Performance tests

telisca Recording

Reference: 180822

sales@telisca.com +33 1 46 45 05 12 www.telisca.com

Table of Contents

1 Performance tests description

1.1 telisca Recording overview and architecture

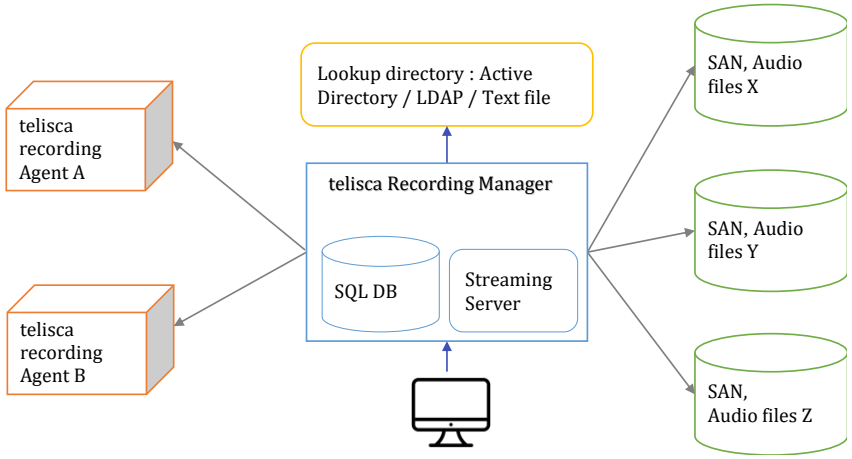
telisca Recording solution is based on a Recording Manager and one or several Recording Agents.

telisca recording Agent:

- Based on IP phones Builtin Bridge or CUBE
- Based on Cisco CUCM Recording Profile settings
- Automatic or selective recording
- Merge calls' segment and converts to MP3
- Support G711, G722, G729 CODECs
- Support consult, transfer, conference calls association
- Support Secured SIP and SRTP
- Support fault tolerant configuration

telisca Recording Manager :

- Archive recordings on efficient network storage,
- Encrypt recordings,
- Automatic purge depending of retention delay (per department/company),
- Annotation of recordings with information provided by the company's internal directory
- Authentication/segmentation of users by Windows or CUCM security groups,
- Web interface for searching recordings by last name, first name, location, with contacts segmented and filtered by department/company,
- Listen to streamed recordings via a web interface, with option to download recording,
- Add notes while listening
- History of recordings which have been reviewed (date, compliance officer, comments, …),



telisca recording is organized in two main modules:

- telisca Recording Agent: Handle a SIP trunk that receive the call sent by the IP Phones Builting Bridge of by CUBE and save it in RAW files. Export the RAW files in mp3 format, merging the different legs of the calls.
- telisca Recording Manager: Manage the metadata database, save and encrypt the mp3 files on SAN. Allow to search and play the recording. Manage directories, CDR, CUCM provisioning …

1.2 Perfomance goal

The goal of the performance tests is to test telisca Recording behavior when the application is stressed and to provide sizing guidelines depending on the number of simultaneous recorded calls and depending of the chosen configuration:

1) Heavy contact center workload of administrative usage
2) telisca Recording Agent & telisca Recording Manager on the same server or on different servers.

3) Conversion & merge of recording media to mp3 immediately or in batch mode
4) Encryption of recording media or not.

1.3 Test platform architecture

We are using telisca Morning Check application to generate calls and play G711 audio to the Recording SIP trunk. This simulates the span call generated from the BIB (Build In Bridge) when recording. However, the difference is that the BIB sends one call for incoming media and one call for outgoing media.

In order to simulate the BIB behaviour, we have evaluated the increased resources to record two media instead of one and to merge/compress two media instead of one. We have found that the recording more than double the CPU resource. However, the merge + compression of two media files use the same CPU resources as for one media file. We have then evaluated the percentage or resources used by recording compare to compression for different number of simultaneous calls. We have doubled the resources used by the recording and kept unchanged the resources used by the compression.

With one media:
- CPU to record = n%
- CPU to compress + database insert = 100 - n%

With two medias, CPU to record doubles, so the resource is increased by n%. So as the increase of global CPU resources is proportional to the number of recorded calls, in order to evaluate the resource used for 100 real BIB calls with two medias, we have to execute the test for 100 + n calls with one media. Or if we test with Morning Check for 100 of calls, then it simulates 100 * 100 / (100 + n) calls.

Morning Check scripts generate reports that we browse to check that the number of calls set are really dialled, answered by the trunk sip and the audio media played without failure.

telisca Recording application agent + Manager are executed on the same or separate VMWare virtual machines with different memory sizing (minimum 4GB) and different numbers of vCPU. The hardware server used is a Dell PowerEdge 430. The CPU is an Intel Xeon E5-2620 v4 (2,1GHz, 8C/16T, 20MB cache memory) with a Passmark of 11 349.

The CPU, memory and disk usage are monitored to get the average value on the heaviest load and to analyse the workload evolution during the test.

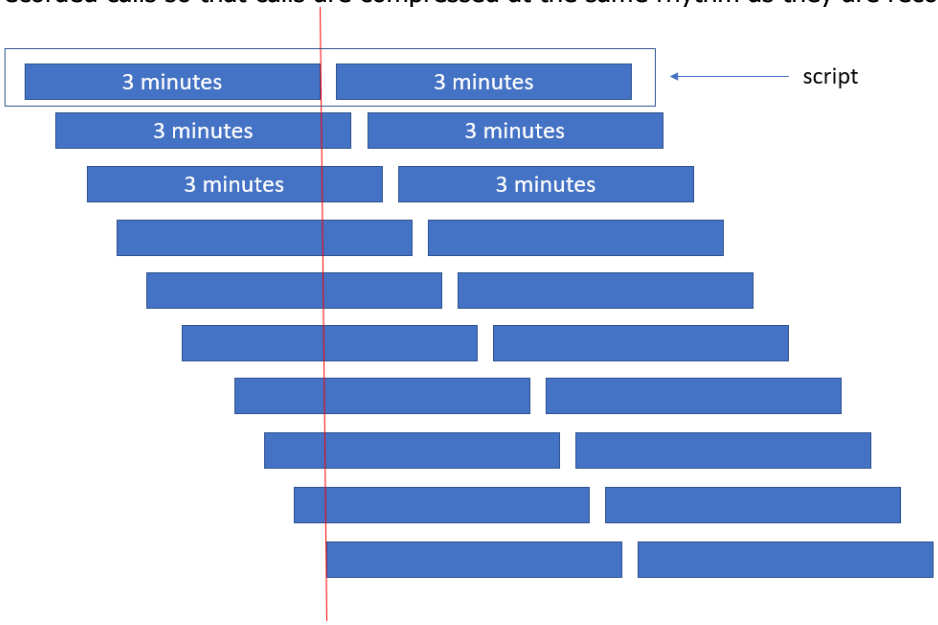
In order to check that workload is not to high we manually test by searching and listening the recorded media. This allows to test:
- The delay to search and play
- The quality of the recorded media
- The quality of the streaming.

