calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager performance with 1000 tenants configured.)	2x Core 2 Quad Xeon x5335, 2.66 GHz	600	Any number	Results occur regardless of the port density and the type of calls being routed. To achieve the peak CAPS, multiple Media Control Platforms might be required. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				Capacity Testing. • Disabled.
SIP Call (Resource Manager performance with MSML embedded in SIP INFO messages.)	2x Core 2 Quad Xeon x5335, 2.66 GHz	300	Any number	To achieve the peak CAPS, multiple Media Control Platforms might be required. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager performance)	4 Virtual Cores, Intel Xeon E5-2695, 2.40 GHz	800	Any number	Tested on TLS only. Results occur regardless of the port density or the type of

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager with Active - Active HA Pair performance)	4 Virtual Cores, Intel Xeon E5-2695, 2.40 GHz	500+500=1000	Any number	Tested on UDP only. Results occur regardless of the port density or the type of calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:**

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				• Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager with Active - Active HA Pair performance)	4 Virtual Cores, Intel Xeon E5-2695, 2.40 GHz	400+400=800	Any number	Tested on TCP only. Results occur regardless of the port density or the type of calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
Resource Manager (Linux)				
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	Using both TCP and UDP. Results occur regardless of the port density or the type of calls routed. Multiple Media Control Platform instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	600	Any number	In this scenario, 100K of DID numbers are configured and mapped to 262 IVR applications, and defined without wild cards or ranges. In other words, ordinary one-to-one mappings. Results occur regardless of the port density or the type of calls routed. Multiple Media Control Platforms required to achieve the peak CAPS. **Reporting Server configured in one of two ways:**

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				• Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	800	Any number	In this scenario, 1 million DID numbers are configured and mapped to 262 IVR applications, and defined in a multi-tenant environment (32 tenants with 30~35K of DIDs per tenant), without wildcards or ranges—In other words, simple one-to-one mappings. Results occurs regardless of the

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				port density or the type of calls routed. Multiple Media Control Platforms required to achieve the peak CAPS. **Reporting Server disabled** (due to the fact that the Reporting Server is unable to support 1 million DIDs).
SIP Call (Resource Manager performance with MSML embedded in SIP INFO messages.)	2x Core 2 Quad Xeon x5355, 2.66 GHz	350	Any number	Multiple Media Control Platforms required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				Capacity Testing. • Disabled
SIP Call (Resource Manager performance)	2x Core 2 Quad Xeon 5355, 2.66 GHz	200	Any number	Tested on UDP only on RHEL 6.4 x64. GVP 8.1.7 or later. Results occur regardless of the port density or the type of calls routed. Multiple MCPs are required to achieve the peak CAPS. **Reporting Server Configured in one of two ways:** • Enabled, but No-DB mode—Without the DB (all data dropped), Reporting Server can afford much higher capacity. • Disabled With Reporting Server and DB enabled the peak CAPS bottleneck will be due to RS (see below).
SIP Call (Resource Manager performance)	4 Virtual Cores, Intel Xeon E5-2695, 2.40 GHz	800	Any number	Tested on TLS only. Results occur regardless of the port density or the type of calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				mode—Without the DB (all data is dropped), the Reporting Server can handle much higher capacities. If both Reporting Server and the DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
SIP Call (Resource Manager with Active - Active HA Pair performance)	4 Virtual Cores, Intel Xeon E5-2695, 2.40 GHz	400+400=800	Any number	Using both TCP and UDP. Results occur regardless of the port density or the type of calls routed. Multiple MCP instances are required to achieve the peak CAPS. **Reporting Server configured in one of two ways:** • Enabled and in No-DB mode—Without the DB (all data is dropped), the Reporting Server can handle much higher

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				capacities. If both Reporting Server and DB are enabled, a peak CAPS bottleneck would occur. See SIP Call (Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server) in Table: Reporting Server Capacity Testing. • Disabled
MRCP Proxy (Windows)				
MRCPv1 requests(MRCP Proxy performance)	2x Core 2 Quad Xeon x5355, 2.66 GHz	1600	N/A	Tested with simulated MRCP servers and clients; calculation is based on MRCP sessions. Tested on Windows 2008 R2.

Reporting Server Capacity Testing

This section describes the capacity of overall system performance when the Reporting Server is tested with multiple Media Control Platform instances.

The tables in this section show the performance of other GVP components individually.

- Use these tables to determine if you encountered any performance limits beyond those already defined in other tables.
- Use these tables if you are interested in determining the overall system limits, which may occur in VoiceXML, media services, reporting, RM, or other functions.

Table: Reporting Server Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
SIP Call Reporting Server in partitioning mode with Microsoft SQL 2008 Enterprise Server	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Microsoft SQL Server DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz	270	Any number 32,400 (~30,000 based on a 120 seconds call duration)	Results occur regardless of the port density or the type of calls processed. Resource Manager and Media Control Platform log information to the Reporting Server using default settings. Increased reporting and logging can reduce Reporting Server capacity. Microsoft SQL database is installed on Windows 2008 Server with the database files residing on a 15k rpm HDD Disk Array.
SIP Call Reporting Server in partitioning mode with Oracle 10g R2 Server	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Oracle DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz	270	Any number	Results occur regardless of the port density or the type of calls processed. Resource Manager and Media Control Platform log information to the Reporting Server using default settings. Increased reporting and logging can reduce Reporting Server capacity. Oracle database is installed on Windows 2003 Server with the database files residing on a 15k rpm HDD Disk Array.
SIP Call	Reporting Server:	300	Any number	Regardless of the

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Reporting Server /w MS SQL Server 2008 R2 Enterprise (partitioning mode)	Quad-Core Xenon 5355 2.66GHz MS SQL DB: Quad-Core Xenon 5355 2.66GHz /w 15 HDD Disk Array			port density and the type of calls being processed with official architecture, which RM and MCP are both logging information to RS and using default setting. Heavier reporting/logging can reduce the RS capacity. MS SQL on Windows 2008 R2 with DB data files reside on a 15 HDD Disk Array (15k rpm)
SIP Call Reporting Server in No DB mode	Reporting Server: 2x Core 2 Quad Xeon E5504, 2.0 GHz, 8 GB RAM	800	Any number	When Reporting Server is configured in No DB mode, data that is sent to it, is dropped. Tested with an actual Resource Manager instance (not a VM) without the Media Control Platform.
SIP Call Reporting Server in partitioning mode with Oracle 11g Server	Reporting Server: 2x Core 2 Quad Xeon x5355, 2.66 GHz Oracle DB: 2x Core 2 Quad Xeon x5355, 2.66 GHz /w 15 HDD Disk Array	300	Any number	Results occur regardless of the port density and the type of calls being processed. RM and MCP both log information to the Reporting Server using default settings. Increased reporting/logging can reduce RS capacity. Oracle DB on Windows 2008 R2 x64 with DB data files that reside on a 15 HDD Disk Array (15k rpm).

CTI Connector and CTI Connector with ICM Capacity Testing

These two tables (one for Windows, one for Linux) describe the capacity testing for overall system performance when the CTI Connector is tested with multiple Media Control Platform instances. Results are provided for CTI applications and treatments using both GVPi and the NGI. In addition, CPUs of varying types and speeds were used for testing on Windows and are specified for each application.

Table: CTI Connector and CTI Connector with ICM Capacity Testing (Windows)

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
CTI Connector				
CTI treatments and bridge transfer application.	2x Quad Core Xeon E5335, 2.0 GHz	25(MCPs w/ GVPi)	800	A call that starts with a prompt-play and route request with 3 treatments and then bridge transfers to an agent. Tested with 5 Media Control Platform instances. GVP 8.1.4 only.
CTI treatments and one-step transfer application with GVPi.	2x Quad Core Xeon E5335, 2.0 GHz	25(MCPs w/ GVPi)	800	A call that starts with a prompt-play and route request with 3 treatments and then transfers to an agent. Tested with 5 Media Control Platform instances. GVP 8.1.4 or earlier.
CTI treatments and bridge transfer application with NGI.	2x Quad Core Xeon E5335 2.0GHz, 4 GB RAM, 270 GB SAS hdd	15(MCPs w/ NGI)	480	A call that starts with a prompt play and route request with 3 treatments, and then a bridge transfer to an agent. Tested with 5 Media Control Platform instances (GVP 8.1.3 or later).
CTI treatments and One-Step Transfer	2x Quad Core Xeon E5335 2.0GHz, 4	25(MCPs w/ NGI)	800	A call that starts with a prompt play

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
application with NGI.	GB RAM, 270 GB SAS hdd			and route request with 3 treatments, and then transfers (in one step via CTIC) out to an agent. Tested with 5 Media Control Platform instances (GVP 8.1.3 or later).
CTIC - ICM				
CTIC-ICM treatments, followed by a bridge transfer. (Call variable event is set to ICM.)	2x Quad Core Xeon E5335 2.0GHz, 2.53 GHz	22(MCPs w/ NGI)	440	Transfer with CED, Call and ECC variable events passing from two MCP instances to a single ICM. Tested on Windows 2008 R2 (GVP 8.1.4 and 8.1.5 releases), with CTIC installed in CRI mode.
CTIC-ICM (CRI mode) treatments, followed by a bridge transfer. Set Call Variable event to ICM (overall system performance, using multiple MCPs, NGi)	2x Quad Core Xeon E5630 2.53GHz	25 CAPS overall system (MCP/NGi)	500 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on Windows 2008 R2. (GVP 8.1.6+)
CTIC-ICM scripts treatments, followed by a cancellation and blind transfer. (Call Variable event is set to ICM.)	2x Quad Core Xeon E5335 2.0GHz	30(MCPs w/ NGI)	600	Different tenants associated with two VRU-PGs; A blind transfer with CED, Call and ECC variable passing from two MCP instances to two ICMs. Tested on Windows 2008 R2 (GVP 8.1.4 or later releases), with CTIC installed in SCI mode.

Table: CTI Connector and CTI Connector with ICM Capacity Testing (Linux)

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comment
CTI Connector				

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comment
CTI treatments and bridge Transfer application (overall system performance, with multiple MCPs, GVPi)	2x Quad Core Xeon E5630 2.53GHz	25 CAPS overall system (MCP/GVPi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments and then bridge transfer to an agent. Tested with 5 MCPs. Support x86 only, GVP 8.1.4+.
CTI treatments and One Step Transfer application (overall system performance, with multiple MCPs, GVPi)	2x Quad Core Xeon E5630 2.53GHz	25 CAPS overall system (MCP/GVPi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments and then transfer out to an agent. Tested with 5 MCPs. Support x86 only, GVP 8.1.4+.
CTI treatments and bridge Transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5410 2.53GHz	20 CAPS overall system (MCP/NGi)	640 (overall system)	Call starts with a prompt play and route request with 3 treatments and then bridge transfer to an agent. Tested with 5 MCPs on EL 5.x x64, GVP 8.1.5 or later.
CTI treatments and bridge transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	20 CAPS overall system (MCP/NGi)	640 (overall system)	Call starts with a prompt play and route request with 3 treatments and then bridge transfer to an agent. Tested with 5 MCPs on EL 6.4 x64, GVP 8.1.7 or later.
CTI treatments and bridge transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5630 2.53GHz	20 CAPS overall system (MCP/NGi)	640 (overall system)	Call starts with a prompt play and route request with 3 treatments and then bridge transfer to an agent. Tested with 5 MCPs. Support x86, GVP 8.1.4

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comment
				or later.
CTI treatments and one step transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5410 2.33GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments and then transfer out to an agent. Tested with 5 MCPs on EL 5.x x64, GVP 8.1.5 or later.
CTI treatments and One Step Transfer application using INFO + INFO model (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments (using INFO + INFO model) and then transfer out to an agent. Tested with 5 MCPs on RHEL 6.4 x64, GVP 8.1.7 or later.
CTI treatments and One Step Transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.33GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments and then transfer out to an agent. Tested with 5 MCPs on EL 6.4 x64, GVP 8.1.7 or later.
CTI treatments and One Step Transfer application (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5630 2.53GHz	25 CAPS overall system (MCP/NGi)	800 (overall system)	Call starts with a prompt play and route request with 3 treatments and then transfer out to an agent. Tested with 5 MCPs. Supports x86, GVP 8.1.4.
CTIC - ICM				
CTIC-ICM treatments, followed by a bridge transfer in CRI mode.	2x Quad Core Xeon E5335 2.0GHz	30 CAPS overall system (MCP/NGi)	600 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured.

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comment
Set call variable event to ICM (overall system performance, with multiple MCPs, NGi)				Tested on GVP 8.1.7 EL 6.4 x64.
CTIC-ICM treatments, followed by a bridge transfer in CRI mode. Set call variable event to ICM (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	25 CAPS overall system (MCP/NGi)	500 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on GVP 8.1.6 EL 5.x x64.
CTIC-ICM treatments, followed by a bridge transfer in CRI mode. Set call variable event to ICM (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	22 CAPS overall system (MCP/NGi)	440 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on GVP 8.1.5 x64.
CTIC-ICM treatments, followed by a bridge transfer in CRI mode. Set call variable event to ICM (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	18 CAPS overall system (MCP/NGi)	270 (overall system)	Bridge Transfer with CED, Call and ECC Variable passing from MCP to ICM. Only one ICM is configured. Tested on GVP 8.1.4 x86.
CTIC-ICM scripts treatments, followed by cancellation and blind transfer. Set Call Variable event to ICM (overall system performance, with multiple MCPs, NGi)	2x Quad Core Xeon E5335 2.0GHz	30 CAPS overall system (MCP/NGi)	600 (overall system)	Different Tenant tied to two VRU-PGs, Blind Transfer with CED, Call and ECC Variable passing from MCP to ICM. Two ICMs are configured. GVP 8.1.4 or later.
CTIC-ICM scripts treatments, followed by cancellation and blind transfer. Set Call Variable event to ICM (overall system	2x Quad Core Xeon E5335 2.0GHz	30 CAPS overall system (MCP/NGi)	600 (overall system)	Different Tenant tied to two VRU-PGs, Blind Transfer with CED, Call and ECC Variable passing from MCP to ICM. Two ICMs are configured.

Application Type (Linux)	Hardware	Maximum CAPS	Tested Ports	Comment
performance, with multiple MCPs, NGi)				Tested with 3 MCPs on RHEL 6.4 x64, GVP 8.1.7 or later.

PSTN Connector and SSG Capacity Testing

These two tables (for PSTN Connector and SSG Capacity) describe the capacity testing for overall system performance when the PSTN Connector or Supplementary Services Gateway components are tested with multiple Media Control Platform instances. In addition, CPUs of varying types and speeds were used for testing on Windows, and are specified for each application.

Table: PSTN Connector Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
PSTN Connector (Windows)				
PSTN Connector *VoiceXML_App1*	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans	Tested with two DMV boards.
PSTN Connector *VoiceXML_App3*	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans	Tested with two DMV boards.
PSTN Connector (Linux)				
PSTN Connector *VoiceXML_App1*	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans (ISDN only)	Tested with two DMV boards. (RHEL 5.8 x86 only)
PSTN Connector *VoiceXML_App3*	2x Xeon 3.0 GHz	N/A	8 T1/E1 spans (ISDN only)	Tested with two DMV boards. (RHEL 5.8 x86 only)

Table: SSG Capacity Testing

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
Supplementary Services Gateway (Windows)				
Supplementary Services Gateway (SSG) outbound call application	2x Core2Quad Xeon E5335 2.53GHz	65	N/A	The SSG makes outbound calls through SIP Server, which becomes the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.5 with SIP Server 8.1.0 or later.
SSG outbound call application	2x Core2Quad Xeon E5335	50	N/A	The SSG makes outbound calls

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
	2.53GHz			through SIP Server which becomes the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.3 or 8.1.4 with SIP Server 8.0.4 or later.
SSG outbound call application	2x Core 2 Quad Xeon x5355, 2.66 GHz	40	N/A	The SSG makes outbound calls through SIP Server which becomes the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. Pre-GVP 8.1.3 with SIP Server 8.0.3.
Supplementary Services Gateway (Linux)				
Supplementary Services Gateway (SSG) outbound call application	2x Quad Core Xeon E5335 @ 2.00GHZ, 4 GB RAM, 67 GB SAS hdd	66	N/A	The SSG makes outbound calls through SIP Server, which becomes the overall system bottleneck. Multiple MCP instances are required to achieve peak capacity. GVP 8.1.5 on RHEL 5.x with SIP Server 8.1.000.54.
SSG outbound call application (overall system) performance, with multiple MCPs)	2x Core2Quad Xeon E5335 2.00GHz	64 CAPS (overall system)	N/A	SSG makes outbound calls via SIP Server which becomes the bottleneck overall system. Multiple MCPs are required to achieve peak capacity. GVP 8.1.7 on RHEL

Application Type	Hardware	Peak CAPS	Peak Ports	Comments
				6.4 x64 with SIP Server 8.1.1.
SSG outbound call application	2x Quad Core Xeon E5335 @ 2.53GHZ, 4 GB RAM, 67 GB SAS hdd	50	N/A	The SSG makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. GVP 8.1.3 and 8.1.4 releases with SIP Server 8.0.4.
SSG outbound call application	2x Core 2 Quad Xeon x5355, 2.66 GHz	40	N/A	The SSG makes outbound calls through SIP Server which become the overall system bottleneck. Multiple Media Control Platform instances are required to achieve peak capacity. Pre-GVP 8.1.3 with SIP Server 8.0.3.

GIR-GVP Port Capacity Tests

Genesys conducted performance tests of various Genesys Interactive Recording (GIR) and GVP Capacities. This section contains test profiles, test results, and analysis.

- GVP-GIR Port Capacity Test Profiles
- GIR-GVP Port Capacity Test Results Summary
- Detailed Studies of GVP Media Server Behavior on Windows
  - Performance Comparison of Physical Server and Virtual Machines
  - Performance Comparison of Different Virtual Machines Configurations
  - Performance Comparison of MP3 only and MP3 + WAV
  - Performance Comparison between SAS HDDs and SSD
  - Data Throughput
  - MCP IOPS
  - MP3 16KBPS Bit Rate Compression
  - MP3 16KBPS Bit Rate Compression with Encryption
- Detailed Studies of GVP Media Server Behavior on Linux
- Multiple Dispatcher Tests

GIR-GVP Port Capacity Test Profiles

- Software (SW) Profiles Used in These Tests
- Hardware (HW) Profiles Used in These Tests
- Virtual Machine (VM) Profiles Used in These Tests

Software (Profiles Used in GIR-GVP Port Capacity Tests)

Note: Unless explicitly noted, all MP3 recordings use stereo channels.

Software Profile 1 **call recording only, MP3 codec (32kbps bit rate) without encryption**	**Software Profile 1a** **call recording only, MP3 codec (16kbps bit rate) without encryption**	***Software Profile 1b'*** **call recording only, MP3 codec (8kbps bit rate Mono) without encryption**
• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = NOT SET • type2 = NOT SET • encryption = disabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC	• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = NOT SET • type2 = NOT SET • encryption = disabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC	• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = NOT SET • type2 = NOT SET • encryption = disabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC • channels = 1 (specifies mono recording)

Software Profile 2 **call recording only, MP3 codec (32kbps) and WAV as dest2 without encryption**	**Software Profile 2a** **call recording only, MP3 codec (16kbps) and WAV as dest2 without encryption**
• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = `http://webdav` • type2 = audio/wav	• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = `http://webdav` • type2 = audio/wav

Software Profile 2 **call recording only, MP3 codec (32kbps) and WAV as dest2 without encryption**	**Software Profile 2a** **call recording only, MP3 codec (16kbps) and WAV as dest2 without encryption**
• encryption = disabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC	• encryption = disabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC

Software Profile 3 **call recording only, MP3 codec (32kbps bit rate) with encryption**	**Software Profile 3a** **call recording only, MP3 codec (16kbps bit rate) with encryption**
• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = NOT SET • type2 = NOT SET • encryption = enabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC	• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = NOT SET • type2 = NOT SET • encryption = enabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC

Software Profile 4 **call recording only, MP3 codec (32kbps) and WAV as dest2 with encryption**	**Software Profile 4a** **call recording only, MP3 codec (16kbps) and WAV as dest2 with encryption**
• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = `http://webdav` • type2 = audio/wav • encryption = enabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC	• dest = S3 or `http://webdav` • type = audio/mp3 • dest2 = `http://webdav` • type2 = audio/wav • encryption = enabled • write interval = 10s • call duration = 210s • callrec_dest = HTCC

Hardware Profiles Used in GIR-GVP Port Capacity Tests

Hardware Profile 1	Specifications & Recommendations	Comment
CPU	Single Hex Core Intel Xeon X5670@ 2.93GHz	
Memory	8 GB or more	4 GB is minimum and 8 GB is recommended
Network	GigaBit Ethernet	100MBit supported
Storage	15k rpm SAS HDD disk storage with at least 72 GB. RAID 0.	15k rpm recommended for maximum performance
OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

Hardware Profile 2	Specification & Recommendation	Comment
CPU	Single Hex Core Intel Xeon X5675@ 3.06GHz	
Memory	16 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP recording cache location. 15k rpm SAS HDD disk storage with at least 136 GB used for all other operations. RAID 0	SSD and 15k rpm SAS HDD are recommended for maximum performance
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 3	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	16 GB or more	8 GB is minimum and recommended
Network	GigaBit Ethernet	100MBit supported
Storage	15k rpm SAS HDD disk storage with at least 72 GB. RAID 0	15k rpm SAS HDD is recommended for maximum performance
OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

Hardware Profile 4	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	32 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP recording cache location.	SSD and 15k rpm SAS HDD are recommended for maximum

Hardware Profile 4	Specification & Recommendation	Comment
	15k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	performance
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 5	Specification & Recommendation	Comment
CPU	Dual Hex Core Xeon X5675 3.06 GHz	
Memory	32 GB or more	4 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	Multiple 15k rpm SAS HDDs disk storage with at least 360 GB used for all other operations. RAID 0.	Split VMs into multiple 15k rpm SAS HDDs.
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 6	Specification & Recommendation	Comment
CPU	Single Eight Core Xeon E5-2640 2.00 GHz	
Memory	64 GB or more	8 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	SSD used for MCP logs and recording cache location. 15k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	SSD and 15k rpm SAS HDD are recommended for maximum performance.
OS	VM vSphere or ESXi 5.x Windows Server 2008 R2 x64 Enterprise Edition SP1	VM vSphere 5.x as Host OS Windows 2008 Server as Guest OS on VM

Hardware Profile 7	Specification & Recommendation	Comment
CPU	Dual 16 core Xeon E5-2683 v4 @ 2.10GHz	
Memory	32 GB or more	8 GB is minimum for each VM
Network	GigaBit Ethernet	100MBit supported
Storage	10k rpm SAS HDD disk storage with at least 360 GB used for all other operations. RAID 0.	SSD and 15k rpm SAS HDD are recommended for maximum performance.
OS	VM vSphere or ESXi 6.x Windows Server 2016/RHEL 7 as Guest OS	VM vSphere 6.x as Host OS Windows Server 2016/RHEL 7 as Guest OS

Virtual Machine (VM) Profiles Used in GIR-GVP Port Capacity Tests

VM Profile 1	Specifications & Recommendations	Comment
Host Hardware	**Hardware Profile 2**	1x X5675@3.06GHz 16 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 2	Specifications & Recommendations	Comment
Host Hardware	**Hardware Profile 4**	2x X5675@3.06GHz , 32 GB RAM
CPU	4 x vCPU	
Memory	8 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance.
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 3	Specifications & Recommendations	Comment
Host Hardware	**Hardware Profile 4**	2x X5675@3.06GHz , 32 GB RAM
CPU	3 x vCPU	
Memory	6 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	10 GB SSD used for MCP recording cache location. At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 4	Specifications & Recommendations	Comment
Host Hardware	**Hardware Profile 4**	2x X5675@3.06GHz , 32 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported

VM Profile 4	Specifications & Recommendations	Comment
Storage	10 GB SSD used for MCP recording cache location.At least 36 GB 15k rpm SAS HDD disk storage used for all other operations.	SSD is recommended for maximum performance
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 5	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 5	2x X5675@3.06GHz , 32 GB RAM
CPU	2 x vCPU	
Memory	5 GB	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	At least 36 GB 15k rpm SAS HDD disk storage.	
Guest OS	Windows Server 2008 R2 x64 Enterprise Edition SP1	

VM Profile 6	Specifications & Recommendations	Comment
Host Hardware	Hardware Profile 7	2x Intel® Xenon® CPU E5-2683 v4@2.10GHz
CPU	2 x vCPU	
Memory	4 GB RAM	4 GB is minimum.
Network	GigaBit Ethernet	100MBit supported
Storage	At least 36 GB 10k rpm SAS HDD disk storage.	
Guest OS	Microsoft Windows Server 2016 or Red Hat Enterprise Linux 7.0	

GIR-GVP Port Capacity Test Results Summary

- Criteria
- Summary of Performance Testing Results
- Parameter Adjustments

Criteria

System Port Capacity is the maximum number of ports (Port Density or PD) or rate (Call Arrivals Per Second or CAPS) that a GIR-GVP system can handle; this number must maximize the usage of hardware resources, while maintaining all criteria within the predefined threshold.

Because CPU usage is the usual deciding factor for peak port capacity, this section presents results that correlate to CPU usage (and other criteria such jitter buffer and max delta from sample recordings) to track the quality of recording.

The following criteria are required for an installation to reach the 95th percentile of quality analysis, from a sample RTP stream:

- Packet Loss <= 1%
- Max Jitter Buffer <= 30ms
- Max Delta <= 200ms

Summary of Performance Testing Results

Table 3: GIR-GVP Port Capacity on Physical Servers

Test Profiles	HW profile	OS	Peak Ports	Comment
SW Profile 1 (32 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	200 (preferred)	
SW Profile 1 (32 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	220 (peak)	If some of audio criteria can be i waived.
SW Profile 1a (16 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	240 (preferred)	
SW Profile 1a (16 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	270 (peak)	If some of audio criteria can be i waived.
SW Profile 1a (16 Kbps bit rate)	HW Profile 6	Windows 2008 R2 x64	350 (preferred)	8 Dispatchers (cores)

SW Profile 1a (16 Kbps bit rate)	HW Profile 6	Windows 2008 R2 x64	450 (peak)	If some of audi criteria can be i waived. 8 Dispa # of cores)
SW Profile 1b (8 Kbps bit rate Mono)	HW Profile 6	Windows 2008 R2 x64	450 (preferred)	8 Dispatchers (cores)
SW Profile 1b Profile 1b (8 Kbps bit rate Mono)	HW Profile 6	Windows 2008 R2 x64	600 (peak)	If some of audi criteria can be i waived. 8 Dispa # of cores)
SW Profile 3a (16 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	210 (preferred)	
SW Profile 3 (16 Kbps bit rate)	HW Profile 1	Windows 2008 R2 x64	270 (peak)	If some of audi criteria can be i waived.
SW Profile 1(32 Kbps bit rate)	HW Profile 3	Windows 2008 R2 x64	240 (preferred)	
SW Profile 1 (32 Kbps bit rate)	HW Profile 3	Windows 2008 R2 x64	360 (peak)	If some of audi criteria can be i waived.
SW Profile 1a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	150 (preferred)	
SW Profile 1a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	210 (peak)	If some of audi criteria can be i waived.
SW Profile 1a (16 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	220 (preferred)	
SW Profile 1a (16 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	240 (peak)	If some of audi criteria can be i waived.
SW Profile 1a (16 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	300 (preferred)	8 Dispatchers (cores)
SW Profile 1a (16 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	360 (peak)	If some of audi criteria can be i waived. 8 Dispa # of cores)
SW Profile 1b (8 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	600 (preferred)	8 Dispatchers (cores)
SW Profile 1a (8 Kbps bit rate)	HW Profile 6	RedHat EL 6.6 x64	650 (peak)	If some of audi criteria can be i waived. 8 Dispa # of cores)
SW Profile 2a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	90 (preferred)	
SW Profile 2a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	150 (peak)	If some of audi criteria can be i waived.

SW Profile 3a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	150 (preferred)	
SW Profile 3a (16 Kbps bit rate)	HW Profile 1	RedHat EL 6.5 x64	210 (peak)	If some of audio criteria can be i waived.

Table 4: GIR-GVP Port Capacity on Virtual Machines (VMs)

SW Profile	HW profile	OS	Port Capacity	Comment
SW Profile 1 (32 Kbps bit rate)	VM Profile 1	VM vSphere 5.1 Windows 2008 R2 x64	300 (preferred)	3 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 1	VM vSphere 5.1 Windows 2008 R2 x64	360 (peak)	3 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 2	VM vSphere 5.1 Windows 2008 R2 x64	360 (preferred)	3 VMs: each VM uses 4 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 2	VM vSphere 5.1 Windows 2008 R2 x64	390 (peak)	3 VMs: each VM uses 4 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 3	VM vSphere 5.1 Windows 2008 R2 x64	520 (preferred)	4 VMs: each VM uses 3 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 3	VM vSphere 5.1 Windows 2008 R2 x64	600 (peak)	4 VMs: each VM uses 3 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 4	VM vSphere 5.1 Windows 2008 R2 x64	600 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 4	VM vSphere 5.1 Windows 2008 R2 x64	660 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1 (32 Kbps bit rate)	VM Profile 6	Windows 2016 x64	130 (preferred)	1 VM: each uses 2vCPU and 1 MCP.
SW Profile 1 (32 Kbps bit rate)	VM Profile 6	Windows 2016 x64	165 (peak)	1 VM: each uses 2vCPU and 1 MCP.
SW Profile 1 (32 Kbps bit rate)	VM Profile 6	RH EL 7.0	100 (preferred)	1 VM: each uses 2vCPU and 1 MCP.
SW Profile 1 (32 Kbps bit rate)	VM Profile 6	RH EL 7.0	160 (peak)	1 VM: each uses 2vCPU and 1 MCP.
SW Profile 1a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5 Windows 2008 R2 x64	720 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1a (32 Kbps bit rate)	VM Profile 4	VM vSphere 5.5 Windows 2008 R2 x64	840 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.

SW Profile 2 (32 Kbps bit rate MP3 + WAV)	VM Profile 4	VM vSphere 5.1 Windows 2008 R2 x64	360 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 2 (32 Kbps bit rate MP3 + WAV)	VM Profile 4	VM vSphere 5.1 Windows 2008 R2 x64	540 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 3a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ Windows 2008 R2 x64	480 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 3a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ Windows 2008 R2 x64	840 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	540 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 1a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	660 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 2a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	480 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 2a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	600 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 3a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	540 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 3a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	660 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 4a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	480 (preferred)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.
SW Profile 4a (16 Kbps bit rate)	VM Profile 4	VM vSphere 5.5/ RH EL 6.5 x64	600 (peak)	6 VMs: each VM uses 2 vCPU & 1 MCP installed per VM.

Parameter Adjustments

These adjustments achieve higher port capacity:

Parameter Value Adjustments Yielding Higher Port Capacity

Parameter	Default Value	Adjusted Value
`mpc.recordnumparallelpost`	30	300
`mpc.recordpostretrybackoff`	120000	15000
`mpc.recordpostretrycount`	3	1
`mpc.mediamgr.recordwritetimeinterval`	1000	10000
`fm.http_proxy`	<not empty>	<empty> (squid bypassed)

Detailed Studies of GVP Media Server Behavior on Windows

- Performance Comparison of Physical Server and Virtual Machines
- Performance Comparison of Different Virtual Machines Configurations
- Performance Comparison of MP3 only and MP3 + WAV
- Performance Comparison between SAS HDDs and SSD
- Data Throughput
- MCP IOPS
- MP3 16 kbps Bit Rate
- MP3 16 kbps Bit Rate with Encryption

Performance Comparison of Physical Server and Virtual Machines

Single Hex Core

With a single hex core CPU, Genesys recommends 200 ports as a reasonable peak port capacity on a physical server with a single X5670, assuming that all criteria have been met. 300 ports can be achieved with a three-VMs configuration of the same hardware, with a single X5675 (performance is slightly better than X5670). The graph below compares overall CPU usage:

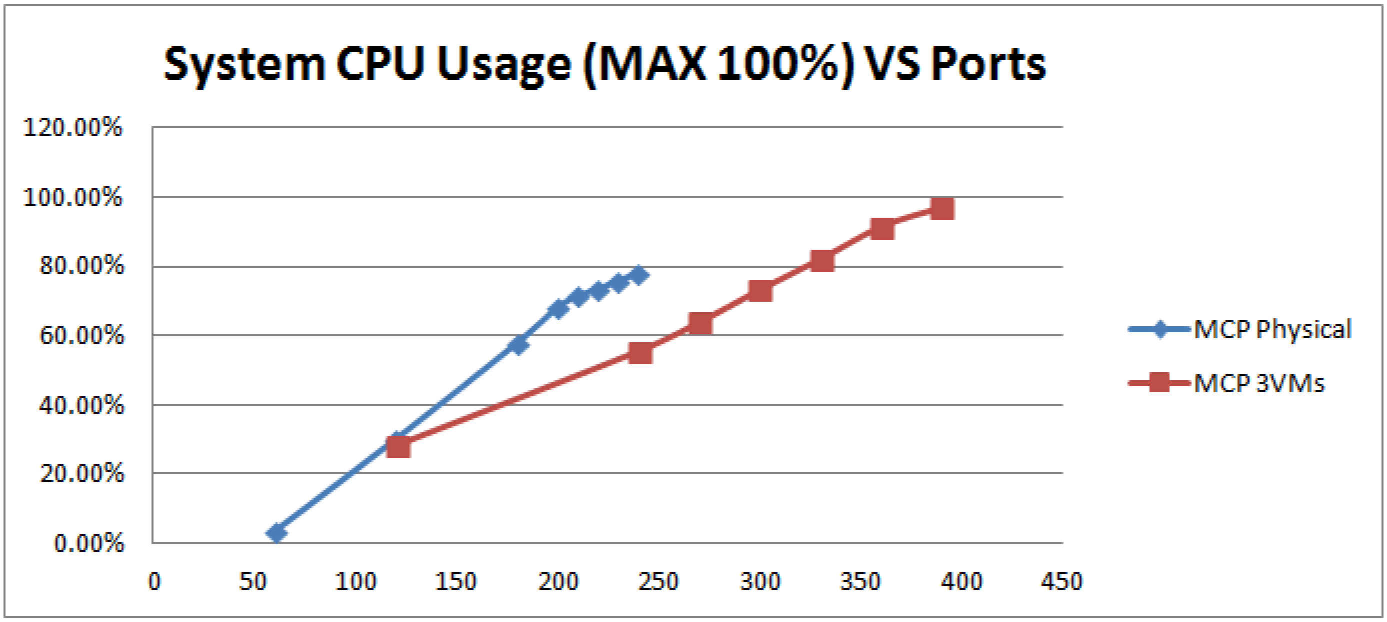


Figure 1: Comparison of System Usage between Physical Server and VM from Single Hex Core

Memory usage for MCP scales linearly against port capacity:

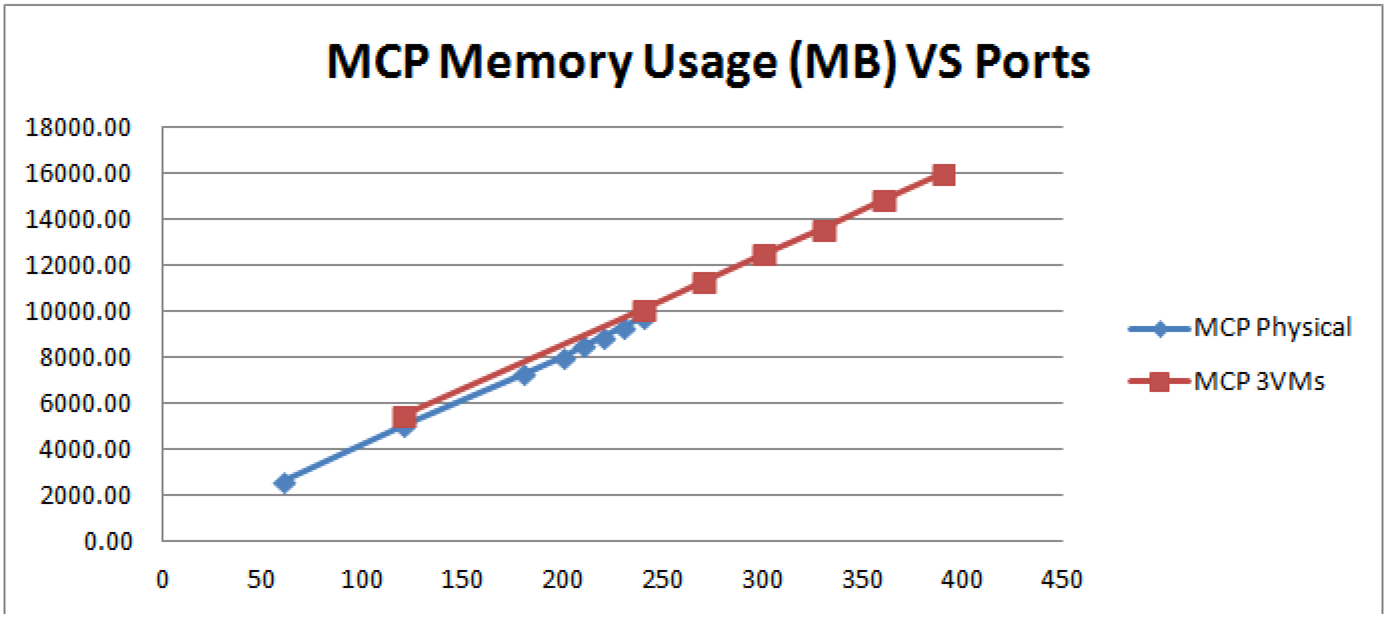


Figure 2: Comparison of MCP Memory Usage between Physical Server and VM from Single Hex Core

The two graphs below compare the 95th percentile value of Max Jitter Buffer and Max Delta, tracking audio quality from a sample RTP stream:

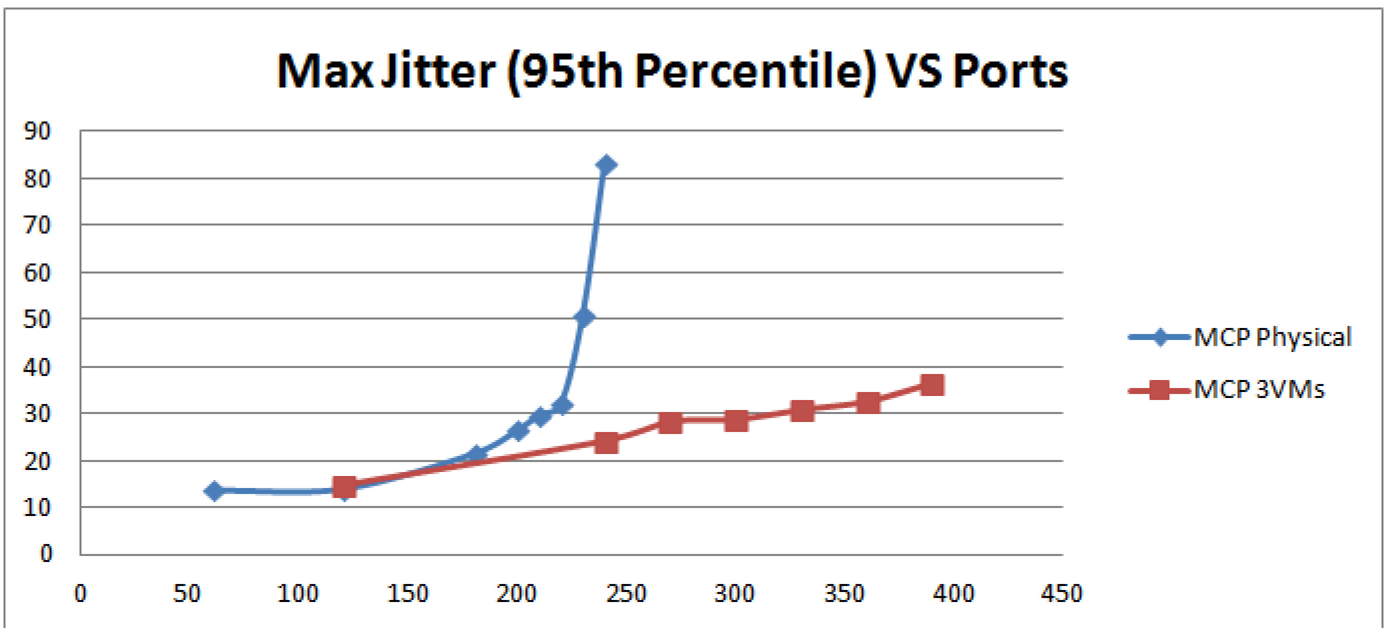


Figure 3: Comparison of Max Jitter between Physical Server and VM from Single Hex Core

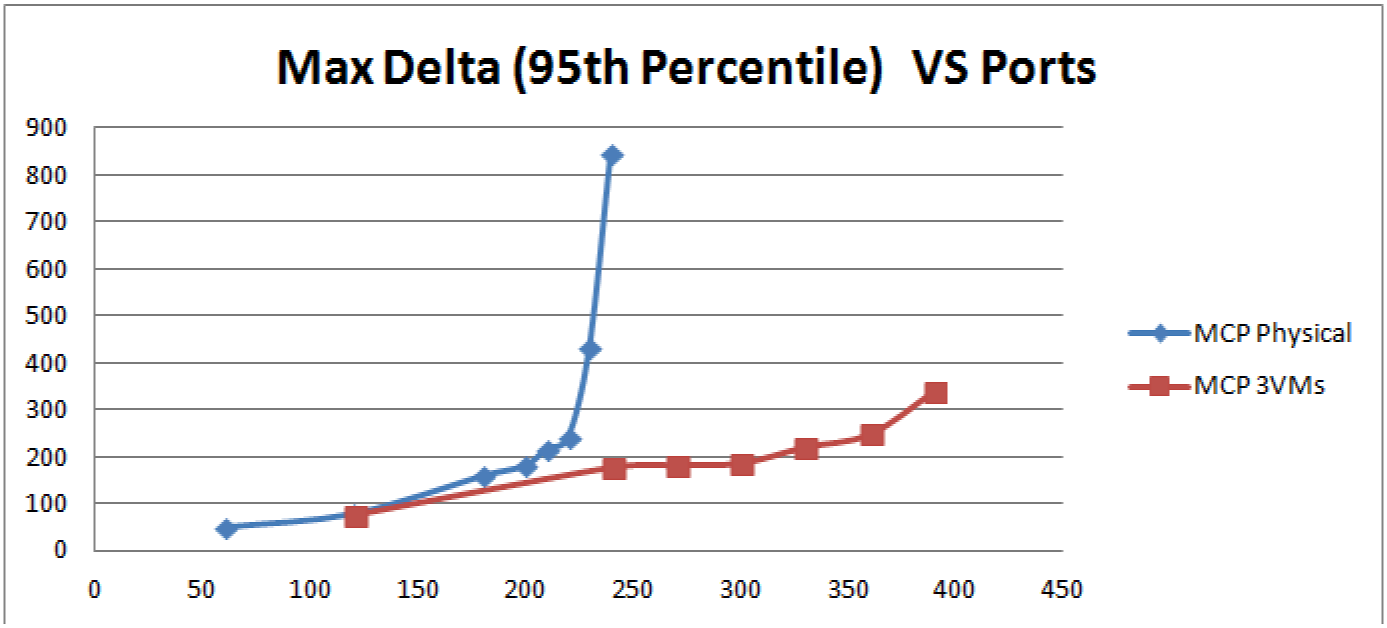


Figure 4: Comparison of Max Delta between Physical Server and VM from Single Hex Core

A strong correlation exists between Max Jitter Buffer and Max Delta, regarding audio quality. A physical server can meet all criteria when its port capacity is 200 or below. Port capacity that is between 200 and 220 may impact audio quality, since both Max Jitter buffer and Max Delta are just slightly beyond the passing criteria. You can consider 220 as peak performance, if audio quality is not strictly required and can be waived. However, when port capacity reaches 230 or beyond, the two values become so big that there is apparent audio quality impact.

For VM configuration: Preferred/Recommended = 300 ports; Peak Port Capacity = 360 ports. With 390 ports, overall system CPU usage is 97%, close enough to 100% that it also observed audio quality impact.

Below are two tables of IOPS for the above two configurations:

Table 1: Disk IOPS of system level from a physical server with a single hex core

Ports	Disk IOPS Physical Server		
Total	Reads	Writes	
60	11.13	0.001	11.13
120	21.82	0.001	21.82
180	32.03	0.001	32.03
200	34.95	0.001	34.95
210	36.53	0.001	36.53
220	37.76	0.001	37.76
230	39.48	0.001	39.48
240	43.33	0.002	43.33

Table 2: Disk IOPS of sum of all VMs of single hex core

Ports	Disk IOPS VMs Overall		
Total	**Reads**	**Writes**	
120	20.68	0.101	20.58
240	36.29	0.070	36.22
270	41.39	0.087	41.30
300	45.57	0.065	45.50
330	48.85	0.000	48.85
360	51.69	0.000	51.69
390	57.82	0.002	57.82

Disk IOPS in Disk IOPS of sum of all VMs of single hex core table is the sum of Disk IOPS from all VMs. Also, IOPS is measured from each VM and then totaled, to determine overall IOPS. The same method is applied to all Disk IO calculations for VM environments in this series of tests.

Also in the above two tables, the IOPS Reads value is quite small because most of the operations are Writes.

The graph below compares the two IOPS tables above:

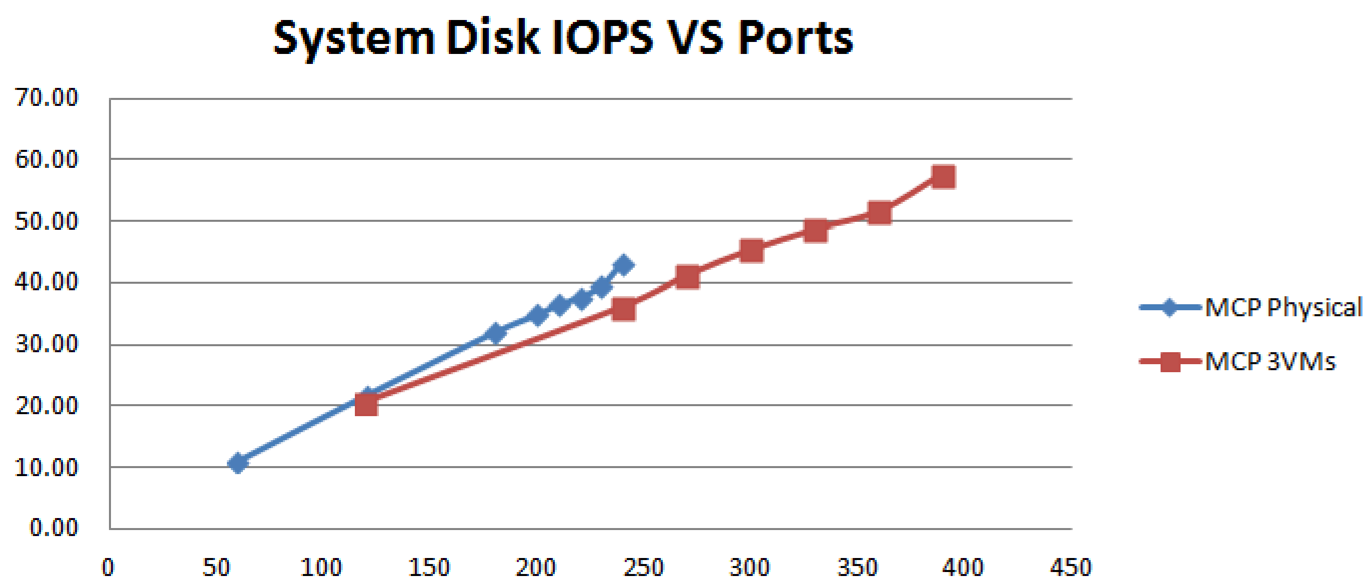


Figure 5: Comparison of System Disk IOPS Physical Server and VM from Single Hex Core

- System level disk IOPS is scaling linearly against port capacity for both physical server and virtual machines.
- SSD is only used on VM env as cache folder of MCP recording while SAS HDD drive is used to installed OS and MCP.

Dual Hex Cores

With a host of dual hex core CPUs (2x X5675@3.06GHz) with 32 GB RAM, we also compare the results from physical server and VM env. In VM env, on same hardware spec, 3 VMs are configured with 4 vCPU and 8 GB RAM assigned to each VM. Only one MCP installed on each VM and a SSD partition is used as cache folder for MCP recording.

The graph below depicts the overall system CPU usage:

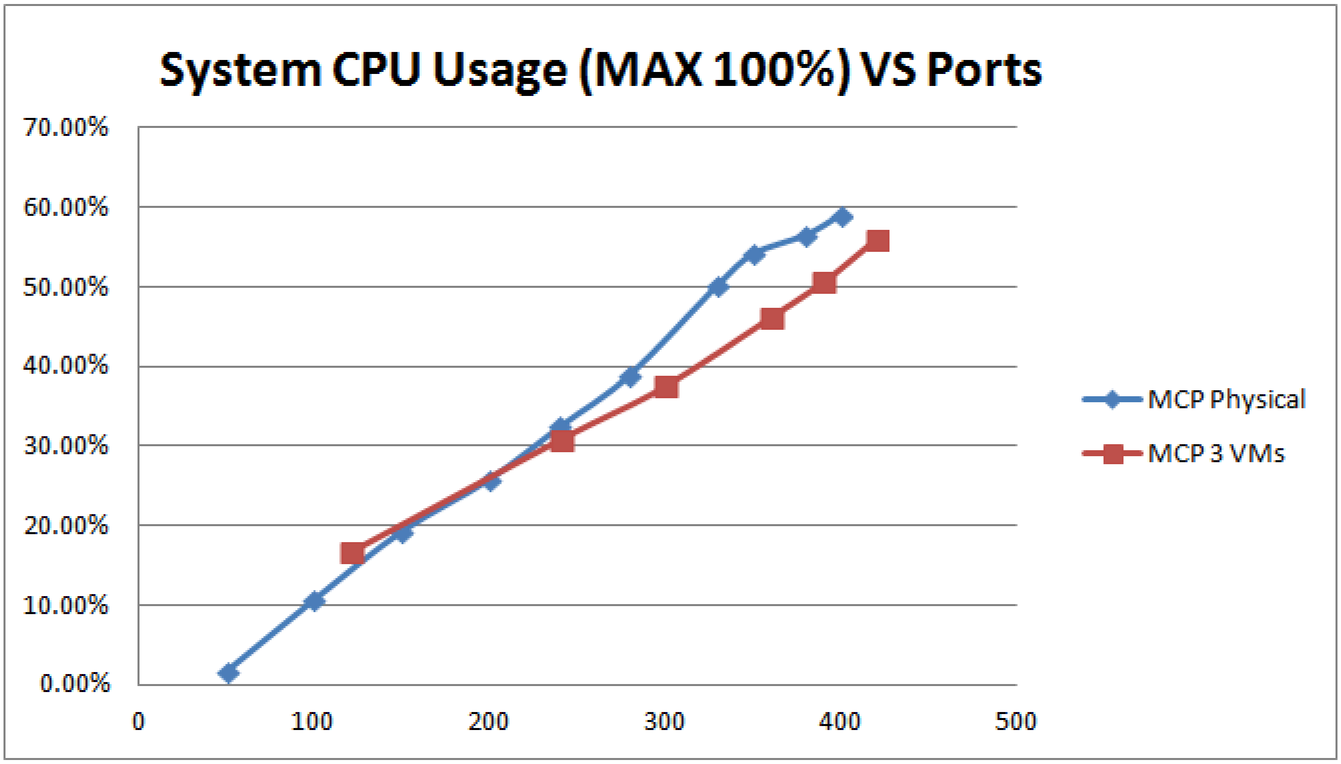


Figure 6: Comparison of System Usage between Physical Server and VM from Dual Hex Cores

The next two graphs show 95 percentile values of Max Jitter and Max Delta from sample RTP stream quality analysis:

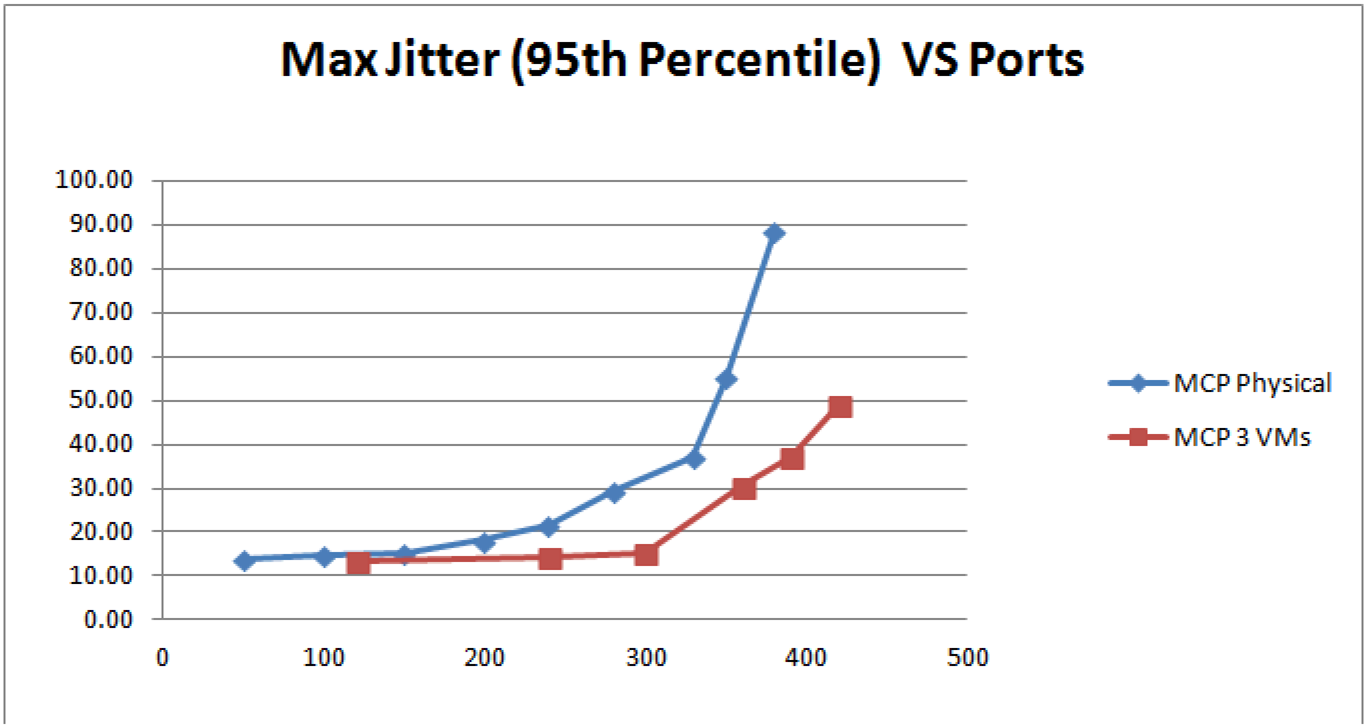


Figure 7: Comparison of Max Jitter between Physical Server and VM from Dual Hex Cores

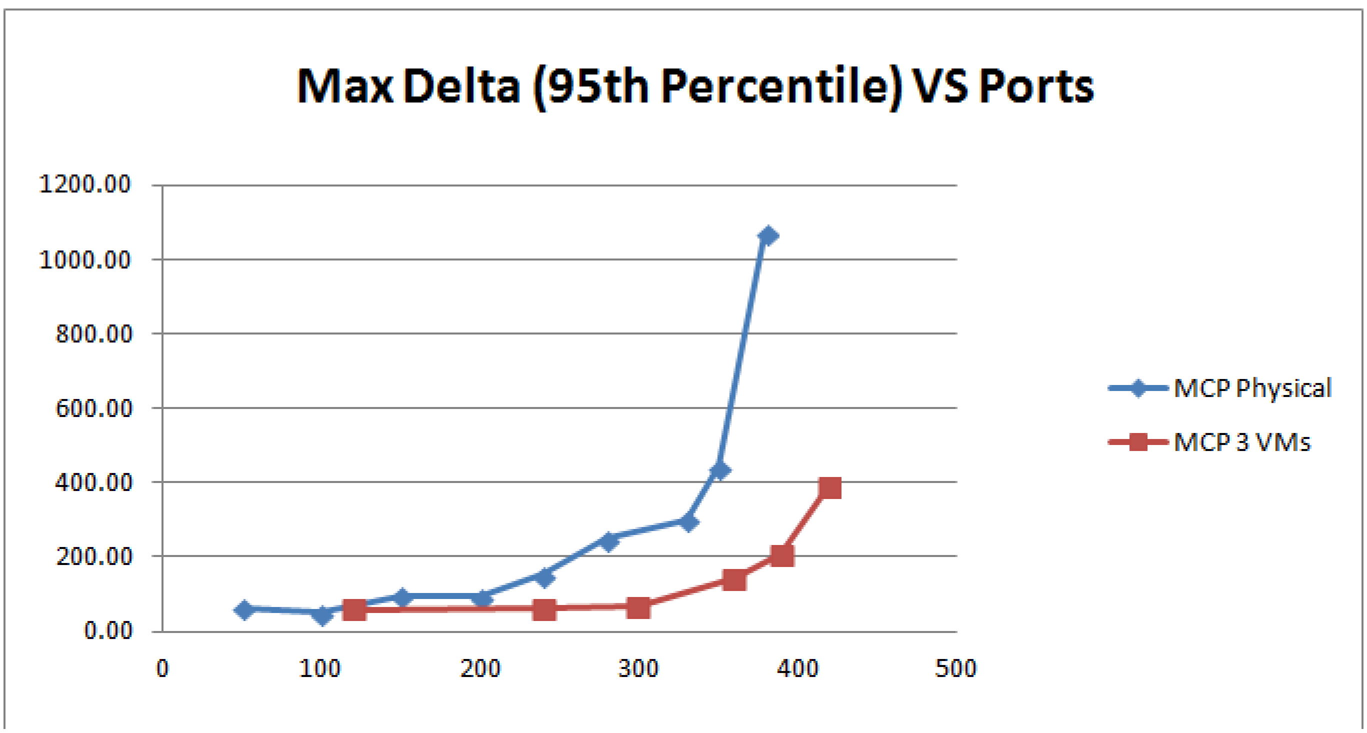


Figure 8: Comparison of Max Delta between Physical Server and VM from Dual Hex Cores

The two tables below show:

- Disk IOPS at system level on a physical server.
  and
- Disk IOPS at system level on a VM environment.

Table 3: Disk IOPS at system level from physical server of dual hex cores

Ports	Disk IOPS Physical Server		
Total	Reads	Writes	
50	9.069	0.000	9.07
100	18.587	0.000	18.59
150	28.598	0.001	28.60
200	37.460	0.001	37.46
240	41.290	0.003	41.29
280	49.031	0.020	49.01
330	53.373	0.001	53.37
350	53.150	0.001	53.15
380	61.456	0.001	61.46
400	64.123	0.001	64.12

Table 4: Disk IOPS of sum of all 3 VMs of dual hex cores

Ports	Overall Disk IOPS		
Total	Reads	Writes	
120	22.38	0.024	22.35
240	38.99	0.012	38.97
300	48.60	0.017	48.59
360	56.05	0.047	56.00
390	60.24	0.002	60.24
420	64.59	0.028	64.57

The graph below compares the above two tables above:

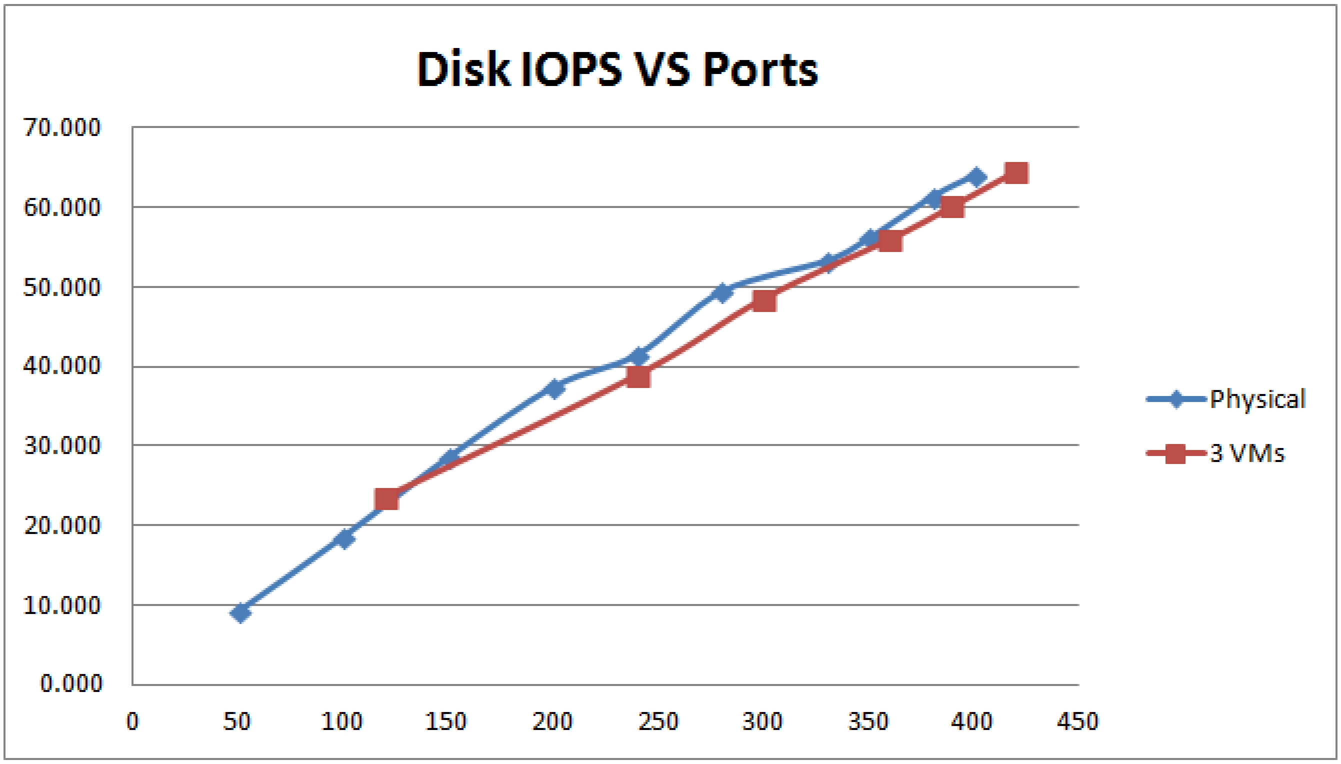


Figure 9: Comparison of System Disk IOPS Physical Server and VM from Dual Hex Cores

- Comparing the figure Comparison of System Disk IOPS Physical Server and VM from Single Hex Core and Comparison of System Disk IOPS Physical Server and VM from Dual Hex Cores: IOPS is linearly related to ports. No big differences exist between the physical server and the VM environment.
- SSD is used only in VM environments, as the cache folder of MCP recordings, while an SAS HDD drive is used to install the OS and MCP.

Performance Comparison of Different Virtual Machines Configurations

Overall CPU usage on a physical server beyond peak port capacity is actually higher than overall CPU usage on virtual machines, while audio quality actually shows a quick downfall on a physical server. So the splitting the load into multiple MCPs in a VM environment will definitely take advantage of hardware resources and will achieve high port capacity with fewer audio quality concerns. There are three different VM configurations on the same hardware spec (counting the dual hex cores, 12 vCPUs in total) that are used for this purpose:

- 3 VMs in total, 4 vCPU are assigned to each VM, only one MCP installed on one VM.
- 4 VMs in total, 3 vCPU are assigned to each VM, only one MCP installed on one VM.
- 6 VMs in total, 2 vCPU are assigned to each VM, only one MCP installed on one VM.

The graph below compares overall system CPU usage.

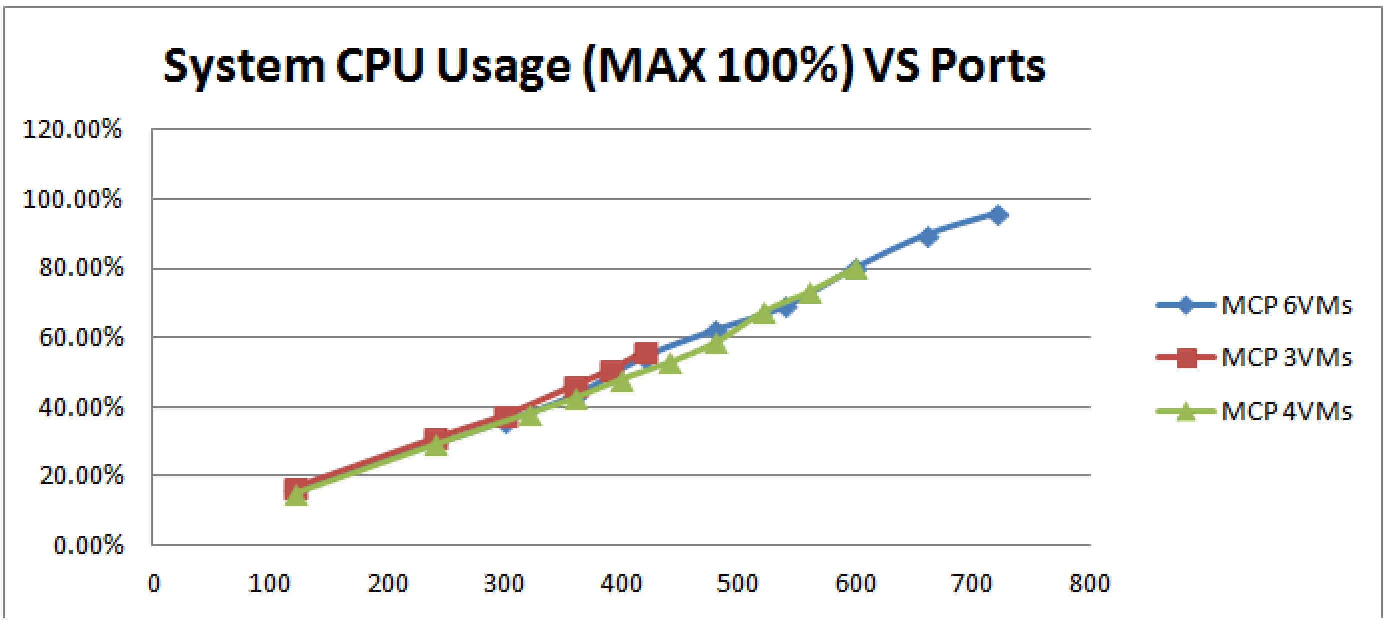


Figure 14: Comparison of System CPU Usage among different VMs configurations

Overall CPU usage scales linearly against port capacity, regardless of how many VMs are configured.

The two graphs below compare RTP stream quality related Max Jitter and Max Delta on these three different VM configurations:

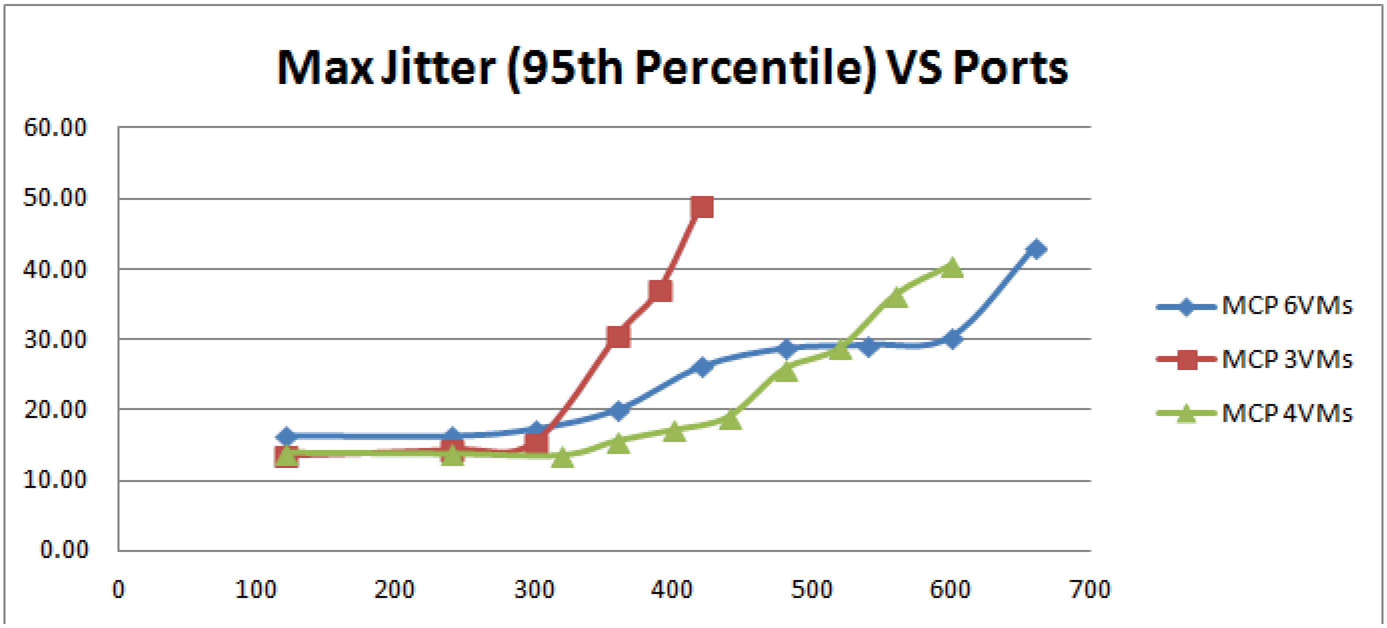


Figure 15: Comparison of Max Jitter among different VM configurations

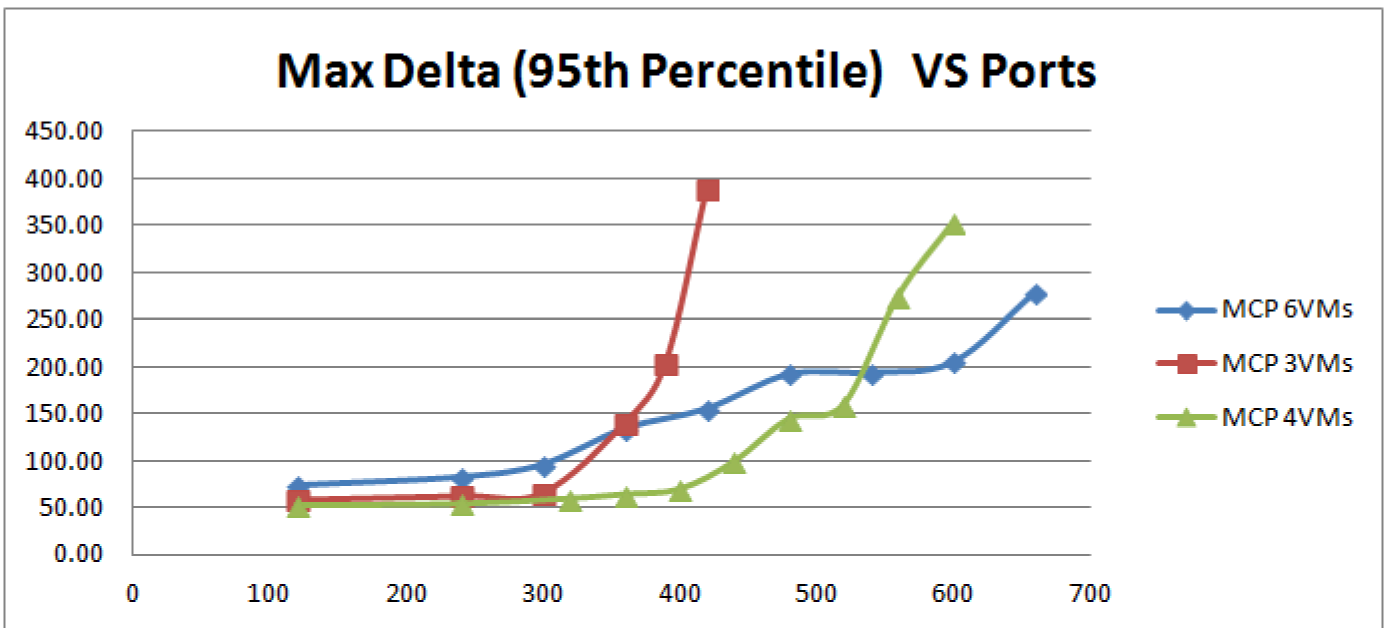


Figure 16: Comparison of Max Delta among different VM configurations

To achieve higher port capacity, configure more VMs and assign less vCPU to each VM. With audio quality criteria considered, Genesys recommends 600 ports as peak for six VM configurations. Six VMs with two vCPUs for each VM is the optimal configuration.

