RUHR-UNIVERSITÄT BOCHUM

RUB

Privacy-preserving Feature Extraction and Classification in Acoustic Sensor Networks

Alexandru Nelus, Rainer Martin

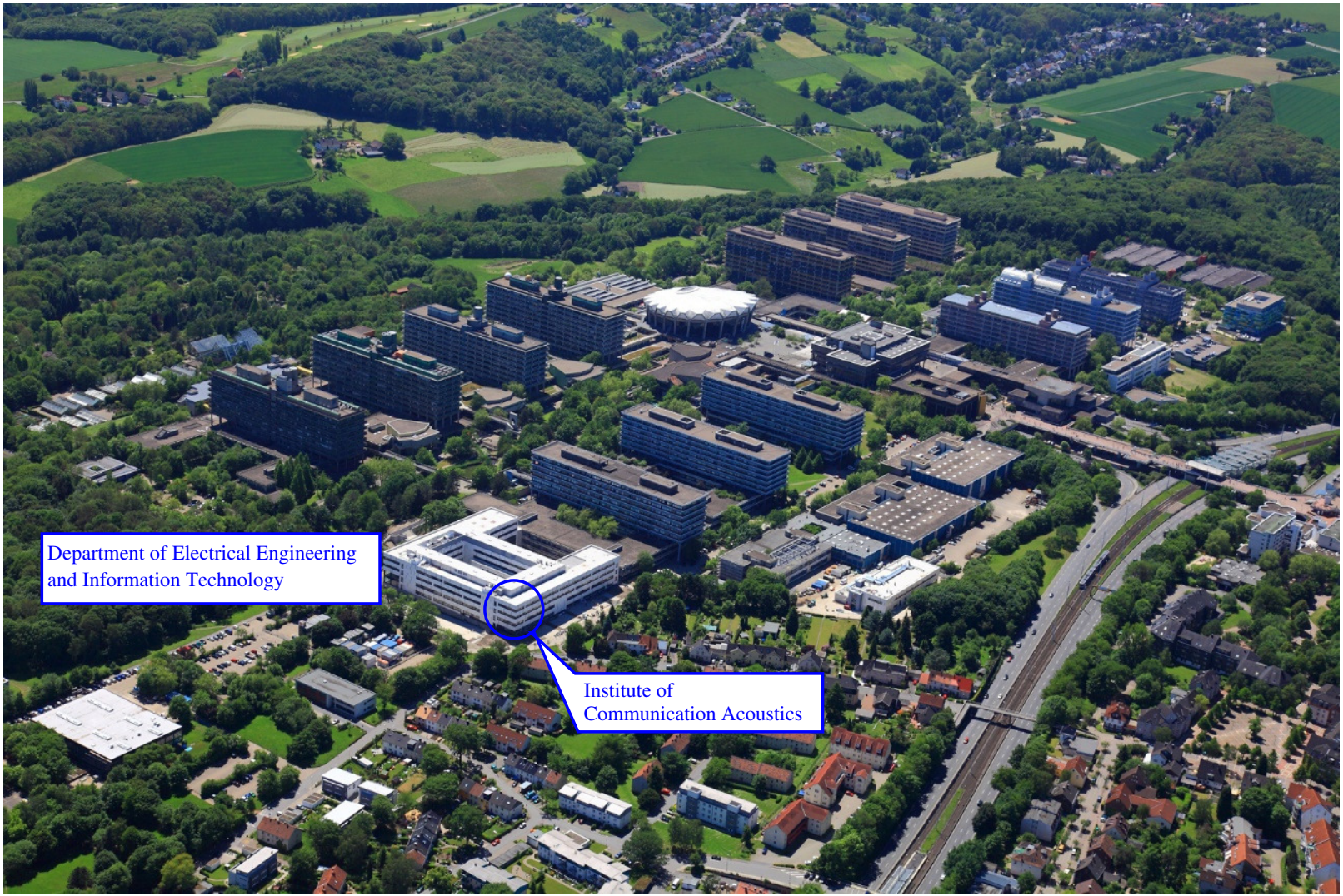
November 2, 2020

Institute of Communication Acoustics

FOR 2457 Acoustic Sensor Networks

Outline

IKA @ Ruhr-Universität Bochum



Institute of Communication Acoustics

Fundamental research and technologies for

human-to-human and **human-to-machine**

communication.