1.4 Scenario

The tests are conducted in a Contact Center context which is the more demanding. The call duration is 3mn and the time between calls is 30s (which is short, because most of the time thee agent has to do a call wrap-up). In this case the BHCA (Busy Hour Call Attempt) is quite high so we will divide by 3 to give also guide lines for an administrative telephony usage scenario.

The calls are generated with a small delay between each other to avoid an abnormal peak on CUCM and recording resource. We make sure however that in the middle part of the test all the calls are recorded at the same time. A second call is generated so that the second call is recorded while the compression is executed. The compression is executed periodically (by default ever 10s) and the

several compressions can be launched. This parameter is changed to adapt to the number of recorded calls so that calls are compressed at the same rhythm as they are recorded.



The script is executed several times for a test duration of 15 to 20mn in order to make sure the workload is stabilized. The average CPU / memory / disk load is recorded for the highest 5 minutes of the test.

1.5 Tests parameters by number of recorded calls

Here are the values of the different parameters of Morning Check and Recording to achieve the number of simultaneous recorded calls simulated.

Nb. calls terminated / minute = Nb. Test calls x 60s / (180s + 30s).

The number of compressed files should higher or equal. It is calculated as = 60 * Nb. Compressed Files / Compress Period.

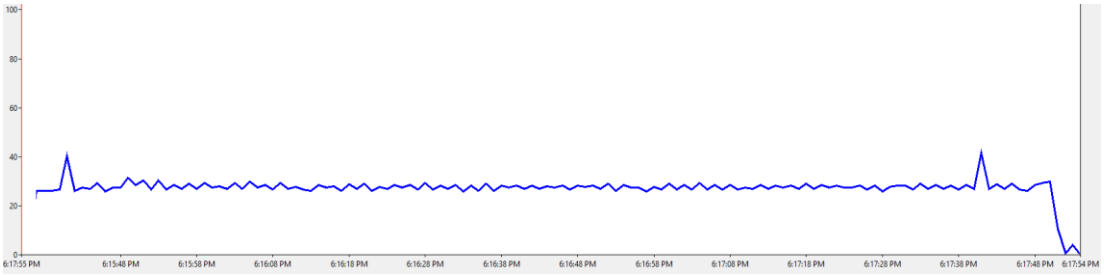
Nb. of simultaneous test calls	Delay between scripts (ms)	Nb. calls / minutes	Import / compress period (s)	Nb. files compressed	Nb. files compressed / minutes
20	9000	5.7	10	2	12
40	4500	11.4	10	2	12
60	3000	17.1	10	3	18
100	1800	28.5	10	5	30
200	900	57	5	5	60
400	450	114	5	10	120
600	300	171	5	15	180

2 Performance tests results

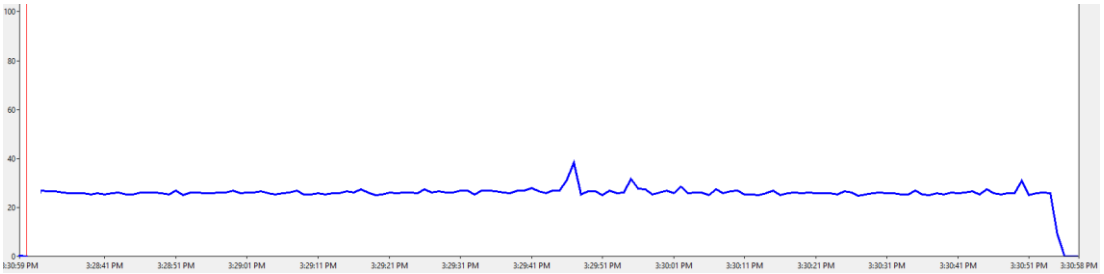
2.1 Compression tests

As the test is based on a one media instead of two per recorded call, we have first tested the increase of resources used when two media are recorded per call and merged into one mp3 file. We have noticed that most of the resources were consumed by the merge/compression module. We have monitored the execution of 200 compressions of 1 RAW file and 200 compressions and merge of 2 RAW files (which is the reality on production environment).

200 Compressions of 1 RAW file, of 3mn audio, with 4 vCPU
All compressions were over in 2 minutes and 24 seconds with 26% usage CPU average.



200 Compressions/merge of 2 RAW files, of 3mn audio each, with 4 vCPU
All compressions were over in 2 minutes and 24 seconds with 26% usage CPU average.



We made the same tests with 100 and 100 to confirm the ratios and we had the same time for each.

We can conclude that the merge + compression of two files, does not use more resources that the compression of one file.

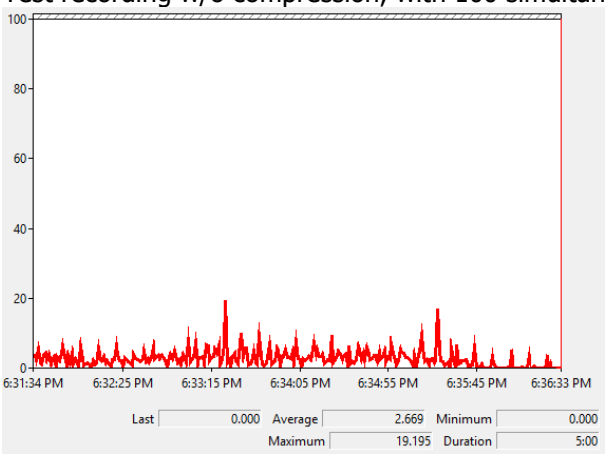
Note: The compression générates 100% CPU peaks. They are executed in very low priority. We have checked it does not alter the quality of the recording, nor the streaming, which are executed in much higher priorities.

2.2 Recording only tests

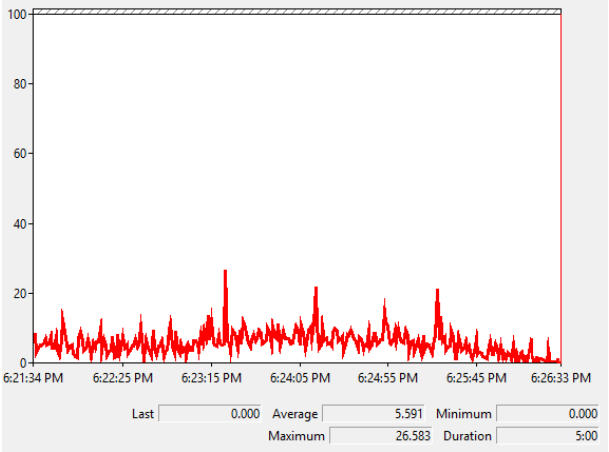
A second test has been executed to measure how much resources are used when recorded two raw files instead of one, per recorded call.