Below is the table of IOPS for 6 VM configurations:

Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only

Ports	Overall Disk IOPS (kbps)		
Total	Reads	Writes	
120	25.18	0.028	25.15
240	42.75	0.052	42.70
300	51.16	0.004	51.15
360	59.61	0.000	59.61
420	67.04	0.000	67.04
480	74.82	0.000	74.82
540	86.30	0.000	86.30
600	94.11	0.000	94.11
660	102.05	0.000	102.04
720	111.30	0.000	111.29

The graph below compares the two tables of IOPS (Table: Disk IOPS of sum of all 3 VMs of dual hex cores for 3 VMs and Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only for 6 VMs), on the same hardware spec of dual hex cores:

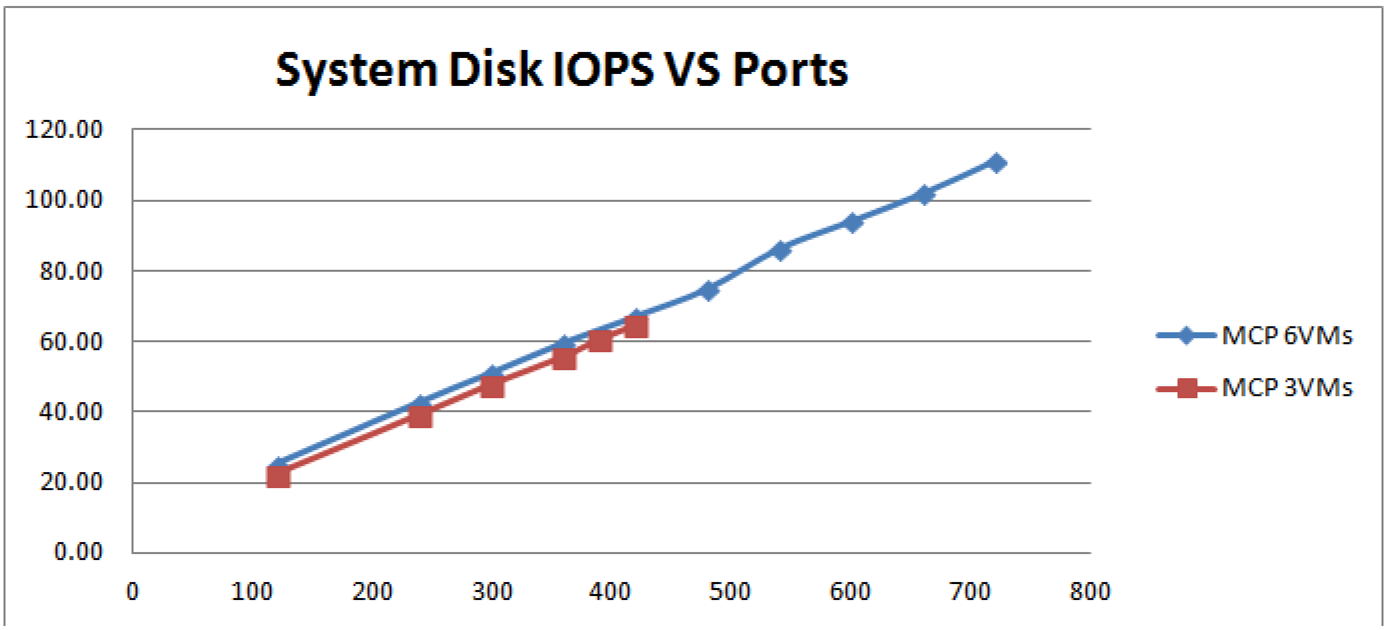


Figure 18: Comparison of System Disk IOPS among different VMs.

- System Disk IOPS scales linearly against port capacity, but not related for VM configurations.
- We ran an additional test with only 1 vCPU assigned to each VM, on a single hex core server Hardware profile 2, with a 6-VMs in total on the one server. We could barely run beyond 150 ports—the single CPU cannot be linearly scaled—which compares with a 3-VMs configuration:

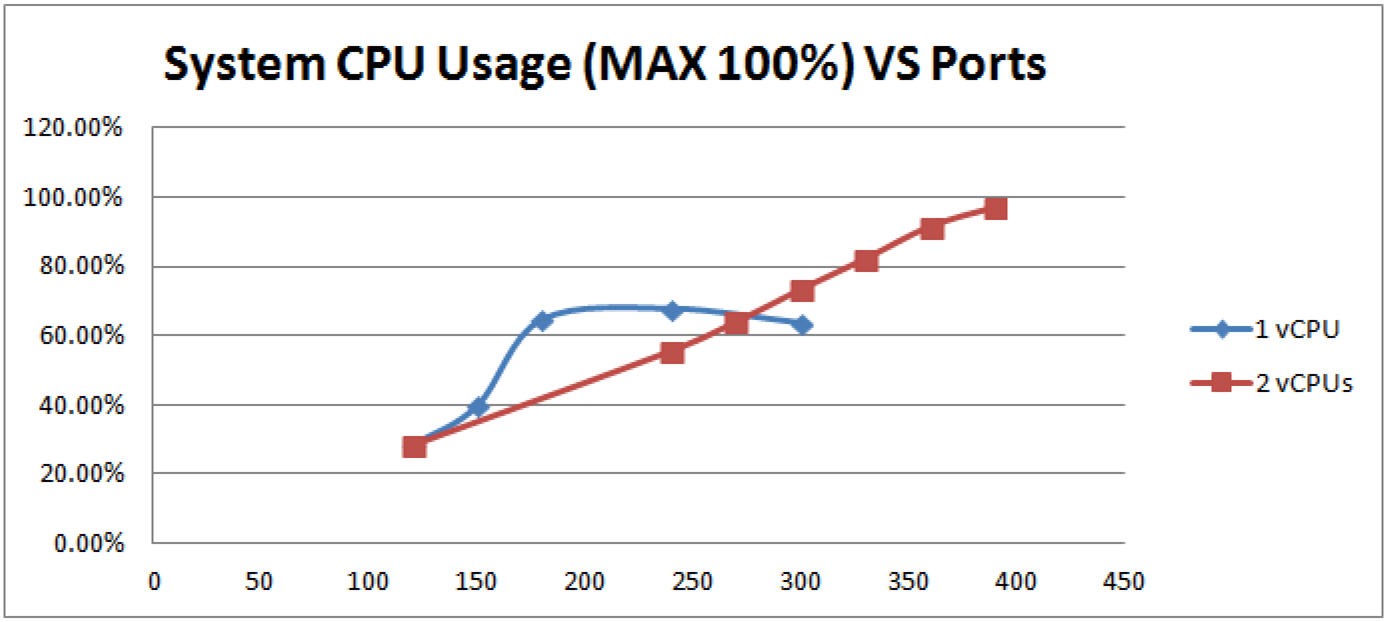


Figure 19: Comparison of System Usage for one vCPU vs two vCPUs VMs configuration

The two graphs below show that both Max Jitter and Max Delta jump significantly beyond 150 ports:

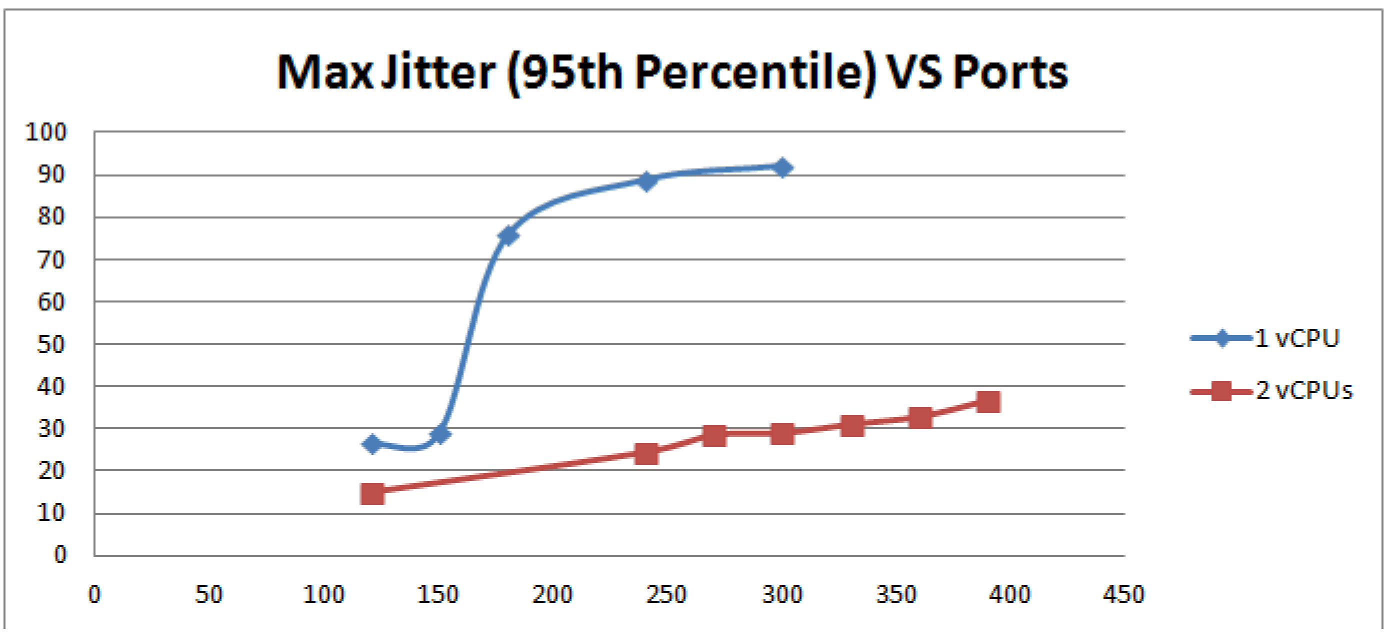


Figure 20: Comparison of Max Jitter for one vCPU vs two vCPUs VMs configuration

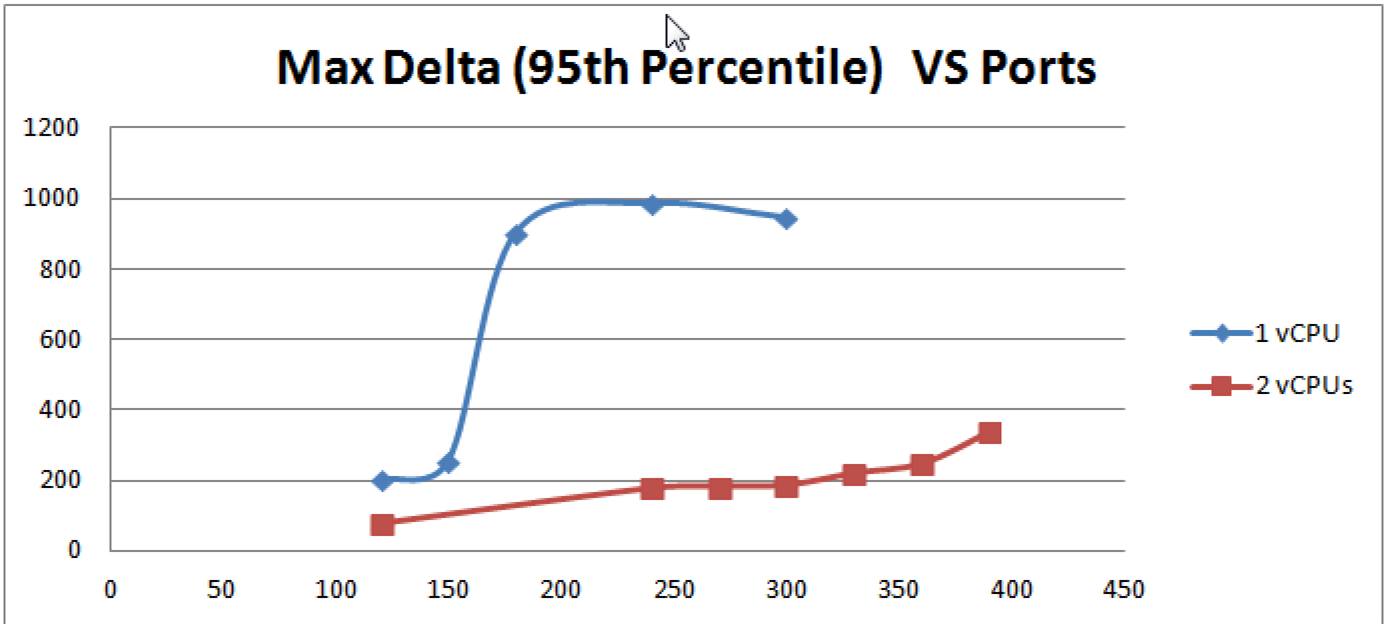


Figure 21: Comparison of Max Delta for one vCPU vs two vCPUs VMs configuration

The comparison indicates that MCP doesn’t perform well on a single vCPU VM.

Performance Comparison of MP3 only and MP3 + WAV

The graph below compares two test profiles (Profile 1 of MP3 only and Profile 2 of MP3 + WAV as dest2) on the same hardware spec with same 6 VM configurations of 2 vCPU per VM. Below is the CPU usage:

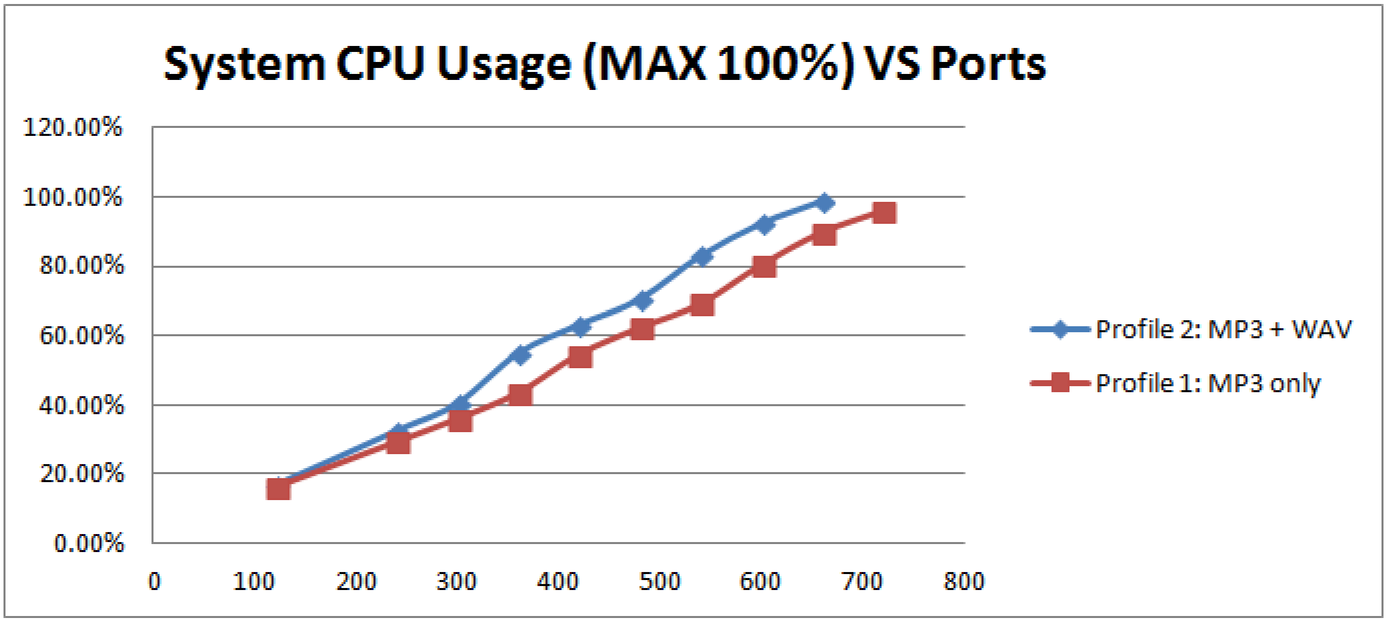


Figure 22: Comparison of System Usage for different test profiles

Overall CPU usage for Software Profile 2 (MP3 + WAV) is slightly higher than for Software Profile 1 (MP3 only).

The two graphs below compare audio quality criteria:

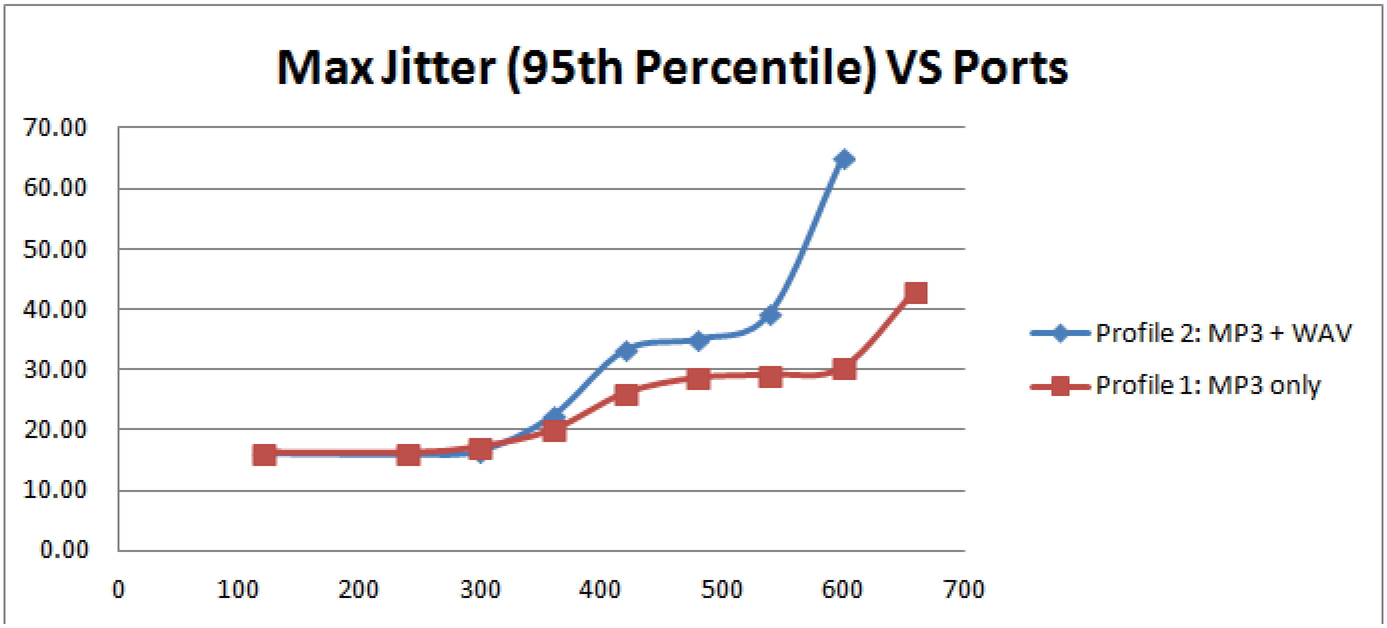


Figure 24: Comparison of Max Jitter among different test profiles

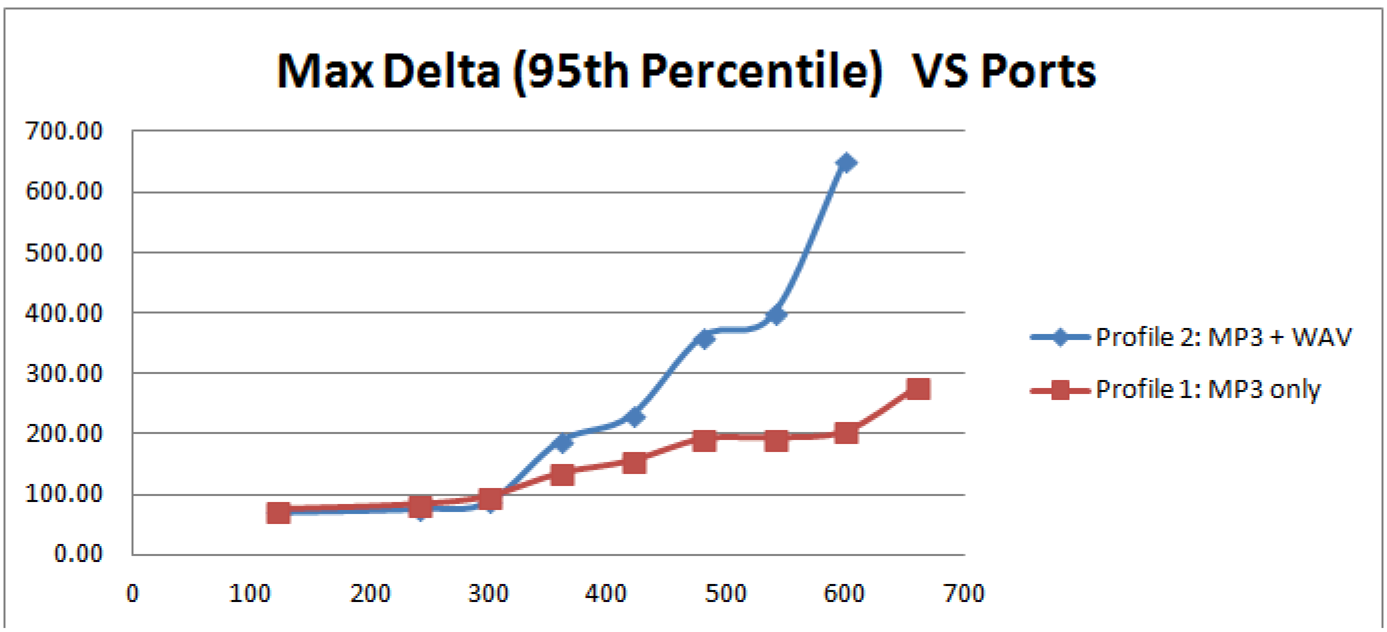


Figure 25: Comparison of Max Delta among different test profiles

- For this test, applying Profile 2 to a 6 VMs configuration: Preferred/Recommended = 360 ports; Peak Port Capacity = 530 ports, if you can ignore some potential impact to audio quality.

The table below shows the IOPS of the sum of all 6 VMs for a test profile of MP3 + wav:

Table: IOPS of sum of all 6 VMs of dual hex cores, MP3 + wav

Ports	Overall Disk IOPS (kbps)		
Total	Reads	Writes	
120	42.64	0.01	42.63
240	77.69	0.00	77.69
300	95.99	0.00	95.99
360	114.28	0.00	114.28
420	130.45	0.00	130.45
480	149.58	0.00	149.58
540	172.49	0.00	172.49
600	194.55	0.00	194.55
660	177.80	0.00	177.80

The graph below compares Table: IOPS of sum of all 6 VMs of dual hex cores, MP3 + wav with Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only:

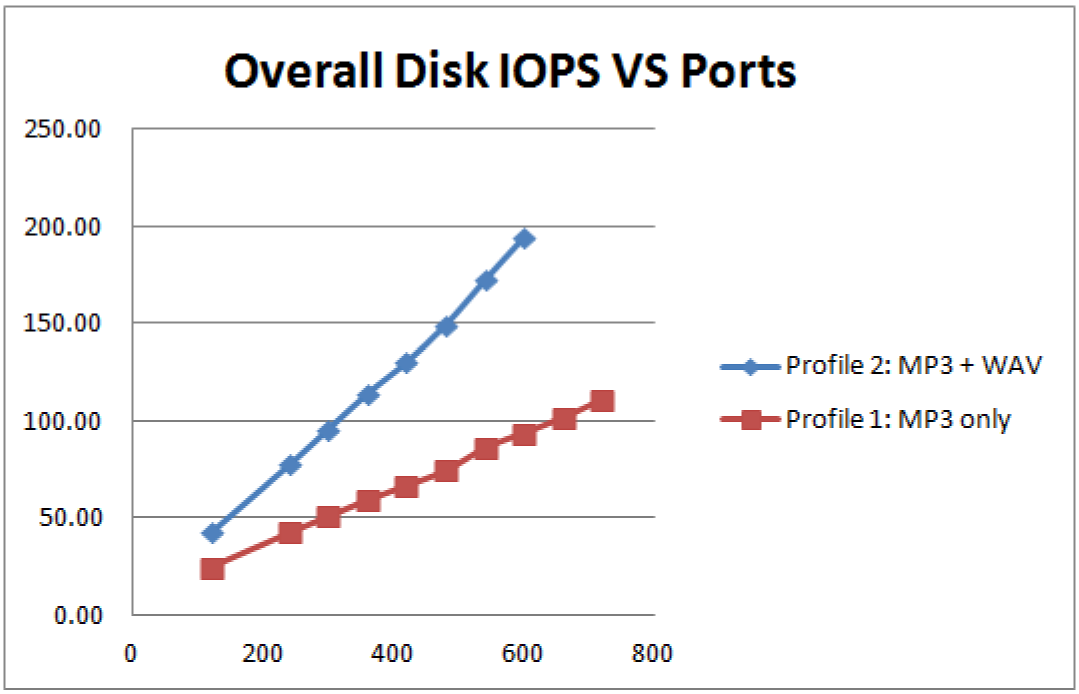


Figure 26: Comparison of System Disk IOPS for different test profiles on VMs

As we have cache folder on a different SSD drive, we can break down disk IOPS for each drive as below:

Table: Disk IOPS Break Down per Drive, Test Profile 1, MP3 only

Ports	Overall Disk IOPS (kbps)			SSD Drive E Disk IOPS (kbps)			HDD Drive C Disk IOPS (kbps)		
Total	Reads	Writes	Total	Reads	Writes	Total	Reads	Writes	
120	25.18	0.03	25.15	20.88	0.00	20.88	4.30	0.03	4.28
240	42.75	0.05	42.70	36.96	0.00	36.96	5.79	0.05	5.74
300	51.16	0.00	51.15	44.63	0.00	44.63	6.53	0.00	6.53
360	59.61	0.00	59.61	52.80	0.00	52.80	6.81	0.00	6.81
420	67.04	0.00	67.04	60.31	0.00	60.31	6.74	0.00	6.74
480	74.82	0.00	74.82	67.85	0.00	67.85	6.97	0.00	6.97
540	86.30	0.00	86.30	79.31	0.00	79.31	6.99	0.00	6.99
600	94.11	0.00	94.11	87.31	0.00	87.31	6.80	0.00	6.80
660	102.05	0.00	102.04	95.12	0.00	95.12	6.92	0.00	6.92
720	111.30	0.00	111.29	104.30	0.00	104.30	6.99	0.00	6.99

Table: Disk IOPS Break Down per Drive, Test Profile 2, MP3 + wav

Ports	Overall Disk IOPS (kbps)			SSD Drive E Disk IOPS (kbps)			HDD Drive C Disk IOPS (kbps)		
Total	Reads	Writes	Total	Reads	Writes	Total	Reads	Writes	
120	42.64	0.01	42.63	38.38	0.00	38.38	4.26	0.01	4.26
240	77.69	0.00	77.69	72.07	0.00	72.07	5.62	0.00	5.62
300	95.99	0.00	95.99	89.04	0.00	89.04	6.95	0.00	6.95
360	114.28	0.00	114.28	107.50	0.00	107.50	6.78	0.00	6.78
420	130.45	0.00	130.45	123.56	0.00	123.56	6.89	0.00	6.89
480	149.58	0.00	149.58	142.65	0.00	142.65	6.92	0.00	6.92
540	172.49	0.00	172.49	165.61	0.00	165.61	6.88	0.00	6.88
600	194.55	0.00	194.55	187.53	0.00	187.53	7.02	0.00	7.02

The two graphs below compare corresponding drives:

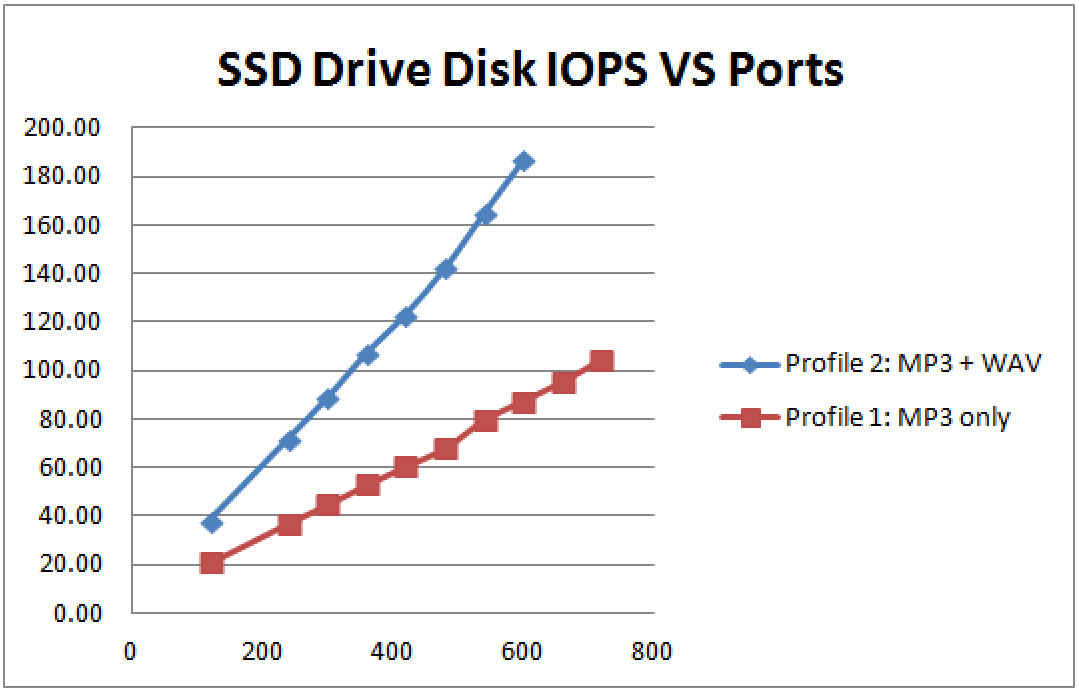


Figure 29: Comparison of cache folder of SSD Drive IOPS for different profiles

This SSD drive is used exclusively as the cache folder for MCP recording. The IOPS for Profile 2 (two dest2, MP3 + wav) is as double as Profile 1 (one dest. MP3 only).

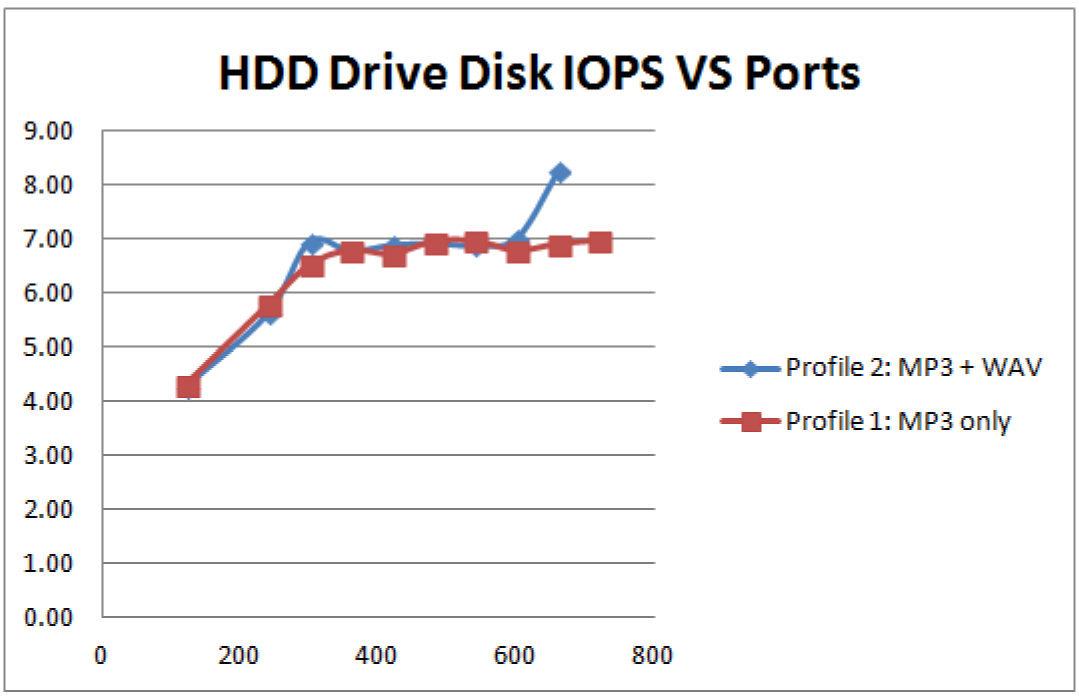


Figure 30: Comparison of HDD Drive IOPS for different profiles

This HDD drive is used for all operations except the cache folder for MCP recording. IOPS is nearly constant at a regular load and below peak. Thus, the IOPS estimating formula can be:

```
IOPS1 = C + k * P (dest only)
IOPS2 = C + 2k * P (both dest + dest2)
Where P = ports, C = 7, k = 0.15
```

Performance Comparison between SAS HDDs and SSD

These tests compare performance between SAS HDDs and SSD for recording, using 6 VMs from the same hardware spec, and these four different HDD and SSD combinations:

- **1 HDD:** all 6 VMs on one 15 krpm SAS HDD drive.
- **2 HDD:** split 6 VMs on two 15 krpm SAS HDD drives, 3 VMs per drive.
- **3 HDD:** split 6 VMs on three 15 krpm SAS HDD drives, 2 VMs per drive.
- **SSD:** all 6 VMs on one 15 krpm SAS HDD while a separate SSD drive used as cache folder only.

The testing was executed with Profile 1, MP3 only. Below is the overall system CPU usage:

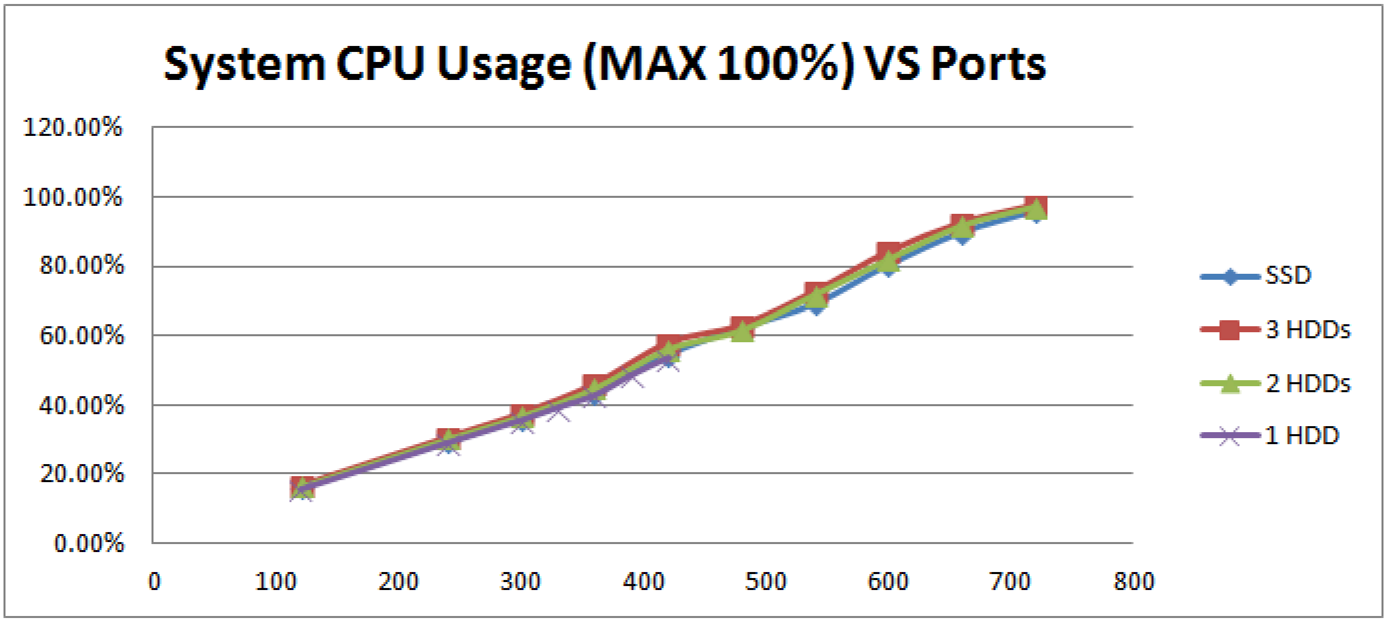


Figure: Comparison of System Usage among different HDD/SSD drive combinations

The overall system CPU usage exhibits no significant different between HDD and SDD.

IOPS is almost the same for these 4 combinations, so these tests use the numbers in Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only.

The graphs below compare max jitter and max delta for HDD/SSD drive combinations:

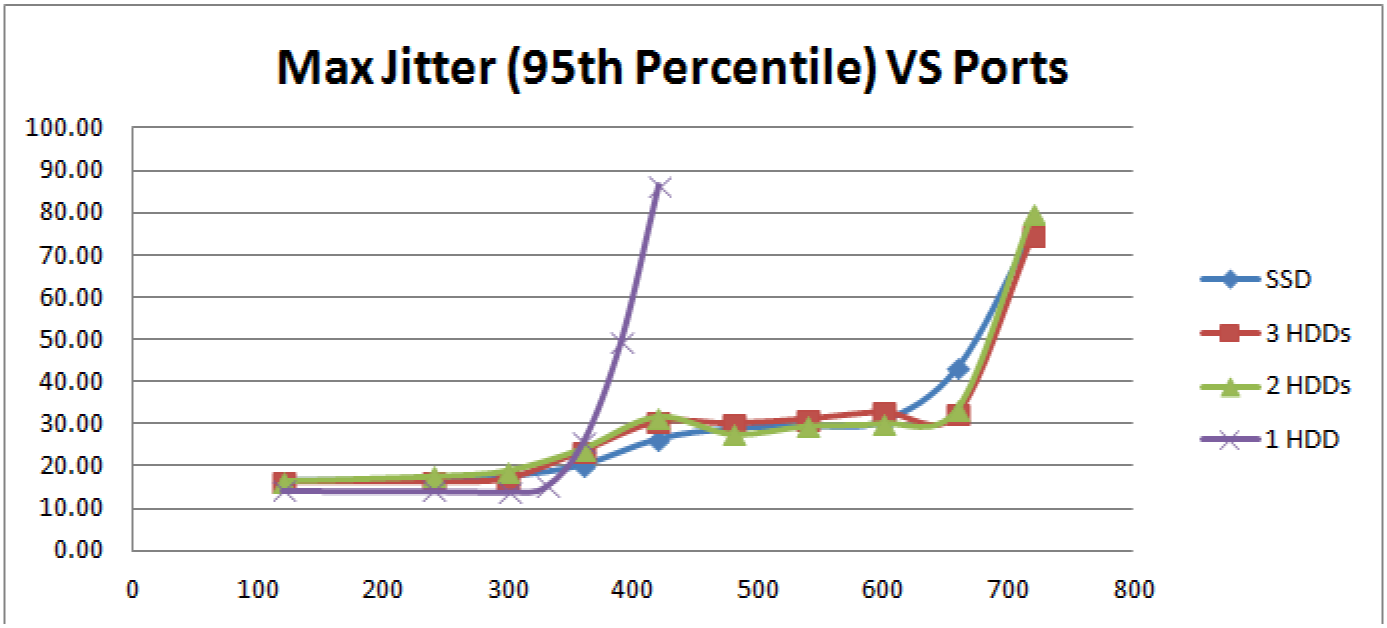


Figure 32: Comparison of Max Jitter among different HDD/SSD drive combinations

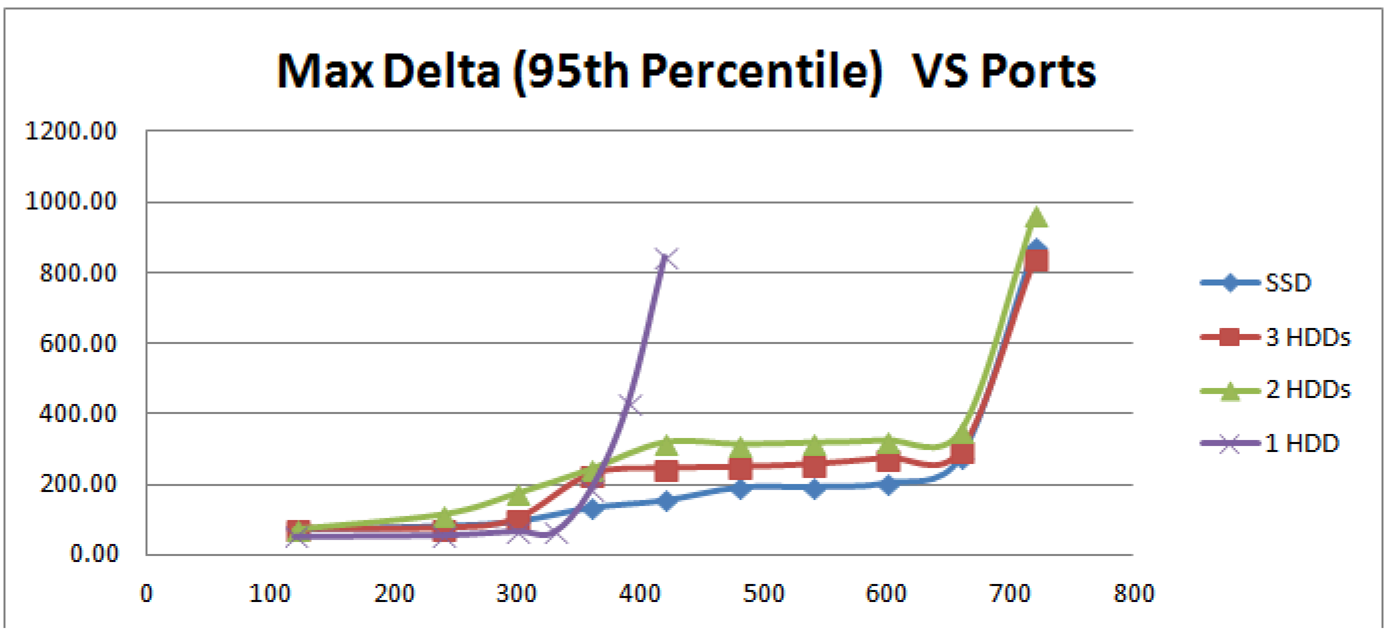


Figure 33: Comparison of Max Delta among different HDD/SSD drive combinations

This graph illustrates the average disk write queue for one drive:

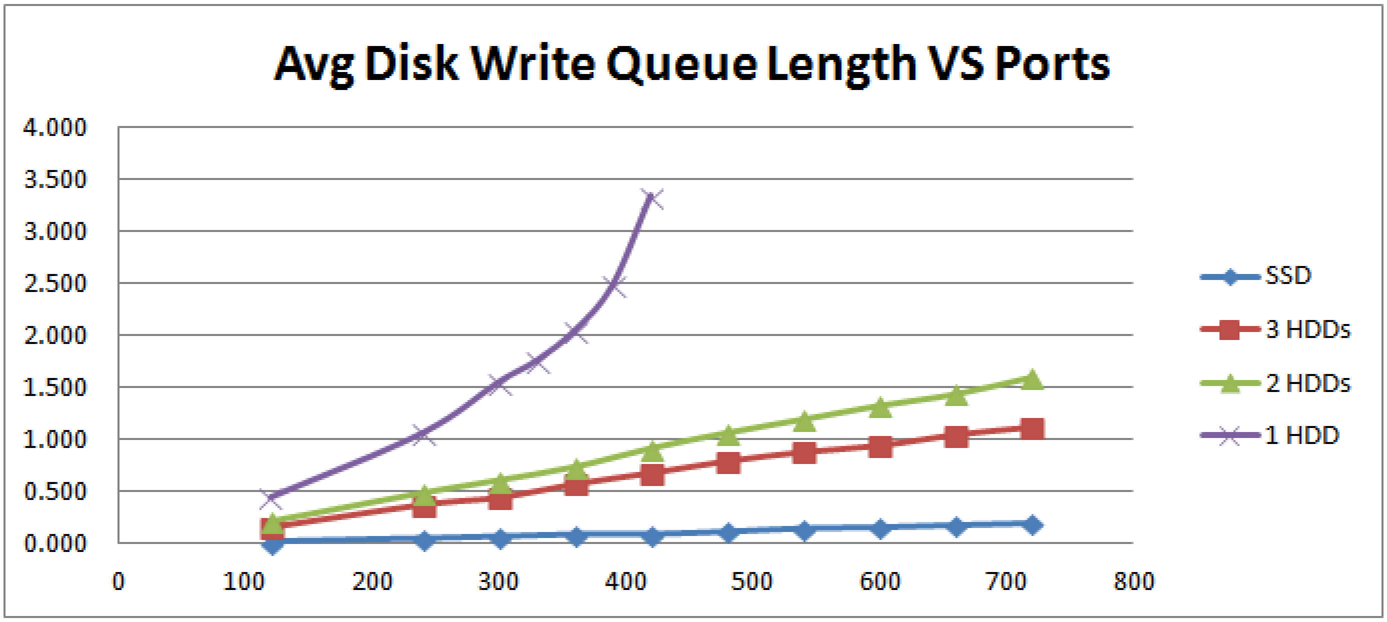


Figure 34: Comparison of Avg Disk Write Queue among different HDD/SSD drive combinations

- The queue starts to increase non linearly around 360 ports, which makes that number close to maximum port capacity of the hard drive.
- In the three graphs above: with only one HDD drive, both max jitter and max delta started to increase dramatically from 330 ports and higher. Thus: Preferable/Recommended = 330 ports; Peak Port Capacity = 360 ports. In Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only, IOPS is 51 for 330 ports; while IOPS is around 60 for 360 ports. Thus: Preferable/Recommended IOPS = 51; maximum IOPS for one 15 krpm SAS HDD = 60.
- With multiple HDDs (2 or 3) to split the load, peak port capacity is nearly the same as SSD—660 ports since the load per drive would be 330 (for 2 HDD drives) and 220 (for 3 HDD drives). Max jitter does not exhibit big differences for these three configurations. But max delta shows a higher delay for 3 HDDs compared to SSD, and 2 HDDs compared to 3 HDDs. Thus: with strict audio quality required in these scenarios, fast media such as SSD will help improve latency and minimize any potential audio quality issues.

Data Throughput

These two formulas estimate data throughputs:

Formula 1 (for MP3 only):
`MP3 bitrate * Ports / 8 = KB/sec`
Or `32kbps * Ports / 8 =KB/sec if MP3 is 32kbps`

Formula 2 (for MP3 + wav):
`(MP3 bitrate + WAV bitrate) * Ports / 8 = KB/sec`
Or `(32 kbps + 128 kpbs) * Ports / 8 = 160 kbps / 8 = KB/sec if 32kpbs MP3 + wav`

Six VM configurations, with SSD as the cache folder for MCP recording, produced the following measurements from testing for test SW Profile 1 (MP3 32 kbps only):

Table: Data Throughputs for MP3 32 kbps only

Ports	Overall Disk (kbps)			SSD Drive Disk (kbps)		
Total	Reads	Writes	Total	Reads	Writes	
120	554.10	0.12	553.98	536.98	0.00	536.98
240	1075.70	0.50	1075.19	1053.19	0.00	1053.19
300	1332.61	0.06	1332.55	1308.69	0.00	1308.69
360	1601.09	0.00	1601.09	1575.02	0.00	1575.02
420	1847.91	0.00	1847.91	1822.30	0.00	1822.30
480	2109.57	0.00	2109.57	2082.49	0.00	2082.49
540	2461.25	0.00	2461.25	2434.04	0.00	2434.04
600	2728.83	0.00	2728.83	2702.57	0.00	2702.57
660	3010.07	0.00	3010.07	2982.84	0.00	2982.84
720	3310.64	0.00	3310.64	3280.45	0.00	3280.45

Apply Formula 1 to the 120-port and 600-port samples from the table above to achieve these results:

```
32 kpbs * 120 / 8 = 480 kb close to 534 in the table (in SSD)
32 kpbs * 600 / 8 = 2400 kb close to 2703 in the table (in SSD)
```

The measurements from real testing are slightly higher than calculations. Because other files such as metadata and JSON files are saved in the same cache folder, the formula might need adjusting.

Below is the table from testing measurement for SW Profile 2 (MP3 + wav) on the same 6 VM configuration with SSD as cache folder of MCP recording:

Table: Data Throughputs for MP3 32 kbps + wav

Ports	Overall Disk (kbps)			SSD Drive Disk (kbps)		
Total	Reads	Writes	Total	Reads	Writes	
120	2444.255	0.25	2444.01	2427.025	0.000	2427.025
240	4881.163	0.01	4881.15	4859.951	0.000	4859.951
300	6083.649	0.00	6083.64	6058.294	0.000	6058.294
360	7380.491	0.00	7380.49	7354.970	0.000	7354.970
420	8547.663	0.00	8547.66	8522.034	0.000	8522.034
480	9828.785	0.00	9828.79	9802.991	0.000	9802.991
540	11093.931	0.00	11093.93	11067.838	0.000	11067.838
600	12335.993	0.01	12335.99	12309.182	0.000	12309.182

Apply Formula 2 to the 120-port and 600-port samples in the table above to achieve these results:

```
(32 kbps + 128 kbps) * 120 / 8 = 2400 kb close to 2427 in the table
(32 kbps + 128 kbps) * 600 / 8 = 12000 kb close to 12309 in the table
```

MCP IOPS

A single HDD local hard drive was used for testing because the HDD itself could become the bottleneck. These tests focus on disk IOPS measurement and calculation, and certain real deployment scenarios require that a local drive not be used. Thus, the measurement for MCP IOPS is useful to calculate overall IO requirement. The three tables below offer three typical MCP IOPS configurations:

Table: MCP IOPS on physical server of single hex core, MP3 only

Ports	Physical MCP IOPS (kbps)		
Total	Read	Write	
60	21.88	14.93	6.95
120	43.25	29.64	13.60
180	63.28	43.37	19.91
200	71.47	49.35	22.12
210	74.95	51.58	23.37
220	78.40	53.93	24.47
230	82.51	56.83	25.68
240	85.28	58.37	26.90

Table: MCP IOPS on 6 VMs of dual hex core, MP3 only

MP3 only Ports	Overall 6 VM MCP IOPS (kbps)		
Total	Read	Write	
120	43.672	29.679	13.993
240	87.477	59.280	28.197
300	109.292	73.848	35.445
360	130.965	88.917	42.047
420	151.706	103.095	48.611
480	171.290	116.053	55.237
540	194.772	132.867	61.905
600	215.101	146.882	68.219
660	236.654	161.990	74.664
720	259.100	177.279	81.820

Table: MCP IOPS on 6 VMs of dual hex core, MP3 + wav

MP3 + wav Ports	Overall 6 VM MCP IOPS (kbps)		
Total	**Read**	**Write**	
120	173.607	146.092	27.515
240	348.369	292.884	55.486
300	434.511	364.898	69.613
360	522.447	439.412	83.035
420	600.880	504.682	96.198
480	693.861	583.965	109.896
540	780.187	656.958	123.229
600	859.359	722.907	136.453
660	790.737	664.024	126.713

The graph below compares the two tables MCP IOPS on physical server of single hex core, MP3 only and MCP IOPS on 6 VMs of dual hex core, MP3 only:

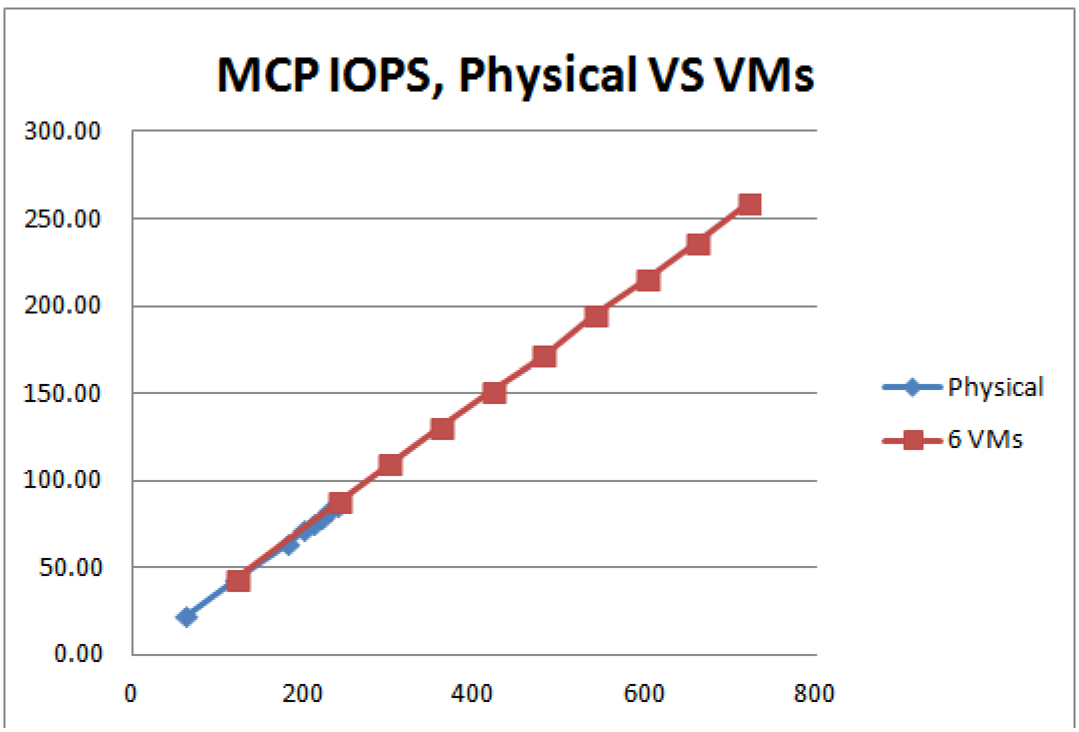


Figure 40: MCP IOPS, Physical VS VMs

The graph below compares the two tables MCP IOPS on 6 VMs of dual hex core, MP3 only and MCP IOPS on 6 VMs of dual hex core, MP3 + wav:

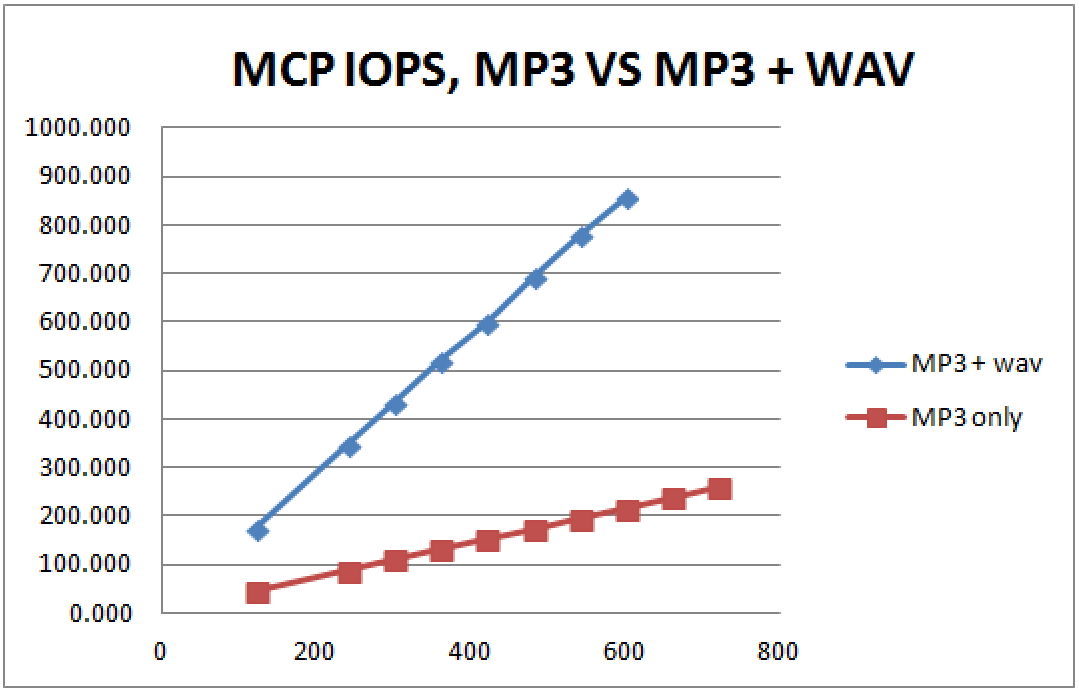


Figure 41: MCP IOPS, MP3 only VS MP3 + WAV

The MCP IOPS is related to the test profile and ports, but on the same physical server and VMs.

MP3 16kbps Bit Rate Compression

Support for MP3 16 kbps bit rate recording compression began with The GVP 8.5.1 release in December 2014. We tested performance on physical server and Virtual Machine (VM) environments, using Windows 2008 R2 x64.

Physical Server on Single Hex Core

Testing was performed on Hardware Profile 1: a physical server on a single hex core of Dell R410. The three graphs below compare system CPU usage and audio quality related metrics, max jitter and max delta.

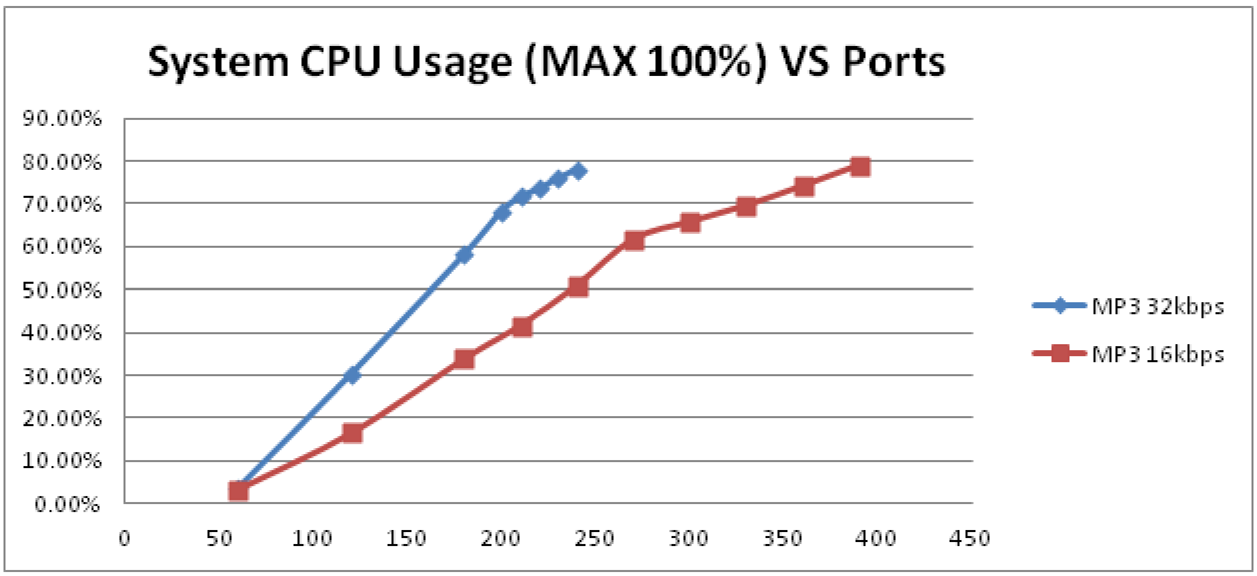


Figure 42: Comparison of System CPU Usage, MP3 16kbps vs 32kbps on Physical Server

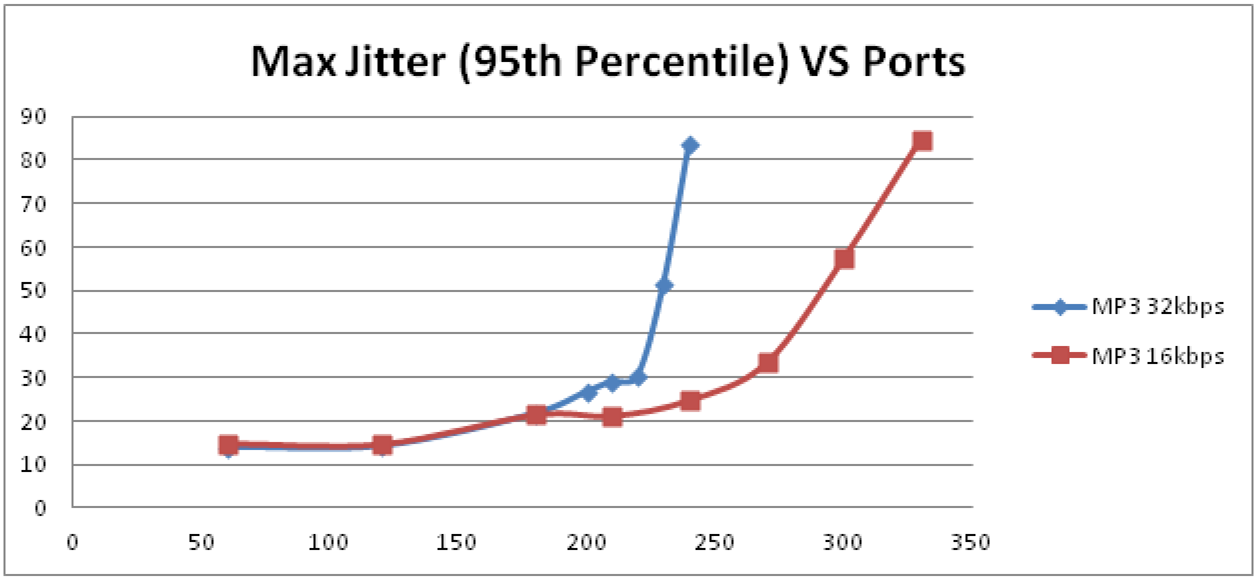


Figure 43: Comparison of Max Jitter, MP3 16kbps vs 32kbps on Physical Server

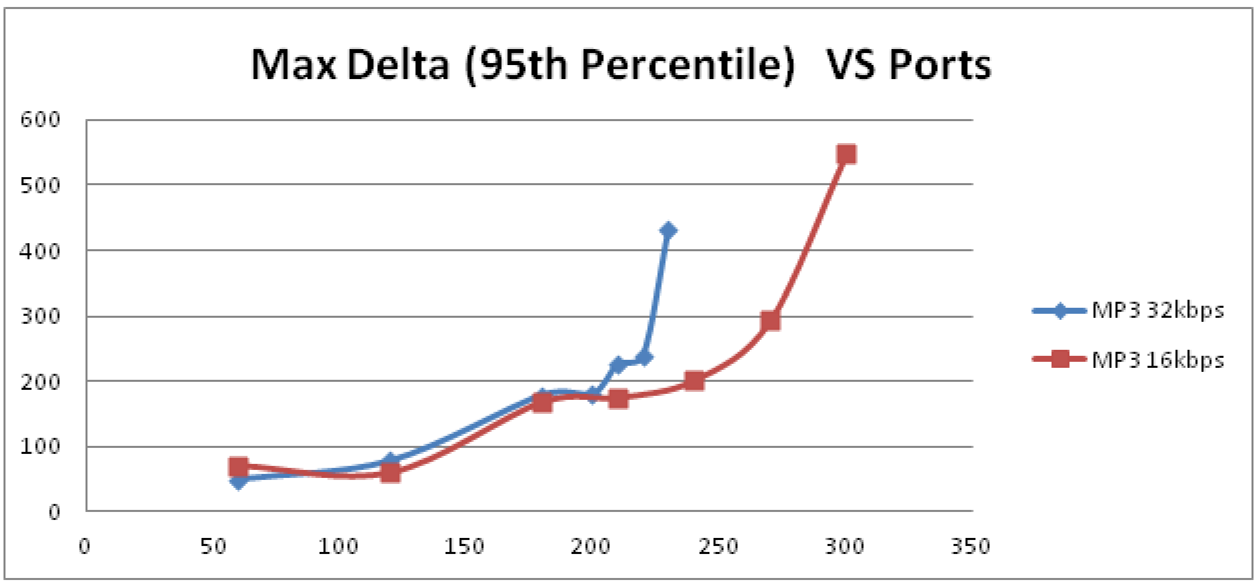


Figure 44: Comparison of Max Delta MP3 16kbps vs 32kbps on a Physical Server

MP3 16kbps consumes less CPU memory, which means higher port capacity. The two graphs above that compare Max Jitter with Max Delta also indicate the higher port capacity of MP3 16kbps. Recommended port capacity for MP3 16kbps: 240 ports (20% higher than the 200 recommended port capacity for MP3 32kbps). Peak port capacity: 270 ports (22.7% higher than the 220 peak port capacity for MP3 32kbps).

The table below lists the system disk IOPS:

Figure 45: System Disk IOPS on Physical Server, MP3 only 16 Kbps

Ports	Physical Server Disk IOPS (kbps)		
Total	Reads	Writes	
60	14.66	0.036	14.62
120	24.00	0.041	23.95
180	33.42	0.029	33.39
210	37.65	0.030	37.62
240	42.21	0.029	42.18
270	47.18	0.036	47.14
300	51.44	0.011	51.43
330	55.81	0.006	55.81
360	60.99	0.002	60.99
390	67.12	0.003	67.11

The graph below compares Table: System Disk IOPS on Physical Server, MP3 only 16 Kbps with Table: Disk IOPS of system level from a physical server with a single hex core, both on a single hex core server:

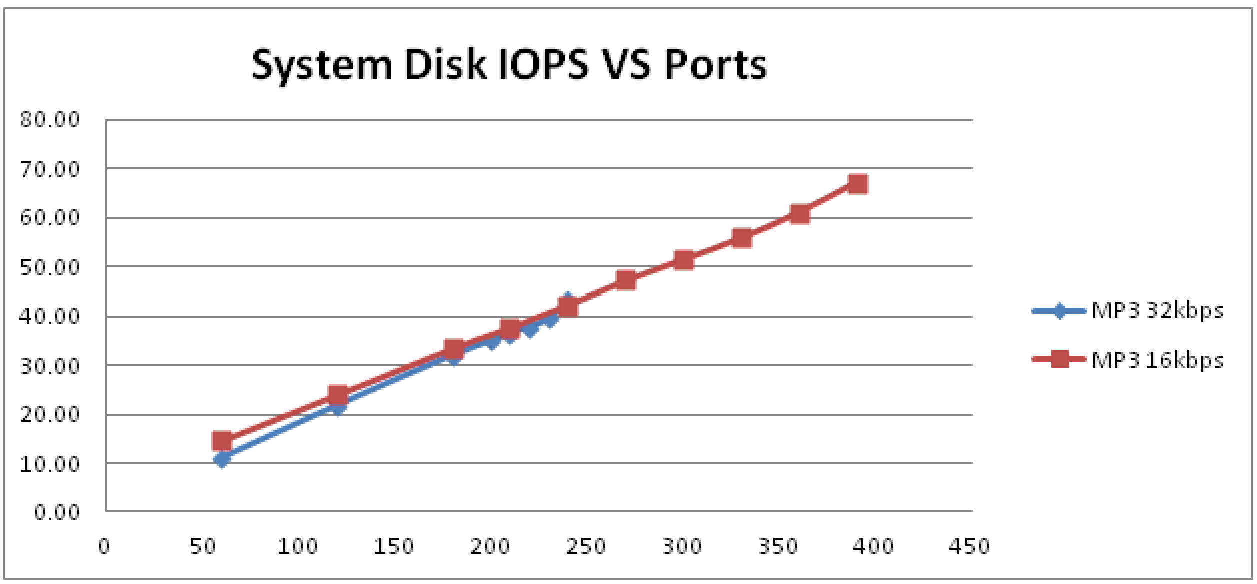


Figure 46: Comparison of System Disk IOPS on Single Hex Core Physical Server, MP3 16 Kbps vs 32 Kbps

The system disk IOPS for MP3 16kbps and 32kbps are nearly identical to each other; reasonable since the disk IO operations should be the same, and at the same port capacity, no matter which MP3 bit rate is chosen.

MCP IOPS is listed here:

Figure 47: MCP IOPS on physical server of single hex core, MP3 only, 16 Kbps

Ports	Physical Server MCP IOPS (kbps)		
Total	Reads	Writes	
60	14.56	7.53	7.04
120	28.64	14.92	13.72
180	42.54	22.29	20.25
210	49.42	25.93	23.48
240	56.41	29.64	26.76
270	63.38	33.34	30.04
300	70.36	36.92	33.44
330	77.53	40.79	36.74
360	85.52	44.46	41.06
390	94.68	48.14	46.54

The graph below compares Table: MCP IOPS on physical server of single hex core, MP3 only, 16 Kbps and Table: MCP IOPS on physical server of single hex core, MP3 only:

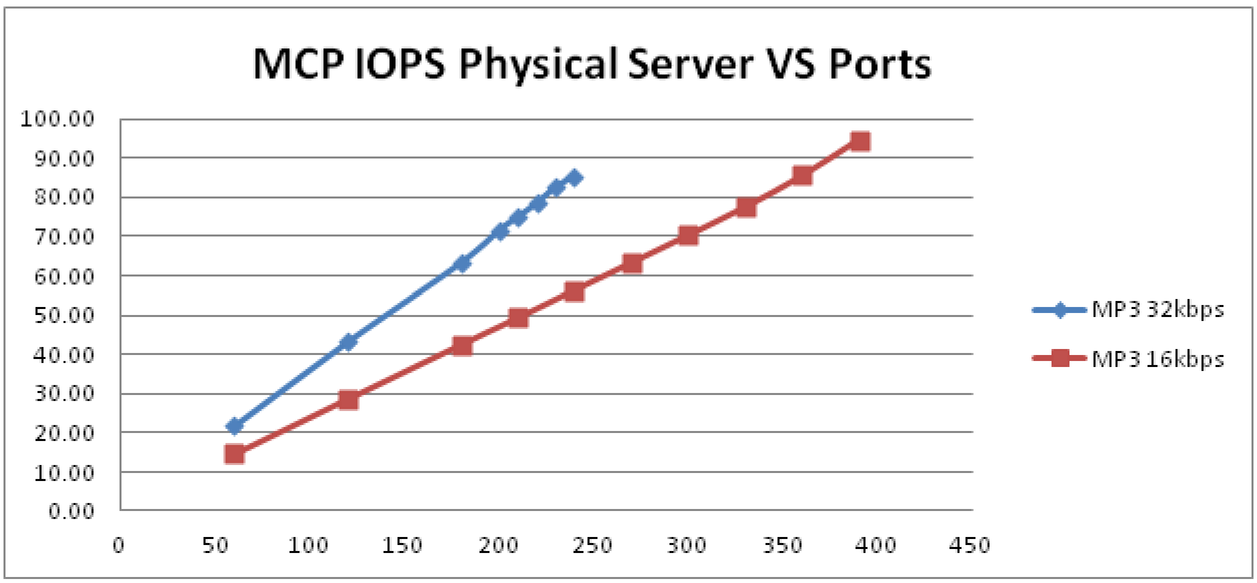


Figure 48: MCP IOPS on Single Hex Core Physical Server, MP3 16Kbps vs 32Kbps

MP3 16kbps uses less IOPS at the process level, probably be due to fewer network operations for MP3 16kbps.

