European Commission | Horizon 2020 European Union funding for Research & Innovation

Big Data technologies and extreme-scale analytics

Multimodal Extreme Scale Data Analytics for Smart Cities Environments

D3.3: E2F2C Privacy preservation mechanisms[†]

Abstract: This document describes the mechanisms deployed in MARVEL to preserve data privacy in the Edge-to-Fog-to-Cloud (E2F2C) computing continuum. This public document describes the efforts carried out under Task 3.1 for the development of "AI-based methods for audio-visual data privacy". The main goal of the task is to work on methods and algorithms capable to run at the edge and ensure protection of personal data in audio-visual recordings. Although the task focused on solutions for audio and video anonymisation capable of preserving the scene information while removing any information about the identity of the citizens being recorded, privacy preservation in the E2F2C computing continuum is ensured in the MARVEL project in multiple ways. Therefore, even though the core part of the document will be on video and audio anonymisation, we will also provide short descriptions of encryptions and cybersecurity methods (T3.4), approaches for porting the AI processing at the edge (T3.3 and T3.5), and security and authentication mechanisms implemented in the MARVEL Data Corpus in WP2. In this way, we aim at highlighting the connections of those approaches to the ones developed in T3.1 towards increased privacy preservation in MARVEL. To better contextualise the motivations and the challenges behind the adopted solutions, the document also revises the pilot-specific privacy issues that the project has to consider.

[†] *The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337.*

Contractual Date of Delivery	31/12/2022
Actual Date of Delivery	30/12/2022
Deliverable Security Class	Public
Editor	*Alessio Brutti (FBK)*
Contributors	All *MARVEL* partners
Quality Assurance	*Alexandros Iosifidis (AU)* *Nikola Simic (UNS)*

The *MARVEL* Consortium

Part. No.	Participant organisation name	Participant Short Name	Role	Country
1	FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS	FORTH	Coordinator	EL
2	INFINEON TECHNOLOGIES AG	IFAG	Principal Contractor	DE
3	AARHUS UNIVERSITET	AU	Principal Contractor	DK
4	ATOS SPAIN SA	ATOS	Principal Contractor	ES
5	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Principal Contractor	IT
6	INTRASOFT INTERNATIONAL S.A.	INTRA	Principal Contractor	LU
7	FONDAZIONE BRUNO KESSLER	FBK	Principal Contractor	IT
8	AUDEERING GMBH	AUD	Principal Contractor	DE
9	TAMPERE UNIVERSITY	TAU	Principal Contractor	FI
10	PRIVANOVA SAS	PN	Principal Contractor	FR
11	SPHYNX TECHNOLOGY SOLUTIONS AG	STS	Principal Contractor	CH
12	COMUNE DI TRENTO	MT	Principal Contractor	IT
13	UNIVERZITET U NOVOM SADU FAKULTET TEHNICKIH NAUKA	UNS	Principal Contractor	RS
14	INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP	ITML	Principal Contractor	EL
15	GREENROADS LIMITED	GRN	Principal Contractor	MT
16	ZELUS IKE	ZELUS	Principal Contractor	EL
17	INSTYTUT CHEMII BIOORGANICZNEJ POLSKIEJ AKADEMII NAUK	PSNC	Principal Contractor	PL

Document Revisions & Quality Assurance

Internal Reviewers

1. *Alexandros Iosifidis, (AU)*
2. *Nikola Simic, (UNS)*

Revisions

Version	Date	By	Overview
0.9	30/12/2022	PC, LB	Final version addressing PC's comments
0.8	21/12/2022	Reviewers, partner contributors, LB	Second round of review and final integration of partner contributions
0.7	15/12/2022	Partner contributions, LB	Integration of PN contributions
0.6	14/12/2022	Partner contributions, LB.	Contributions from the partners, integration and form fixing by the LB
0.5	08/12/2022	Reviewers, LB	Comments from the first review, fixing by LB and request for fixing to contributors
0.4	25/11/2022	LB	Merging and harmonisation of the contributions plus minor adjustments from contributors
0.3	16/11/2022	Partners contribution, LB	Integration of the first round of contributions
0.2	20/10/2022	STPM, WPL, LB	ToC updated based on comments from WPL and STMP
0.1	14/10/2022	Alessio Brutti	ToC submitted to WPL and STMP

Disclaimer

The work described in this document has been conducted within the MARVEL project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957337. This document does not reflect the opinion of the European Union, and the European Union is not responsible for any use that might be made of the information contained therein.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AD	Automatic Directory
AI	Artificial Intelligence
AS	Authentication Service
API	Application Programming Interface
ART	Accountability, Responsibility, Transparency
ASR	Automatic Speech Recognition
AudioAnony	Audio anonymisation
CATFlow	Data acquisition Framework
CCTV	Closed-Circuit TV
CIA	Confidentiality, Integrity and Availability
CRNN	Convolutional Recurrent Neural Network
CNN	Convolutional Neural Network
devAIce	audio AI technology integrated in a device
D#.#	Deliverable
DNN	Deep Neural Network
DPO	Data Protection Officer
DynHP	Compressed models
EB	Ethics Board
EC	European Commission
EdgeSec	Security services at the edge
EER	Equal Error Rate
EU	European Union
E2F2C	Edge-to-Fog-to-Cloud
FedL	Framework and implementation of ML algorithms – Federated learning
FID	Frechet Inception Distance
FSGAN	Face Swapping GAN
GA	Grant Agreement
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
GPURegex	GPU Pattern Matching Framework
GUI	Graphical User Interface
HDFS	Hadoop Distributed File System

HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
KDC	Key Distribution Center
ID	Identifier
IoT	Internet of Things
Karvdash	Kubernetes CARV dashboard
LDAP	Lightweight Directory Access Protocol
LPC	Linear Prediction Coding
LSTM	Long Short-Term Memory
M#	Month #
MAE	Mean Absolute Error
MCU	Microcontroller Unit
MEMS	Micro Electro-Mechanical Systems
ML	Machine Learning
MOPS	Million Operations per Second
MVP	Minimum Viable Product
NAT	Network Address Translation
openSMILE	open-source Speech and Music Interpretation by Large-space Extraction
PESQ	Perceptual Evaluation of Speech Quality
PD	Personal Data
PFLD	Practical Face Landmark Detection
RAM	Random Access Memory
RP1	First Reporting Period
RPi	Raspberry Pi
R1	First prototype release
RTSP	Real-Time Streaming Protocol
SED	Sound Event Detection
SED@Edge	Sound Event Detection at the Edge
SGX	Software Guard Extensions
SALS	Simple Authentication Security Level
SLA	Service Level Agreement
SOTA	State-of-the-Art
speakerID	Speaker Identification
SSL	Secure Sockets Layer

STFT	Short-Time Fourier Transform
STOI	Short-Time Objective Intelligibility
SW	Software
TCP	Transmission Control Protocol
TEE	Trusted Execution Environment
TGS	Ticket Grant Server
TGT	Ticket Grant Ticket
TLS	Transport Layer Security
T#.#	Task #.#
UART	Universal Asynchronous Receiver-Transmitter
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VC	Voice Conversion
VideoAnony	Video anonymisation
VPN	Virtual Private Network
WER	Word Error Rate
WP	Work Package

Executive Summary

The goal of this document is to describe the approaches adopted in MARVEL to ensure that the privacy of the citizens is preserved in the Edge-to-Fog-to-Cloud (E2F2C) processing infrastructure, given the guidelines and regulations provided by the EU, the GDPR, and by the pilots' Data Protection Officers (DPOs).

Therefore, **the core of the document** is the description of the video and audio anonymisation components. Thanks to the fact that they remove information related to the citizens' identity (faces, voices, and car plates) as the first processing components of the pipeline, the data can be shared across the different processing tiers and components without (or with limited) threats to citizens' privacy. A reader already familiar with the MARVEL project and its other deliverable can focus on the section related to anonymisation and skip the other parts of the documents.

In addition to these processing stages, the project implements other mechanisms to ensure that data are transferred and processed in the E2F2C infrastructure in a secure way. These include: edge processing and cybersecurity mechanisms such as encryption and authentication. For some components, edge processing solutions are adopted so that data do not need to be transferred through the platform. However, only a small set of services can be deployed in this way due to the limitations of the edge devices and the complexity of AI models. These approaches, although developed in other WPs and described in the related deliverables, are briefly presented here for completeness and for highlighting connections between those approaches to the ones developed in Task 3.1.

Finally, the document also reviews the most relevant privacy issues related to smart cities and to the specific use cases in the MARVEL project.

1 Introduction

In the smart city scenarios addressed in MARVEL, where multiple cameras and microphones are employed to pervasively monitor and understand the city, preserving the privacy of the citizens is of paramount importance towards the development of ethical and responsible AI-based solutions. In particular, since collected data are transferred to the local public authority's data centres and to third parties for audio-visual analytics, it is crucial to remove as much information that can be used to automatically identify the citizens in the recording as possible.

In MARVEL, Task 2.5 and WP9 are in charge of monitoring privacy and ethics aspects and providing guidelines and requirements to ensure that the MARVEL framework complies with the related regulations. These guidelines are reported in D1.2[1]. The goal of T3.1 is to develop and release algorithmic solutions that allow the project to be compliant with the aforementioned guidelines and requirements. In particular, the goal of the task is "to work on methods and algorithms capable to run at the edge and ensure protection of personal data in general audio-visual recordings", as included in the Grant Agreement (GA).

This deliverable describes the algorithmic outcomes of the work conducted under T3.1 for audio and video anonymisation. Considering that the task ends at M24 but efforts towards edge deployment and optimisation will be allocated under T3.5, the document describes also the expected future developments of the algorithms for their deployment in the final prototype.

Finally, since MARVEL implements multiple mechanisms to ensure the preservation of the privacy of the citizens, the document also describes the cybersecurity solutions and the edge processing approaches adopted in the MARVEL framework.

1.1 Contribution to WP3 and project objectives

The contribution of T3.1 to WP3 is described in the GA as: "(i) define and execute measures ensuring privacy preservation during data processing". In terms of overall project goals, the results of T3.1 (described in this document) mainly contribute to the achievement of **Objective O1**, i.e., to "*Leverage innovative technologies for data acquisition, management and distribution to develop a privacy-aware engineering solution for revealing valuable and hidden societal knowledge in a smart city environment"* via the **Enabler E3** "*Privacy preservation and assurance: Deliver privacy assurance mechanisms to ensure privacy preservation at critical data exchange points along the data path*."

In addition, the task contributes also to **Objective O5** "*Foster the European Data Economy vision and create new scientific and business opportunities by offering the MARVEL Data Corpus as a free service and contributing to BDVA standards*" by allowing Enabler **E2 "***Sharing of the dataset"*, which will not be possible without anonymisation. Finally, note that the algorithms developed under T3.1 and described here also contribute to **Objective O4** "*Realise societal opportunities in a smart city environment by validating tools and techniques in real-world settings*" as the realisation of the use cases is possible only if privacy requirements from the pilots' DPOs are satisfied. Overall, the results of this task are of crucial importance for the success of the project.

[1] MARVEL D1.2: MARVEL's experimental protocol, 2020. Confidential.

1.2 Contribution to KPIs

Being mainly related to Objective O1, the approaches described in this deliverable contributes to two project KPIs:

KPI-O1-E3-1: Number of incorporated safety mechanisms (e.g., for privacy, voice anonymization) ≥ 3:

Audio and video anonymisation components are two safety mechanisms incorporated in the MARVEL framework. Other mechanisms related to cybersecurity, authentication in the MARVEL Data Corpus and edge processing have been also developed in WP2 (T2.4), WP3 (T3.3, T3.4, T3.5), and WP5 (T5.3). Therefore, the KPI is achieved.

KPI-O1-E3-3: Video and voice anonymisation expected to improve by at least 10%:

This KPI is interpreted in two different ways for audio and video anonymisation. For video anonymisation, the 10% improvement refers to the computational complexity reduction towards edge deployment. The achievement of this KPI involves efforts under T3.1 and T3.5. The use of neural network parameters quantisation as well as the development of specific solutions for microcontrollers contributed to the completion of this KPI. Nevertheless, further developments are expected under T3.5 in the following period of the project. For audio anonymisation, since very lightweight methods already exist, the 10% improvement refers to the amount of information about the acoustic scene preserved with respect to full removal of speech segments. Thanks to the use of voice conversion strategies instead of the full audio removal, the acoustic information is preserved and hence the KPI is also completed.