So, we executed the test of by disabling the compression, so that only the recording of the media is tested. We have kept the highest 5 minutes of the test and made a graph hereafter.

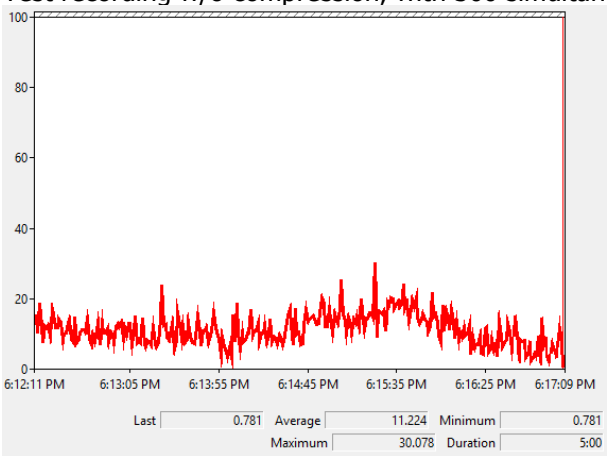
Test recording w/o compression, with 100 simultaneous recording, 4vCPU



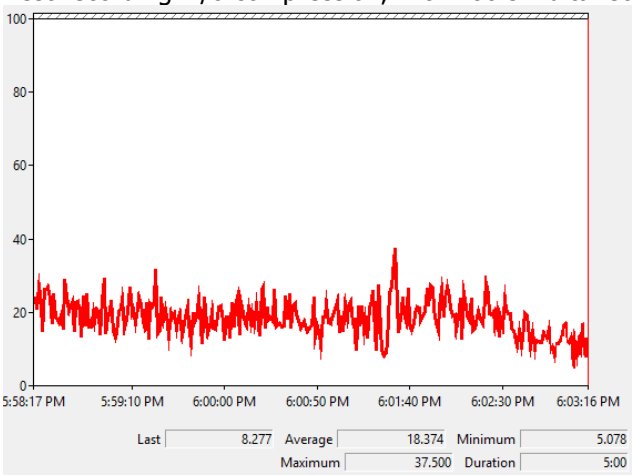
Test recording w/o compression, with 200 simultaneous recording, 4vCPU



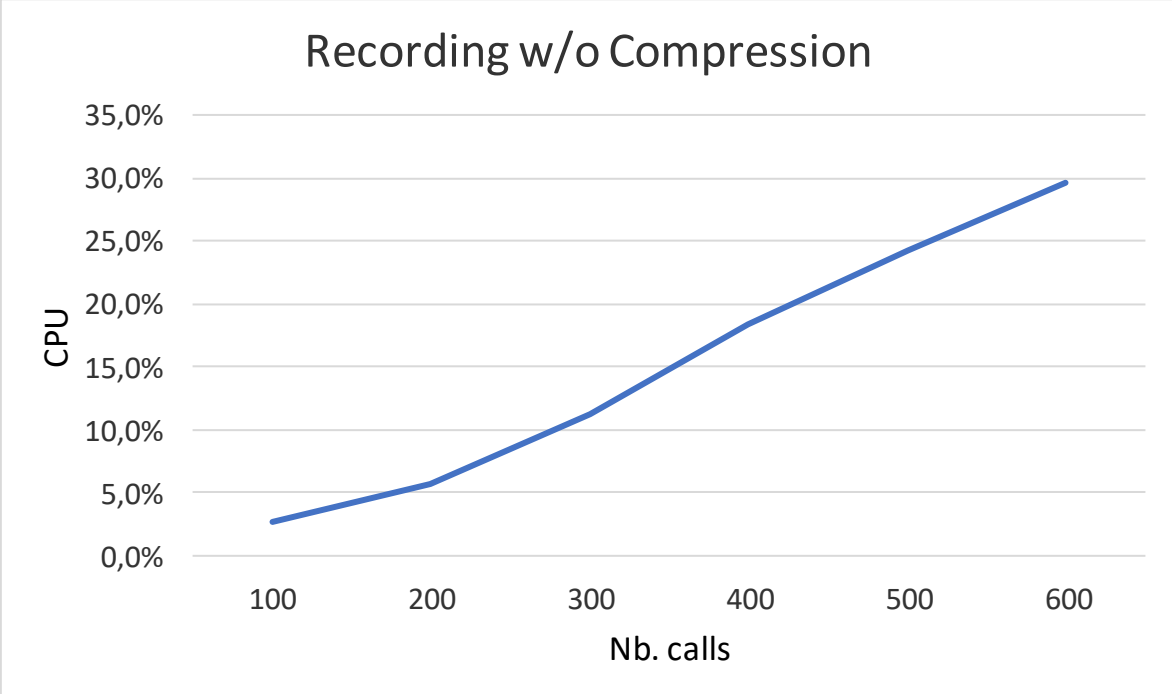
Test recording w/o compression, with 300 simultaneous recording, 4vCPU



Test recording w/o compression, with 400 simultaneous recording, 4vCPU



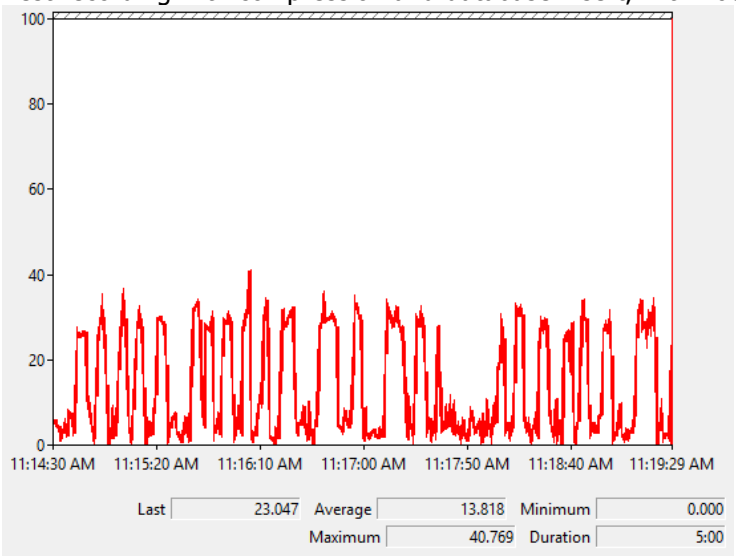
Simultaneous recording	Average CPU	Maximum CPU
100	2.7%	19.2%
200	5.6%	26.6%
300	11.2%	30.1%
400	18.4%	37.5%
500	24.2%	48.8%
600	29.6%	55.9%