VMs on Dual Hex Cores Server

The testing for MP3 16kbps was conducted on VM Profile 4 (based on Hardware Profile 4, which is a dual hex cores server). 6 VMs were configured, while only one MCP was installed on each Windows VM. The three graphs below compare overall CPU usage, audio quality related max jitter and max delta for MP3 16kbps vs 32kbps:

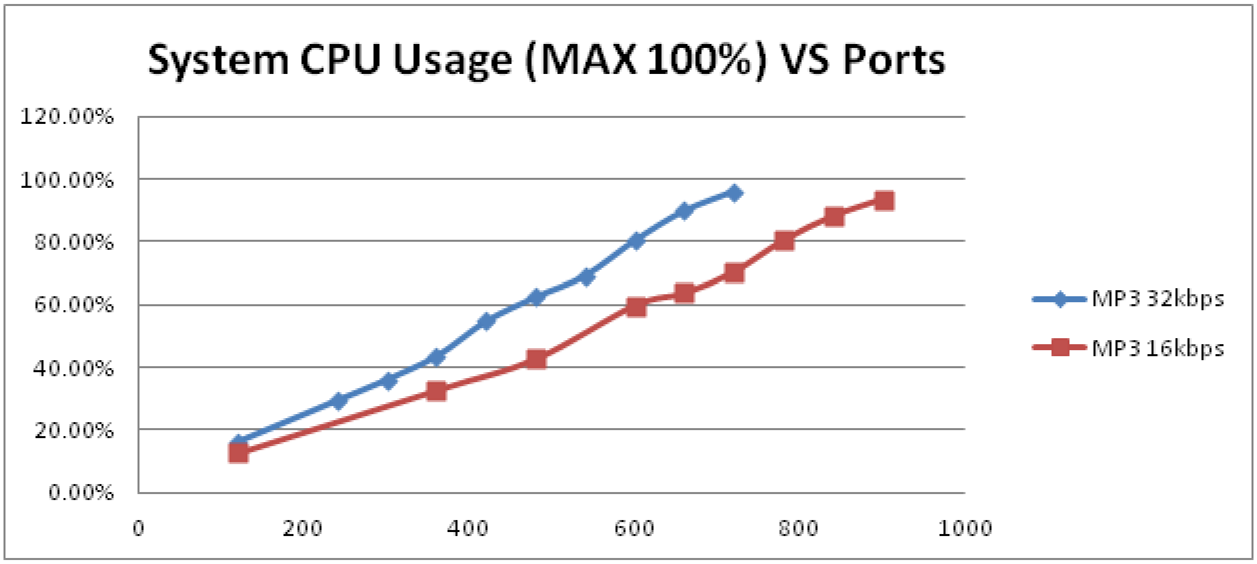


Figure 49: Comparison System CPU Usage of MP3 16kbps vs 32kbps on VM env

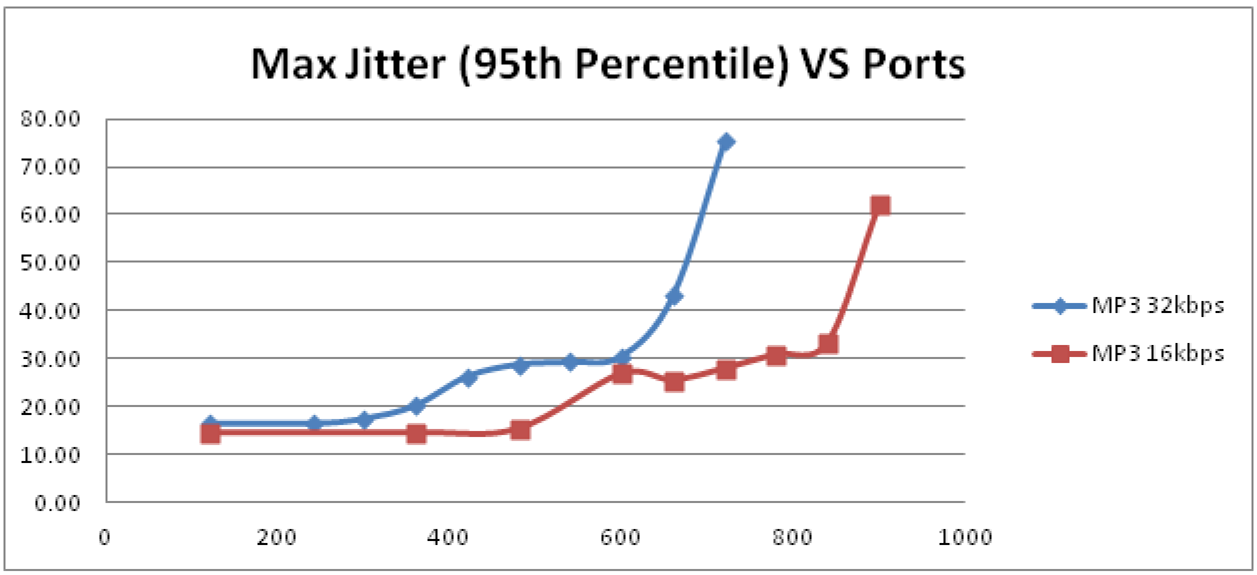


Figure 50: Comparison of Max Jitter, MP3 16kbps vs 32kbps on VM env

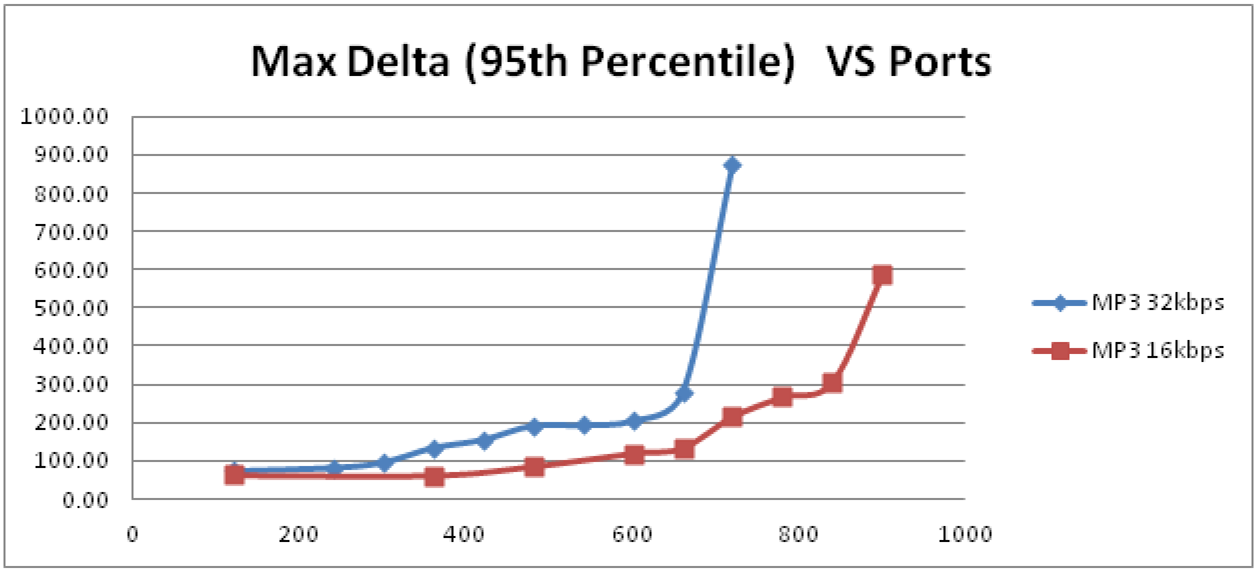


Figure 51: Comparison of Max Delta, MP3 16kbps vs 32kbps on VM environment

MP3 16kbps consumes less CPU memory, which matches test results on a physical server in Figure: Comparison of System CPU Usage, MP3 16kbps vs 32kbps on Physical Server. Both Max Jitter and Max Delta also show a higher port capacity for MP3 16kbps compression, which also matches test results on a physical server from Figure 4: Comparison of Max Jitter, MP3 16kbps vs 32kbps on Physical Server & Figure: Comparison of Max Delta MP3 16kbps vs 32kbps on a Physical Server. Preferred/Recommended port capacity for MP3 16 kbps: 720 ports (20% higher 600 ports for than MP3 32kbps). It’s the same increase as observed from a physical server. Peak port capacity for MP3 16kpbs can be as high as 840 ports (27.3% higher than 660 peak port capacity for MP3 32kbps.

The table below illustrates system disk IOPS:

Figure 52: Overall Disk IOPS on all 6 VMs of dual hex cores, MP3 only, 16 Kbps

Ports	Overall 6 VMs Disk IOPS (kbps)		
Total	Reads	Writes	
120	26.57	0.13	26.44
360	63.47	0.13	63.34
480	80.66	0.15	80.51
600	93.73	0.04	93.69
660	109.53	0.14	109.39
720	118.76	0.13	118.62
780	126.15	0.07	126.08
840	134.12	0.04	134.09
900	142.21	0.09	142.12

The graph below compares overall disk IOPS of all 6 VMs for MP3 16kpbs against 32kbps in Table: Disk IOPS of sum of all 6 VMs of dual hex cores, MP3 only':

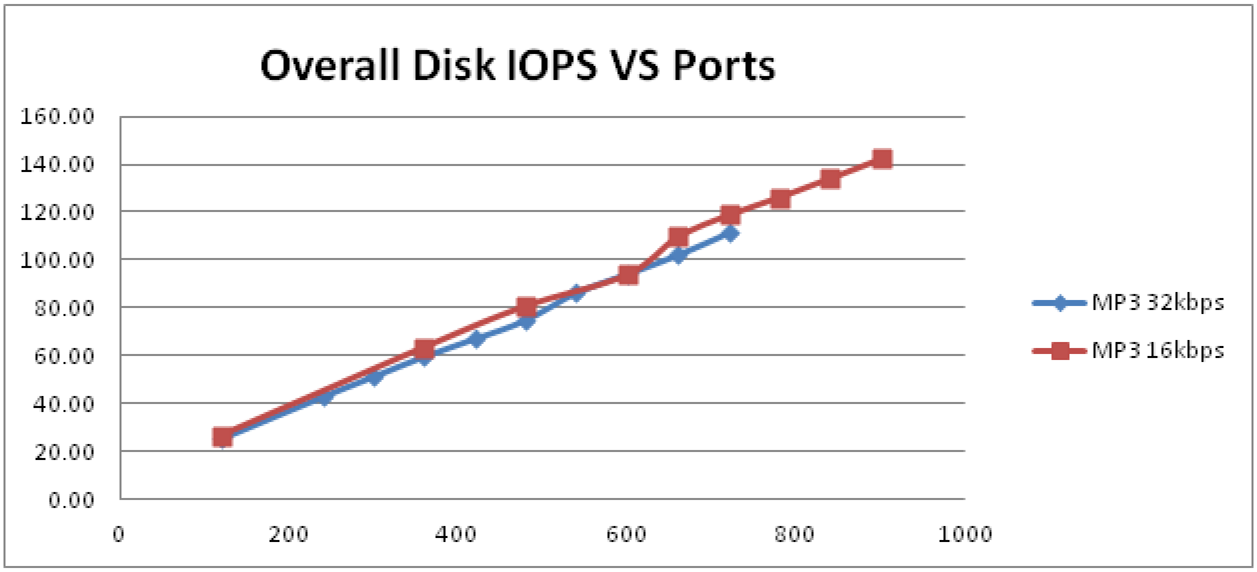


Figure 53: Comparison of Overall 6 VMs Disk IOPS MP3 16 kbps vs 32 kbps

The IOPS from both MP3 16kbps and 32kbps are inline with each other, as in the physical server tests.

Data throughput for MP3 16kbps is listed in following table:

Figure 54: Data Throughputs for MP3 only, 16 kbps

Ports	Overall Disk (kbps)			SSD Drive Disk (kbps)		
Total	Reads	Writes	Total	Reads	Writes	
120	318.17	0.68	317.49	296.313	0.001	296.312
360	892.94	0.52	892.42	856.077	0.001	856.076
480	1175.63	0.79	1174.84	1132.997	0.001	1132.996
600	1537.43	0.19	1537.23	1510.543	0.000	1510.543
660	1729.45	0.58	1728.87	1680.374	0.003	1680.371
720	1890.48	0.58	1889.90	1837.492	0.000	1837.492
780	2045.34	0.35	2045.00	1995.239	0.004	1995.235
840	2191.98	0.15	2191.83	2142.373	0.002	2142.371
900	2349.18	0.75	2348.44	2298.426	0.004	2298.422

Using this formula:
`MP3 bitrate * Ports / 8 = kbps`
...where MP3 bitrate=16kbps and Ports = `120` and `720` from the table above,

The results...
`16 kpbs * 120 / 8 = 240` kbps (compared to 296 in the table -- in SSD)
and
`16 kpbs * 720 / 8 = 14400` kbps (compared to 1837 in the table -- in SSD)

...from real testing for MP3 16kbps are slightly higher than calculations predict, due to other files such as metadata and JSON files being saved in the same cache folder. So the formula still stands.

The following table lists MCP IOPS:

Figure 55: Overall MCP IOPS from 6 VMs of dual hex core, MP3 only, 16kbps

Ports	Overall 6 VMs MCP IOPS (kbps)		
Total	Reads	Writes	
120	28.931	14.915	14.016
360	86.517	44.456	42.061
480	114.574	59.153	55.421
600	142.112	73.730	68.382
660	156.495	81.359	75.136
720	170.237	88.660	81.577
780	184.173	96.048	88.125
840	197.767	103.263	94.504
900	211.644	110.545	101.099

The graph below compares Overall MCP IOPS with MP3 32k MCP IOPS, and shows the same trend of physical server results that appeared in Figure: MCP IOPS on Single Hex Core Physical Server, MP3 16Kbps vs 32Kbps:

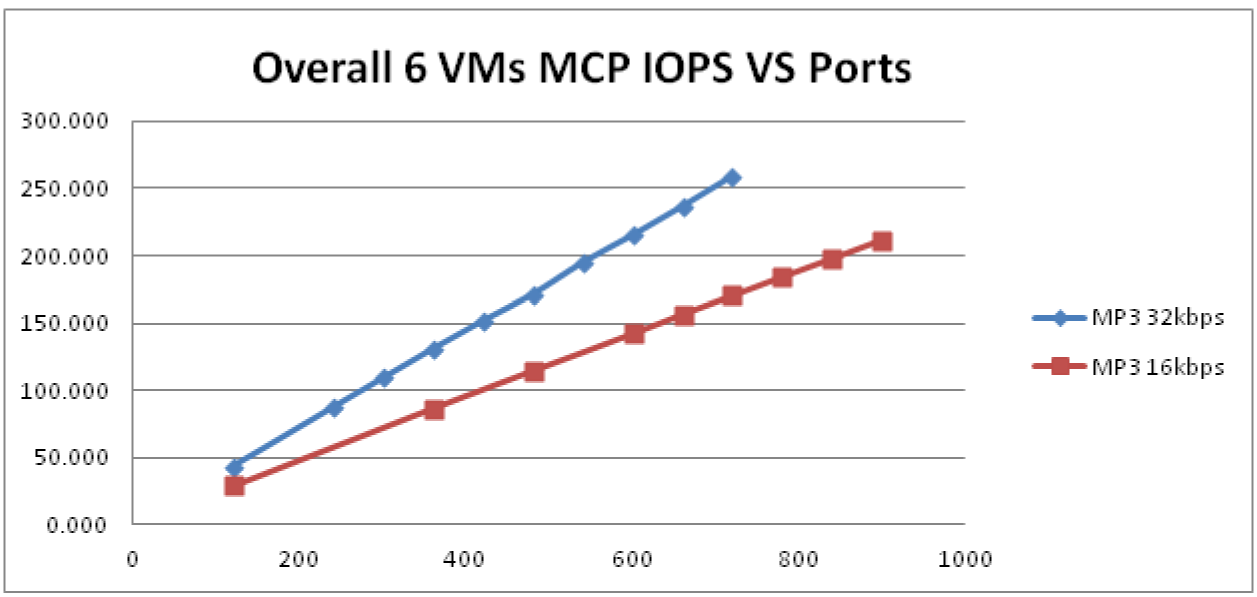


Figure 56: MCP IOPS on 6 VMs of Dual Hex Cores, MP3 16Kbps vs 32Kbps

MP3 16 kbps Bit Rate with Encryption

We tested the MP3 16 kbps bit rate with encryption, using the `dest2` physical server and Vitual Machine (VM) environments, which compares with results of non-encryption from MP3 16 kbps Bit Rate without Encryption. The OS remained Windows 2008 R2 x64.

Physical Server on Single Hex Core

These tests were performed on Hardware Profile 1: a physical server on a single hex core of Dell R410. The three graphs below compare system CPU usage and audio quality-related metrics, max jitter and max delta.

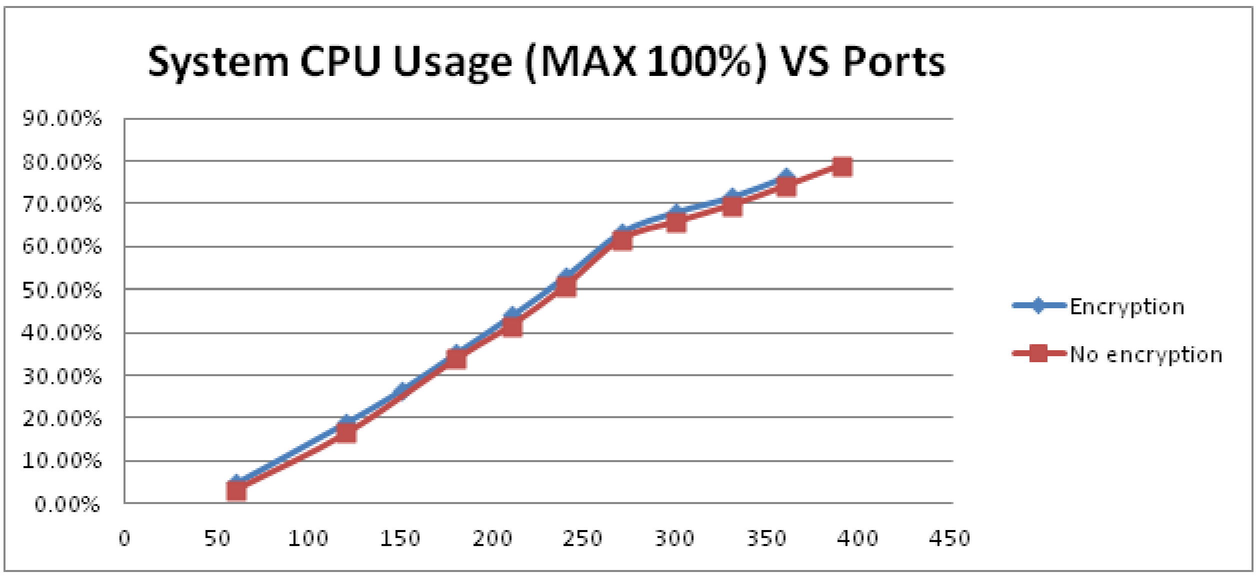


Figure 57: Comparison of Physical Server System CPU Usage of MP3 16kbps encryption vs non-encryption

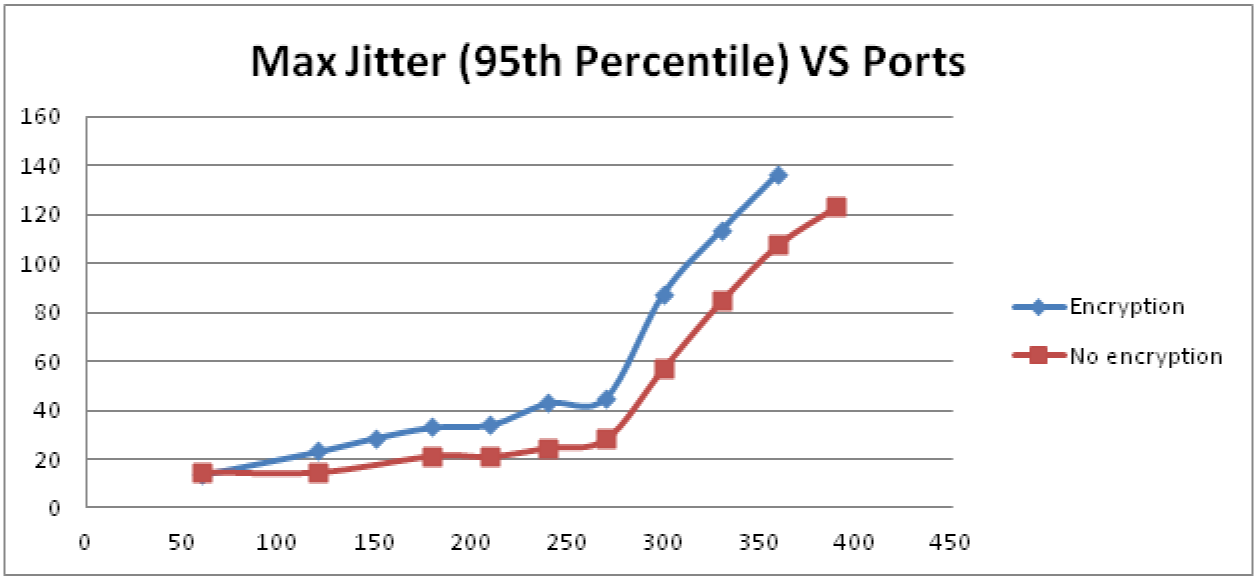


Figure 58: Comparison of Physical Server Max Jitter of MP3 16kbps encryption vs non-encryption

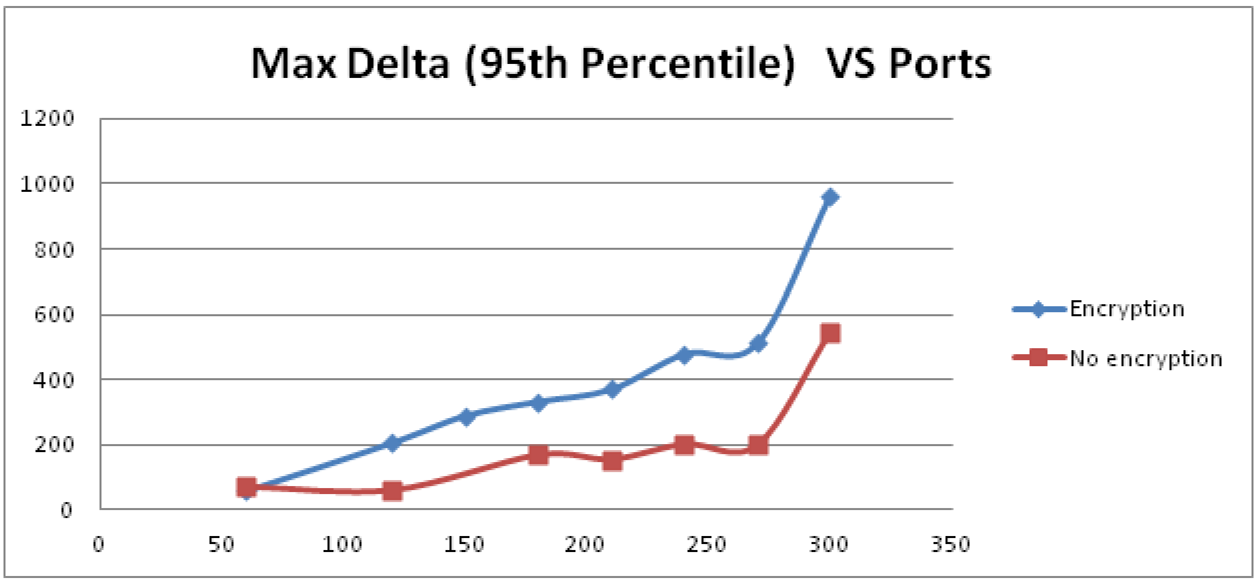


Figure 59: Comparison of Physical Server Max Delta of MP3 16kbps encryption vs non-encryption

In the graphs above, encryption consumes slightly higher system CPU than does non-encryption. Max Jitter and Max Delta consume much more CPU with encryption, than without. If a slightly higher delay due to latency introduced by encryption is acceptable, then recommended and preferred port capacity would be 210 ports—only a 12.5% reduction from the peak capacity of 240 ports offered by non-encryption. If the audio quality strictly applies, then the recommended port capacity can be as low as 120 ports. Peak port capacity could be the same 270 ports as non-encryption, if the delay is acceptable.

The table below lists system disk IOPS:

Figure 60: IOPS on physical server of single hex core, MP3 only, 16 Kbps, encryption

Ports	Physical Server Disk IOPS		
Total	Reads	Writes	
60	14.66	0.036	14.62
120	24.00	0.041	23.95
180	33.42	0.029	33.39
210	37.65	0.030	37.62
240	42.21	0.029	42.18
270	47.18	0.036	47.14
300	51.44	0.011	51.43
330	55.81	0.006	55.81
360	60.99	0.002	60.99
390	67.12	0.003	67.11

The graph below compares system disk IOPS on a physical server IOPS with non-encryption:

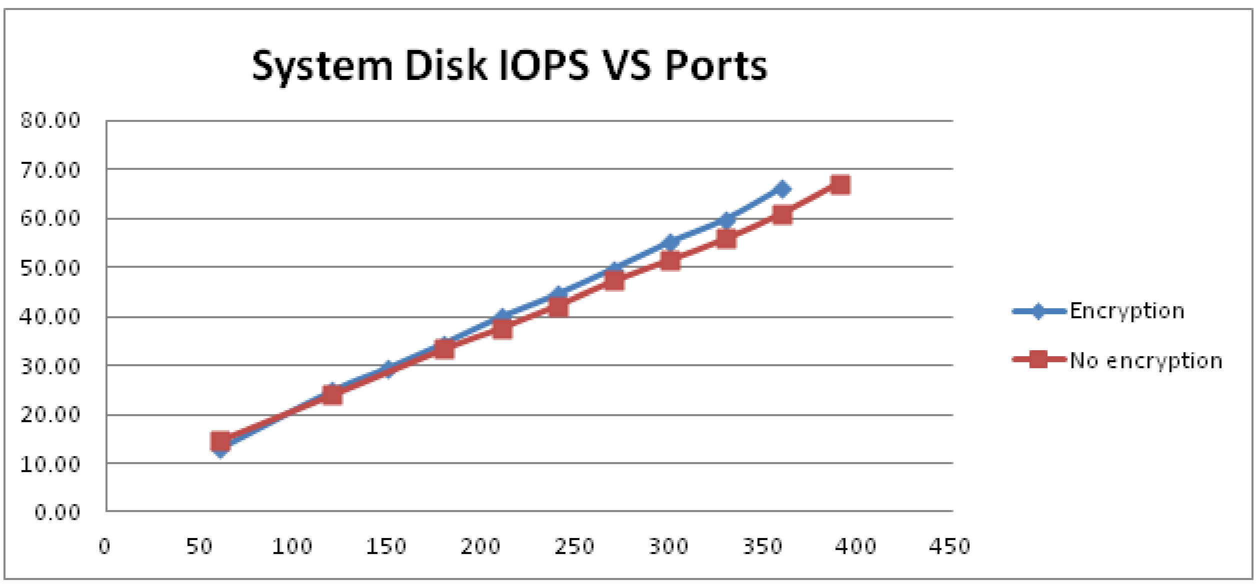


Figure 61: Comparison of System Disk IOPS on Single Hex Core Physical Server, MP3 16kbps encryption vs non-encryption

System disk IOPS is nearly the same for encryption and non-encryption; both increase slightly at a higher port capacity. Some of that can be attributed by other disk IO operations, such as encryption key files.

The table below lists MCP IOPS:

Figure 62: MCP IOPS on physical server of single hex core, MP3 only, 16 Kbps, encryption

Ports	Physical Server MCP IOPS		
Total	Reads	Writes	
60	16.53	8.88	7.65
120	32.59	17.69	14.91
150	40.40	21.96	18.44
180	48.46	26.46	22.01
210	56.35	30.83	25.52
240	64.32	35.24	29.08
270	72.28	39.64	32.64
300	80.06	43.95	36.11
330	88.61	48.53	40.07
360	100.48	52.91	47.57

The graph below compares total MCP IOPS between encryption and non-encryption:

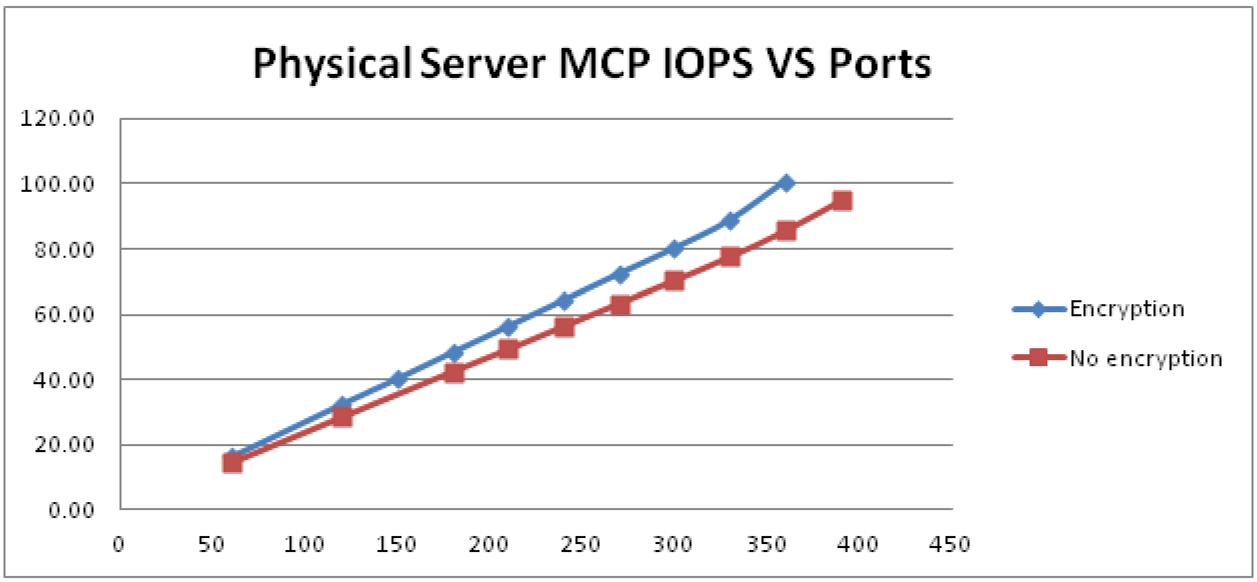


Figure 63: MCP IOPS on Single Hex Core Physical Server, MP3 16Kbps encryption vs non-encryption

MCP IOPS for encryption increases when port capacity increases. As seen in Figure: Comparison of System Disk IOPS on Single Hex Core Physical Server, MP3 16kbps encryption vs non-encryption, increase for disk IOPS is much smaller for encryption, so here the increase should be attributed to network IOs.

VMs on Dual Hex Cores Server

The testing for MP3 16kbps with encryption was conducted on the VM Profile 4 based on Hardware Profile 4 of a dual hex cores server, same as non-encryption in the 16knps tests VMs on Dual Hex Cores Server. Six VMs were configured while only one MCP was installed on each Windows VM. Below are three graphs comparing overall CPU usage, audio quality related max jitter and max delta for MP3 16kbps encryption vs non-encryption:

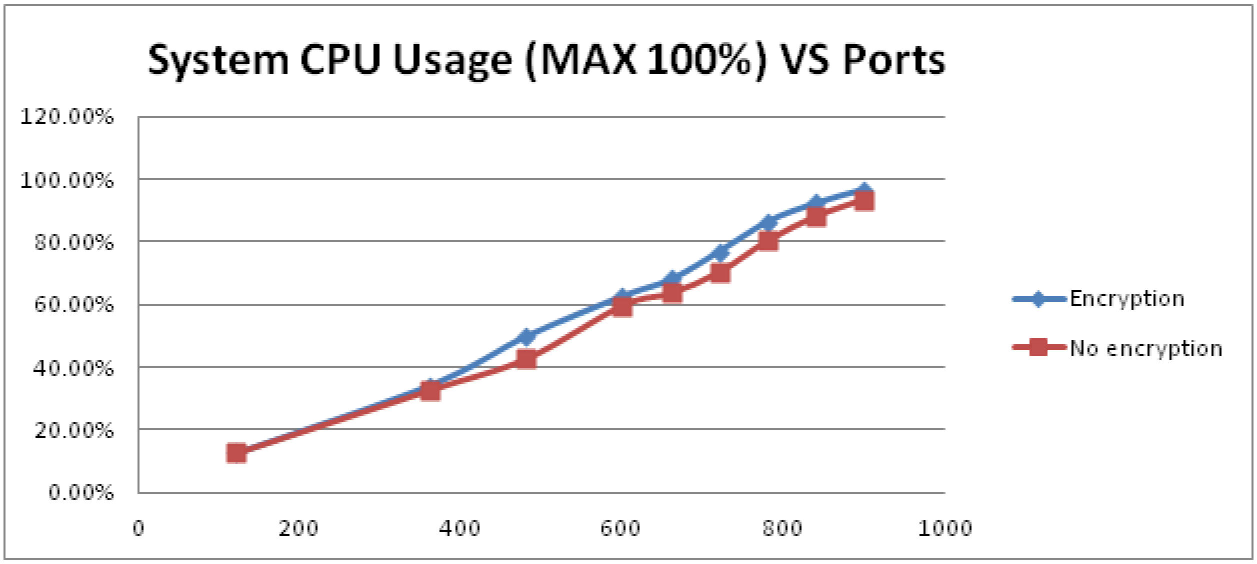


Figure 64: Comparison of Overall VMs CPU Usage of MP3 16kbps encryption vs non-encryption

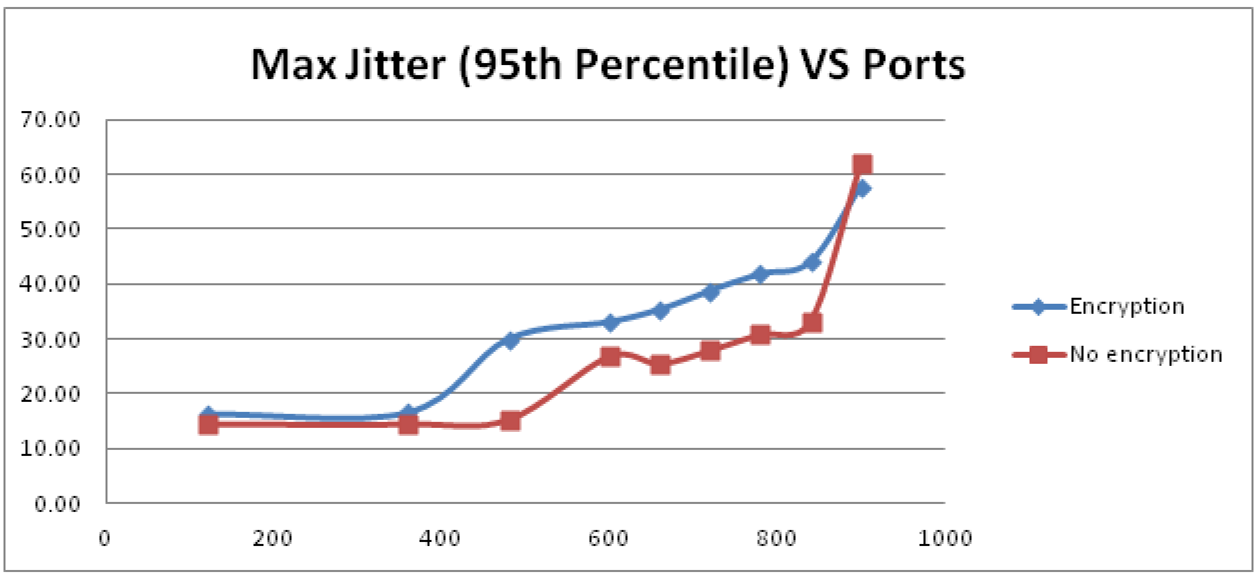


Figure 65: Comparison of Overall VMs Max Jitter of MP3 16kbps encryption vs non-encryption

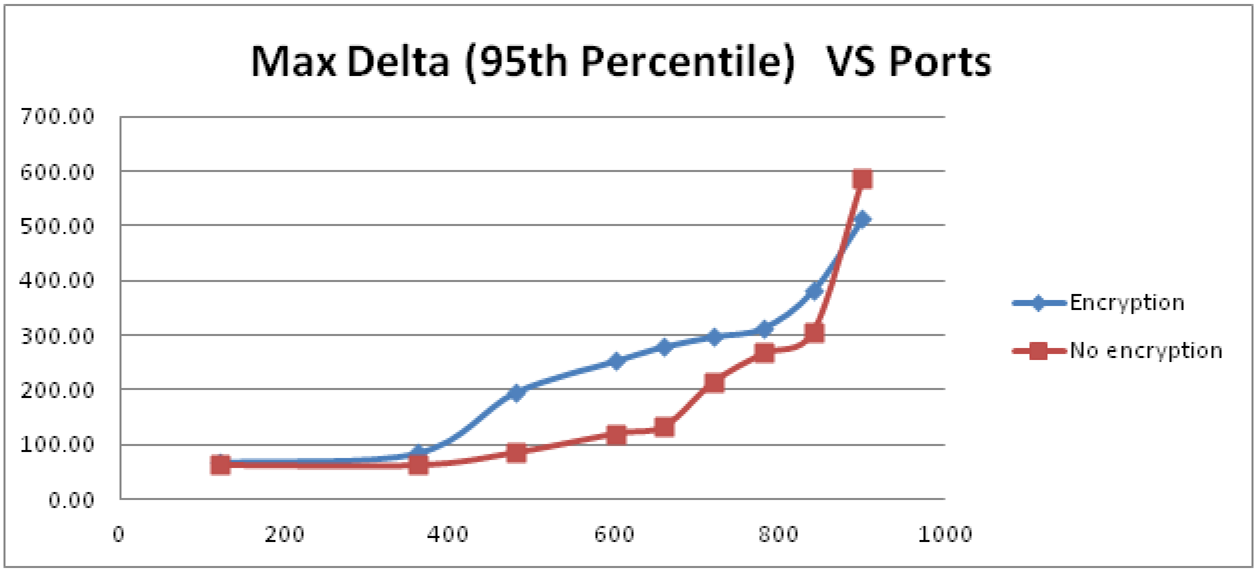


Figure 66: Comparison of Overall VMs Max Delta of MP3 16kbps encryption vs non-encryption

The VM environment exhibits a similar trend: slightly overall CPU usage for the encryption profile, and much higher for max jitter and max delta. Applying the same criteria from the physical server results, if a slightly higher delay (due to latency introduced by encryption) is acceptable, then the recommended and preferred port capacity could be 600 ports—only a 16.7% reduction of the peak 720 ports with non-encryption. If audio quality strictly applies, the recommended ports can be as low as 480 ports. And if some delay is acceptable, then the peak port capacity can be the same 840 ports as non-encryption.

The overall system disk IOPS for all 6 VMs is listed below:

Figure 67: Overall Disk IOPS on all 6 VMs of dual hex cores, MP3 only, 16 Kbps, encryption

Ports	Overall 6 VMs Disk IOPS			SSD Drive Disk IOPS		
Total	Reads	Writes	Total	Reads	Writes	
120	28.70	0.004	28.69	21.881	0.000	21.881
360	67.46	0.004	67.46	56.238	0.000	56.238
480	87.56	0.026	87.54	74.903	0.000	74.903
600	108.01	0.015	107.99	93.647	0.000	93.647
660	119.49	0.005	119.48	104.304	0.000	104.304
720	128.76	0.020	128.74	114.441	0.000	114.441
780	137.68	0.015	137.66	123.210	0.002	123.209
840	146.99	0.009	146.98	132.646	0.002	132.644
900	154.68	0.025	154.66	140.145	0.002	140.143

The graph below compares system disk IOPS with encryption and with non-encryption, on the same VM environment:

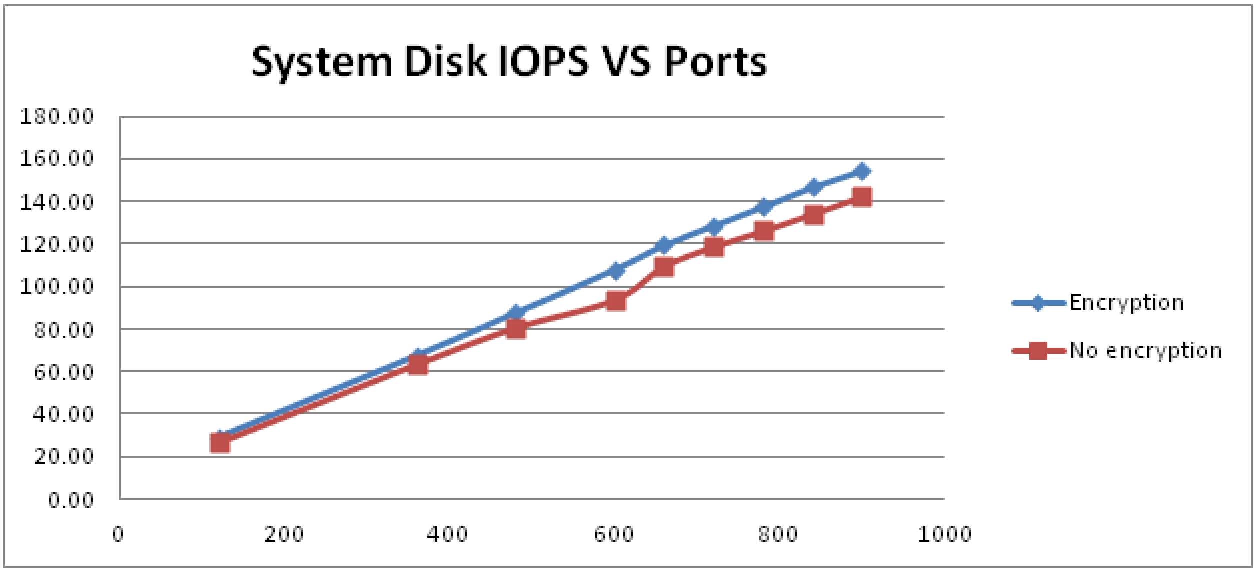


Figure 68: Comparison of Overall 6 VMs Disk IOPS MP3 16 kbps encryption vs non-encryption

As with the physical server tests, encryption increases as port capacity increases. Also as with the physical server tests, some of that can be attributed to extra disk IO operations.

The table below lists Data throughputs for encryption:

Figure 69: Data Throughputs for MP3 only, 16 kbps, encryption

Ports	Overall Disk KB/sec			SSD Drive Disk KB/sec		
Total	Reads	Writes	Total	Reads	Writes	
120	387.99	0.02	387.97	304.229	0.000	304.229
360	1096.82	22.54	1074.28	876.599	0.000	876.599
480	1344.60	107.95	1236.65	1191.403	0.006	1191.397
600	2187.50	348.40	1839.09	1532.171	0.000	1532.171
660	2024.16	35.09	1989.07	1652.232	0.000	1652.232
720	1955.33	99.81	1855.51	1803.207	0.006	1803.201
780	2572.79	205.15	2367.64	1982.733	0.024	1982.709
840	2534.97	28.65	2506.32	2097.871	0.043	2097.829
900	2851.85	119.47	2732.38	2297.264	0.007	2297.257

Using Formula 1...

```
MP3 bitrate * Ports / 8 = KB/sec
Or 16kbps * Ports / 8 =KB/sec if MP3 is 16kbps
```

...take two samples (120 & 720) from the above table above, and apply them to Formula 1:

```
16 kpbs * 120 / 8 = 240 kb close to 304 in the table (in SSD)
```

```
16 kpbs * 720 / 8 = 14400 kb close to 1803 in the table (in SSD)
```

The measurements from real testing for MP3 16kbps encryption are slightly higher than these calculations predict, due to other file, such as metadata and JSON files, being saved on the same cache folder.

The graph below compares overall data throughputs with no encryption:

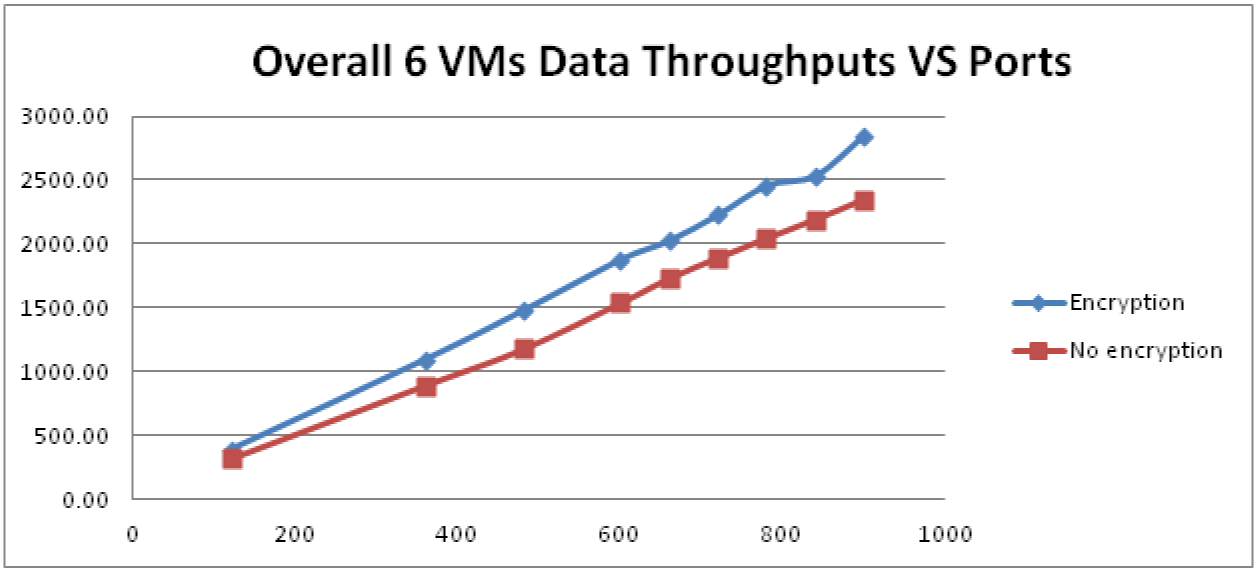


Figure 70: Comparison of Overall 6 VMs Data Throughputs MP3 16 kbps encryption vs non-encryption

The data throughputs for encryption increase slightly when port capacity increases, matching a similar trend with system disk IOPS.

The table below lists overall MCP IOPS from all 6 VMs:

Figure 71: Overall MCP IOPS from 6 VMs of dual hex core, MP3 only, 16kbps, encryption

Ports	Overall 6 VMs MCP IOPS		
Total	Reads	Writes	
120	34.874	17.638	17.236
360	102.624	52.900	49.724
480	130.285	70.377	59.909
600	168.849	87.766	81.083
660	186.175	96.882	89.293
720	193.248	105.171	88.077
780	219.395	114.398	104.997
840	235.730	123.009	112.720
900	252.198	131.682	120.516

The graph below compares performance of the same configuration, except with non-encryption:

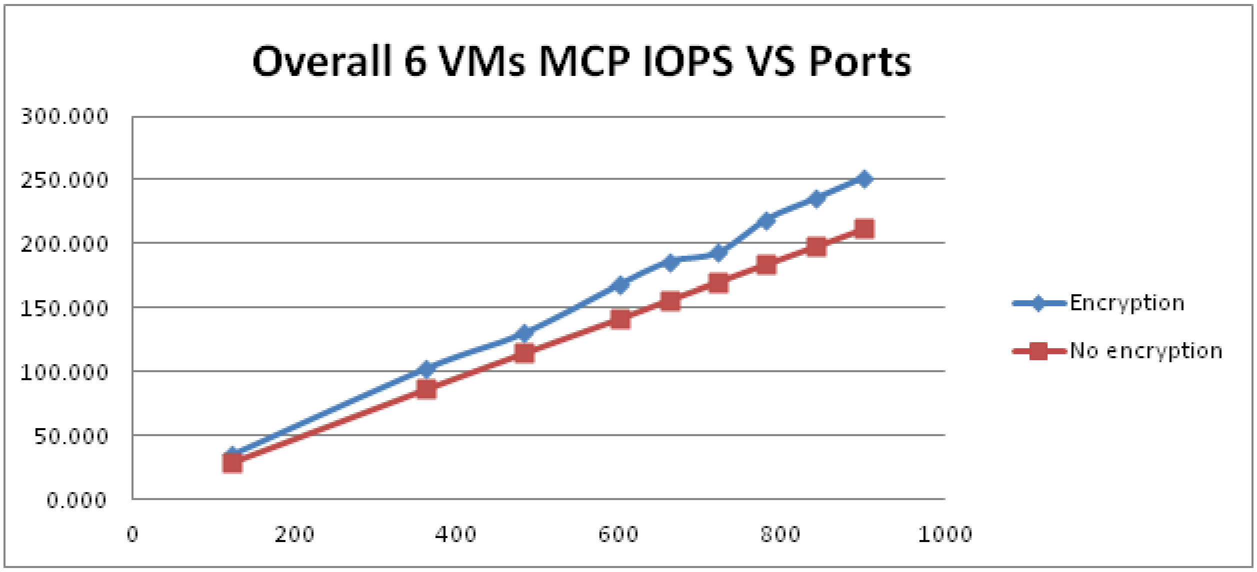


Figure 72: MCP IOPS from 6 VMs of dual hex core, MP3 only, 16kbps, encryption vs non-encryption

MCP IOPS performance is affected slightly by encryption, similar to the trend expressed in the physical server results.

Detailed Studies of GVP Media Server Behavior on Linux

Because MP3 16kbps produces better results than 32kbps on Windows, Linux tests focused on MP3 16kpbs testing profiles. Also based on previous results on Windows, we chose two typical Hardware Profiles for Linux testing: Hardware Profile 1 for physical server testing, and Hardware Profile 4 with Virtual Machine Profile 4 for virtual machine environment testing.

Parameter Adjustments

These adjustments achieve higher port capacity:

Parameter	Default Value	Adjusted Value
`mpc.recordnumparallelpost`	30	300
`mpc.recordpostretrybackoff`	120000	15000
`mpc.recordpostretrycount`	3	1
`mpc.mediamgr.recordwritetimeinterval`	1000	10000
`fm.http_proxy`	(empty)	(squid bypassed)

Comparisons with Windows

Physical Server on Single Hex Core

These tests use Software Profile 1a on Hardware Profile 1 for a physical server. Here are three graphs illustrating overall system CPU usage, MCP CPU usage and memory usage:

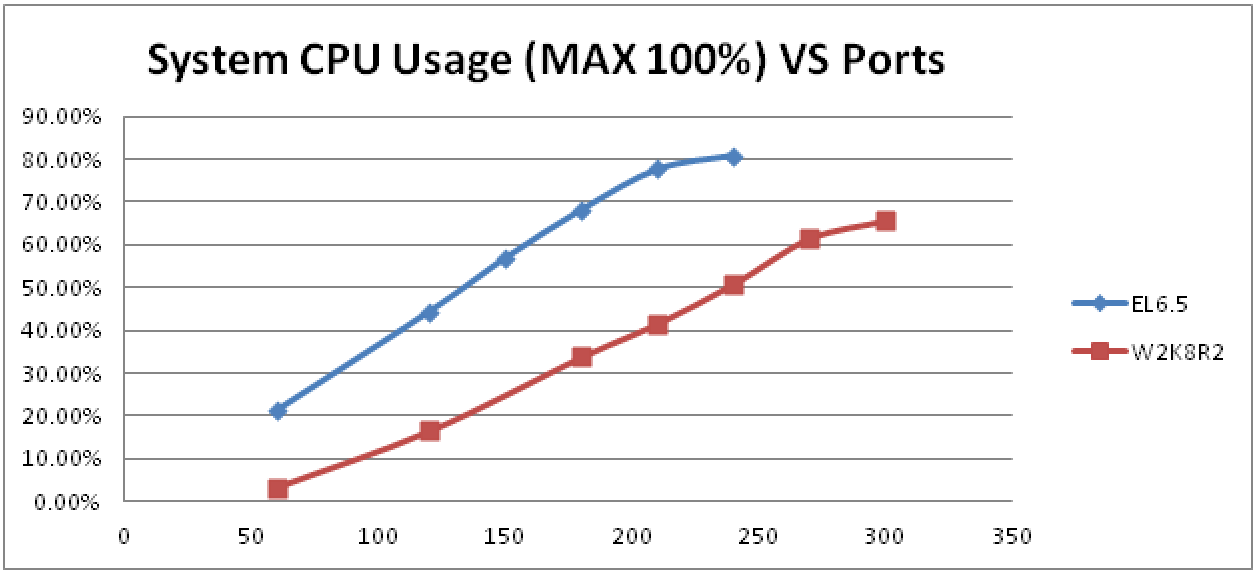


Figure 73: Comparison of System CPU Usage on a physical server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

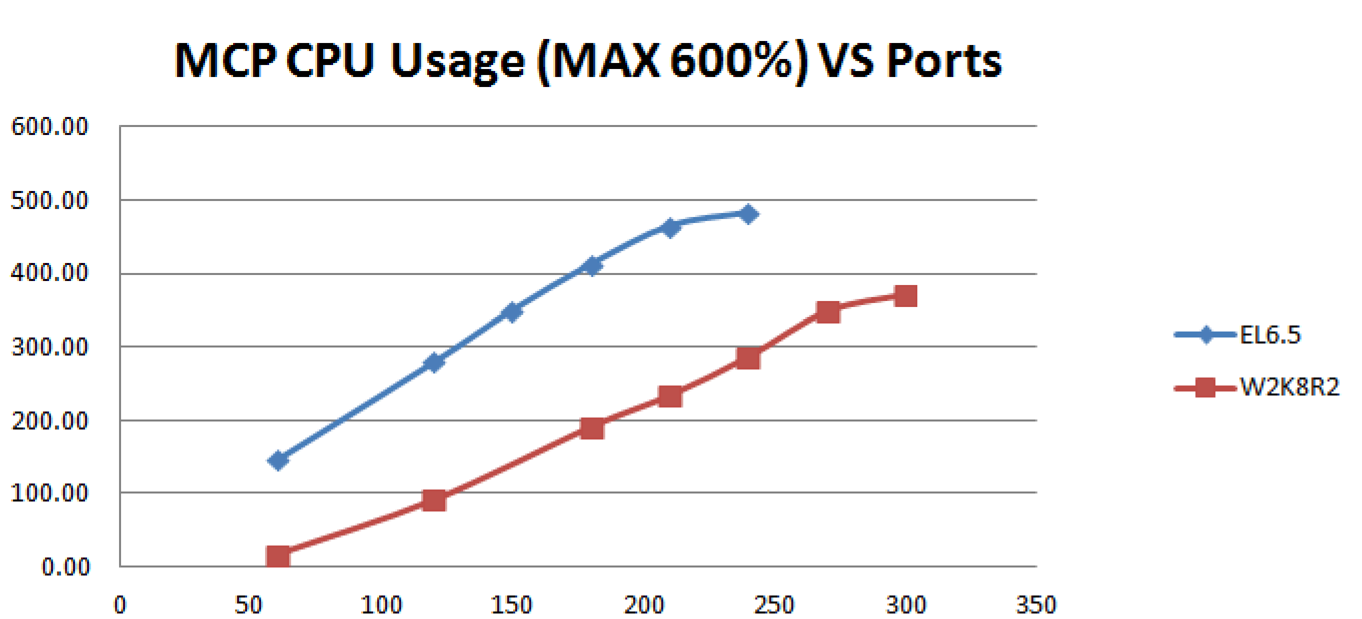


Figure 74: Comparison of MCP CPU Usage on a physical server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

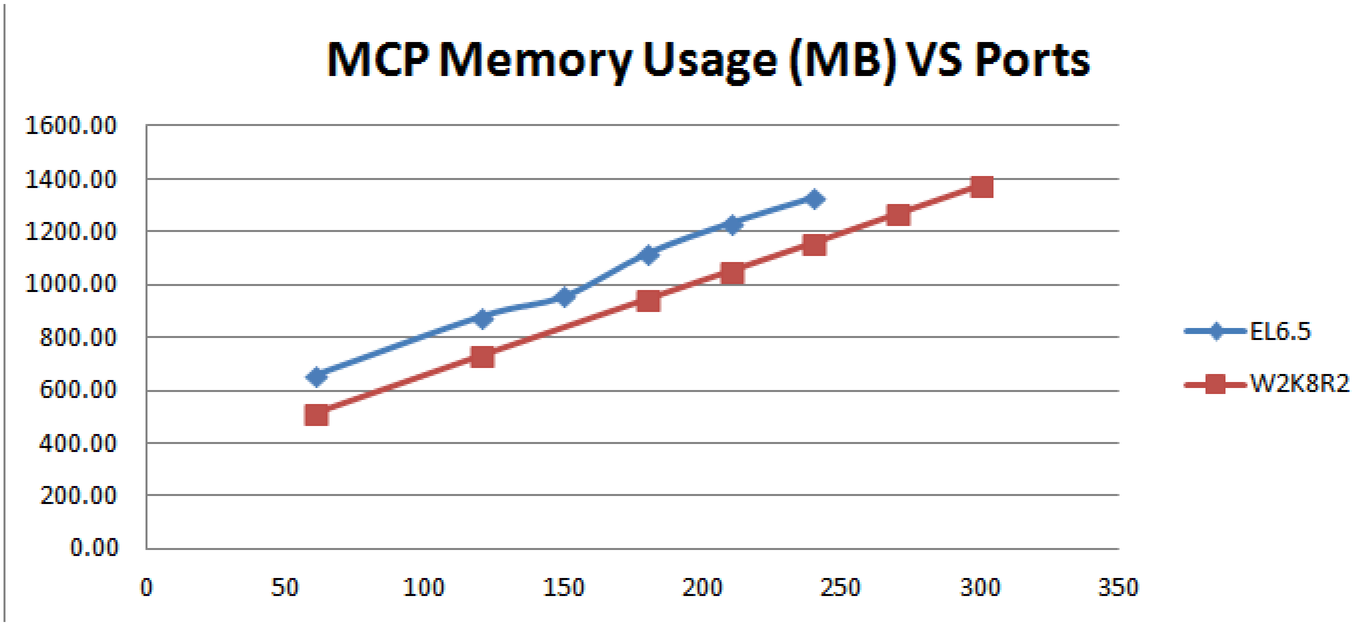


Figure 75: Comparison of MCP Memory Usage on a physical server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

Linux uses more resources (CPU, memory etc) than Windows, and so lower capacity is achieved on Linux with a 37.5% reduction (150 vs. 240) for preferred ports and a 22.2% reduction (210 vs. 270) for peak ports.

The two graphs below compare audio quality in terms of max jitter and max delta:

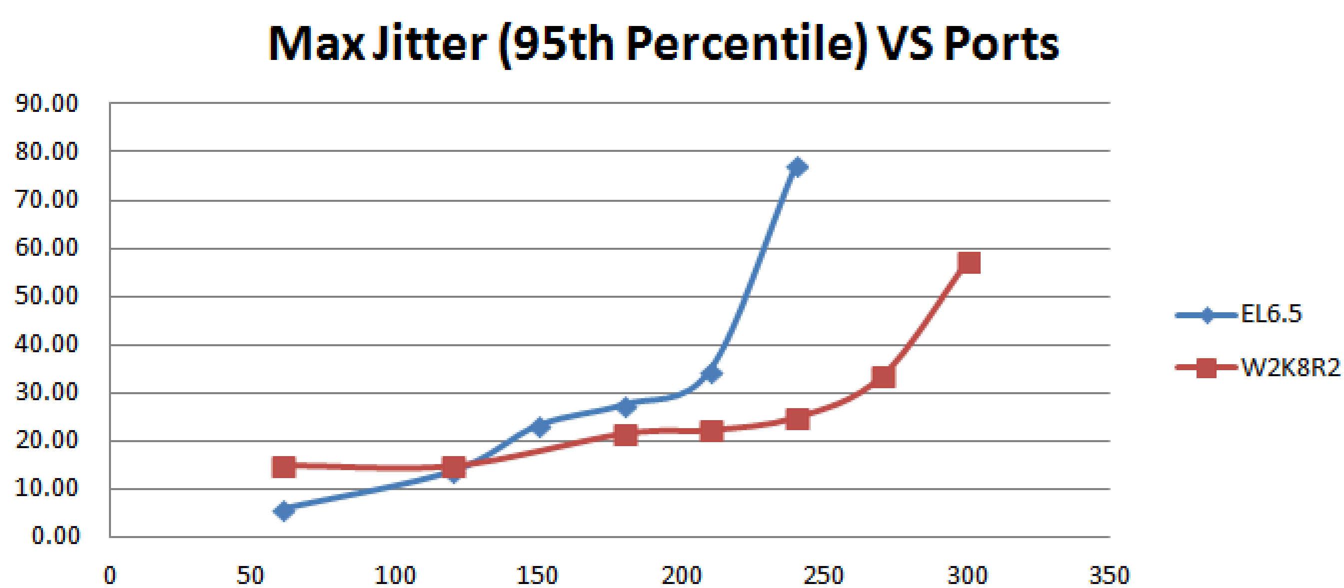


Figure 76: Comparison of Max Jitter on A physical server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

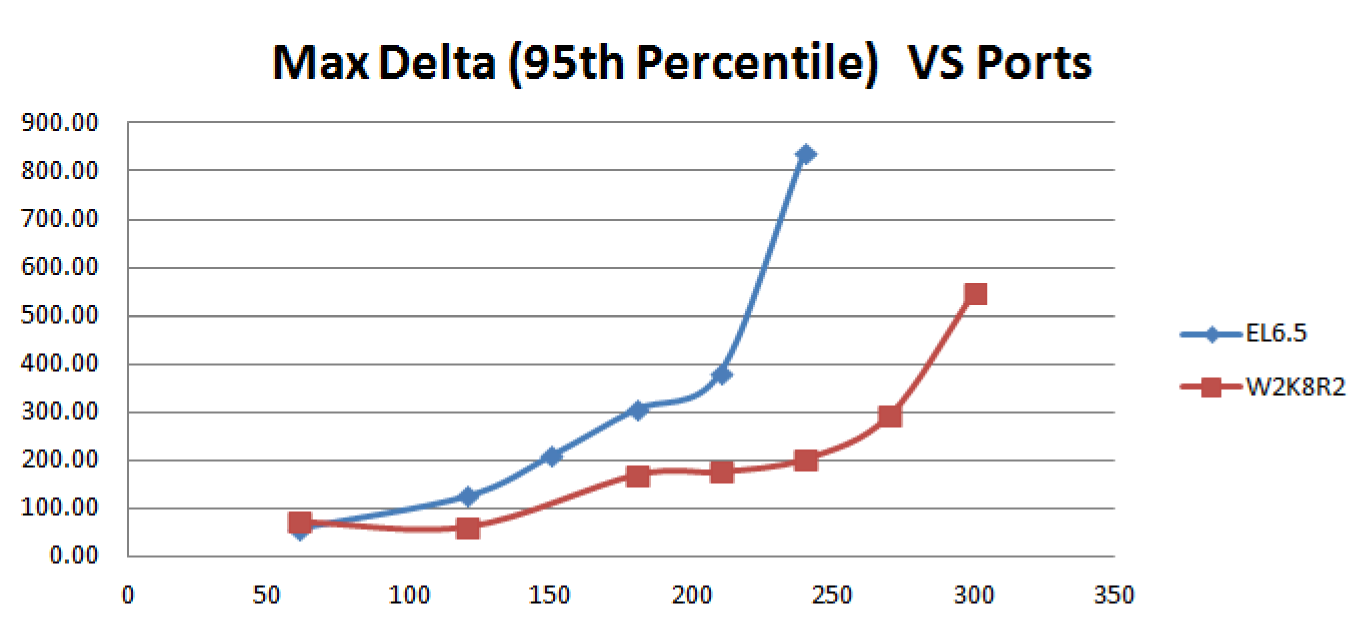


Figure 77: Comparison of Max Delta on A physical server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

Note that Max Jitter is somewhat close between Windows and Linux. But Linux has a lower value at lower ports and a slightly higher value on higher ports. The Max Delta shows that Linux has the higher value even though it is nearly the same for both Windows and Linux at lower ports.

System disk IOPS is illustrated in this table for Linux EL 6.5:

Figure 78: System Disk IOPS on a physical server of single hex core on EL 6.5, MP3 only 16 kbps

Total	Reads	Writes		
Ports	Disk IOPS Physical Server			
	60	12.75	0.000	12.754
120	23.12	0.000	23.117	
150	27.65	0.000	27.645	
180	32.15	0.000	32.150	
210	36.73	0.000	36.729	
240	41.57	0.000	41.568	

The graph below compares System Disk IOPS performance on Linux and Windows physical servers:

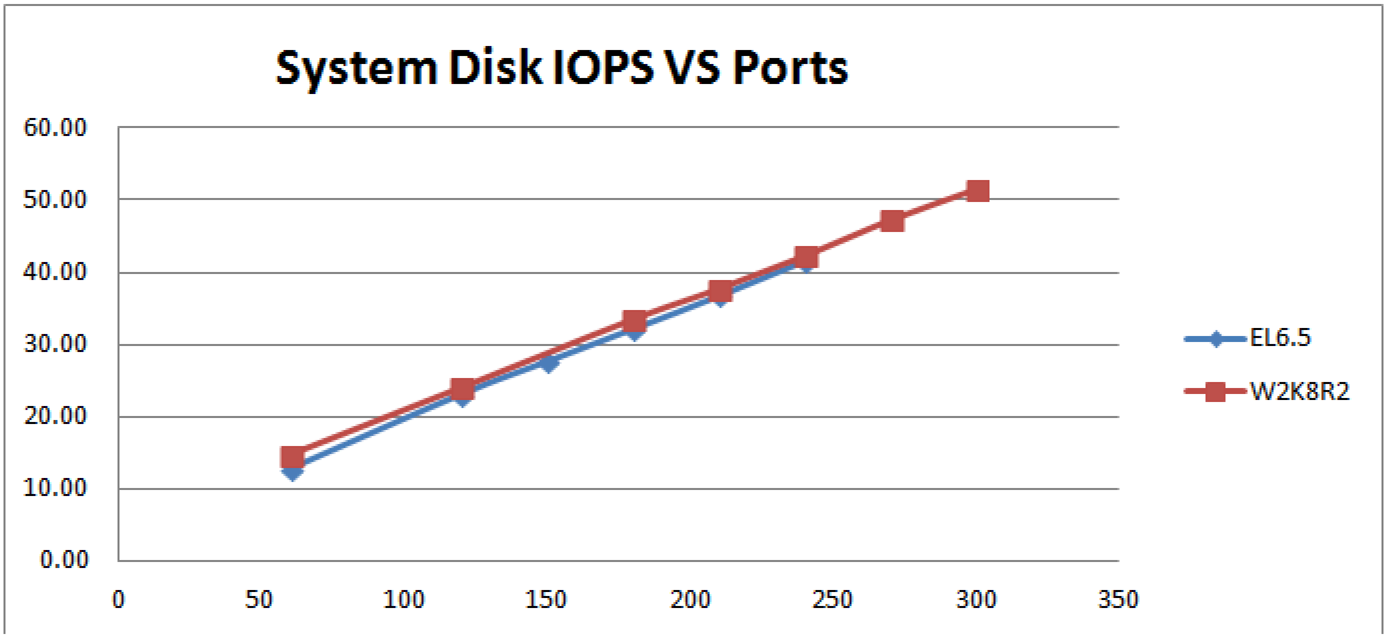


Figure 79: Comparison of System Disk IOPS on a Physical Server, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

Note that IOPS on both Windows and Linux are similar; and so Disk IOPS is related to the test profile, and irrelevant to a particular OS. So the IOPS numbers from the previous Windows testing can be used generally for both Windows and Linux.

VMs on Dual Hex Cores Server

These tests use SW Profile 1a on HW Profile 1 with VM Profile 4 for virtual machine environment testing. Below are three graphs illustrating overall system CPU usage, MCP CPU usage and memory usage:

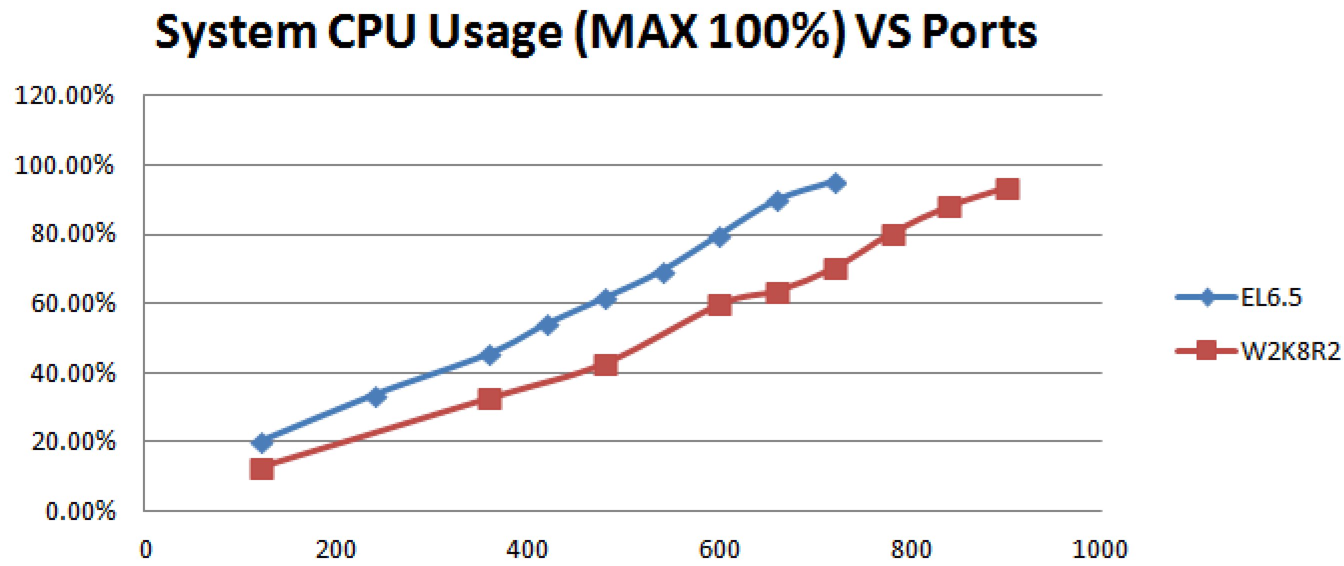


Figure 80: Comparison of System CPU Usage on VM env, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

You can observe the same trend as with the physical server results in the previous section Linux consumes more CPU resources. Below are two graphs of audio quality-related metrics that show the same thing.

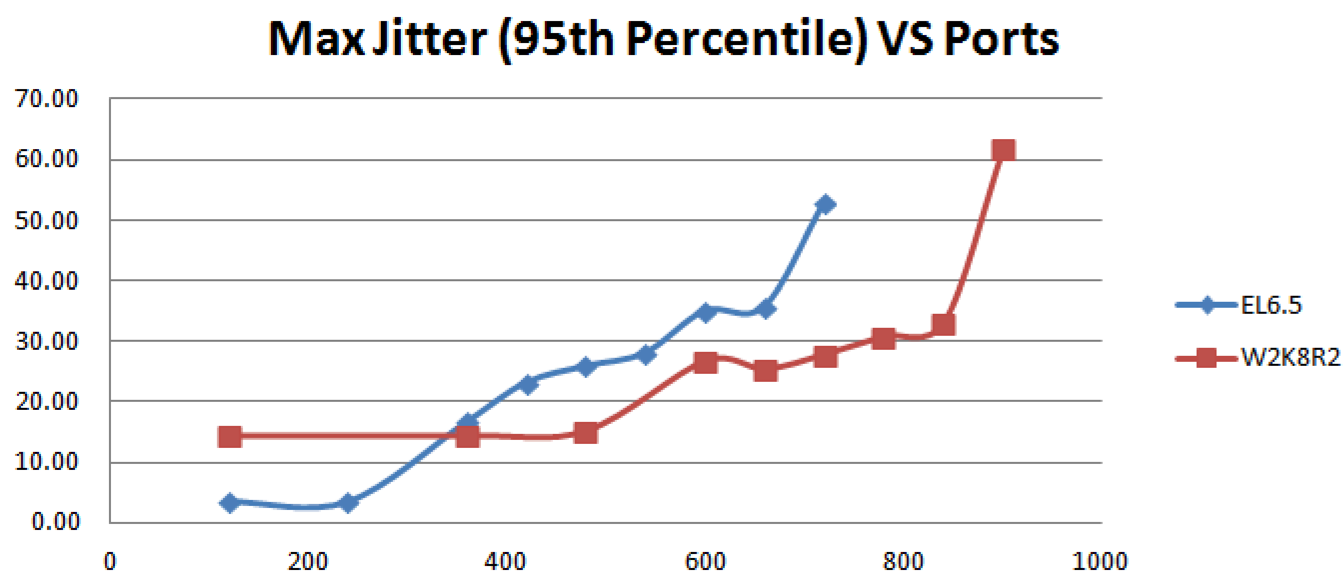


Figure 81: Comparison of Max Jitter on VM env, MP3 16kbps without encryption, RH EL 6.5 vs. Win 2008 R2

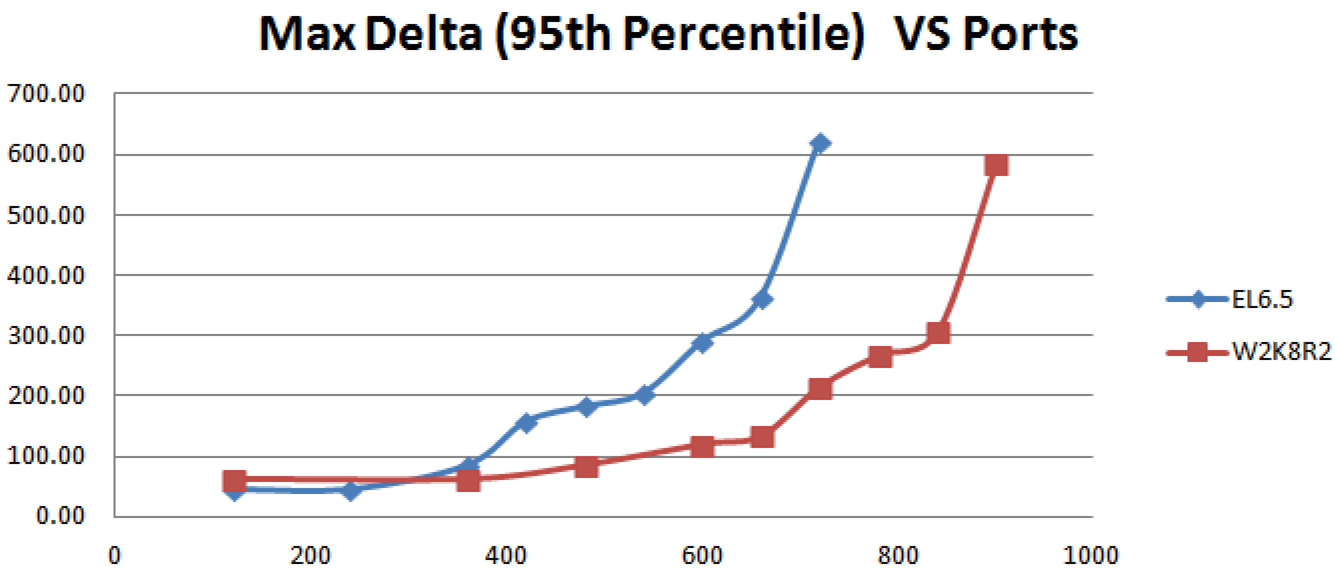


Figure 82: Comparison of Max Delta on VM env, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

As observed on the above graphs, 540 ports are recommended and preferred. This is a 25% reduction compared with Windows 2008 R2 (540 vs. 720). Peak capacity would be 660 ports, which is a 21.4% reduction compared to Windows 2008 R2 (660 vs. 840). Similar reductions were also observed on physical server tests in the previous section.