- **Research groups:**
  - **Speech and Audio Signal Processing**
    Prof. Dr.-Ing. Rainer Martin
  - **Cognitive Signal Processing**
    Prof. Dr.-Ing. Dorothea Kolossa
  - **Adaptive Systems of Signal Processing**
    PD Dr.-Ing. Gerald Enzner

Speech and Audio Signal Processing

Fundamental research and technologies for

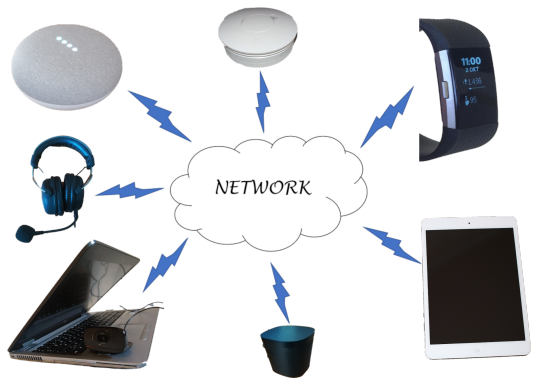
human-to-human and **human-to-machine**

communication.

- **Signal processing, estimation, and machine learning for**
  - voice communication systems
  - hearing aids and cochlear implants
  - recording, processing, and rendering of audio signals
  - acoustic scene analysis and environmental surveillance
  - acoustic source localization, separation, and tracking
  - acoustic sensor networks and IoT.

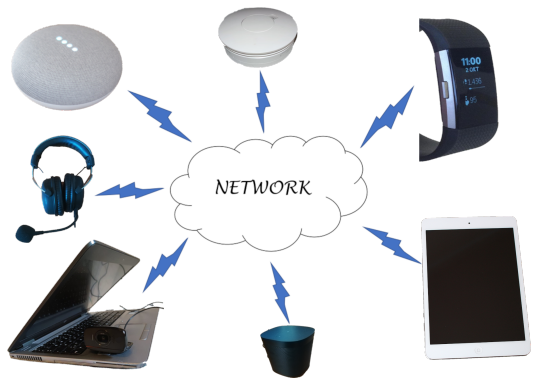
Acoustic Sensor Network (ASN)

- Many applications make use of acoustic sensors
  - smart watches / smart phones / smart speakers
  - surveillance and security devices
  - typically, all of these are connected (WiFi, BT $\rightarrow$ IoT)



Acoustic Sensor Network (ASN)

- Many applications make use of acoustic sensors
  - smart watches / smart phones / smart speakers
  - surveillance and security devices
  - typically, all of these are connected (WiFi, BT $\rightarrow$ IoT)



- ASNs open new opportunities for acoustic signal processing and learning
  - speech signal enhancement
  - acoustic event detection
  - acoustic source localization
  - speech recognition

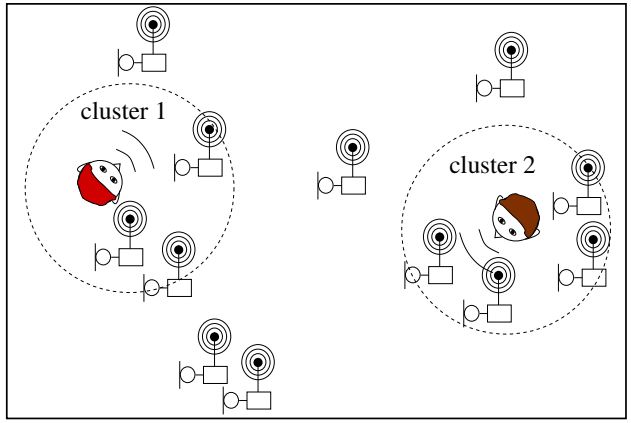
- Increasing demands for privacy (see e.g. EU GDPR)