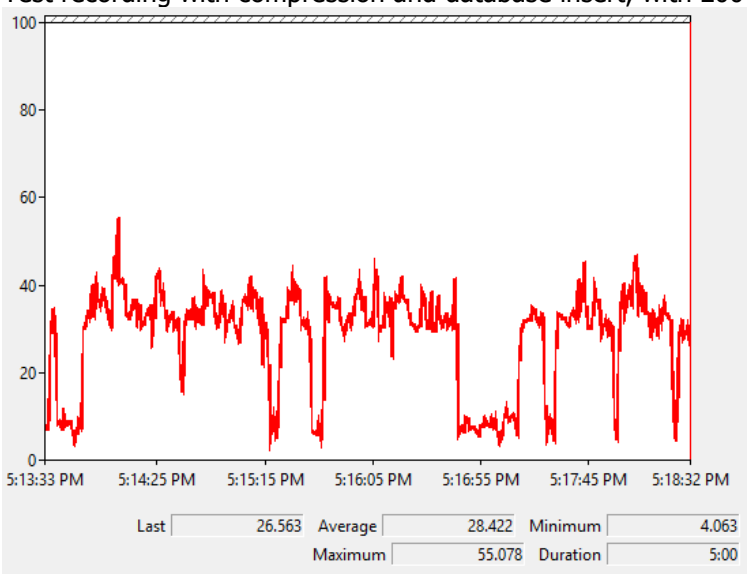
We can conclude that the CPU resource increase more than the number of recorded media.

2.3 Full features tests (recording, compression, database insert)

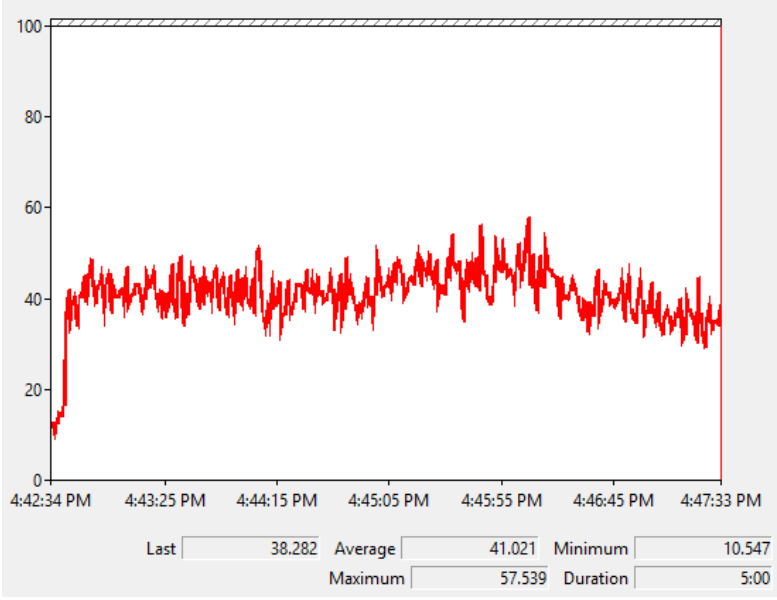
Test recording with compression and database insert, with 100 simultaneous recording, 4vCPU



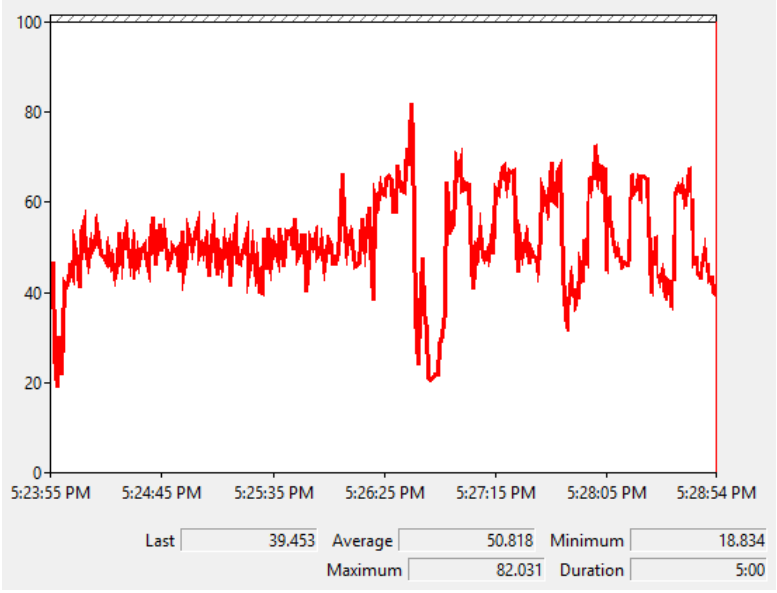
Test recording with compression and database insert, with 200 simultaneous recording, 4vCPU



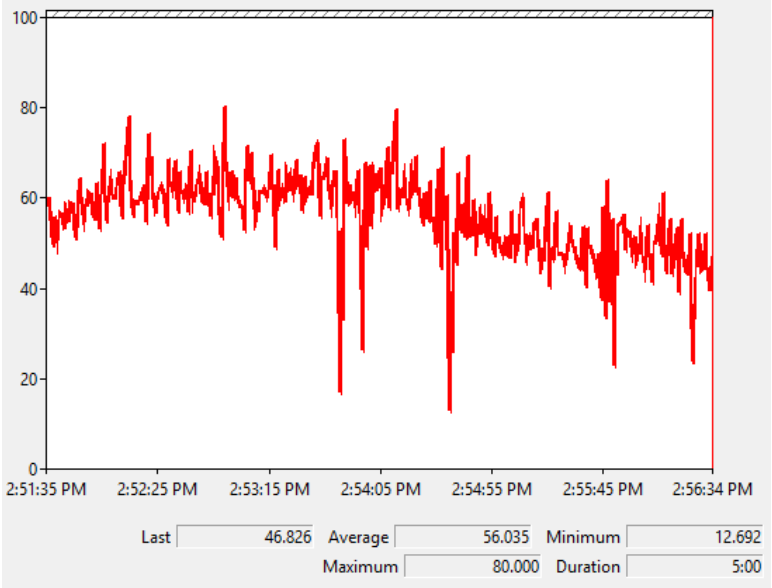
Test recording with compression and database insert, with 300 simultaneous recording, 4vCPU



