VoiSentry

Speaker Verification User Guide

MAN1263 - Revision 1.19 | 2022

aculab

PROPRIETARY INFORMATION

Document Revision

Rev	Date	By	Detail
1.0d	15/03/18	df	Formatting
1.2	16/03/18	pl	Updated following QA review
1.3	24/04/18	ebj	Reformatted and diagrams added.
1.4	25/04/18	ac	Added licence usage information.
1.5	25/04/18	ac	Added Technical Support page information.
1.6	19/06/18	ac	Add info about browser support, disk import, web system log, and Verifications page. Added info for the NTP status banner change on the status page.
1.7	14/09/18	rs/ac	Update to reflect changes for thresholds and sensitivity. Remove the mention of licence renewals.
1.8	16/05/19	rs/ac/ ebj	Added a bit about identification performance. Added information about Docker image.
1.9	05/08/19	rs	Removed all references to threshold, added sensitivity. More consistent usage of the term's likelihood and confidence.
1.10	30/09/19	rs/ebj	Added a paragraph about spoof detection
1.11	25/11/19	ac	Update memory requirements
1.12	26/11/19	rs/ebj	Renamed spoof to presentation attack to be consistent with the API guide.
1.12V	19/02/20	ebj	Style changes implemented.
1.13	03/06/20	rs	Added to Identification performance.
1.14	22/02/21	ac	Add information about licensing resilience. Add missing items in gui.
1.15	26/02/21	ac	Update title page to show correct revision.
1.16	01/04/21	rs	Removed or changed text that is not relevant to VoiSentry or concerns inner workings of VoiSentry.

1.17	15/04/21	ac	Changed back page to new version which no longer references voisentry.com. Remove some dead links to vmware site.
1.18	14/07/21	ac	Add PUBLIC to footer.
1.19	13/04/22	ac	Add MAN1263.

CONTENTS

1 INTRODUCTION AND OVERVIEW

Call centres frequently need to positively verify a caller's identity. This often includes a series of 'security questions' of a kind that only the genuine caller would be expected to be able to answer. However, imposters may know enough about their targets to answer such questions, so it is useful to use measures that an imposter cannot easily replicate.

Speaker verification using voice biometric information (or 'voiceprints') has proven to be a valuable method of confirming a caller's identity, removing much of the dependency upon detailed security questions. Other benefits include a reduction in call duration, reduced loading on the call centre agent, and an improved caller experience.

Aculab VoiSentry offers a speaker verification service in a highly scalable, resilient, multi-tenanted, virtualised software package.

NOTE

VoiSentry is also available as a Docker image, allowing deployment in environments where creating a VM is not possible. The rest of this document will mostly refer to only the VM, but the information should hold true for the Docker image.

VM Only will show where functionality is not available in the Docker image.

Speaker verification needs to be implemented carefully if it is to deliver the full benefit it offers, so this User Documentation describes not only the usual installation, administration and API documentation, but also how the technology works and guidance on 'best practice'.

1.1 Call Centre Environment

Call centres typically depend upon a data processing environment deploying well thought-out strategies in respect of equipment practice and virtualisation of component services.

In order to integrate into such an environment, the Aculab VoiSentry Speaker Verification system is provided as a downloadable Virtual Appliance software package suitable for deployment into either a VMware vSphere, or a Microsoft Hyper-V virtualisation environment.

The term 'Virtual Appliance' means that, despite being a software download, the product is essentially an all-in-one software package in the manner of a 'black box' - there is no need to install or tend to an operating system, or install individual separate software packages. The user simply deploys the VM onto the hypervisor (vSphere or Hyper-V), and only interacts with the service via a web-based administration User Interface (UI) and various Web Services (REST) APIs.

The prime reason for delivery of VoiSentry as a VM is to allow easy deployment into existing equipment. However, any high-volume Speaker Verification system will be CPU-intensive, and in most cases this will mandate that VoiSentry is the only VM running on each hardware server.

1.2 Call Centre Integration

VoiSentry is intended to be integrated with, and used in conjunction with, a call centre that already provides a level of automation in the form of a front-end Interactive Voice Response (IVR) system that is able to capture caller voice recordings of passphrases for verification purposes, and send them as standard voice (.wav) files to VoiSentry for verification against an identity claim.

VoiSentry is intended to be integrated into the call centre's automation and call-flow at least in part via that IVR system.

1.2.1 Data Retention

VoiSentry does not itself provide any voice recording capabilities, nor does it store voice recordings that are sent to it. It only stores derived biometric (voiceprint) metadata used for verification purposes, which is accessed against a unique identifier (or key) provided by the controlling application. The voiceprint data itself is not readily retrievable, and in any event, it is not possible to use the voiceprint to reconstruct the original voice recording.

1.2.2 Data Reporting

By policy, VoiSentry only retains enrolment and verification transaction counts for the purpose of accounting against licensed limits and assigned quotas. It does not provide reports of successful or unsuccessful verification attempts, as those outcomes are readily accumulated by the call centre (via the application or service that is using VoiSentry).

2 VOISENTRY ARCHITECTURE

The VoiSentry product is supplied as a Virtual Appliance to be deployed as a VM onto a hardware platform with an installed virtualisation hypervisor (VMware vSphere or Microsoft Hyper-V).

2.1 System Scalability and Clustering

A single VoiSentry VM is a complete stand-alone speaker verification system, but there may be circumstances where it is desirable to 'spread the load' over more than one system (or 'Node'), in which case, additional Nodes may be attached to the initial server to form a 'Cluster'.

When clustered, each individual Node contains an identical up-to-date replica of the entire VoiSentry database, and if it suffers an unexpected partitioning from the rest of the cluster, it will continue providing verification services as before the partitioning.

Reasons for clustering several servers may be:

- Scalability: where average or peak demand exceeds the service capacity of a single Node, or response times at peak usage become unacceptable.
- Resilience: where there is a requirement to provide redundancy for 'fail-over' protection.
- Flexibility in maintenance: where individual servers or VMs are being maintained or upgraded, additional Nodes may be added to the cluster to stand in for the Node being maintained.

NOTE

When an installation consists of several nodes in a cluster, it is normal to distribute Web Service API requests across the cluster using a load-balancing reverse proxy in an identical manner to running a traditional distributed web platform.

Alternatively, applications that request service from VoiSentry may poll any known Node for a list of all the Nodes in the cluster, and distribute requests across the cluster by themselves.

In terms of the administrative User Interface (UI), the entire cluster may be monitored and managed from an administrator login at a single Node.

2.2 Data Scalability

The VoiSentry VM has a self-contained datastore volume within the virtual appliance that is suitable for initial evaluation and testing. For real-world applications, it is usual that a larger data volume is required, in which case an additional virtual disk can be administratively assigned to the VM (via the hypervisor), and the data migrated to that larger external volume VM Only.

If an existing external data volume is subsequently determined to be too small, the database can readily be moved onto even larger virtual disks by attaching and migrating onto new Nodes with a larger datastore volume.

3 VOISENTRY SERVICES

In order to implement an advanced speaker verification subsystem as an adjunct to a call centre, VoiSentry provides the following services:

- An HTML based administration User Interface (UI) that (for most operations) allows administrative control and status visibility over the entire cluster.
- A web-services based administration REST interface (referred to here as WS), so that VoiSentry may be remotely administered from third party data equipment.
- A web-services REST based Application Programming Interface (API) to provide Speaker Verification and Identification services:
  - Caller enrolment with the provision of one or (ideally) more initial recordings of a passphrase.
  - Enrolment updates to allow continuous improvement of an existing enrolment by the provision of further (validated) recordings of the original caller.
  - Verification (or otherwise) of an identity claim against an existing enrolment using a new recording of the passphrase.
  - Identification of a speaker from a new recording.
  - Automatic Speech Recognition (ASR) of digits, ‘yes’, and ‘no’ to assist in the recognition of pin-numbers or account codes.
  - Recognition of DTMF keypad input in case that is used in place of spoken digits.

4 VOISENTRY LICENCES

VoiSentry is licensed in respect of the number of verifications or identifications per day. Licences are obtained in the first instance from the Aculab Web-based licence management website, and may be deployed onto one or more ‘Licence Server’ that is available to the cluster.

In order to provide flexibility and choice in respect of scalability and resilience strategies, there are several options for deploying licences to the benefit of a particular cluster:

- Each node in a cluster contains an embedded licence server, so any licences installed on any given node will be available to the entire cluster.
- Resilient licensing – the licences are not tied to the node in a cluster they are installed on. Each node will have knowledge of all licences in a cluster, such that if a node is restarted, or shutdown, the number of licenced verifications/identifications will remain the same. At the point there is a single node cluster, you will not be able to decluster until all licences have been removed. This is to ensure you do not accidently decluster and lose you licences.
- Licence Server software is available as a software download that may be installed onto one or more Linux or Windows hosts, which may either be ‘bare-metal’ or virtualised systems. Such ‘external licence servers’ may be made known to a cluster to make any licences installed there available to the entire cluster.

NOTE

Note that each licence server and the licences it contains may only be utilised by a single cluster.

5 VOISENTRY USER HIERARCHY AND MULTI-TENANTING

VoiSentry has two separate account types (administrator and tenant), and each type has two trust levels.

5.1 Administrator Account Types

The ‘senior’ administrator account type permanently has a single member (‘superuser’) that is present with a default password (‘password’) when the system is first started.

The superuser may be used to create or modify either ‘junior’ administrator account types (‘admin’) or ‘tenant’ accounts.

The ‘admin’ accounts have all of the authority of superuser except the ability to create or modify admin accounts.

The responsibility of superuser and admin accounts is limited to platform administration.

When a tenant account is created (or subsequently), it may be assigned quotas in respect of the maximum number of enrolments it may employ in its datasets, and the maximum number of verifications per day it may use across its datasets.