By definition, the components described in this document allow the deployment of GDPR-compliant and Trustworthy AI. Therefore, they contribute to the achievement of the **related use case non-functional requirements** (as described in D1.2), i.e., cybersecurity, data protection, Ethical, Trustworthy & Responsible AI (which are used to evaluate the overall trustworthiness of the **MARVEL framework**).

1.3 Relation to other WPs, deliverables and activities

The content of this deliverable and the activities of T3.1 are tightly related to the following WPs:

- WP1: D1.2 reports guidelines for Ethical, Trustworthy & Responsible AI, based on EU regulations, that pose constraints on where and how anonymisation has to be performed.
- WP2: as described later, two out of three pilots require that data are anonymised before being stored in the MARVEL Data Corpus. Note that, in this case, the anonymisation process is not subject to real-time constraints. Details about the guidelines for privacy assurance related to the MARVEL Corpus are reported in D2.1[2] and in D2.2[3].
- WP4: since anonymisation has to be performed as close as possible to the sensors, it is a crucial component for audio-visual data capturing. In particular, a specific audio device was developed in MARVEL, consisting of the IFAG MEMS microphones and a Raspberry Pi, which hosts the audio anonymisation component (jointly developed with VAD), as described in D4.4[4].

[2] MARVEL D2.1: Collection and analysis of experimental data, 2021. https://doi.org/10.5281/zenodo.5052713

[3] MARVEL D2.2: Management and distribution Toolkit – initial version, 2022. https://doi.org/10.5281/zenodo.6821195

[4] MARVEL D4.4: Optimal audio-visual capturing, analysis and voice anonymization, 2022. To appear.

- WP5: the audio and video anonymisation in combination with the other security mechanisms developed in MARVEL are related to the framework design and to the data processing flow defined under WP5.
- WP9: D9.2[5] and D9.4[6] discuss risks related to ethics and privacy and provide GDPR-related guidelines for anonymisation techniques.

1.4 Structure of the Document

Section 2 provides a review of the privacy issues in the MARVEL application scenarios, with the specific problems of each use case. The section also summarises the overall strategy adopted by MARVEL, besides the audio-video anonymisation.

The core of the document where the innovative and novel contribution is reported consists of Section 3 and Section 4. These sections present the audio and video anonymisation approaches with the problem formulation, state-of-the-art (SOTA), and the evolution from what is currently deployed in the R1 prototype (1st Release or R1) and what the consortium will achieve in the final integrated solution. Section 3 presents the approach based on McAdams coefficients deployed in the initial stage of the project, in combination with VAD, as well as the new strategy, based on voice conversion neural models, which will be employed in the final implementation. For the latter, the section provides an experimental analysis that supports the implementation choices. Analogously, Section 4 details the algorithmic solutions adopted for video anonymisation, explaining the first implementation of the component and the developments towards the final deployment. The section presents also preliminary results towards the deployment of VideoAnony on MCU boards. Finally, the section presents visual crowd counting results on anonymised video of the MT uses cases, confirming that the proposed approach is capable of removing citizen's identity while preserving enough content for other components.

Section 5 and Section 6 briefly discuss other approaches which help improving privacy protection, i.e., edge processing and authentication, and security mechanisms for data communication and data storage. Although these solutions are described in detail in the related deliverables of WP2 and WP5, they are also described here in order to provide a global view of all the mechanisms adopted in the MARVEL project to preserve the privacy of the citizens. A reader already familiar with the MARVEL project and its deliverables can avoid Section 5 and Section 6. Finally, Section 7 wraps up the deliverable with conclusions and final remarks.

[5] MARVEL D9.2: POPD - Requirement No. 4, 2021. Confidential.

[6] MARVEL D9.4: OEI - Requirement No. 8. Confidential.

2 Privacy preservation in Smart Cities

Privacy preservation and ethical requirements are two key aspects of any solution aiming at a Trustworthy AI. In the context of the MARVEL project, these are crucial issues as the developed framework processes data collected by multiple surveillance cameras and microphones which record unaware citizens in public spaces.

Therefore, the project includes T2.5 and WP9 dedicated to monitoring and defining guidelines and requirements for the development of Responsible and Trustworthy AI solutions. These guidelines with the related risks are described in D9.5[7] and in D1.2 which defines the experimental protocol. This chapter revises the privacy requirements taking into account the peculiarities of MARVEL as well as the specific issues of each pilot, with restrictions introduced by the respective DPOs, and explains how the privacy mechanisms applied in the framework help delivering a privacy-preserving solution.

2.1 Privacy protection as legal and ethical requirements

Research and innovation projects such as MARVEL produce large sets of data and/or process various data categories, including sensitive data. The data processed during the project realisation might relate to numerous objects as well as people. To satisfy legislative and ethics requirements, the project must be realised in accordance with national and international laws, as well as specific sources of regulation applicable to the project such as the FAIR principles[8] as described in the MARVEL Data Management Plan under D8.2[9], and internal procedures and processes. For this reason, during the realisation of the project, the consortium must ensure that legal and ethical requirements concerning data processing are respected.

One of the greatest regulatory challenges is to satisfy the requirements of the European data protection regulatory framework including the General Data Protection Regulation 2016/67, known as GDPR[10]. This source of regulation enshrines sets of obligations to entities that process personal data, including implementation of appropriate technical and organisational measures that should ensure the protection of personal data against unlawful and/or unauthorised access, modification, or other personal data processing. The GDPR is a technically neutral law and hence it does not explicitly impose what kind of technical and organisational measures should be implemented to secure data (but provides a list of potential solutions). Therefore, organisations that process personal data should assess relevant factors (such as the nature and scope of data, their sensitivity, potential threats, available technologies for protecting data, their cost of implementation etc.).

Considering the nature of the MARVEL project, there is no doubt that the project consortium has to consider many factors to implement appropriate technical and organisational measures to avoid processing personal data within the project.

[7] MARVEL D9.5: M - Requirement No. 9, 2021. Confidential.

[8] European Commission. (2016). H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. European Commission Directorate-General for Research & Innovation. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[9] MARVEL D8.2: MARVEL Data Management Plan, 2021. Confidential.

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance

As the implementation of MARVEL can lead to a set of arising ethical questions, the project has been the subject of an ethics evaluation that has been conducted through the Ethics Appraisal Scheme[11]. The ethics evaluation has imposed the project consortium **to give a particular focus on the implementation of anonymisation/pseudonymisation techniques to decrease the possibility of any harm potentially caused by data processing**. In this way, the evaluation emphasised the importance of protecting personal data and personal privacy.

2.2 MARVEL and anonymisation/pseudonymisation of data

Pseudonymisation and anonymisation are two similar (but not the same) concepts that have been extensively discussed since the GDPR entered into force. Both are used to protect personal data and subsequently personal identities. According to the GDPR Art 4(5), pseudonymisation is the "*processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.*"

Thus, pseudonymisation enables keeping the de-identified data separately from the "additional information" that is needed for identification purposes (once this data is linked with 'pseudonymised' data). Therefore, pseudonymisation permits data handlers to use personal data more freely, without fear of infringing the rights of data subjects. From a legal point of view, a pseudonym is an identifier that is associated with an individual. The benefit in terms of data protection is that the data will be inaccessible to unauthorised users which will effectively lower the risk of a potential data breach and increase the security of personal data.

The recital 26 of the GDPR defines anonymised data as "*data rendered anonymous in such a way that the data subject is not or no longer identifiable.*" This means that anonymisation can be defined as a data processing activity resulting in the permanent removal of personal identifiers, both direct and indirect, that may lead to an individual being identified. Therefore, the main difference between the two privacy-enhancing techniques is that pseudonymisation is reversible while anonymisation is not. Where 'de-identified' or pseudonymised data is in use, there is a residual risk of re-identification. Since a pseudonym still remains considered to be personal data according to the GDPR, the adequate safeguards would have to apply. On the contrary, anonymisation being irreversible due to removal of personal identifiers, the data will not fall within the scope of the GDPR. When data is anonymised, the application of the data protection regulation is not of concern, as it might be in the case of pseudonymised data.

MARVEL's Anonymisation/Pseudonymisation strategy is a highly valuable solution for smart cities. Massive amounts of data generated and collected from connected IoT devices, and the availability of high-performance computing resources may result in several privacy issues. Privacy protection mechanisms such as anonymisation and pseudonymisation are necessary to be applied to ensure GDPR compliance and protect the privacy of data subjects. For this reason, the ethics strategy of the MARVEL project is based on the prevention of direct collection of personal data apart from the personal data that are required to be gathered for administrative purposes (e.g., contact details for dissemination purposes). In case data providers will collect any kind of personal data, the preferred strategy of the consortium is the application of

[11] More about available at https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm

anonymisation techniques to such data by the data providers before making data available to the partners for further processing.

Within MARVEL, audio and video anonymisation techniques are developed aiming to protect the privacy of individuals that are being recorded. There are two different approaches:

- offline anonymisation where audio-video data are anonymised before being shared with the consortium;
- online anonymisation where data processing is performed in the first data elaboration stage, or as close as possible to the edge, output of which does not contain any personal data.

There are different technical considerations when it comes to audio and video anonymisation. In both cases, personal identifiers (e.g., face and voice markers useful for identification) must be removed or processed to make the recorded person(s) unidentifiable.

2.3 Specific privacy issues in MT use cases

The goal of the Municipality of Trento is to improve citizens' well-being and quality of life by making a better and effective use of the multimodal data daily provided by the audio-visual sensors distributed in the city. However, providing even better and customisable services must necessarily be combined with the protection of the data and privacy of everyone.

The protection of data and citizen privacy is one of the cornerstones of the MARVEL framework whose prerogative is to increase urban security through the analysis of anomalous and potentially dangerous situations by notifying local authorities and emergency services of potential anomalous events that may lead to dangerous situations, such as:

- crowded areas;
- criminal/anti-social behaviours;
- monitoring of parking places;
- analysis of a specific area for better urban planning.

The MARVEL framework is not a traditional surveillance technology based on the use of (sometimes) invasive audio and video equipment. The project is intended to be a support to local authorities, not an automatic security system. Multi-modal data are processed in real-time, using machine learning algorithms to identify situations that require further analysis by the public authorities. All these features of the MARVEL framework complied with the principles established by the GDPR 2016/679 and were underlined by the MT DPO in the document for the appointment of FBK as the Data Treatment Responsible. In particular, the restrictions on the audio and video data that are analysed by the system are:

- the data are anonymised at the source, near the devices used to collect such information;
- the raw data are accessible only by FBK via a data processing agreement, using dedicated machines;
- the raw data are not shared with third parties, not even among the consortium partners;
- the data are not used for the profiling of citizens.

Moreover, the MARVEL framework is not able to analyse or understand conversations, but only to associate audio and video with situations considered dangerous or anomalous.

In addition to the data-related restrictions, since the use cases employ the operational surveillance network of MT, further constraints apply to the accessibility of the devices and sensors. As a consequence, edge devices cannot be part of the MARVdash framework and are

only accessible by FBK with specific machines. This restriction required the design of a very specific Fog infrastructure which is detailed in WP5 deliverables. Table 2.1 reports the data sharing limitations and the device accessibility for the MT use cases.

Table 2.1: Data sharing limitation and device availability in the MT use cases

	Consortium	FBK
Raw data	No	IT-managed machines
Anonymised data	Yes	Yes
Edge Devices	No	Yes
Fog Devices	Yes/No	Yes

2.4 Specific privacy issues in GRN use cases

GRN expects to provide stakeholders operating in the transport arena, such as transport authorities, transport consultants, and researchers in transport with new services covering two time spans: short-term and long-term decision-making. For this purpose, GRN is collecting data from various audio-visual, visual and audio devices installed around the island to collect road traffic data. Inevitably the data sources will collect personal data related to the general public that needs to be removed from the data to comply with GDPR legislation.

GRN has identified three sources of personal data that could be present in video and audio data and therefore would need to be removed from the data:

- video data: faces and vehicle number plates are features that would enable individuals to be identified;
- audio data: GRN would need to prevent the identification of any pedestrian or driver from the recordings, and the removal of personal information present in the voice signal.

These functionalities would be in line with the DPO's requests to preserve the privacy of any citizen who is recorded by the MARVEL framework.