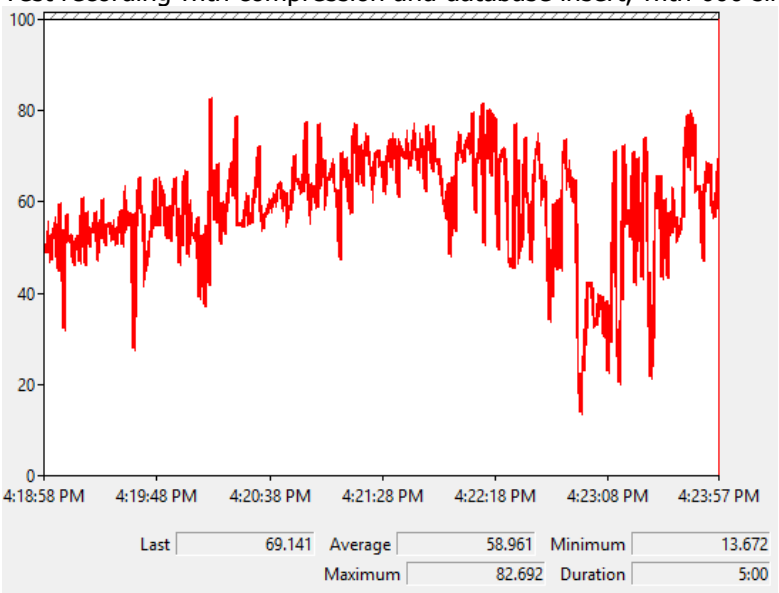
Test recording with compression and database insert, with 400 simultaneous recording, 4vCPU



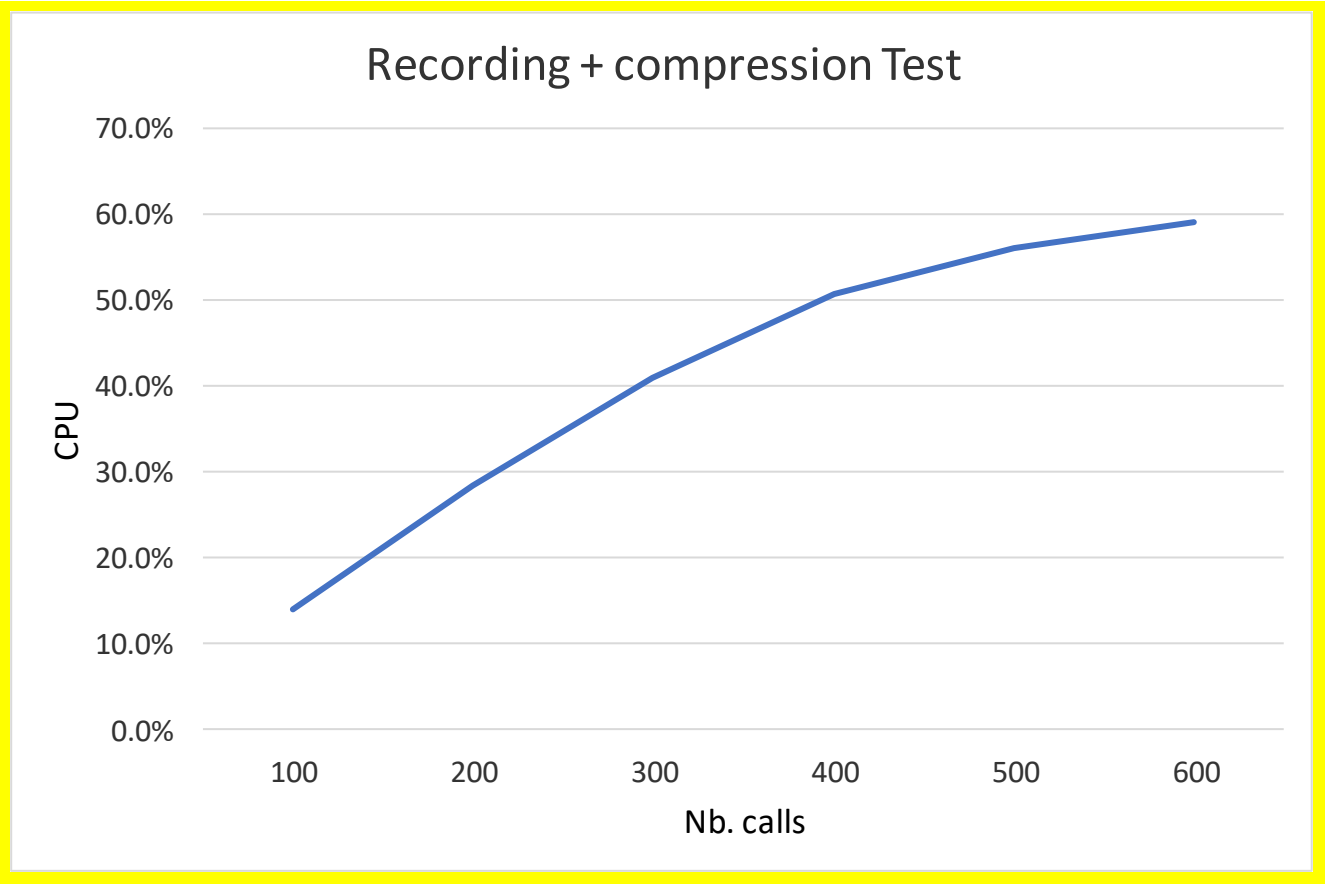
Test recording with compression and database insert, with 500 simultaneous recording, 4vCPU



Test recording with compression and database insert, with 600 simultaneous recording, 4vCPU



Simultaneous recording	Average CPU	Maximum CPU
100	13.8%	40.8
200	28.4%	45%
300	41%	55.5%
400	50.8%	64.7%
500	56%	80%
600	59%	82.7%



2.4 Comparison of tests with and w/o compression

The goal is to separate the % of CPU used by recording and by compression/database insertion. This information is useful to take into account the double of media sent which increase the CPU workload of the recording, but not of the compression or not the database insert.

Simultaneous Morning check calls	Average w/o compression	Average with compression	% of CPU used by recording vs total CPU used
100	2.7%	13.8%	19.5%
200	5.6%	28.4%	19.7%
300	11.2%	41%	27.3%
400	18.4%	50.8%	36.2%
500	24.2%	56%	43.2%
600	29.6%	59%	50.2%

For the moment, we are not confident with the 500 and 600 ratios as the tests with compression for 500 & 600 are suspicious.