The disk IOPS is displayed here:

Figure 83: Disk IOPS from overall 6 VMs of dual hex core, MP3 only 16 kbps, on EL 6.5

Ports	Overall 6 VMs Disk IOPS			SSD Drive Disk IOPS		
	Total	Reads	Writes	Total	Reads	Writes
120	28.17	0.00	28.17	24.011	0.000	24.011
240	49.78	0.00	49.78	44.590	0.000	44.590
360	71.11	0.00	71.11	65.747	0.000	65.747
420	81.59	0.00	81.59	76.058	0.000	76.058
480	92.37	0.00	92.37	86.767	0.000	86.767
540	102.96	0.00	102.96	97.305	0.000	97.305
600	112.33	0.00	112.33	106.727	0.000	106.727
660	122.06	0.00	122.06	116.440	0.000	116.440
720	130.82	0.00	130.82	125.121	0.000	125.121

The graph below compares the above table with the corresponding table for Windows, for the same MP3-only 16 kbps profile:

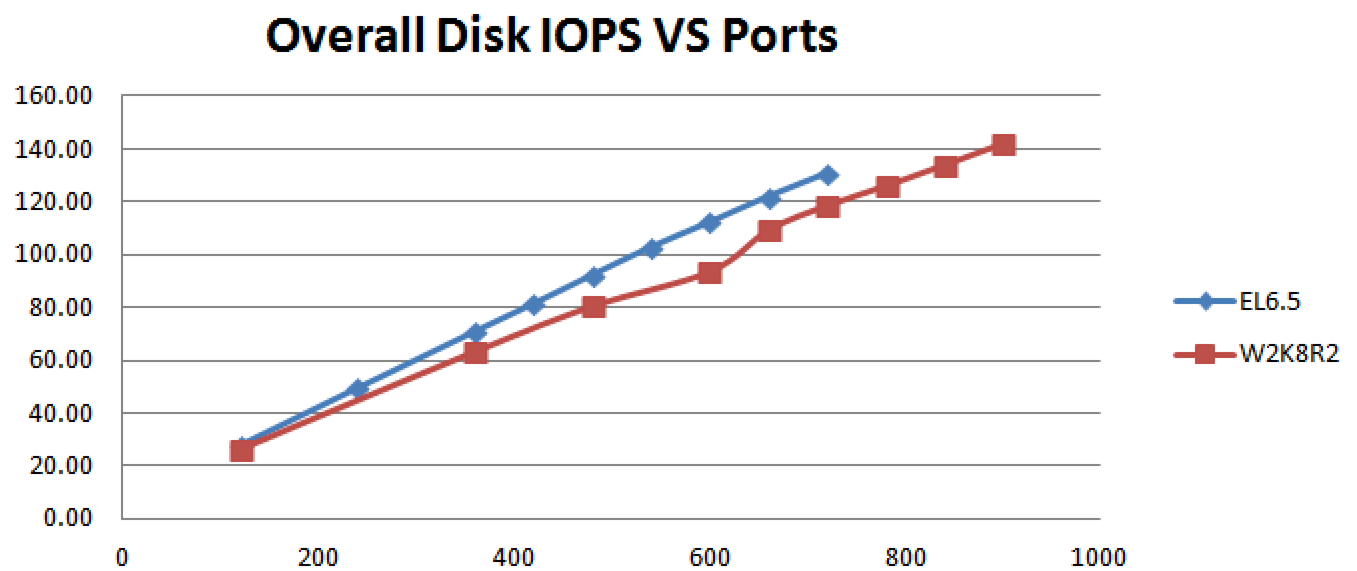


Figure 84: Comparison of System Disk IOPS on VM env, MP3 16kbps without encryption, RH EL 6.5 vs. Windows 2008 R2

Note that disk IOPS results for Linux and Windows are very close, and corresponds to the results on a physical server in the previous section.

The data throughput for this MP3-only profile on EL 6.5 is illustrated below:

Figure 85: Data throughputs from overall 6 VMs of dual hex core, MP3 only 16 kbps, on EL 6.5

Total KB/sec	Read KB/sec	Write KB/sec	Total KB/sec	Read KB/sec	Write KB/sec	
Ports	Overall Disk KB/sec			SSD Drive Disk KB/sec		
120	417.70	0.00	417.70	389.474	0.000	389.474
240	788.58	0.00	788.58	751.418	0.000	751.418
360	1145.77	0.00	1145.77	1104.237	0.000	1104.237
420	1317.38	0.00	1317.38	1274.484	0.000	1274.484
480	1496.20	0.00	1496.20	1451.114	0.000	1451.114
540	1677.83	0.00	1677.83	1627.798	0.000	1627.798
600	1843.65	0.00	1843.65	1795.706	0.000	1795.706
660	2023.36	0.00	2023.36	1974.070	0.000	1974.070
720	2193.62	0.00	2193.62	2142.769	0.000	2142.769

Comparing MP3 only and MP3 + WAV

Physical Server on Single Hex Core

This test uses SW Profile 2a (MP3 16 kbps + wav without encryption) on HW Profile 1 for a physical server, compared which SW Profile 1a (MP3 only 16 kbps without encryption) is used as a baseline for comparison. Below are three graphs illustrating overall system CPU usage, MCP CPU usage and memory usage:

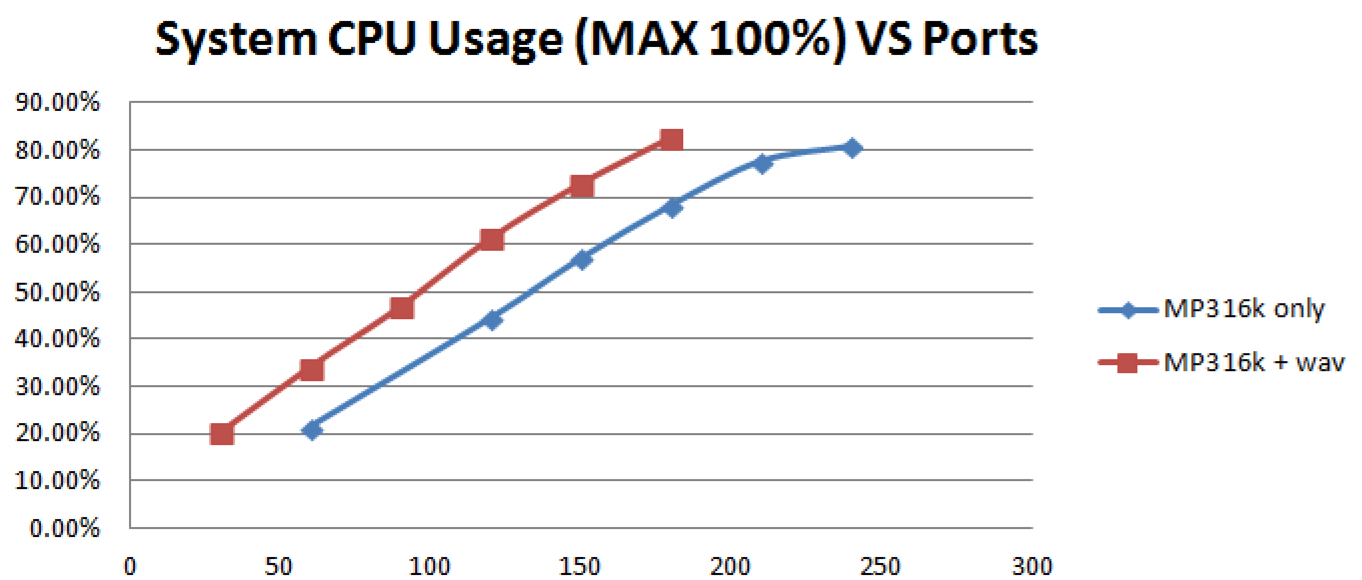


Figure 86: Comparison of System CPU Usage on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

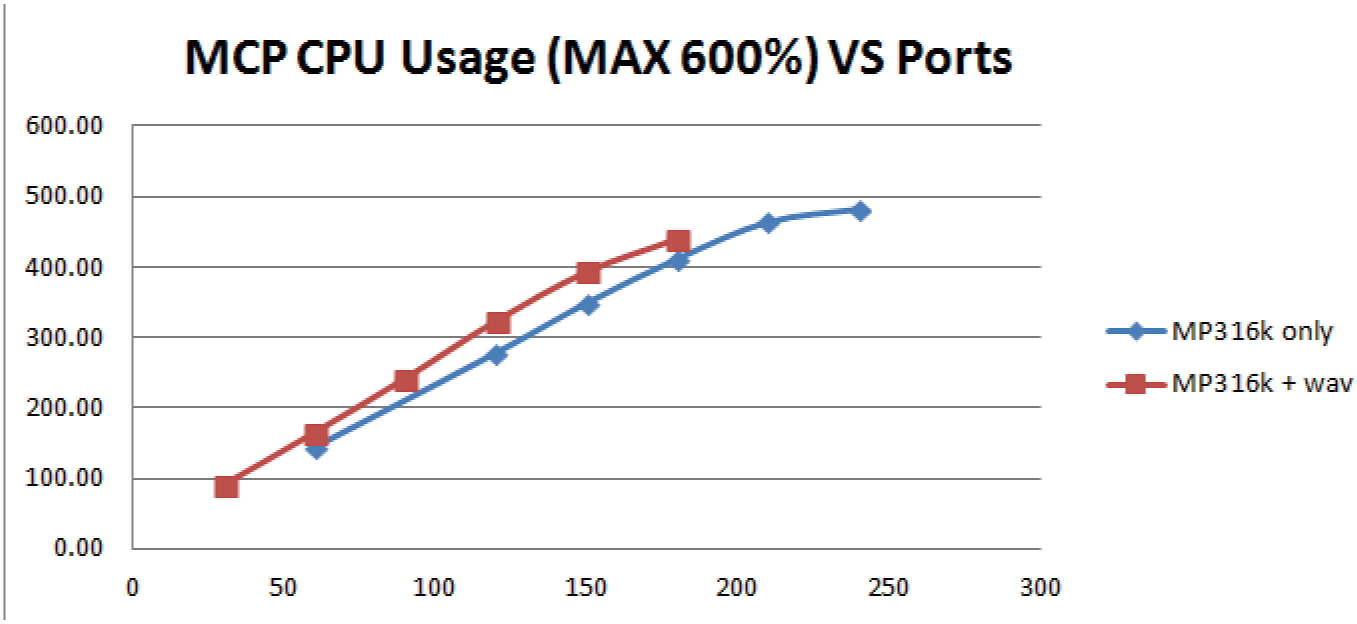


Figure 87: Comparison of MCP CPU Usage on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

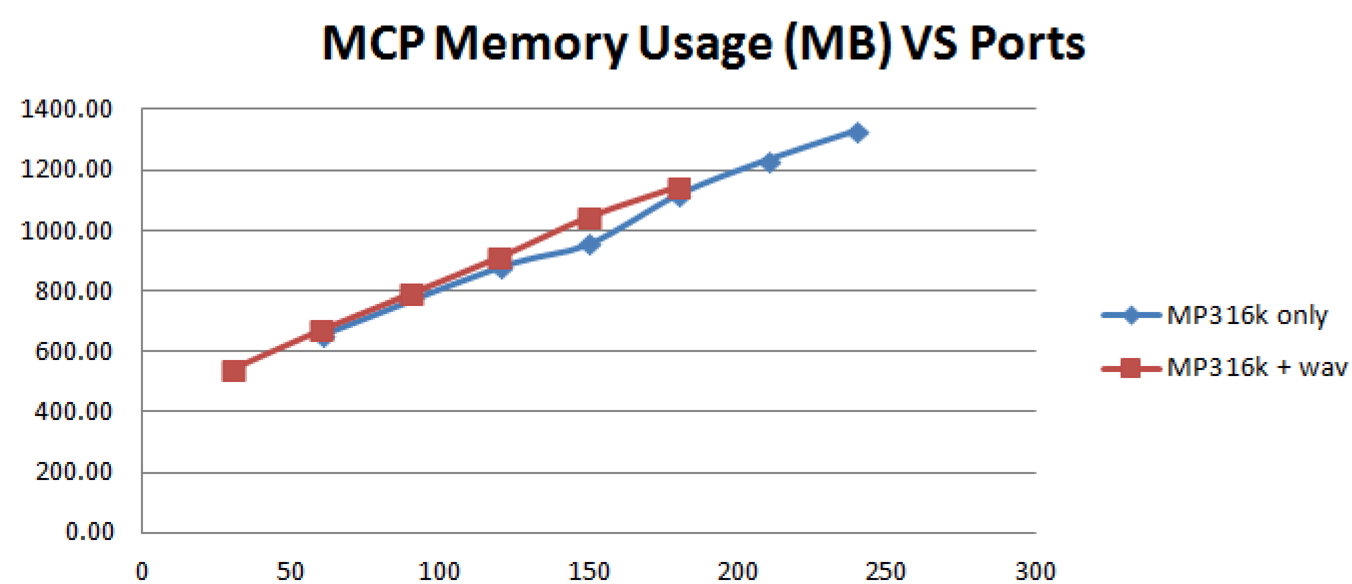


Figure 88: Comparison of MCP Memory Usage on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

The comparison shows apparent higher MCP usage and overall system CPU usage for the MP3 + wav profile. However, the MCP memory usage is not significantly higher.

The audio quality metric also shows some differences, below:

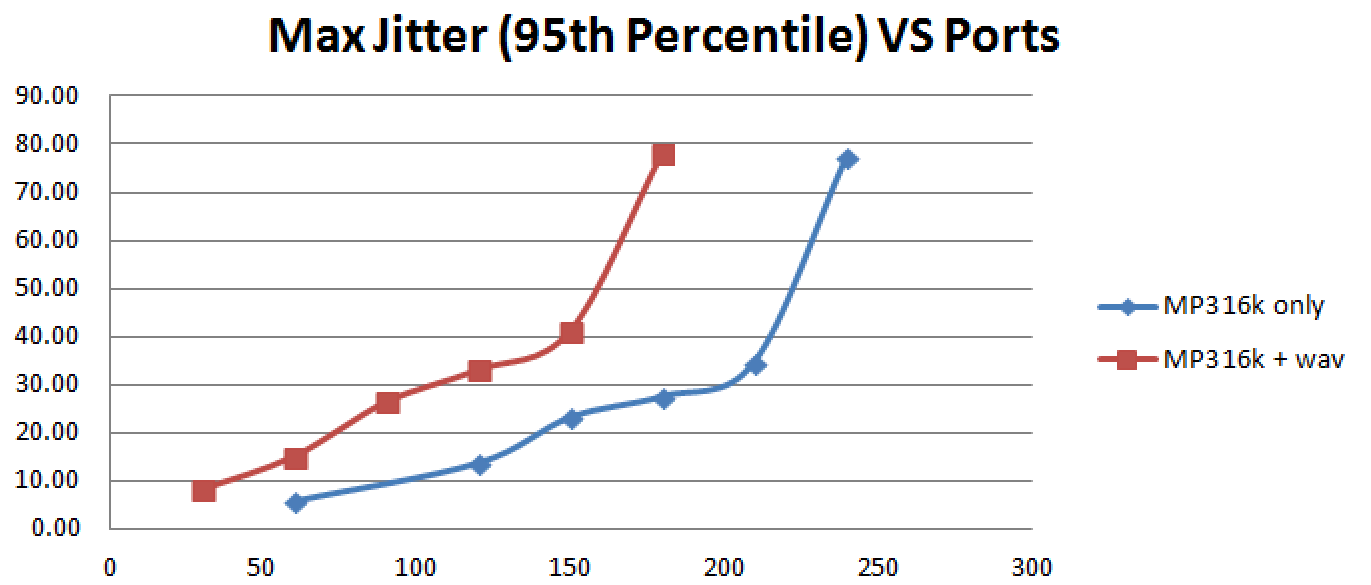


Figure 89: Comparison of Max Jitter on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

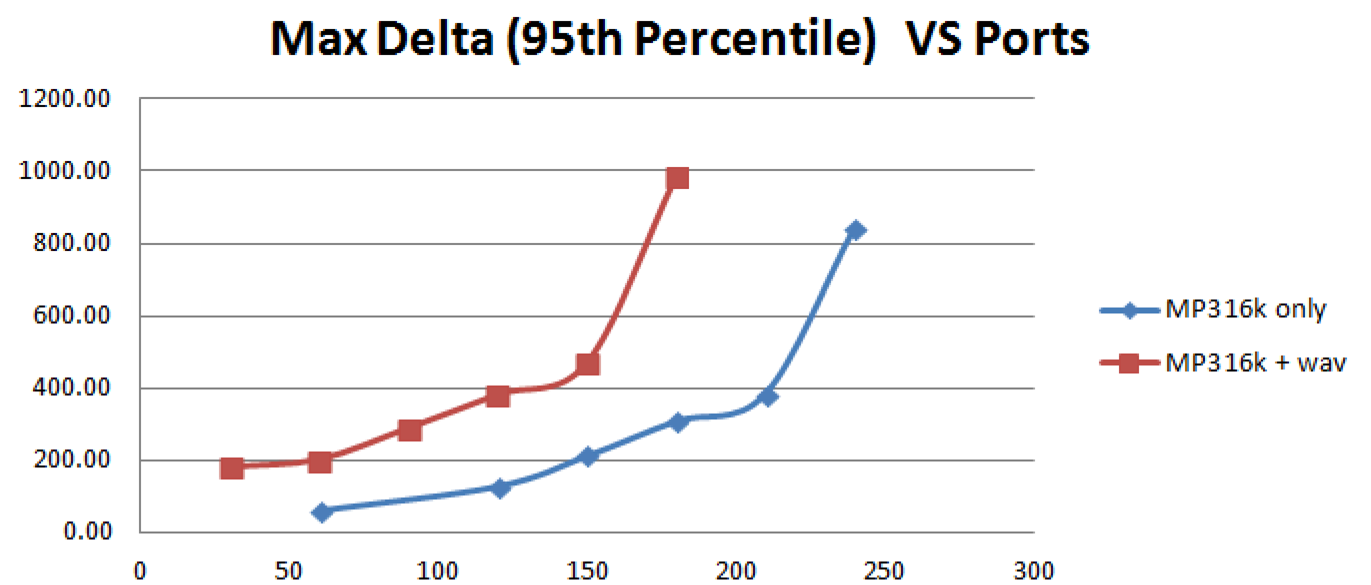


Figure 90: Comparison of Max Delta on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

Note that lower capacities would be achieved for the MP3 + WAV profile. 90 ports would be recommended and preferred, a 40% reduction (90 vs. 150) compared with the MP3-only profile, while 150 ports would be peak capacity a 28.6% reduction (150 vs. 210).

System disk IOPS is listed in the following table:

Figure 91: System Disk IOPS on a physical server of single hex core on EL 6.5, MP3 16 kbps + wav

Ports	Physical Server Disk IOPS		
	Total	Reads	Writes
30	15.18	0.008	15.17
60	26.70	0.000	26.70
90	35.53	0.003	35.53
120	46.04	0.002	46.04
150	55.44	0.000	55.44
180	65.50	1.520	63.98

The graph below compares disk IOPS with the MP3-only profile:

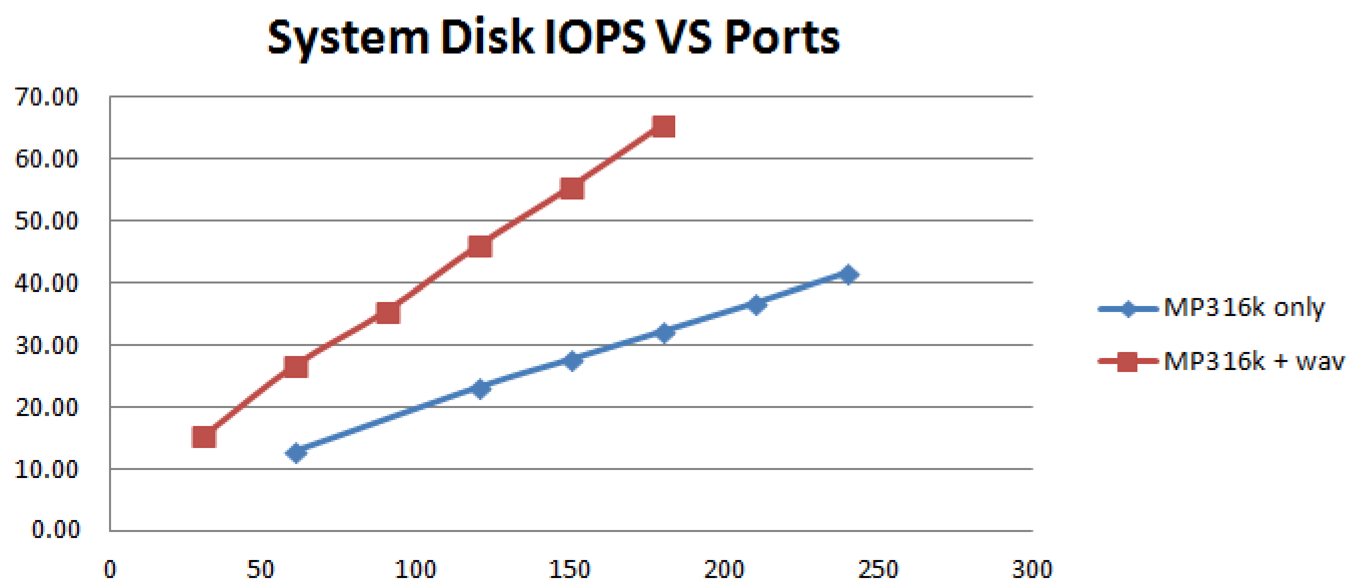


Figure 92: Comparison of System Disk IOPS on a Physical Server, MP3 + wav vs. MP3 only, on RH EL 6.5

The MP3-only profile is almost double the disk IOPS for MP3 + wav profile, as observed in the Windows testing.

VMs on Dual Hex Cores Server

A similar trend of overall CPU usage occurs in the Virtual Machine environment.

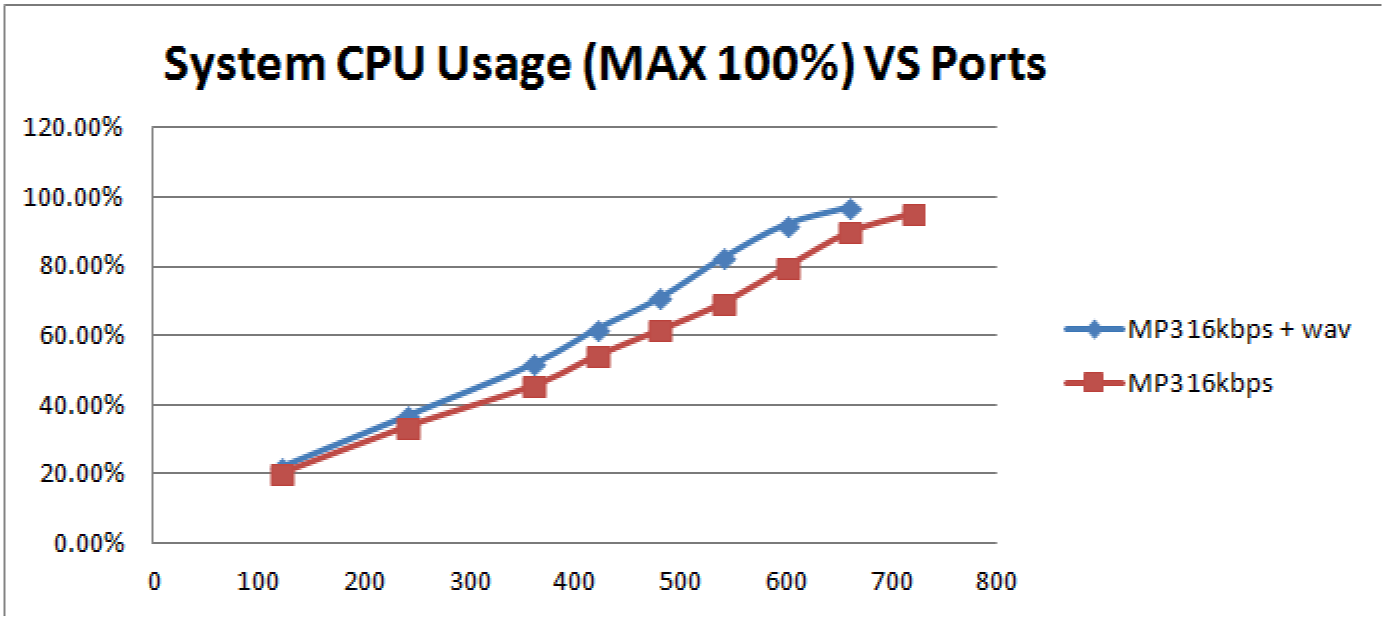


Figure 93: Comparison of System CPU Usage on VM env, MP3 + wav vs. MP3 only, on RH EL 6.5

The audio quality metrics shows similar trends as on a physical server.

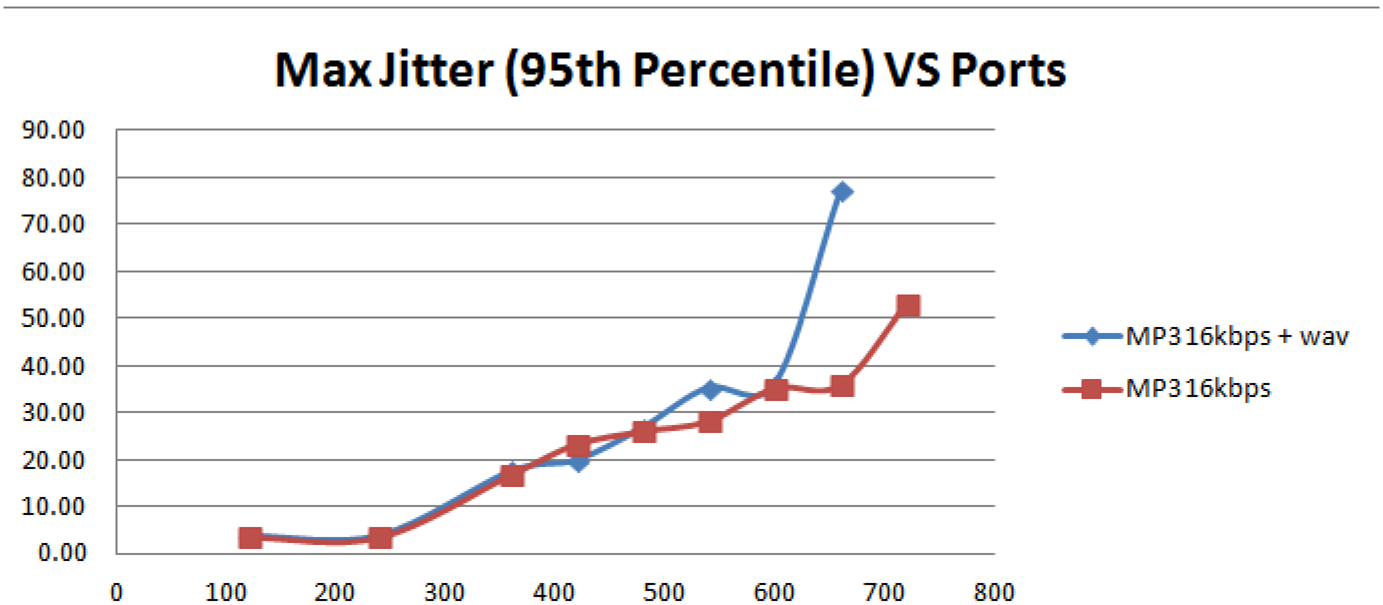


Figure 94: Comparison of Max Jitter on VM env, MP3 + wav vs. MP3 only, on RH EL 6.5

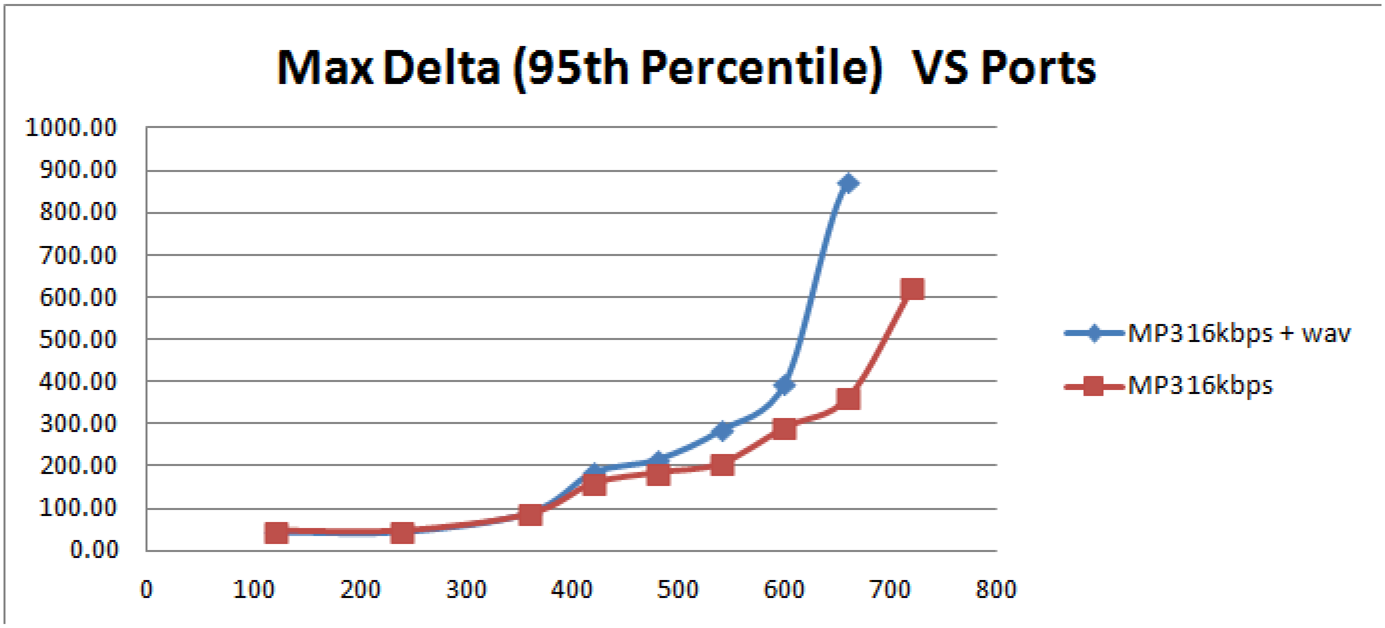


Figure 95: Comparison of Max Delta on VM env, MP3 + wav vs. MP3 only, on RH EL 6.5

480 ports are recommended and preferred for this MP3 + wav profile, an 11.1% reduction (480 vs. 540 for MP3 only); peak would be 660, a 9.1% reduction (600 vs. 660 for MP3 only).

Below is a table to illustrate overall 6 VMs disk IOPS:

Figure 96: System Disk IOPS on a VM environment of dual hex cores on EL 6.5, MP3 16 kbps + wav

Ports	Overall 6 VMs Disk IOPS			SSD Drive Disk IOPS		
	Total	Reads	Writes	Total	Reads	Writes
120	52.99	0.00	52.99	48.728	0.000	48.728
240	100.50	0.00	100.50	95.174	0.000	95.174
360	144.34	0.00	144.34	138.864	0.000	138.864
420	164.65	0.00	164.65	158.979	0.000	158.979
480	183.45	0.00	183.45	177.711	0.000	177.711
540	207.27	0.00	207.27	201.564	0.000	201.564
600	224.97	0.00	224.97	219.197	0.000	219.197
660	275.49	0.00	275.49	269.584	0.000	269.584
720	187.34	0.00	187.33	179.984	0.001	179.983

Compared with the MP3-only profile, overall 6 VM disk IOPS for MP3 + wav profile shows almost double IOPS, as in the previous physical server section.

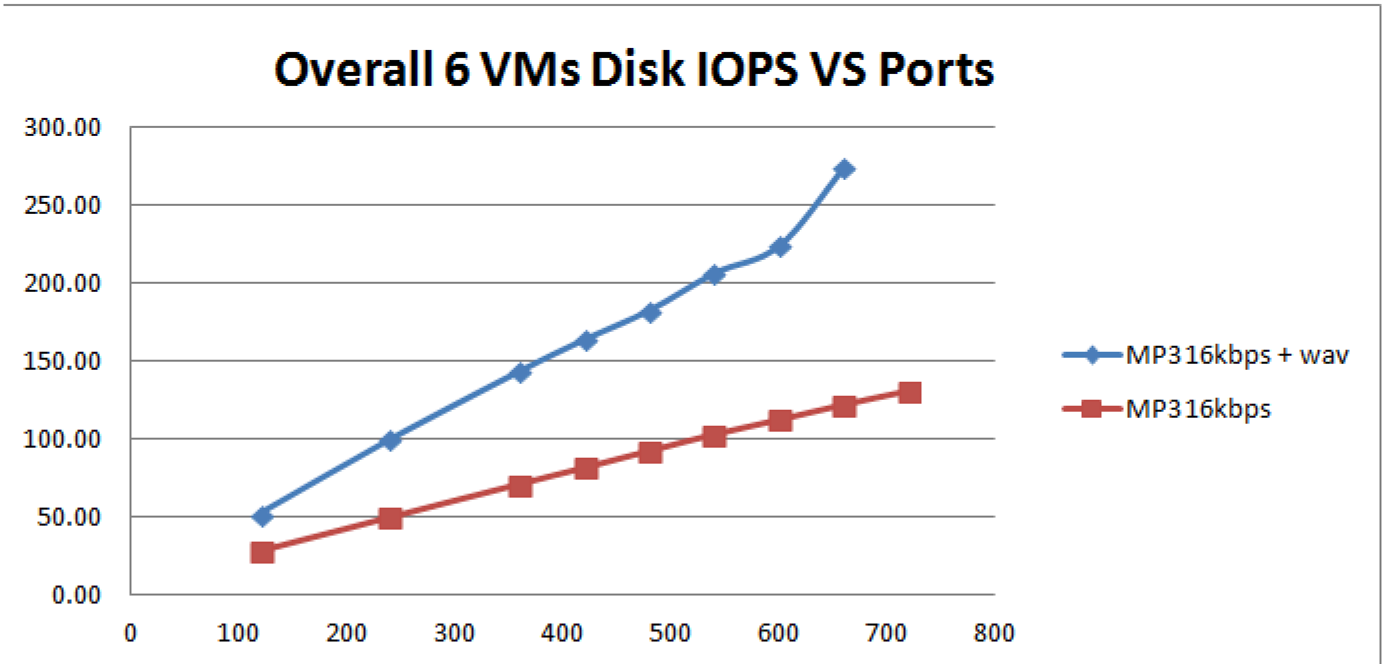


Figure 97: Comparison of System Disk IOPS on VM env, MP3 + wav vs. MP3 only, on RH EL 6.5

The table below illustrats overall data throughput for this MP3 + wav profile on VMs of RH EL 6.5 environment.

Figure 98: Data throughputs from overall 6 VMs of dual hex core, MP3 16 kbps + wav, on EL 6.5

	Overall Disk KB/sec			SSD Drive Disk KB/sec		
Ports	**Total KB/ sec**	**Read KB/ sec**	**Write KB/ sec**	**Total KB/ sec**	**Read KB/ sec**	**Write KB/ sec**
120	2376.30	0.00	2376.30	2347.222	0.000	2347.222
240	4684.79	0.00	4684.79	4646.371	0.000	4646.371
360	6975.83	0.00	6975.83	6933.441	0.000	6933.441
420	8100.79	0.00	8100.79	8056.843	0.001	8056.842
480	9242.32	0.00	9242.32	9195.871	0.001	9195.871
540	10391.78	0.00	10391.78	10344.249	0.001	10344.249
600	11512.54	0.00	11512.54	11462.150	0.001	11462.149
660	12804.19	0.01	12804.18	12752.305	0.001	12752.304
720	9380.58	0.00	9380.58	9336.194	0.003	9336.191

Encryption

MP3 16 kbps Only on a Physical Server of Single Hex Core

This is SW Profile 3a (MP3 16 kbps only with encryption) on HW Profile 1 for a physical server which SW Profile 1a (MP3 only 16 kbps without encryption) is used as baseline to compare with. Here are the three graphs illustrating overall system CPU usage, MCP CPU usage and memory usage:

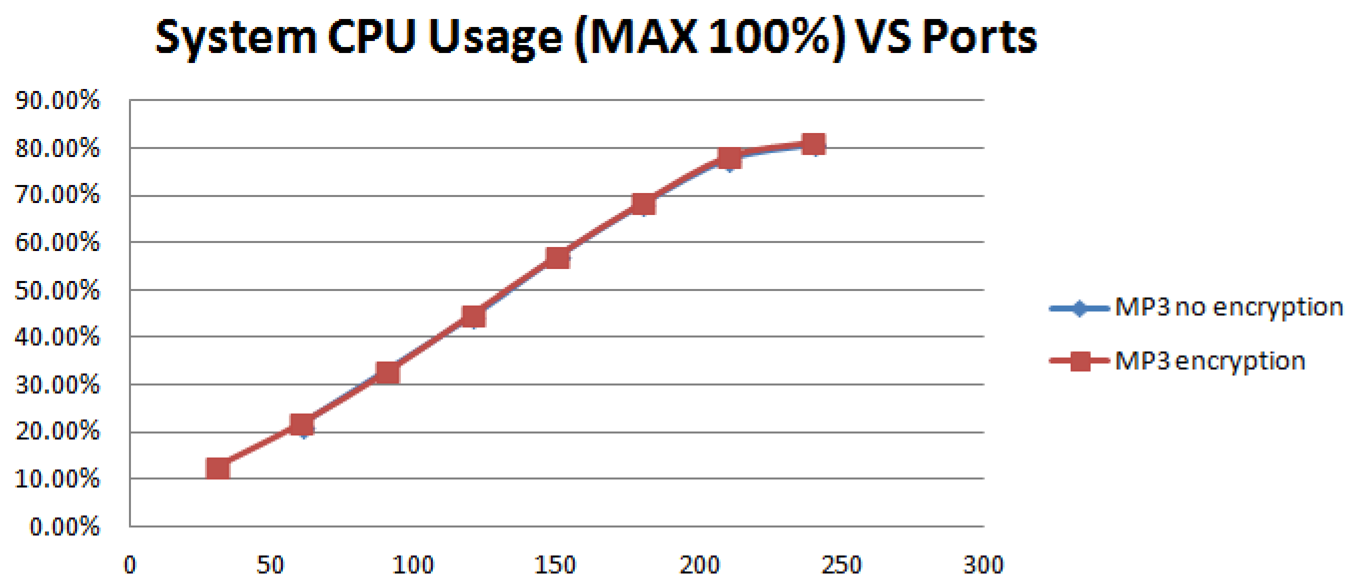


Figure 99: Comparison of System CPU Usage on a Physical Server, MP3 only 16 kbps encryption vs. non-encryption, on RH EL 6.5

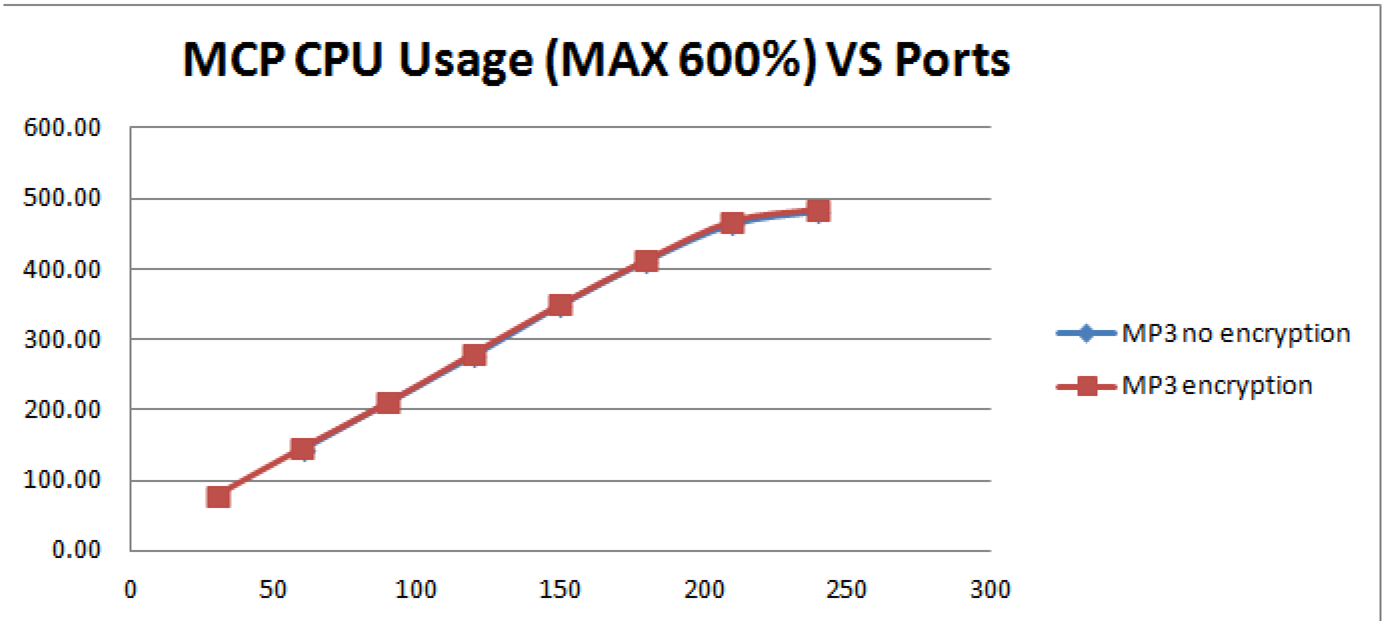


Figure 100: Comparison of MCP CPU Usage on a Physical Server, MP3 only 16 kbps encryption vs. non-encryption, on RH EL 6.5

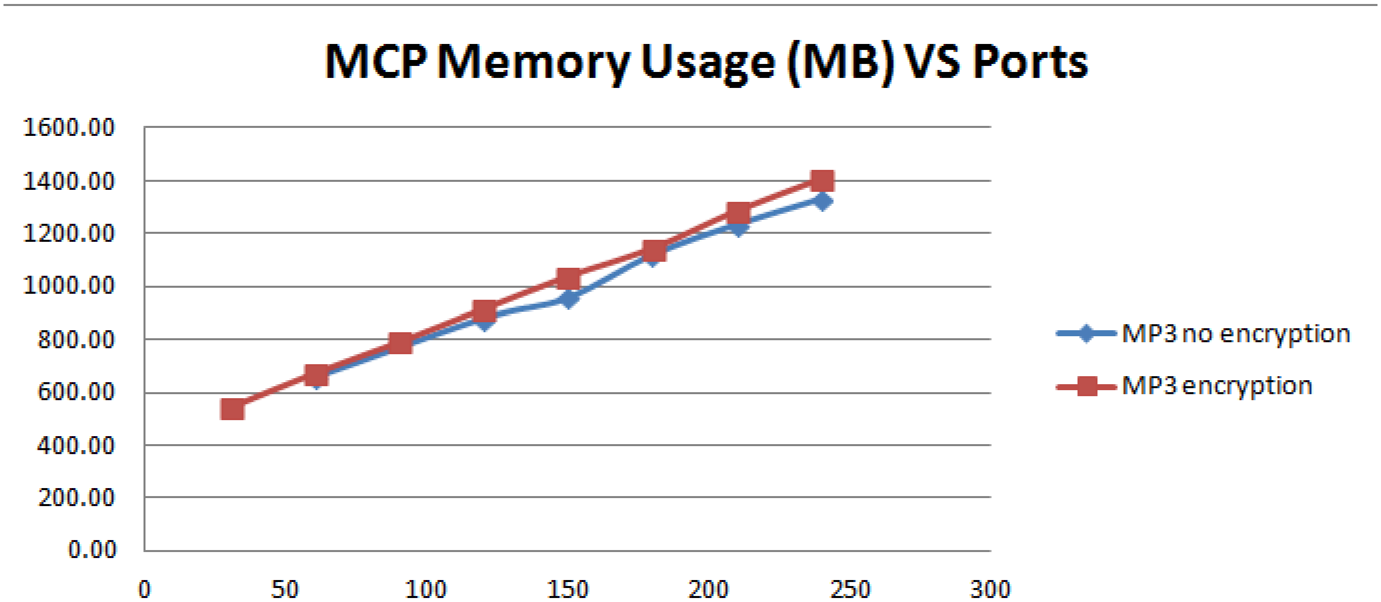


Figure 101: Comparison of MCP Memory Usage on a Physical Server, MP3 only 16 kbps encryption vs. non-encryption, on RH EL 6.5

It can be observed that both system CPU and MCP CPU are quite inline to each other between encryption and non-encryption profiles while MCP memory for encryption is slightly higher than non-encryption.

Let us look at audio quality metrics further:

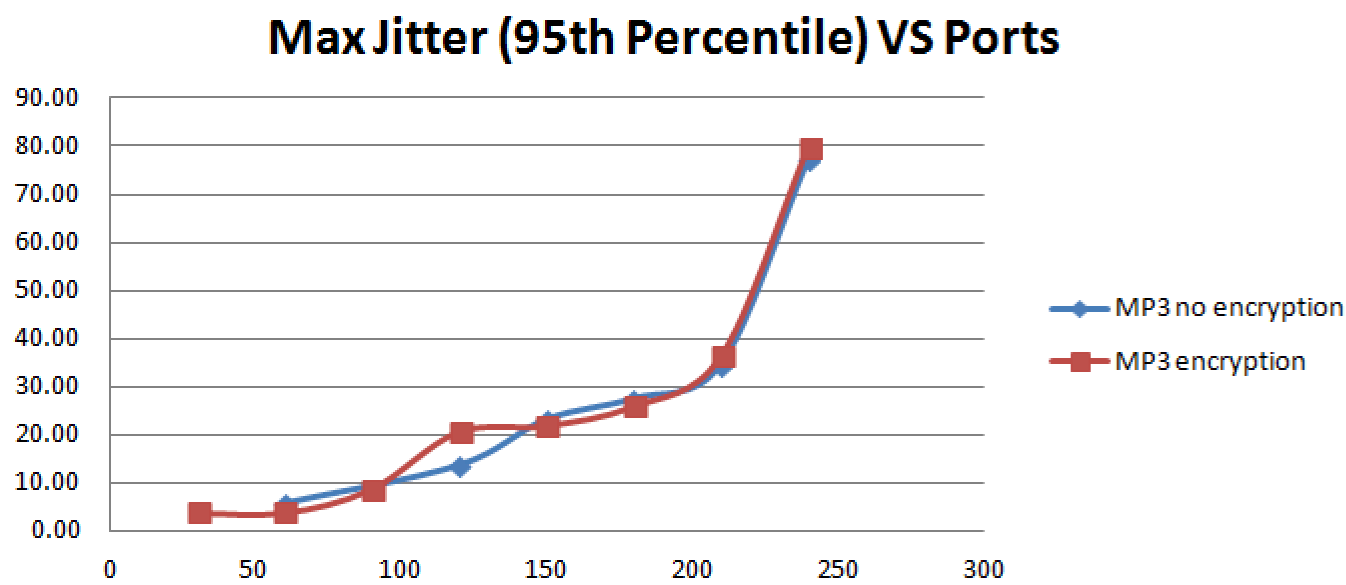


Figure 102: Comparison of Max Jitter on a Physical Server, MP3 only, Encryption vs. Non-encryption on EL 6.5

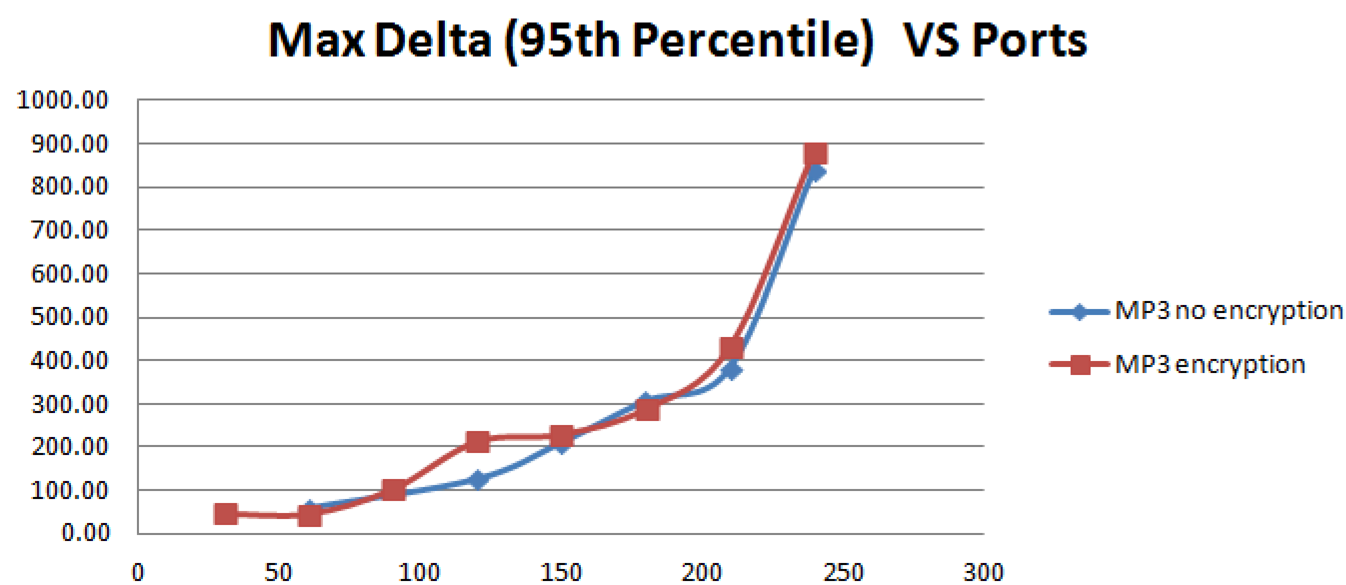


Figure 103: Comparison of Max Delta on a Physical Server, MP3 only, Encryption vs. Non-encryption on EL 6.5

Max Jitter is similar for both encryption and non-encryption scenarios, as are the Max Delta metrics. Thus, the preferred ports (540) and peak ports (660) for encryption are the same as for non-encryption.

System disk IOPS is illustrated below:

Figure 104: System Disk IOPS on a Physical Server on EL 6.5, MP3 16 kbps only, Encryption

Ports	Physical Server Disk IOPS		
	Total	Reads	Writes
30	8.12	0.000	8.122
60	14.22	0.000	14.220
90	19.98	0.000	19.975
120	25.12	0.000	25.122
150	30.62	0.000	30.621
180	35.07	0.000	35.074
210	39.83	0.000	39.828
240	44.74	0.000	44.739

The graph below compares encryption with non-encryption:

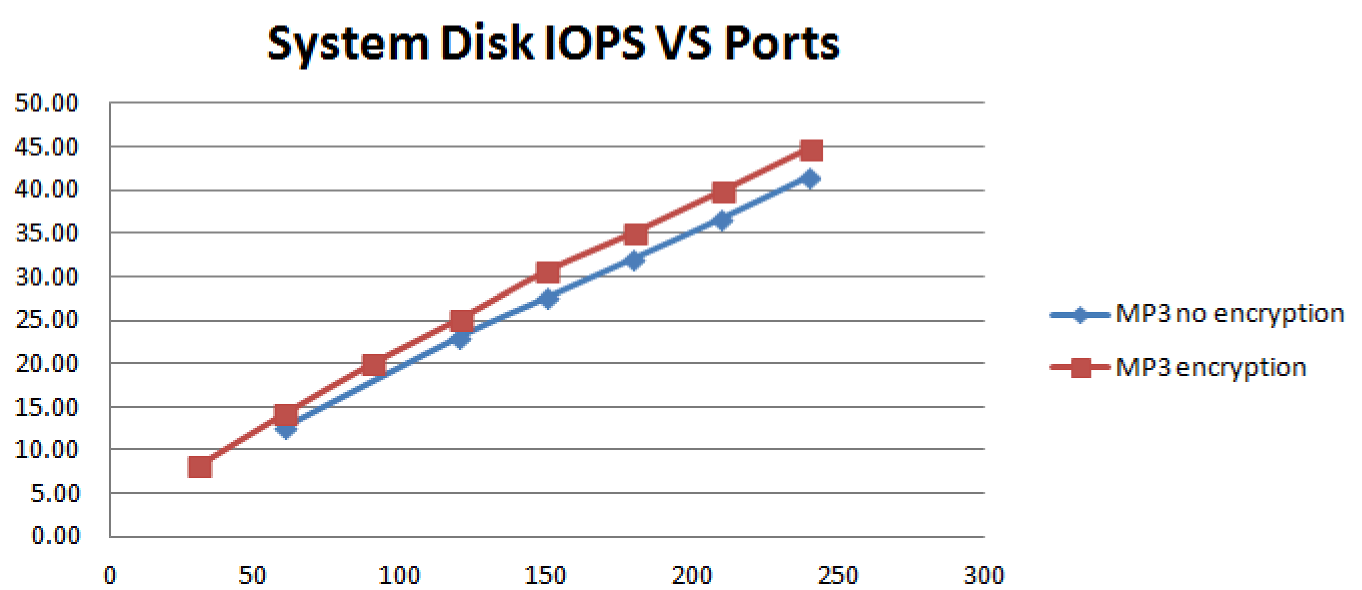


Figure 105: Comparison of System Disk IOPS on a Physical Server, MP3 only, on EL 6.5, Encryption vs. Non-encryption

Slightly higher system disk IOPS occurs in the encryption scenario, likely caused by the extra key/pem files required for encryption.

MP3 16 kbps Only on VMs of Dual Hex Cores

This test uses SW Profile 3a (MP3 16 kbps only with encryption) on VM Profile 4 configured as HW Profile 0 for a VM environment, compared with SW Profile 1a (MP3 only 16 kbps without encryption) on the same hardware specification. Below are graphs illustrating overall system CPU usage and memory usage:

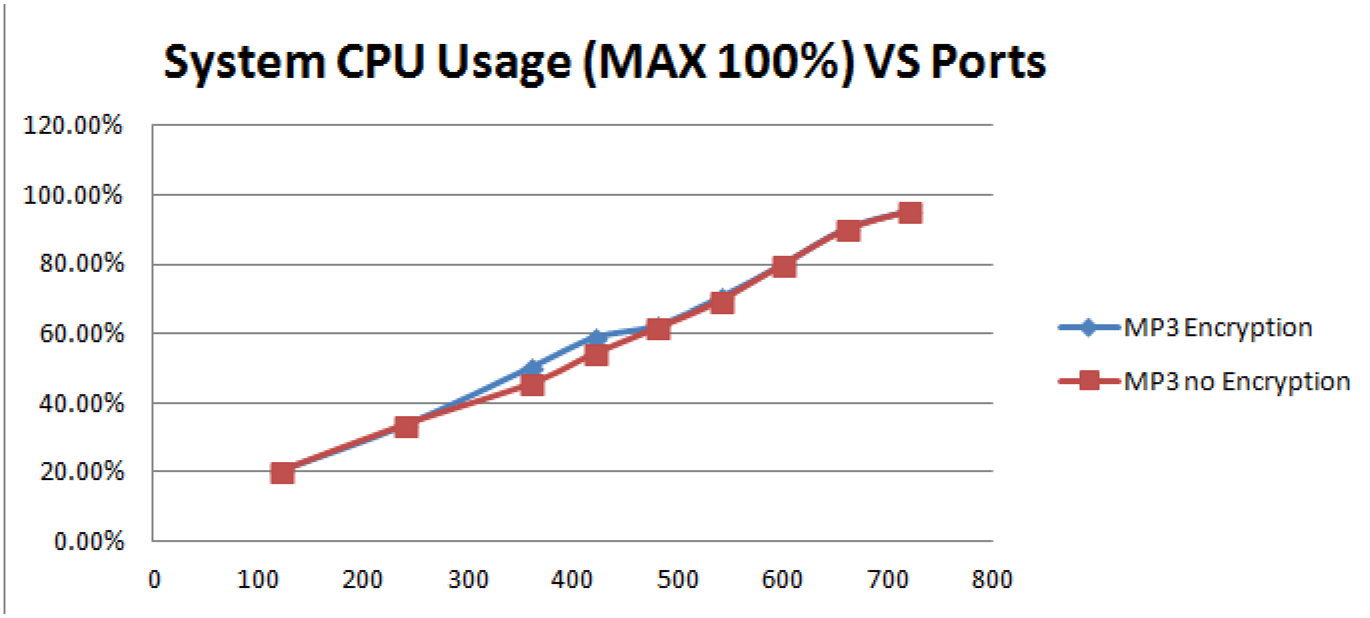


Figure 106: Comparison of System CPU Usage on VMs, MP3 only 16 kbps encryption vs. non-encryption, on RH EL 6.5

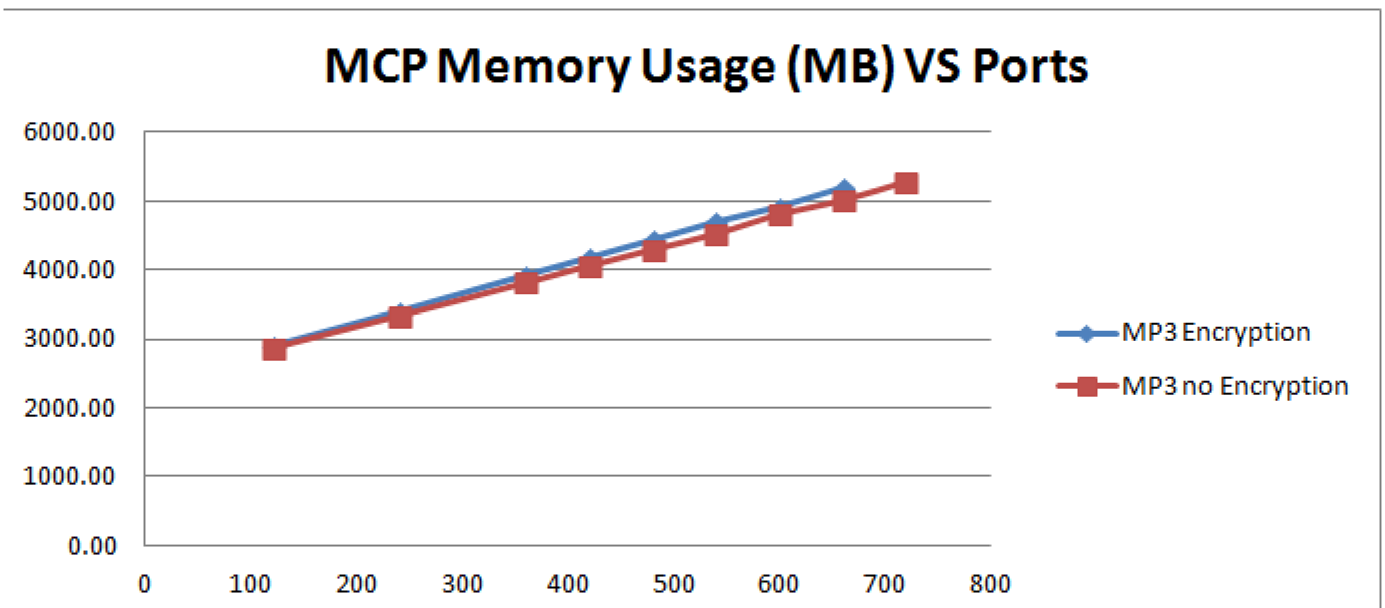


Figure 107: Comparison of MCP Memory Usage on VMs, MP3 only 16 kbps encryption vs. non-encryption, on RH EL 6.5

As observed in previous physical server graphs, CPU usage is almost the same for both encryption and non-encryption, while MCP memory usage is slightly higher for encryption.

Consider audio quality metrics:

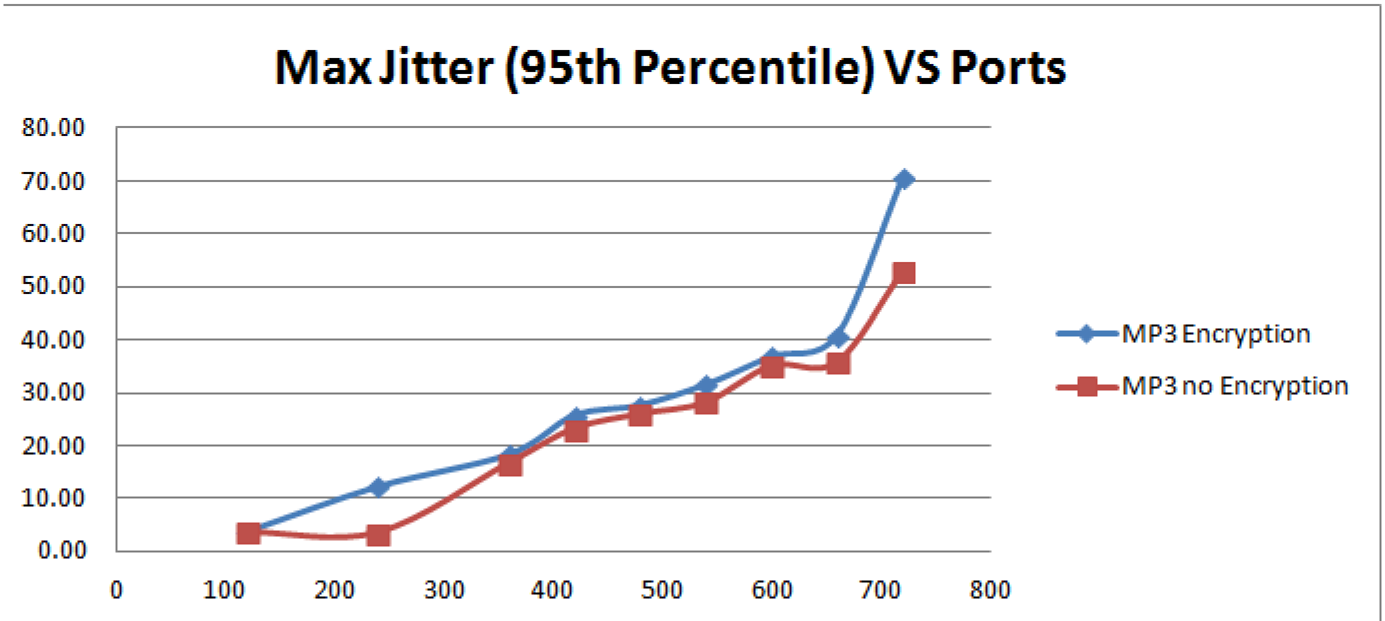


Figure 108: Comparison of Max Jitter on VMs, MP3 only, Encryption vs. Non-encryption on EL 6.5

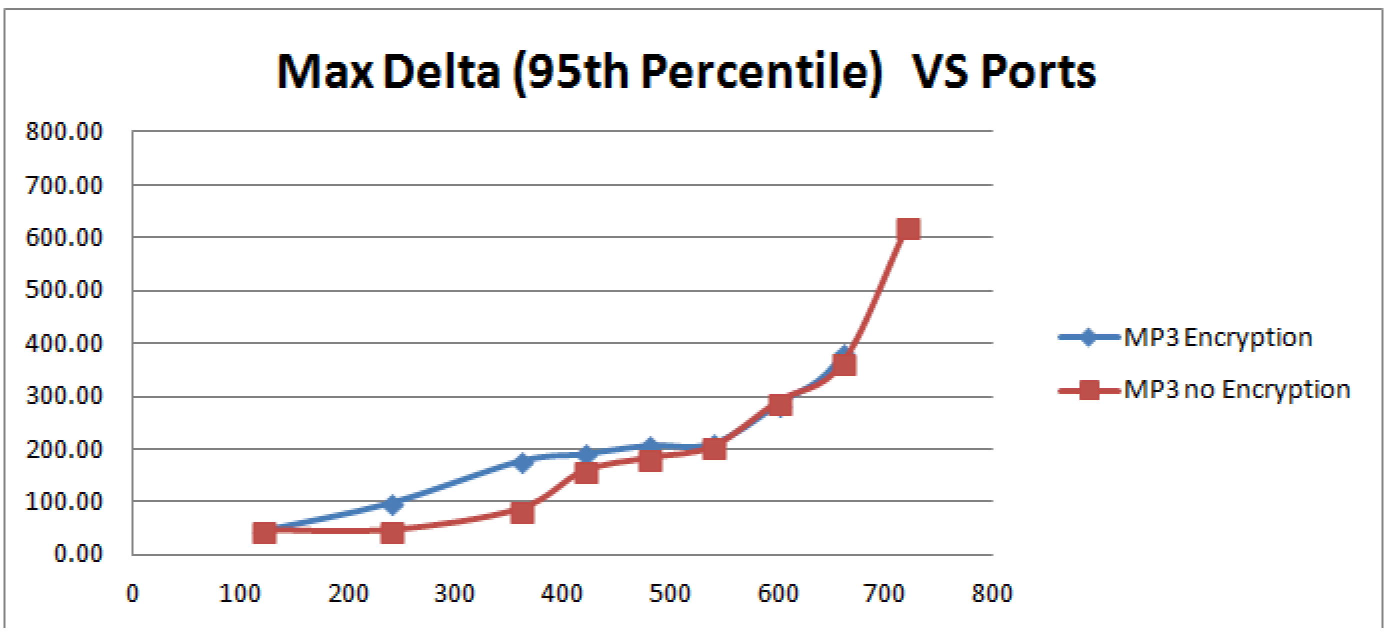


Figure 109: Comparison of Max Delta on VMs, MP3 only, Encryption vs. Non-encryption on EL 6.5

Similar trends can be observed in the previous physical server section that both encryption and non-encryption achieved similar value for both Max Jitter and Max Delta. So the preferred ports (540) and peak ports (660) for encryption would be the same as non-encryption.

Figure 110: Overall System Disk IOPS on VMs of EL 6.5, MP3 16 kbps only, Encryption

Ports	Overall 6 VMs Disk IOPS			SSD Drive Disk IOPS		
Total	Reads	Writes	Total	Reads	Writes	
120	30.44	0.00	30.44	25.997	0.000	25.997
240	53.41	0.00	53.41	47.939	0.000	47.939
360	75.57	0.00	75.57	70.011	0.000	70.011
420	86.37	0.00	86.37	80.600	0.000	80.600
480	97.32	0.00	97.32	91.564	0.000	91.564
540	108.20	0.00	108.20	102.393	0.000	102.393
600	117.95	0.00	117.95	112.132	0.000	112.132
660	127.85	0.00	127.85	121.911	0.000	121.911
720	136.85	0.00	136.85	130.951	0.000	130.951

The graph below compares encryption with non-encryption:

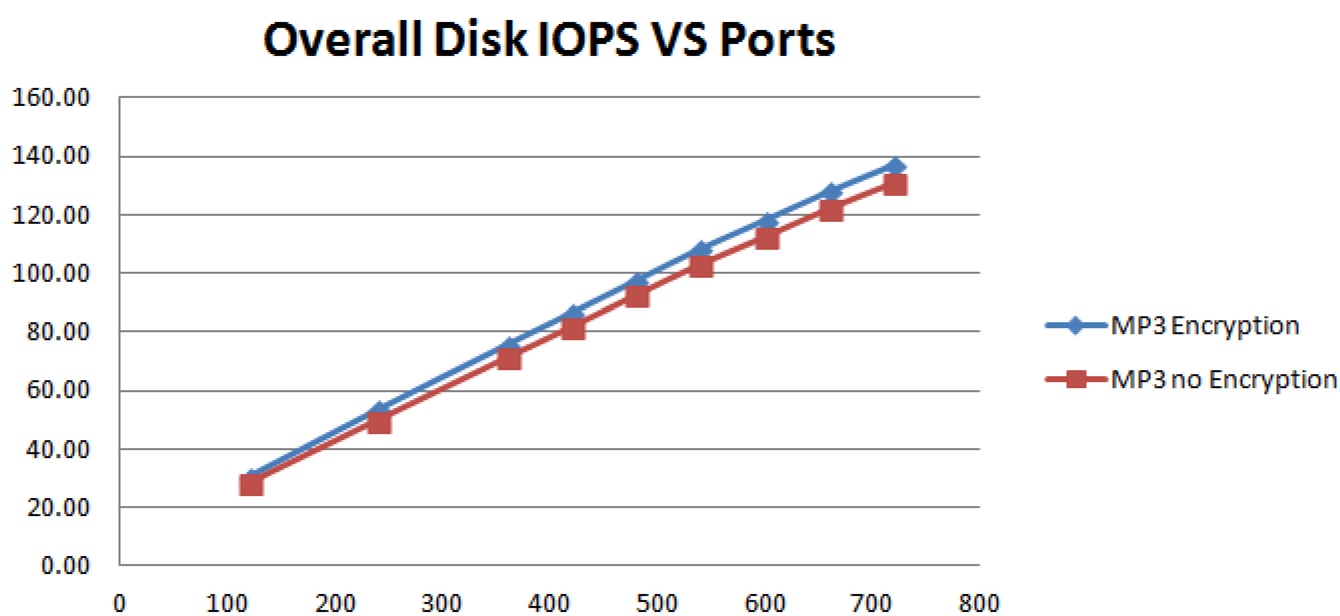


Figure 111: Comparison of System Disk IOPS on VM env, MP3 only, on EL 6.5, Encryption vs. Non-encryption

As in the previous physical server section, system disk IOPS for encryption is slightly higher than non-encryption.