GRN devices are all in the Kubernetes cluster and hence they are accessible by all partners. Some partners, such as FBK and FORTH, have been given direct access to the Fog and Edge devices to facilitate the deployment of their components. Raw data is accessible only by the VideoAnony components which obtains the IP address of the raw video streams from an encrypted file. Therefore, none of the partners has direct access to the raw data. Figure 1 shows this data flow. The VideoAnony component can either be executed on the edge or fog layer, depending on the availability of hardware. Table 2.2 reports the data sharing limitations and the availability of the processing devices in the GRN use cases.

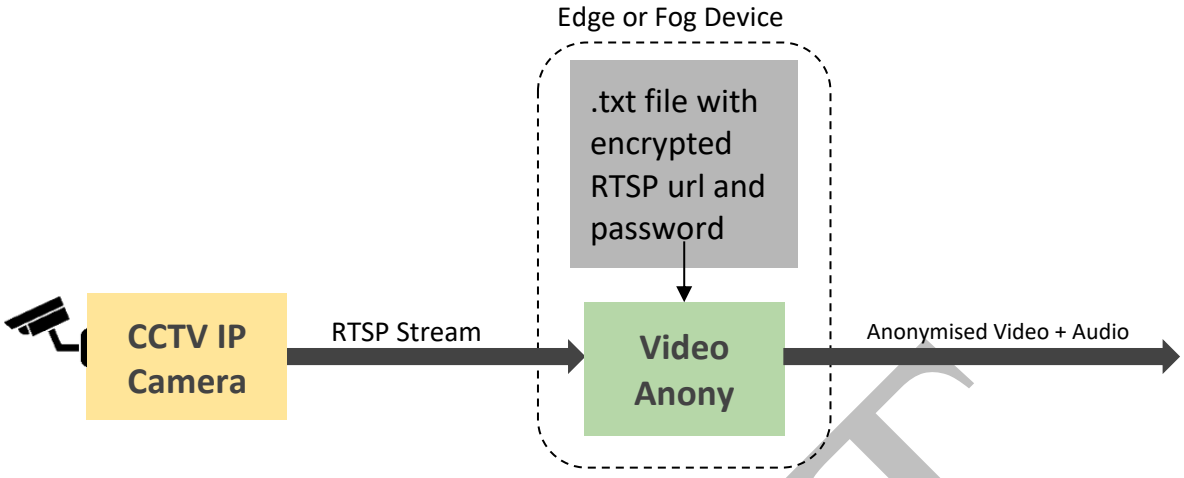

Figure 1: Data flow of raw video data to be anonymised by VideoAnony and shared with the rest of the components

Table 2.2: Data and device availability for GRN use cases

	Consortium	GRN
Raw data	No	Yes
Anonymised Data	Yes	Yes
Edge Devices	Yes	Yes
Fog Devices	Yes	Yes

2.5 Specific privacy issues in UNS use cases

The focus of the UNS1 use case is to evaluate the potential of drones for monitoring large public events. The experiments will be held in a strictly controlled environment within a staged recording process. All participants have to sign a consent form. The risk of tracking pedestrians and vehicles is very low within the experiment. Legal oversight and privacy compliance are ensured by the DPO. As the focus of the UNS1 use case is to evaluate the potential of drones in smart city situational awareness, which implies an analysis of several setups and scene configurations, robustness of the setup is a specific issue that is carefully considered. For this issue, modularity is considered as a specific non-functional variable. Besides that, secure transmission and end-user experience are examined periodically.

UNS2 use case deals with audio-visual emotion recognition, which includes recording of face and speech of participants within a staged recording process. Such a use case has to guarantee the safety of recorded personal data, which represents a specific issue for the use case. Data protection, secure transmission and the support of distributed learning are defined as specific non-functional variables to overcome this issue. Modularity and end-user experience conclude the list of non-functional variables for the UNS2 use case.

In general, given the nature of the UNS use cases which are all based on staged recordings, whose participants gave their consents to audio-visual recordings, privacy regulations do not directly apply and no specific issues are present. However, the use case implementations

account for restrictions that would apply in real deployments and, therefore, apply audio and video anonymisation on edge devices. Nevertheless, data are accessible by all partners and all devices are accessible via MARVdash.

2.6 Solutions deployed in MARVEL E2F2C infrastructure

To preserve confidentiality, integrity, and availability of data (CIA triad) that is the essence of information security, as well as to properly enforce information security management, MARVEL applies multiple security measures. They serve to protect sensitive information that could be found, namely, at rest, in use, and in transit, based on GDPR requirements as well as those specifically required by the pilot's DPOs to be met. In particular, three different strategies are applied as shown in Figure 2.

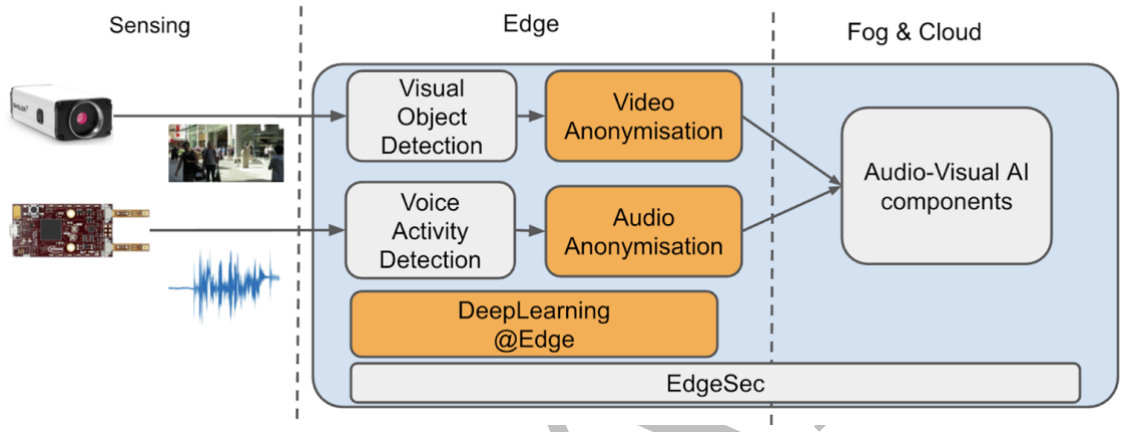

Figure 2: Security mechanisms in MARVEL

Audio and video anonymisation by applying face blurring/swapping and voice conversion as the first processing stage of any use case pipeline or prior to sharing the recorded data within the MARVEL Data Corpus. These solutions are described in detail in the following sections. The use of this type of anonymisation approaches allows deploying the following AI processing stages in a transparent manner and prevents both automatic, as well as human, re-identification of the targets. Note that pseudoanonymisation is always applied, as no citizen identity labels are associated with the data.

Edge (or Fog) processing is privacy-preserving by design as data carrying sensible information never leaves the proximity of the sensor that collected them. This drastically reduces the risks related to leakage of personal information or to illegal or inappropriate use of the data. Nevertheless, this solution only partially addresses the privacy preservation issue as edge AI components still get access to personal information and not all components are suitable for this type of processing.

Cybersecurity and authentication mechanisms are finally implemented to ensure the security of the processing components and the data travelling in the E2F2C continuum, and the protection of the data stored in the MARVEL Data Corpus. Data processing in environments that do not provide a high-level of security (e.g., unencrypted networks) might endanger the confidentiality, integrity, and availability of data. This could generate higher risks when sensitive data are processed. Therefore, within MARVEL, sensitive data are encrypted before being transmitted over the network and remain in this form until arrival at the receiving node. Standard cryptographic protocols like Secure Sockets Layer (SSL) and/or Transport Layer Security (TLS) will be used to provide endpoint encryption, and a hash algorithm will be used

to add an additional layer of security. Also, MARVEL will leverage the security features of Trusted Execution Environments (TEE) and apply specific measures to protect the CIA of data at rest.

3 Audio Anonymisation

This section focuses on audio anonymisation, presenting the problem from a scientific and practical point of view, describing the solutions deployed in the first integrated prototype of the MARVEL framework, and presenting the scientific advancements towards the implementation of an advanced component in the final prototype. Note that the goal of the audio anonymisation component is not only to remove the speaker's identity from the audio signals but also to preserve as much of the surrounding context as possible for further analysis.

3.1 Problem formulation

The recording and analysis of voices of citizens raise privacy concerns related to the risk of voice cloning, theft of personal information, frauds or misuse by public authority. One popular approach to mitigate these issues is to develop privacy by design solutions which perform the data processing on the edge devices, close to where the data are generated. This solution avoids transferring or storing sensitive information on cloud infrastructures. However, the limited computational resources of these nodes typically require the compression of AI models in order to reduce their memory and computation requirements. An alternative solution, that does not require any modification of the audio processing algorithms, is voice anonymisation. This solution is a special case of voice conversion (VC): the goal is to remove speaker information from a speech utterance while leaving unaltered the other acoustic attributes. In this way, the recordings are safely stored since the identity of the involved speakers cannot be recovered while the content is preserved for future analysis.

3.2 State-of-the-Art

Besides pure signal processing approaches (Patino et al., 2021), where specific filters transform pitch and formants of the original waveform, recent techniques for VC operate by disentangling the speaker information from the associated acoustic features (Qian et al., 2022) and synthesising similar speech based on the characteristics of other speakers. Many systems synthesise speech with good quality, as long as both the source and target voice data are available in the training data. On the other hand, handling the condition of a target voice unseen during training (any-to-any voice conversion) is still an open problem. Zero-shot voice conversion is one of the challenges, and most VC systems are not robust enough for addressing it. Generative adversarial networks (GAN) and auto-encoders are the common techniques used in VC. StarGAN is a GAN-based VC framework proposed by (Wang et al., 2020), while (Qian et al., 2019) and (Chen et al., 2021) are auto-encoder-based systems where the model is trained to reconstruct the acoustic feature from the encoded representations in a target-speaker-dependent manner. Other recent approaches combine bottleneck features with a sequence-to-sequence based synthesis module. In the conversion phase, given the speech and speaker representations, the model generates the converted acoustic features, which are then sent to a neural vocoder, e.g., HiFi-GAN (Kong et al., 2020), to synthesise the new waveform.

3.3 Solution deployed in MARVEL first prototype

Although originally VAD and AudioAnony were handled as two separate components, during the course of the project the consortium decided to introduce a single joint component to be deployed on the edge devices connected to the microphones. The decision was based on the fact that the two components are tightly correlated as audio anonymisation is applied on speech

segments only and VAD is itself a privacy preservation method. More details about the practical implementations are available in D5.4[12].

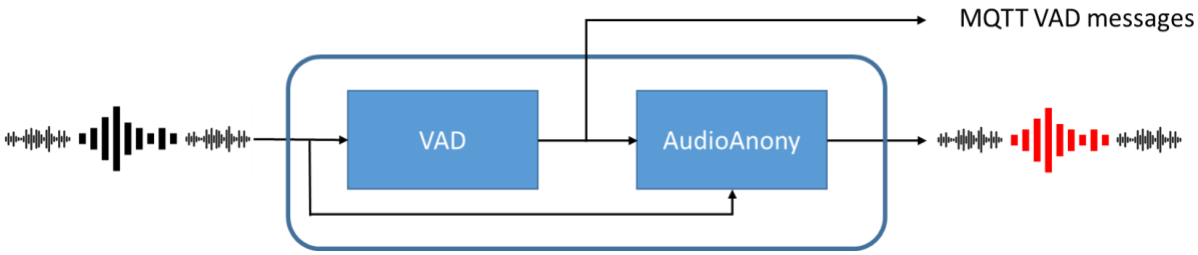

Figure 3: Block diagram of the joint VAD+AudioAnony component

3.3.1 Voice Activity Detection

The goal of the voice activity detection component is to identify audio segments containing human speech. It represents the first layer in the audio anonymisation pipeline after the input devices (MEMS microphones), as shown in Figure 3. It serves as a pre-processing step where speech segments will be detected first and anonymised later on by AudioAnony, all this happening in the order of milliseconds on the deployed edge devices (Raspberry Pis or Intel NUC). The identified speech boundaries are not only used by AudioAnony but they are also made available to other processing components and to the Decision-Making Toolkit (DMT) component for further processing. Figure 4 shows an overview of the audio analytics pipeline.