5.2 Tenant Account Types

Each tenant is able to create one or more datasets (against which callers are enrolled and verified) and one or more access keys (which are used by external application programs to access a particular dataset) against each dataset. Each access key may be assigned quotas in terms of the maximum number of enrolments and the maximum number of verifications per day.

A tenant may also create or modify ‘user’ account types, each of which have all of the authority of the tenant account that created it, except the ability to create or modify further user accounts.

A particular tenant has no visibility of other tenant accounts or their assets (datasets or access keys).

6 VOISENTRY INDICATIVE PERFORMANCE

In provisioning VM resources, it should be noted that the VoiSentry Virtual Appliance is shipped configured for 4 (virtual) CPU cores and 6GB RAM, and the VM's internal datastore volume will hold in the region of 20,000 enrolments.

This configuration is suitable for initial testing and evaluation purposes, but an absolute minimum configuration for production deployment or performance testing should be 8 CPU cores and 16GB RAM. This can be administratively changed via the hypervisor whenever the VM is stopped. Potentially, a larger external datastore volume may be suitable for the number of enrolments to be used for testing (for guidance, one enrolment takes up around 140kB of disk space, and to provide for database management, you should target using less than 70% of the external datastore volume for enrolment data).

With regard to CPUs, there is an identifiable trend in server CPU evolution towards providing more CPU cores, but at a lower CPU clock rate, so one CPU's cores may not be comparable in performance to another. Also, up to a point, any increase in CPU cores to a system should be accompanied by an increase in RAM.

Another performance-critical element will be the speed of the disk subsystem, and the use of SSDs (both for the VM and for any external datastore volume) is highly recommended. Simplest to describe and evaluate is a local SSD (i.e. mounted in the underlying server that constitutes the Node), but data centre practice may mandate other arrangements, in which case the access speed of any extended datastore subsystem will be particularly critical.

As has been suggested here, actual throughput performance can depend upon a number of different factors, and the only way of determining that a particular configuration is suitable for a given task is by proper pre-deployment testing.

6.1 Verification Throughput

For guidance, a VoiSentry cluster consisting of a single Node running as a sole guest under VMware ESXi 6.0 on an Intel i7 at 3.2GHz, and assigned all 4 cores (8 hyperthreaded cores) and 16GB RAM with a local SSD has been shown to achieve in excess of 40,000 successful verifications per hour, with an average latency of less than 0.5 seconds. This was achieved against a test dataset using the passphrase "It is obvious who I am, even on the telephone". Adding further nodes to a cluster can be expected to scale overall throughput approximately linearly.

6.2 Identification Throughput

The identification service will compare a given signal to a number of speaker models. The latency for the first comparison is almost the same as for a single verification. Subsequent comparisons are much faster. To identify one of fifteen speakers will take four times as long as a single verification; one of thirty will take seven times as long; and one of sixty, thirteen times.

6.3 Verification Accuracy

The same system described above has been shown to have an Equal Error Rate (EER) of 2% (the proportion of false acceptances and false rejections being 2%) against a test dataset.

This accuracy will not be affected by the same considerations as throughput performance, but may be affected by the following factors:

- Length and phonetic complexity of the passphrase, and whether a specific passphrase is used rather than an arbitrary segment of speech
- Caller phone type and call quality
- Caller noise environment

Please read the section “VoiSentry Operational Considerations” that refers to “Verification Accuracy” for a more complete consideration of verification performance.

6.4 Resistance to Presentation Attacks (Spoofing)

VoiSentry offers some protection against three types of presentation attacks. These are labeled TypeA, TypeB and TypeC in the API guide (please see the API guide for a description of the three types). Although VoiSentry attempts to detect these three types, some attacks can be extremely sophisticated and it is inadvisable to rely on VoiSentry’s detection alone for protection against them.

Using VoiSentry as part of a multi-factor authentication system is recommended (this is expanded on later in this document). For instance, multi-factor authentication can help to establish that the speaker is live and not a recording, thus eliminating one type of presentation attack.

6.5 Identification Accuracy

The accuracy of an identification task depends very much on the number of speakers to choose from. If there is only one choice, the accuracy is the same as for verification. As more speakers are added to the list, the accuracy will drop. The drop cannot be absolutely predicted as it will depend on a large number of factors, including the gender and age distribution of the speakers to choose from, and whether the identification task is text dependent or text independent.

Identification will work best in a task where each speaker is trained on a personal pass-phrase, and the audio presented for identification is a speaker’s pass-phrase. This is a text dependent scenario – the audio to identify contains the same utterances as the audio used to train the corresponding speaker’s model.

In text independent mode, the text of the supplied test audio could be very different to the text of the audio used when training the corresponding speaker’s model. When this happens, identification can be less reliable; especially if another speaker’s model is based on text that is more similar to the test audio than the target speaker’s.

When choosing to do text independent identification, it is recommended that the speaker models are each trained on sufficient speech to cover the all sounds that might be present in the test speech

Given a typical distribution of speakers (half male and female) the following results can be expected from a text dependent application:

The correct speaker has the highest confidence

Number of speakers	1	5	15	30	60
Percent	100	100	99	99	96

7 MINIMUM VM SYSTEM REQUIREMENTS FOR DEPLOYMENT

- 8 (virtual) CPU cores
- 16 GB RAM for the VM (perhaps 20GB total for the virtualisation server)
- 200MB of external datastore volume (Virtual hard disk) per 1000 enrolments
- Hypervisor, one of:
  - VMware ESXi 6.0 or ESXi 6.5
  - Windows Server 2012 or 2016 running Hyper-V

As has been suggested above, actual throughput performance can depend upon a number of different factors, and the only way of determining that a particular configuration is suitable for a given task is by proper pre-deployment evaluation.

8 SUPPORT FOR IPV6

Each VoiSentry Node is required to be supplied with an IPv4 static address, it cannot function without it.

It may also be furnished with an IPv6 static address by which individual Nodes may be accessed, and IPv6 addresses that are automatically assigned (and accessible via the hypervisor) may also be used. This includes access to the administrative web UI, the administrative web services (/ws) and the REST API (/api).

9 AUTOMATIC SPEECH RECOGNITION

VoiSentry provides an automatic speech recognition (ASR) service. ASR is the identification of the words spoken, irrespective of who spoke them. The ASR provided by VoiSentry is fixed vocabulary, grammar-based (often referred to as connected word recognition, or CWR).

VoiSentry offers a CWR system for digit recognition and yes/no recognition in several languages. It assumes that the spoken words will always fit a known grammar (a finite set of rules defining allowed sequences of words, set by the user). CWR systems are easy to use, computationally efficient, and quick to respond. They are best suited to selecting items from a menu or for inputting a short sequence of digits. The interpretation of the resulting text string is quick and easy.

For example, a grammar for recognising a PIN might allow any combination of four digits. The CWR system will then always return some combination of four digits, regardless of what the speaker actually said, and it is up to the application software to decide whether those digits are feasible or not.

10 SPEAKER VERIFICATION AND IDENTIFICATION

10.1 What are Speaker Verification and Speaker Identification?

The terms speaker verification and speaker identification are often used interchangeably, but they serve different purposes. Other terms are also used when referring to speaker verification or identification, such as speaker recognition, or voice recognition.

VoiSentry provides services for speaker verification and speaker identification. To help understand the difference between the two, a definition for each is given below.

Speaker Identification (*SI*) is the process of identifying who is speaking. One speaker is selected from a population of known speakers. There is no prior claim about the speaker's identity - it is left up to the system to decide. The system analyses a sample of speech and returns its best estimate of the speaker's identity, usually with a score to indicate how confident it was in making the decision.

In practice, the accuracy and reliability of SI systems is highly dependent on the size and diversity of the population of known speakers.

Speaker Verification (*SV*) is the process of verifying a speaker's identity claim. The speaker will claim to be a specific person enrolled on the system, and the system will use a sample of the speaker's speech to accept or deny that claim. Speaker verification always analyses a single speaker model and, so, is much faster than identification. The number of speakers enrolled on the system does not affect the speed or accuracy of verification.

There are two main classes of speaker verification: text-*dependent* and text-*independent*:

***Text-dependent* SV** requires the user to say a specific phrase, or a specific sequence of words in order to be verified. The same phrase or word sequence will have been used when creating the speaker model. The phrase used for text-dependent verification must never change and the whole phrase must be said during a verification attempt. The phrase does not have to be the same for each speaker.

Text-dependent systems have one main advantage: they can be more demanding about exactly what sound characteristics are acceptable at each point in time. This means that they can provide verification results with a high degree of confidence.

***Text-independent* SV** will operate regardless of which words are spoken. The advantage of text independent verification is that it can be used at any point, or several times, during an extended dialogue.

10.2 What can Speaker Verification do?

Speaker Verification is primarily used to restrict access to confidential information and secure computer systems. It normally forms part of a larger access control system, which combines information from different sources to separate out genuine identity claims from those of impostors. This is termed Multi-Factor Authentication, or MFA.

There are many ways to restrict access to secure systems, but very few which allow automatic verification of an individual's identity over the telephone. In that scenario, the information which is available can be summarised as:

- the voice characteristics of the individual
- the sequence of words they speak
- any telephone keys they may press
- any metadata related to the call

The metadata, along with a combination of Speaker Verification, Automatic Speech Recognition, and key-press (DTMF) provides complete coverage of all these factors.

For example, the speaker would initially enter an account or user number (either using the telephone keypad, or by speaking the numbers). This forms their identity claim. They would then speak a passphrase and the vocal characteristics of the speaker are inspected by the SV system; or, the speaker might be asked to say certain digits from their PIN, and ASR is then used to verify that the correct digits were spoken, while SV inspects the vocal characteristics. The results of the analysis are then used confirm or deny the identity claim.

Beyond this example, significantly improved security can be achieved by continuing the verification process throughout the rest of the call. In that way, not only is it possible to improve the system's confidence in its initial decision about the speaker's identity, but it can also detect any changes in the voice of the speaker. This can be a very valuable method for detecting a replay attack or, in extreme cases, coercion of a speaker.