Data throughput is illustrated in this table:

Figure 112: Data throughputs from overall 6 VMs of dual hex core, MP3 16 kbps only, encryption, on EL 6.5

Ports	Overall Disk KB/sec	SSD Drive Disk KB/sec

	Total KB/ sec	Read KB/ sec	Write KB/ sec	Total KB/ sec	Read KB/ sec	Write KB/ sec
120	435.55	0.00	435.55	403.192	0.000	403.192
240	822.06	0.00	822.06	780.379	0.000	780.379
360	1186.43	0.00	1186.43	1140.874	0.000	1140.874
420	1359.14	0.00	1359.14	1311.668	0.000	1311.668
480	1549.49	0.00	1549.49	1500.982	0.000	1500.982
540	1719.89	0.00	1719.89	1669.506	0.000	1669.506
600	1905.09	0.00	1905.09	1853.208	0.000	1853.208
660	2081.23	0.00	2081.23	2027.495	0.000	2027.495
720	2269.56	0.00	2269.56	2214.658	0.000	2214.658

MP3 16 kbps + wav on VMs of Dual Hex Cores

This test uses SW Profile 4a (MP3 16 kbps + wav with encryption) on VM Profile 4 configured as HW Profile 1 for a VM environment to compare with SW Profile 2a (MP3 16 kbps + wav without encryption) on the same HW spec. Below are two graphs illustrating overall system CPU usage and memory usage:

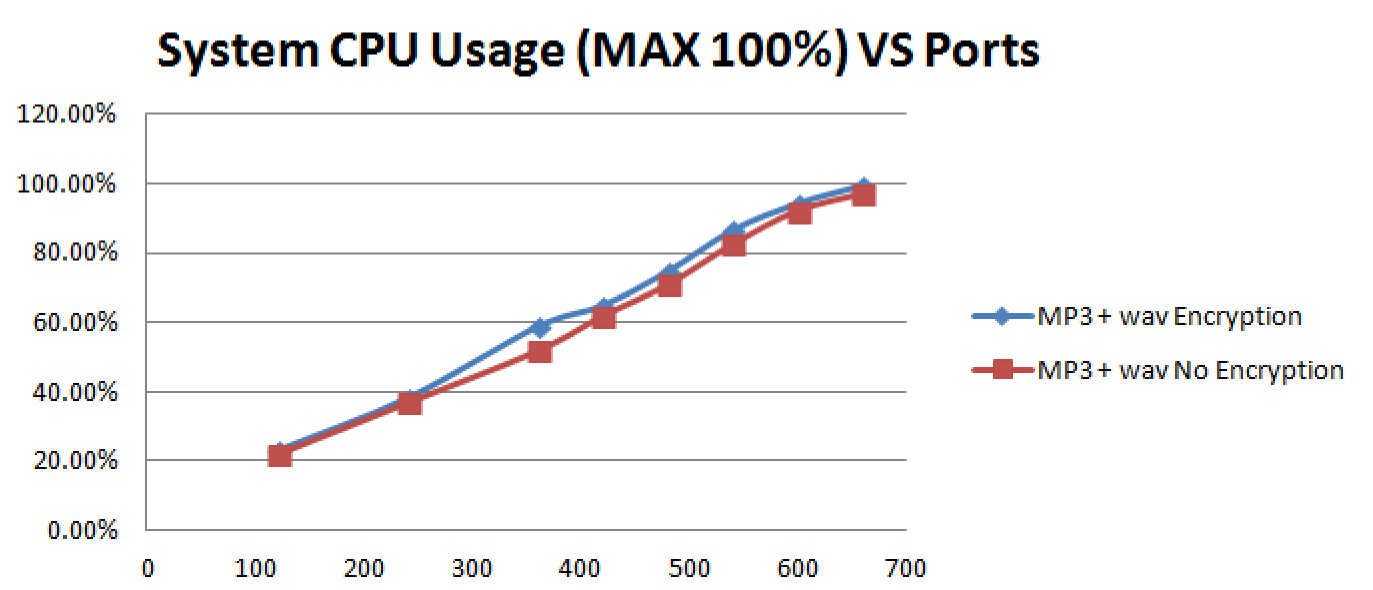


Figure 113: Comparison of System CPU Usage on VMs, MP3 16 kbps + wav encryption vs. non-encryption, on RH EL 6.5

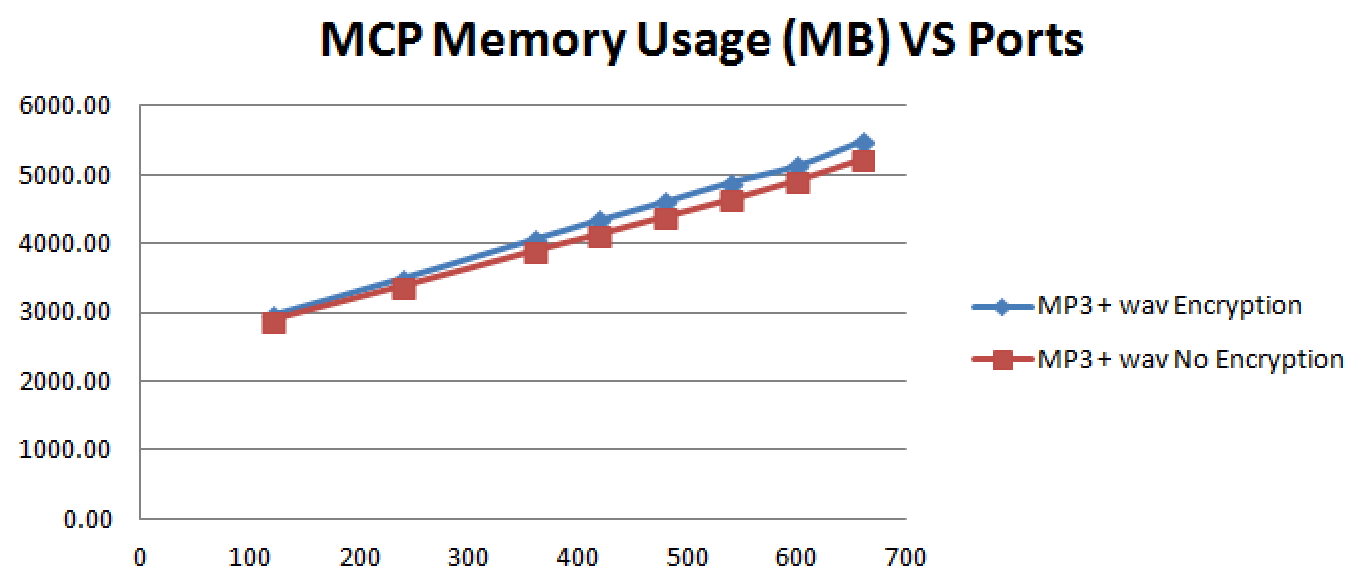


Figure 114: Comparison of MCP Memory Usage on VMs, MP3 16 kbps + wav encryption vs. non-encryption, on RH EL 6.5

System CPU usage is quite close to each other for both encryption and non-encryption, while MCP memory usage for encryption is slightly higher than for non-encryption, similar to the previous MP3 only test scenarios.

The audio quality metrics of Max Jitter and Max Delta also show similar trends.

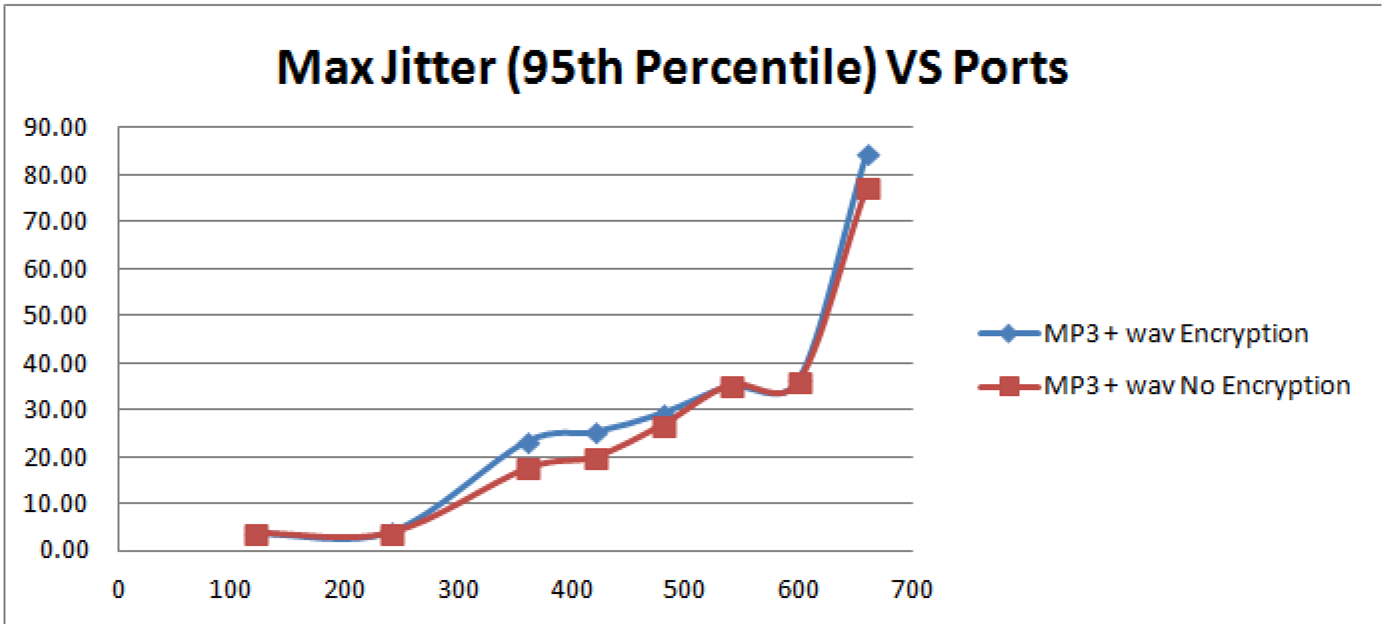


Figure 115: Comparison of Max Jitter on VMs, MP3 + wav, Encryption vs. Non-encryption on EL 6.5

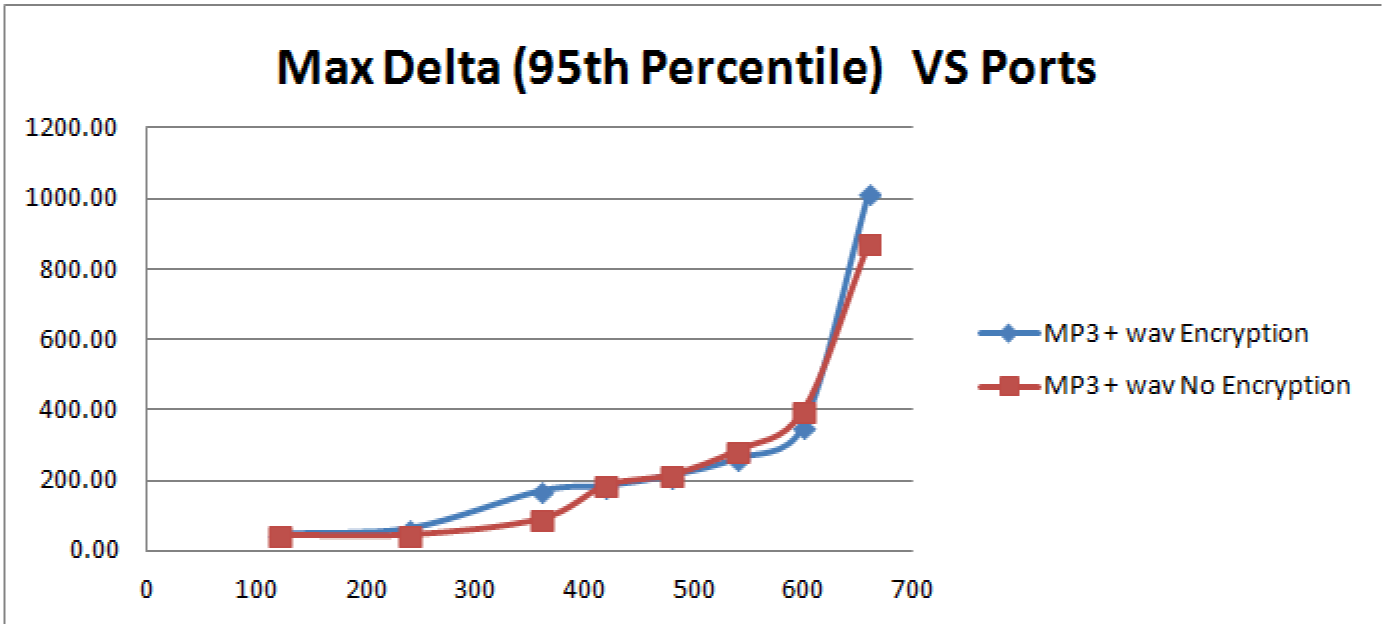


Figure 116: Comparison of Max Delta on VMs, MP3 + wav, Encryption vs. Non-encryption on EL 6.5

The recommended and preferred ports for encryption of MP3 + wav would be 480 the same as non-encryption of MP3 + wav, as is 600 for peak ports.

The table below shows overall system disk IOPS, for reference:

Figure 117: Overall System Disk IOPS on VMs of EL 6.5, MP3 16 kbps + wav, Encryption

Ports	Overall 6 VMs Disk IOPS			SSD Drive Disk IOPS		
	Total	Reads	Writes	Total	Reads	Writes
120	53.97	0.00	53.97	49.506	0.000	49.506
240	102.98	0.00	102.98	97.468	0.000	97.468
360	149.87	0.00	149.87	144.235	0.000	144.235
420	171.89	0.00	171.89	166.144	0.000	166.144
480	196.97	0.00	196.97	191.140	0.000	191.140
540	223.52	0.01	223.51	217.663	0.000	217.663
600	246.26	0.03	246.22	240.216	0.000	240.216
660	296.60	0.00	296.60	290.582	0.000	290.582

The graph below compares encryption with non-encryption, and shows the same trend as observed previously in this section:

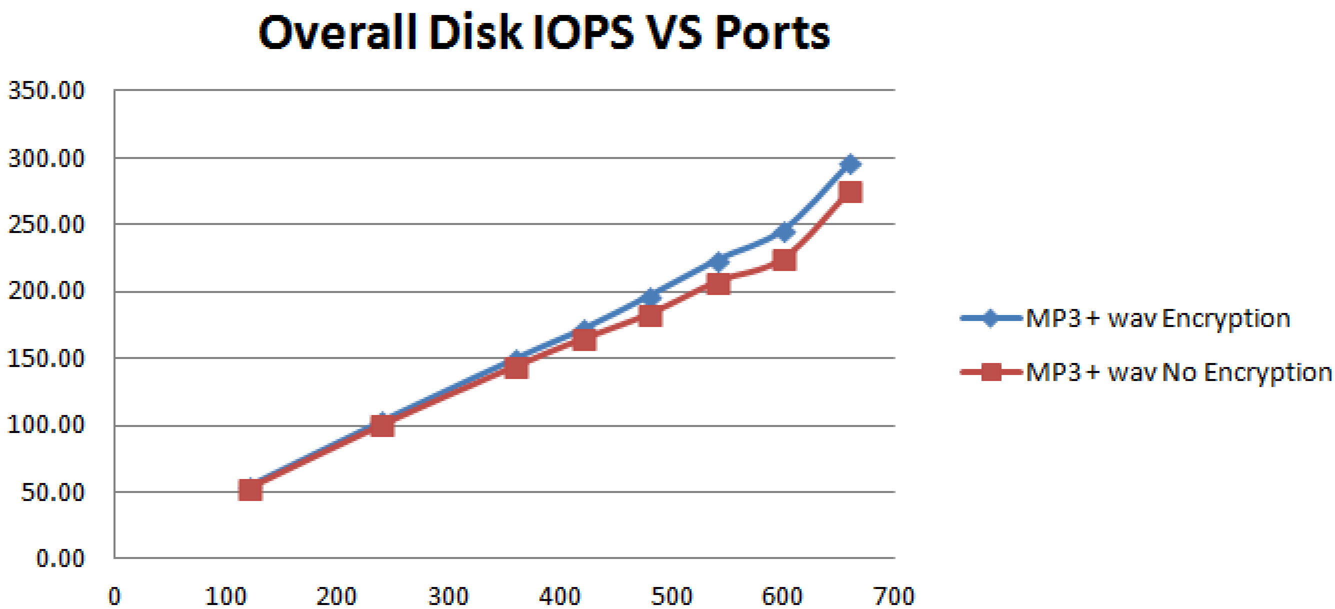


Figure 118: Comparison of System Disk IOPS on VM env, MP3 + wav, on EL 6.5, Encryption vs. Non-encryption

Data throughput is also listed below as reference:

Figure 119: Data throughputs from overall 6 VMs of dual hex core, MP3 16 kbps + wav, encryption, on EL 6.5

Ports	Overall Disk KB/sec			SSD Drive Disk KB/sec		
	Total KB/sec	Read KB/sec	Write KB/sec	Total KB/sec	Read KB/sec	Write KB/sec
120	2421.76	0.00	2421.76	2373.612	0.000	2373.612
240	4756.37	0.00	4756.37	4699.737	0.000	4699.737
360	7065.62	0.00	7065.62	7004.491	0.000	7004.491
420	8179.23	0.00	8179.23	8116.591	0.000	8116.591
480	9366.53	0.00	9366.53	9301.426	0.000	9301.426
540	10489.26	0.14	10489.12	10423.230	0.000	10423.230
600	11647.29	0.78	11646.51	11574.973	0.000	11574.973
660	12976.30	0.06	12976.24	12905.764	0.001	12905.763

Multiple Dispatcher Tests

The GVP 8.5.1 August 2015 release added support for multiple dispatchers and improved latency performance on physical servers. Capacity increased due to the latencies improvements; CPU usage can now be increased as well.

Below are results from three tests from configurations of one dispatcher, four dispatchers, and eight dispatchers on a single 8-core physical server using SW Profile 1a (16Kbps MP3 stereo only without encryption).

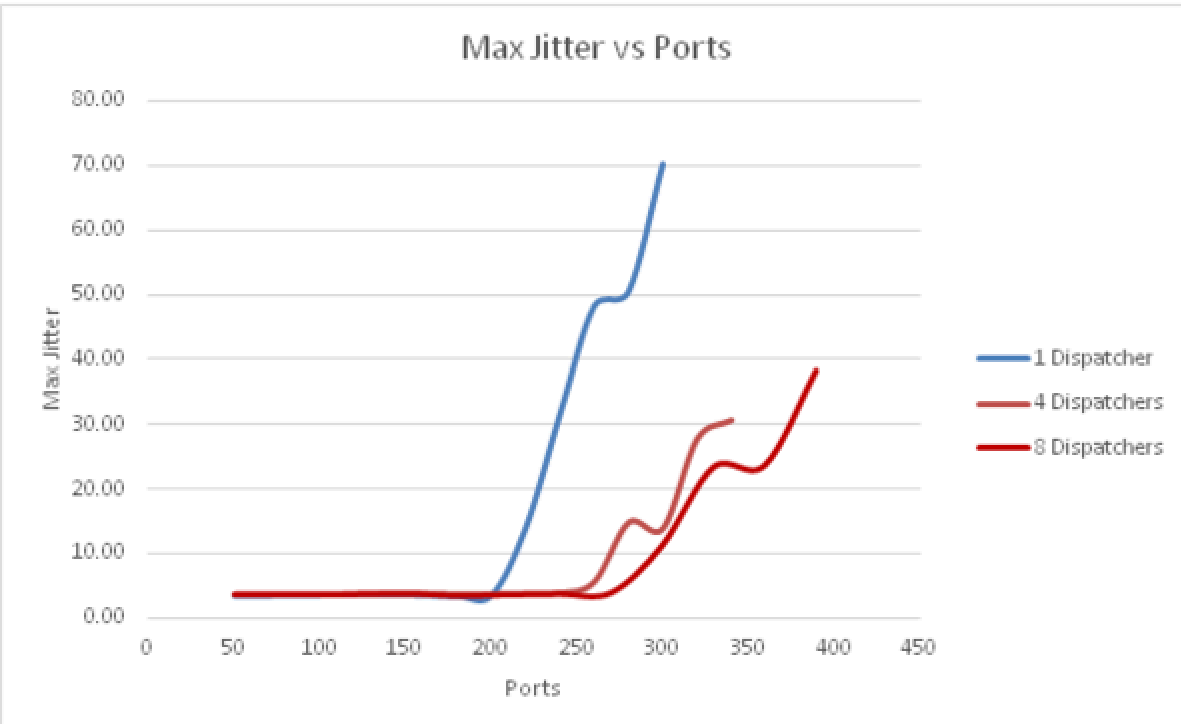


Figure 120: Max Jitter, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on EL 6.6

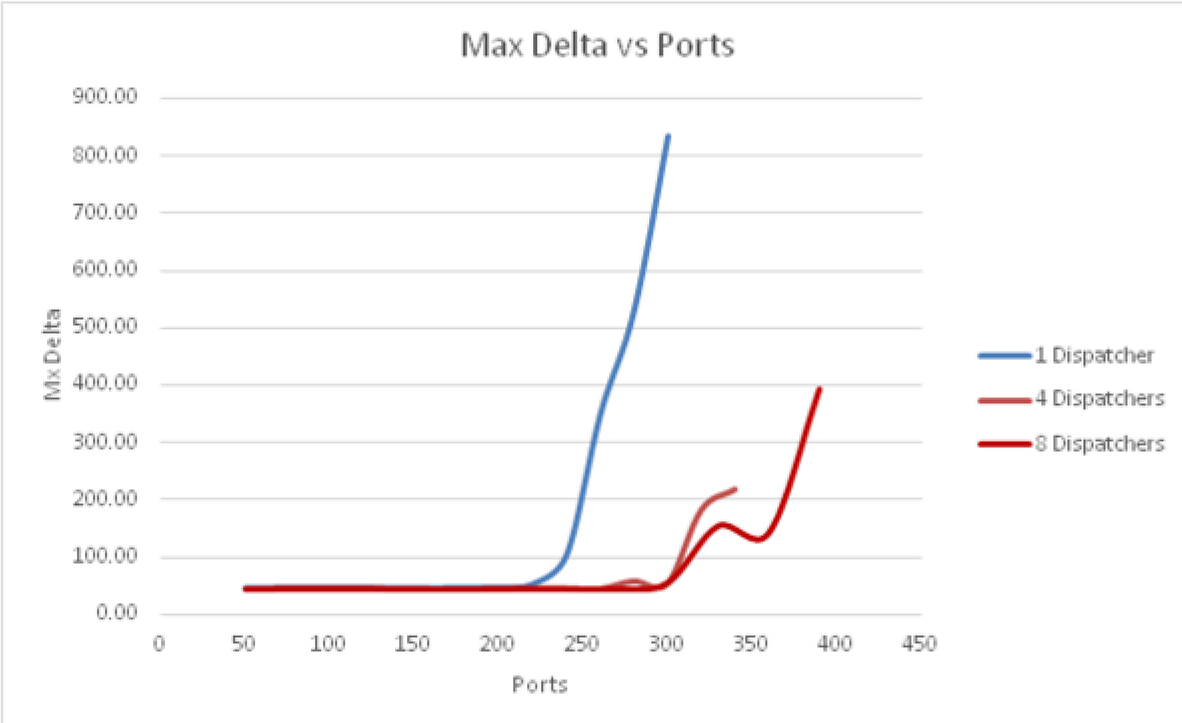


Figure 121: Max Delta, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on EL 6.6

Results from four-dispatcher and eight-dispatcher configurations show significant improvement, when compared against the default of one dispatcher. Additionally, the results with eight dispatchers show slightly better results than with four dispatchers.

The next results are for system CPU usage:

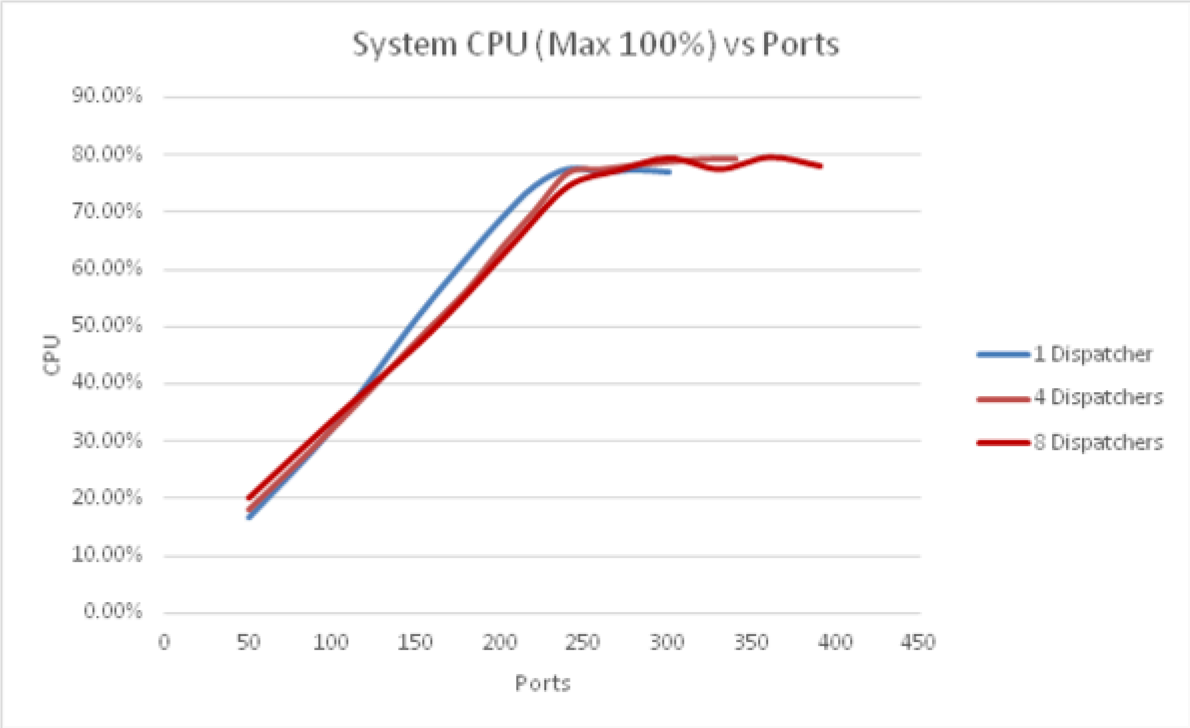


Figure 122: System CPU, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on EL 6.6

CPU usage for one dispatcher is slightly higher than four and eight dispatchers, while CPU usage flattens out near peak capacity. The four-dispatcher and eight-dispatcher configurations may drive the CPU slightly higher, and thus achieve the higher capacity.

The next results are for memory usage:

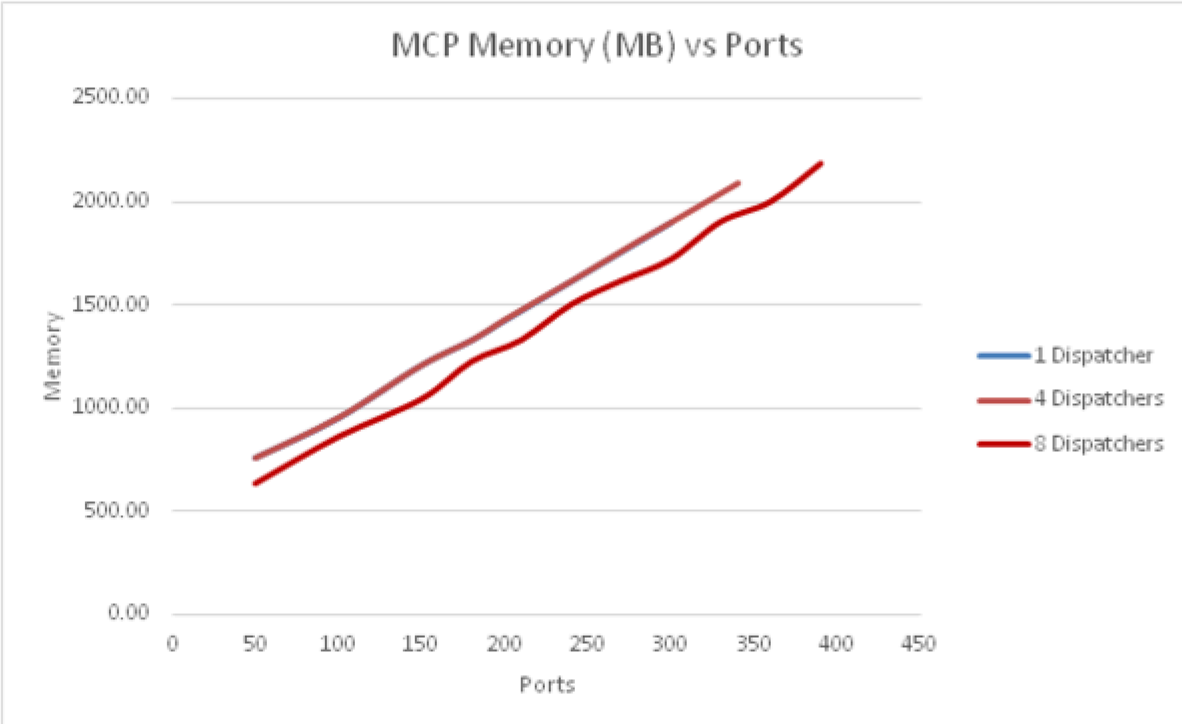


Figure 123: MCP Memory Usage, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on EL 6.6

Memory usage for four-dispatcher and eight-dispatcher configurations is similar, with eight dispatchers using slightly less memory. Genesys recommends the eight-dispatcher configuration. Comparing eight dispatchers to the default one dispatcher reveals a 36% increase in preferred capacity (300 vs 220 ports); eight dispatchers achieved a 50% increase in peak capacity.

Jitter and Delta Tests

The results below were achieved using the same four-dispatcher and eight-dispatcher configurations, on a Windows 2008 R2 server with the same hardware specifications as the previous tests.

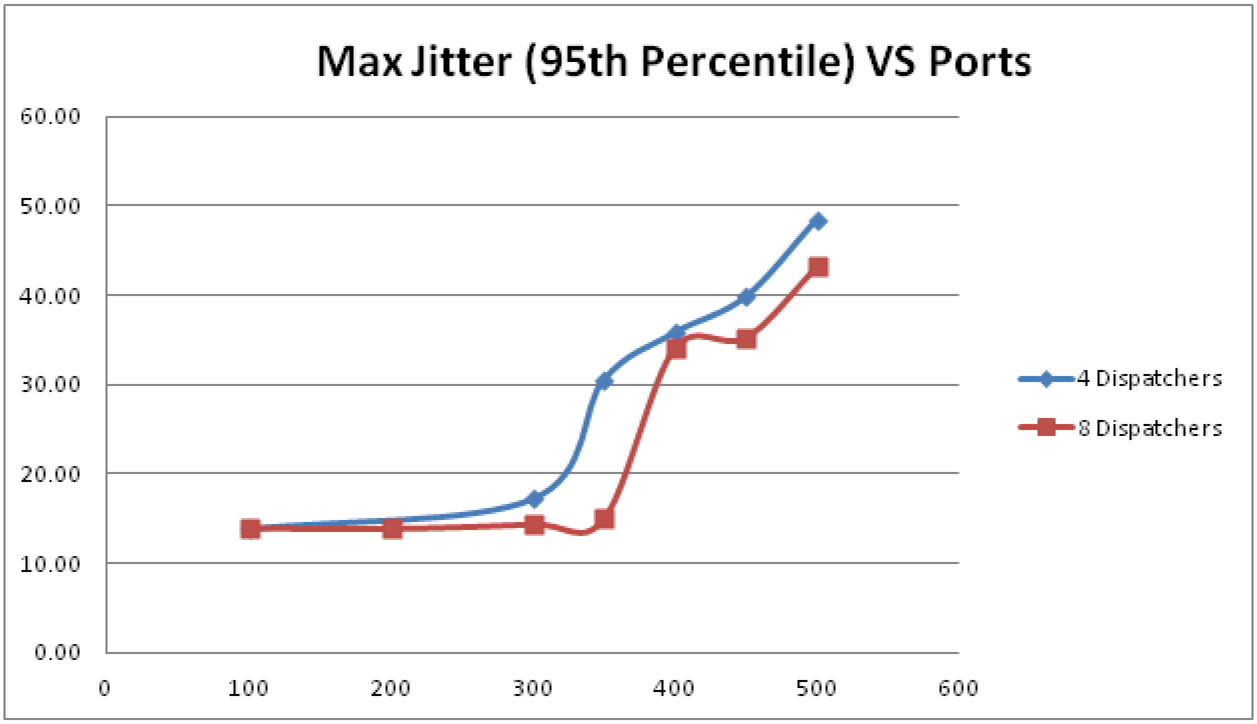


Figure 124: Max Jitter, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on W2K8 R2

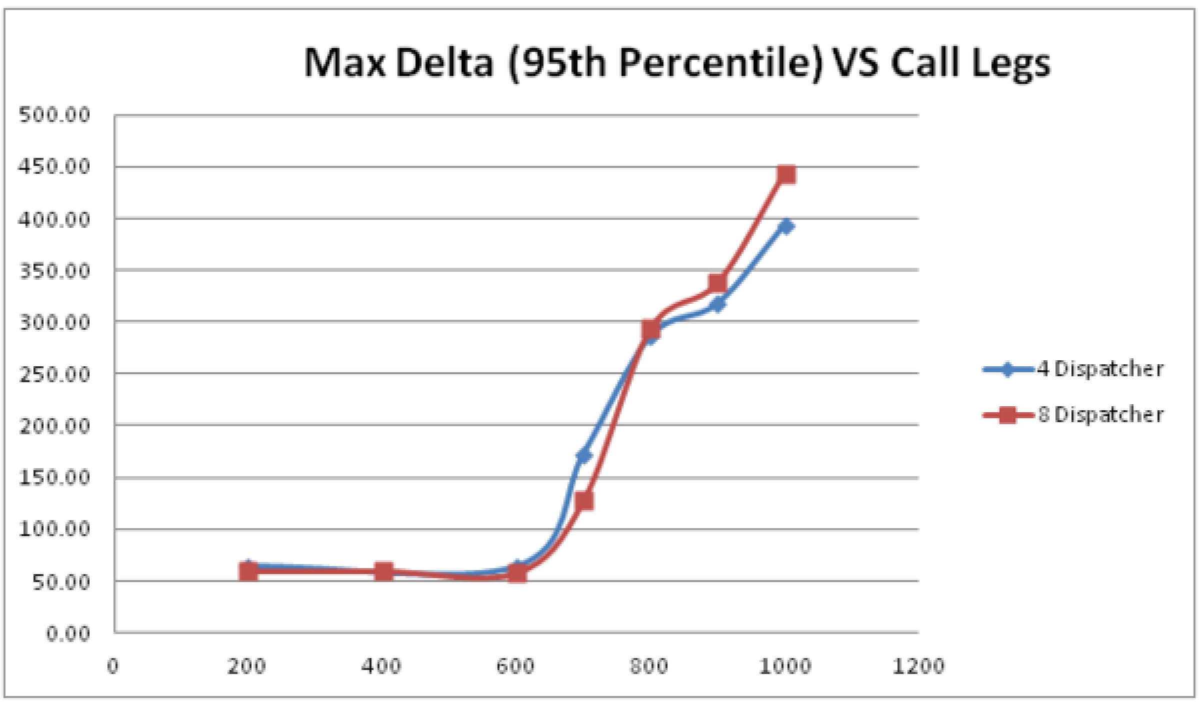


Figure 125: Max Delta, Different # of dispatchers of 16Kbps MP3 stereo only, No encryption, on W2K8 R2

These results are the same as were seen on similar Linux systems—eight-dispatcher configurations achieved slightly better results than four-dispatcher configurations. Genesys recommends that the number of dispatchers and the number of cores should be the same in physical server configurations.

MP3 8 Kbps Mono Support

Physical Server Windows 2008 R2

The GVP 8.5.1 August 2015 release added support for MP3 Mono with an 8Kbps bit rate. Genesys tested the recommended eight dispatchers configurations on physical servers running Windows 2008 R2 Server and Linux EL 6.6, to compare with the results for 16Kbps MP3 stereo configurations.

Below are graphs of Max Jitter and Max Delta for latencies on a Windows 2008 R2 Server:

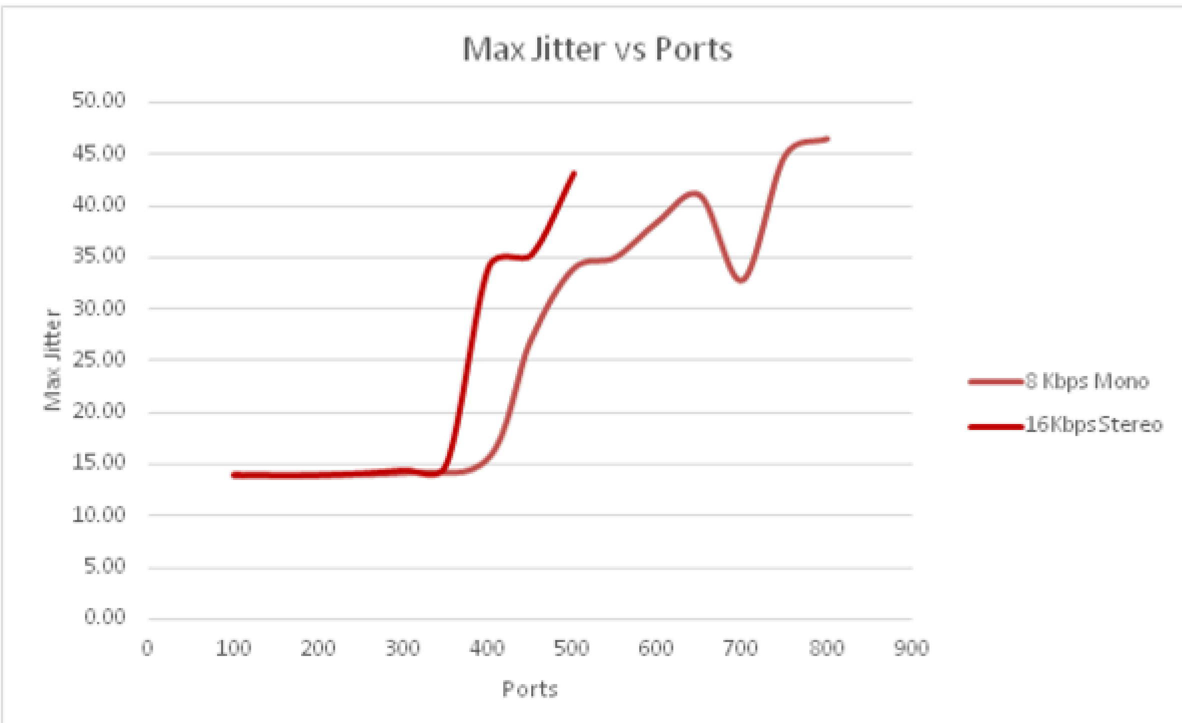


Figure 126: Max Jitter, 16Kbps MP3 stereo vs 8Kbps Mono, No encryption, on W2K8 R2

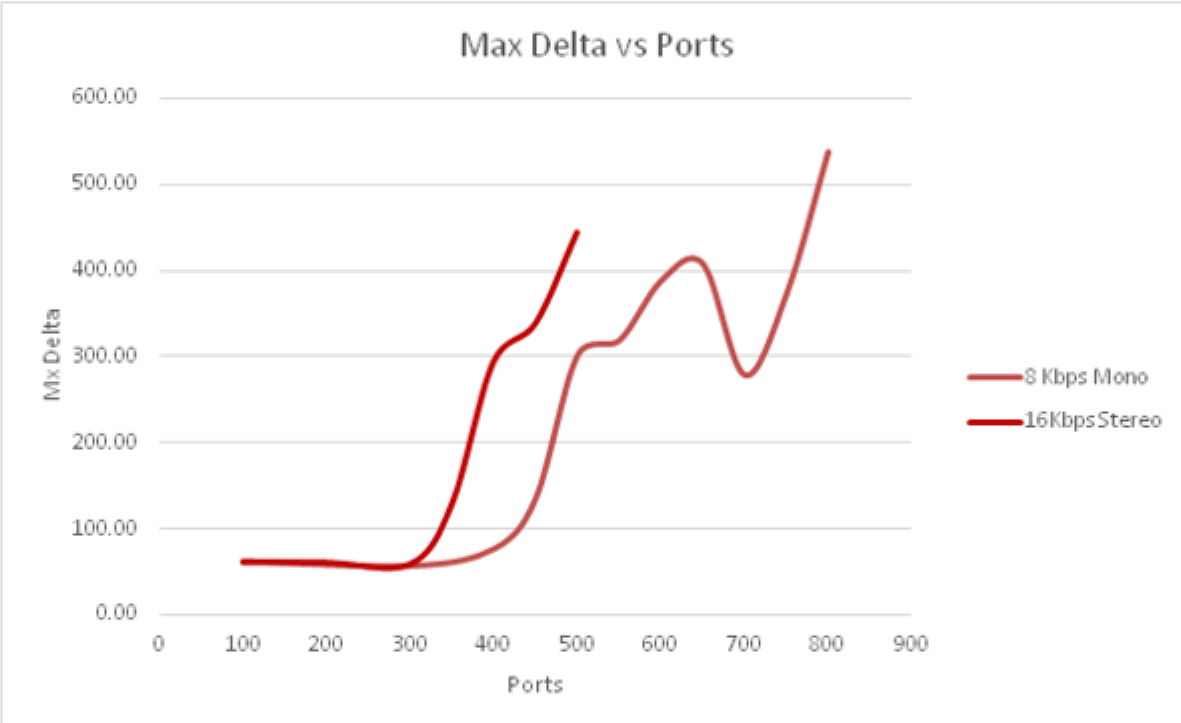


Figure 127: Max Delta, 16Kbps MP3 stereo vs 8Kbps Mono, No encryption, on W2K8 R2

Below are CPU and Memory usage results for a Windows 2008 R2 Server configuration:

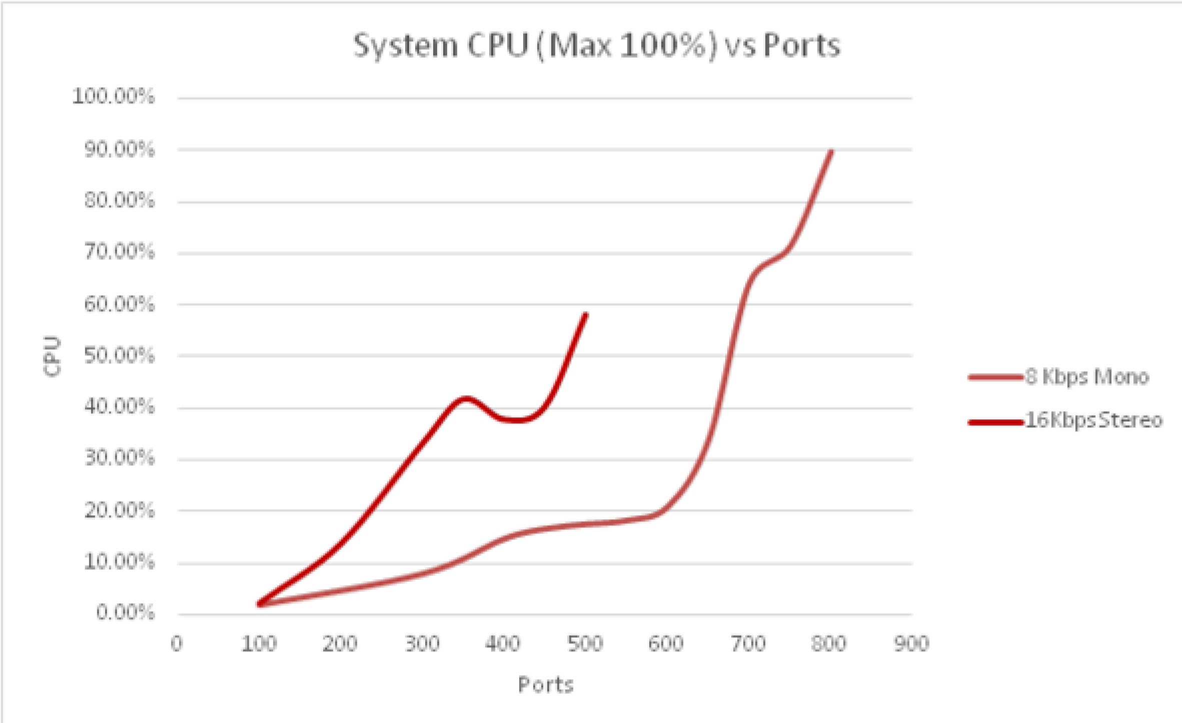


Figure 128: System CPU, 16Kbps MP3 stereo vs 8Kbps Mono, No encryption, on W2K8 R2

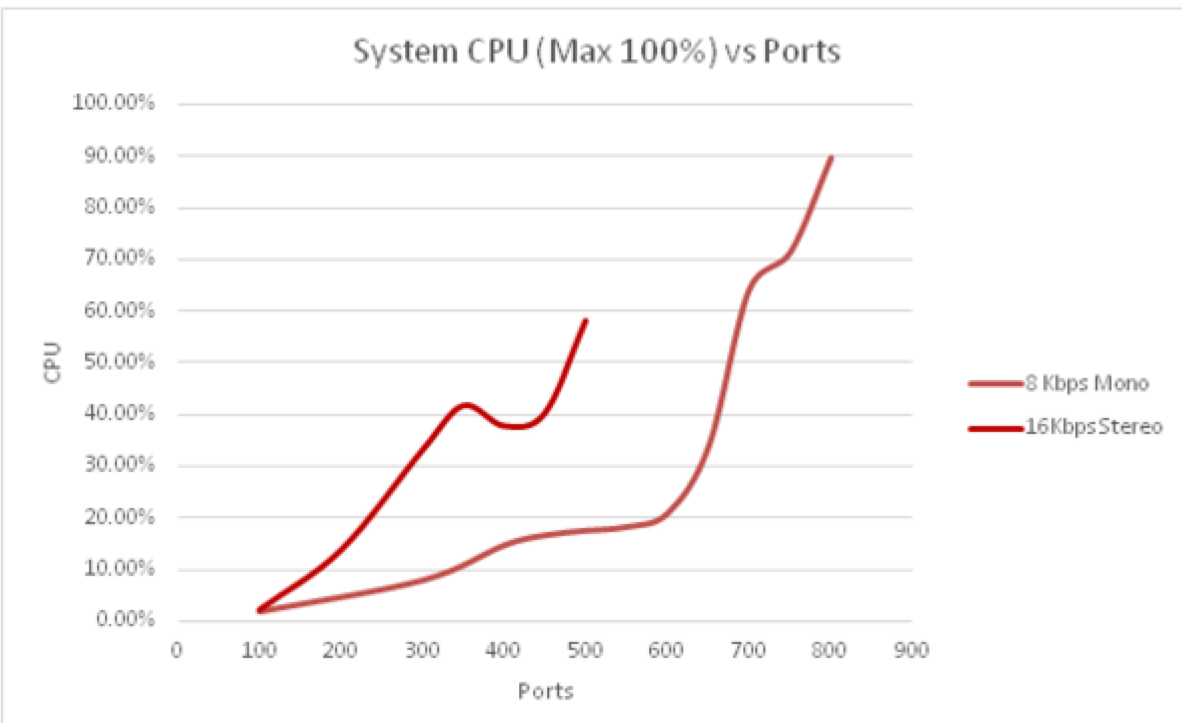


Figure 129: MCP Memory Usage, 16Kbps MP3 stereo vs 8Kbps Mono, No encryption, on W2K8 R2

The above results show that an 8 Kbps configuration uses less CPU and memory, with lower latencies of observed for Max Jitter and Max Delta, which achieves higher capacity. For example, there was a 28.6% increase in capacity (450 vs 350 ports) for preferred ports, and a 33.3% increase in peak capacity (600 vs 450 ports) when using 8 Kbps Mono compared to 16Kbps Stereo.

Figure 130: Disk IOPS, 8 Kbps Mono, No encryption, on W2K8 R2

Ports	Physical Server Disk IOPS		
	Total	**Reads**	**Writes**
100	60.35	0.00	60.35
300	168.82	0.00	168.82
400	224.52	0.00	224.52
450	252.20	0.00	252.20
500	280.28	0.00	280.28
550	306.22	0.00	306.22
600	333.88	0.00	333.88
650	359.84	0.00	359.84
700	388.11	0.00	388.11
750	415.87	0.00	415.87
800	441.96	0.00	441.96

Physical Server Linux EL 6.6 Tests

These tests were performed on a Linux EL 6.6 physical server. The graphs below compare Max Jitter and Max Delta for latencies with an MP3 8Kbps bit rate Mono configuration with using a 16 Kbps MP3 stereo configuration.

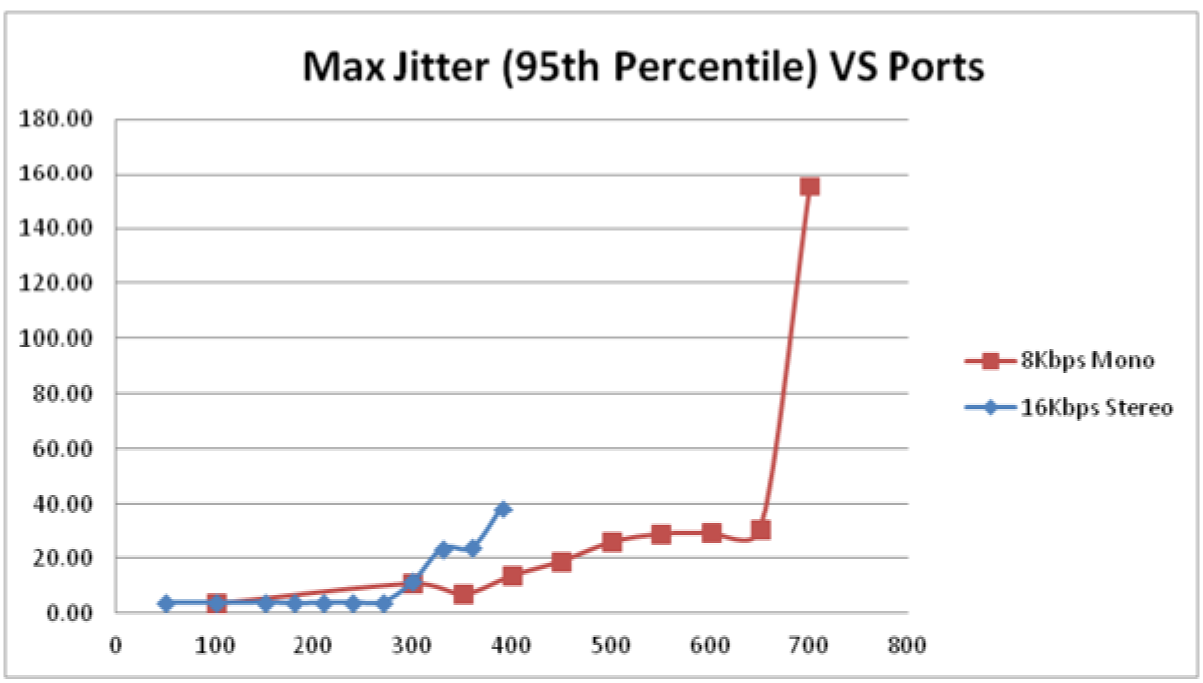


Figure 131: Max Jitter, 16Kbps MP3 stereo vs 8Kbps Mono, No encryption, on EL 6.6

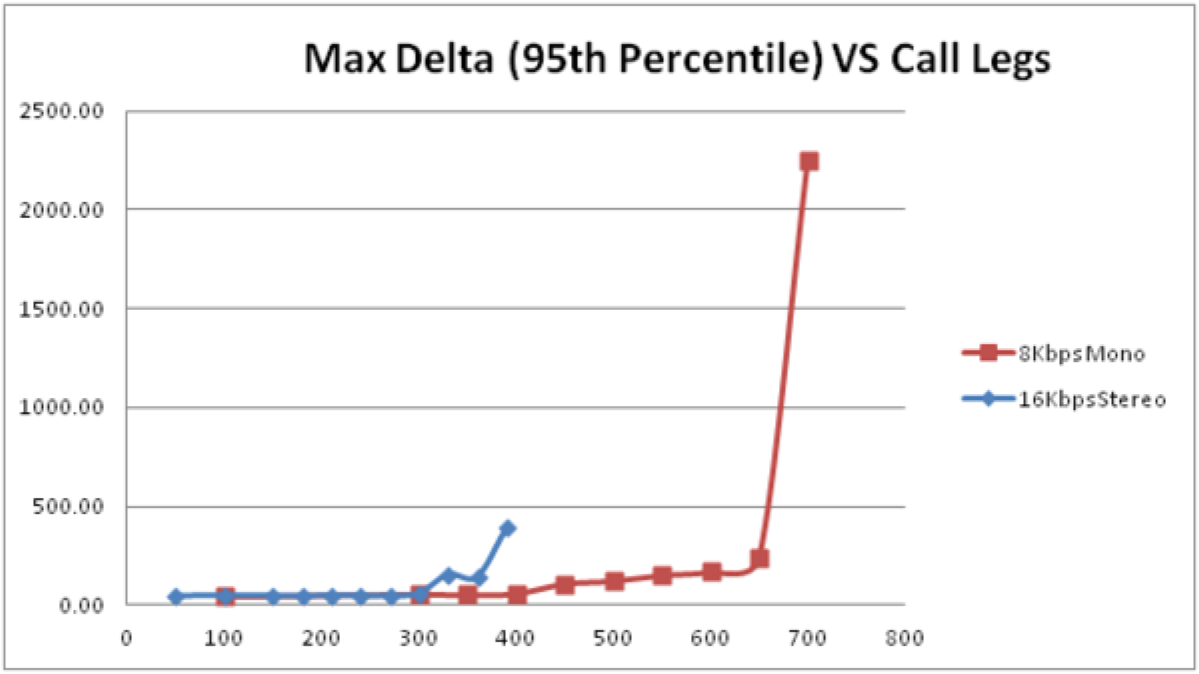


Figure 132: Max Delta, 16 Kbps MP3 stereo vs 8Kbps Mono, No encryption, on EL 6.6

Overall system CPU usage is shown below:

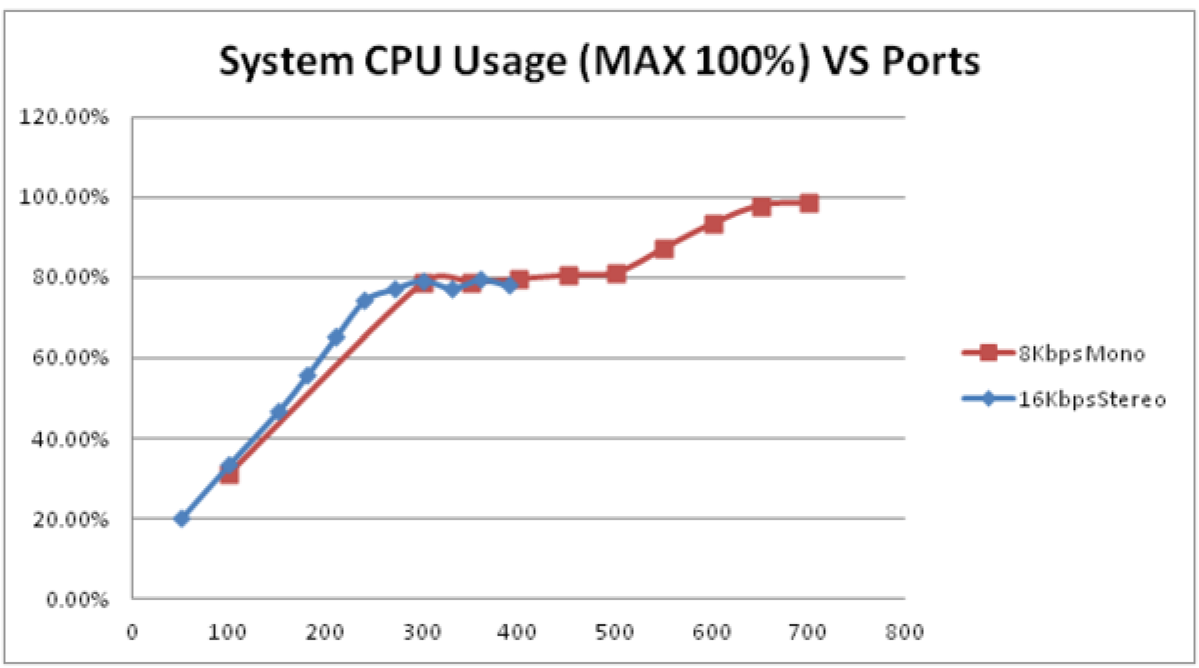


Figure 133: System CPU, 16 Kbps MP3 stereo vs 8Kbps Mono, No encryption, on EL 6.6

The same trend was seen on Linux as was seen on Windows—slightly lower CPU usage for 8 Kbps

Mono, lower latencies for both Max Jitter and Max Delta, thereby achieving higher capacity.

In fact, there is a 100% increase in preferred ports capacity (600 vs 300 ports), and an 80% increase in peak capacity (650 vs 360 ports) when using 8 Kbps Mono on Linux (with no encryption).

Figure 134: Disk IOPS, 8 Kbps Mono, No encryption, on EL 6.6

Ports	Physical Server Disk IOPS		
	Total	**Reads**	**Writes**
100	49.95	0.00	49.95
300	138.42	0.00	138.42
350	157.07	0.00	157.07
400	178.74	0.00	178.74
450	207.88	0.00	207.88
500	225.21	0.00	225.21
550	248.20	0.00	248.20
600	272.85	0.00	272.85
650	299.11	0.00	299.11
700	332.05	0.01	332.04

Performance and Scalability Comparisons

This section compares performance and scalability between GVP 8.x and previous releases, using the profiles VoiceXML_App1 and VoiceXML_App2.

- Performance Comparisons
- Scalability Comparisons
- High Performance Configuration

Performance Comparisons

Tested with VoiceXML_App1	Tested with VoiceXML_App2

Peak capacity of GVP 8.x:
• with NGI:
  • ~50% higher than VG 7.2
  • ~90% higher than GVP 7.6
• with GVPi, equivalent to GVP 7.6
Comparing GVP 8.x and GVP 7.6 (with GVPi) to GVP 8.1:
• 8.1 uses significantly fewer CPU cycles (relatively 30%)
• 8.1 uses less memory (relatively 30%)
In 8.1, the peak capacity is identical to previous releases (using identical temp file management mechanisms), as the bottleneck is due to disk I/O.
	Peak capacity of GVP 8.x:
	• with NGI:
	  • ~66% higher than VG 7.2
	  • ~100% higher than GVP 7.6
	• with GVPi, equivalent to GVP 7.6
	In the use case with GVPi, the peak capacity for GVP 8.x is identical to GVP 7.6 (using identical temp file management mechanisms), because the bottleneck is due to disk I/O.

Scalability Comparisons

For applications that are CPU-dependent (or applications in which bottlenecks occur due to CPU cycles) GVP 8.x can use additional CPU cycles and cores. Use case results showed that peak port densities scaled upward linearly relative to an increase in CPU clock speed.

Table: Examples of Peak Capacity using VoiceXML_App1

Processor	Total Clock Speed	Peak Port Density
2x Core 2 Quad, 2.66 GHz	21.28 GHz	1300
2x Core 2 Dual, 3.00 GHz	12 GHz	700

Processor	Total Clock Speed	Peak Port Density
1x Core 2 Dual, 3.00 GHz	6 GHz	400

Figure: CPU Clock Speed Versus Peak Capacity is a graphical depiction of the peak port density in Table: Call Control Platform Bandwidth Usage.

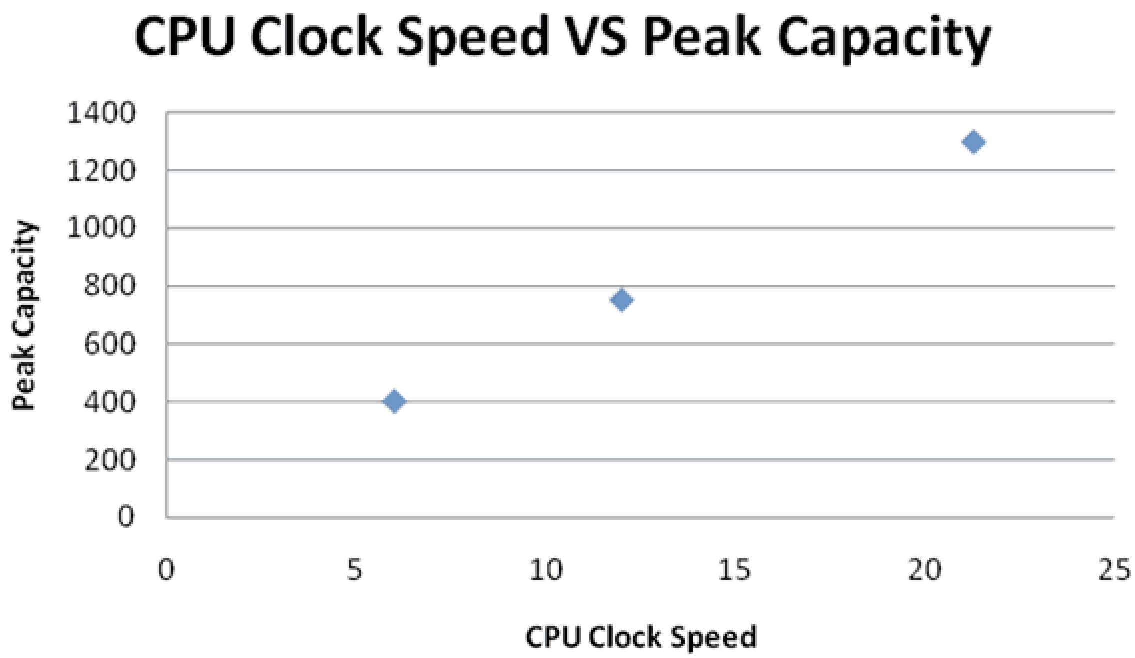


Figure: CPU Clock Speed Versus Peak Capacity

CPU Clock Speed Versus Peak Capacity

To increase the total clock speed by 100%, the peak capacity would have to increase by ~90 to 100%, assuming:

- The type of CPUs are the same as the ones in Table: Call Control Platform Bandwidth Usage.
- The VoiceXML_App1 application is used.
- The overall system bottleneck CPU cycles remain the same.

High Performance Configuration

The Media Control Platform can support more than 400 ports on a single host, however, some configuration changes are required. Use Genesys Administrator to configure the Media Control Platform for high performance by modifying the options and default values in the table below, and configure the Windows Registry on the Media Control Platform to support either the NGI, GVPi, or both.

Table: High Performance Configuration for Media Control Platform

Section	Option/Key	Default Value	High Performance Value
Media Control Platform with NGI			
mpc	maxmediathreads	32	16
vxmli	max_num_documents	5000	10,000 (> 1000 ports)
Windows Registry key: `HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\Tcpip\Parameteres\TcpTimedWaitDelay`		None	Type = DWORD Value = 30 or 1e (hex)
Media Control Platform with GVPi			
mpc	maxmediathreads	32	32
PageCollector	maxpoolthreads	512	>= Port Density
PopGateway1	maxpoolthreads	512	>= Port Density
Windows Registry key: `HKEY_LOCAL_MACHINE\Software\CallNet\CnlnetSettings\MaxThreadPool`		None	Type = DWORD Value >= Port Density /2

top | toc

Application Test Cases

The following application test cases are described in this section:

- NETANN Announcement
- MSML Announcement
- Transcoding
- Secure RTP
- Conference Performance
- HR Timer
- MSML CPA/CPD Performance

NETANN Announcement

When the Media Control Platform acts as a media announcement server, high call rates can be sustained. Using a NETANN announcement application, it can sustain up to 200 CAPS (~1100 ports) for a typical audio playback of 3.3 seconds, however, call setup and tear down latency increases.

Figure: CAPS versus Call Duration (Announcement) shows call durations at various CAPS. When CAPS reaches its peak (200 CAPS), the setup and tear down latency can reach 3.3 seconds. Optimally, call setup and tear down latency should be maintained at <1sec (or 500 ms each) with CAPS at 150 (with 600 ports).

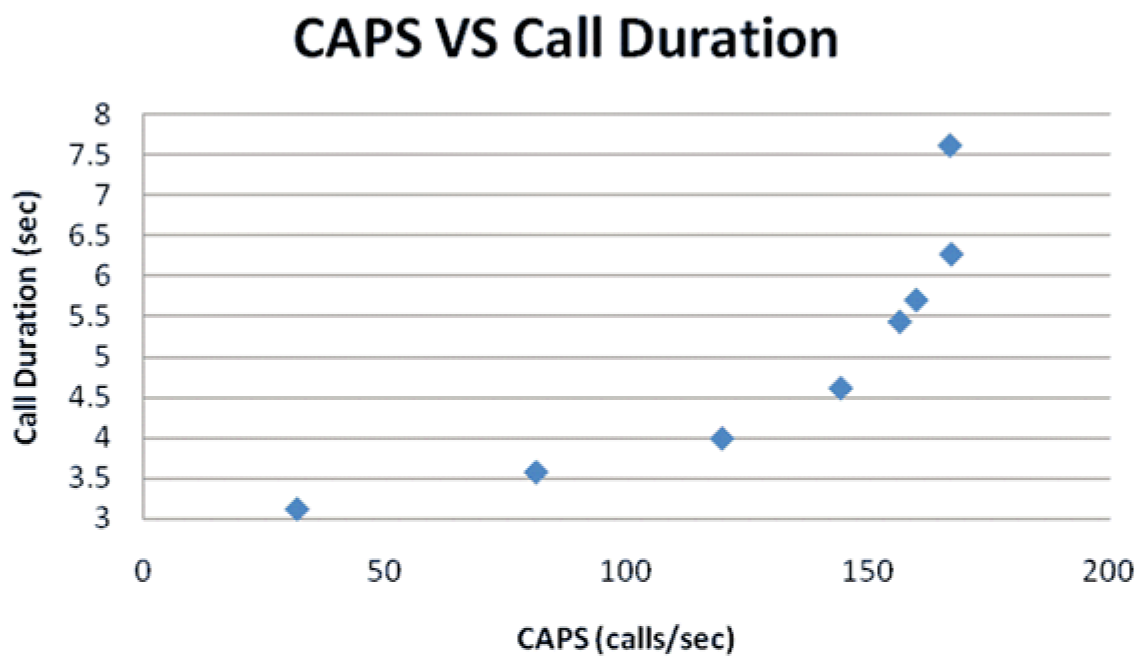


Figure: CAPS versus Call Duration (Announcement)

Figure: Port Density versus CAPS (Announcement) illustrates that, as the call duration increases with higher port density, the additional call setup and tear down latency prevents the CAPS from scaling linearly in relation to the port density.

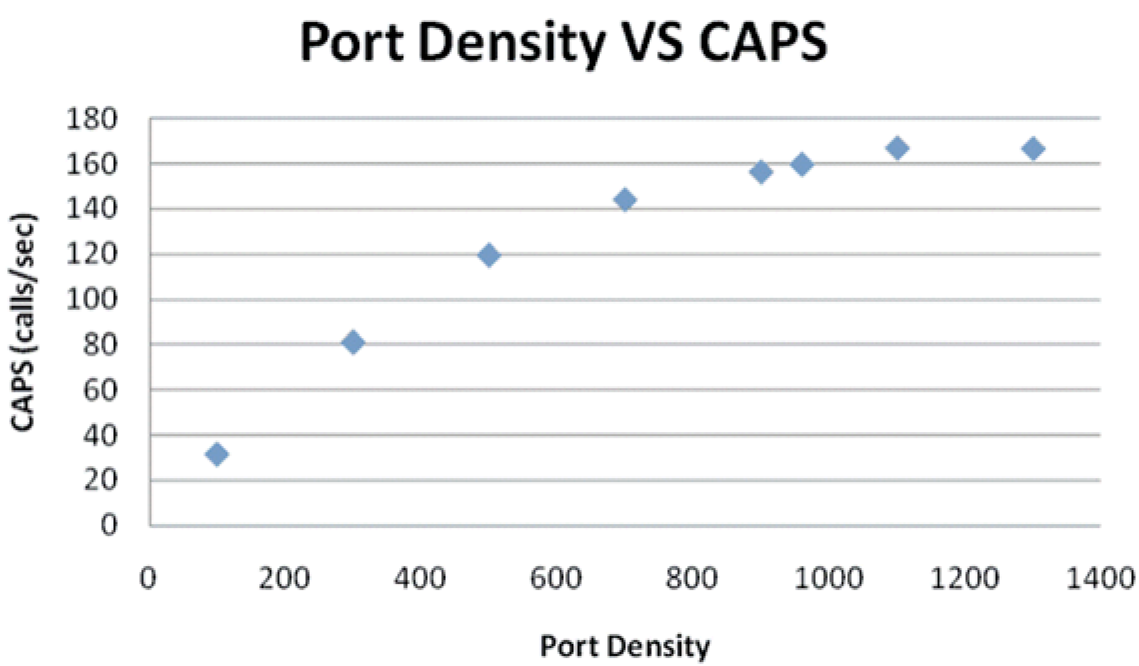


Figure: Port Density versus CAPS (Announcement)

In Figure: CAPS versus Call Duration (Average), a bottleneck is caused by the media streaming. Shorter audio announcements increase the time spent on call setup and tear down and, although the load on the system decreases, shorter audio prompts cause the peak CAPS to increase. The graph depicts a use case where a one-second audio announcement drives the peak CAPS to ~235. Optimally, in this use case, call setup and tear down latency should be maintained at <1sec and CAPS at 200 (with ~500 ports).

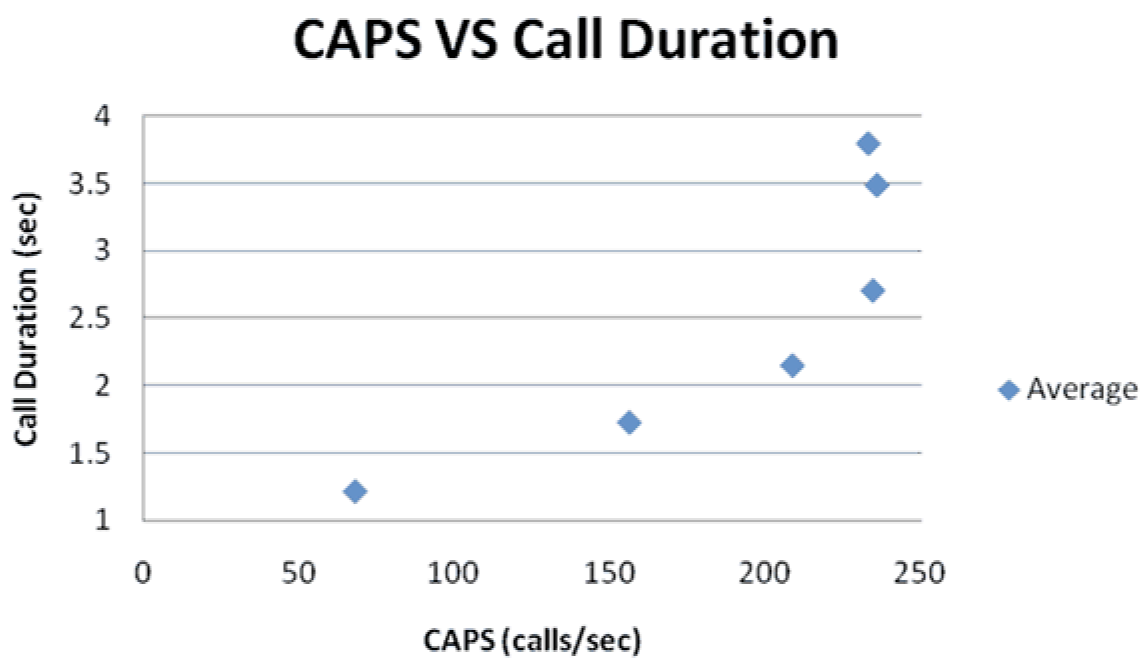


Figure: CAPS Versus Call Duration (Average)

MSML Announcement

MSML announcement applications of 3, 10, and 20 second durations were tested on RHE Linux 5, Update 4, x64. Announcement applications were tested to compare:
1 prompt/1 request vs. **2 prompts/1 request** vs. **2 prompts/2 requests** vs. **3 prompts/3 requests**, in the following scenarios:

MSML Announcement - 3 second duration
1 audio file with 1 prompt (SIP INFO)

MSML Announcement - 10 second duration
1 audio file (10s) with 1 prompt (SIP INFO)
2 audio files (4s and 6s); 1 prompt (SIP INFO)
2 audio files (per prompt); 2 prompts (SIP INFO)

MSML Announcement - 20 second duration

1 audio file (20s) with 1 prompt (SIP INFO)
3 audio files (4s, 6s, 10s) with 1 prompt (SIP INFO)
3 audio files (per prompt) with 3 prompts (SIP INFO)

A 3-second audio file, with a single prompt (SIP INFO) and `gvp:precheck` turned on, resulted in a peak capacity of 80 CAPS or 260 ports, which is lower than the NETANN preferred capacity of 120 CAPS or 500 ports. Figure: CAPS Versus Port Density—NETANN and MSML Announcement (Linux) provides a comparison of the test scenarios.