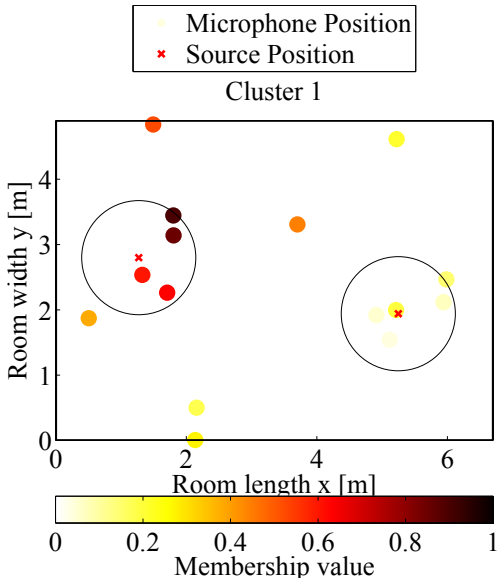
- DFG collaborative research project FOR 2457

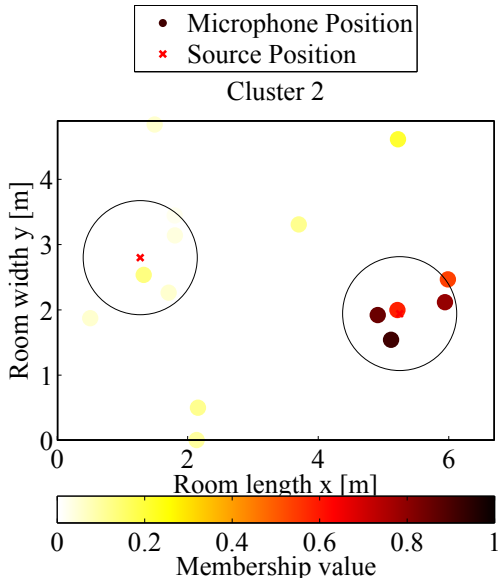
FOR 2457 Acoustic Sensor Networks

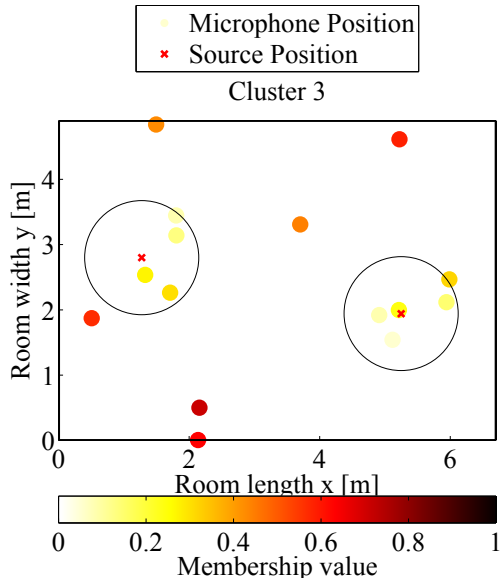
Clustering and Enhancement in ASNs

Assignment of acoustic sensors to source-dominated clusters via fuzzy clustering.