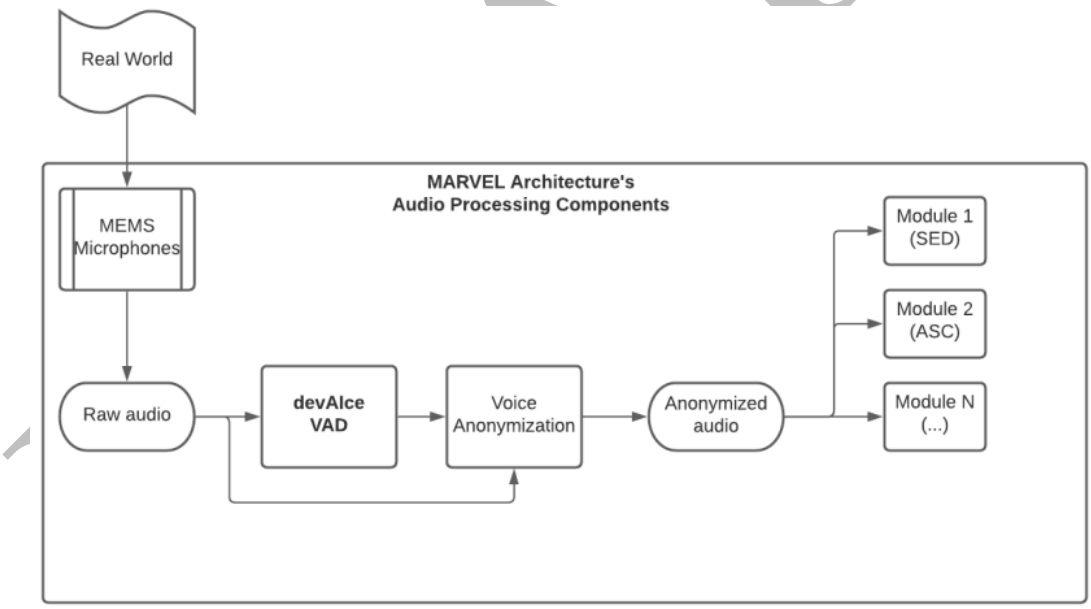

Figure 4: Overview of the audio analytics pipeline

To operate and be deployable on edge devices, the VAD module has been developed by adapting a lightweight Long-Short-Term-Memory (LSTM) based architecture. The upgraded VAD module also adopts the same architecture, by adding on top of it a CNN-based attention module on both time and frequency domains, leading to a novel SOTA approach. This attention module improves noise robustness and enhances the VAD performance in unrestricted (in the

[12] MARVEL D5.4: MARVEL Integrated framework – initial version, 2022. Confidential.

wild) experimental scenarios and is illustrated in Figure 5, where H is the hidden state of dimension T (sequence length) and D (number of hidden features). Dual attention is the combination of both attention modules.

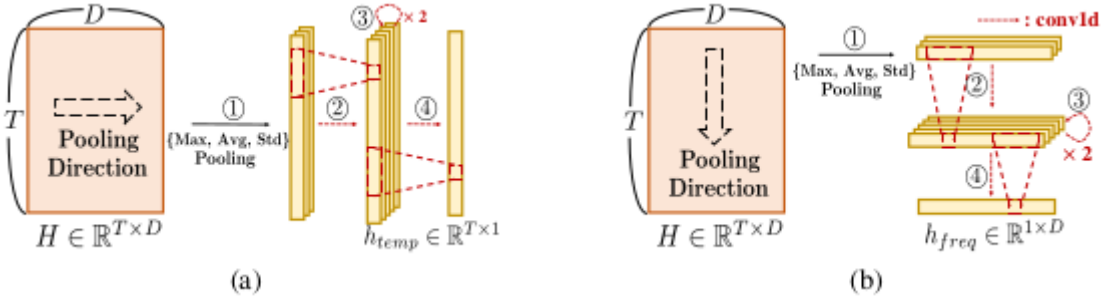

Figure 5: (a) Temporal attention (b) Frequential attention

The VAD module operates on the frame-level. However, to produce the boundaries of speech segments on which the anonymisation module will operate, an algorithmic solution has been applied as a post-processing step. This solution aggregates the activated frames according to a certain sensitivity level which can be adjusted, producing in the end the relevant speech segments.

To ensure maximum efficiency during deployment, the VAD module is part of the devAIce platform, which is AUD's modular technology. Being part of devAIce, the VAD module can be installed as a Python solution and be used within the same code as AudioAnony. This approach guarantees a low latency between both components by avoiding extra network delays, such as, for example, time establishing connection between different containers and exchanging information.

3.3.2 AudioAnony with McAdams Coefficients

The anonymisation approach currently deployed in the first release of the MARVEL integrated prototype implements the method based on McAdams coefficient introduced by (Patino et al., 2021). This purely signal-processing approach has been used as one of the baselines of the 1st Voice Privacy Challenge[13]. The implementation deployed in MARVEL is derived from the official repository[14].

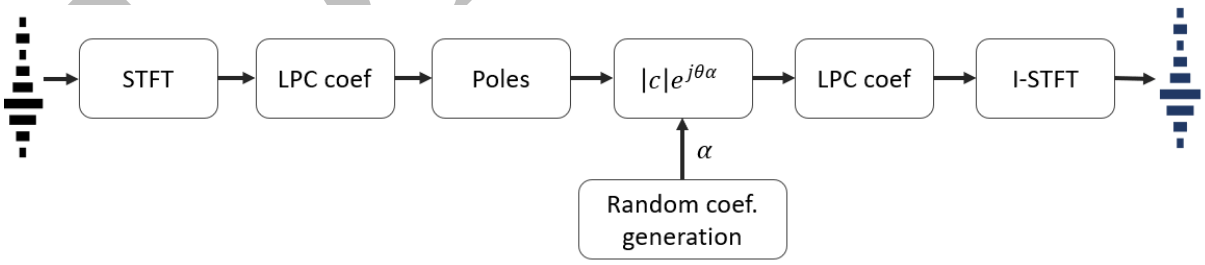

Figure 6: Processing pipeline of the approach based on McAdams coefficients

Figure 6 shows the block diagram of the processing pipeline. Once the linear predictive coding (LPC) coefficients have been derived from the short-time Fourier transform (STFT) of the

[13] https://www.voiceprivacychallenge.org/

[14] https://github.com/Voice-Privacy-Challenge/Voice-Privacy-Challenge-2020

signal, the phases of the complex positive poles are shifted by using a random coefficient. The new LPC coefficients are used to reconstruct the signal. This transformation leads to shifting the formants of the speech waveform toward low frequencies when the coefficient is smaller than 1 and toward higher frequencies when the coefficient is above 1, as shown in the equation below:

$$\hat{c} = |c|e^{j\theta\alpha},$$

where c is the generic pole, θ is the phase of the pole and α is the conversion coefficient.

As described in D4.4, the audio anonymisation component is deployed in combination with VAD, therefore the processing described above is applied only to audio segments including some speech content. In all the other cases the audio signal is not processed.

3.3.3 Role in RP1 and associated use cases

The role of this component is to remove speaker identities from audio streams collected by outdoor microphones. The component is deployed in two use cases of the R1 prototype: MT3 – Monitoring of Parking Places and UNS1 – Drone Experiment. As mentioned in Section 2, in the latter case, audio anonymisation is not necessary as data includes only persons who provided their consent to the recordings. Nevertheless, the component is deployed to simulate a realistic application scenario. In both cases, the component includes also VAD and it is deployed on the edge devices. In the UNS use case, the component is deployed via MARVdash using a proper dockerisation. Multiple instances exist that can run on the Cloud, Linux Workstations, Intel NUC, and Raspberry Pi. In this way, the deployment of the component can be changed depending on the implementation needs. For what concerns the MT use case, since the devices are not in the Kubernetes cluster, the component is manually installed on the Raspberry Pi.

3.4 Current anonymisation approach and future plans towards R2 prototype

The voice anonymisation technique explored in MARVEL follows the voice conversion paradigm and is based on the model presented by (Liu et al., 2021) that learns to map the speech content, expressed as bottleneck features related to an ASR task, constrained by the target speaker characteristics (computed as an x-vector). An attention mechanism induces the correct alignment to the target mel-spectrogram and, finally, a Generative Adversarial Network (GAN)-based vocoder generates the converted waveform.

We modified the original architecture by replacing the Automatic Speech Recognition (ASR) bottleneck and the x-vector speaker representation with pre-trained speech and speaker representations, namely WavLM (Chen et al., 2022) and TitaNet (Koluguri et al., 2022). Pre-trained speech representations have shown remarkable performance for a variety of speech processing tasks (e.g., speech recognition and separation, speaker diarization and verification). Very recently, the anonymisation task has also been investigated with respect to semi-supervised learning models (Miao et al., 2022). In particular, for the speech content representation, we performed an experimental comparative study considering a variety of pre-trained models as well as different variations of WavLM.

In addition, we also investigated different strategies for the selection of the target speaker representation to increase the de-identification capabilities of the model. As a result, the direct usage of these embeddings can avoid training ad-hoc features. This allows an effective selection and suitable combination of pre-trained representations, taking into account the ultimate goal

of audio anonymisation where speech content should be preserved and speaker identity instead removed.

The approach is built upon the Phonetic PosteriorGram-Based Voice Conversion (PPG-VC) model (Liu et al., 2021), in which the content extractor is based on popular pre-trained representations inserted in a sequence-to-sequence based module that can be regarded as an encoder-decoder model. The encoder combines the content features with pitch and voice/unvoiced information along with the speaker representation through the TitaNet model. A multi-speaker location-relative attention-based synthesis model is then trained to reconstruct spectral features from these features, conditioning on speaker representations for speaker identity control in the generated speech. The decoder of the synthesis model adopts an auto-regressive network structure. Moreover, a mixture of logistics attention mechanisms guarantees a monotonic alignment process. In Figure 7, the training and generation phases are illustrated. During training, the incoming signal is decomposed into a content representation while the speaker information is encoded using the TitaNet model. At the generation phase, this information is substituted by another speaker (encoded by TitaNet). In this way, the model generates a new waveform inducing a voice conversion from the source speaker to the target one.

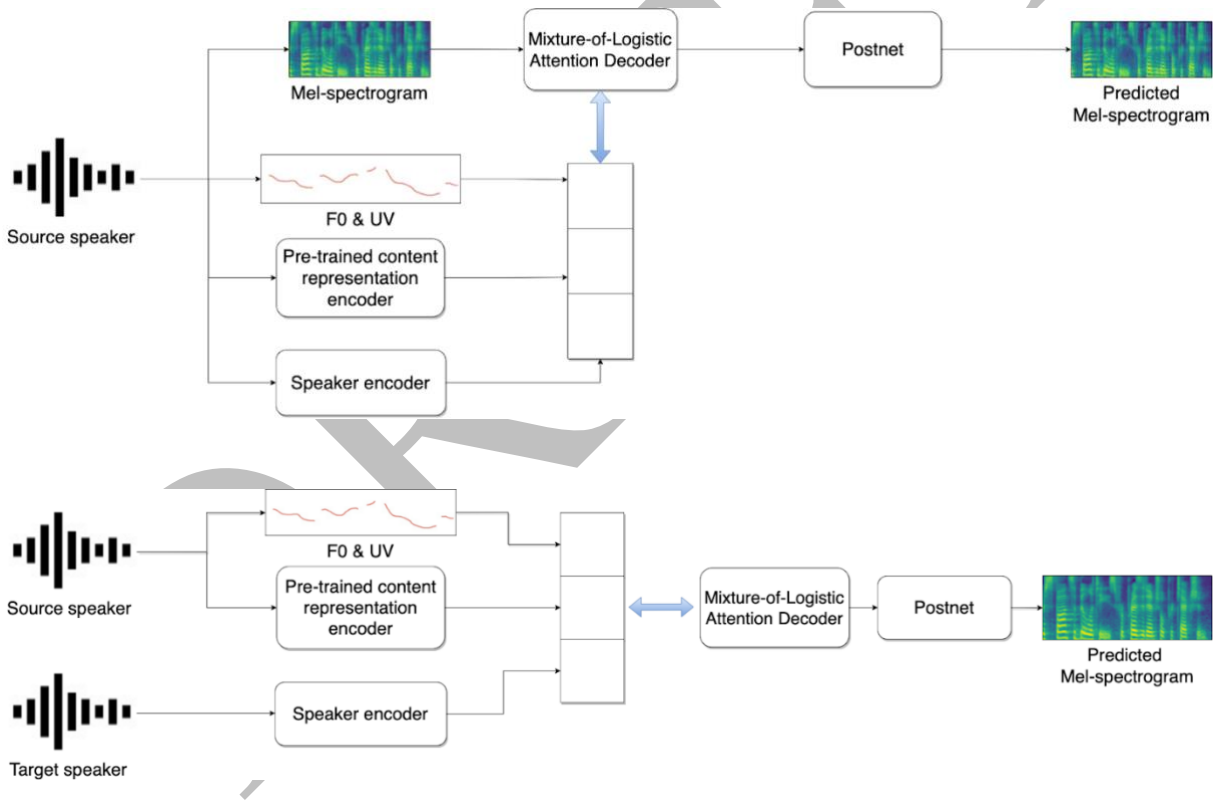

Figure 7: Audio anonymisation based on voice conversion: training and generation phases

Experiments on the LibriSpeech dataset (Panayotov et al., 2015) (see Table 3.1) show the effectiveness of the proposed method, obtaining great flexibility in the selection of the target speaker and allowing to select (even randomly) the characteristic of the generated voice since virtually any potential speaker can be represented via embeddings, both in terms of identity and content. The table reports performance in terms of:

- Equal Error Rate (EER), which measures the speaker identification error given two speech utterances, the higher the better. The performance is measured using the Nvidia speaker recognition baseline using TitaNet.
- Perceptual Evaluation of Speech Quality (PESQ) and Short-time objective intelligibility (STOI), which measure the signal intelligibility and distortion. For both metrics, a higher value indicates better performance.
- Word Error Rate (WER) on the converted signals using the Kaldi Librispeech baseline[15].