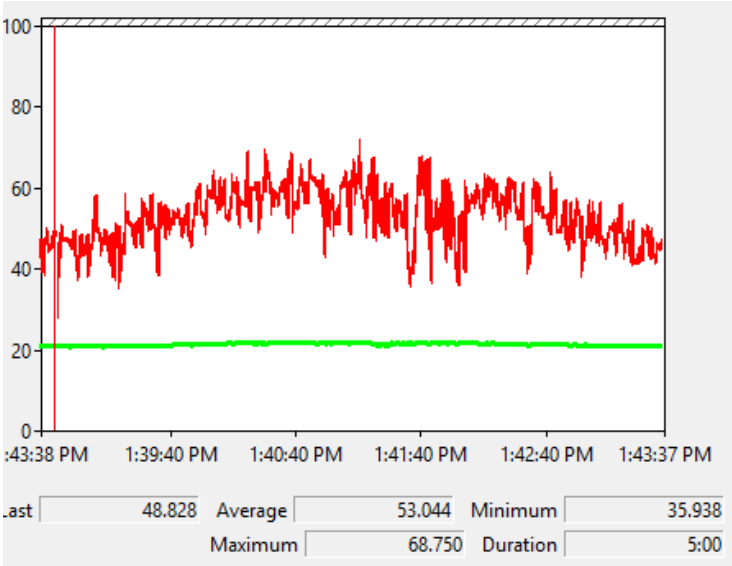
This result provides a guideline to calculate how many real calls are simulated by our tests, considering the fact that, with real recorded calls, there are two calls sent (one for each leg) to the recording Trunk SIP instead of one for the test. We have seen that the part of the CPU used for Recording will double as the part of the CPU used by the compression will remain unchanged.

Real recorded calls = Morning Check Calls / (1 + %CPU recording seul)

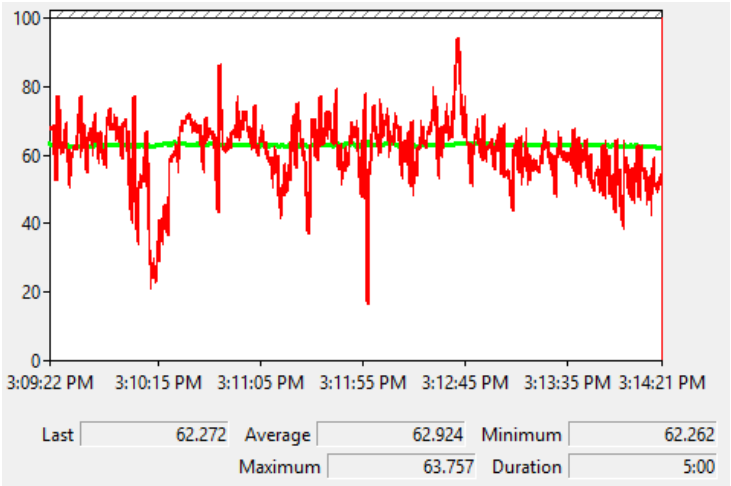
Simultaneous Morning check calls	% of CPU used by recording vs total CPU used	Real recorded calls simulation
100	19.5%	83
200	19.7%	167
300	27.3%	235
400	36.2%	293
500	43.2%	342
600	50.2%	385

2.5 Comparison of tests Recording Agent with and without Recording Manager

Test telisca Recording Agent alone (Recording + compression w/o database insert), with 600 simultaneous recording, 4vCPU



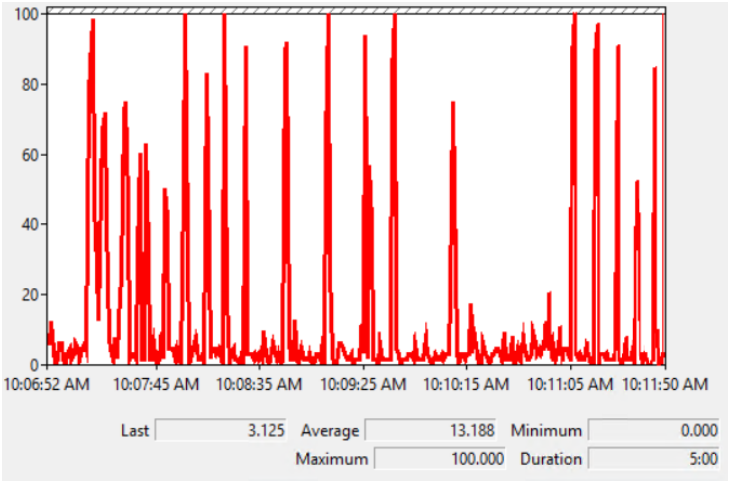
If we compare with the test with Recording Manager (which was generating database inserts) the average CPU workload is 63%.



If telisca Recording Agent is installed on a separate server the workload is reduced by 20% / 63% = 31%. Or the number of calls supported can be increased by 31%.

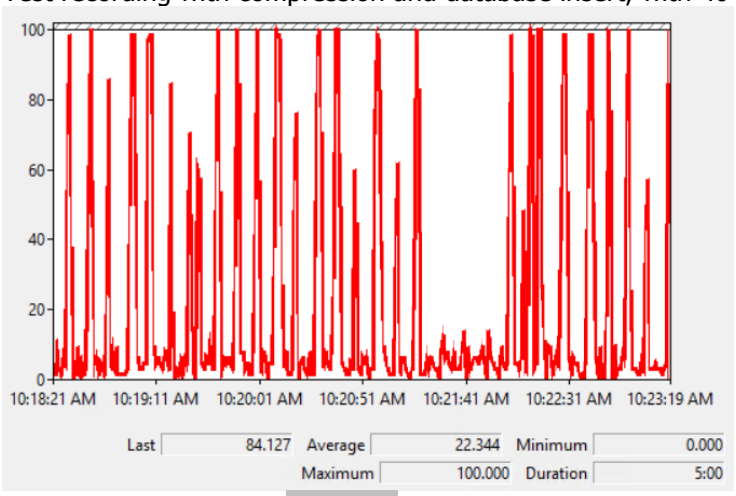
2.6 Full features tests for 1 VCPU and 4GB RAM

Test recording with compression and database insert, with 20 simultaneous recording (16 real calls)