A Multi-Factor Authentication system will include checks on many "factors" - an account number, PIN, cryptographic response code, a caller's CLI, behavioural statistics, etc. - as well as speaker verification. If any of these factors casts doubt on the speaker's identity then there is generally a set of fallback checks, designed to clear up any level of doubt.

Fallback checks may take more time and require interaction with a live operator, so to minimise costs it is important that the MFA system can make use of not just the results, but also any confidence measures associated with them. Additional checks should only be invoked when there is a high level of confidence that the initial decision was wrong.

If all the component parts of an access control system are well integrated, the use of speaker verification can simultaneously reduce costs and increase security.

10.3 How does Speaker Verification work?

10.3.1 Enrolment

Enrolment involves "teaching" the Speaker Verification system about the range of sounds which are characteristic of that speaker. This is done by providing the system with recordings of the speaker's voice. The system analyses these recordings and builds a Speaker Model so that it can estimate a confidence that subsequent sounds were produced by the same person.

The accuracy of the SV system is primarily dependent on the data used for enrolment. This should be sufficient in terms of coverage of speech sounds, or 'phonemes', to match the range of sounds expected to be encountered during verification.

10.3.2 Verification

In practice, the robustness of the verification result can be affected by noise and channel distortion. The verification confidence is formed by estimating a number of similarities between the audio sample under test and the Speaker Model, if the signal is corrupted with noise or distortion, then the similarities will drop. Although VoiSentry is robust against noise and other distortions, clean signals will give the best results.

10.3.3 Adaptation

There is one further process required of a Speaker Verification system: adaptation of the speaker models. The end-user is not normally aware of this, but it is essential to improving and, ultimately, maintaining verification accuracy.

Adaptation is required to counter the effects of insufficient or atypical enrolment data. Immediately after the initial enrolment, there is usually very little data available to the SV system to determine the acoustic variations to be expected in routine use. Often that data is also measurably different from the subsequent data used for verification.

A speaker's voice will vary very little when prompted to speak several times in quick succession (as is commonly done during enrolment). However, their voice will be noticeably different when they call from a different location or when they have been using the system for verification for many months, and have become blasé with the whole process. This can give rise to false rejections of the genuine caller.

This situation can easily be improved however. The more the system is used for verification, the more examples of the speaker's voice parameters become available, and these can be used for re-training or adaptation of the Speaker Model. This will improve the model's coverage and make it more specific to the individual voice. Obviously, it is essential that the models are only updated once the speaker's identity has been thoroughly checked and confirmed, to avoid "hijacking" of the speaker's model by an impostor.

It is also desirable to continue adaptation well beyond the initial usage period. Just as the acoustic characteristics of any individual's speech are not strictly unique, they also vary over time. As a person ages, their voice will change. Even over a shorter time-frame, they may develop new vocal habits and mannerisms. Allowing the speaker's model to track changes in their voice, through adaptation, will ensure that the system remains accurate.

10.4 Best Practice

There are some general principles which will affect any speaker verification (SV) system's performance. The most important part of the process is enrolment.

10.4.1 Enrolment

For an SV system based on a fixed passphrase, enrolment will usually involve analysing several repetitions of that passphrase. For a text-independent system the enrolment audio needs to include most of the words that the speaker is likely to say. This normally involves analysing several complete sentences.

As a general rule, any recording used for enrolment should contain at least 4 syllables, preferably more. Recordings with more syllables will generally produce more precise models, and thus better verification accuracy. A pass-phrase such as, "my voice will let me in", is six syllables.

In VoiSentry the enrolment process consists of two stages. Stage one is a call to the enrol service which will take a number of recordings and create a speaker model. We recommend three individual recordings, each containing the whole of the speaker's pass-phrase (if text dependent) or a word rich sentence. Stage two is a verification-update loop. At least two calls to the verify service should be made, each with a new recording, and each followed by a call to update.

Stage two is essential. At least two verifications should be done. The verifications must be done using new recordings. Each verification must be followed by an update using the recording that was used for verification. After each verification, take note of the confidence score. If the speaker is failing to be verified, continue to perform the

verification-update sequence, each time with a new recording. Once the speaker consistently passes verification, the enrolment process can be considered complete.

10.4.2 Verification

Verification does not typically require as much data as enrolment, but, as with enrolment, the longer and more phonetically varied the speech, the more accurate will be the verification result. If a speaker is using a pass-phrase in a text dependent environment, the pass-phrase should be at least four syllables, preferably more.

10.4.3 Adaptation

Once a speaker model is fully enrolled, it may become necessary, in time, to adapt the speaker model to improve accuracy. The update service can be used to adapt a speaker's model using the recording used for a recent verification attempt. However, this should only be done once the caller's identity has been confirmed by security questions or other independent methods, even if the speaker passed verification.

In general, adaptation should be employed if a speaker is found to consistently fail verification, indicating that there has either been insufficient training or some systematic change in the speaker's voice or the acoustic environment. Adaptation need not be performed after good verification scores, or on the basis of one or two failed attempts. More than one call to the update service may be required, each call must be with a different recording. A particular recording must not be used more than once to update a speaker model.

10.4.4 Dialogue Design

The key to a successful SV deployment lies in the details of the system design, and, most importantly, in the handling of the dialogue.

There are some decisions which need to be made at the outset. Firstly, the reason for the use of SV must be understood. Typically, there are only a few common reasons, but they need to be prioritised for each deployment:

- To reduce the time taken to complete a transaction.
- To reduce the need for direct interaction between the caller and any human operator.
- To make the security process simpler and more natural for the caller.
- To make the caller more confident that every precaution has been taken to secure their data.
- To reduce the success rate of impostors trying to access the system.

These priorities will determine trade-offs between speed of enrolment, ease of access for genuine callers, reliable identification of impostors, and perceived security.

10.4.5 Security Scenarios

If usability and minimal disruption to the caller are the most important aims, then Speaker Verification can still provide high security, provided best practice for enrolment is followed. As described earlier: three recorded utterances should be used to create the initial model. This should be followed with at least two verification attempts, each using a new recording, and each followed by an update. The verification plus update loop should be continued until the speaker consistently passes verification.

If correct identification of impostors is deemed critically important, it is not enough to base the verification decision on a single, short utterance such as a pass-phrase. Instead, the speech must be analysed continuously throughout the call. The best way to achieve this, without placing an unacceptable burden on the caller, is to use a hybrid text-dependent and text-independent approach and to carefully tune the structure of the dialogue.

For the hybrid system, the speaker will require two models: a text dependent model for verification of the speaker's pass-phrase, and a text independent model for verification at points during the call. The models will require their own enrolment phase that follows the best practice described above. The text dependent model will be based on repetitions of the pass phrase, the text independent model will be based on general speech recordings.

For example, a verification call could begin with the entry of an account number using the telephone keypad - this is the identity claim - followed by a simple text-dependent Speaker Verification transaction where the caller is asked to say their pass-phrase. Then, during the call with an operator, all the speech is collected and used for text-independent SV at regular intervals. Additionally, and especially in the case of a spoken PIN or a numerical passphrase, the spoken words can be identified by a CWR system. The combination of all these factors should be sufficient to reject the vast majority of impostors without unduly delaying access to a legitimate caller.

The dialogue of the call should be arranged such that any particularly critical transactions are delayed until sufficient examples of the caller's speech have been verified. If, by that point, the caller has been flagged as a possible impostor, the system would prompt them to answer additional security questions.

By designing the call dialogue carefully, a significant amount of speech can be collected without the caller becoming impatient or frustrated, and this can improve the impostor detection rate.

11 VOISENTRY INSTALLATION: QUICKSTART GUIDE

For initial installation, it will be necessary to have a working DHCP server on the local area network, at least temporarily.

11.1 Installation on a 'Bare Metal' Server

There is no option to install VoiSentry directly onto a 'bare metal' server. If this is the preference, it is practically equivalent (for testing at least) to installing the free (for evaluation) VMware ESXi hypervisor onto the server (a very straight-forward process), and deploying VoiSentry into that.

The free version of ESXi allows use of unlimited virtual cores for the first 60 day evaluation period, after which the hypervisor must be licensed. If you license it with the free licence key, it then only supports one physical CPU, and 8-way virtual SMP from that point onwards. Please refer to VMware documentation for the terms and conditions of use.

11.2 Installation on VMware ESXi

VoiSentry can be deployed either within a vCenter managed data centre, or onto standalone ESXi hosts.

NOTE

Important: Please read the section of "VoiSentry Operational Considerations" that refers to "VMware client selection and operation".

ESXi is compatible with a wide range of bare metal servers, but if support from VMware is anticipated, it may be advantageous to use a server that is on VMware's "supported products" list.

- Install ESXi onto the server, and set a static IP address for the ESXi host interface.
  - Preferably download and install the vSphere Client for Windows. Alternatively, open the VMware Host Client.
  - Via the ESXi client, ensure that the NTP service has a list of suitable NTP servers configured (the NTP project provides "0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org, 3.pool.ntp.org"), and that the NTP service is running.
  - Place the VoiSentry Virtual Appliance (.ova file) in a known location on your local PC.
  - Via the ESXi host client, and select "Host, Create/Register VM".
    - In the window that appears, select "Deploy a virtual machine from an OVF or OVA file" and click "Next".
    - Enter a unique name (unique on this hypervisor) for the VM you are about to create, and either drag the .ova file to the drag/drop area (if using the Web client), or click through to browse and select it in your local file-system; click "Next".
    - Select the datastore you wish to use for the VM. If this is a standalone ESXi installation, the only choice is likely to be the local filesystem, "datastore1"; click "Next".
    - Select "Thick provisioned, eager zeroed" (see the section "VMware client selection and operation" for clarification) and then click "Next".

- At the summary screen, click “Finish”, and wait until it is clear that the file has been uploaded.

- Click on “Virtual Machines” in the left-hand window, observe that the new VM has been created, power it up if necessary.