Table 3.1: Results using various Voice Conversion techniques on Librispeech test-clean set; metrics considered are EER, PESQ, STOI, and WER

Method	EER (%) ↑	PESQ ↑	STOI ↑	WER (%) ↓
Original signal	1.2	-	-	5.29
McAdams	5.6	2.7	0.96	11.74
AGAIN-VC (Chen et al., 2021)	44.4	1.07	0.28	44.37
PPG-VC (Liu et al., 2021)	46.1	1.09	0.16	12.93
WavLM+TitaNet	33.0	1.11	0.74	7.50

The proposed approach is compared against the McAdams coefficient currently deployed in the first integrated prototype, a popular many-to-many voice conversion method (AGAIN-VC) and the baseline PPG-VC from which our proposed approach is derived. Note that AGAIN-VC gives very high EER but at the cost of detrimental speech distortions. On the other hand, the proposed approach, although it still carries some information about the speaker, considerably improves the signal quality with only minor degradation in terms of WER with respect to the original signals.

Table 3.2: Results using different WavLM representations, varying the dimension of the used hidden state

WavLM Models	EER (%) ↑	PESQ ↑	STOI ↑	WER (%) ↓
WavLM (HS 768)	33.0	1.11	0.74	7.50
WavLM (HS 128)	33.3	1.10	0.74	7.07
WavLM large (HS 1012)	18.8	1.17	0.80	6.70

Other advantages of the approach include its scalable architecture, e.g., for distributing the computational pipeline (edge/fog/cloud), the possibility of controlling the "distance" between the source-target speaker and the adoption of speaker-customised synthesisers. Indeed, Figure 8 and Table 3.2 show that different pairs of WavLM representations and target selection strategies control the resulting performance. As a result, according to a specific requirement in terms of content preservation (WER) or speaker anonymisation (EER), the appropriate model can be selected. Ideally, the pairs on the right lower quadrant provide the best combination.

[15] https://kaldi-asr.org/

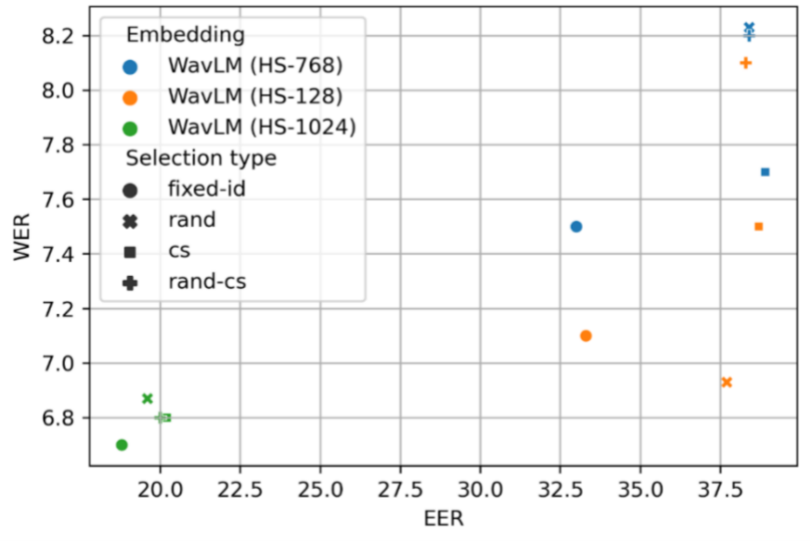

Figure 8: Anonymisation performance expressed as a combination of content preservation and speaker identity confusability: models and target selection strategies can control the effectiveness of the resulting anonymisation

The contribution of the AudioAnony component to the project and use case KPIs have been already discussed in Section 1.2. For what concerns the asset-specific KPIs, the following are defined in D1.2 for AudioAnony:

- At least 50% EER (contribute to **KPI-O1-E3-1)**
- 20% PESQ improvement over baseline (contribute to **KPI-O1-E3-1)**
- 20% SED improvement over baseline (contribute to **KPI-O1-E3-1)**
- 10% complexity reduction (contribute to **KPI-O1-E3-3)**

As can be observed by the results reported above, the component is still short in terms EER, while the signal quality considerably increases with respect to other voice conversion approaches. For what concerns PESQ, since it is based on similarity with respect to the original signal, improving it while changing the voice is not possible. Therefore, we will reformulate this KPI in terms of WER or STOI.

Finally, in the remainder of the project, we will work in collaboration with T3.5 on increasing the efficacy of the AudioAnony algorithm to ease its deployment on low-end edge devices and its combination with VAD under WP5. Therefore, the latter KPI will be addressed at a later stage. Note that the complexity reduction will be achieved not only by working directly on the conversion module but also on the auxiliary models (for speaker and speech embeddings).

4 Video Anonymisation

This section focuses on the video anonymisation task, from a scientific and practical point of view, and describes the solutions deployed in the first integrated prototype. The section illustrates the scientific advancements obtained under T3.1 which will lead to the implementation of an advanced component in the final prototype.

4.1 Problem formulation

Visual privacy preservation is mostly achieved via video redaction methods by obfuscating the personally data (PD) of a subject whose, in the context of visual data analytics in a smart city environment, face and car number plate are among the most identity-informative parts. Classic face anonymisation techniques, e.g., blurring (Du et al., 2019) or pixelation (Gerstner et al., 2013), can effectively remove PD. However, this comes at the high cost of degrading other vision-related tasks, particularly for action/emotion recognition in which poses play an essential role. Nevertheless, not all visual attributes serve for the identification task, i.e., to answer the question *which person (whose face) is this*? Indeed, some visual attributes are common for defining *what a person (face) is*, which are non-identifiable and essential for the detection task.

4.2 State-of-the-Art

We discuss recent face obfuscation techniques for anonymising visual data.

Visual Anonymisation often refers to irreversible obfuscation for removing PD of the data subject in visual content, also known as de-identification in some works (Gafni et al., 2019). Many works anonymise visual data by obfuscating the faces, the most privacy-concerning content, using classic techniques such as blurring via filters (Du et al., 2019), pixelation by enlarging the pixels (Gerstner et al., 2013), mosaicing by merging small blocks of pixels from different regions or masking by simply replacing the visual information with random data (Asghar et al., 2019). Recently, with the progress in GANs, face anonymisation techniques have advanced by generating realistic faces of a different identity, leaving intact most of the non-identifiable visual and geometrical attributes (Hukkelås et al., 2019; Gafni et al., 2019; Sun et al., 2018; Sun et al., 2018b; Maximov et al., 2020). Sun et al., 2018, proposed a two-step face inpainting technique for anonymisation by first detecting the 68 facial landmarks, and then synthesising the faces guided by the landmarks. With a blurred face as condition, the generated faces have a rather high visual quality, yet resemble the original face. DeepPrivacy (Hukkelås et al., 2019) exploits the generator of StyleGAN (Karras et al., 2019) to generate a face of a fake identity, conditioned both on the context image with the face masked out and on seven facial landmarks. The generated faces are limited in anonymisation and pose preservation. Conditional GANs have also been proposed to explicitly control the identity of the generated faces (Sun et al., 2018; Maximov et al., 2020). Recently, CIAGAN (Maximov et al., 2020) has introduced an identity discriminator to enforce the generated faces to be different from the source image, achieving a better anonymisation performance, while the visual naturalness and pose preservation are not yet satisfactory although the subset of landmarks is carefully chosen. Moreover, CIAGAN cannot be easily applied to unknown condition identity, as the condition identities are encoded within the network during the training.

4.3 Solution deployed in RP1 prototype

4.3.1 Overview

The goal of VideoAnony is to anonymise faces and car number plates detected in the raw video streams coming from the CCTV cameras of each pilot site. The anonymisation solution deployed in the R1 prototype is performed by means of blurring methods and will be extended to more advanced GAN-based face-swapping techniques in the R2 prototype. Specifically, the component receives the incoming raw video stream via RTSP, processes it with the face and licence plate detection and anonymisation modules, and finally transmits the anonymised videos via a customised RTSP server. The component has been successfully implemented in both fog machines and edge devices with sufficient processing power and enables real-time processing.

4.3.2 Internal Operation & Technologies

With the use of the OpenCV library, the current deployable version of VideoAnony can read and anonymise several RTSP raw streams simultaneously, instantiating as many threads in charge of reading the raw video stream as CCTV cameras. VideoAnony employs the YOLOv5 detector for face and car plate detection, which is finetuned using a related public dataset and pilot-provided annotations. After the detection of the regions of interest, the component blurs them by applying a Gaussian filter, and finally replaces the original faces and number plates with the blurred ones in the original image. The component finally transmits the anonymised videos to an RTSP server, from which subsequent AI models read the anonymised stream, using GStreamer.

4.3.3 Role in R1 and associated use cases

The role of this component is to anonymise video streams from pilot sites that will expose car plates and faces of citizens. The component was used in the Minimum Viable Prototype (MVP) delivered at M12 to provide anonymised videos into the Data Corpus in an offline manner and was integrated into the MARVEL platform in RP1 in all pilot cases where visual anonymisation was required. In particular, the component was deployed on devices acting as a fog layer (e.g., use cases in MT) or as an edge layer (e.g., use cases in GRN and UNS).

4.4 Current anonymisation approach and future plans towards R2 prototype

Motivated by the need to maintain various facial attributes (such as pose and expression) so as not to degrade subsequent video-related tasks, we are also developing a lighter version of a SOTA GAN-based face-swapping model (Chen et al., 2020) on top of the current deployable version of VideoAnony. Since GAN-based solutions are notoriously computationally demanding, we are currently focusing on reducing the computational complexity of the model. The ongoing efforts are mainly guided by a couple of directions, i.e., replacement of convolutional blocks with more efficient depth-wise separable convolutions (Howard et al., 2017) and quantisation of the trained model from 32-bit floating point numbers to 8-bit signed integers. Preliminary results showed that, after replacing the convolutional layers, we were able to achieve a 37% reduction in the number of parameters and an 88% reduction in the number of multiplication-addition operations, while the quantisation of the model allowed us to achieve a 73% reduction in model size. This model reduction allows achieving the specific asset KPI as well as the related project KPIs (see Sec. 1 for more details).

Concerning the other asset KPIs, namely: image naturalness, face detection rate and face de-identification, results are reported in (Dall'Asen et al., 2021) for the uncompressed model. The evaluation of the KPIs on the compressed model will be carried out as part of the benchmarking activity.

4.4.1 MCU implementation

In parallel to the work described above, we explored at research level the possibility to perform face-swapping on a microcontroller, therefore with more stringent requirements towards lowering the complexity of the neural network. We based our approach on PhiNets (Paissan et al., 2022) applied to GAN for face-swapping. In particular, we targeted a resource-constrained platform based on a low-power microcontroller, a Kendryte K210, with a RISC V dual core working at 400MHz and overall consuming less than 300mW, achieving processing speeds over 15fps with an FID score under 150.