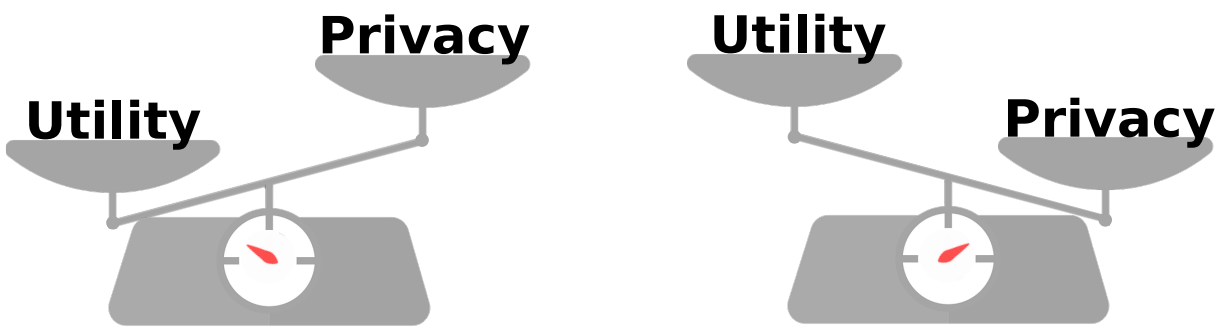


Privacy in Speech and Audio

- Speech signals convey information on several levels (linguistic, paralinguistic).
- In speech and audio privacy, protection strategies are not well established [Nautsch et al., 2019] - and are difficult to deal with!

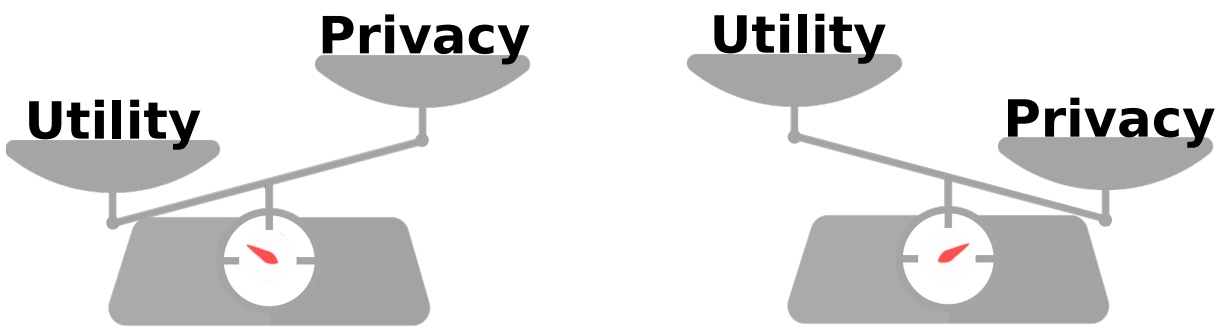
Privacy in Speech and Audio

- Speech signals convey information on several levels (linguistic, paralinguistic).
- In speech and audio privacy, protection strategies are not well established [Nautsch et al., 2019] - and are difficult to deal with!
- Difficult utility - privacy trade-off [Nelus et al., 2016]
  - **utility**: maximize performance of desired task
  - **privacy**: minimize revealed task-extraneous information



Privacy in Speech and Audio

- Speech signals convey information on several levels (linguistic, paralinguistic).
- In speech and audio privacy, protection strategies are not well established [Nautsch et al., 2019] - and are difficult to deal with!
- Difficult utility - privacy trade-off [Nelus et al., 2016]
  - **utility**: maximize performance of desired task
  - **privacy**: minimize revealed task-extraneous information



⇒ How can we utilize acoustic sensors for novel applications without compromising privacy?

Recent Publications

- Follow privacy-by-design approach via **information minimization**
  - privacy requirements and feature aggregation
    [Nelus et al., 2016; Nelus et al., 2017]

Recent Publications

- Follow privacy-by-design approach via **information minimization**
  - privacy requirements and feature aggregation
    [Nelus et al., 2016; Nelus et al., 2017]
- Explore **DNN-based privacy-preserving feature extraction** schemes
  - variational information, siamese, adversarial
    [Nelus & Martin, 2019; Nelus et al., 2019; Nelus & Martin, 2018]

Recent Publications

- Follow privacy-by-design approach via **information minimization**
  - privacy requirements and feature aggregation
    [Nelus et al., 2016; Nelus et al., 2017]
- Explore **DNN-based privacy-preserving feature extraction** schemes
  - variational information, siamese, adversarial
    [Nelus & Martin, 2019; Nelus et al., 2019; Nelus & Martin, 2018]
- Improve **trade-off between privacy and utility** in ASN-based clustering and classification tasks
  - domestic activity monitoring, smart home
    [Nelus, Ebbers, et al., 2019]
  - clustering and local enhancement of signals.
    [Gergen et al., 2018]