On the VM console, you should see a blue Aculab console screen displayed, make sure it is expanded to see all of the screen.

From this point, installation is common with that for Hyper-V - please continue at the “Post-VM Installation” section below.

11.3 Installation on Hyper-V

You will need to create a virtual Ethernet switch if one does not already exist. You will need to perform this step once only on a particular Hyper-V instance:

- run "Hyper-V Manager"
- click on "Virtual Switch Manager"
- select "External" from the selection list
- click on "Create Virtual Switch"
- enter a name in the "Name:" box (it doesn't really matter what name)
- select an Ethernet card from the "External network" selection box
- click OK

The virtual appliance is provided for Hyper-V as a Hyper-V export file. To install it:

- run "Hyper-V Manager"
- click "Import Virtual Machine"
- follow the wizard

NOTE

You may need to select the virtual Ethernet switch you created previously.

11.4 Post VM Installation VM Only

On the blue Aculab console screen (accessible via the hypervisor), you will observe that (provided there is an operational DHCP server in the local network) the VM will have obtained an IP address for the VoiSentry 'Admin' interface via DHCP.

If you need to change the VM configuration from the default of 4 virtual cores and 6GB, do this now. Power-down the VM, change the configuration using the hypervisor tools, then power-up again so that the blue console screen reappears.

Before you can run the VoiSentry service, or indeed sensibly access the VM, you will need to set the 'Admin' static IP address (which is the item at the top of the IP table). Enter a suitable IP address and Netmask, press <enter>, then cursor down to "Confirm and Restart Immediately" (you may usually leave all of the other fields you see unchanged), type "y" and <enter>. The VM will restart.

After restart, once the red "Initialising" message has disappeared from the top of the display, you may proceed to log on to your selected IP address using a web browser.

The VoiSentry website will only support modern browsers, and does not support any version of Internet Explorer, but should work with Edge.

Initially, the HTML web UI utilises an SSL certificate that has been self-signed by Aculab, and will not relate in any way to the domain or IP regime in which you have started the VM. As a result, your browser will inform you that this website is not to be trusted. Accordingly, to proceed, you will need to tell the browser to ignore the problem and let you through to the website itself (you can deal with the certificate issue later).

You should now see the "Node Login" screen, and as this is a newly created Node, it will have a single default account named "superuser" with password "password". Log in using the superuser account (you should change the superuser password from the default as soon as possible).

Once logged in, you should see the main 'Status Monitor' page, which provides an overview of the Node.

As the Node is not yet in the 'clustered' state and the VoiSentry service is still not yet running, there is little to see, but one thing you will see is whether or not the Node has managed to access an operational NTP server. A coordinated time reference is essential to the operation of the clustered database, and the Node initially has installed a list of NTP servers from ntp.org. If access to the external internet is denied by firewalling, you must provide at least one accessible NTP server, and make that known to the Node via the administration page 'System Administration, Timezone & NTP Servers'.

To put the Node into an operational mode where verification requests may be serviced, it is necessary to 'cluster' it. Create a 'cluster of one' by accessing the administration page 'System Administration, Cluster Create/Join/Leave' and assigning it a name and password. See the detailed section below for more information.

Initially, the Node will indicate a status of 'Blocked', which is essentially a maintenance mode indicating that it should not be sent verification traffic. Set it to 'Active' by accessing the administration page 'System Administration, Node Active/Block/Power'.

The Node is now ready for you to create tenants by accessing the administration page 'System Administration, Login Create/Modify/Delete'. Login as a tenant and use 'Service Management' to create at least one dataset and one access key for that dataset.

11.5 Installation for Docker

11.5.1 Host Requirements

There can only be one VoiSentry container per Docker host, and it will use the host IP address. The host requirements are the same as for the VM deployments.

The host must be running NTP for a clustered solution, and should ideally be running for a non-clustered solution.

11.5.2 Deployment

Please see the VoiSentry Docker Installation Guide.

11.6 Post installation for Docker

Once the container has been created, and is running, the initialization will be done, and the website will become available on https://<host-ipaddr>:443 . This will take a few minutes.

The VoiSentry website will only support modern browsers, and does not support any version of Internet Explorer, but should work with Edge.

Initially, the HTML web UI utilises an SSL certificate that has been self-signed by Aculab, and will not relate in any way to the domain or IP regime in which you have started the VM. As a result, your browser will inform you that this website is not to be trusted. Accordingly, to proceed, you will need to tell the browser to ignore the problem and let you through to the website itself (you can deal with the certificate issue later).

You should now see the “Node Login” screen, and as this is a newly created Node, it will have a single default account named “superuser” with password “password”. Log in using the superuser account (you should change the superuser password from the default as soon as possible).

Once logged in, you should see the main ‘Status Monitor’ page, which provides an overview of the Node.

As the Node is not yet in the ‘clustered’ state and the VoiSentry service is still not yet running, there is little to see.

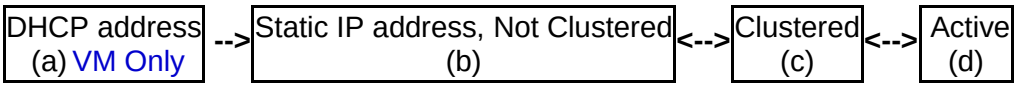
To put the Node into an operational mode where verification requests may be serviced, it is necessary to ‘cluster’ it. Create a ‘cluster of one’ by accessing the administration page ‘System Administration, Cluster Create/Join/Leave’ and assigning it a name and password. See the detailed section below for more information.

Initially, the Node will indicate a status of ‘Blocked’, which is essentially a maintenance mode indicating that it should not be sent verification traffic. Set it to ‘Active’ by accessing the administration page ‘System Administration, Node Active/Block/Power’.

The Node is now ready for you to create tenants by accessing the administration page ‘System Administration, Login Create/Modify/Delete’. Login as a tenant and use ‘Service Management’ to create at least one dataset and one access key for that dataset.

11.7 Node Operational States

There is a sequence of states that a particular Node may move between (in a linear fashion), from installation through to the fully operational ‘Active’ state, as follows:



VM Only In state (a), the web UI will not be running, so the only access is via the blue ‘console’ screen, where you must set a static IP address.

In state (b), the web UI will be running, but the verification services won’t be until you set the Node into ‘clustered’ mode.

In state (c) and (d), the speaker verification services will now be running, although it is strongly recommended to set the Node into ‘active’ mode so that it can correctly indicate it’s state of readiness to any applications using it.

From this, it can be seen that to change the IP address of an Active Node, it must be first set to ‘Blocked’, then declustered, and only then may the IP address be changed (declustering and changing the IP address will both involve a system restart).

A number of administrative operations on a Node will involve putting the Node into the ‘Blocked’ state, and some require the Node to be explicitly declustered.

12 VOISENTRY ADMINISTRATION VIA THE WEB UI

The web User Interface (UI) is intended to allow the entire cluster to be administered by logging into a single Node.

For an administrator, the ‘Service Management’ menu allows limited controls over datasets and access keys - mainly in respect of disabling and deleting those assets. That menu is really targeted at the tenant, as it allows creation and other manipulation of datasets and access keys.

Most of the administrator options are on the 'System Administration' menu.

12.1 Status Monitor – Status Page

After login, the page on view will be the main ‘Status monitor’ Status page, which (for the administrator only) provides a variety of information about the Node, and if there is more than one Node in the cluster, an overview of the cluster, including alarm indications from Nodes other than the one currently being viewed.

It should be understood that much of the information presented on this page has been queued and propagated across Nodes, and also queued between processes upon each Node, then suffering further delays before being updated to the browser itself - so such information may be somewhat delayed before aggregation and presentation.

For a tenant viewing this page, the Status Monitor only provides information about their own licence quota entitlements and usage, along with similar information regarding their own access key utilisation.

12.1.1 Cluster Status

The first table on the Status page provides information about Nodes in the cluster. Each Node entry consists of three lines of data, with a blue indication in the left-hand margin showing the Node currently logged in to.

For rapid assessment, the left and the right areas of each Node's entry have coloured backgrounds: green is 'good', orange is 'good, but in maintenance mode (Blocked) or otherwise needing some consideration', and red indicates that the Node requires attention, possibly urgently.

The central area of the display, to the left shows the Node serial number and IP addresses, the centre shows data disk volume size and % utilisation, and separately the amount of live data on the Node. The database system requires an absolute minimum of 10% spare disk capacity to ensure that it can accomplish certain maintenance tasks, and exceeding that represents a critical situation. Accordingly, VoiSentry sets an 'orange' warning at 70% data volume utilisation, and a 'red' warning at 85% utilisation (the red warning also indicates that the Node has stopped accepting further enrolments).

To the right of centre are some system performance metrics, provided so as to indicate the short-term behaviour of the VoiSentry service. The column marked 'TPM' provides an indication of 'Transactions Per Minute' for Enrolments (above), Verifications (middle), and Identifications (bottom), extrapolated from a count taken over a 5-second period. To the right of that are columns showing average and maximum transaction processing time as seen by the VoiSentry server, calculated within each 5-second sampling period.

On the right, the column headed 'Heartbeat', when green indicates that the Node is running and that various internal monitoring function are active, but when red indicates that the Node is no longer updating its status and so is regarded as 'Unavailable' within the cluster.

The column headed 'Node Alerts', when red and displaying the 'Alert' message, provides a tool-tip displaying specific local concerns upon each Node.

Specific concerns indicated for Node Alerts are:

- NTP status seen from that Node VM Only
- Database volume capacity warnings (if the database on that Node is in danger of filling the disk volume assigned to it).

12.1.2 Cluster Maintenance Messages

Immediately below the 'Cluster Status' table is an area where any important cluster maintenance messages are displayed. This table is only present when such messages are presented, and because of the nature of those messages, they always have an orange background. Usually, the presence of one or more of these messages means that further cluster operations (clustering, declustering, restarting or powering off individual Nodes) is not allowed.