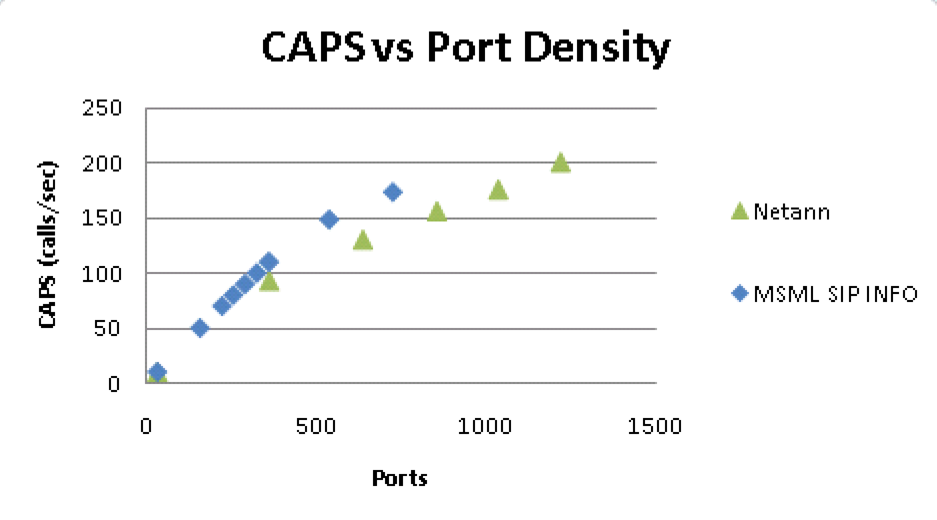


Figure: CAPS versus Port Density—NETANN and MSML Announcement (Linux)

In Figure: CAPS versus Port Density—NETANN and MSML Announcement (Linux) (above), testing starts to fail beyond 80 CAPS in the MSML test case. The call duration deviation is better than in the NETANN test case. See Figure: Port Density versus ACD—NETANN and MSML Announcement (Linux) (below).

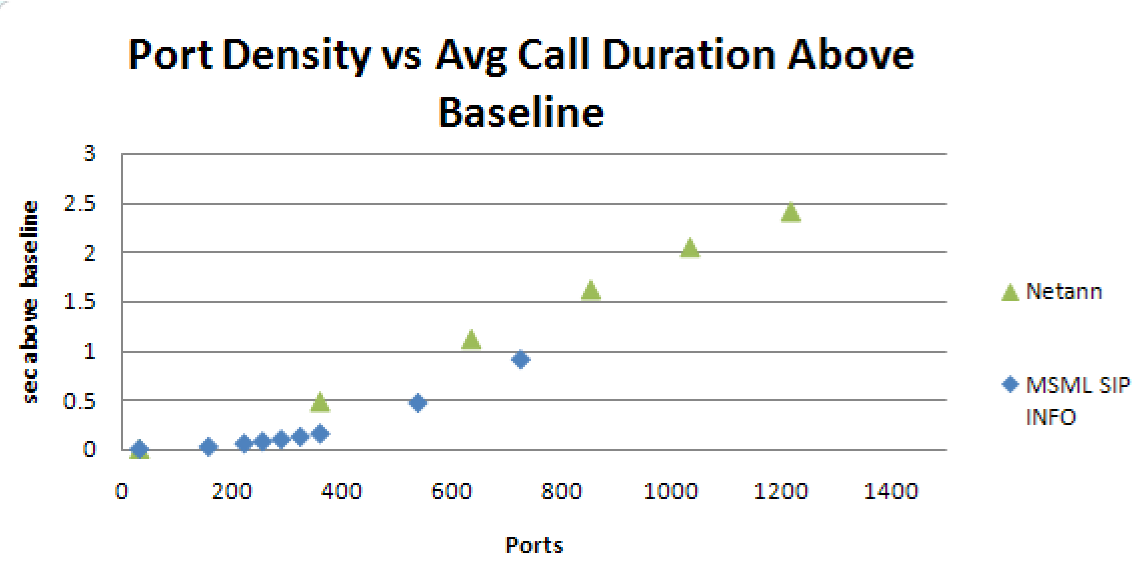


Figure: Port Density versus ACD—NETANN and MSML Announcement (Linux)

Overall system CPU usage is illustrated in the graph in Figure: Port Density versus System CPU Usage—MSML and NETANN (Linux). Overall CPU usage is quite similar, but MSML test case is slightly higher than NETANN at high ports, which is beyond peak capacity.

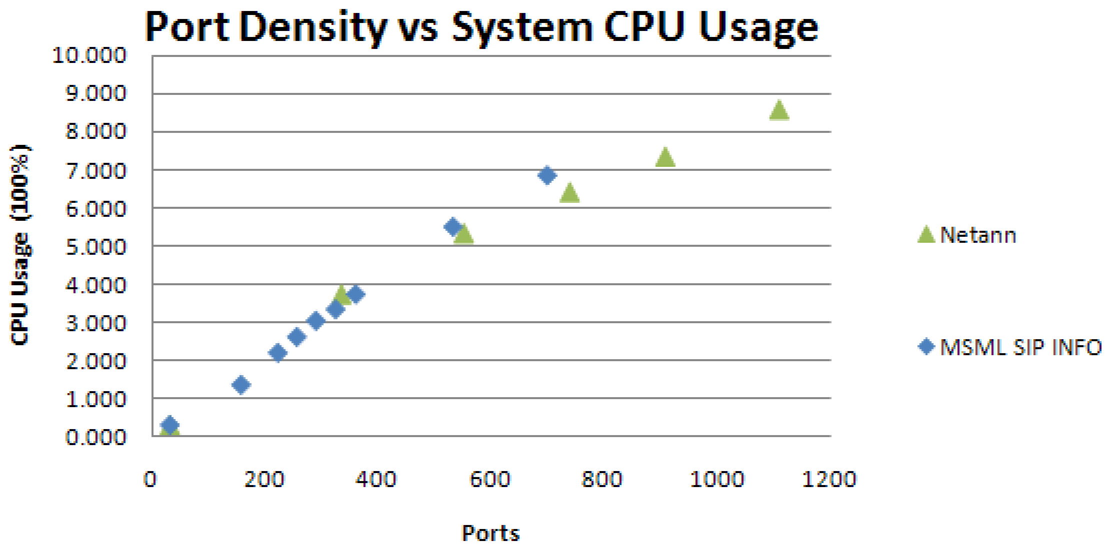


Figure: Port Density versus System CPU Usage—MSML and NETANN (Linux)

As indicated by the graph in Figure: Port Density versus ACD—MSML Application (Linux), performance for the 10-second announcement application when gvp:precheck is turned off, is almost the same with 1 or 2 audio files in a single prompt (200 CAPS or 2000 ports) while two prompts (SIP INFO) only achieve 130 CAPS or 1400 ports.

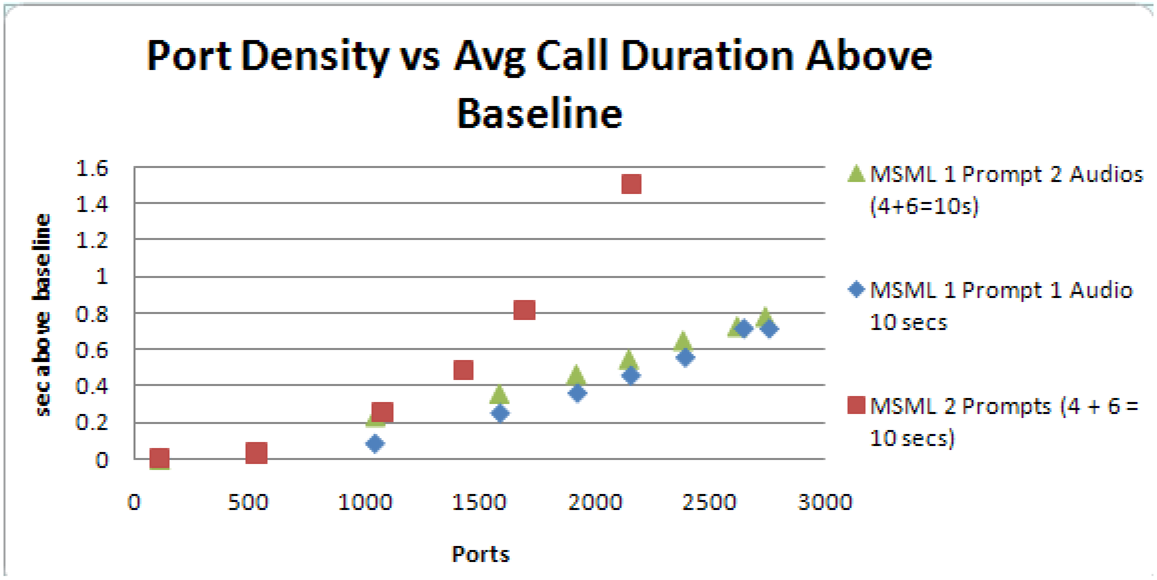


Figure: Port Density versus ACD—MSML Application (Linux)

In Figure: Port Density versus ACD—MSML Application (Linux) The performance of 2 single prompt test case results are quite similar, while the call duration increases significantly for the 2 prompts scenario. There is some trending in the overall CPU usage in Figure: Port Density versus MCP CPU Usage—MSML Application (Linux).

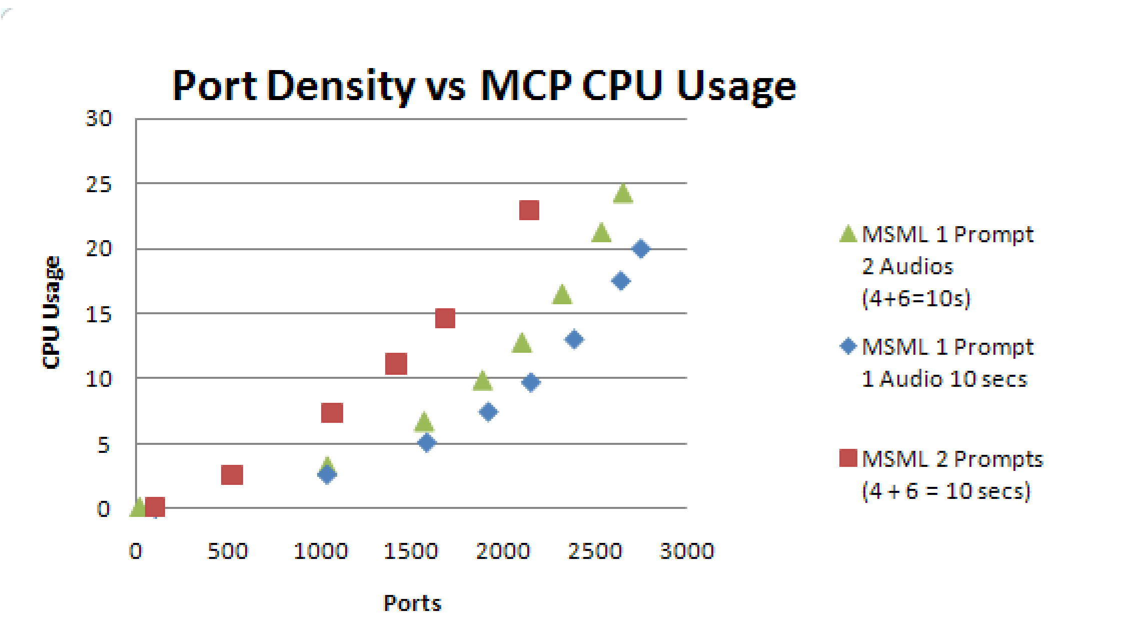


Figure: Port Density versus MCP CPU Usage—MSML Application (Linux)

In the 2 prompts test case in Figure: Port Density versus MCP CPU Usage—MSML Application (Linux), the CPU usage is significantly higher than in the 2 single-prompt test.

In a 20-second announcement scenario with `gvp:precheck` turned off, when there are more audio files, even in a single prompt, performance is impacted: 130 CAPS or 2600 ports for three audio files in a single prompt versus 150 CAPS or 3000 ports for one audio file in a single prompt.

Multiple prompts may also downgrade the performance, in that only 100 CAPS, or 2000 ports are achieved in the 3 prompts test case (3 SIP INFO messages).

MSML Announcement Latency

Latencies for MSML Play Announcement are measured using the following metrics: call setup latency (from INVITE to first RTP packet), SIP INFO (with MSML) response latency and Inter-prompt latency. The background load is a prompt (G.711), audio-only, lasting 120 seconds.

Below are two graphs for call setup latency of MSML Play Announcement on a four-VM setup of vSphere using two Xeon E5620 CPUs (eight cores) with 16GB RAM. Each VM uses two vCPUs and 4GB RAM. The guest OS is Windows 2008 Server R2 Enterprise. The graph in Figure:CAPS vs. Call Setup Latency shows the latency (in milliseconds) based on call rate and the other shows port density. A small jump in latency occurs when the load goes above 50 cps or 6000 ports. The latency is still below acceptable criteria (500 ms). The overall CPU usage approaches 70% when the CAPS rate is 50.

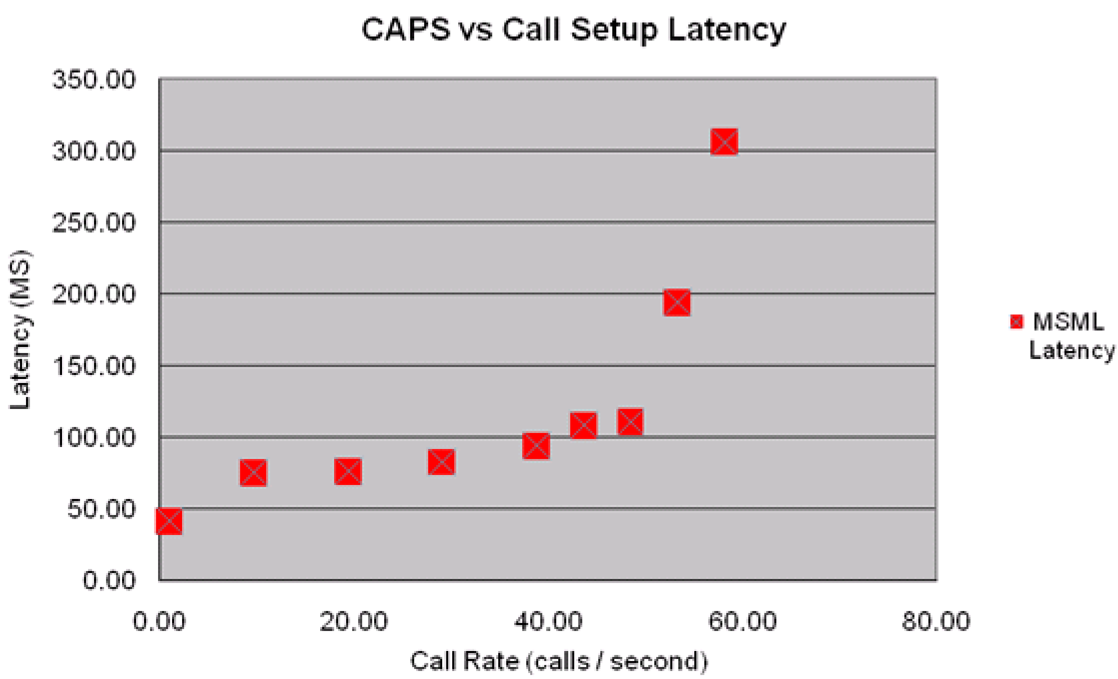


Figure:CAPS vs. Call Setup Latency

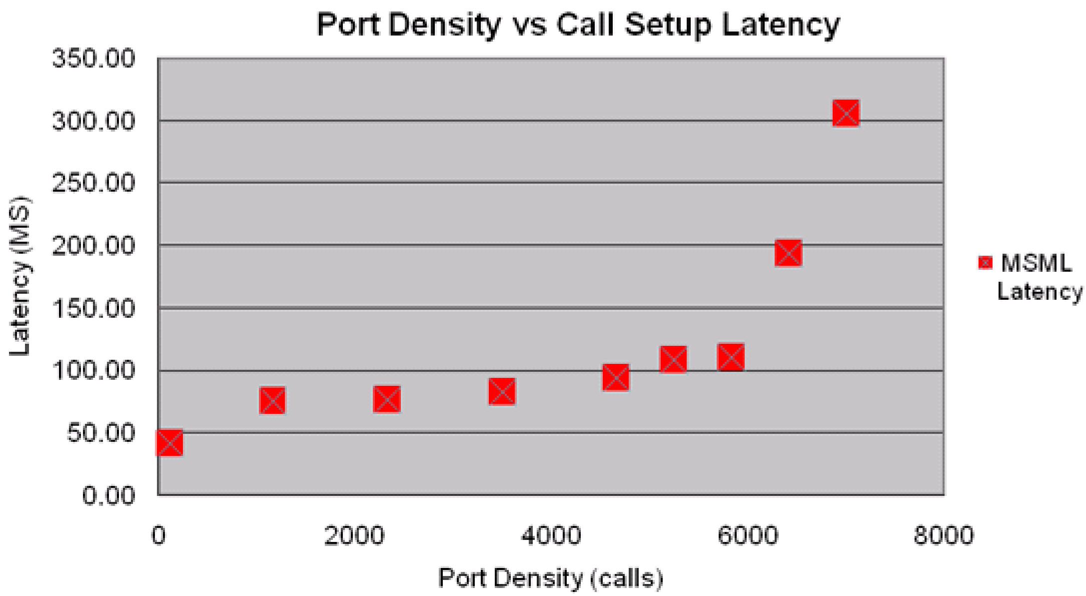


Figure: Port Density vs. Call Setup Latency

Figure: Inter-Prompt Latency vs. Ports, displays a small jump in inter-prompt latency (using audio files) when the load goes beyond 6000 ports.

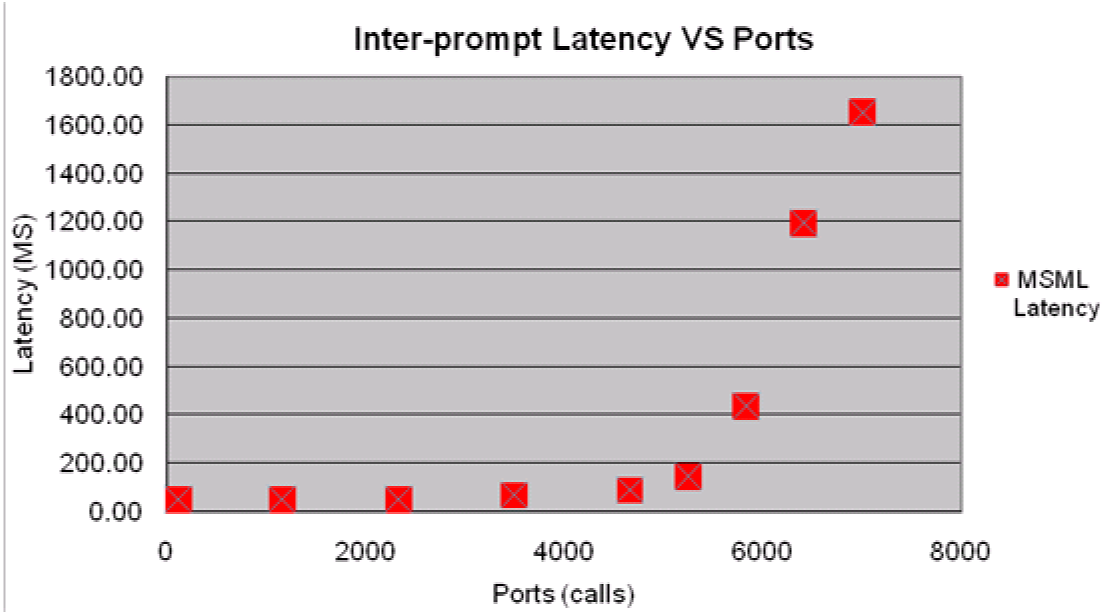


Figure: Inter-Prompt Latency vs. Ports

Figure: SIP INFO -> 200OK Latency vs Ports shows the SIP INFO (with MSML embedded) response (200 OK) latency.

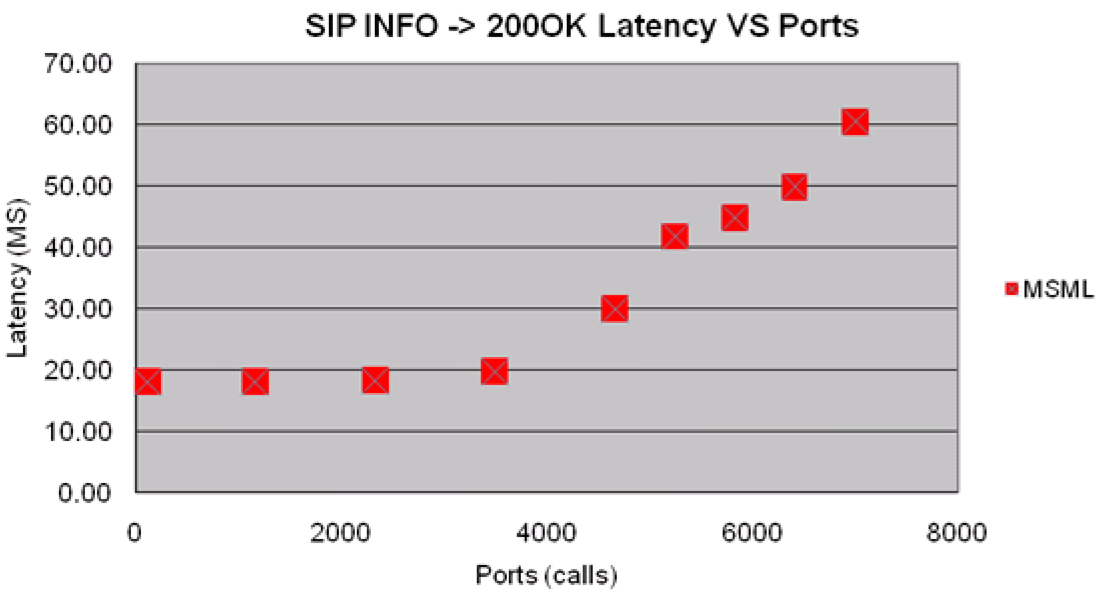


Figure: SIP INFO -> 200OK Latency vs Ports

MSML Video Play Performance

Several factors can affect the performance of Video Play using MSML Announcements such as resolution, bit rate, and frame rate. For H.263 video (tested using the AMR audio codec), the following tests were chosen for analysis and comparison:

- CIF, 512 Kbps bit rate (high), 30 fps (high)
- CIF, 128 Kbps bit rate (low), 10 fps (low)
- 4CIF, 2 Mbps bit rate (high), 30 fps (high)
- 4CIF, 512 Kbps bit rate (low), 10 fps (low)

Tests were conducted using three VMs under vSphere 5.0 on a single hex-core machine using Xeon X5670. Each VM was assigned two vCPUs with only one MCP installed on each VM. The results of each test follow.

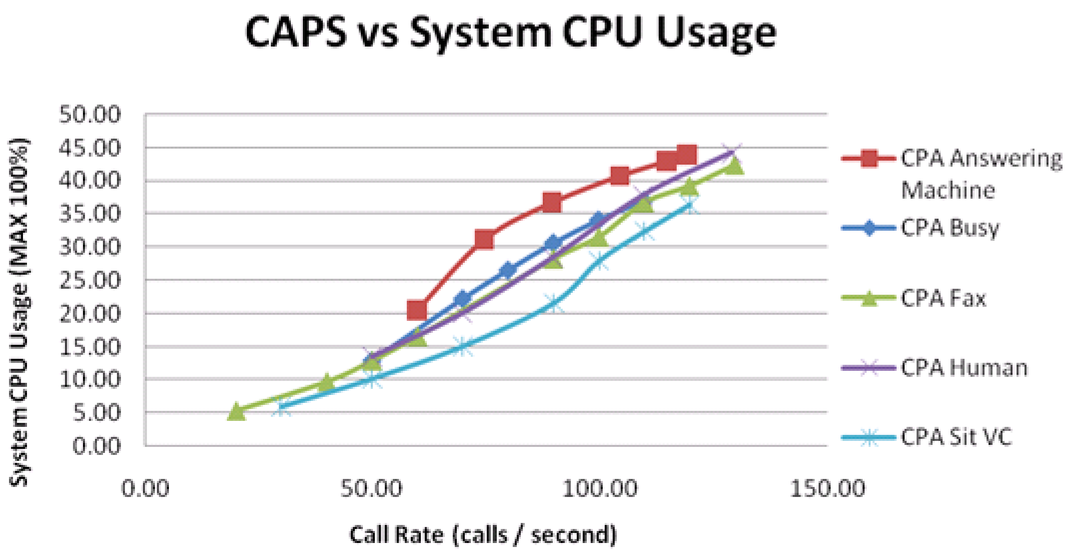


Figure: CAPS vs. System CPU Usage 1

In CAPS vs. System CPU Usage 2 and CAPS vs. System CPU Usage 3 (below), the dominant factor of peak capacity is the frame rate, while the impact from both bit rate and resolution is small. The CPU is the apparent bottleneck in these tests, meaning that additional capacity can be reached with more powerful CPUs.

For H.264 video (with AMR audio), using finer granularity resulted in more groups of combinations:

- CIF, 512 Kbps bit rate (high), 30 fps frame rate (high)
- CIF, 256 Kbps bit rate (middle), 15 fps frame rate (middle)
- CIF, 128 Kbps bit rate (low), 10 fps frame rate (low)

- 4CIF, 2 Mbps bit rate (high), 30 fps frame rate (high)
- 4CIF, 1 Mbps bit rate (middle), 15 fps frame rate (middle)
- 4CIF, 512 Kbps bit rate (low), 10 fps frame rate (low)

- 720P, 4 Mbps bit rate (high), 30 fps frame rate (high)
- 720P, 2 Mbps bit rate (middle), 15 fps frame rate (middle)
- 720P, 1 Mbps bit rate (low), 10 fps frame rate (low)

A similar trend is evident when testing H.264. The first of the following graphs shows how varying bit rate and frame rate, while keeping the resolution constant (4CIF), affects CPU usage. The second graph shows how varying the resolution and bit rate, while keeping the frame rate constant (15 fps), affects CPU usage.

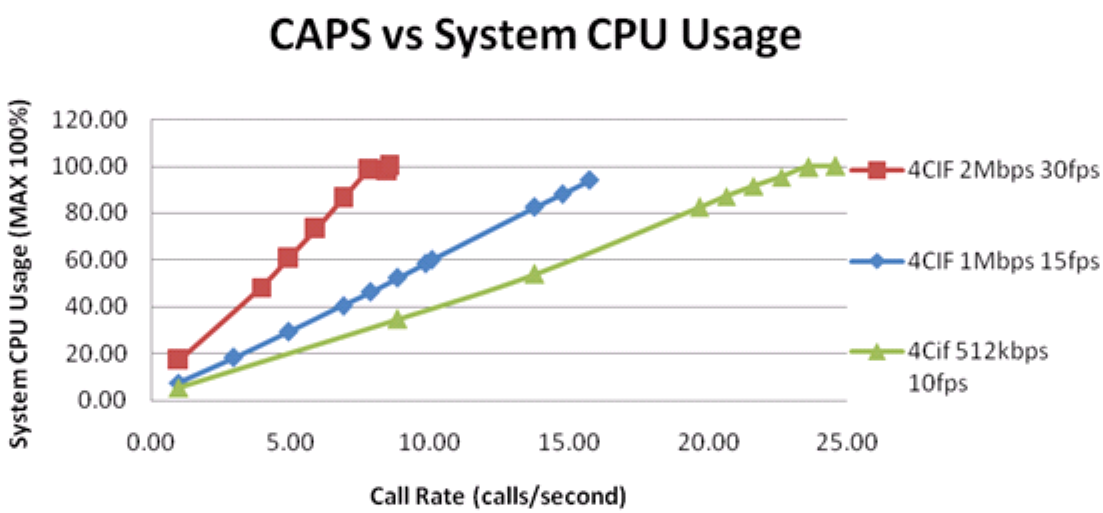


Figure: CAPS vs. System CPU Usage 2

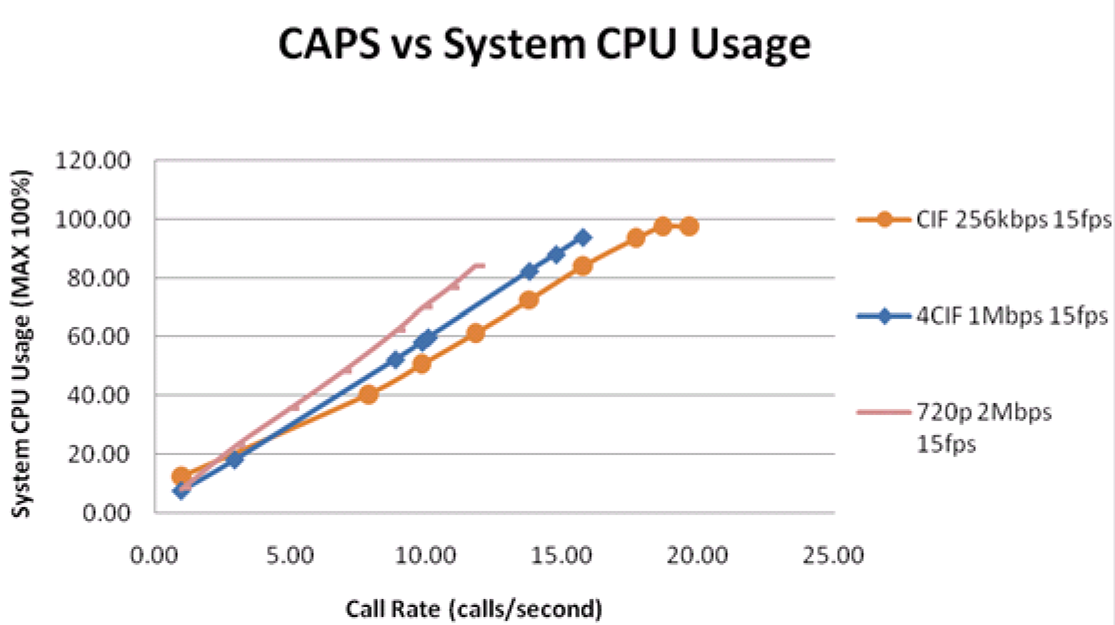


Figure: CAPS vs. System CPU Usage 3

Keeping the resolution constant and varying the frame rate and bit rate causes larger variations in CPU usage and peak capacity, while keeping the frame rate constant and varying the resolution and bit rate does not.

MSML MP3 Play & Recording Performance

MP3 playback using MSML Announcements was tested using the G.711 ulaw audio codec on the RTP channel. The MCP was required to transcode from MP3 to G.711 ulaw during this test. Two types of MP3 files were used in these tests:

- 96K bit rate, 32KHz sampling stereo
- 320K bit rate, 44.1KHz sampling stereo

Testing was conducted with three VMs running under vSphere 5.0 on a single hex-core Xeon X5670 processor machine. Each VM was assigned two vCPUs and each ran a single MCP instance.

Transcoding involves additional CPU resources. Lower bit rate and lower sampling rates will use fewer CPU resources and achieve a higher peak capacity. The chart CAPS vs. System CPU Usage 4 depicts overall CPU usage vs. call rate for the above mentioned MP3 files:

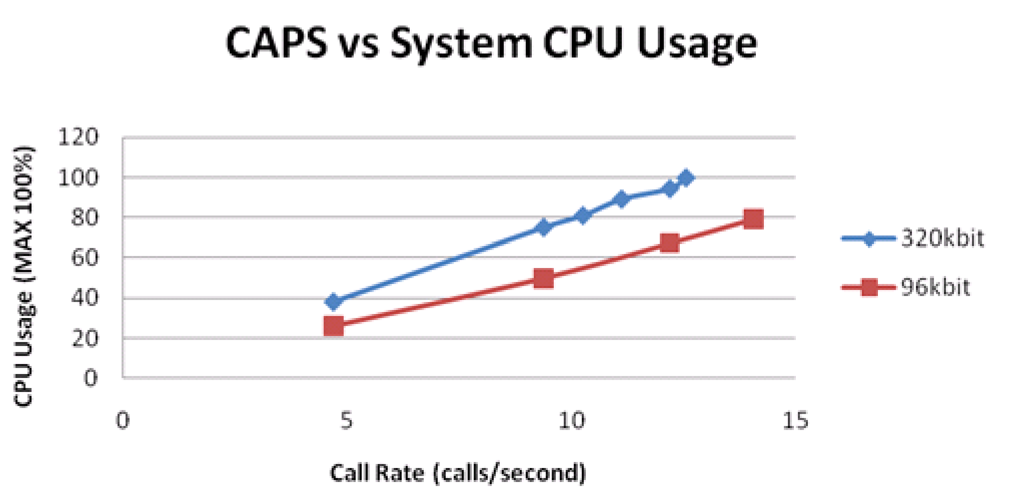


Figure: CAPS vs. System CPU Usage 4

An MP3 recording using the same two bit rates resulted in a lower peak capacity, since a disk speed bottleneck was reached before CPU saturation occurred.

Transcoding

The Media Control Platform can transcode various media codecs in real time. The impact on performance from transcoding overhead varies, depending on the codec that the Media Control Platform is transcoding to or transcoding from. Other variables that contribute to transcoding overhead are the number of audio prompts played by GVP and the amount of customer input received.

The worst case scenario occurs when the Media Control Platform is constantly transcoding between two codecs during the entire call. (Most VoiceXML applications require minimal encoding [G711u to AMR]). In Figure: Port Density versus CPU Usage (G711u and G711a), the least amount of transcoding overhead is between G711u and G711a codecs, where the peak capacity drops by ~25%.

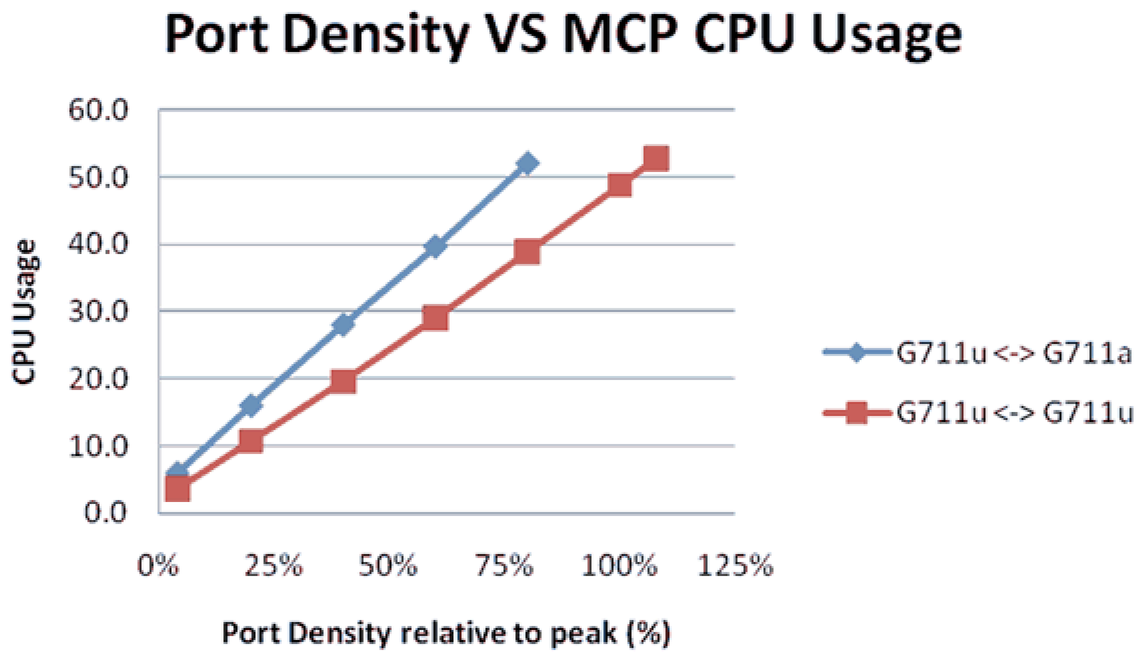


Figure: Port Density versus CPU Usage (G711u and G711a)

Figure: Port Density versus CPU Usage (G711u and AMR) illustrates the impact of transcoding overhead. It is greater between the G711u and AMR codecs, where the peak capacity drops by ~75%.

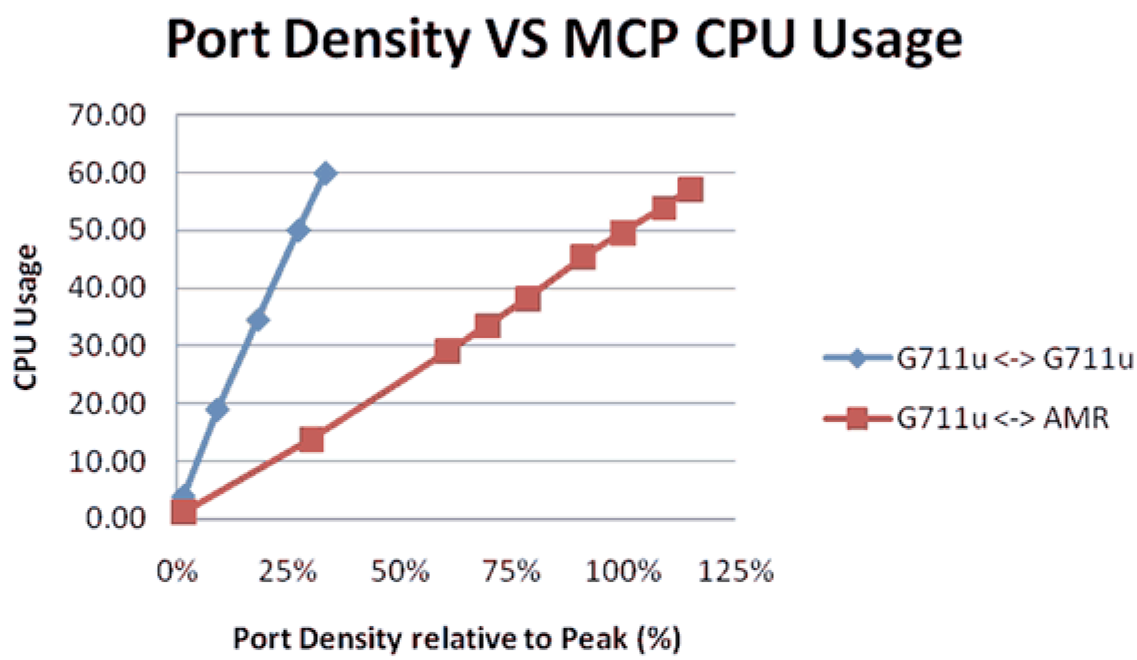


Figure: Port Density versus CPU Usage (G711u and AMR)

The transcoding test cases in this section depict a worst case scenario involving constant transcoding between two codecs. However, in a call flow scenario where the audio stream was suppressed or silent, and the application was waiting for user input 50% of the time, transcoding overhead would be reduced by 50%.

Video Transcoding Performance

Beginning with release 8.1.5, the MCP is capable of performing video transcoding. Video Transcoding Performance was measured using bridge transfers containing video and audio streams. Video transcoding requires more CPU resources than audio-only transcoding. A video bridge transfer without transcoding can achieve hundreds of ports on a machine containing 3 VMs running on vSphere 5.0, with a single hex-core Xeon X5675 processor. On the same machine, video bridge transfers that involve video transcoding can range from single digit port numbers to a few dozen. Peak capacity is affected by resolution, frame rate and bit rate.

Transcoding was tested with video codec H.264 and AMR audio codec. The testing was divided into groups that consist of resolution downscaling, frame rate downscaling and bit rate downscaling. Figure: Port Density vs. System CPU 1 displays the group of resolution downscaling from VGA to CIF and QCIF, respectively, with the same frame rate of 30. It shows the transcoding performance drops up to ~80%.

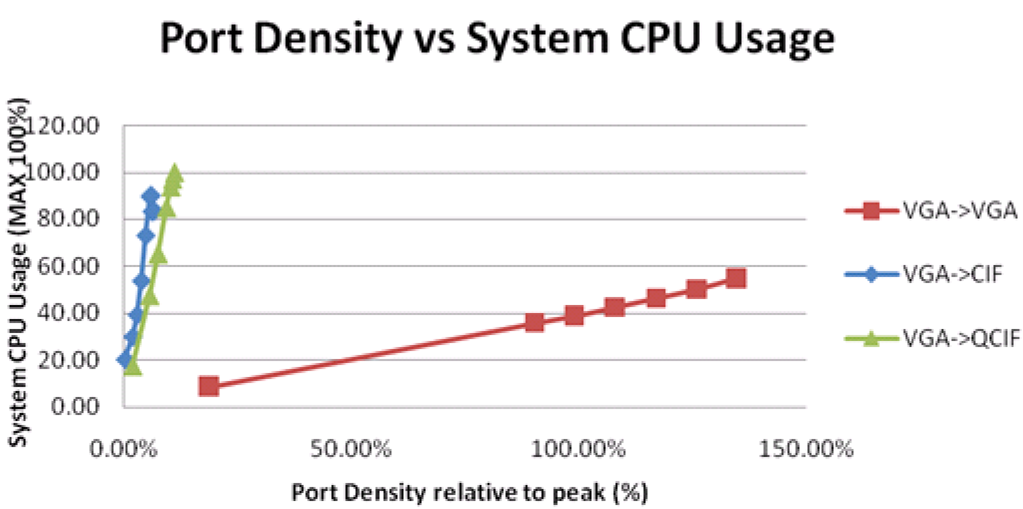


Figure: Port Density vs. System CPU 1

Figure: Port Density vs. System CPU 2 displays the focus for transcoding only:

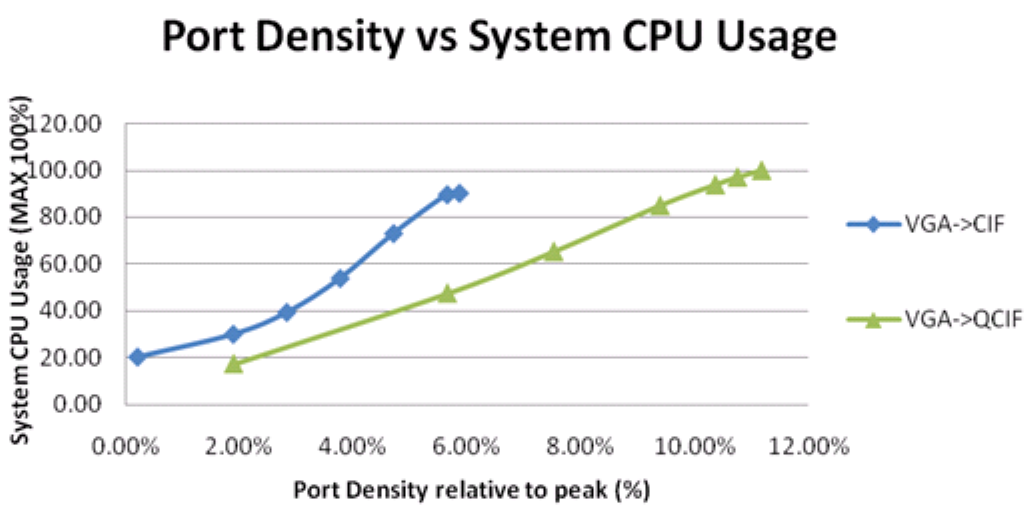


Figure: Port Density vs. System CPU 2

With higher resolution, downscaling transcoding of the performance drops even further. Figure: Port Density vs. System CPU 3 is a graph for transcoding from 720P to CIF and QCIF. Performance drops up to ~90% for VGA to QCIF while ~95% for VGA to CIF transcoding.

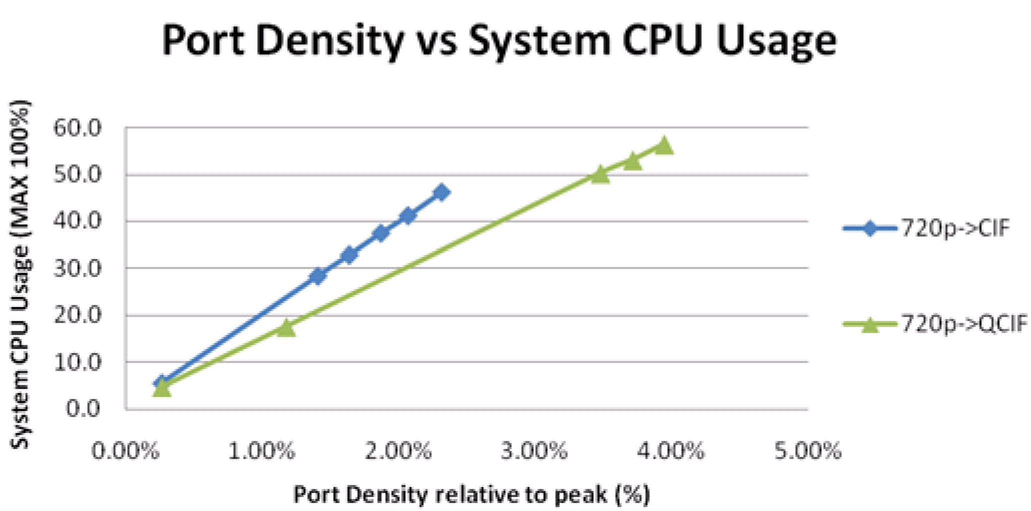


Figure: Port Density vs. System CPU 3

in Figure: Port Density vs. System CPU 4, performance dropped more than 95% when tested with a downscaled frame rate and the same resolution (VGA).

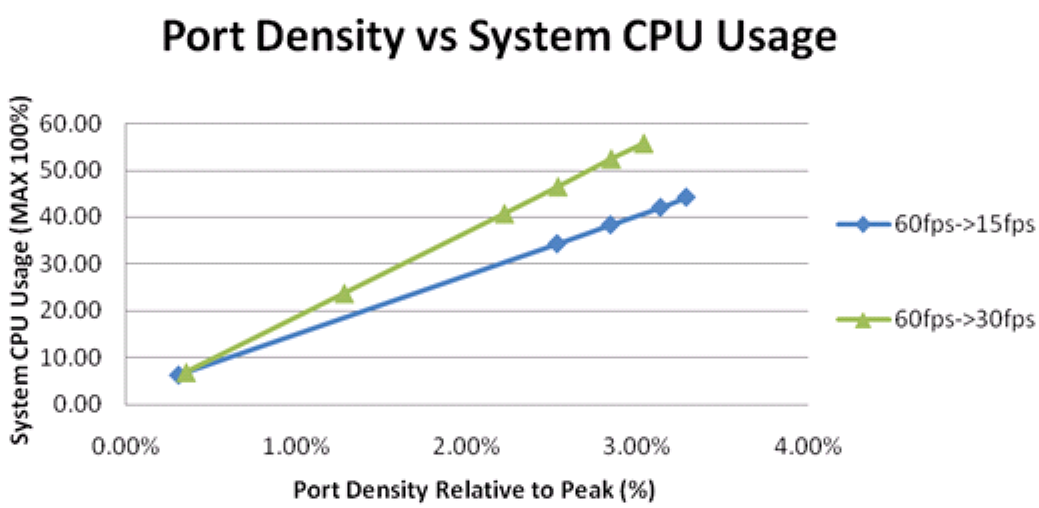


Figure: Port Density vs. System CPU 4

Secure RTP (SRTP)

Secure Real Time Protocol performance testing was conducted by using two bridge transfer scenarios with unidirectional RTP/SRTP streams; one included PCMU audio-only, and the other, a 3gp container of H.264 video plus AMR audio. The PCMU audio-only transfer was tested on both Windows and Linux, while the video plus audio transfer was tested on Linux only.

Tests were conducted with SRTP in the following scenarios (and one with RTP-only) to provide comparison of results:

- Baseline of RTP-only (without SRTP)
- Default SRTP mode (encrypted and authenticated) encryption
- Default SRTP mode (encrypted and authenticated) decryption
- Unencrypted SRTP mode (authenticated only) encryption
- Unencrypted SRTP mode (authenticated only) decryption
- Unauthenticated SRTP mode (encrypted only) encryption
- Unauthenticated SRTP mode (encrypted only) decryption

The test results suggest that peak capacity is almost the same for SRTP and RTP, regardless of the SRTP mode used. The audio-only tests resulted in 1200 ports achieved on Windows and 1500 ports on Linux, and 400 ports for the video + audio test case (on Linux only).

Capacity results were also the same regardless of SRTP mode. However, CPU utilization results varied.

Figure: System CPU Usage versus Port Density—Test Cases 1, 2, 3 depicts the audio-only test case on

Windows, comparing CPU usage in testing scenarios 1, 2, and 3.

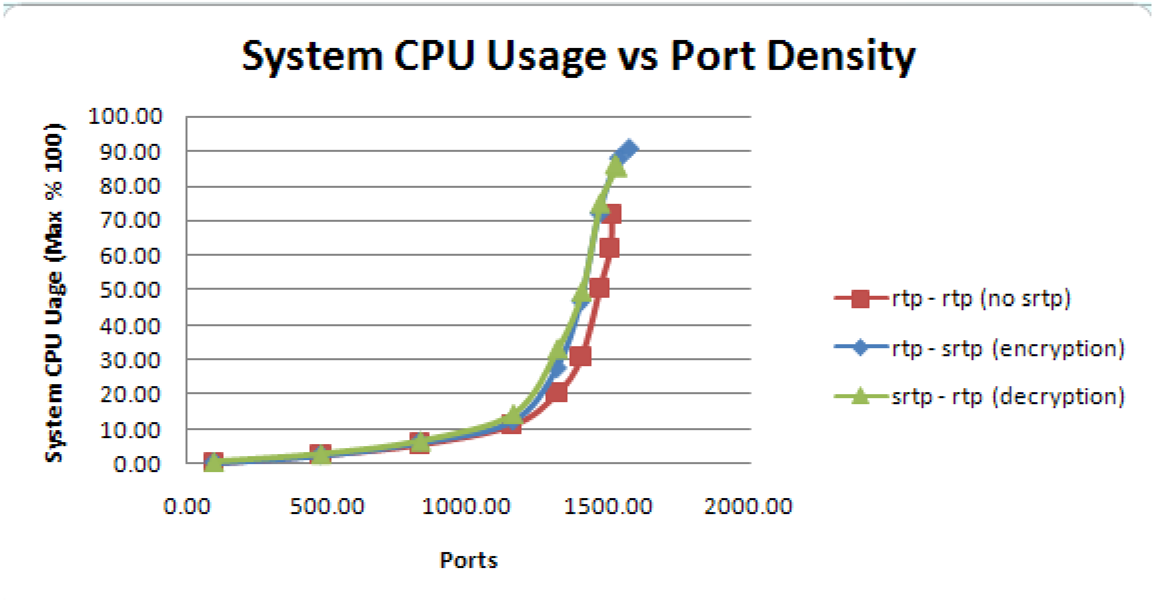


Figure: System CPU Usage Versus Port Density—;Test Cases 1, 2, 3

Figure: System CPU Usage Versus Port Density—Test Cases 2, 4, 6 depicts the audio-only test case on Windows, which compares CPU usage in testing scenarios 2, 4, and 6.

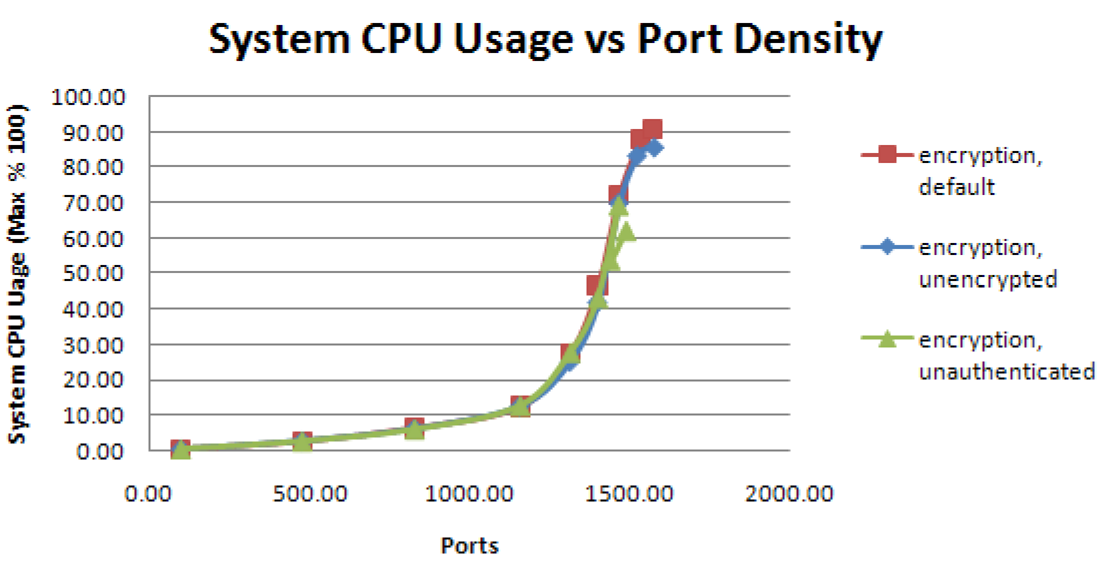


Figure: System CPU Usage versus Port Density—Test Cases 2, 4, 6

In Figure: System CPU Usage Versus Port Density—Test Cases 2, 4, 6, CPU usage increases beyond 1200 ports (launching 1300 ports), meaning the peak capacity is also 1200 ports. The following additional results were observed:

- In the encryption scenario, the overall system CPU usage increased from 11.4% to 12.5%—a 10%

increase at 1200 ports.

- In the decryption scenario, the overall system CPU usage increased from 11.4% to 14.4%—a 26% increase at 1200 ports.

The difference in CPU usage is negligible whether SRTP is configured in default (encrypted and authenticated), unencrypted, or unauthenticated mode.

In Figure: System CPU Usage versus Port Density—Audio-Only Test Case, the audio-only test case (on Linux), the CPU usage is more linear than on Windows, therefore, 1500 ports (launching 1700 ports) is considered peak capacity. The following additional results were observed:

- In the encryption scenario, the overall CPU usage increased from 22.8% to 33.1%—a 45% increase.
- In the decryption scenario, the overall CPU usage increased from 22.8% to 31.4%—a 38% increase.

The difference in CPU usage is negligible whether SRTP is configured in default (encrypted and authenticated), unencrypted, or unauthenticated mode.

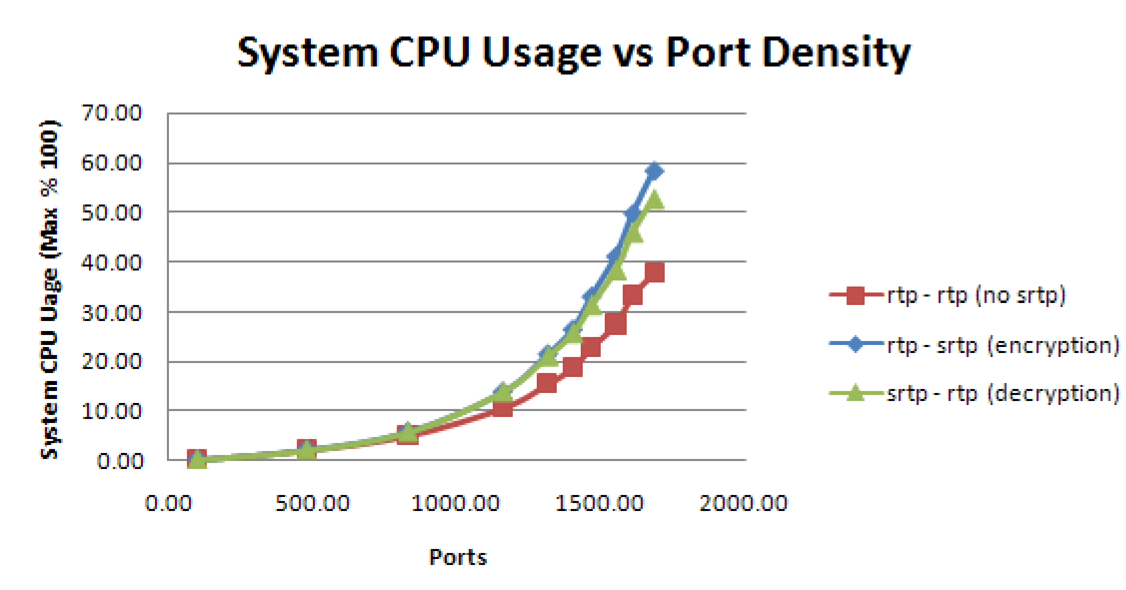


Figure: System CPU Usage versus Port Density—Audio-Only Test Case

In Figure: MCP Memory Usage versus Port Density, in the video + audio test case, memory usage is causing the bottleneck. The graph depicts a comparison of virtual memory usage when default encryption, default decryption SRTP mode, and an RTP-only scenario is tested. All of these test case results approach the 3GB limit when ports reach 400. Even in the RTP-only test case, the virtual memory is only slightly lower. Therefore, 400 ports is considered peak capacity.

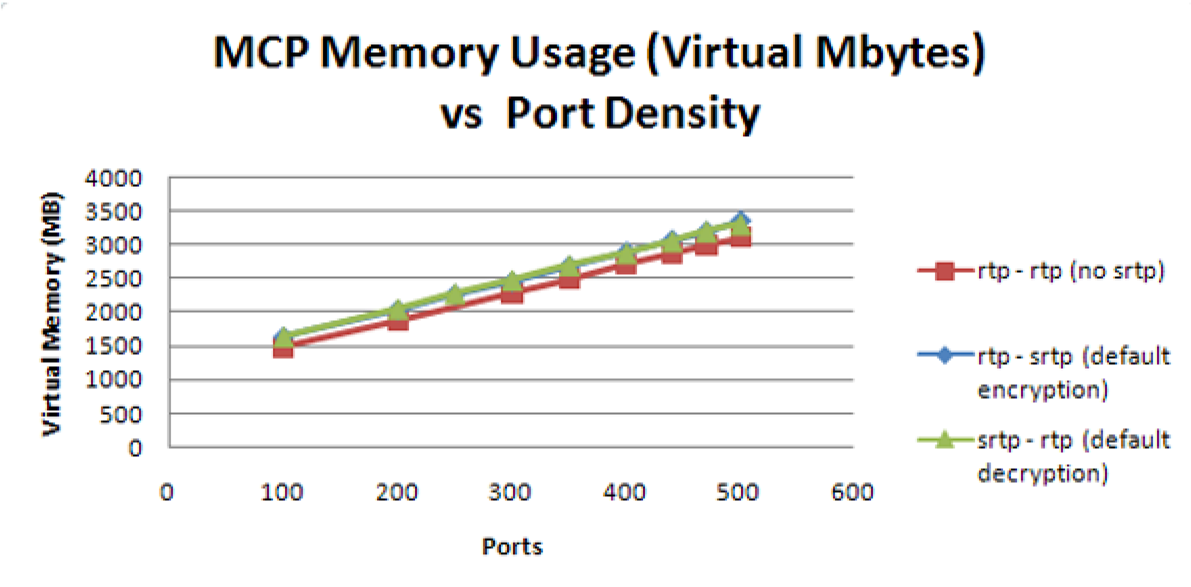


Figure: MCP Memory Usage versus Port Density

The graphs in Figure: System CPU Usage versus Port Density—Default Encryption, Decryption, Figure: System CPU Usage versus Port Density—Encryption, and Figure: System CPU Usage versus Port Density—Decryption provide comparisons of the system CPU usage in various encryption and decryption test case scenarios:

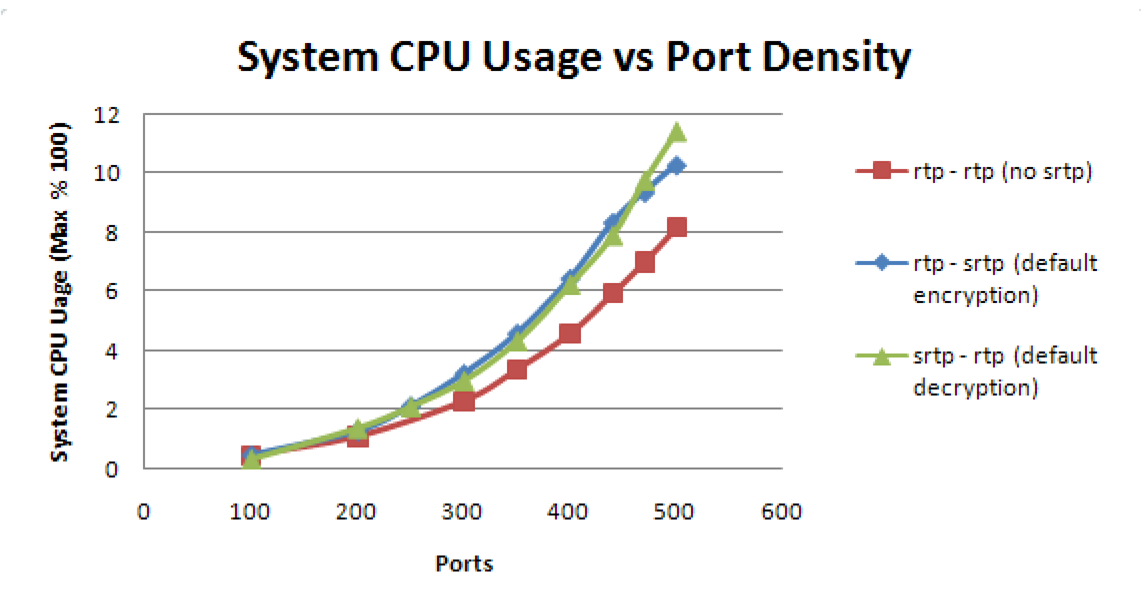


Figure: System CPU Usage versus Port Density—Default Encryption, Decryption

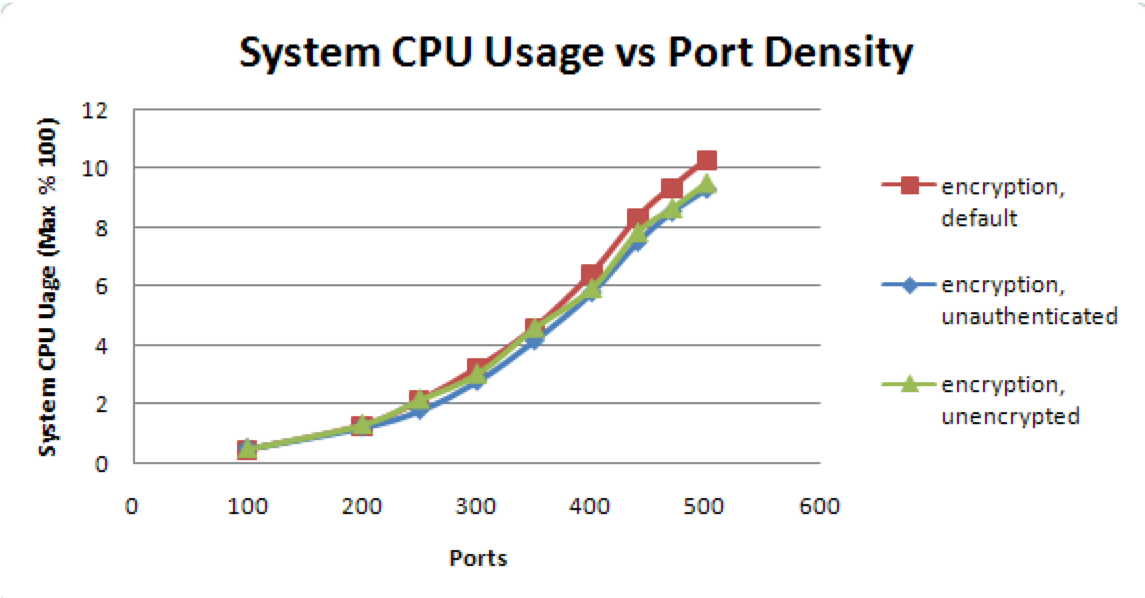


Figure: System CPU Usage versus Port Density—Encryption

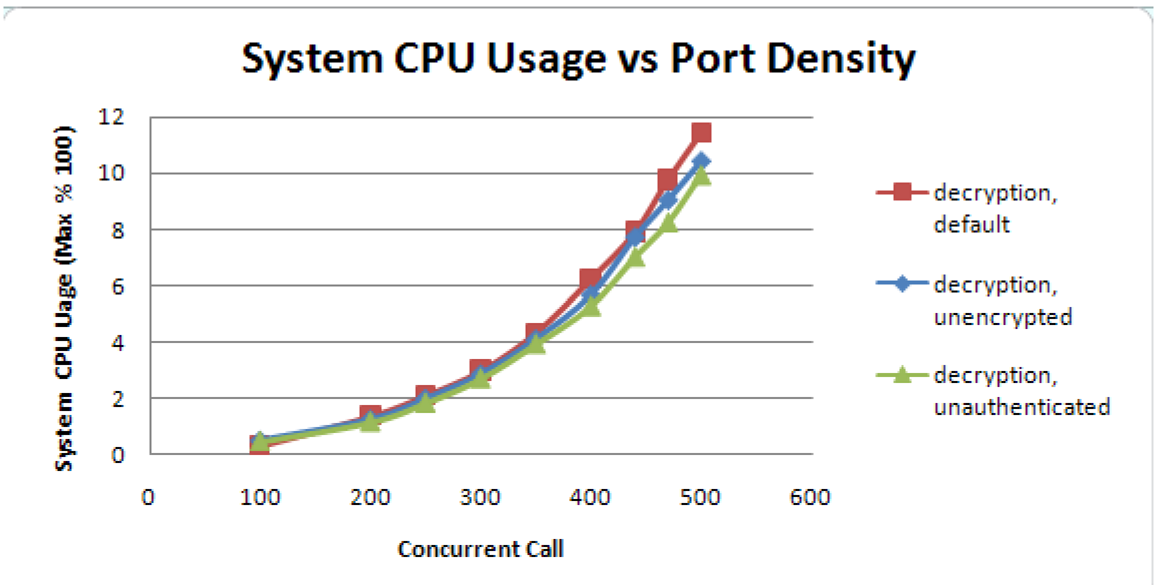


Figure: System CPU Usage versus Port Density—Decryption

The SRTP/RTP test results suggest these conclusions:

- In the encryption scenario, the overall system CPU usage increased from 4.6% to 6.5%—a 41% increase at 400 ports.
- In the decryption scenario, the overall system CPU usage increased from 4.6% to 6.2%—a 35% increase at 400 ports.
- In the encryption scenario, the unencrypted and unauthenticated mode test cases indicates lower CPU usage than the default mode test cases—89% and 91% respectively at 400 ports.

In the decryption scenario, the unencrypted and unauthenticated mode test cases indicated lower CPU usage than the default mode test cases—85% and 92% respectively at 400 ports.

Play Cache

Enabling the play cache functionality increases overall capacity. The transcoding occurs just once during the first call. The transcoded contents is cached and reused in all subsequent calls, and resources normally used for transcoding are no longer needed.

Figure: System CPU Usage vs. Port Capacity (audio-only), Figure: System Disk Usage vs. Port Capacity, and Figure: System Memory Usage vs. Port Capacity are graphs of data derived from MP3 file playback via MSML play:

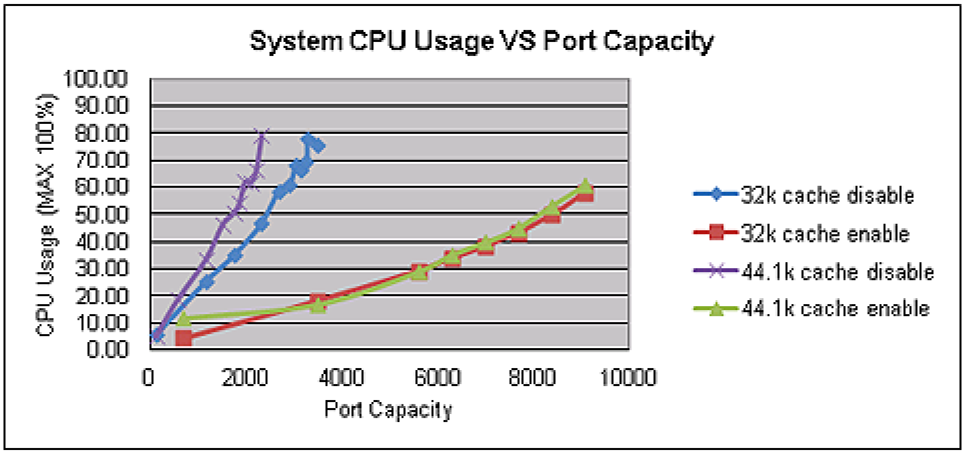


Figure: System CPU Usage vs. Port Capacity (audio-only)

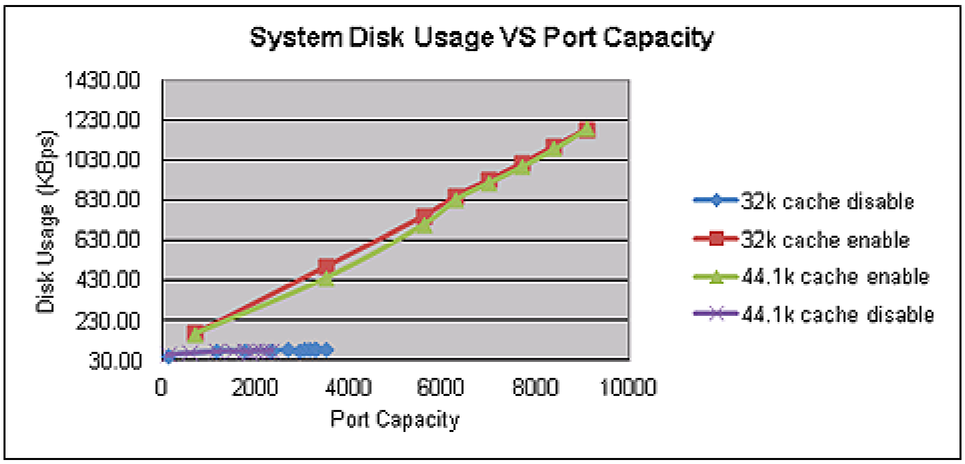


Figure: System Disk Usage vs. Port Capacity

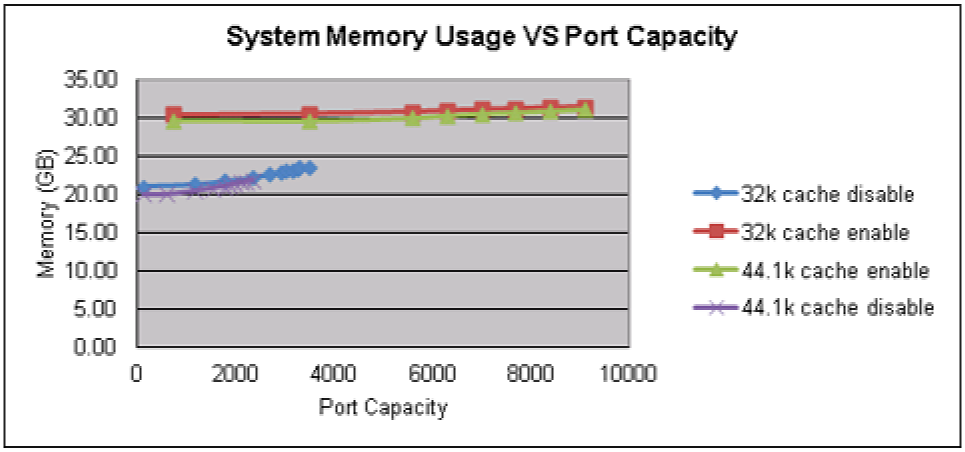


Figure: System Memory Usage vs. Port Capacity

By storing and then accessing a transcoded codec, the play cache both conserves and expends system resources in a mostly positive trade-off.

Play Cache Enabled

Transcoding (and its strong demand on system capacity) occurs only during the first call; system performance improves noticeably on all subsequent calls.

The play cache consumes system memory when enabled, which increases disk traffic and affects system performance.

Play Cache Disabled
CPU usage is intensive during transcoding, which occurs for every call. System performance is noticeably affected.

Figure: System CPU Usage vs. Port Capacity (video and audio) (below) compares the results for transcoding video and audio with the play cache enabled and disabled. The video stream is transcoded from 720p (30fps, 4Mbps, High profile and level 3) to CIF (Main profile and level 2). The audio is AMR. The source file is 3gp. Note that capacity is even further impacted than with audio-only.