Recent Publications

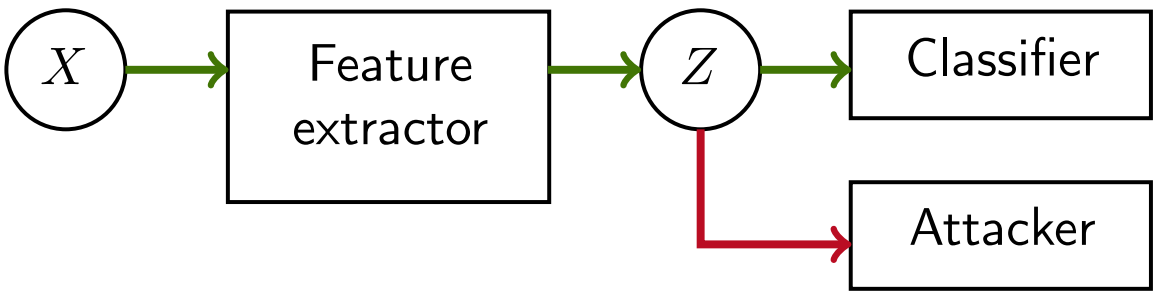
- Follow privacy-by-design approach via **information minimization**
  - privacy requirements and feature aggregation
    [Nelus et al., 2016; Nelus et al., 2017]
- Explore **DNN-based privacy-preserving feature extraction** schemes
  - variational information, siamese, adversarial
    [Nelus & Martin, 2019; Nelus et al., 2019; Nelus & Martin, 2018]
- Improve **trade-off between privacy and utility** in ASN-based clustering and classification tasks
  - domestic activity monitoring, smart home
    [Nelus, Ebbers, et al., 2019]
  - clustering and local enhancement of signals.
    [Gergen et al., 2018]

$\Rightarrow$ New insights into the functionality of DNN-based privacy-preserving feature extraction in **real-world scenarios**.

Feature Interception Attacks in ASNs

- Consider current technological trends:
  - Internet of Things (IoT) $\rightarrow$ distributed processing
  - prevalence of DNN-based solutions

- Highlight privacy risks at data pipeline level in ASN scenarios
  - feature extractor converts low-level representation X into processed feature vector Z
  - interception of DNN-based feature representation Z