The processing pipeline is made of 4 steps (see Figure 9), each based on the PhiNets neural architecture combined with a SOTA solution for each task. In particular, first, we perform face detection using PhiNets with a YOLO-V5 detection head, which can generate a very lightweight CNN capable of locating the target faces to swap. Secondly, Landmark detection is obtained using PhiNets with a PFLD (Practical Face Landmark Detection) (Guo et al., 2019) detection head. The same is applied for the third step, i.e., image alignment, using the detected landmarks. Finally, face generation and face swapping are performed, where PhiNets are used in a GAN configuration to generate an artificial face reproducing pose, expression, lighting and colour of the original face. This approach enables high-fidelity face generation in real-time on resource-constrained devices.

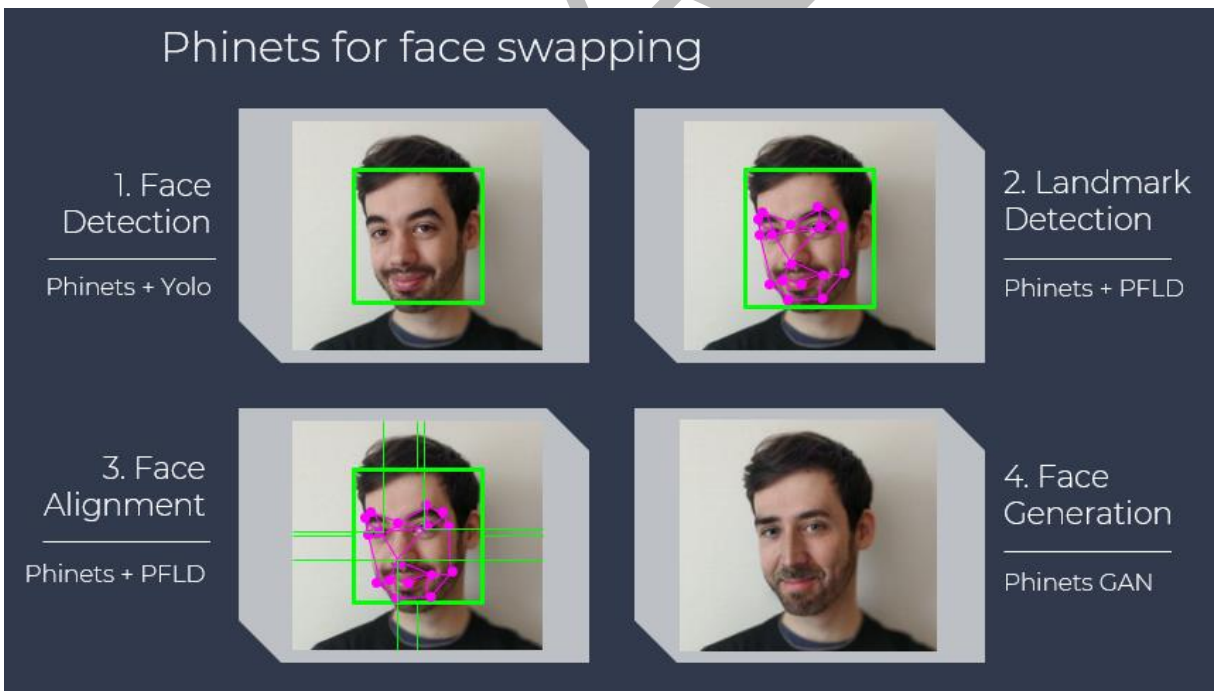

Figure 9: Processing steps of the MCU-based face-swapping component

This last step is based on FSGAN (Face Swapping GAN) (Nirkin et al., 2019) which was one of the first GANs proposed in the literature but it is limited to the generation of only one target identity. We replaced the encoder and decoder networks of FSGAN with Phinets and modified

the approach to be fully convolutional for better efficiency on microcontroller runtimes since the fully connected layer present in FSGAN is inefficient.

We could fit the entire pipeline from face detection to face swapping into our target platform, the K210, with an overall 46mJ/frame and 6fps@280mW. Still, the solution is in prototype form and we are working to improve these results.

4.5 AI-based vision on anonymised streams

One of the crucial requirements of VideoAnony is to obfuscate person faces and car plates while preserving the visual content, so that other processing components can still achieve their tasks. To validate this functionality, AU has trained the Visual Crowd Counting components on the anonymised data from the MT1 use case (Monitoring of Crowded Areas). The frames collected in that dataset come from 6 different scenes, with a wide variety in the number of people present, camera angles, and weather conditions. The dataset contains 710 images in total. 426 have been assigned to the training set, 142 were used for validation, and 142 for testing. Figure 10 presents examples of anonymised test data.

Figure 10: Piazza Fiera – 3 and Piazza Fiera – 4 - anonymised test images

The prototype model achieved a satisfactory performance of 27.55 MAE (Mean Absolute Error) when trained on the anonymised data. The impact of the video anonymisation on the quality of the dataset, however, cannot be measured as the original data, before anonymisation, were not available. The model relies on detecting heads in the image to determine the number of people present, and by visually inspecting the test images it would seem that anonymisation did not impact its ability to do it adversely. The resulting heatmaps corresponding to the images shown in Figure 10 are shown in Figure 11.

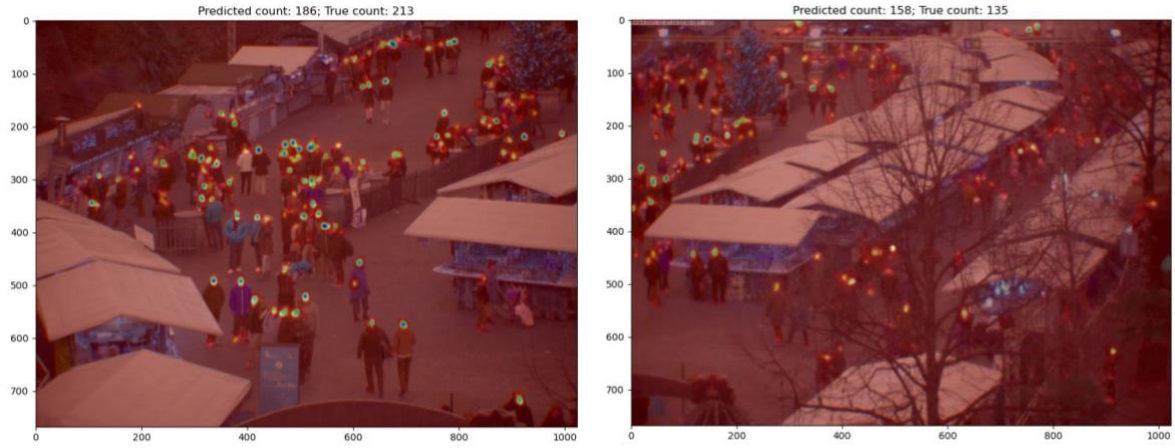

Figure 11: Output heatmaps of Piazza Fiera – 3 and Piazza Fiera – 4

5 Edge-Fog Processing

Besides anonymising audio and visual streams, a relevant privacy preserving mechanism in MARVEL is the development of algorithms which are deployable at the edge and fog tiers, therefore limiting the data transmission (with related security risks) and avoiding transferring sensitive information to third party computational platforms. As mentioned in the Executive Summary and in Section 1, these algorithms are developed under other tasks than T3.1. The goal of this section is to give to a reader not familiar with the MARVEL project a view of what has been implemented in terms of privacy preservation via edge processing.

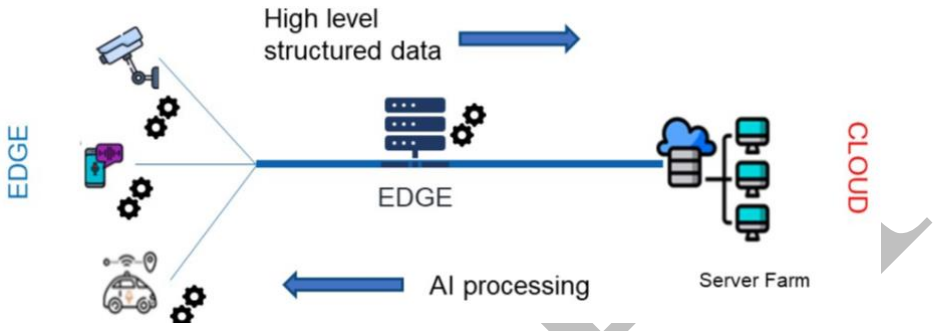

Figure 12: Edge-Fog processing: AI algorithms are deployed on the edge devices or on the fog, while only structured information is transferred to the cloud

This section presents an overview of the SED, SED@Edge, and CATFlow AI-based methods that have been adapted or modified in order to be deployable on edge and fog devices, typically characterised by limited resources. In addition, the section describes DynHP, a methodology developed under T3.5 for unstructured pruning, and FedL, a technique for federated learning developed under T3.4.

DynHP is a methodology developed for training and compressing, at the same time, a Deep Neural Network (DNN) model. DynHP has been designed for being executed (at least in principle) by resource-limited edge devices. During training, DynHP incrementally prunes the parameters considered "less useful" for the learning task at hand. The pruning adopted in DynHP is structured. Such a design choice is guided by the fact that structured pruning is more effective for memory reduction and computational and communication efficiency. The alternative pruning approach not considered in DynHP is termed "unstructured", which falls in the category of sparsification techniques, i.e., these are beneficial mainly for the efficient communication of models over the network. Due to its edge-oriented design choices, DynHP can have an impact also in terms of privacy preservation.

Since DynHP methodology is designed to run effectively in a fixed memory budget, it is possible to execute it directly on the edge device without the need to transfer the data to centralised and remote locations. Note that in this configuration, the amount of data available for training might reduce significantly, possibly affecting the effectiveness of the final compressed models. Therefore, to overcome this issue, DynHP has been extended to work in federated settings (activity performed in T3.5 and initially reported in D3.4[16]). Although the results are still preliminary, the feasibility study conducted on a limited number of nodes (2)

[16] MARVEL D3.4: MARVEL's federated learning realization, 2022. To appear.

shows that distributed compression is viable in federated settings (which is, by definition, a framework for training ML models in a privacy-preserving fashion).

FedL is a component developed to enable federated learning of deep learning models. In federated learning, data are not exchanged among clients, which became an important demand in recent years due to user-privacy laws. Instead of aggregating data at a centralised server, which is a typical procedure for training models, federated learning leaves data on distributed devices, whereas only shared models are averaged on the server. Within the MARVEL project, FedL develops a customised Federated Learning strategy, which optimises the learning process for flaky client-server communication. The strategy allows clients to be temporarily unavailable during training and to share updates only if the client data is valuable to global model, based on several metrics including number of client data points, model metrics such as accuracy, model gradient variance, client availability history and others.

CATFlow is an object detector and tracker subsystem with a number of configurable parameters. In particular, the number of frames processed per second and the detector size can be specified in the settings file. This configurability lends itself useful when dealing with real-time processing requirements in the presence of limited resources at the edge of the fog. With appropriate settings, the CATflow component was able to process two video streams in real-time on the Fog machine and one real-time stream on the Edge device.

The development of the **Sound Event Detection (SED)** component by TAU is mainly focusing on deployment at the cloud level of infrastructure. To enable the scaling into a large number of monitored audio streams at the cloud level, the focus is on solutions requiring limited computational resources. This approach also enables transparent deployment of the component to all levels of the MARVEL infrastructure. Most of the SOTA pre-trained audio embeddings are computationally heavy and not well-suited for the previously described setup. Thus, development focuses on methods that are learned fully on the use case-specific data to have full control of neural network architecture. For the GRN4 use case (Junction Traffic Trajectory Collection), vehicle type detection task, two SED approaches were studied: one applying sound event detection on consecutive non-overlapping 10-second segments, and one applying audio tagging inside consecutive non-overlapping one-second segments. For the sound event detection-based approach, CRNN neural network architecture containing convolutional and recurrent layers was introduced in (Cakir et al., 2017). In this architecture, convolutional layers act as feature extraction layers and recurrent layers learn the temporal dependencies in the sequence of features produced by previous layers. A typical CRNN neural model has 3.5M parameters, and parameters of the convolutional layers (e.g., number of channels) and weight matrices of recurrent layers contribute directly to this. A computationally lighter CRNN model was produced by replacing convolutional neural network blocks with depth-wise separable convolutions and recurrent neural network blocks with dilated convolutions (Drossos et al., 2020). The resulting neural network model (CRNN-DESED) has 290K parameters. In the evaluations with the MAVD-traffic dataset (Zinemanas et al., 2019), these two CRNN approaches produced comparable performance: 44.1kHz signals CRNN had 40.3 macro-averaged F1-score and CRNN-DESED 39.6. The audio tagging-based approach used simpler architecture based on only convolutional layers and fully connected layers. Two CNN approaches based on similar neural network architectures used in PANNs audio embeddings (Kong et al., 2020) were evaluated: six-layer CNN (CNN6) based on AlexNet with 4.57M parameters, and 10-layer CNN (CNN10) based on VGG-like CNNs with 4.95M parameters. In the evaluations, the audio tagging-based approaches outperformed sound event detection-based

ones: 44.1kHz signal CNN6 had 46.5 macro-averaged F1-score and CNN10 44.6. A more detailed description of the systems and the evaluations can be found in D3.1[17].