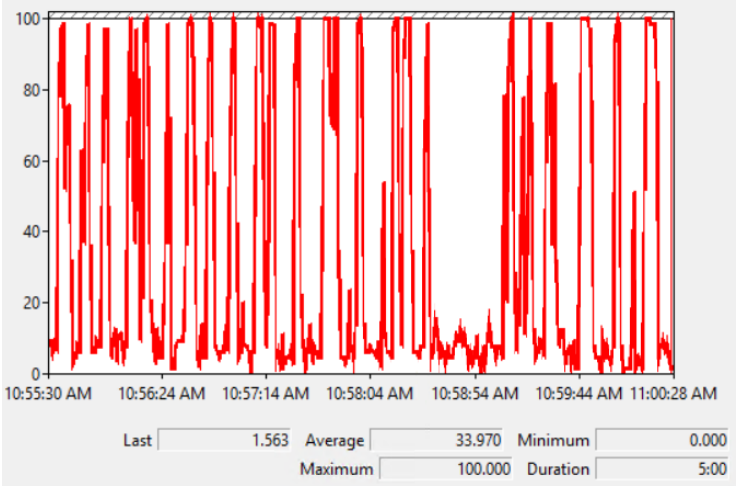
14% CPU

Test recording with compression and database insert, with 40 simultaneous recording (33 real calls)



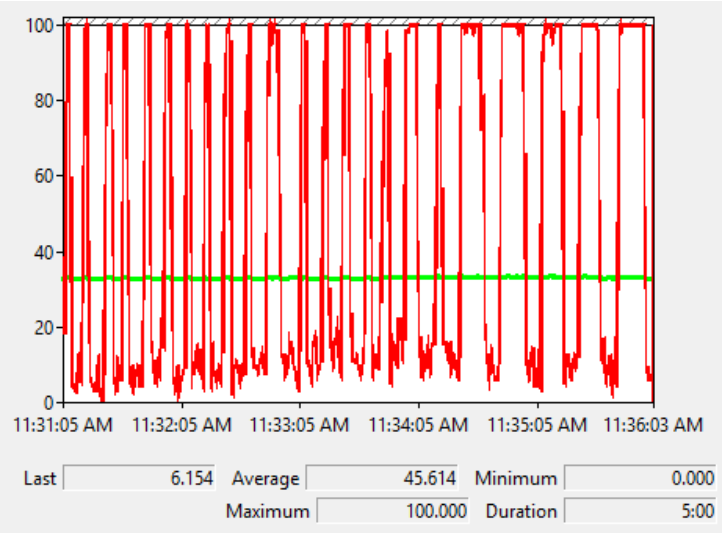
22% CPU

Test recording with compression and database insert, with 60 simultaneous recording (50 real calls)



34% CPU

Test recording with compression and database insert, with 116 simultaneous recording (100 real calls)

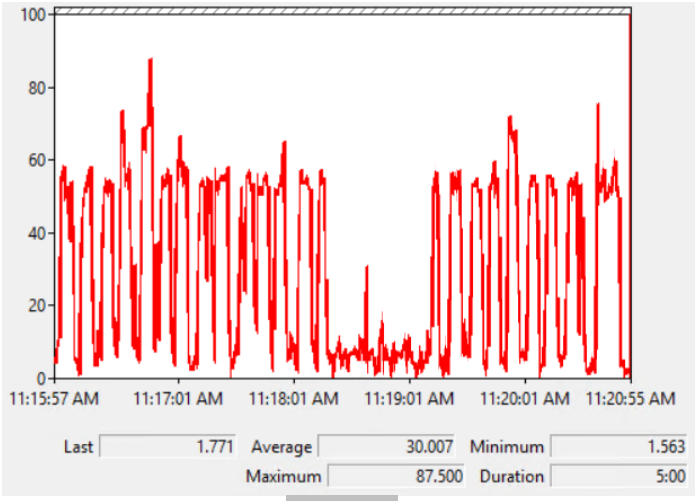


45% CPU

Note: The 100% CPU peaks are generated by MP3 compression processes. They are executed in very low priority. We have checked it does not alter the quality of the recording, nor the streaming which are in high priorities.

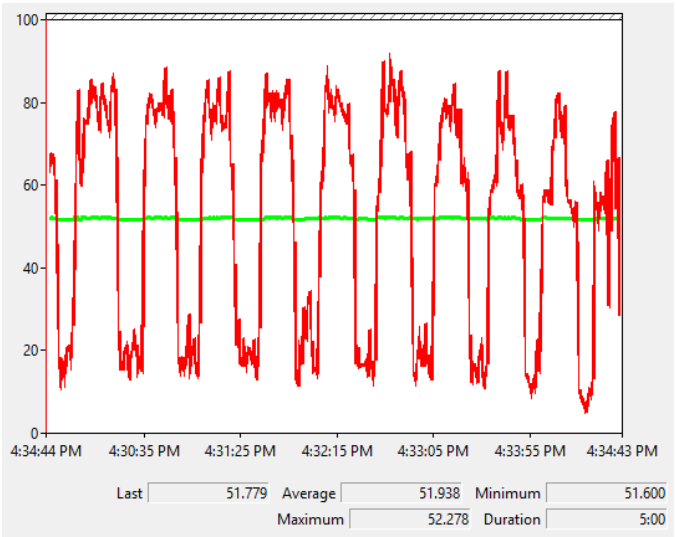
2.7 Full features tests for 2 VCPU and 4GB RAM

Test recording with compression and database insert, with 100 simultaneous recording (83 real calls).



30% CPU

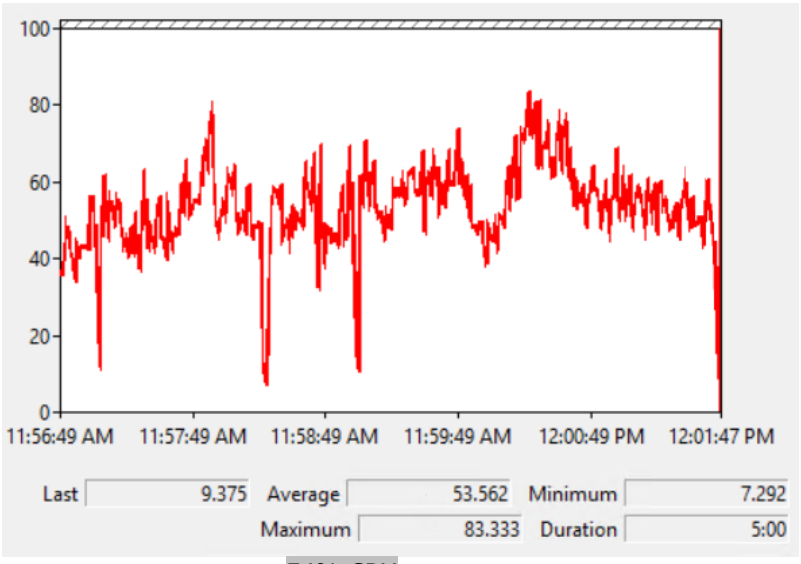
Test recording with compression and database insert, with 255 simultaneous recording (200 real calls)