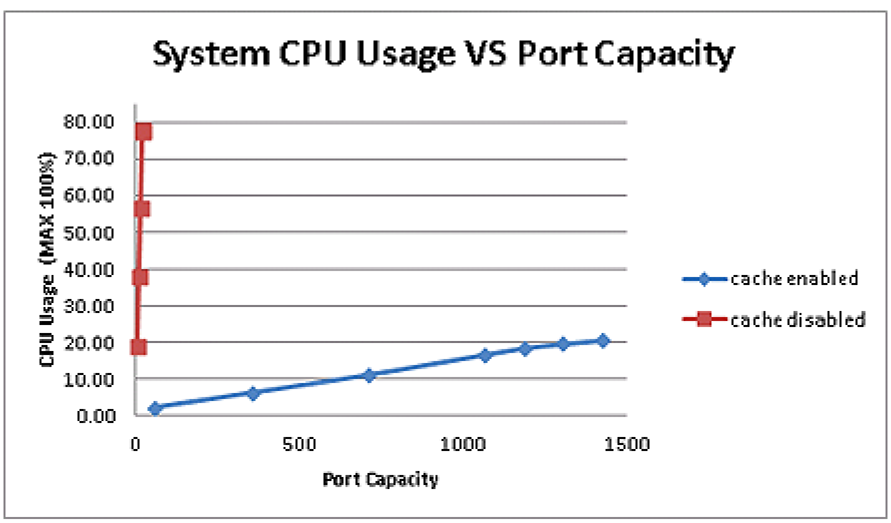


Figure: System CPU Usage vs. Port Capacity (video and audio)

Conference Performance

In the following conference use cases, three variables affect Media Control Platform performance:

- The number of simultaneous conferences.
- The number of participants per conference.
- The number of speaking participants.

As the graph Figure: CPU Usage Versus Total Number of Participants illustrates, the variable impacting performance the most is the total number of participants hosted by GVP (the number of conferences multiplied by the number of participants per conference).

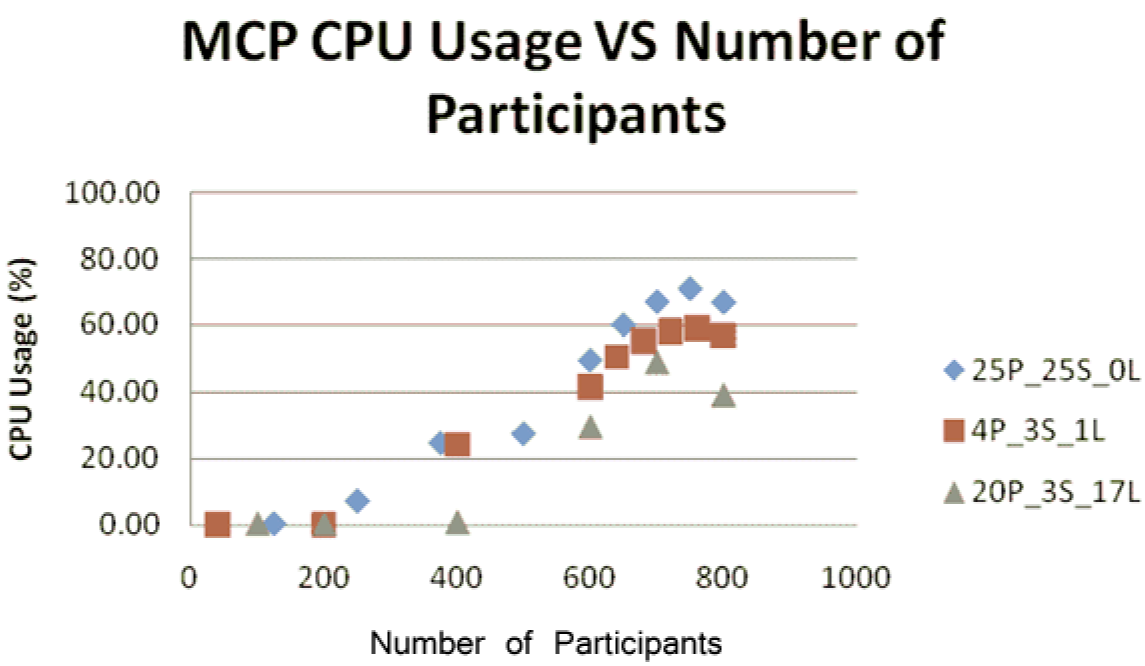


Figure: CPU Usage Versus Total Number of Participants

The symbols and legend in the graph in Figure: CPU Usage Versus Total Number of Participants are explained below:

- 25P_25S_0L = 25 participants per conference (25 active speakers + 0 listen only)
- 4P_3S_1L = 4 participants per conference (3 active speakers + 1 listen only)
- 20P_3S_17L = 20 participants per conference (3 active speakers + 17 listen only)

Overall, the CPU usage increases with a higher percentage of actively speaking participants. However, regardless of the conference configuration, the system bottleneck occurs when the total number of participants reaches ~600 (on a 2x Xeon5160 @3.0GHz server).

The test was conducted using pre-8.1.5 GVP versions (on a physical 2x Xeon5160 @3.0GHz server that NetAnn used to create and join the conference).

For GVP 8.1.5, there is no limit to the number of participants. Two types of testing were conducted: a conference with 32 participants and a conference with as unlimited participants. Both test types used MSML to create and join the conference.

The first test type (32 participants) used four VMs on vSphere 5.0 on 2x Xeon E5620 (8 cores). Each participant was an active speaker for five minutes (300 seconds). The higher number of participants (768 participants from 24 conferences of 32 participants each) was handled successfully. The overall system CPU usage was not as high as before, since the bottleneck was the call to join an MSML conference.

The second test used a physical machine—a Xeon X5675 @3.06GHz—since only one conference would be created. The testing was conducted with two types of codecs: G.711u audio-only and H.263 video (video switching) + G.711u audio. The newly introduced MCP parameter `conference.threadedoutput` had to be enabled (default off) for a larger conference, otherwise, MCP could not achieve such a high number of participants. There were only three active speakers in the conference while all other participants were listeners. Each participant would stay 30 minutes (1800 seconds) in the conference. The Figure: System CPU Usage vs. Participants (below) shows the overall system CPU usage:

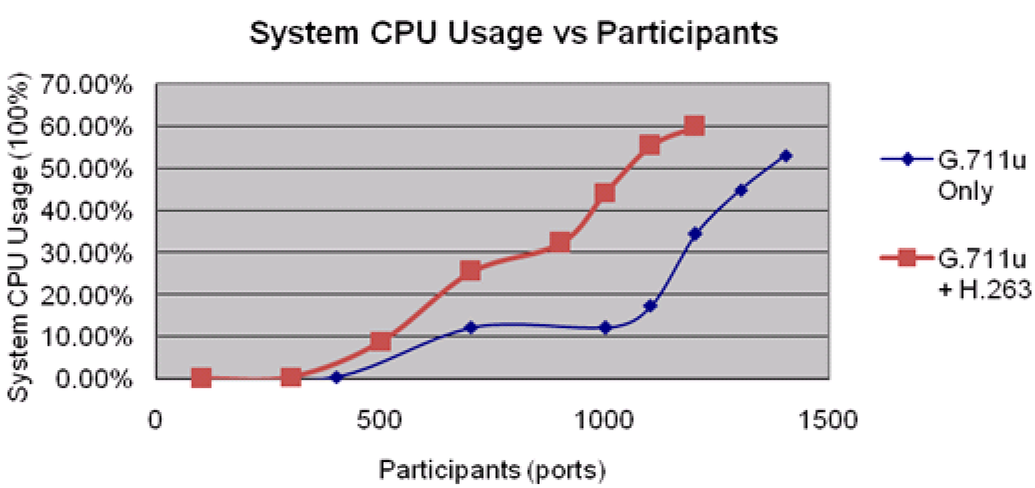


Figure: System CPU Usage vs. Participants

Figure: 95th Percentile vs. Ports (below) shows that CPU usage jumps beyond 1100 participants for the G.711u-only case and beyond 900 participants for the G.711 + H.263 case; and that 48-hour load testing can survive 1300 participants for the G.711u-only case and 1100 participants for the G.711u + H.263 case. The ninety-fifth percentile of call duration shows that CPU usage would jump beyond 1300 participants for the G.711u-only case and beyond 1100 for the G.711u + H.263 case.

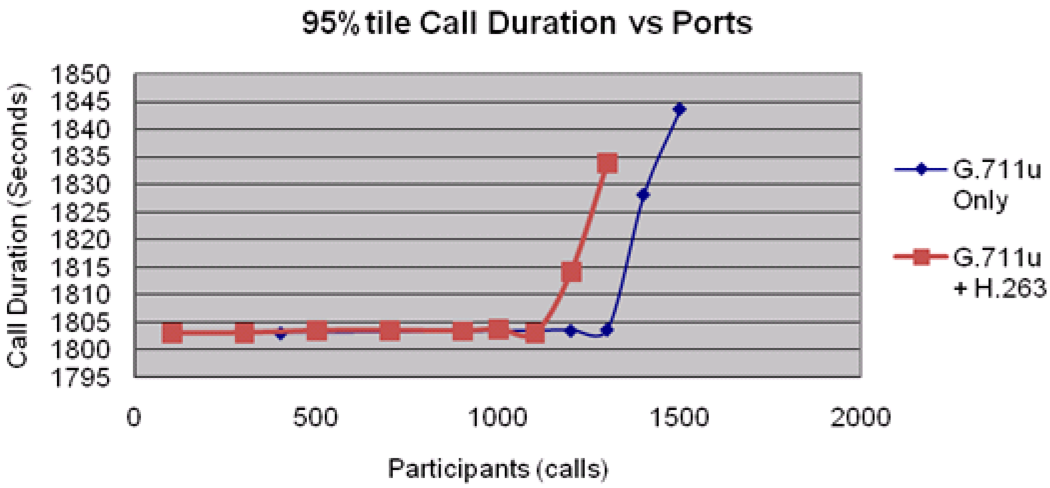


Figure: 95th Percentile vs. Ports

The graph below illustrates memory usage:

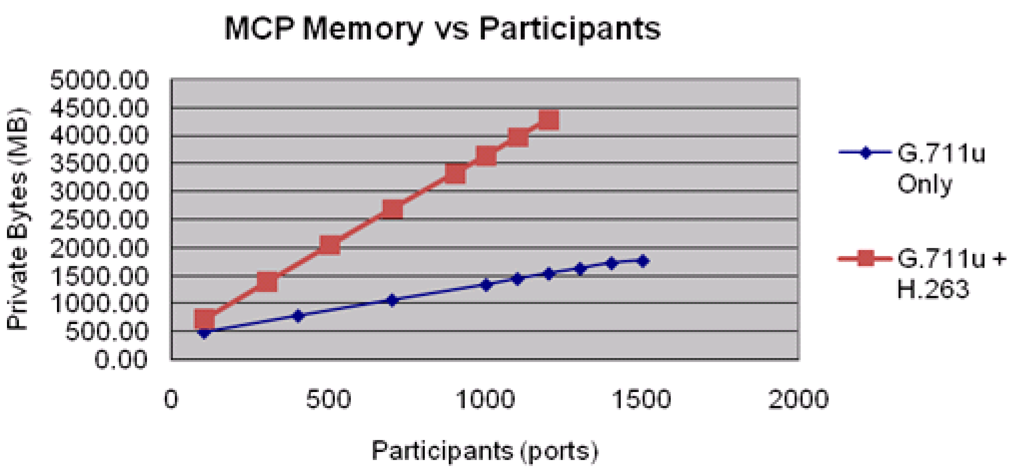


Figure: MCP Memory vs. Participants

When tested on a 64-bit OS, memory usage for the G.711u + H.263 case exceeded 3GB virtual memory (2GB beyond 500 participants). This usage crashed a 32-bit OS. Thus, the peak capacity for G.711u + H.263 is 500.

HR Timer

Two parameters: HR Timer (specifically for Windows) and Gain Control, impact the performance of the conference. Figure: System CPU Usage vs. Capacity (hr timer - gain) (below) compares the performance in terms of system CPU usage from combinations of different values of these two parameters:

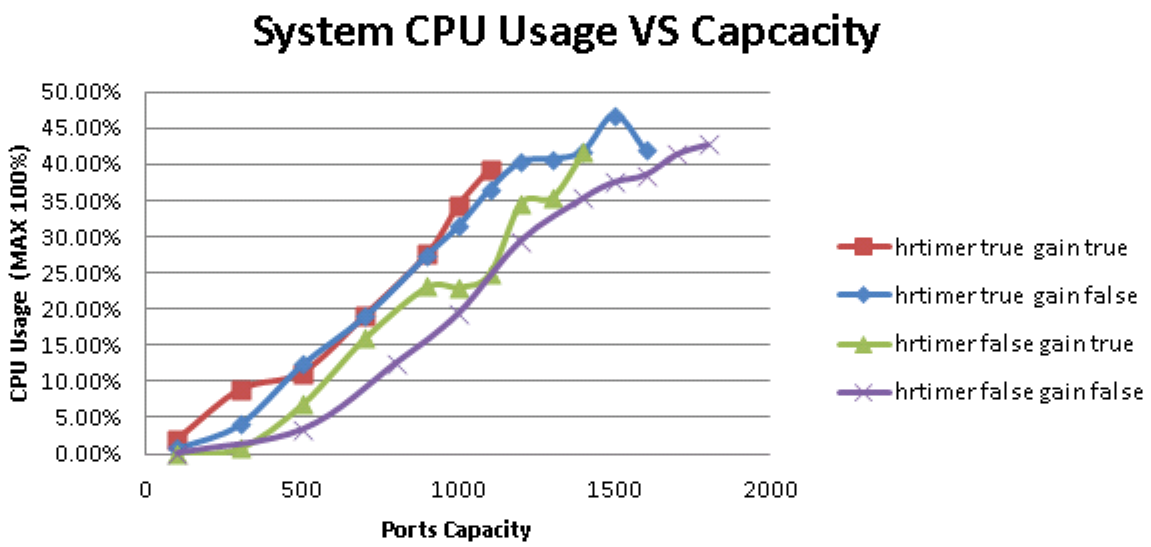


Figure: System CPU Usage vs. Capacity (hr timer - gain)

The highest port capacity (measured in Participants) with the least CPU usage can be achieved when both HR Timer and gain control are turned OFF. Conversely, the lowest port capacity with the highest CPU usage is achieved when both parameters are turned ON. Table: Port Capacity (measured in participants) below documents these results.

Port Capacity (measured in participants)

Gain Control	HR Timer	Participants
ON	ON	1000
OFF	ON	1300
ON	OFF	1400
OFF	OFF	1800

Prior to 8.1.6, gain control was hard-coded to turned on and there was no HR timer. So in 8.1.6, gain control on and HR time off was compatible with previous releases.

MSML CPA/CPD Performance

CPA/CPD is requested through MSML; therefore performance is measured by call rate. Testing was conducted with different tones such as Answering Machine, Busy, Fax, Human, and SIT VC. Below is a graph for all of the above tones:

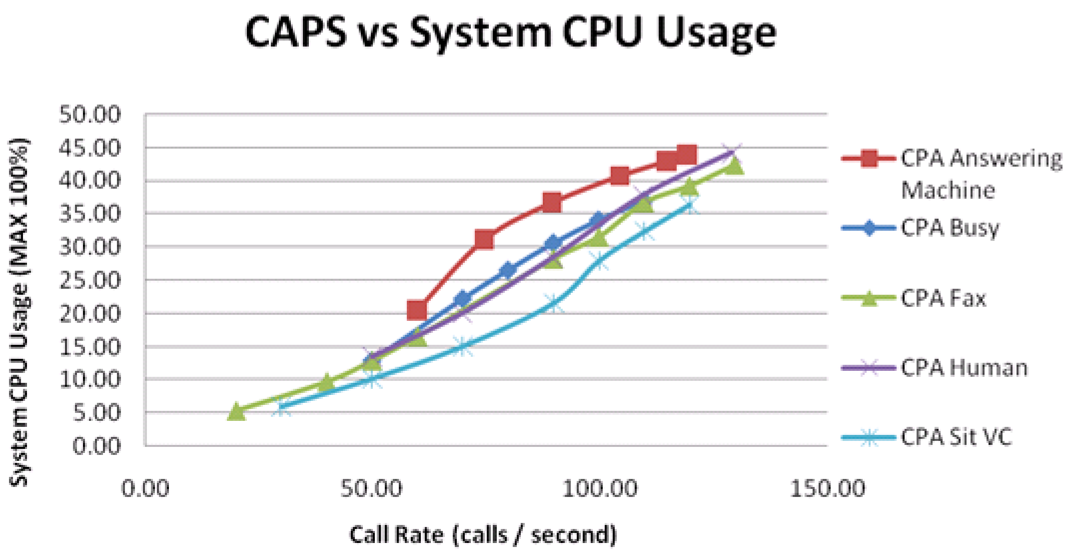


Figure: CAPS vs. System CPU Usage 5

The call duration varies, depending upon the type of tone and the length of recognition, and the peak call rates are quite close one another for each tone. In other words, call rate—and not ports—is a major factor determining peak capacity.

top | toc

Component Test Cases

The following component test cases are described in this section:

- Media Control Platform on Windows
- Media Control Platform on Linux
- Resource Manager
- MRCP Proxy
- PSTN Connector
- CTI Connector
- CTI Connector/ICM
- Supplementary Services Gateway
- Reporting Server

Media Control Platform on Windows

Testing was conducted on Windows 2003, Windows 2008 (x86 and x64), and Windows 2008 R2. But not all testing was executed in the same release. In general, performance results were similar when the Media Control Platform was installed on either version of Windows. The next Figure shows slight differences.

Figure: Call Setup Latency Versus Concurrent Calls (MCP on Windows) depicts the call setup latency metrics for each Windows version when the VoiceXML_App1 was executed as the background load.

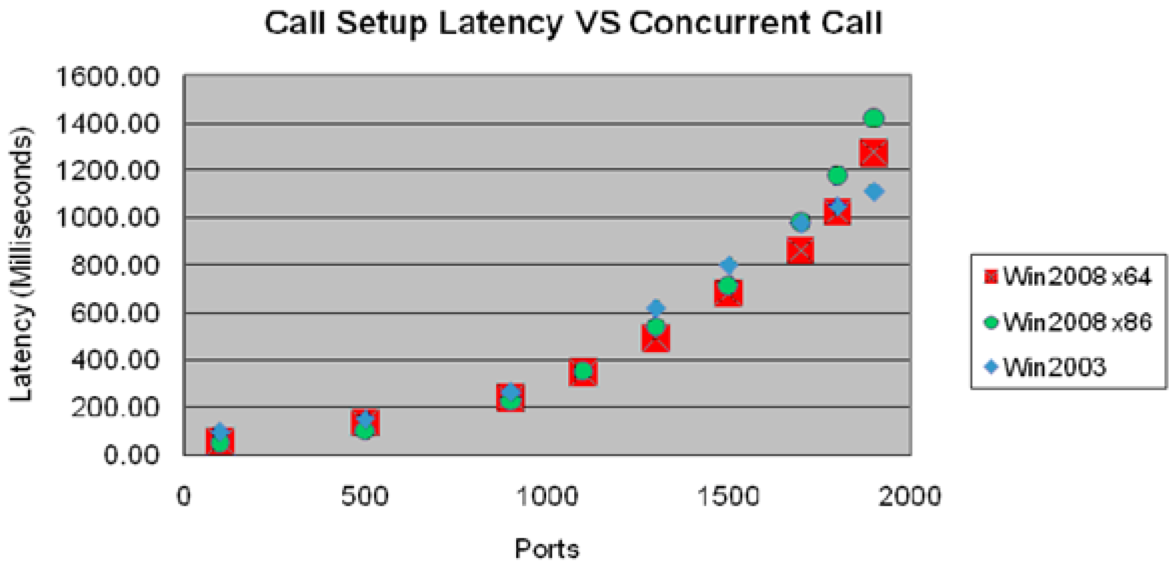


Figure: Call Setup Latency Versus Concurrent Calls (MCP on Windows)

The latency numbers are quite close for ports below 1300, while they are slightly higher on Windows 2003 when ports are 1300 or higher. This trend continues up to 1800 ports, at which point latency on Windows 2008 exceeds those on Windows 2003.

In Figure: System CPU Usage Versus Concurrent Calls (MCP on Windows), the graph depicts CPU usage for the overall system on each Windows version, when testing was performed by using VoiceXML_App1 and the results were scaled.

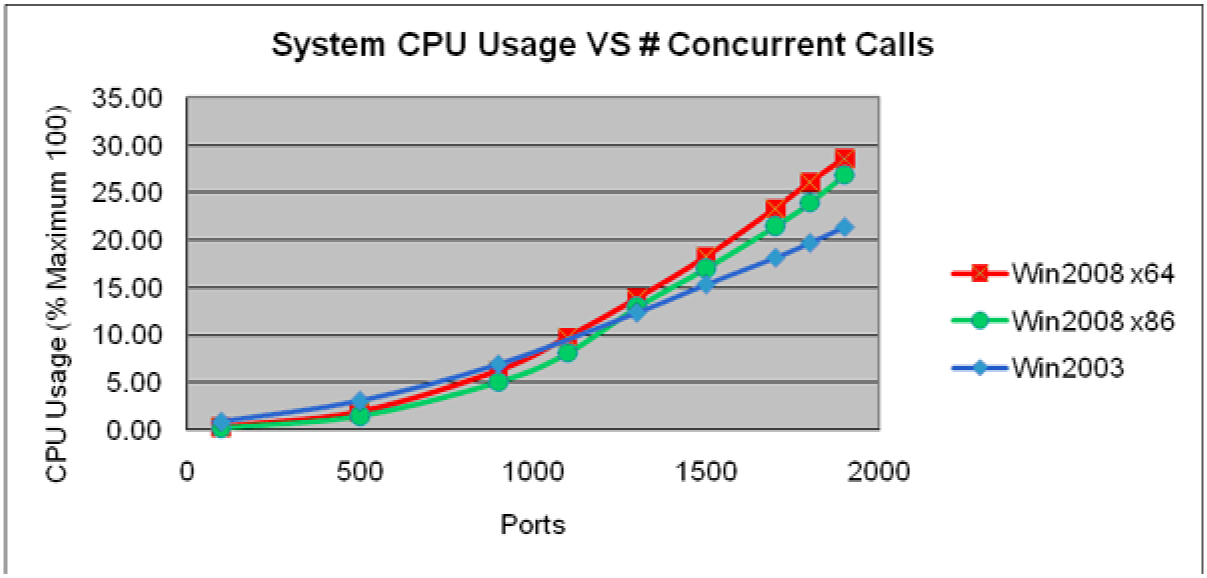


Figure: System CPU Usage Versus Concurrent Calls (MCP on Windows)

The CPU usage for both versions of Windows 2008 (x86 and x64) are almost identical to each other, trending a little higher than Windows 2003. Also, Windows 2003 can sustain higher ports than the preferred 1300—ignoring other factors such as call setup latency. Beyond 1300 ports, the Windows 2008 call pass rate drops below Genesys QA pass criteria of 99.99%. However, on Windows 2003, 1800 ports can be achieved within the pass criteria if call setup latency is ignored.

Media Control Platform on Linux

Testing was conducted on Red Hat Enterprise Linux 4, RHEL 5 (x86 and x64), and RHEL 6 x64. However, not all testing was executed in the same release. There were no significant differences in performance on RHEL 4 & RHEL 5 (both x86 & x64), while the performance was better on RHEL 5 than on RHEL 6.

Below is the graph of call setup latency measured in different Red Hat Linux systems on physical servers, when RHEL 6.x x64 (VXML_App1 was executed as background load: Figure: Call Setup Latency Versus Concurrent Calls (MCP on Linux) depicts the call setup latency metrics for RHEL 4 and RHEL 5 on physical servers when VoiceXML_App1 was executed as the background load.

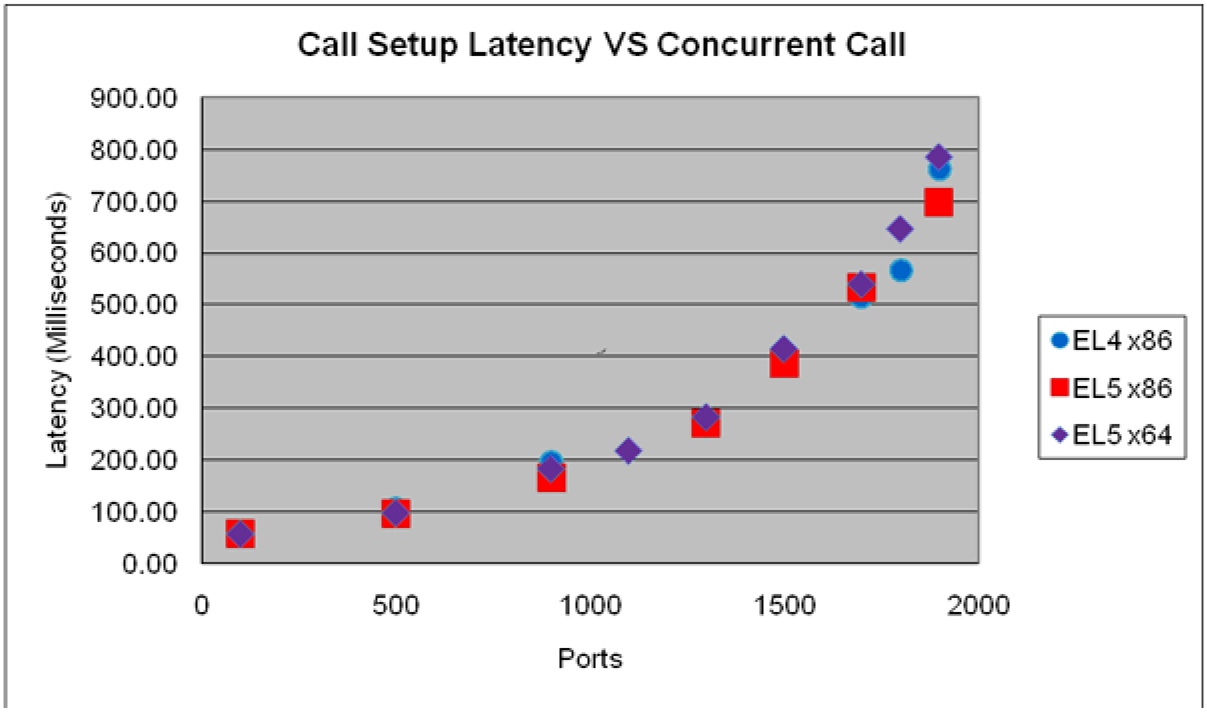


Figure: Call Setup Latency Versus Concurrent Calls (MCP on Linux)

All three Linux versions showed latency results that were almost in line with one another at 1700 ports or lower. Above 1700 ports, which is beyond Genesys QA preferred peak capacity, there were some differences.

In Figure: CPU Usage Versus Concurrent Calls (MCP on Linux), the graph provides a comparison of the overall CPU usage when VoiceXML_App1 is executed.

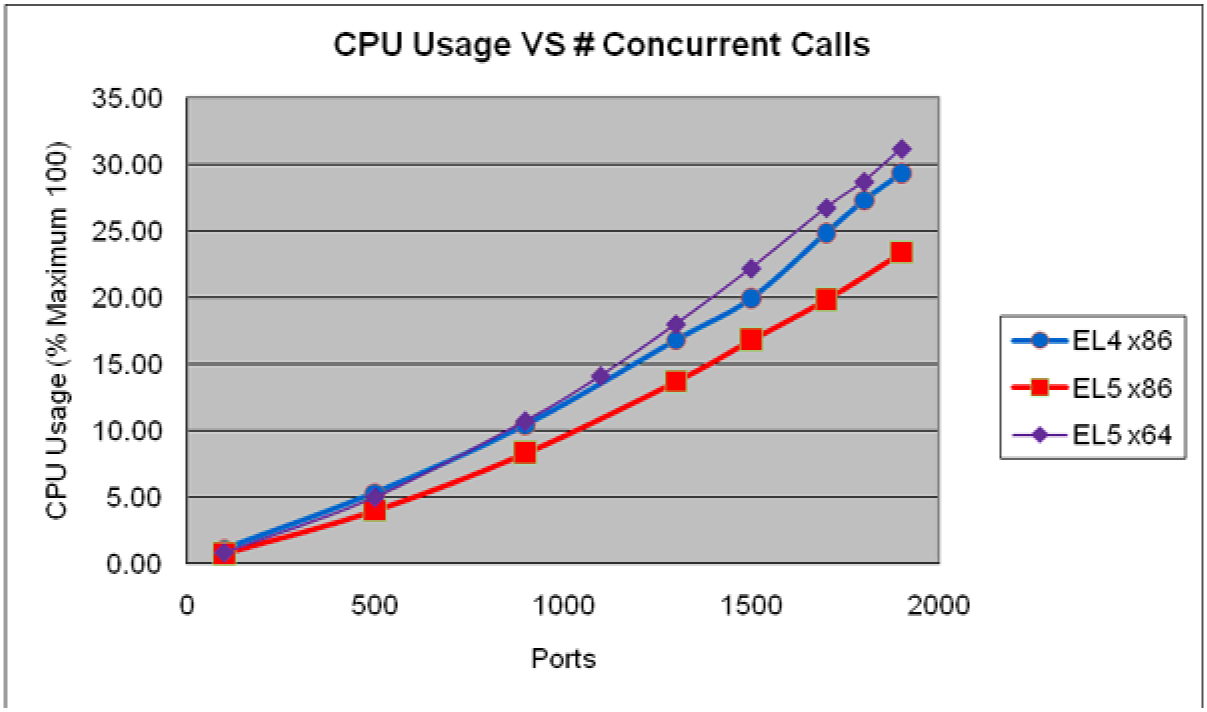


Figure: CPU Usage Versus Concurrent Calls (MCP on Linux)

There were no significant differences between Linux versions for overall CPU usage, and overall performance was similar, even when multiple simultaneous factors, such as call setup latency and call pass rate, were considered.

Performance differences between Linux and Windows depends on specific use cases. The maximum number of concurrent ports on Linux is slightly higher than on Windows in some test cases in which there were more ports, such as MSML with CPD, but worse for other test cases, such as those in which G.711 and G.729 transcoding was used.

GVP overall performance on Linux and Windows is quite similar, and although the test cases performed on both Windows and Linux were not identical, the peak capacity was not significantly different.

Because performance suffers on RHEL 6.4 x64, the virtual environment was used only to test MCP on RHEL 6.x x64 as a guest OS on ESXi 5.0. Below is the comparison of call setup latency between EL5 x64 and EL6 x64 while both were on virtual environment as guest a OS:

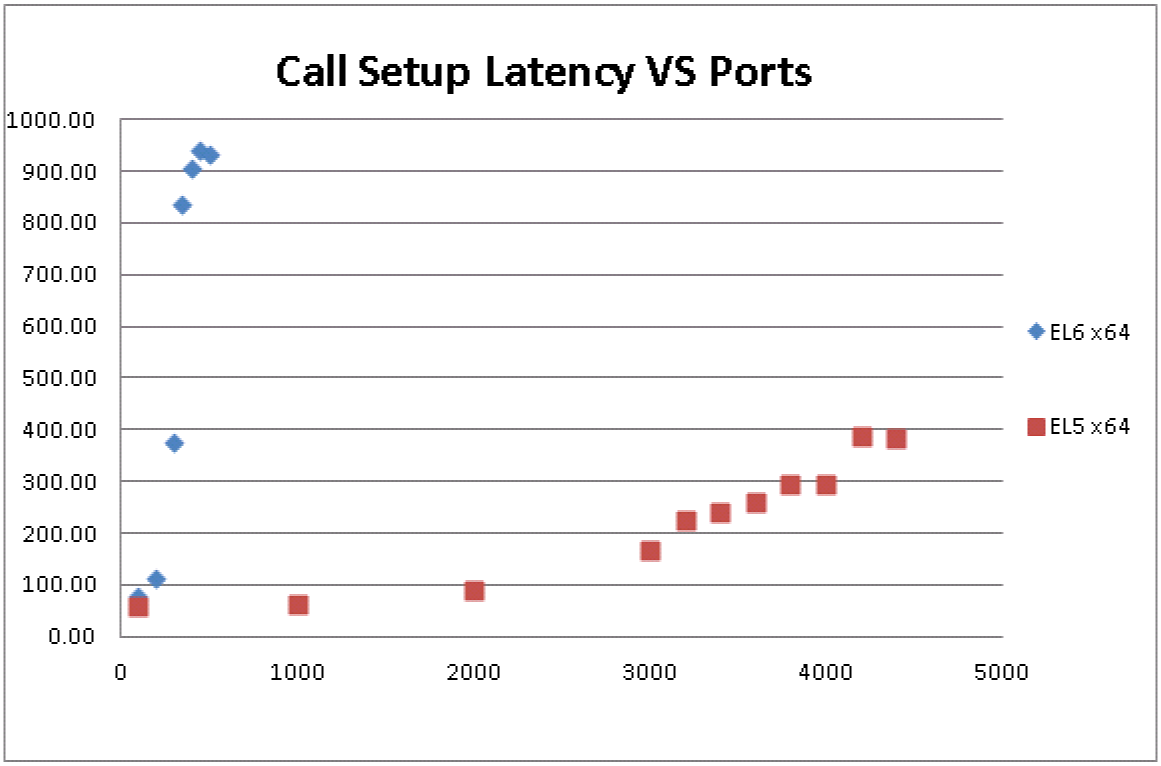


Figure: Call Setup Latency vs. Ports

Figure: Call Setup Latency vs. Ports (above) shows that call setup latency increased significantly on EL6 x64, compared with EL5 x64.

Resource Manager

Resource Manager performance was measured in terms of Call Arrivals Per Second (CAPS). Performance was not affected by the number of simultaneous calls held by Resource Manager. Resource Manager performed most efficiently when multiple Media Control Platforms were used.

The effect on Resource Manager performance differs, depending on the type of call being processed (for example, conference versus announcement calls), but generally, a peak of 800 CAPS can be sustained for call types such as call routing and conference, and regardless of whether it is in an HA or non-HA configuration. This applies to all Windows versions and most RHEL servers except RHEL 6 x64.

CPU consumption on Resource Manager is very low. The 800 CAPS limit mentioned previously is due to the use of virtual memory, which exceeds the 3GB limit (configured at the OS level) when Resource Manager is running consistently beyond 800 CAPS.

The same capacity results were achieved when the Resource Manager was tested using both UDP and TCP due to a bottleneck when it reached the 3GB virtual memory limit.

Figure: CAPS Versus CPU Usage (Resource Manager) depicts CPU usage when Resource Manager is installed on Red Hat Enterprise Linux 5 x86.

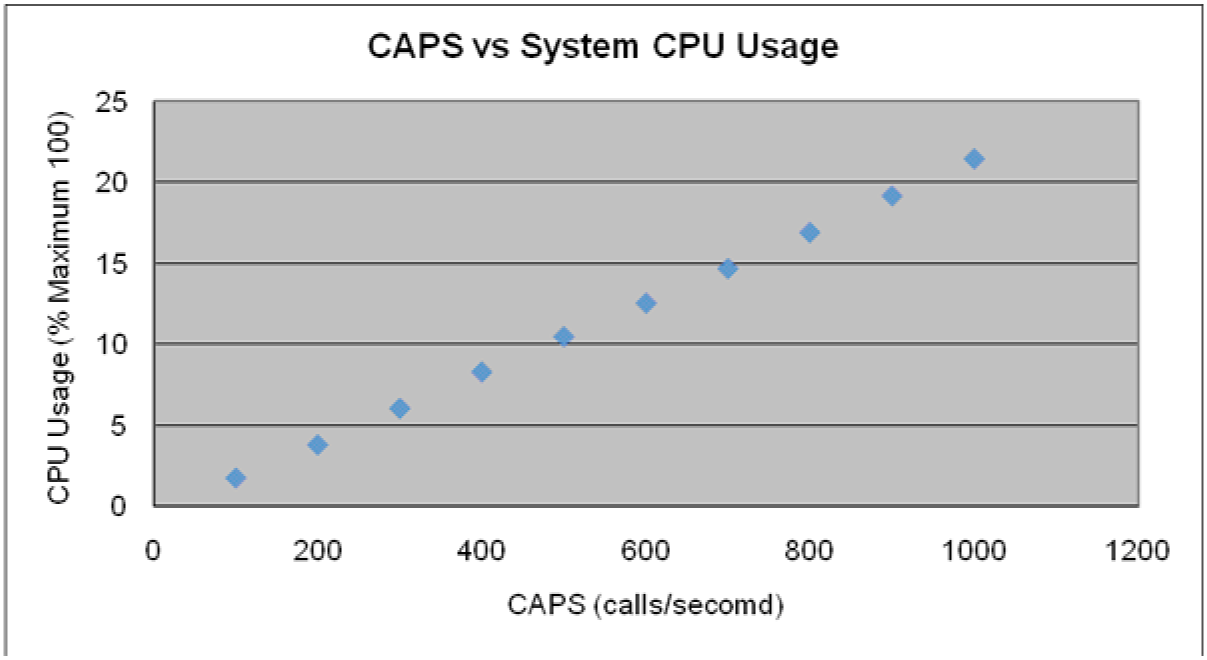


Figure: CAPS Versus CPU Usage (Resource Manager)

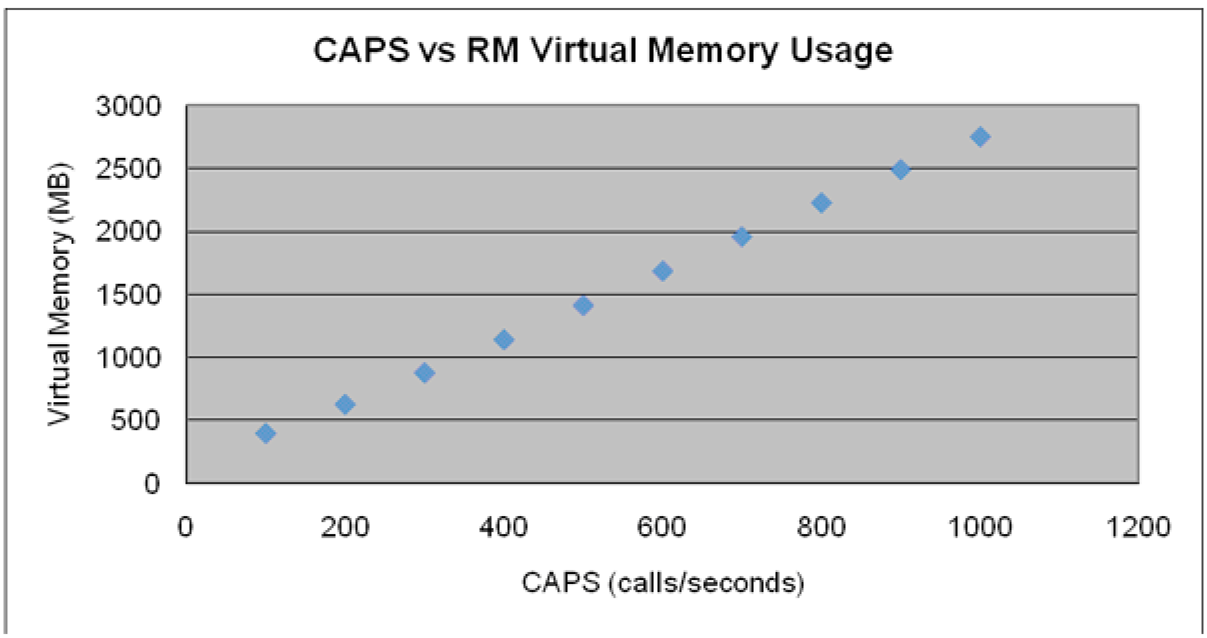


Figure: CAPS Versus Memory Usage (Resource Manager)

Figure: CAPS Versus Memory Usage (Resource Manager) shows that the Resource Manager can sustain 900 CAPS. However, since the memory usage almost reaches the 3GB limit, a 800 CAPS peak capacity seems more appropriate.

When configured with the Reporting Server, Resource Manager sustained 800 to 900 CAPS, but the Reporting Server performance invariably caused a bottleneck to occur. See Reporting Server.

When a single tenant (Environment by default) is used, 800 CAPS is achieved (see Figure: CAPS Versus Memory Usage (Resource Manager)). When multiple tenants are configured, the Resource Manager performance is slightly degraded. When tested with 1000 tenants, each configured with 1 child tenant, the Resource Manager performance achieves 600 CAPS of peak capacity.

When a large number of Direct Inward Dialing (DID) numbers are configured in a single tenant, the Resource Manager performance, again, is slightly degraded. When 100,000 DID numbers are tested with 262 IVR Profiles (without the use of DID ranges or wild cards, for example, a long, simple mapping list), peak capacity is 600 CAPS.

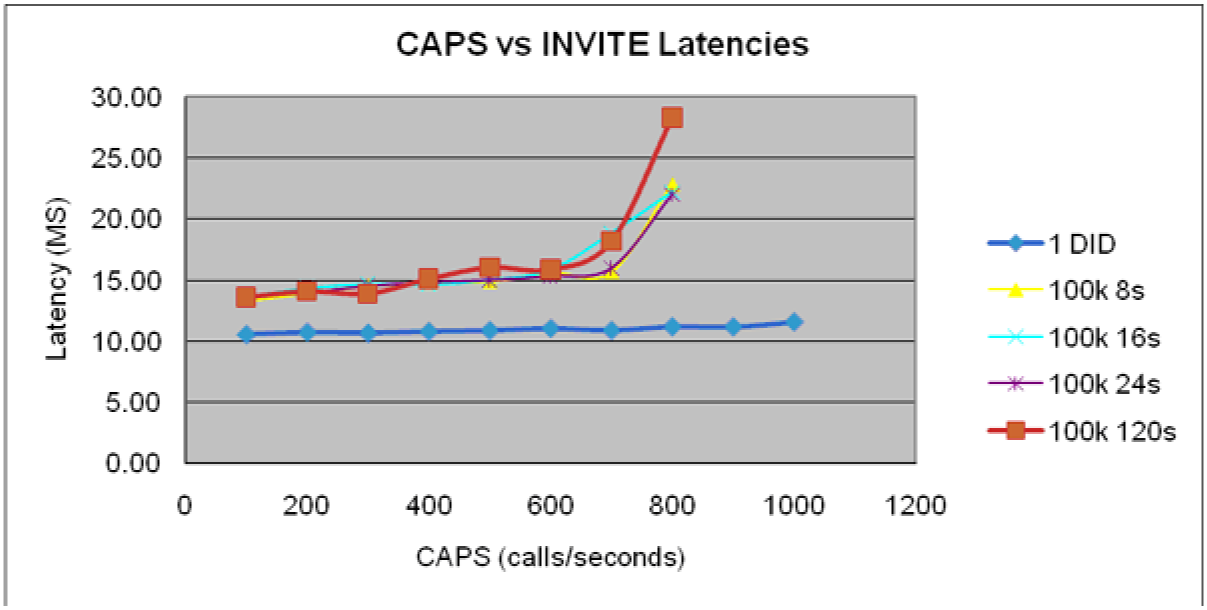


Figure: CAPS Versus SIP INVITE Latencies (Resource Manager)

Figure: CAPS Versus SIP INVITE Latencies (Resource Manager) depicts the propagation of SIP message latencies between 1 and 100,000 DID numbers with various call durations, and shows higher latencies for SIP INVITE messages for 100,000 DID numbers versus the 1 DID baseline, while there is not much difference in latencies with SIP ACK messages (see Figure: CAPS versus INVITE Latencies (Resource Manager) and Figure: CAPS Versus SIP BYE Latencies (Resource Manager)). The delay likely occurs when Resource Manager searches for mappings upon receiving SIP INVITE messages. The testing also indicates that call duration is not relevant to Resource Manager performance.

The two previous scenarios (1000 tenants with one DID entry each and 100,000 DID in a single tenant) produce the worst results. Resource Manager can achieved better performance results when multiple tenants are configured with a small number of DID entries per tenant. Resource Manager was tested with the requirement of 1 million DIDs distributed among 32 tenants, each containing 30-35 K of DID entries. (A 4MB size limitation exists for Management Framework objects). Even in this configuration, the Resource Manager still achieved 800 CAPS.

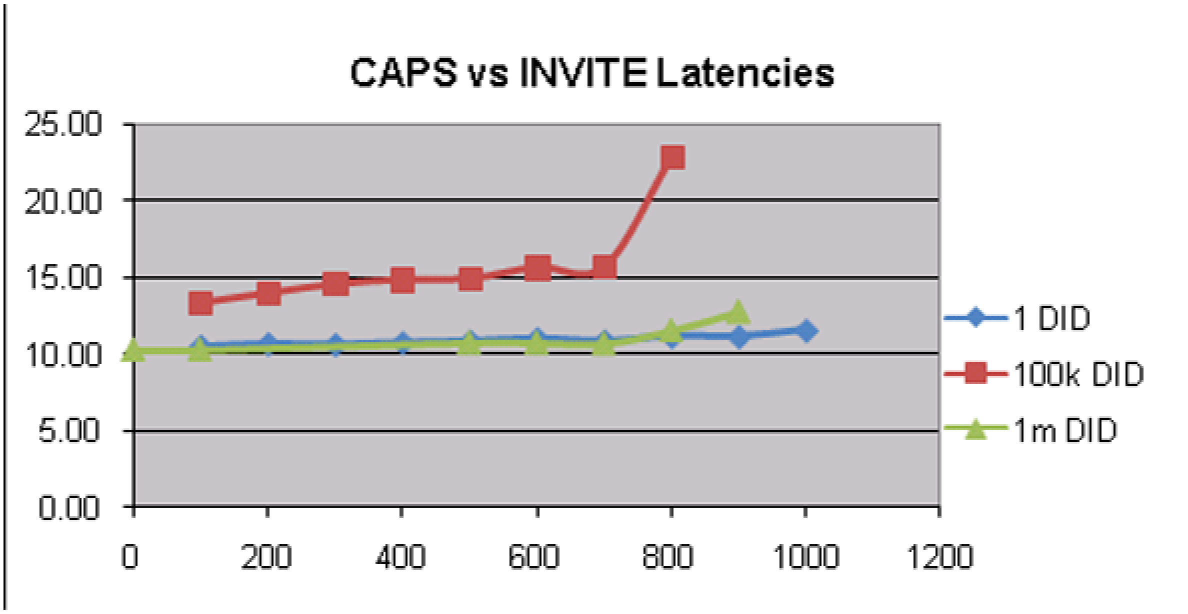


Figure: CAPS versus INVITE Latencies (Resource Manager)

In Figure: CAPS Versus SIP INVITE Latencies (Resource Manager), the SIP INVITE latency is almost in line with one DID entry until capacity reaches 900 CAPS, then it increases. The virtual memory is also close to the 3GB limit.

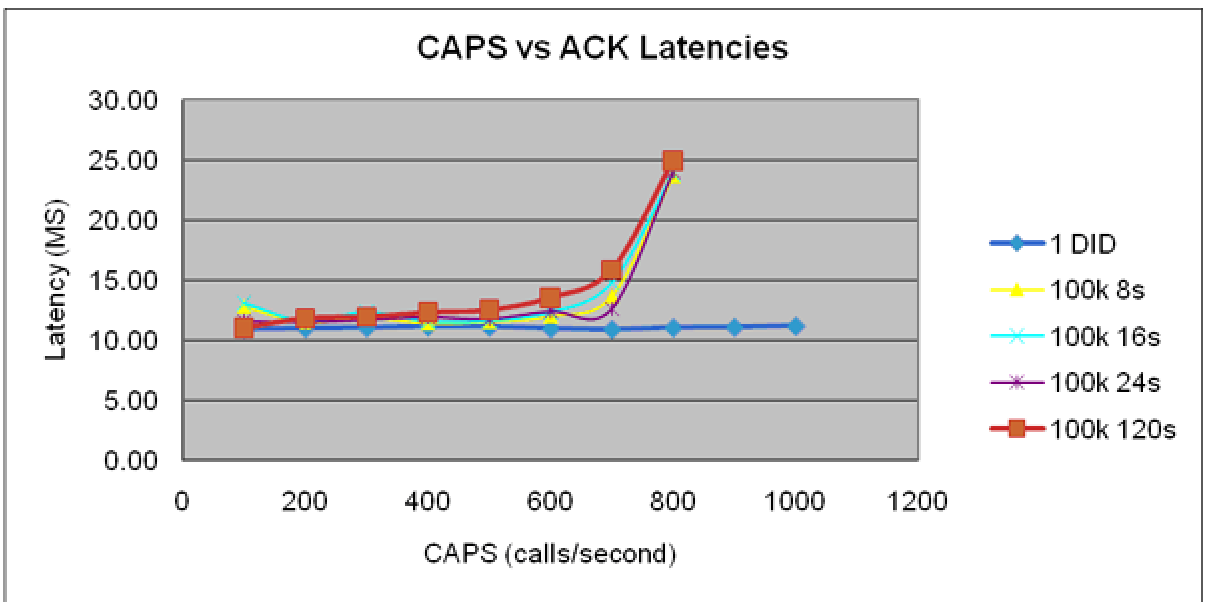


Figure: CAPS Versus ACK Latencies

Performance superiority of RHEL 5 x86 over RHEL 6 x64 was observed during RM testing. Below is the graph for propagating latency of INVITE to compare RHEL 6 x64 and RHEL 5 x86:

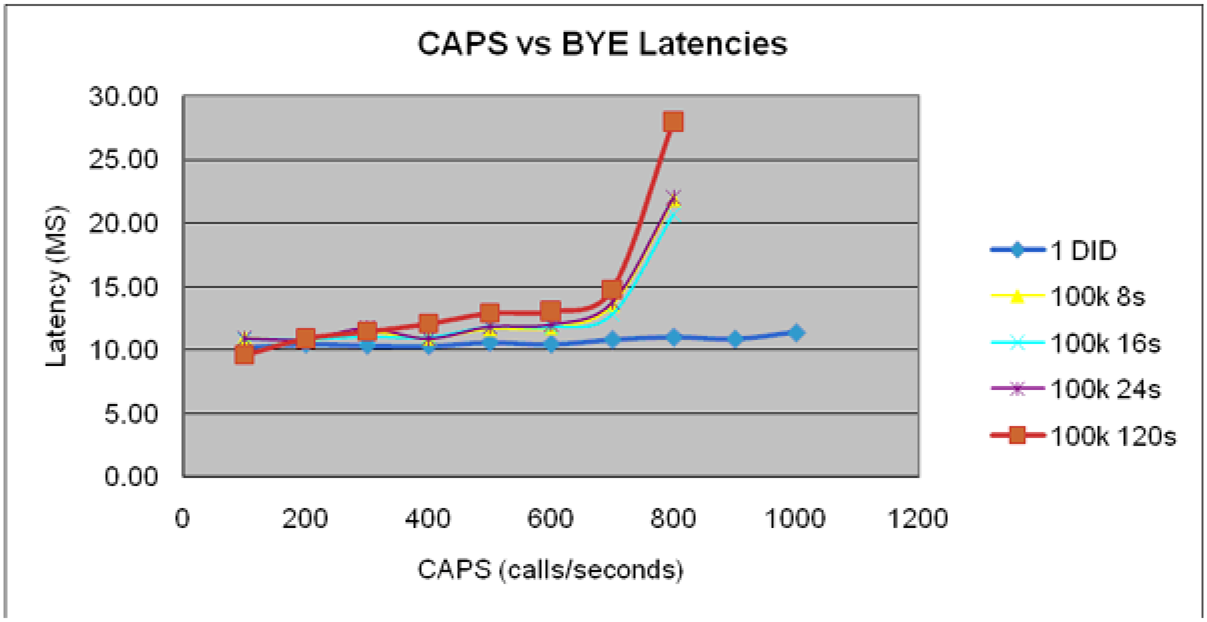


Figure: CAPS vs. BYE Latencies

Latency increased significantly on RHEL 6 x64, even at low CAPS.

MRCP Proxy

The MRCP Proxy performance is measured in terms of MRCP sessions and is benchmarked by using simulated MRCP v1 servers and clients. A typical 10 second MRCP session for either an ASR or TTS request used for testing.

The peak capacity is achieved at 1,600 concurrent MRCP requests per second (half ASR and half TTS) in CAPS, but the MCRP Proxy can hold 16,000 MRCP sessions in total. Beyond 1600 CAPS, it might still respond. However, the entire session becomes quite lengthy and will eventually time out. Figure: ASR Call Duration Versus CAPS—MRCP Proxy (MRCPv1) and Figure: TTS Call Duration Versus CAPS—MRCP Proxy (MRCPv1) depict the ASR and TTS 95th percentile of call duration. The results indicate that the call duration beyond 800 CAPS more than doubles.

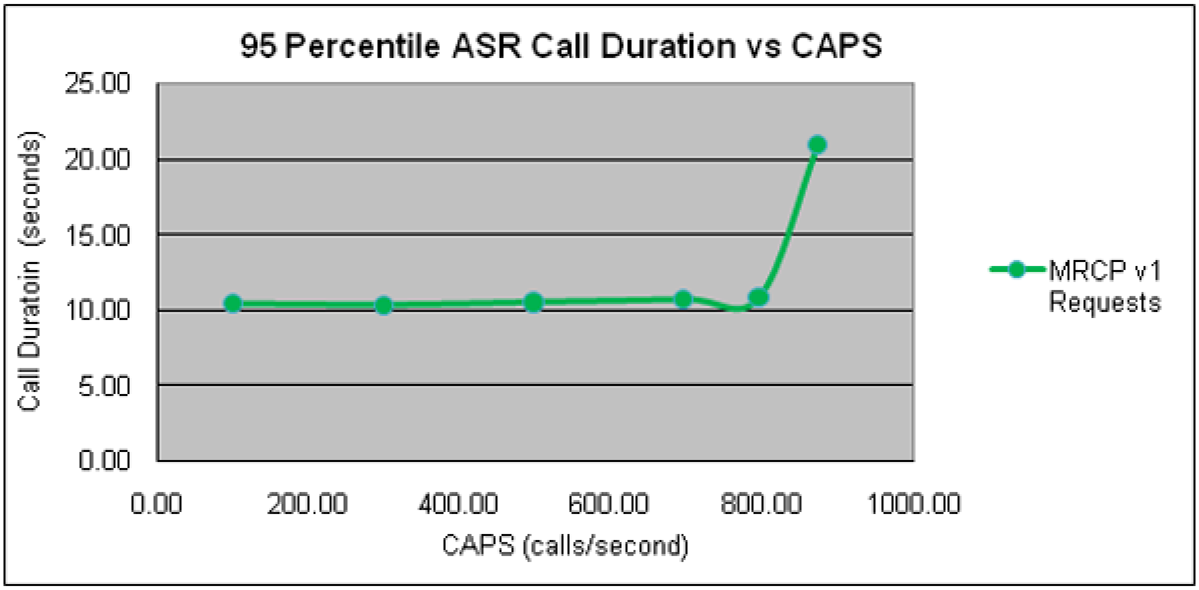


Figure: ASR Call Duration Versus CAPS—MRCP Proxy (MRCPv1)

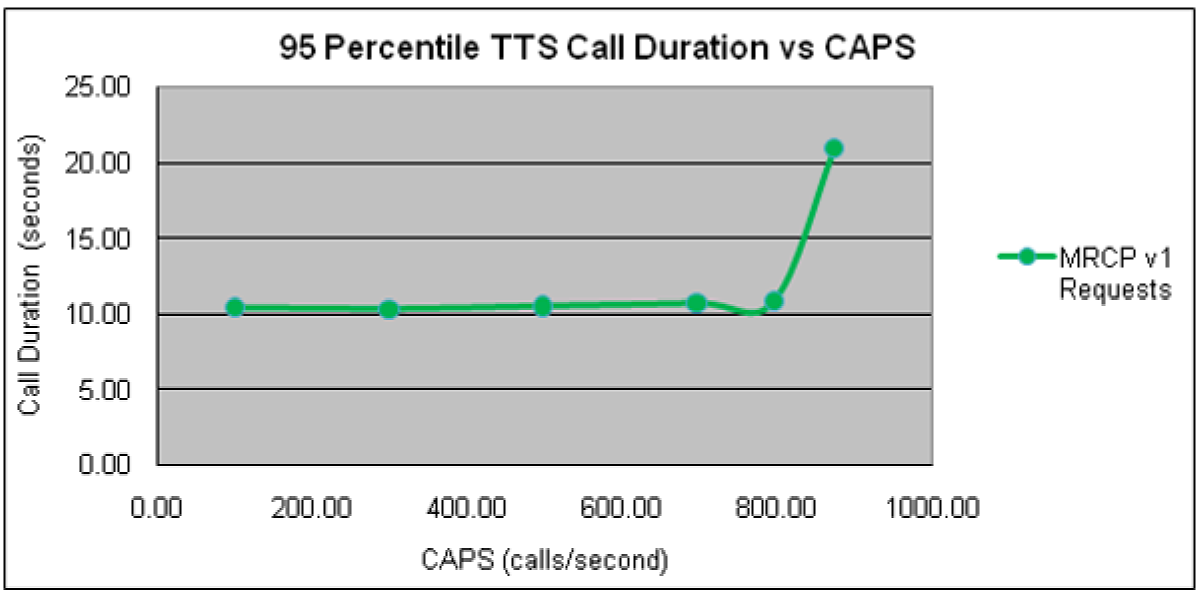


Figure: TTS Call Duration Versus CAPS—MRCP Proxy (MRCPv1)

Figure: TTS Call Duration Versus CAPS—MRCP Proxy (MRCPv1) and Figure: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy) depict the overall CPU usage for the MRCP Proxy. The CPU usage increases substantially beyond 1600 CAPS.

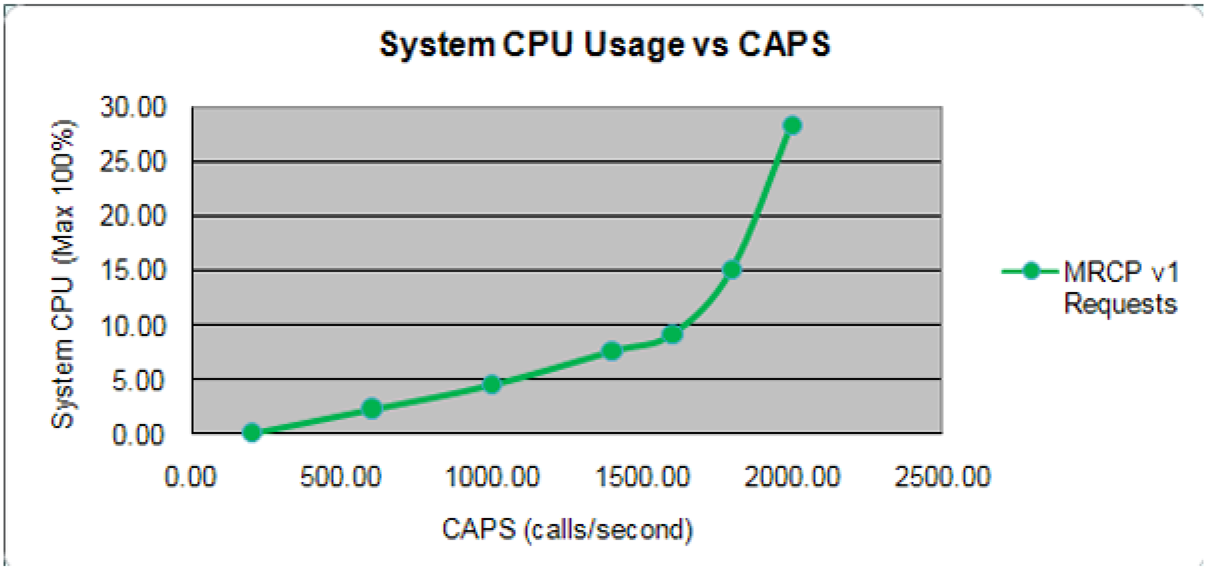


Figure: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy)

As shown in Figure: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy), the same significant increase in memory (private bytes) consumption beyond 1600 CAPS, is indicated.

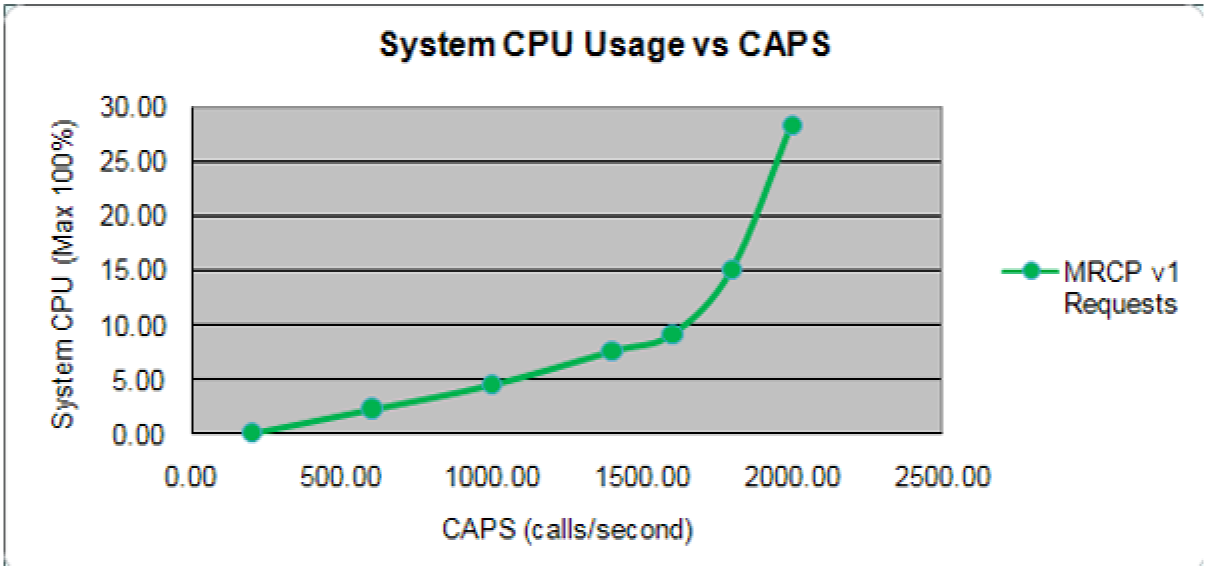


Figure: System CPU Usage Versus CAPS—MRCPv1 (MRCP Proxy)

PSTN Connector

The performance of PSTN Connector is measured in terms of T1/E1 spans. Two Dialogic DMV Media Boards, which provide 8 T1/E1 spans, were tested in a machine with a Dual Xeon, 3.0GHz CPU. Also, all other components such as the Media Control Platform, Resource Manager, and SIP Server were installed off board.

- Two different protocols were used—ISDN and CAS/RB Wink.

- Two application profiles were used—VoiceXML_App1 and VoiceXML_App3.
- The overall CPU idle was approximately 80% for both applications.

CTI Connector

CTI Connector performance was tested in two scenarios, in which a Play Treatment application was used with two different transfer types (a bridge transfer and a blind transfer)NGI. The Media Control Platforms were configured to use NGI and GVPi, respectively.

Two test cases where GVPi was used produced 25 CAPS on Windows and Linux. A test case in which the NGI was used in a blind transfer scenario produced 25 CAPS on Windows and Linux, while a bridge transfer produced only 15 CAPS on all supported Windows versions and 20 CAPS on Linux.

Beyond peak capacity, the overall call-pass-rate dropped below the 99.95% criteria. Figure: PD Versus CPU Usage (CTI Connector) is a sample of the CTI Connector overall CPU usage versus port density CAPS when NGI is used in a blind transfer scenario:

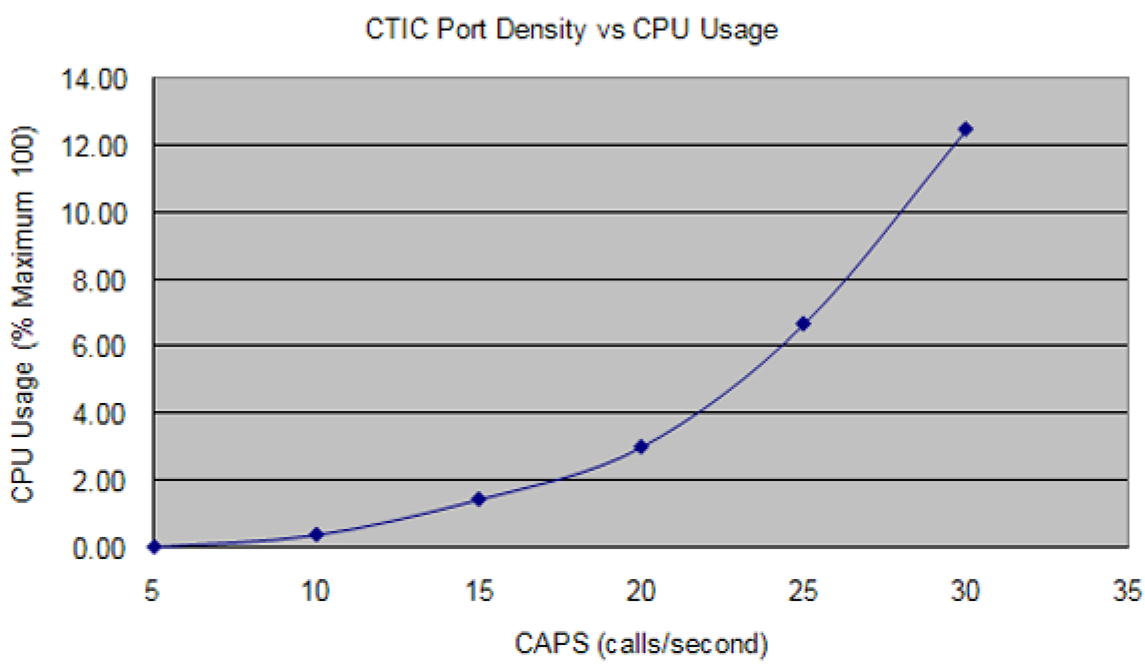


Figure: PD Versus CPU Usage (CTI Connector)

CTI Connector/ICM

The CTI Connector 8.1.4 is integrated with Cisco Intelligent Contact Management (ICM), enabling customers to choose between two Computer Telephony Integration (CTI) deployment modes—Genesys CTI or Cisco CTI.

In these test cases, CTI Connector/ICM performances measured in CAPS and testing was conducted by using two ICM modes of operation—Service Control Interface (SCI) and Call Routing Interface (CRI).

Two bridge transfer scenarios and a blind transfer scenario were tested with CED, Call, and ECC variables passing from the Media Control Platform to ICM. Multiple Media Control Platform instances, configured to use NGI only, achieved the following results:

- One ICM configured in CRI mode achieved 22 CAPS on Windows and 18 CAPS on Linux with GVP 8.1.4 or earlier, and 22 CAPS with GVP 8.1.5.
- Two ICMs configured in SCI mode on both Windows and Linux achieved 30 CAPS.
- Test results indicated a bottleneck on the ICM side of the platform. The graphs in Figure: CPU Usage Versus CAPS—Blind Transfer (CTI Connector/ICM) and Figure: CPU Usage During Blind Transfer (CTI Connector/ICM) depict a sample of the CPU usage when CTI Connector/ICM performs a blind transfer in SCI mode on Windows 2008 R2.

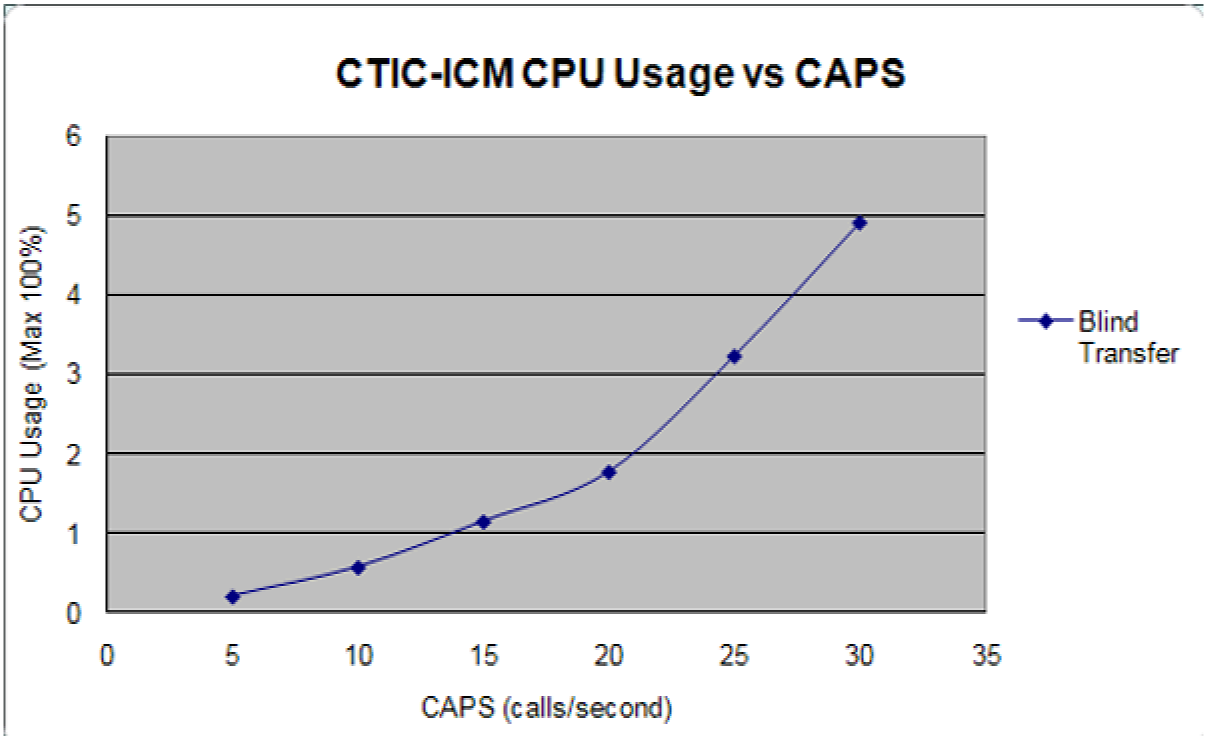


Figure: CPU Usage Versus CAPS—Blind Transfer (CTI Connector/ICM)

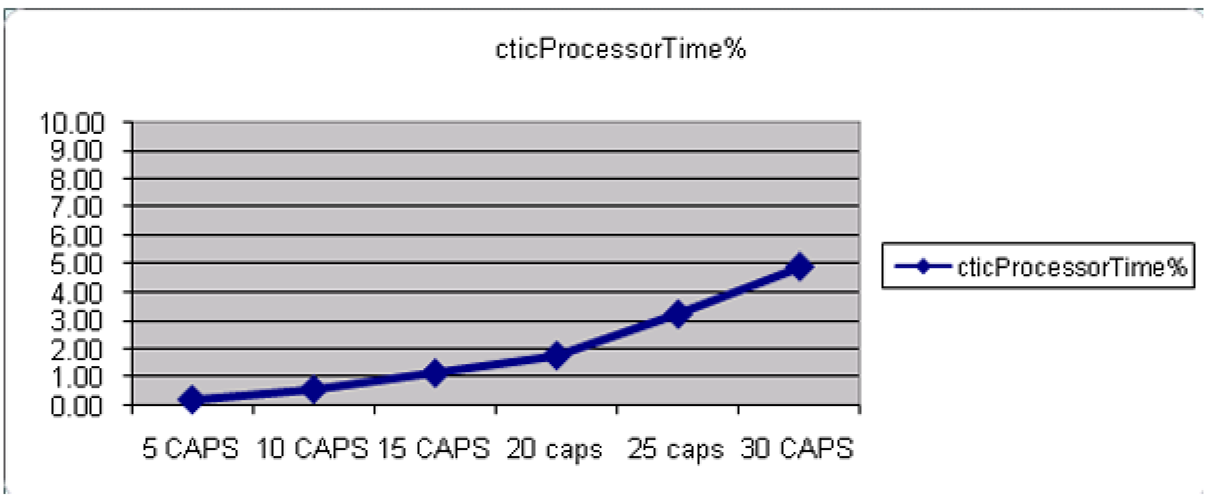


Figure: CPU Usage During Blind Transfer (CTI Connector/ICM)

Supplementary Services Gateway

The Supplementary Services Gateway makes outbound calls through SIP Server, therefore, the call rate (or CAPS) is used to measure the Supplementary Services Gateways performance. Figure: HTTP Request CAPS Versus Notification CAPS provides a comparison of launching call rates (HTTP Requests or targets) and notification of completed calls: (real or achieved CAPS).