12.1.3 NTP Server Status VM Only

As NTP is so important to correct operation of the cluster, the second table on the Status display show the status of individual NTP servers as seen from the current Node.

It should be noted that immediately after a Node has been started, it may take several minutes for NTP synchronisation to be achieved and the NTP alert message to disappear. NTP alerts at other times though are potentially a real cause for concern.

In principle, if one Node can get a good connection to the NTP servers, then all Nodes should be able to get a response from each nominated NTP server in the list, but each Node may have connected to different IP addresses as each nominated contact is usually (at ntp.org at least) a pool of servers rather than a single machine.

The default is to how just the NTP status banner. To see the information per server the banner needs to be pressed, and it will expand. Pressing again will collapse to the default banner view.

12.1.4 Licences by Licence Server

The next table is ‘Licences by licence server’, which displays all of the licence servers currently known to the cluster.

The licences are shared across all nodes in a cluster. This means there is no restriction on installing licences to specific nodes in a cluster.

12.1.5 Cluster Usage Against Licences

The next table provide an assessment of current usage against licences on a cluster-wide basis, so this is how much of the cluster’s licence capacity is being consumed by all of the tenants on the cluster.

12.1.6 Tenant Usage Against Quotas

The next table provides an assessment by tenant of current usage against quotas. This is significant to an administrator, as it may become necessary to assess or assign additional quota capacity on a per-tenant basis.

12.1.7 Datasets by Tenant

This table is simply a listing of all of the datasets on the cluster (by Tenant), and the number of enrolments stored within each dataset.

12.1.8 Access Key Usage Against Quotas

This table provides some visibility of how the tenant is using the platform, as access keys are really the business of the tenant and not the administrator.

12.1.9 Account Logins

The final table is again for information only, and provides the administrator with a view of who is logged in to the cluster, from where, and to which Node.

12.2 Status Monitor – Verifications Page

The 'Status Monitor' Verifications page will show a table of the verifications per day for each node in the cluster. There is also a 'Graphs' button, which when pressed will show a graph for all tenants cumulative verifications per day, and one graph for each tenant showing the verifications per access key.

12.3 Status Monitor – Identifications Page

The 'Status Monitor' Identifications page will show a table of the identifications per day for each node in the cluster. There is also a 'Graphs' button, which when pressed will show a graph for all tenants cumulative verifications per day, and one graph for each tenant showing the verifications per access key.

12.4 Common Administrative Operations

12.4.1 Setting a Node's Active/Blocked Status

A Node may be set either to the 'Active' or 'Blocked' (maintenance) state via the administration page 'System Administration, Node Active/Block/Power'.

Setting a Node to 'Blocked' does not actually prevent it from processing verification requests, but this status is indicated in all API results, and Nodes advertised as 'Blocked' should not be used considered as an active part of the Cluster. In particular, changing the status of a Node from 'Active' to 'Blocked' will prevent the Node from powering down or restarting for a set period of time, typically 60 seconds. This feature ensures that any pending API calls results can be completed and their results returned.

A number of administrative operations on a Node (power down, restart, decluster, etc.) require it to be set to 'Blocked' before they may be carried out.

12.4.2 Powering Down and Restarting a Node

A Node may be either restarted or powered down VM Only via administration page 'System Administration, Node Active/Block/Power'. A Node may only be powered down or restarted when set to maintenance mode ('Blocked').

VM Only There are also facilities for powering down and restarting on the blue 'console' screen (available via the hypervisor), and the hypervisor itself has similar facilities. Please do not use these methods except in the event that the administrative UI is unavailable.

12.4.3 Creating a New Cluster

Creating a 'cluster of one' from a currently unclustered Node was briefly covered in the Quickstart Guide, above - essentially this is achieved via the administration page 'System Administration, Cluster Create/Join/Leave'.

Two pieces of information are required when creating a new cluster - these are a cluster name (which is subsequently displayed on the top bar of the Cluster Status table on the main Status page), and a cluster password, which is not displayed anywhere. This is to avoid the security concern of someone attaching a new Node to a cluster for nefarious purposes. Accordingly, this password must be kept safe. Two pieces of information are required when creating a new cluster - these are a cluster name (which is subsequently displayed on the top bar of the Cluster Status table on the main Status page), and a cluster password, which is not displayed anywhere. This is to avoid the security concern of someone attaching a new Node to a cluster for nefarious purposes. Accordingly, this password must be kept safe.

12.4.4 Adding a Node to an Existing Cluster

Adding a currently unclustered Node to an existing cluster (whether a 'cluster of one' or a 'cluster of many') can be achieved via the administration page 'System Administration, Cluster Create/Join/Leave'.

Select 'Join a Cluster' and enter the information requested there, which is the name of the target cluster, the cluster password of the target cluster, and the IP address of any Node in the target cluster. Clicking 'Submit' will cause the Node to exchange information with the cluster, and if the cluster name and cluster password is correct, the Node will be restarted and in due course will come up added to the cluster (but initially in maintenance mode).

In joining the cluster, the Node will acquire a local copy of the cluster database, meaning that it will accept all of the logins known by the cluster.

In a cluster that already contains a very large number of enrolments, it will take a considerable amount of time to assemble the replica database (typically several hours), and neither the admin UI nor the various web service interfaces will be available on the new Node during that process. A progress report will be displayed on the status page of existing Nodes. This database replication will produce a significant, albeit controlled amount of additional traffic from existing Nodes in the cluster to the new Node, so in a production system it is advised to schedule the operation for a time when the cluster is less busy.

To allow it to indicate that the Node is ready to start accepting requests, you must set it to 'Active' mode.

12.4.5 Removing a Node from a Cluster (Declustering)

Removing a Node from a cluster can be achieved via the administration page 'System Administration, Cluster Create/Join/Leave'. Select 'Decluster a Node' and select the Node that you wish to decluster (you do not need to be logged in to that Node to decluster it). Declustering a Node will involve an immediate Node restart.

The caution given previously regarding declustering is repeated here:

NOTE

Important! When a Node is declustered, its local copy of the cluster database is deleted.

NOTE

You will not be allowed to decluster a single node cluster that contains licences. You will need to remove the licences first. This is to stop accidently losing you licences by declustering.

If a Node is in a 'cluster of one' - i.e. it is the last Node of a cluster, declustering it will discard all of the datasets, access keys and account logins owned by the cluster - all of the cluster assets will be deleted.

Accordingly, if a stand-alone Node needs to be declustered for maintenance (for example, to add an extended Node Database Volume), the cluster data can be preserved by creating a new Node, adding that Node to the cluster in order to create a copy of the cluster database, then declustering the original Node for maintenance.

In respect of impact on a cluster, declustering a Node is a low-impact operation compared to adding a Node (clustering).

12.4.6 Forcing out an Unavailable Node from a Cluster

The final option on the administration page 'System Administration, Cluster Create/Join/Leave' is 'Force-out a Node'.

This function is only required when a Node has been physically removed from a cluster (and possibly destroyed), but has not been formally removed from the cluster (declustered). This is useful as the cluster retains knowledge of all Nodes in the cluster, even if they are not currently available - it is essentially a 'housekeeping' activity that should ideally never be required, as Nodes should normally be removed by declustering.

12.4.7 Changing a Node IP Address VM Only

The static IP address of a Node may be changed via the administration page 'System Administration, Node IP Address'. A Node is known to a cluster primarily via its IP address, so in order to change that address, it must first be removed from the cluster by declustering it.

Changing the IP address will involve a Node restart.

12.4.8 Creating Admin and Tenant Logins

The 'superuser' can create multiple 'admin' accounts in order to delegate almost all administrative tasks, and may also create 'tenant' accounts for users who intend to avail themselves of speaker verification services - both of these are available via the administration page 'System Administration, Login Create/Modify/Delete'.

Initially, the page displays a list of existing logins along with (in the case of tenants) the quotas assigned to them. This list provides embedded links that enable the administrator to delete, or enable or disable individual logins, and also (in the case of tenants) to edit the quotas assigned to the account.

To create a new 'admin' account, click on 'Create New Admin Account' and enter the information requested there.

Similarly, to create a new 'tenant' account, click on 'Create New Tenant Account' and enter the requested information, which in this case includes quotas for both Enrolments and Verifications per day.

For completeness, only a tenant can create ordinary 'user' accounts, which allow them to delegate just their own tenant related tasks (and that is the only 'System Administration' option available to a tenant).

12.4.9 Managing Licences

System subscription and Verification/Identification licences may be added, or deleted, for any of the node's internal licence servers, or any of the configured external licence servers. The 'Manage Licences' tab on the 'System Administration, Cluster Licence Servers' administration page is used to add, or delete licences. There is a separate section for each node and external licence servers, which can be expanded or collapsed by clicking on the relevant Licence server header.

Depending upon the installation, licensing changes may require offline use of the Aculab Licence Tool (ALT) where the Aculab licence servers cannot be reached due to network restrictions, i.e. firewalling. When adding or deleting a licence which requires offline use of the ALT, click the 'Install Licence/Key' button to enter the licence to add, or click the 'Delete' button, and a new key will be shown in the ALT Key column. This ALT key must be entered into the ALT tool (from a location that can reach the Aculab licence servers), and the returned key entered using the 'Install Licence/Key' button on the VoiSentry 'Manage Licences' page.

If an installation does not need to use the offline ALT, then adding, and deleting licences can all be done from the 'Manage Licences' tab.

12.4.10 Adding and Deleting External Licence Servers

External (non-Node) licence servers may be added or modified via the 'External Licence Servers' tab on the administration page 'System Administration, Cluster Licence Servers'. Click on 'New Server' and enter the IP address and port number (if the active port has been changed from the default port of 7270), then click 'Submit'. Licence servers may be deleted by clicking 'Delete'.

External licence servers and the licences they present are documented on the main Status page.

Rationale regarding the use of external licence servers has been provided in the 'Introduction and Overview' section under 'VoiSentry Licensing', and in the 'Licences by Licence Server' when describing the main Status page.