SED@Edge is a component for sound event detection designed to run on very low-end IoT devices featuring microcontrollers with very limited resources in terms of computation and memory.

Figure 13: SED@Edge on an STM board

SED@Edge employs the PhiNets networks (Paissan et al., 2022), which is a family of modular scalable backbones that can be easily tuned using few hyperparameters to match the memory and computational resources available on different embedded platforms. The main convolutional block used in the architecture is a modified version of the inverted residual block used in MobileNetV2 and MobileNetV3 (Howard et al., 2017, Howard et al., 2019) architectures. This block is composed of a sequence of three operations, namely: a pointwise expansion convolution, a depth-wise convolution, and a squeeze-and-excitation block. The proposed model achieved SOTA performance on the UrbanSound8k dataset (Salamon et al., 2014) for spectrogram classification while using an extremely low number of parameters.

[17] MARVEL D3.1: Multimodal and privacy-aware audio- visual intelligence – initial version, 2022. https://doi.org/10.5281/zenodo.6821318

6 Security mechanisms to ensure privacy preservation during processing, transit and storage

Although this document mostly focuses on the results of the efforts of T3.1, other very important security mechanisms are in place in the MARVEL framework, as discussed in Section 2. In particular:

- Cybersecurity approaches to ensure authentication and encryption of the data being involved in the live MARVEL processing.
- Protection of the data stored in the MARVEL Data Corpus.

Therefore, to give to the reader a more comprehensive view of the overall privacy preservation mechanics developed in the MARVEL project for the E2F2C infrastructure, this section reviews the aforementioned solutions, which are the results of the efforts of other WPs and are hence described in detail in other deliverables.

6.1 Security Mechanisms

The security mechanisms of the MARVEL E2F2C framework focus on the confidentiality of data at rest, in transit, and in use. These mechanisms include authorisation, authentication, and encryption and are described in the following sub-sections.

6.1.1 Authorisation / authentication

MARVdash offers a registration wizard, where a user fills in a username and password. Based on these credentials a user private namespace is created and a private folder is mounted. In that way, user's services and data are isolated. Internally, these services can access any resource in the E2F2C-private network. Persistent data is directed to MARVdash-supplied storage that can also be accessed via the MARVdash web interface. MARVdash is described in detail in D3.2[18].

Kubernetes offers an API object called ingress, which exposes HTTP(S) routes to the cluster services. MARVdash adds authentication directives to all ingress resources. All services are exposed on subdomains of the main dashboard domain. The subdomain is created by composing the service name and the user's name, so they can always be the same, allowing the user to bookmark the location. An ingress controller (NGINX ingress controller) is required for the services to be exposed to the outside world (Figure 14).

Each respective ingress is configured to perform a single sign-on through MARVdash. The default deployment integrates Vouch Proxy[19] as an OAuth 2.0/OIDC client to the dashboard, which in turn provides credentials to the NGINX-based web proxy implementing the ingress. Consequently, each service can only be accessed by its owner and no external parties can visit a user's service frontend without appropriate credentials.

[18] MARVEL D3.2: Efficient deployment of AI-optimised ML/DL models – initial version, 2022. https://doi.org/10.5281/zenodo.6821232

[19] https://github.com/vouch/vouch-proxy

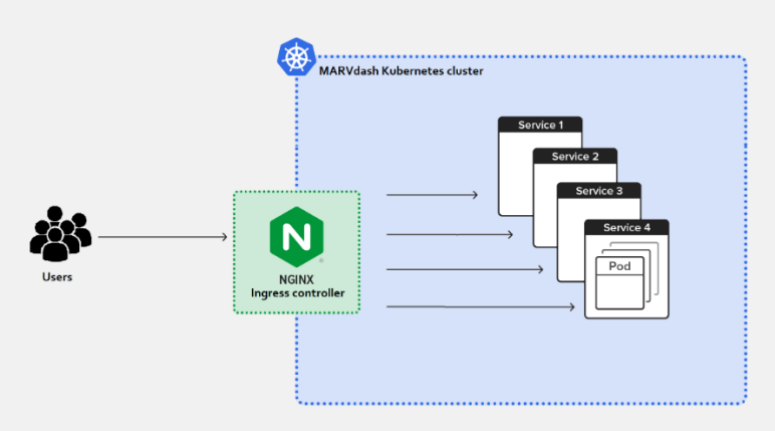

Figure 14: NGINX ingress controller topology

MARVdash allows external, secure access to the Data Corpus API services, by integrating a web proxy service. A new ingress endpoint is created at the Kubernetes side (in order to get an HTTPS-compliant URL) and is configured to require authentication using the MARVdash-registered credentials before forwarding requests to the backend.

6.1.2 Encryption

To achieve encryption, two components are developed, namely EdgeSec VPN and EdgeSec TEE. Both of these tools are described in detail in D4.2[20].

Figure 15: MARVEL E2F2C framework nodes communicating through EdgeSec VPN

EdgeSec VPN is a dockerised version of n2n[21], tailored for MARVEL needs. EdgeSec VPN is aligned with the n2n architecture, which is based on two key components: edge and super nodes. Super nodes are used by the edge nodes for discovering other edge nodes, and routing the traffic when the nodes are behind symmetrical firewalls. The n2n, and therefore the EdgeSec VPN, is a peer-to-peer VPN that works on the second layer of the OSI model, allowing the peers to maintain reachability across Network Address Translations (NATs) and firewalls. Edge nodes that participate in the same virtual network form a community. Super nodes are able to serve more than one community allowing a single computer to join multiple communities. Encryption of the traffic is performed on edge nodes using encryption keys defined by the users. Super nodes, on their turn, forward packets based on a clear-text packet header without inspecting the decrypted packet payload. In that way, data encryption is managed by the users and is not

[20] MARVEL D4.2: Security assurance and acceleration in E2F2C framework – initial version, 2022. https://doi.org/10.5281/zenodo.6821254

[21] https://www.ntop.org/products/n2n/

delegated to infrastructure. Every MARVEL E2F2C framework node has become an edge node in the n2n architecture described above (Figure 15). A super node is also deployed in the PSNC infrastructure.

EdgeSec TEE is based on the SCONE platform[22], inheriting its functionality, therefore it supports the encryption of the network traffic as well as file encryption in a transparent way. EdgeSec TEE allows the configuration of programs with secrets secured from attackers even if they have control of the operating system or the hypervisor. With EdgeSec TEE, programs authenticate themselves ensuring that only the correct ones are executed in an SGX enclave, a separated and encrypted region for code and data. In that way, malwares are prevented from attaching to programs. An application deployed in an SGX enclave protects its data from accesses by all other software including the operating system. As a result, such an application's data are protected against adversaries with root access, since they cannot dump the main memory of an application and get access to its keys. EdgeSec TEE makes use of existing native Python images for the creation of a confidential container (Figure 16). The resulting image runs on remotely-attested Intel SGX enclaves and has an encrypted filesystem.

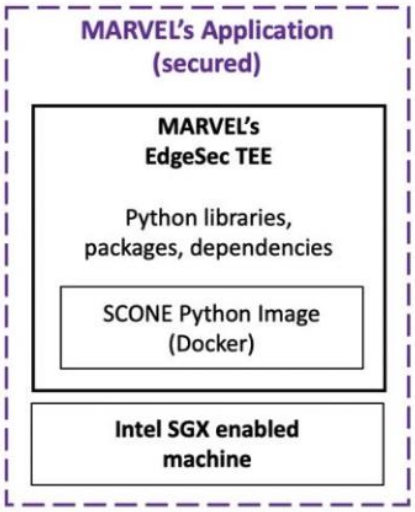

Figure 16: An overview of EdgeSec TEE

6.2 Data storage in MARVEL Data Corpus (from the aspect of privacy)

MARVEL Data Corpus data storage is based on Hadoop Distributed File System (HDFS) fault-tolerant data storage file system (Sivaraman and Manickeckezian, 2014). There are various tools and solutions available to secure the HDFS environment, and each of them has different features and different effectiveness under a different context. The security tools and solutions can be divided into four categories (Saraladevi et al., 2015; Seonyoung and Youngseok, 2013), which are i) encryption, ii) authentication, iii) authorisation, and iv) auditing.

6.2.1 User Authentication for data corpus

Authentication refers to the verification of system or user identity for accessing the system, or in other words, it is the procedure of confirming whether the user is the person they claimed to

[22] https://scontain.com/

be. Two common authentication technologies are (Zhang, 2014): i) Lightweight Directory Access Protocol (LDAP) for the directory, identity, and other services (Ruckcharti et al., 2017), and ii) Kerberos (Algaradi and Rama, 2019).

Authorisation is the process of determining the access rights of the user, specifying what they can do with the system. As Hadoop mixes various systems in its environment, it requires numerous authorisation controls with different granularity. In Hadoop, the process of setup and maintenance of the authorisation control is simplified and can be done by dividing users into groups by specifying them in the existing LDAP or Active Directory (AD). Other than that, authorisation can also be set up by giving role-based access control for connection methods that are alike. The popular tool for authorisation control is Apache Sentry (Zhang, 2014).

Currently, user authentication for the MARVEL Data Corpus is performed through MARVdash allowing external, secure access to the Data Corpus API services, by integrating a web proxy service. The transmission of data is secured with HTTPS and the underlying TLS. The authorisation level is defined during the registration of the user account (e.g., simple user, administrator, etc.).

6.2.2 Authentication mechanism in Data Corpus / HBase / Hadoop

HBase can be configured to provide User Authentication, which ensures that only authorised users can communicate with HBase. The authorisation system is implemented at the RPC level and is based on the Simple Authentication and Security Layer (SASL) (Kanyeba and Yu, 2016), which supports (among other authentication mechanisms) Kerberos. SASL allows authentication, encryption, negotiation, and/or message integrity verification on a per connection basis.

The most popular mechanism for authentication is Kerberos, which is also the primary authentication for Hadoop developed by MIT (Neuman and Ts'o,1994). Kerberos protocol provides secure communications over a non-secure network by using secret-key cryptography. The protocol of Kerberos is shown below:

- The client will first need to request Ticket Grant Ticket (TGT) from the Authentication Server (AS) of the Key Distribution Centre (KDC).
- After the client received the TGT, the client will have to request Service Ticket (ST) from Ticket Grant Server (TGS) from Key Distribution Centre (KDC) (the Authentication Server and Ticket Grant Server are the components of Key Distribution Centre).
- The client can use the ST to authenticate a Name Node.

The TGT and ST will be renewed after a long-running of jobs. The greatest benefit of Kerberos is that the ticket cannot be renewed if it was stolen. Kerberos provides powerful authentication for Hadoop. Instead of using a password alone, the cryptographic mechanism is used when requesting services (Algaradi and Rama, 2019; Kanyeba and Yu, 2016).

6.2.3 Privacy assessments for Data Corpus

In general, the datasets that are stored in the Data Corpus are meant to become public. Therefore, data will have been priorly anonymised by the rest MARVEL platform components (i.e., VideoAnony, AudioAnony, etc.). Nevertheless, there can exist datasets that are in progress and/or the data owner (pilot) has not yet realised/authorised their public use. Three dissemination levels are considered. The ***public datasets*** can be viewed by everyone (including external users). The ***private datasets*** can be viewed only by the MARVEL partners. The ***in-***

progress datasets may not have been anonymised yet and are visible by the MARVEL partners. Data owners (pilots) can set the visibility level as they are processing their datasets. Then, upon user registration and login to the platform, the authentication/authorisation policy for a user type is implemented by the abovementioned security mechanisms (see subsections 6.2.1 and 6.2.2).