52% CPU

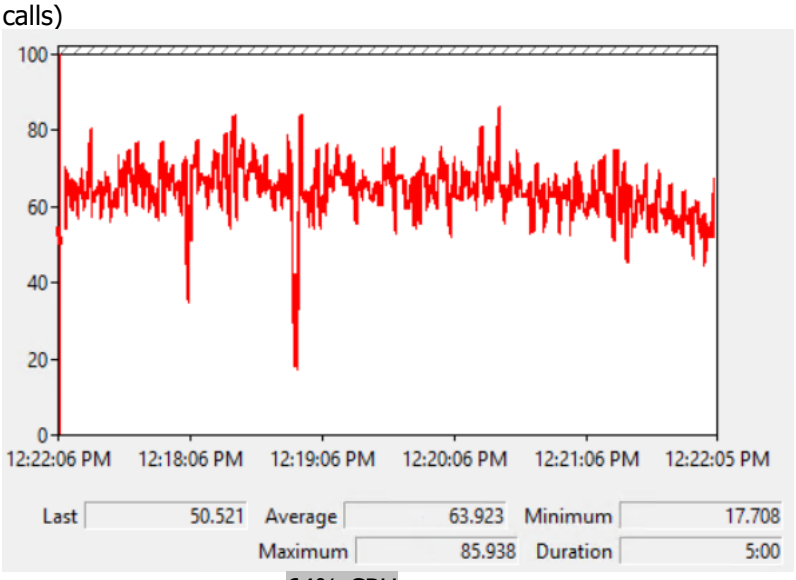
2.8 Full features tests for 3 VCPU and 4 GB RAM

Test recording with compression and database insert, with 300 simultaneous recording (235 real calls)



54% CPU

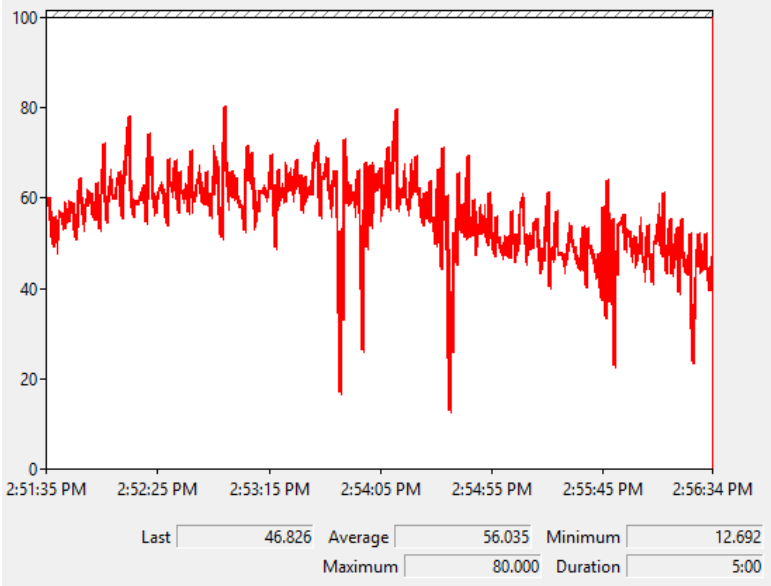
Test recording with compression and database insert, with 400 simultaneous recording (293 real calls)



64% CPU

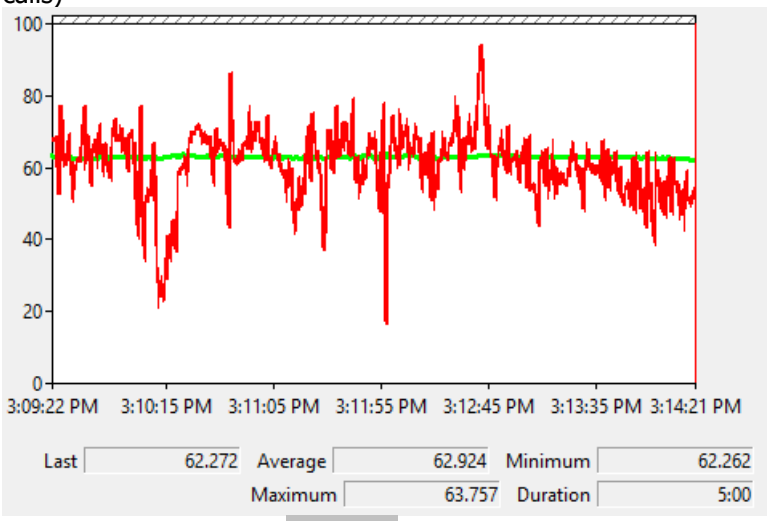
2.9 Full features tests for 4 VCPU and 4GB RAM

Test recording with compression and database insert, with 500 simultaneous recording (342 real calls)



56% CPU

Test recording with compression and database insert, with 600 simultaneous recording (385 real calls)



63% CPU

3 Sizing guidelines

We can conclude from the above tests the following guide lines to size telisca Recording Virtual Machine according to the number of simultaneous recording.

The performance tests with 1 to 4 vCPU give the following result when converting to real simultaneous calls.

We have found that 4GB of memory was enough for the Recording Agent. However, the memory sizing takes into account that, with more recorded agents, the database size will increase. It is important to have more memory available then to increase the database performances.

Note: The processor used for the test was Xeon E5-2620 v4 (2,1GHz, 8C/16T, 20MB cache memory), with Passmark 11 349.

3.1 Recording Agent + Manager on same server (contact center / trader)

Nb. vCPU	RAM	Maximum real recorded simultaneous calls
1	4 GB	100
2	6 GB	200
3	6 GB	300

3.2 Recording Agent + Manager on separate servers (contact center / trader)

By comparing the workload tests between TRA alone and TRA + TRM, we have seen that the TRA alone performance is increased by 30%.

telisca Recording Agent sizing

Nb. vCPU	RAM	Maximum real recorded simultaneous calls
1	4 GB	130
2	4 GB	250
3	4 GB	350
4	4 GB	500

telisca Recording Manager sizing

Nb. vCPU	RAM	Maximum real recorded simultaneous calls
1	4 GB	150
1	6 GB	300
2	8 GB	600

3.3 Recording Agent + Manager on same server (Administrative environment)

With a BHCA (Busy Hour Call Attempt) of 6, the workload will be one third of the Contact Center workload tested. The number of recorded lines supported will be as follow.

Nb. vCPU	RAM	Maximum number of agents
1	4 GB	300
2	6 GB	600
3	6 GB	900