NOTE

Note that, if either a licence server contained in a Node or an external licence server becomes unavailable, the licences it issued are still considered available to the Cluster for a time-out period (normally one hour). This provides a 'grace period' that enables essential maintenance to be performed without affecting service capacity.

NOTE

Also note that from the standpoint of an external licence server, individual Nodes in a Cluster periodically request licences from it, and it considers those licences as 'Issued' for a period of one hour from last use, and so will not issue those same licences to another requesting entity until that time-out period has elapsed.

12.4.11 Licence Administration - Support Information

There is a 'Support Information' tab. This provides information that Aculab support may ask for when contacting them about any licence related questions for your deployment. It will display information about the licences installed on the cluster, and whether these licences are 'subscription' or 'perpetual', and whether they have expired.

12.4.12 Licence Administration - Update/Check Support

There is an 'Update/Check Support' tab. Aculab support may provide you with a file to upload to either check or update your supported licence information.

12.4.13 Changing Timezones

The timezone of the entire cluster may be modified via the administration page 'System Administration, Timezone & NTP Servers'. Internally, Nodes use a fixed UTC time reference obtained from NTP servers, and the timezone setting only affects time as presented on various status pages.

The 'Configure Timezone' tab provides the ability to select by 'Continent or Ocean', and then by 'City or Region', which will normally also select 'daylight saving' as appropriate to the culture of the region.

Alternatively, the last entry on the 'Continent or Ocean' select box is 'UTC-GMT-Offsets', which allows for fixed offset from UTC (GMT) in increments of 1 hour that will not be subject to daylight saving modifications.

12.4.14 Changing NTP Servers VM Only

As already mentioned, an accurate time reference is critical to operation of the cluster, and every Node as initially installed contains a list of NTP servers provided by ntp.org.

This list may be modified or added to via the administration page 'System Administration, Timezone & NTP Servers'. Changes to the list on one Node will be propagated across the cluster. Attempts to delete all of the servers in the list will result in the original default list of NTP servers being reinstated.

If access to the external internet is prevented by firewalls or otherwise, you must provide a local NTP server, and make that known to the cluster via this admin page. Lack of a working NTP time reference will result in unpredictable behaviour between Nodes in the cluster.

12.4.15 Creating a Technical Support package

If Aculab support requires information to investigate a problem then they will probably ask for a technical support package from the VoiSentry node, or nodes.

To create a technical support package navigate to the administration page 'System Administration, Technical Support Package'. Select the node for which the package is to be generated for, and also the time span to collect files, then press the Create button. Only one package may be generated for the cluster at any time. Once the package has been generated the Snapshot available section will show that a file has been created and will provide a Download button. Press the Download button to download the package. Once downloaded, the process can be repeated for other nodes if required.

12.4.16 Viewing the web system message log

When performing certain actions using the website, there may be system messages show for a short amount of time. These are also stored in a log file which can be

viewed from the ‘System Administration – Cluster Log Files’ page.

12.4.17 Installing Website SSL Certificates

As mentioned in the Quickstart Guide, the admin web UI (which is essentially a web server) has a certificate installed that has been issued and self-signed by Aculab plc. This certificate operates perfectly well for securing HTTPS traffic between the web server and the browser, but will not satisfy most modern browsers in the respect that it is not traceable to a well-known CA and nor does it explicitly authenticate the user domain.

As a result, it is likely that most browsers will hesitate to load the first page of the site, and will require an explicit authority or exception before it will do so, and even then they are likely to permanently display some kind of complaint in the region of the 'https:' part of the URI, or thereabouts to show that the site isn't trusted.

If any of this is a problem, the solution may be to obtain a proper certificate for the domain or public IP address you wish to use for the admin web UI. Navigate to the administration page 'System Administration, Node Site Certificates & SSH'.

At the line that is labelled 'Select new private key and site certificate to upload', click on 'Browse' and browse on your local machine to the file containing the certificate you wish to install. Enter the passphrase for the certificate (if one is required) and click on 'Submit' to upload it.

If the certificate passes inspection, it will be installed immediately. To inspect the new certificate, refresh the page (important!), the click on the 'Info' bar to view the derived information, or click on 'Certificate' to see the certificate itself.

NOTE

Note that this process installs the certificate on the current Node only, and does not propagate it across the cluster.

12.4.18 Enabling and Disabling SSH Access VM Only

SSH login is disabled by default, but may be enabled either via the blue console display (accessible via the Hyper-V or ESXi hypervisor), or via the administration page 'System Administration, Node Site Certificates & SSH'. Enabling or disabling SSH access affects just the one Node upon which the operation is performed.

The main reason for enabling SSH access is to allow Aculab support engineers to access a Node for diagnostic purposes, so there is really no reason for it to be left enabled by default.

12.4.19 Adding an External (Extended) Database Volume VM Only

The VoiSentry VM contains an internal database volume of around 4GB in size, and this volume provides sufficient database capacity for many smaller applications, up to perhaps 20,000 enrolments. The main Status page will show a Node Alert when about 70% of this capacity is exceeded, and will prevent further enrolments after about 85% of capacity is exceeded (some spare capacity is always reserved for routine maintenance purposes). Explicit usage information is available by selecting 'System Administration, Node Database Volume'.

To convert a Node to use an 'external' disk volume for database storage, you must first ensure the Node is declustered, then create a new, unformatted virtual disk volume of suitable size using the Hyper-V or VMware hypervisor administrative functions. Then add it to the VM so that it will be seen as an external disk (the VM will typically need to be powered down in order to add a new disk to it, then restarted to successfully recognise it).

The recommended practice is that the new virtual disk is created in an SSD - if so, when using VMware, be absolutely certain to use 'thick provisioned, eager zeroed' to create the new disk volume - it will take more time to create than when using 'thin provisioning' (and much longer again under ESX 6.5 than ESX 6.0) it but it will provide much better disk performance when the volume is actually in use. 'Thin' provisioning can cause severe operational problems in the form of hugely increased disk latency as additional disk space is assigned.

If the new virtual disk is created in a remote backing store (as is often common practice in a data centre), then 'thick' provisioning might not be so important, although for the absence of doubt, thick provisioning is always recommended.

Also, if using such a remote backing store, it is vital to test overall performance before putting the system into production. This is even more important if a multi-node Cluster is to be supported using the same backing store, particularly in respect of the performance of VoiSentry itself, but also for resilience (it creates a single point of failure), and for the impact upon that datastore from the standpoint of other workloads it may be supporting.

Even in a data centre where the use of such back-end storage for VMs is absolutely standard practice, deploying VoiSentry upon high-capacity SSDs installed in the compute Node itself may still be the most efficient choice - performance of VoiSentry is highly dependent upon the performance of the underlying data storage.

Regardless of how the virtual disk is provisioned, to add an extended disk volume, navigate to 'System Administration, Node Database Volume', and click on the tab 'Add Extended Database Volume', which will display any disks that the VM can use for the extended volume.

Select the candidate disk and click on 'Submit', which will restart the Node using the new volume, whereby it may then be added back into the Cluster.

12.4.20 Importing (recovering) an External (Extended) Database Volume VM Only

This option allows you to attempt the import (recovery) of an existing Extended Volume Disk, along with all its data.

If successful, this Node will end up in a new cluster, with all of the attributes of the cluster from which the Volume originated – including all accounts and passwords.

12.4.21 Upgrading Node System Software VM Only

NOTE

Note that there may be some special procedures or caveats that apply to specific upgrades, so before attempting any kind of upgrade, be certain to read the accompanying release notes.

One method of upgrading a Node is simply to create a new VoiSentry Node using an updated VM image, adding that to the Cluster, and retiring the old Node by declustering.

NOTE

If you upgrade by creating a new node with the updated image, and adding it to the cluster, you **MUST** ensure it has fully synchronised the data from the cluster before retiring the old node.

Alternatively, you may upgrade a Node 'in-place' by uploading a suitable upgrade file to it. Upgrade files are provided in a file identified in the pattern 'ast_package_XXXXX.zip', where 'XXXXX' consists of version information. This file must be renamed or copied to filename 'ast_package.zip' and placed in a suitable location on the PC being used for admin UI access.

Browse to 'System Administration, Node Software Upgrade', click on the 'Browse' button, and navigate to where the file 'ast_package.zip' is stored. Select it and click on 'Submit', which will start the upload process which should take just a few seconds.

The upgraded software will be installed the next time the Node is restarted. The status page will show that a node has a pending software upgrade until it is restarted.

13 VOISENTRY ADMINISTRATION VIA REST

Most of the administrative functions that are available via the HTML (web) User Interface are also achievable via Web Services (REST) calls. This enables administrative functions to be integrated into third-party applications, as used in call centres.

The REST interface is accessible at the following URI, where $TARGET is the IP address of the Node, and <REST_function_invocation> represents the function name and query string of the REST function to invoke:

https://$TARGET:/ws/<REST_function_invocation>

Specific examples are shown using the ubiquitous command-line program 'curl', which is available for most operating systems.

As administrative tasks are notionally being undertaken on behalf of a particular administrator or tenant, along with every REST call you must provide a valid username and password that is appropriate to the request being made.

REST calls return a JSON object as shown in the REST documentation.

14 VOISENTRY OPERATIONAL CONSIDERATIONS

This section presents some matters for consideration when planning and operating a VoiSentry implementation.

14.1 VMware Virtual Disk Deployment

There is an important distinction in the way in which virtual disks are provisioned - both the base VM (from the .ova file) and any external datastore volume (external virtual hard disk) that might be attached to the VM.

There are essentially three choices:

- '**Thin provisioned**', whereby only the storage actually used by the virtual disk is committed by ESXi
- '**Thick provisioned, lazy zeroed**', whereby all of the virtual disk is notionally committed, but the actual zeroing of the physical disk space is deferred, and performed 'lazily"
- '**Thick provisioned, eager zeroed**', whereby the physical disk space is committed and zeroed before the virtual disk is released for use.