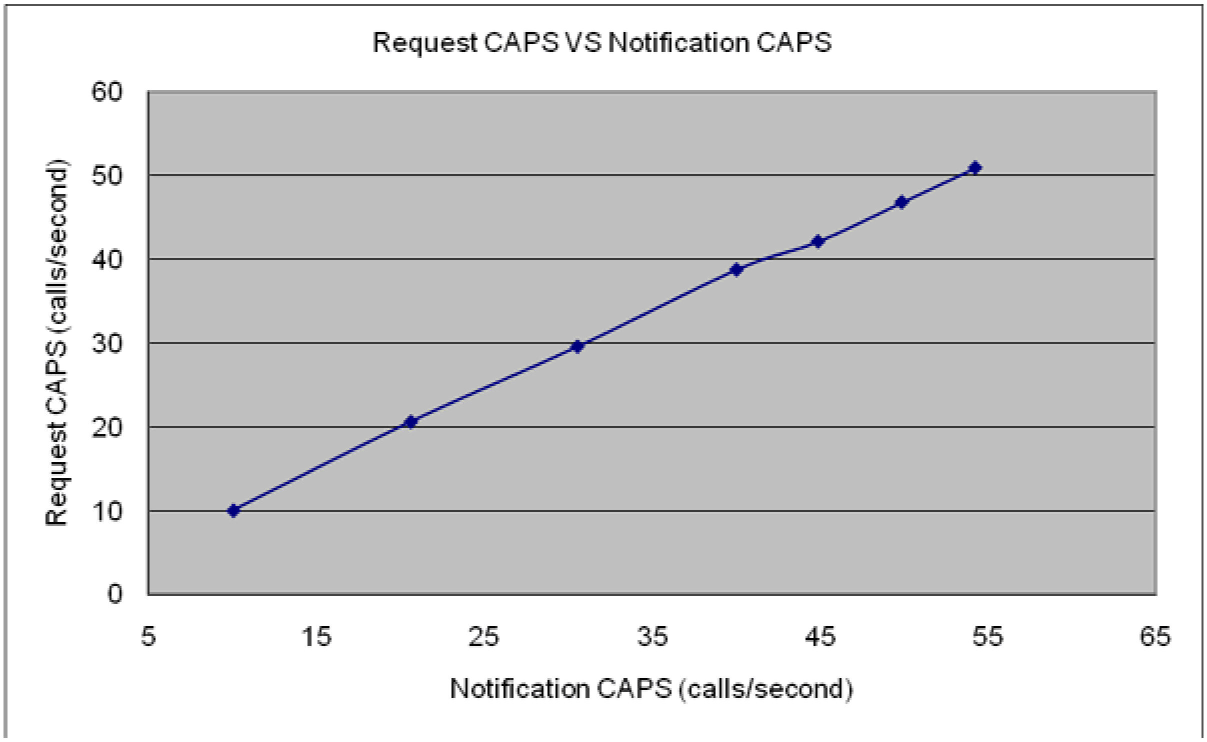


Figure: HTTP Request CAPS Versus Notification CAPS

The Supplementary Services Gateways peak capacity result is 65 CAPS (using GVP 8.1.5 with SIP Server 8.1.0), 50 CAPS (using GVP 8.1.3 or later and SIP server 8.0.4) and 40 CAPS (using pre-GVP 8.1.3 and pre-SIP Server 8.0.4). These results are due to a bottleneck on the SIP Server side of the network--multiple Media Control Platforms are used to provide a sufficient number of ports to handle VoiceXML applications, regardless of their complexity. The call rate can exceed SIP Servers peak capacity, but the requests (which are stored on the database server) tend to accumulate. If egress rate is not high enough, the stored request records can easily reach the database limit of 100,000 records.

Reporting Server

Like the Resource Manager, Reporting Server performance is measured in terms of CAPS. The number of simultaneous calls being processed by GVP does not affect performance and there are no known performance bottlenecks with the Reporting Server software. However, performance can be affected by the database setup. When the Reporting Server is tested without the Media Control Platform in `No DB` mode (Reporting Server drops all call data), it can achieve 800 CAPS.

Capacity reached 800 CAPS when the Reporting Server was tested in No DB mode with the Resource

Manager only (no MCP), and in that configuration, the Reporting Server drops all received call data.

A use case was conducted on Microsoft SQL 2008 and Oracle 10g R2 Enterprise Servers, with the Resource Manager and the Media Control Platform streaming information (including CDR, upstream, and SQA data) to the Reporting Server for each call with a default log level. The result was a peak capacity of 270 CAPS.

The same use case was conducted on an Oracle 11g Enterprise Server only; the result was a peak capacity of 300 CAPS.

> **Tip**
>
> For this test case, the Reporting Server was installed on a Dual Quad Core Xeon computer with a 2.66GHz CPU, separate from the database server. The Microsoft SQL and Oracle Database Servers were installed on a 15-disk Disk Array computer with a Dual Quad Core Xeon, 2.66GHz CPU. For practicality, a simulator was used instead of a real MCP (or a Reporting Server client inside MCP) to run the tests. Also, simulated data (five MCP metrics per call) was used for the MCP simulator to submit the data to the Reporting Server.

The Reporting Server can connect to other GVP components by using TCP or TLS.

Single Server Test Cases

Single-server performance testing was conducted on hardware slightly different from the suggested hardware requirements. The servers used for the performance test cases had the following hardware specifications: 1x Intel Xeon 5160, with a 3.0 GHz CPU, 8GB of RAM, and a 73GB SAS HD.

The following software components were installed:

- Windows 2008 Enterprise Server, SP2, x86 or Windows 2008 Enterprise Server R2, x64
- Microsoft SQL Server 2008 Standard version
- Microsoft Internet Information Server (IIS) configured as a Web Application Server (WAS)
- Management Framework 8.0.3 (Database Server, Configuration Server, Solution Control Server, Message Server)
- Genesys Voice Platform 8.1.3 or 8.1.4 (Resource Manager, Media Control Platform [Squid], Reporting Server)
- SNMP Master Agent 8.0.2
- Genesys Administrator 8.0.3
- SIP Server 8.0.4
- An ASR/TTS Server (Nuance Recognizer 9.0.12, RealSpeak 4.5, Nuance Speech Server 5.0.9)

The following test results indicate higher performance metrics than GVP 7.6 with 48 ports has achieved:

- 600 ports - VoiceXML_App1 (DTMF)
- 100 ports - VoiceXML_App2 (ASR with MRCP v1)
- 160 ports - VoiceXML_App3 (AS ASR/TTS with MRCP v1)
- 120 ports - VoiceXML_App3 (AS ASR/TTS with MRCP v2)

Figure: Port Density Versus CPU Usage (Single Server) depicts the trend of overall CPU usage versus ports density of the VoiceXML_App1 profile.

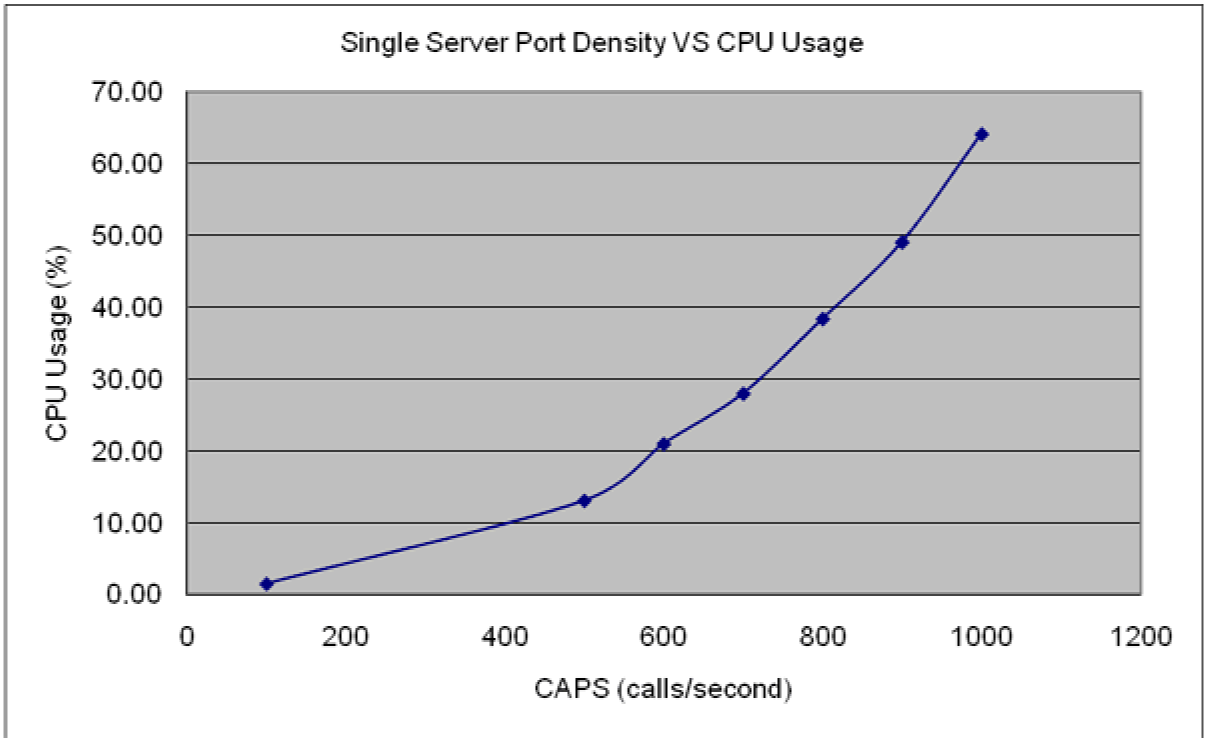


Figure: Port Density Versus CPU Usage (Single Server)

Figure: Call Setup Latency Versus Port Density (MCP only) depicts call setup latency versus concurrent calls in a Media Control Platform only configuration.

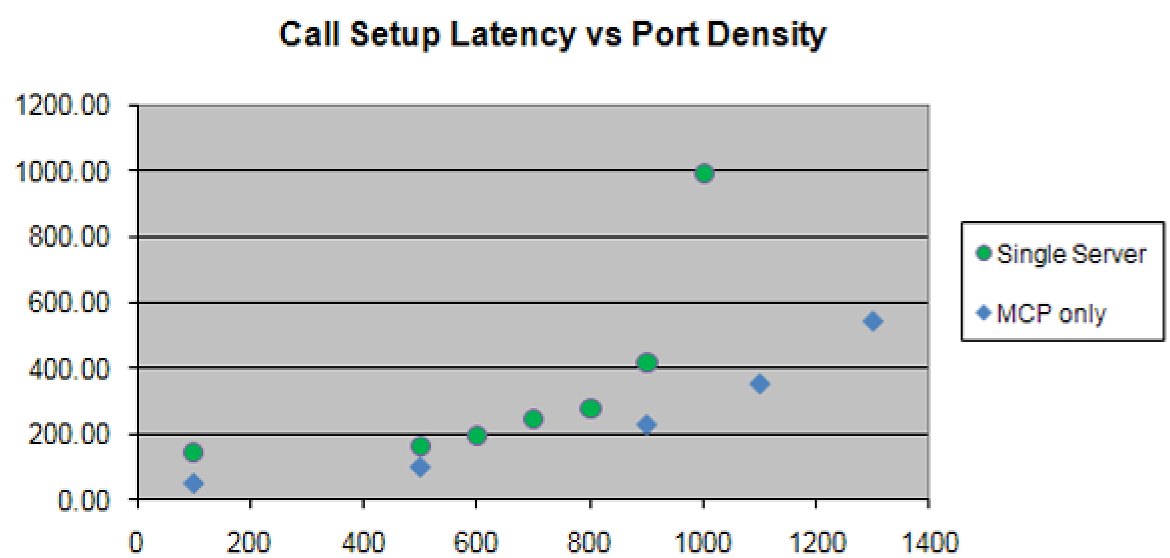


Figure: Call Setup Latency Versus Port Density (MCP only)

In this test case, the latency aligns with the Media Control Platform only configuration with fewer ports configured. Here, the latency is slightly than higher, because in a single-server configuration, the Resource Manager and SIP Server are configured before Media Control Platform. Latency jumps beyond peak capacity after 800 ports.

top | toc

Multiple MCP Instances and Virtual Machines Test Cases

The Media Control Platform-only configuration is also measured with ESXi 4.1 Hypervisor VMs and multiple instances on the same host to compare performance metrics. The testing was conducted on a 2x Core 2Quad Xeon x5355, with a 2.66 GHz CPU, and 8 Cores with 12GB of RAM (higher than recommended).

- First Series: Multiple MCPs
- Second Series: Multiple MCPs
- Third Series: Multiple MCPs
- Jitter Quality on Virtual and Actual MCP Machines

First Series: Multiple Media Control Platforms

The first series of performance tests were conducted on servers with 1, 2, 4, and 8 VM images installed, and only one Media Control Platform on each VM, with the following hardware configuration:

- 1 VM 8 virtual CPUs 12 GB RAM for the VM
- 2 VM 4 virtual CPUs 6 GB RAM for each VM
- 4 VM 2 virtual CPUs 3 GB RAM for each VM
- 8 VM 1 virtual CPUs 1.5 GB RAM for each VM

To provide comparisons, the operating system used for the VMs was Windows 2008 Enterprise SP2, x86, which was also the operating system that is installed on the host used to test multiple Media Control Platform instances (1, 2, 4, and 8, respectively).

The VoiceXML_App1 (DTMF) was used as the standard application profile. Figure: Port Density—Virtual Machines Versus Media Control Platforms depicts the peak capacity that was obtained from the configurations that were described above.

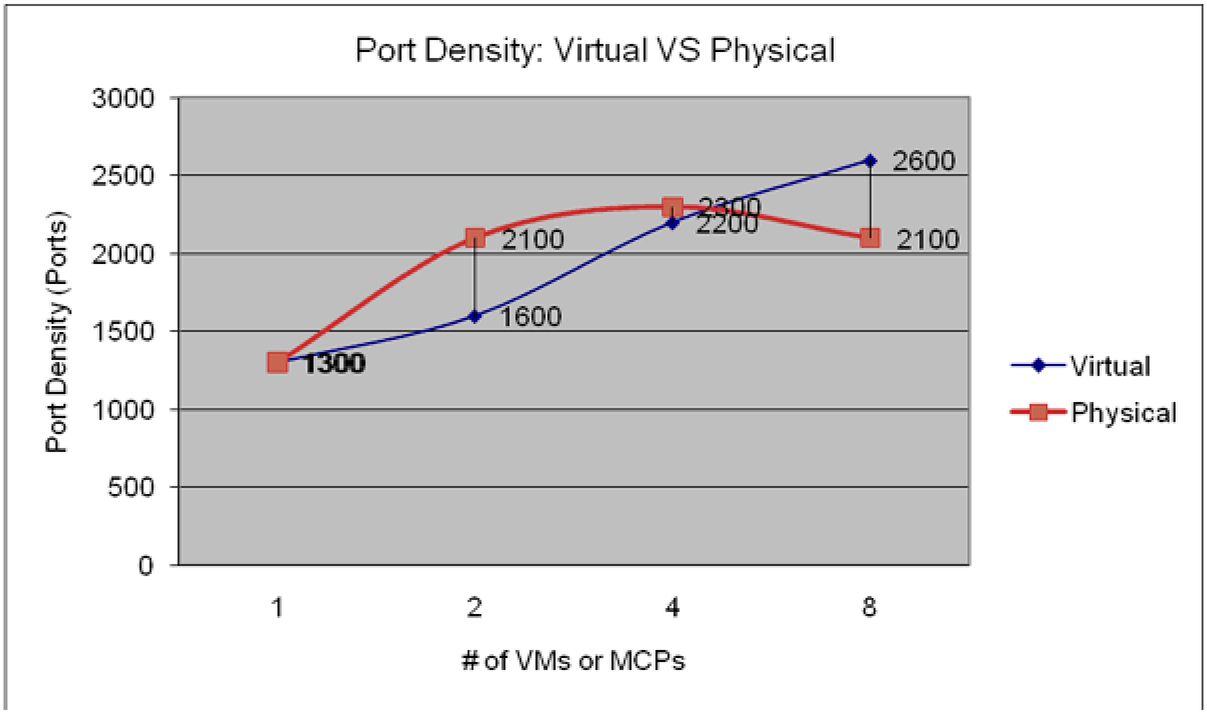


Figure: Port Density—Virtual Machines Versus Media Control Platforms

Figure: CPU Usage—Virtual Machines Versus Media Control Platforms shows the highest peak capacity when 8 VMs (2600 ports) are configured, while the highest peak capacity when multiple Media Control Platform instances (2300 ports) are configured is at four instances (peak capacity is actually lower when eight Media Control Platform instances are configured). You can utilize the greatest number of ports when the number of VMs corresponds to the number of CPUs (Cores). However, using the multiple VM configurations results in a higher percentage of CPU usage. Figure: CPU Usage—Virtual Machines Versus Media Control Platforms depicts the number ports when CPU usage is measured during testing.

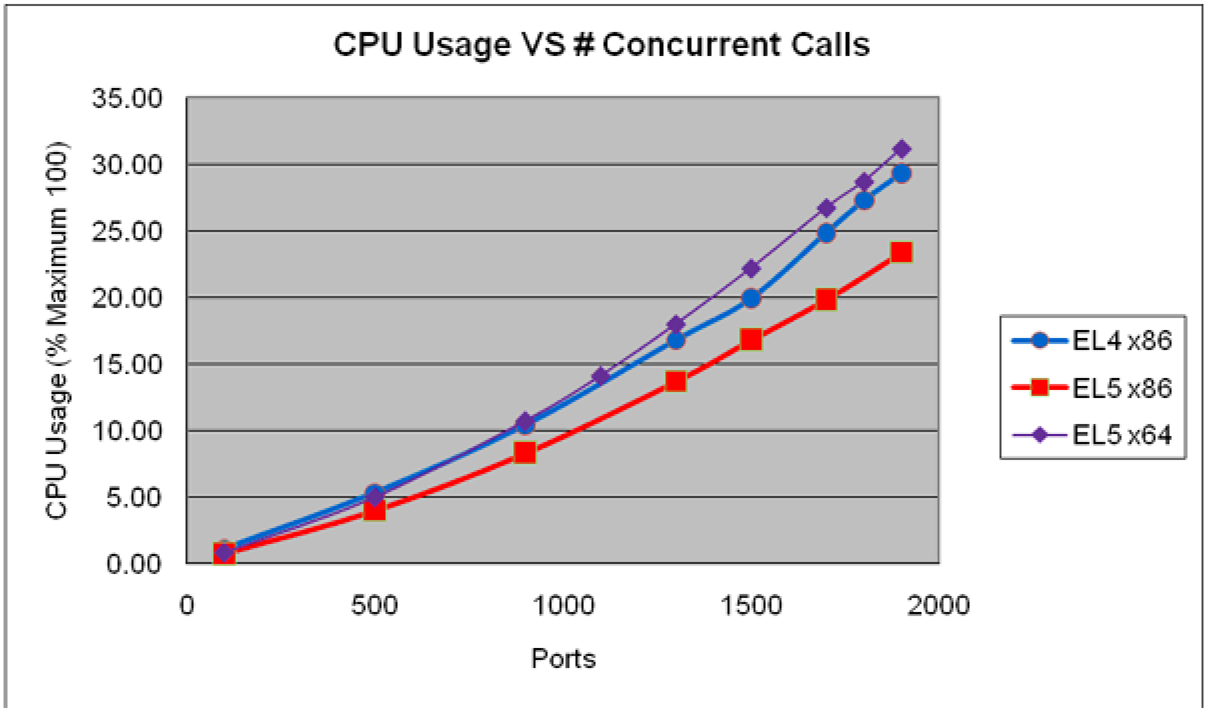


Figure: CPU Usage—Virtual Machines Versus Media Control Platforms

The graphs in Figure: Call Setup Latency Versus Port Density (Virtual Machines) and Figure: Call Setup Latency Versus Port Density (Actual MCPs) provide a comparison of the call setup latency when multiple VMs are configured versus multiple Media Control Platform instances.

The latency is lower when more VMs or more Media Control Platform instances are at the same port density, because with more VMs or more Media Control Platforms, fewer calls are distributed to each VM or Media Control Platform instance.

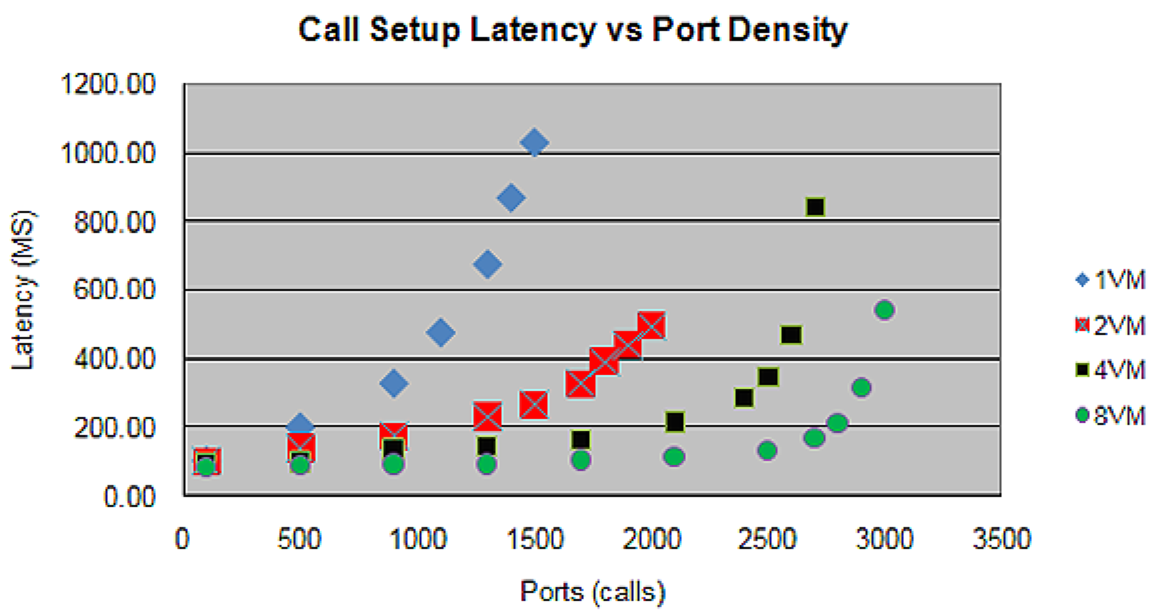


Figure: Call Setup Latency Versus Port Density (Virtual Machines)

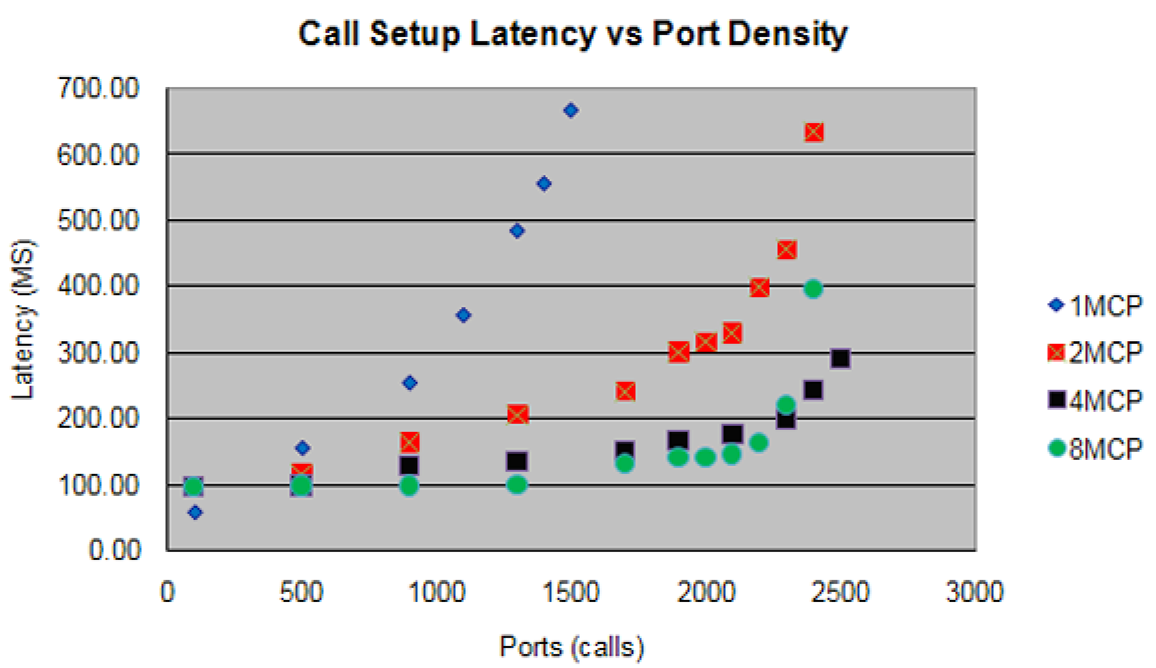


Figure: Call Setup Latency Versus Port Density (Actual MCPs)

The next three graphs provide a comparison between multiple VMs and multiple Media Control Platform instances. The graphs depict 1-to-1, 2-to-2, 4-to-4, and 8-to-8 comparisons, respectively.

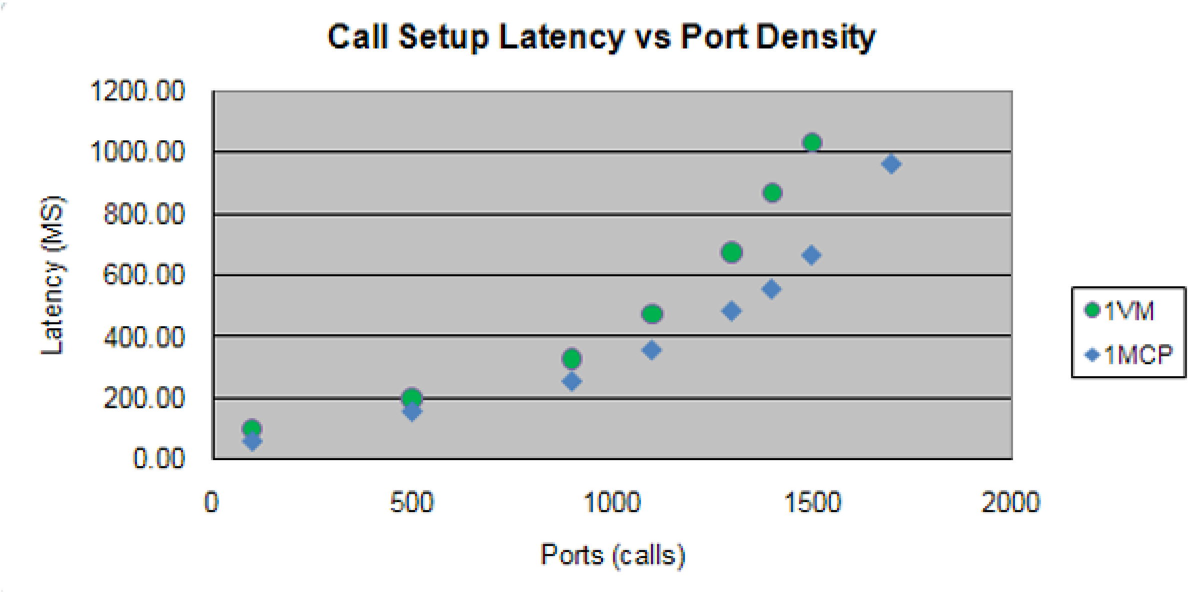


Figure: Call Setup Latency Versus Port Density (1 VM-to-1 MCP)

In Figure: Call Setup Latency Versus Port Density (1 VM-to-1 MCP), the same peak ports is achieved when comparing 1 Media Control Platform instance to 1 VM, however, the Media Control Platform produced lower call setup latency.

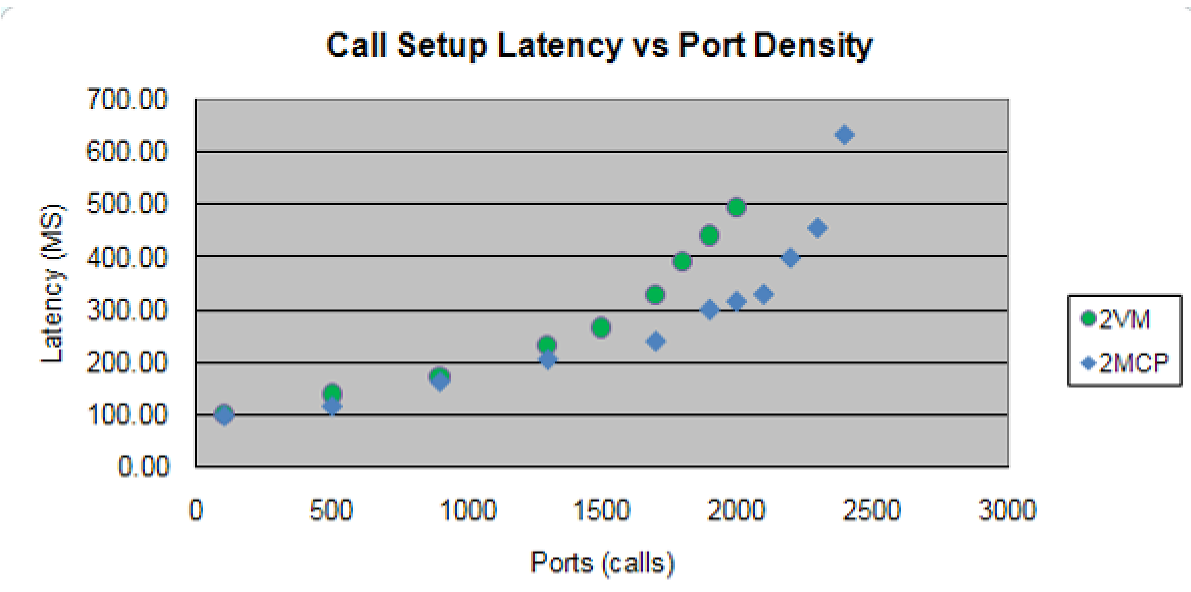


Figure: Call Setup Latency Versus Port Density (2 VM-to-2 MCP)

Figure: Call Setup Latency Versus Port Density (2 VM-to-2 MCP) indicates that 2 Media Control Platform instances perform better than 2 VMs with lower latency and higher peak ports.

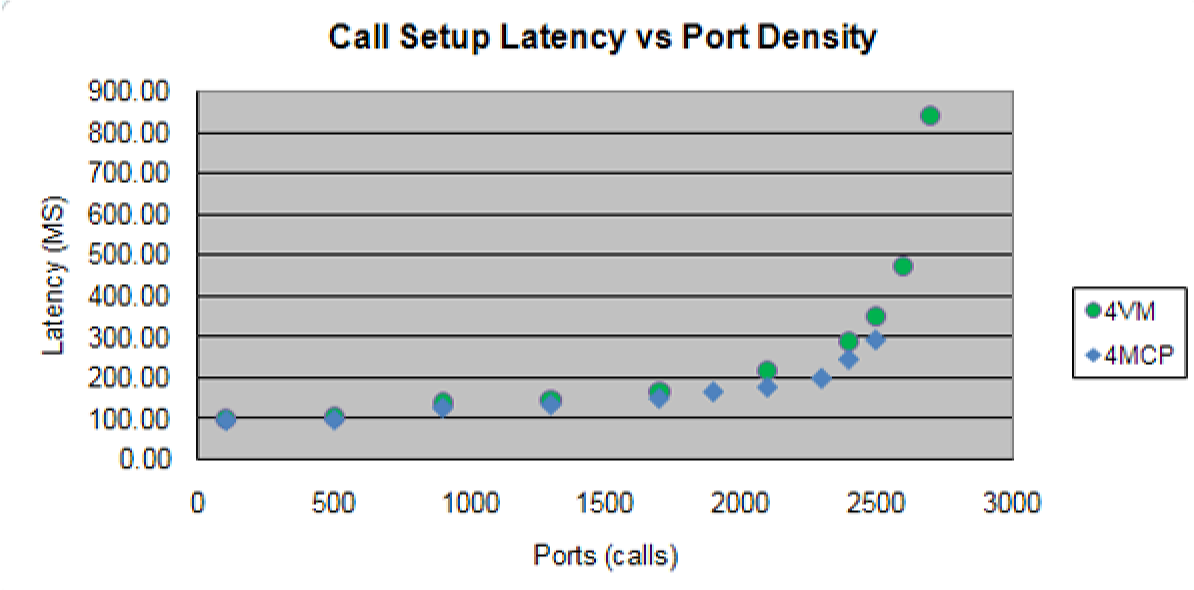


Figure: Call Setup Latency Versus Port Density (4 VM-to-4 MCP)

Figure: Call Setup Latency Versus Port Density (4 VM-to-4 MCP) indicates that the performance of the 4 VMs and 4 Media Control Platform instances are quite close. The peak capacity and the latency trends are almost the same.

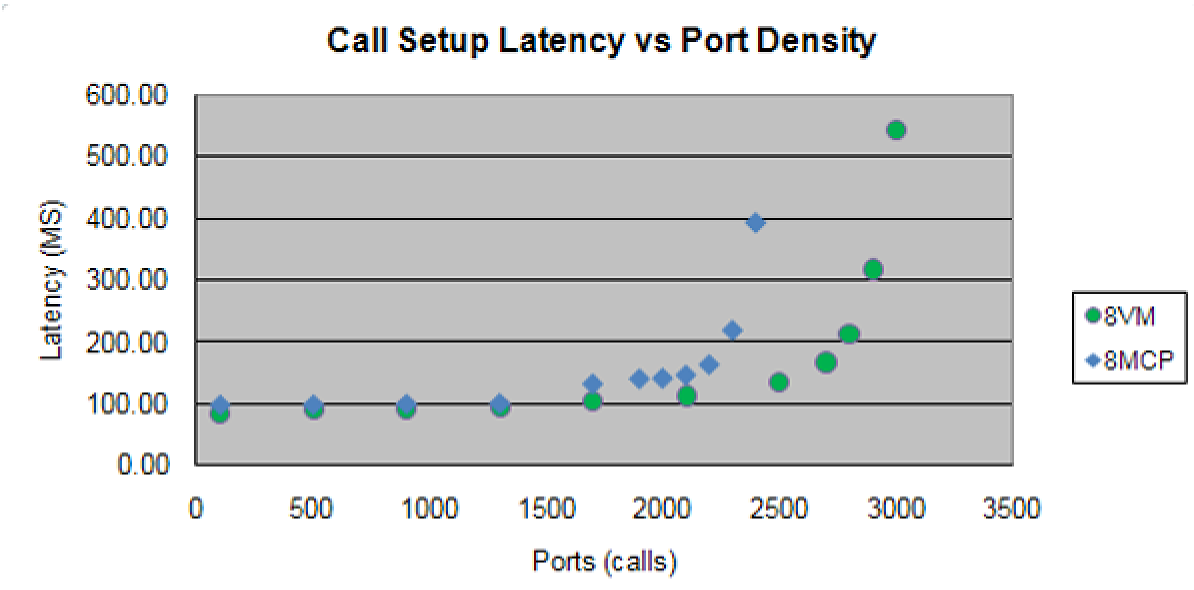


Figure: Call Setup Latency Versus Port Density (8 VM-to-8 MCP)

Figure: Call Setup Latency Versus Port Density (8 VM-to-8 MCP) indicates that 8 VMs perform better than 8 Media Control Platform instances. The VMs produce lower latency and higher peak ports.

Second Series: Multiple Media Control Platforms

The second series of performance tests were conducted on servers with 1 VM with 1 Media Control Platform instance using 1 CPU and then, the same configuration using 2 CPUs. These tests were executed by using 8 VMs and a 325-port load as the baseline, which is the highest peak capacity we could attain.

As shown in Table: CPU Usage—1 Media Control Platform per VM, the test conducted by using 2 CPUs (per VM) resulted in slightly higher CPU usage than the test with 1 CPU, while both results were only about 1/8 of the overall CPU usage when 8 VMs were configured.

Table: CPU Usage—1 Media Control Platform per VM

VMs	MCPs per VM	CPUs per VM	Ports	Overall CPU Usage
8	1 (8 total)	1 (8 total)	2600	83.69%
1	1 (1 total)	1 (1 total)	325	9.45%
1	1 (1 total)	2 (2 total)	325	9.76%

Third Series: Multiple Media Control Platforms

The third series of performance tests were conducted on servers with 1 VM with 2 Media Control Platform instances and then, the same configuration with 2 VMs and 4 VMs, respectively. See Table: CPU Usage—2 Media Control Platforms per VM.

Table: CPU Usage—2 Media Control Platforms per VM

VMs	MCPs per VM	CPUs per VM	Ports	Overall CPU Usage
2	1 (2 total)	4 (8 total)	1600	69.14%
2	2 (4 total)	4 (8 total)	1600	69.22%
4	1 (4 total)	2 (8 total)	2200	80.10%
4	2 (8 total)	2 (8 total)	2200	77.94%

Test results do not indicate higher ports as capacity peaks, because CPU usage is already high. However, the call setup time gets shorter. See Figure: Call Setup Latency Versus Port Density—2 VMs and Figure: Call Setup Latency Versus Port Density—4 VMs.

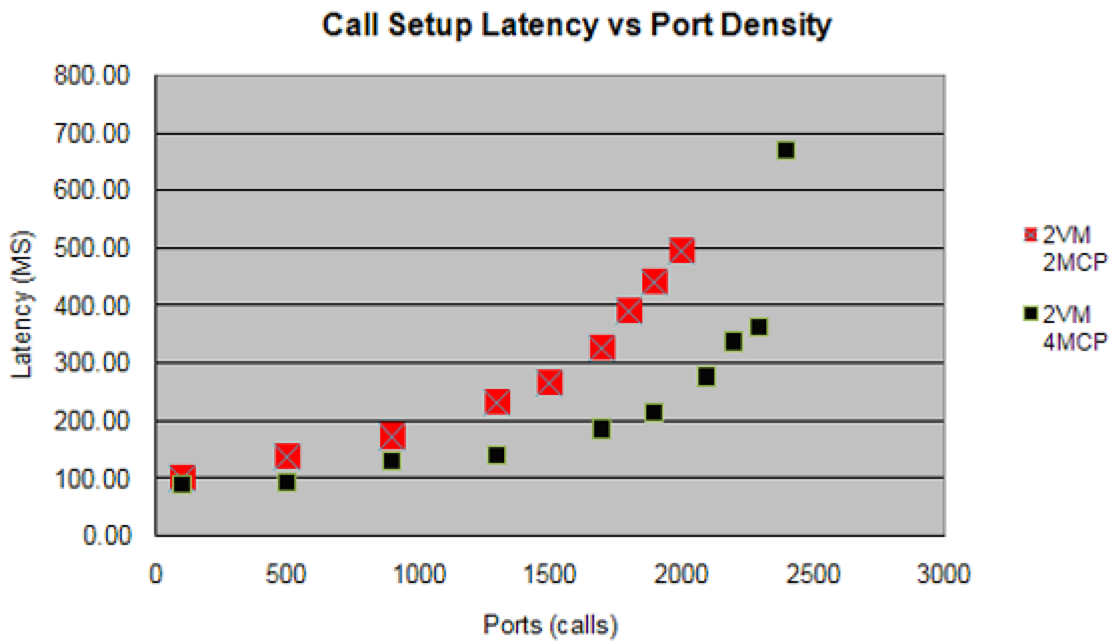


Figure: Call Setup Latency Versus Port Density—2 VMs

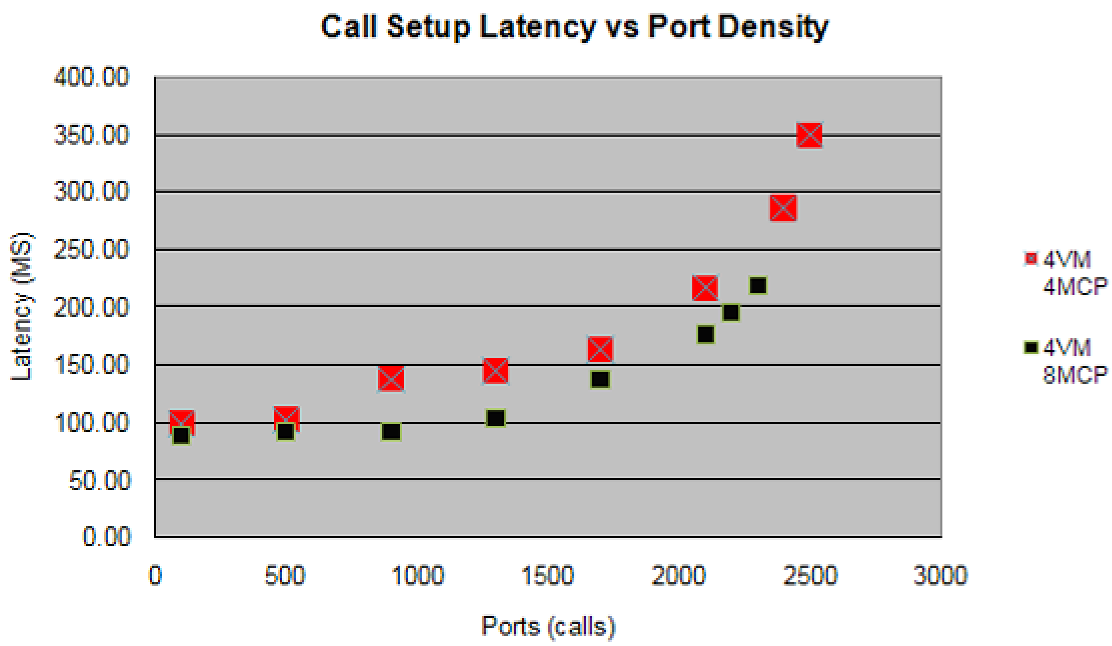


Figure: Call Setup Latency Versus Port Density—4 VMs

Results indicated somewhat shorter call durations, as indicated in the next two graphs showing the 95 percentile of call duration.

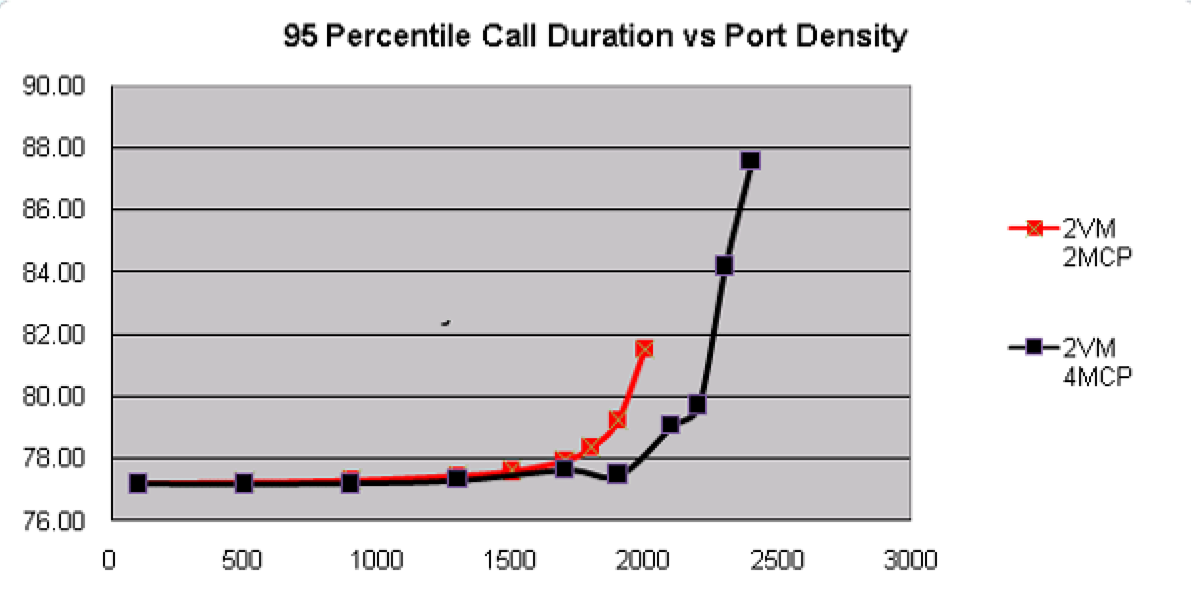


Figure: Call Duration Versus Port Density—2 VMs

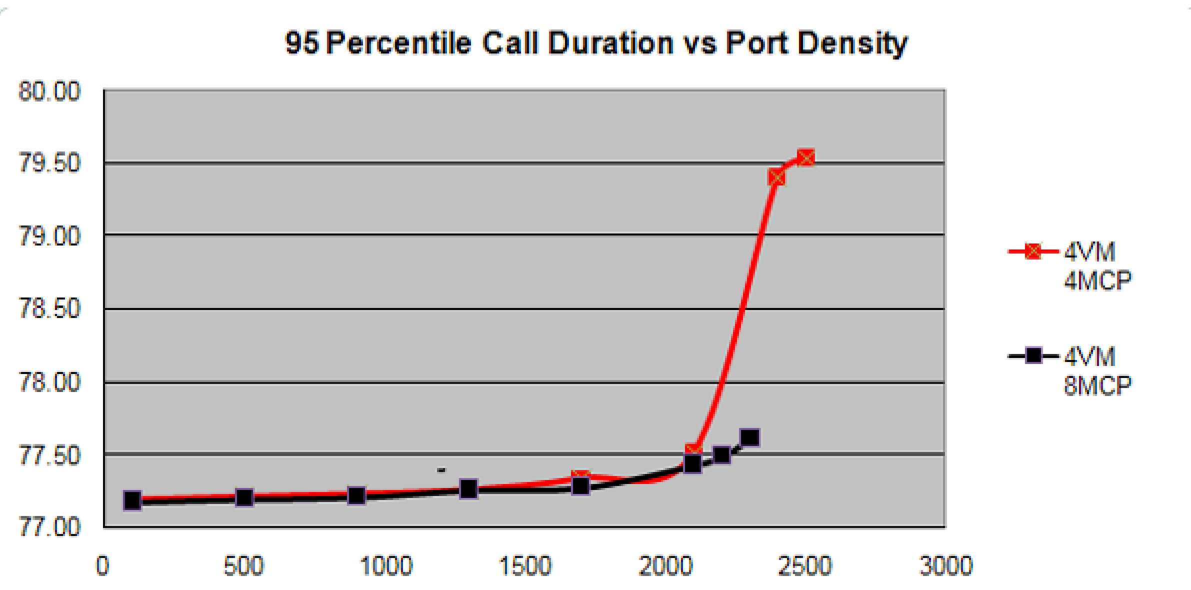


Figure: Call Duration Versus Port Density—4 VMs

The same test was repeated for GVP 8.1.5 on Windows 2008 Server R2 x64, with the latest hardware, which obviously achieves higher peak capacity. From the graph below, you can see that the latencies of x64 Win and x86 Win systems are quite in line with one another at lower ports, but a marked difference appeared at higher ports around peak capacity.

Jitter Quality on Virtual and Actual Media Control Platform Machines

Two metrics are used to measure jitter quality—Jitter Average (the weighted average of a stream's packets) and Jitter Max (the maximum number streamed packets). Two VMs and 2 Media Control

Platform instances were used to test jitter quality. As expected, the results revealed some differences between the virtual and actual machines: See the next two graphs.

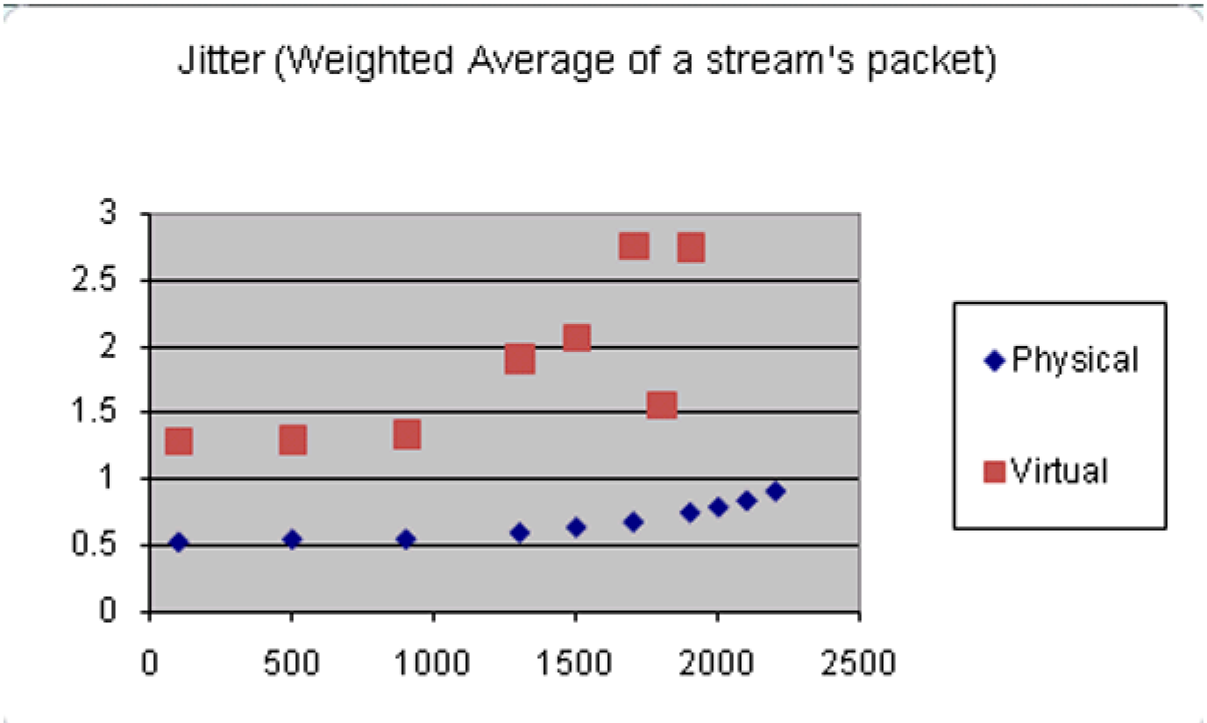


Figure: Jitter (Weighted Average)

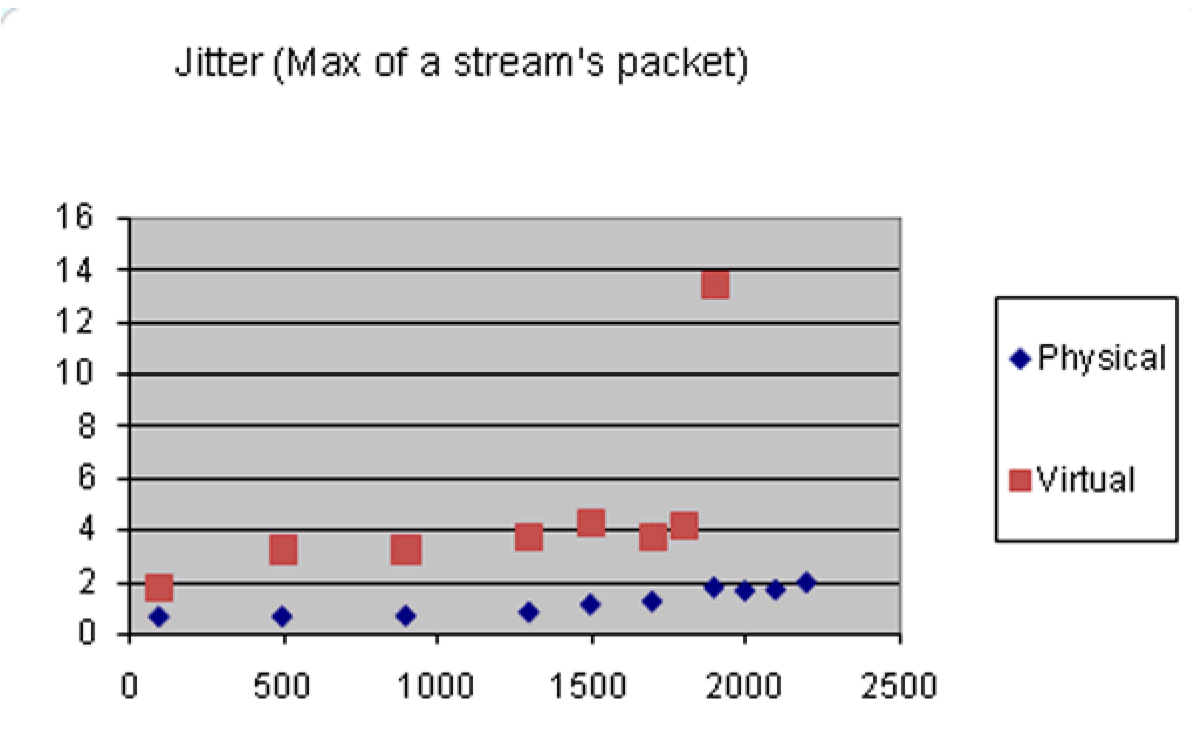


Figure: Jitter (Maximum of Streamed Packets)

From the perspective of media latency, the difference is minor (less than or equal to 5%). In Figure: Speech Resource by Audio Latency, speech response latency was tested with 1000 words of grammar.

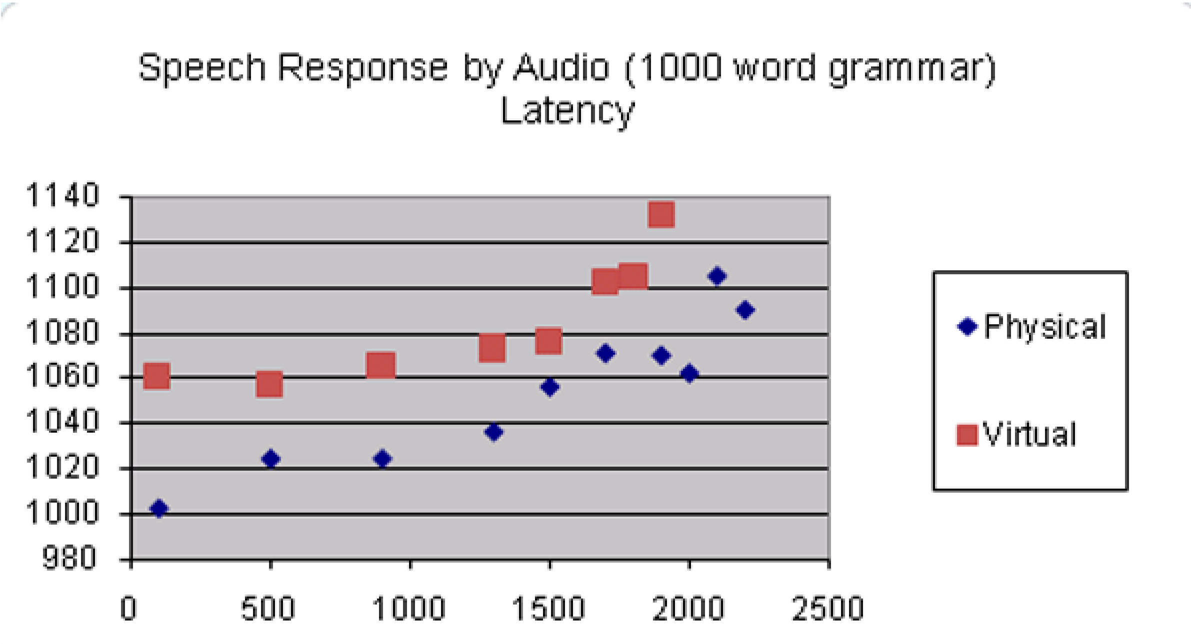


Figure: Speech Resource by Audio Latency

In Figure: Speech BargeIn to TTS Latency, speech bargein is compared to TTS latency the difference is between 20 and 50 milliseconds.

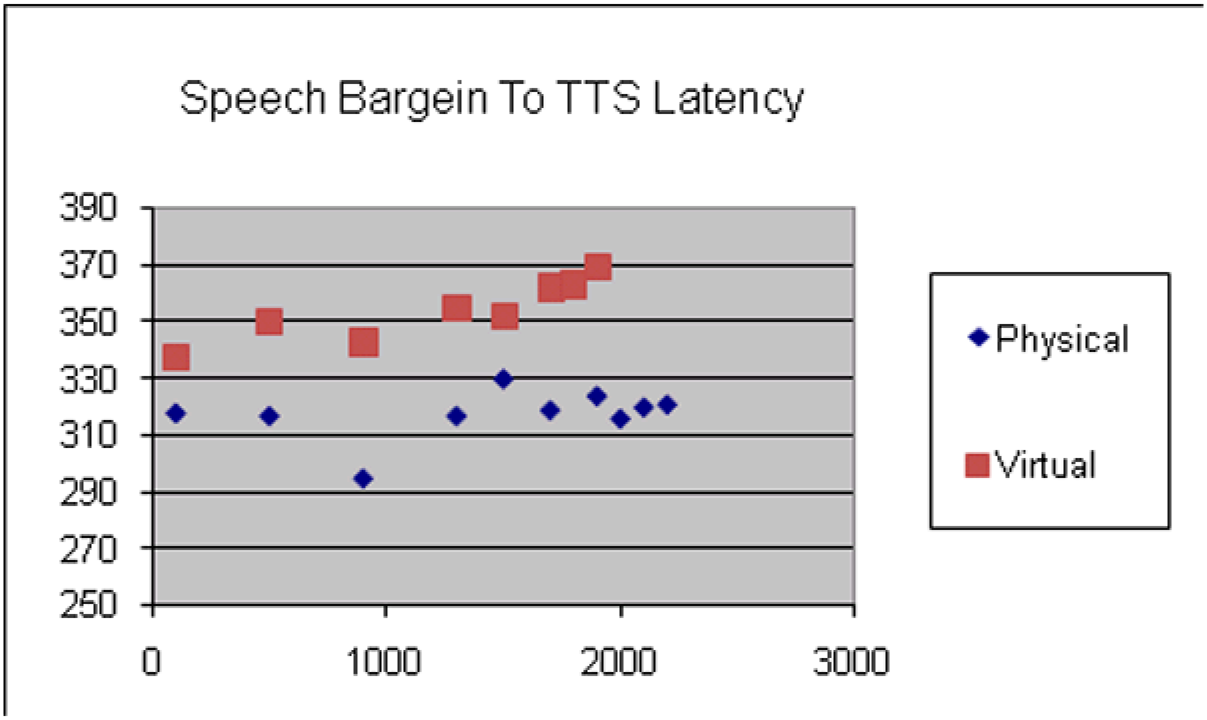


Figure: Speech BargeIn to TTS Latency

top | toc

Call Setup Latency Test Cases

In the following test cases, maximum capacity was achieved within the constraints of specific thresholds. However, the system was also tested beyond the recommended capacity to determine the extent of performance degradation.

The test case in Figure: Port Density versus Call Setup Latency uses the VoiceXML_App1 profile (see VoiceXML Application Profiles) to show how the CSL increases as the PD increases. The rate at which the CSL increases is relatively constant until the system reaches a bottleneck—for example, when the system load is beyond peak capacity.

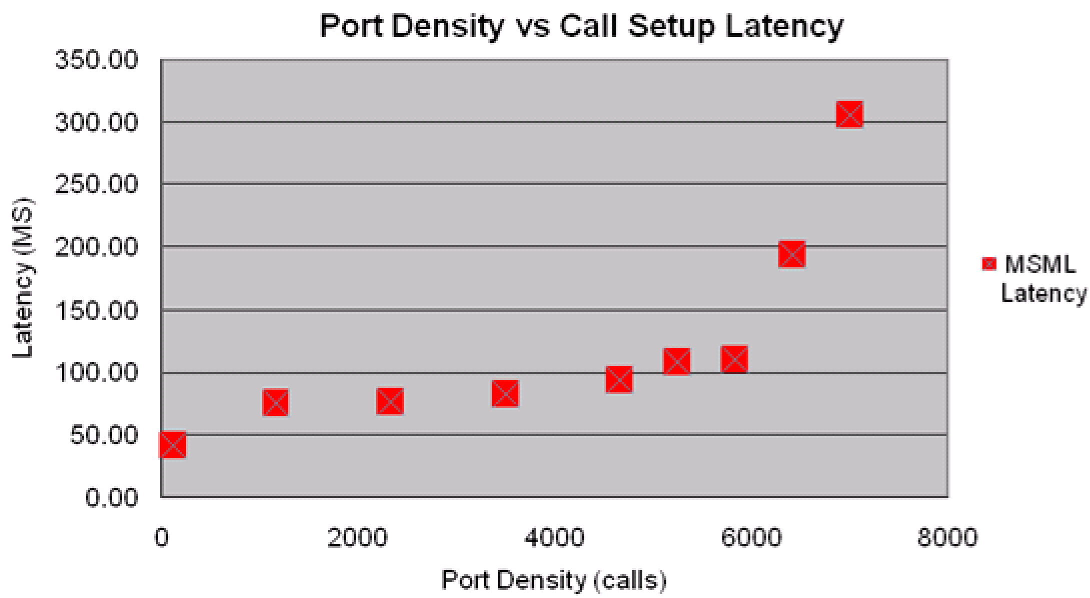


Figure: Port Density versus CSL

Caller Perceived Latency Test Case

The graph in Figure: Port Density versus DTMF shows the DTMF response-to-audio-prompt latency at various port densities (relative to the peak capacity indicated in Table: GVP VOIP VXML/CCXML Capacity Testing). Notice that the TTS prompts produce ~300 ms more latency than the audio file prompts. This is due to the beginning silence played by the TTS engine.

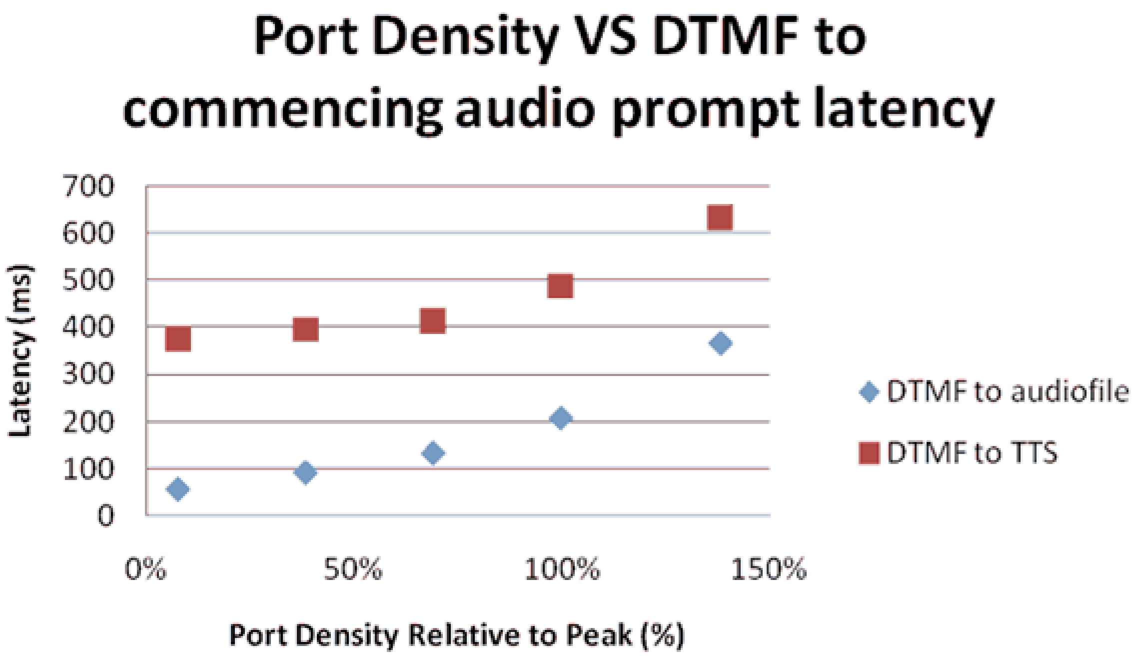


Figure: Port Density versus DTMF

When there is speech input, additional latency is usually caused by the ASR engine. In Figure: Port Density versus Speech, the latency result from 1000 words of grammar using the Nuance OSR3 MRCP version 1 (MRCPv1) engine. The result can vary, depending on the type of MRCP engine used, the type of speech grammar used, and the load on the speech engine.

The performance results in Figure: Port Density versus Speech were obtained from isolated ASR engines supporting the same number of recognition sessions at all Media Control Platform port densities; the MRCP engines did not cause a bottleneck. Therefore, depending on the load on the Media Control Platform, it can add as much as ~100 ms of latency.

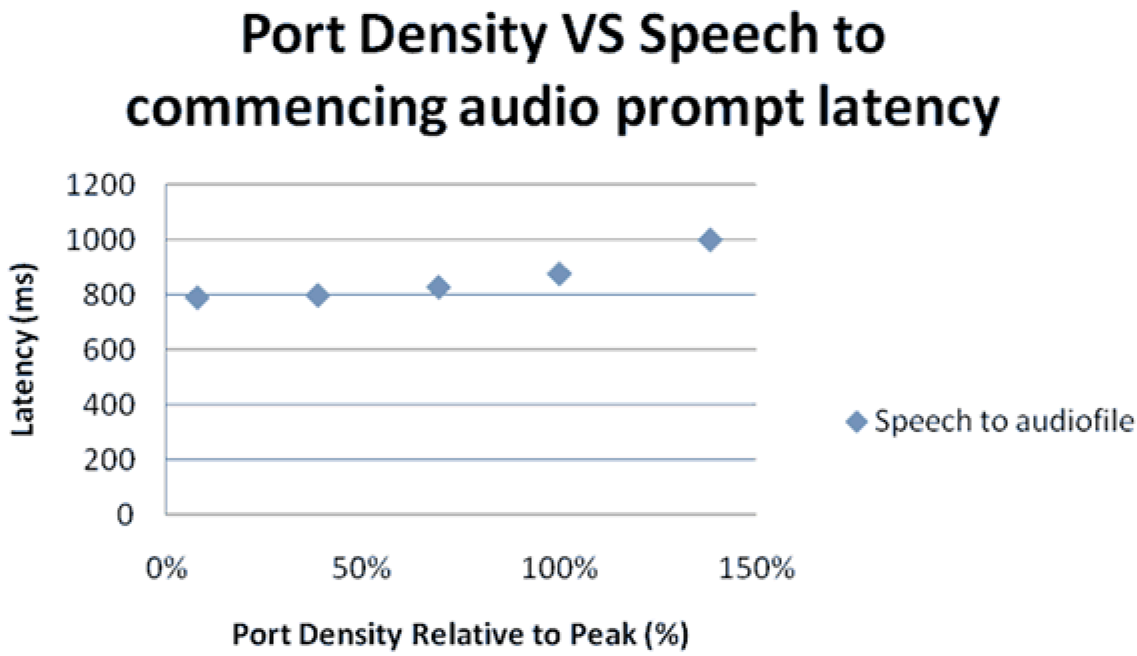


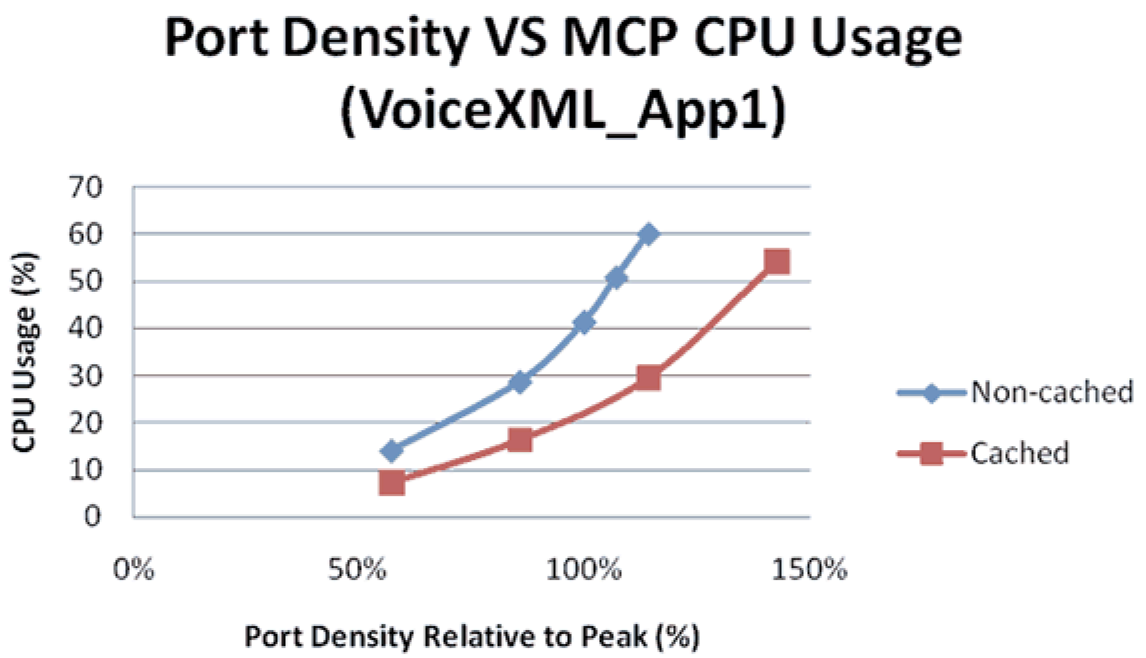
Figure: Port Density versus Speech

top | toc

Cachable VoiceXML Content Test Cases

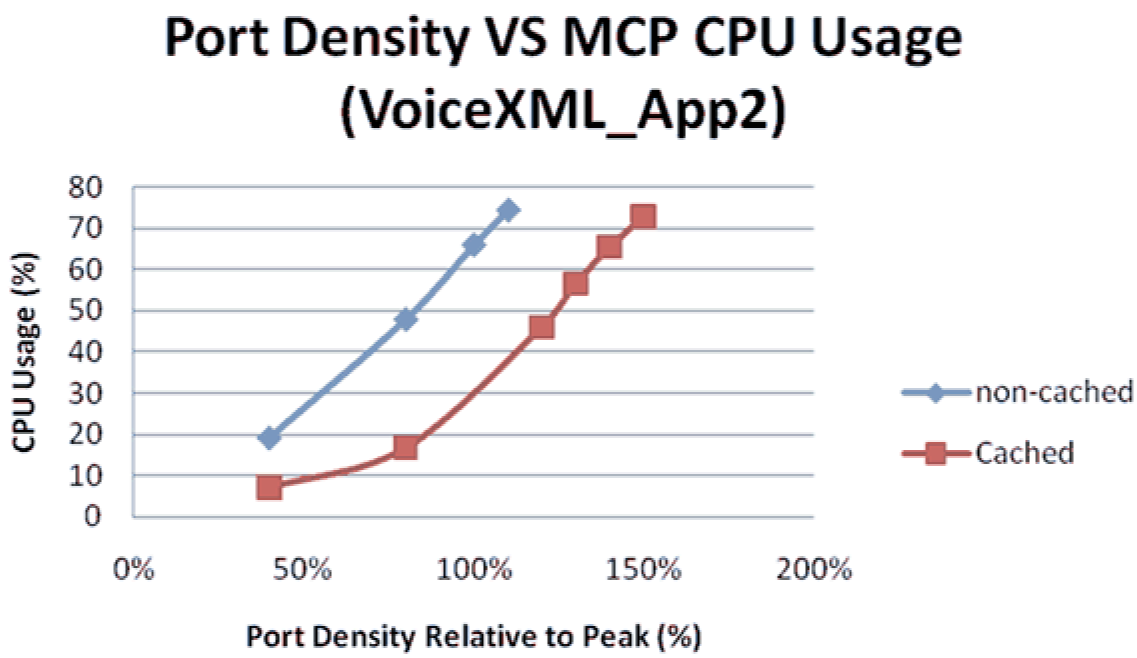
In the following test cases, maximum capacity was achieved within the constraints of specific thresholds. However, the system was also tested beyond the recommended capacity to determine the extent of performance degradation.

GVP can cache internal, compiled VoiceXML objects. Caching VoiceXML objects saves a significant amount of compilation time, resulting in less CPU usage. The VoiceXML_App1 application (see VoiceXML Application Profiles) was used for the test case in Figure: Port Density vs. CPU (VoiceXML_App2) and was based on the peak capacity indicated in Table: GVP VOIP VXML/CCXML Capacity Testing.



Port Density vs. CPU (VoiceXML_App1)

The more complex the VoiceXML content, the greater the benefit of having cachable content. The test case in Figure: Port Density vs. CPU (VoiceXML_App2) (below) is similar to the one in Figure: Port Density vs. CPU (VoiceXML_App1) (above), except that the more complex VoiceXML_App2 application was used (see VoiceXML Application Profiles).



Port Density vs. CPU (VoiceXML_App2)

In Figure: Port Density vs. CPU (VoiceXML_App1) and Figure: PD vs. CPU (VoiceXML_App2), the processing of cachable and non-cachable content are compared with the Media Control Platform using the same level of CPU consumption for both applications. The following results show the benefits of using cachable content:

CPU Consumption—Media Control Platform at peak capacity:

- 15% less consumption than non-cached content using VoiceXML_App1.
- ~30% less consumption than non-cached content using VoiceXML_App2.

Port Density—CPU consumption at same level for both applications:

- ~30-35% greater than non-cached content using VoiceXML_App1.
- ~50% greater than non-cached content using VoiceXML_App2.

top | toc