Apart from that build-in controls, the **Assurance Platform** (Lakka et al., 2022; Smyrlis et al., 2021) has been deployed in the MARVEL backend. This platform can additionally perform a security/risk analysis for the Corpus components, as well as penetration testing elements. Moreover, *Monitoring Assessment Profiles* can be developed in order to continuously monitor the real-time operation of the system in terms of security and/or privacy. For each profile, a set of *Event Captors* gather information from the system, such as event logs. The main capturing mechanisms are implemented with Elastic Beats, while customised events captors are also supported. These sensing mechanisms are sending information back to the platform and its *Monitor* subcomponent, which analyses the events with a rule-based logic (implemented in the Drools rule engine). ML evaluations can be also supported for cases that required more complex analysis. Henceforth, security and/or privacy assessments can be performed. Two assessment examples are considered so far:

- **Availability of service:** Event captors are periodically checking wherever a service is up or down. In the latter case, the monitoring module can calculate metrics, like the average downtime time, the mean time to repair/restore, or the mean time to respond to an incident. Thus, the Assurance Platform can assess the availability of the internal Hadoop/HBase services, as well as the web-based Graphical User Interface (GUI) that is offered to the external users.
- **User access and Data Ownership:** Event captors can collect the user-related actions from logfiles. Therefore, the Assurance Platform can assess the applicability of privacy controls for *accessing* the data. For example, the platform can reason wherever private or in-progress datasets have been viewed/downloaded by an unauthorised external user (an attacker that gained access to the system) and raise an alert to the system administrator. The platform can also measure the number of violations in case of such an attack, which may also reflect the time that the administrator is required to fix/mitigate the problem (i.e., block the malicious or compromised user account). Similarly, the platform can audit the *processing* of aspects for the datasets, i.e., wherever an unauthorised user account managed to ingest new data or update/delete existing datasets.

The rulesets for the deployed assessment profiles form the basis for the Service Level Agreement (SLA) driven verification of the system. The aforementioned metrics (e.g., mean time to respond, number of violations, etc.) can be utilised, with thresholds specifying the fair use of the system.

7 Conclusions

This document provided a description of the audio and video anonymisation methods deployed in the 1st MARVEL integrated prototype, as well as of the advanced approaches developed as part of Task 3.1 for the final prototype. The section on VideoAnony focused in particular on quantization and compression of the solutions currently deployed. The section described also an alternative very compact version of the component which is suitable for deployment on IoT microcontrollers. An integration example where anonymised videos from the MT1 use case are successfully processed by the VCC component is also reported at the end as a showcase of a typical processing pipeline in the MARVEL framework. Concerning AudioAnony, the deliverable presented details about the first solution implemented in the R1 and the new component based on sequence-to-sequence voice conversion strategies using pre-trained models. This new solution will be integrated with VAD and deployed in the R2. The document described also plans for further developments under other tasks (Task 3.5) towards the final deployment (R2) in the MARVEL framework. The design of both the audio and video anonymisation components was driven by guidelines and regulations provided by the EU and by the GDPR as well as by specific requirements from the DPOs of every single pilot. Their deployment and real-time processing in the prototype are essential to allow the execution of the use cases.

To provide the reader with the general context within which the activities under the Task 3.1 have been carried out, the most relevant issues and limitations related to privacy were also discussed in this document, with a specific focus on smart cities and on the MARVEL's use cases. Finally, the deliverable also provided an overview of other mechanisms necessary to preserve the citizens' privacy by ensuring security in the E2F2C processing as well as preventing unauthorised accesses to the MARVEL Data Corpus.

8 References

T. S. Algaradi, and B. Rama, "Static knowledge-based authentication mechanism for Hadoop distributed platform using Kerberos," International Journal on Information Technology, Advanced Science Engineering, vol. 9, issue 3, pp. 772-780, 2019.

M.N. Asghar, N. Kanwal, B. Lee, M. Fleury, M. Herbst, and Y. Qiao, "Visual surveillance within the EU general data protection regulation: A technology perspective". *IEEE Access, 7*, pp. 111709-111726, 2019.

E. Cakir, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen. "Convolutional recurrent neural networks for polyphonic sound event detection*". Transactions on Audio, Speech and Language Processing: Special issue on Sound Scene and Event Analysis*, 25(6):1291–1303, 2017.

Y.-H. Chen, D.-Y. Wu, T.-H. Wu, and H.-y. Lee, "Again-VC: A one-shot voice conversion using activation guidance and adaptive instance normalization," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2021.

R. Chen, X. Chen, B. Ni, and Y. Ge, "Simswap: An efficient framework for high fidelity face swapping". In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2003-2011, 2020.

S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao, J. Wu, L. Zhou, S. Ren, Y. Qian, Y. Qian, J. Wu, M. Zeng, X. Yu, and F. Wei, "Wavlm: Large-scale self-supervised pre-training for full stack speech processing", IEEE Journal of Selected Topics in Signal Processing, 2022.

K. Drossos, S. Mimilakis, S. Gharib, Y. Li, and T. Virtanen, "Sound event detection with depthwise separable and dilated convolutions". In International Joint Conference on Neural Networks, pages 1–7, 2020.

L. Du, W. Zhang, H. Fu, W. Ren, and X. Zhang, "An efficient privacy protection scheme for data security in video surveillance". *Journal of Visual Communication and Image Representation, 59*, pp.347-362, 2019.

O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video". In *Proceedings of the IEEE/CVF International Conference on Computer Vision,* pp. 9378-9387, 2019.

T. Gerstner, D. DeCarlo, M. Alexa, A. Finkelstein, Y. Gingold, and A. Nealen, "Pixelated image abstraction with integrated user constraints". *Computers & Graphics*, *37*(5), pp.333-347, 2013.

X. Guo, S. Li, J. Yu, J. Zhang, J. Ma, L. Ma, W. Liu, W., H. Ling. "PFLD: A Practical Facial Landmark Detector". arXiv. https://doi.org/10.48550/arXiv.1902.10859, 2019

Y. Nirkin, Y. Keller and T. Hassner, "FSGAN: Subject Agnostic Face Swapping and Reenactment," in IEEE/CVF International Conference on Computer Vision, 2019 pp. 7183-7192

Howard A.G., M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications". arXiv preprint arXiv:1704.04861, 2017.

Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," in IEEE/CVF International Conference on Computer Vision, 2019, pp. 1314-1324

N. Dall'Asen, Y. Wang, H. Tang, L. Zanella, E. Ricci, "Graph-based generative face anonymization with pose preservation", In *Proceedings of the International Conference on Image Analysis and Processing*, 2021.

H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization". In *International symposium on visual computing*, pp. 565-578, Springer, Cham, 2019.

M. Kanyeba, and L. Yu, "Securing authentication within Hadoop," International Conference on Electrical, Mechanical and Industrial Engineering, Atlantis Press, Phuket, Thailand, pp. 100-103, 2016.

T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks". In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pp. 4401-4410, 2019.

N. R. Koluguri, T. Park, and B. Ginsburg, "TitaNet: Neural Model for speaker representation with 1D Depth-wise separable convolutions and global context," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2022.

Z. Khan, Z. Pervez, and A. G. Abbasi, "Towards a secure service provisioning framework in a smart city environment". *Future Generation Computer Systems*, *77*, pp.112-135, 2017.

J. Kong, J. Kim, and J. Bae, "HiFi-GAN: Generative Adversarial Networks for efficient and high-fidelity speech synthesis," CoRR, vol. abs/2010.05646, 2020.

Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-scale pretrained audio neural networks for audio pattern recognition". IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28:2880–2894, 2020.

E. Lakka, G. Hatzivasilis, S. Karagiannis, A. Alexopoulos, M. Athanatos, S. Ioannidis, M. Chatzimpyrros, G. Kalogiannis, and G. Spanoudakis, "Incident handling for healthcare organizations and supply-chains," IEEE Symposium on Computers and Communications (ISCC), Rhodes, Greece, pp. 1-7, 2022.

S. Liu, Y. Cao, D. Wang, X. Wu, X. Liu, and H. Meng, "Any-to-many voice conversion with location-relative sequence-to-sequence modeling," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, 2021.

M. Maximov, I. Elezi, and L. Leal-Taixé, "CIAGAN: Conditional identity anonymization generative adversarial networks". In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5447-5456), 2020.

X. Miao, X. Wang, E. Cooper, J. Yamagishi, and N. Tomashenko, "Language-Independent Speaker Anonymization Approach Using Self-Supervised Pre-Trained Models," in The Speaker and Language Recognition Workshop (Odyssey), 2022.

C. Neuman, C., and T. Ts'o, "Kerberos: An authentication service for computer networks," IEEE Communications Magazine, pp. 33-38, 1994.

F. Paissan, A. Ancilotto, E. Farella, "PhiNets: a scalable backbone for low-power AI at the edge," ACM Transactions on Embedded Computing Systems, 2022.

V. Panayotov et al., “Librispeech: An ASR corpus based on public domain audio books,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2015, pp. 5206–5210.

J. Patino, N. A. Tomashenko, M. Todisco, A. Nautsch, and N. W. D. Evans, “Speaker anonymisation using the McAdams coefficient,” in Interspeech, 2021.

K. Qian, Y. Zhang, S. Chang, X. Yang, and M. Hasegawa-Johnson, “AutoVC: Zero-shot voice style transfer with only autoencoder loss,” in International Conference on Machine Learning, 2019.

K. Qian, Y. Zhang, H. Gao, J. Ni, C.-I. Lai, D. Cox, M. Hasegawa-Johnson, and S. Chang, “ContentVec: An improved self-supervised speech representation by disentangling speakers,” in Proceedings of the 39th International Conference on Machine Learning, 2022.

U. Ruckcharti, W. Krathu, N. Atiratana, and C. Arpnikanondt., “Hadoop as a service for academic purposes,” 9th International Conference on Management of Digital Ecosystems (MEDES’17), ACM, Bangkok, Thailand, pp. 81-87, 2017.

J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” in 22nd ACM International Conference on Multimedia (ACM-MM’14), Orlando, pp. 1041–1044, 2014.

B. Saraladevi, N. Pazhaniraja, P. Victer Paul, M.S. Saleem Basha, and P. Dhavachelvan, “Big Data and Hadoop-A study in security perspective,” 2nd International Symposium on Big Data and Cloud Computing (ISBCC’15), Elsevier, Procedia Computer Science, vol. 50, pp. 596-601, 2015.

P. Seonyoung, P., and L. Youngseok, “Secure Hadoop with encrypted HDFS,” International Conference on Grid and Pervasive Computing (GPC), Springer, LNTCS, vol. 7861, pp. 134-141, 2013.

E. Sivaraman, E., and R. Manickeckezian, “High performance and fault tolerant distributed file system for Big Data storage and processing using Hadoop,” International Conference on Intelligent Computing Applications, IEEE, Coimbatore, India, pp. 32-36, 2014.

M. Smyrlis, I. Somarakis, G. Spanoudakis, G. Hatzivasilis, and S. Ioannidis, “CYRA: A model-driven cyber range Assurance Platform,” Applied Sciences, MDPI, vol. 11, issue 11, article 5165, pp. 1-28, 2021.

Q. Sun, L. Ma, S. J. Oh, L. Van Gool, B. Schiele, and M. Fritz, “Natural and effective obfuscation by head inpainting”. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 5050-5059, 2018.

Q. Sun, A. Tewari, W. Xu, M. Fritz, C. Theobalt, and B. Schiele, “A hybrid model for identity obfuscation by face replacement”. In *Proceedings of the European conference on computer vision (ECCV)* pp. 553-569, 2018.

R. Wang, Y. Ding, L. Li, and C. Fan, “One-Shot Voice Conversion using Star-Gan,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2020.

K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X.S. Shen, “Security and privacy in smart city applications: Challenges and solutions”. *IEEE Communications Magazine*, *55*(1), pp.122-129, 2017.

X. Zhang, “Secure your Hadoop cluster with Apache Sentry,” ed: Cloudera, April, 2014.

P. Zinemanas, P. Cancela & M. Rocamora, "MAVD: A Dataset for Sound Event Detection in Urban Environments", Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop, pages 263–267, New York University, 2019.