If the ESXi instance is deployed within a data centre that utilises an intelligent back-end storage subsystem, and ESXi identifies that such a back-end datastore provides 'Hardware Acceleration', then it *might* be possible to use either 'Thin provisioned' or 'Thick provisioned, lazy zeroed' without any ill-effects, depending upon the nature and efficiency of the subsystem.

However, if the ESXi instance is using storage without hardware acceleration, and any kind of local disk system (SSD or otherwise) will *not* provide such acceleration, then it is vital to select 'Thick provisioned, eager zeroed' when creating VMs or virtual disks for use as external datastore volumes. This is certainly always the safer option.

The consequence of choosing the wrong type of provisioning will be extremely poor datastore performance during disk writes, to the extent that the Node or Cluster may demonstrate various failure conditions (Node dropouts, cross-Cluster communication problems, etc). VoiSentry will certainly not reach optimum performance.

The problem appears to occur because of the time it takes for the hypervisor to provision, and then zero the required disk data blocks before the data block is released for new data to be written to the disk. A further observation is that simply choosing 'Thick provisioned, eager zeroed' seems to take much longer on ESXi 6.5 than on ESXi 6.0. It may be that, given enough time since provisioning, ESX will have overcome its laziness and complete the zeroing of the virtual disk, but the various management Clients give no indication of how long that would take, or when it is complete.

Refer to the following section on "VMware Client selection and operation" for additional information regarding how to ensure the correct type of provisioning.

14.2 VMware Client Selection and Operation

At the time of writing, there were two VMware ESXi versions available for evaluation, being version 6.0 and 6.5.

With an evaluation user account registered with VMware, you can download an iso for the latest version:

https://my.VMware.com/en/web/VMware/evalcenter?p=free-esxi6

There are some very significant usage issues that may arise from the choice you might make between these versions, particularly your ability to ensure the proper type of virtual disk provisioning (see the previous section on "Virtual disk deployment" to understand the problem).

ESXi version 6.0 provides options on the ESX login page of:

- Download vSphere Client for Windows
- Open the VMware Host Client

By contrast, ESXi version 6.5 takes you directly to the "VMware Host Client" login page, with no ability to download the "vSphere Client for Windows".

There are significant differences between these two clients, and important differences between versions of the "VMware Host Client", as detailed below.

14.2.1 vSphere Client for Windows

This is a Java package for recent versions of Windows, and purports to only support features of VMware ESX up to and including those supported in version 5.0.

Regardless of that claim, from the standpoint of deploying and managing VoiSentry VMs and virtual disks, it has proven to provide a very stable and capable management console, and it's highly recommend its used for that purpose.

Even though it is only available for download from an ESXi 6.0 installation, it is equally capable of managing either ESXi 6.0 or 6.5 hypervisors.

One particular benefit is that, when deploying a VoiSentry VM (a .ova file), and also when creating a virtual Hard Disk for use as an external datastore volume, it always allows you to explicitly select "Thick Provisioned, Eager Zeroed".

14.2.2 VMware Host Client

This is a direct HTML based user interface for managing the ESXi hypervisor. The interface provided on ESX 6.5 seems more stable than that provided by ESXi 6.0.

The important point is that when deploying a VM from a .ova file, both Host Clients only allow the selection of either "Thin provisioning" or "Thick provisioning", with no explicit selection of "Lazy" or "Eager" zeroing.

It also appears that the default under ESXi 6.0 for "Thick provisioned" is "Thick provisioned, lazy zeroed", whereas the default for (at least the latest update of) ESXi 6.5 is the much more desirable "Thick provisioned, eager zeroed", so only the host client associated with the latest update of ESXi 6.5 is recommended for use with the "VMware Host Client".

A further observation is in relation to creating a virtual hard disk for use as an external datastore volume. When editing setting on a VM and you select "Add Hard Disk", a panel opens that allows you to set the capacity of the hard disk you wish to create. Immediately to the left of the "New Hard Disk" label is a tiny downwards-pointing triangle, and only by clicking on that do you see the options of where to create the virtual disk, and the style of provisioning. In this case, all three provisioning options are presented, so select "Thick provisioned, eager zeroed".

In summary: if your workstation is running Windows, you may either use ESXi version 6.0 (or prior version) and install the 'vSphere client for Windows', or use the latest update of ESXi 6.5 and use the 'VMware Host Client' (HTML interface).

Otherwise (e.g. if your workstation is running Linux) your only option is to use the latest update of ESXi 6.5 and the 'VMware Host Client' (HTML interface). To download this version, you will need to register on the VMware website.

14.3 Configuring VMs on VMware ESXi

There are a number of parameters to consider when selecting a virtualisation host, and configuring a VM for VoiSentry - many of which will have a bearing on Speaker Verification performance.

There are perhaps two aspects to performance, being:

- Verification Accuracy
- Verification Latency

This section considers optimising for Verification Latency, being the elapsed time before sending a voice file to VoiSentry, and receiving a valid response - it has no bearing upon Verification Accuracy.

Speaker Verification is simultaneously a CPU-intensive task, and both a memory intensive and disk intensive task (in that enrolment records are quite large), so selection of a server with appropriate characteristics is important in order to minimise latency.

For the purpose of discussing latency here, a latency of 0.5 seconds or less is regarded as acceptable. In reality, an acceptable latency is entirely determined by the application, and applications can be constructed that are either intolerant or accommodating of higher latencies.

14.3.1 Selecting a Virtualisation Server

For low levels of verification traffic (for example, only one verification request at a time), verification latency will depend almost entirely upon the CPU clock rate, in that (all other things being equal) a faster CPU clock will result in lower latency.

There has been a readily observed trend over time in server CPUs for more recent CPUs to have a generally slower clock rate (which will increase latency), but also to have more CPU cores, and more CPU cores tend to support more parallelism, meaning the ability to support more simultaneous verifications.

At the same time, more recent CPU architectures seem to have slightly improved their efficiency in processing numerical operations, which improves latency a little, so it is very difficult to make specific recommendations - other than to say that within a given CPU family, a faster CPU clock will result in less latency.

14.3.2 Configuring VMs on the Server

There are some slightly curious observations to be made regarding the observed performance of VoiSentry VMs on a modern virtualised server that are due to the combined CPU, memory and disk intensive nature of the workload, and perhaps the nature of Hyper-threaded cores versus non Hyper-threaded cores, and how VMware deploys them under particular workloads.

Without trying to rationalise the observations any more than that, here are the recommendations (where vCPUs refers to ‘hyper-threaded cores’):

1. For reasonable performance, the VM should not have fewer than 8 vCPUs.
2. Beyond 8 vCPUs, optimum results are often obtained when the VM has around half the number of vCPUs on the server - so if the server has 24 vCPUs, for best performance the VoiSentry VM should have 12 (or maybe 14) vCPUs.
3. Two VoiSentry VMs each having 8 vCPUs will give significantly better performance in terms of TPH (Transactions per Hour) and request latency on a single server than one VM having 16 vCPUs.
4. A VM should have 1.5GiB RAM per vCPU, although more than 16GiB provides no benefit (although too little RAM is very much a problem).

Point 3 in particular is something of a surprise, but in practice if you intend to deploy two VMs, it makes a lot of sense from a resilience standpoint to place them on separate servers.

14.4 External Datastore Volume Disk-space Management

VoiSentry requires around 10% of the total disk capacity for database maintenance and management, with an additional 5% reserve to ensure that the system never runs out of space. As a result, a hard limit is set at 85% of disk capacity such that when that limit is reached, no further enrolments are accepted, and a red alarm will be displayed on the UI Status.

The enrolment hard limit is based upon the capacity of the Node that is closest to the limit - enrolments are replicated across the Cluster so if datastore capacities are not all the same, the smallest capacity is the determining factor for the onset of alerts and limiting.

However, it is inadvisable to allow such a hard limit to arrive without warning, so when 70% of disk space has been used, a “Yellow Alert” will be displayed on the UI status display, and an alert indication (diskalert) will be returned in all API calls.

Despite this, it would be a mistake to believe that upon hitting the 70% Alert, there is still 15% to use before any real problems occur. If a particular system has hit 70%, and then database management starts up on one Node and consumes an additional 10% of disk space, for a while at least there will only be around 5% of margin before hard-limiting.

Therefore, the 70% “Yellow Alert” should be treated very seriously, and additional capacity organised ideally even before that point is reached.

14.5 Load Sharing Across a VoiSentry Cluster

A VoiSentry installation may consist of a single Node running in ‘clustered’ mode, or several Nodes clustered together. In the latter case, in order to achieve the best (lowest) latency, it is advisable to share the total verification load evenly across the Cluster.

In an ordinary web-server environment, load-sharing is often achieved by using a reverse-proxy as the initial target of all service requests, and maintaining that proxy with knowledge of all the servers able to process requests so that it distributes the processing load in any required manner. This can be readily achieved with VoiSentry by using software such as Nginx (along with a few scripts), but the approach does have the disadvantage of making the reverse proxy itself a single point of failure.

Alternatively, in the absence of a reverse proxy, provided a client application is furnished with a means of identifying the IP address of at least one Node in the Cluster, a single ‘ping’ API call to that Node will return a list of all IP addresses of Nodes currently displaying an active (‘A’) node status.

Clients should obtain this list periodically, and use it to load-share requests across the cluster, and provided all Nodes are of equal processing capability, a round-robin approach to distribution is probably as good as any other.

Each Node will respond to all requests whether in active (‘A’) or blocked (‘B’) mode, but it is intended that Nodes currently in use should be administratively set to the active (‘A’) mode, and only set to blocked (‘B’) as a precursor to powering them down or otherwise removing them from service.

Upon such a change from active to blocked, an internal timer prevents power down, restart or declustering for a period of 60 seconds.

Accordingly, it is recommended that each client retrieves the list of active Nodes at least every 30 seconds in order to ensure that subsequent requests are only presented to active Nodes.

In addition, every API request returns the current active or blocked status of the Node that processed the request.

14.6 VoiSentry Verification Accuracy

When performing a speaker verification task, VoiSentry is presented with some audio and an identity claim. The system will then calculate a confidence that the audio matches the identity. In order for the verification attempt to pass, the confidence must meet a predetermined measure. In VoiSentry, the predetermined measure is represented by the value 1, and the confidence must be greater than or equal to 1 for the verification attempt to pass. The VoiSentry user must not use a different measure – see `sensitivity` below.

14.6.1 Sensitivity

The accuracy of a speaker verification system is judged by the number of false identity claims that fool the system; these imposters manage to get a high enough confidence to pass as the speaker. VoiSentry provides a method (called sensitivity) to increase or decrease the strictness of the system. If too many imposters are fooling the system, the sensitivity can be raised to counteract this.

A second measure by which a speaker verification system is judged is the number of true identity claims that are rejected. These are genuine speakers who fail to pass. If too many true speakers are rejected, the sensitivity can be lowered.

The sensitivity is the primary tool that affects the speaker verification system's performance. Increasing the sensitivity makes it more difficult to fool the system, but also makes it more difficult for genuine speakers to pass. Decreasing the sensitivity makes it easier for genuine speakers to pass, but will also allow more imposters through.

14.6.2 Equal Error Rate

A common measure for judging a speaker verification system is the Equal Error Rate (EER). This is the level where the percentage of false claims that succeed (false acceptances) is equal to the percentage of true claims that fail (false rejections). If a system claims an EER of 2%, it means that it is expected that two percent of imposters will pass and two percent of true speakers will fail. The default sensitivity level for VoiSentry should provide an EER between 2 and 3 percent.

The EER advertised for a given system is, however, not a guaranteed rate for a deployed system. When speaker verification is tested, the test outcome is subject to a fairly large set of variables, each of which can affect the EER that is calculated. Some of these variables are discussed below.

When calculating the confidence for an identity claim, a comparison is made between the audio data that is offered as part of the claim and the voice model of the claimed identity. There are several variables that can affect the confidence:

1. If the spoken text used to train the voice model is the same as the spoken text making the identity claim (text dependent), the confidence will be higher than if the text is different (text independent). For instance, given a voice model trained on repetitions of the phrase "My voice can let me in", the score will be higher if the same phrase is used during verification than a different one like, "I am who I say I am."

2. If the audio device (e.g. a mobile telephone) connected to the verification system is the same during model training and identity verification, the confidence will be higher than if a different device is used. For instance, given a model trained using a mobile phone, the score will be higher if the same phone is used during verification than if a SIP phone or Skype call is used.
3. The environment can have an effect. If the voice model is trained in a very quiet environment, a verification attempt from a noisy environment, such as a car, will produce a lower confidence.
4. The amount of data (length of utterances) used for training and verification will have an effect. As a general rule, increasing the amount of data for training will result in a better (and more representative) voice model. Also, long verification utterances will perform better than short ones.

When testing a verification system with the aim to calculate an EER, the variables described above are taken into consideration. A text dependent test (training and verification done using the same wordy phrase) where the speakers always use the same phone in the same environment will give the best EER. But this might not be representative of the real-world scenario where the system is to be deployed. If the deployed system is going to allow speakers to say whatever they want, from wherever they want, the performance may well be significantly lower than advertised.

It is also important to understand that relying on the EER is not best practice for every deployment, it is merely a useful benchmark to look at. An EER of 2% may sound good, but out of every 100 true verification requests 2 could fail. For a product where being customer friendly is vital this might not be acceptable, in which case lowering the sensitivity to decrease false rejections but allow more false acceptances might be a solution. Similarly, a high-security deployment might want almost no false acceptances, but be content with more false rejections.

Designing the real-world system to cope with different audio devices, different environments and even different phrases is simply a matter of knowing when to update the speaker's voice model. It is good practice to keep a voice model current by updating it with the latest audio provided by the speaker during a verification attempt, providing that new audio has been positively validated. Naturally, the voice model should only be updated if the speaker is not an imposter. However, if a speaker fails a verification attempt they are not necessarily an imposter - they might be in a different environment or be using a different phone. The verification system should have a fall-back mechanism whereby the speaker can verify who they are through a different verification procedure. If they pass the fall-back procedure, then the audio can be used to update the model. The next time the speaker phones in, the model will contain the new information and they will be more likely to pass.

In conclusion, there are many factors that can affect the performance of a speaker verification system. It may be that not all of these factors were taken into consideration when calculating the EER of the system; so, the EER should not be taken as a performance guarantee.

14.7 VoiSentry Verification Versioning

At the time of writing, a VoiSentry Node has exactly one set of internal algorithms for enrolment and verification, for reference purposes set as “version 1”. It is anticipated that over a period of time, there are likely to be improvements to the system in the form of more advanced algorithms that provide better results. Provided these do not use fundamentally incompatible data formats, it is intended that new releases of the VoiSentry VM will contain the original version alongside such subsequent versions.

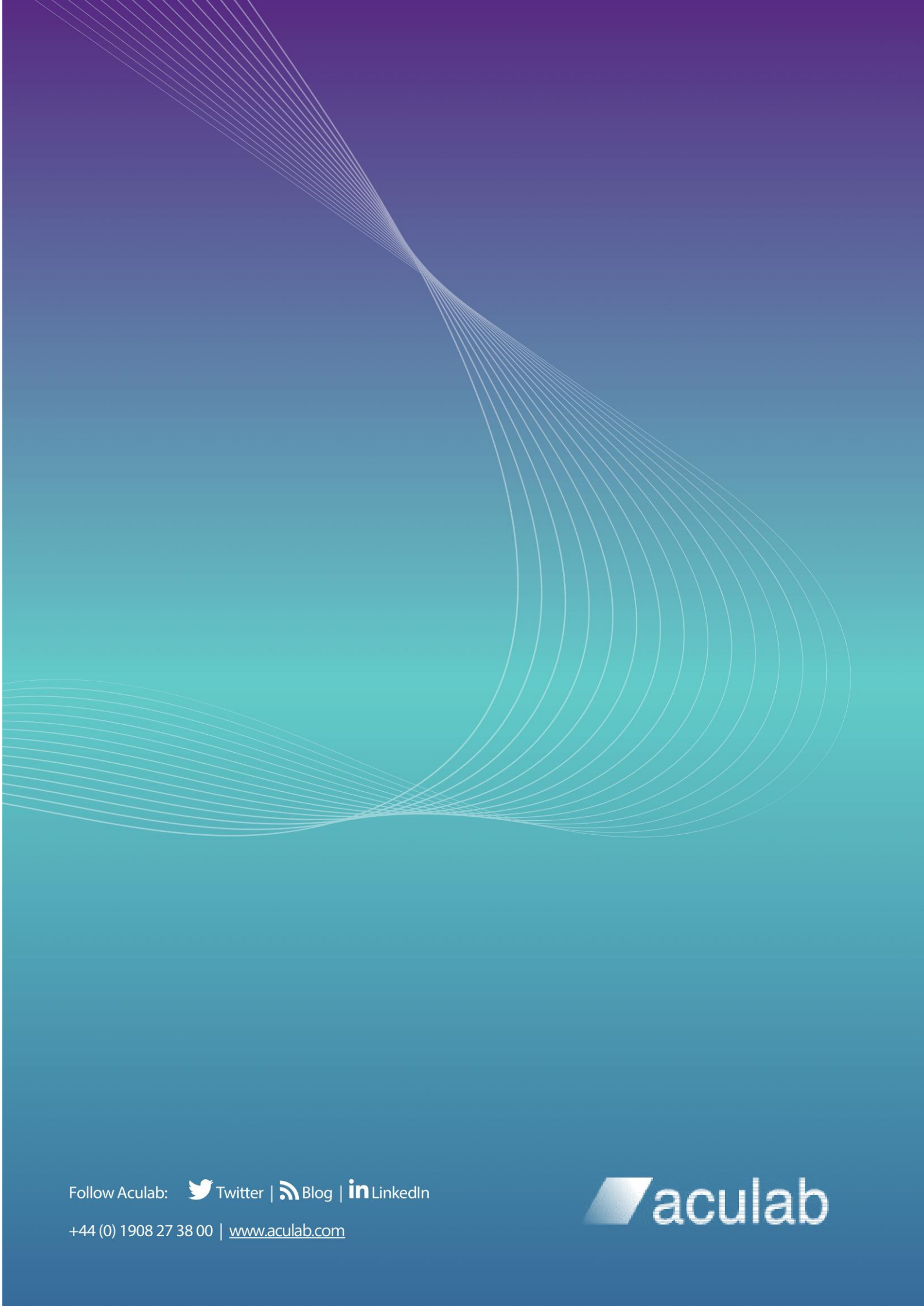
VoiSentry has been designed to run all versions of its speaker verification engine concurrently and invisibly. All new enrolments will automatically use the latest version of the engine, all speaker verifications will use the same version of the engine that was used during the speaker's enrolment.

Any verification attempts on existing voice models will continue to use the version that created them (this avoids any incompatibility issues). New enrolments will use the latest version. Over time, it is possible that two or more versions of the engine are being used for verifications.

Such an upgrade to the verification engine may introduce improvements that will benefit existing voice models. To update these models to the latest version the speakers must be re-enrolled. Re-enrolment can be done without inconveniencing the speaker provided their original recordings are still available.

Accordingly, it is advised that the original set of audio files that were used to enrol a speaker is stored. Over the lifetime of a voice model it will be regularly updated with new audio, particularly on the occasion when a verification fails and the speaker is verified by other means. These audio recordings should also be stored.

The speaker can be re-enrolled at any time using the stored files. The new voice model will automatically use the latest version of the verification engine.



Follow Aculab: Twitter | Blog | LinkedIn

+44 (0) 1908 27 38 00 | www.aculab.com

aculab