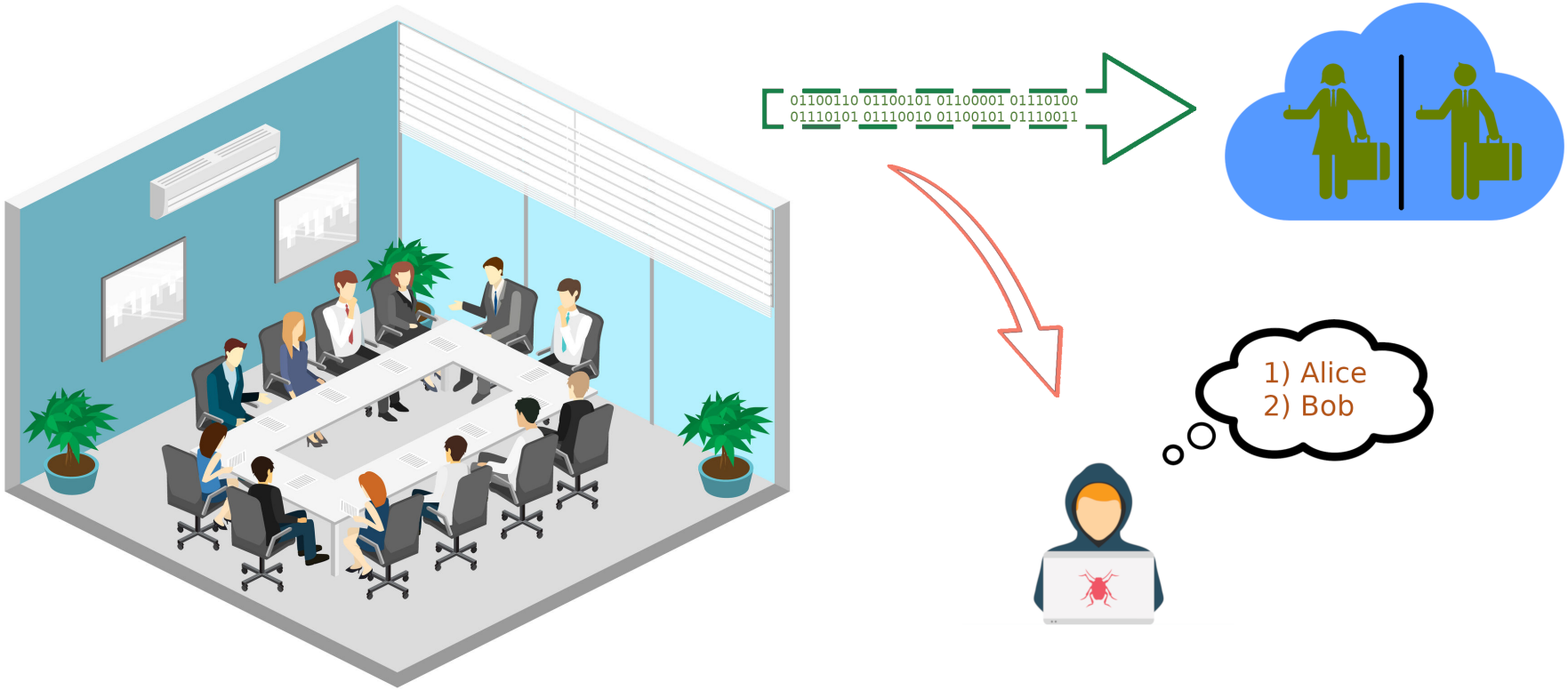


- Propose to tackle them using:
  - information minimization - variational information feature extraction
  - Euclidean-distance-based - siamese feature extraction

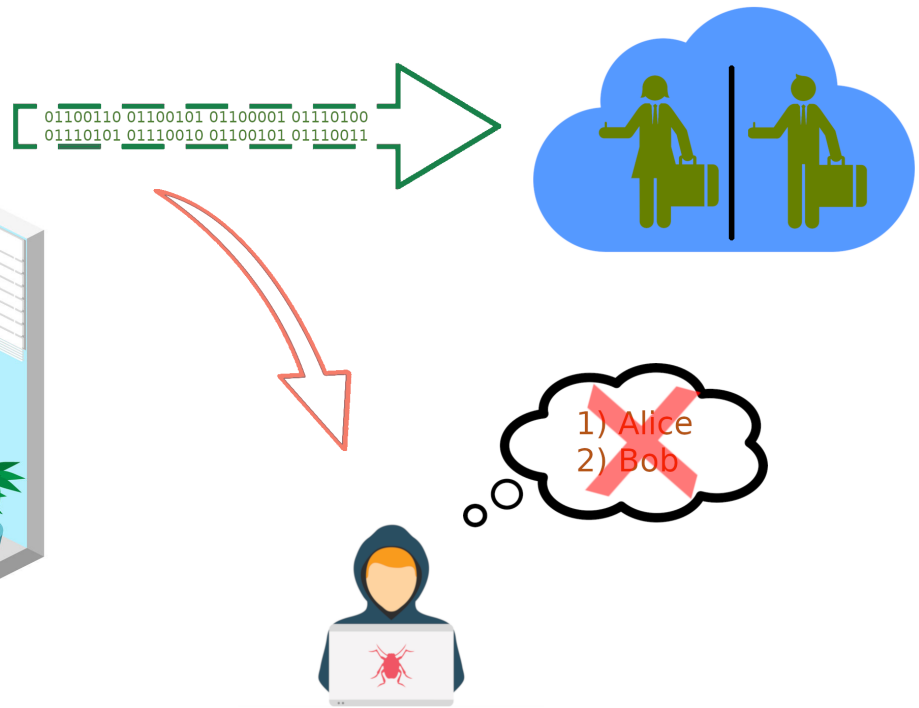
Gender Recognition vs. Speaker Identification

Gender Recognition vs. Speaker Identification

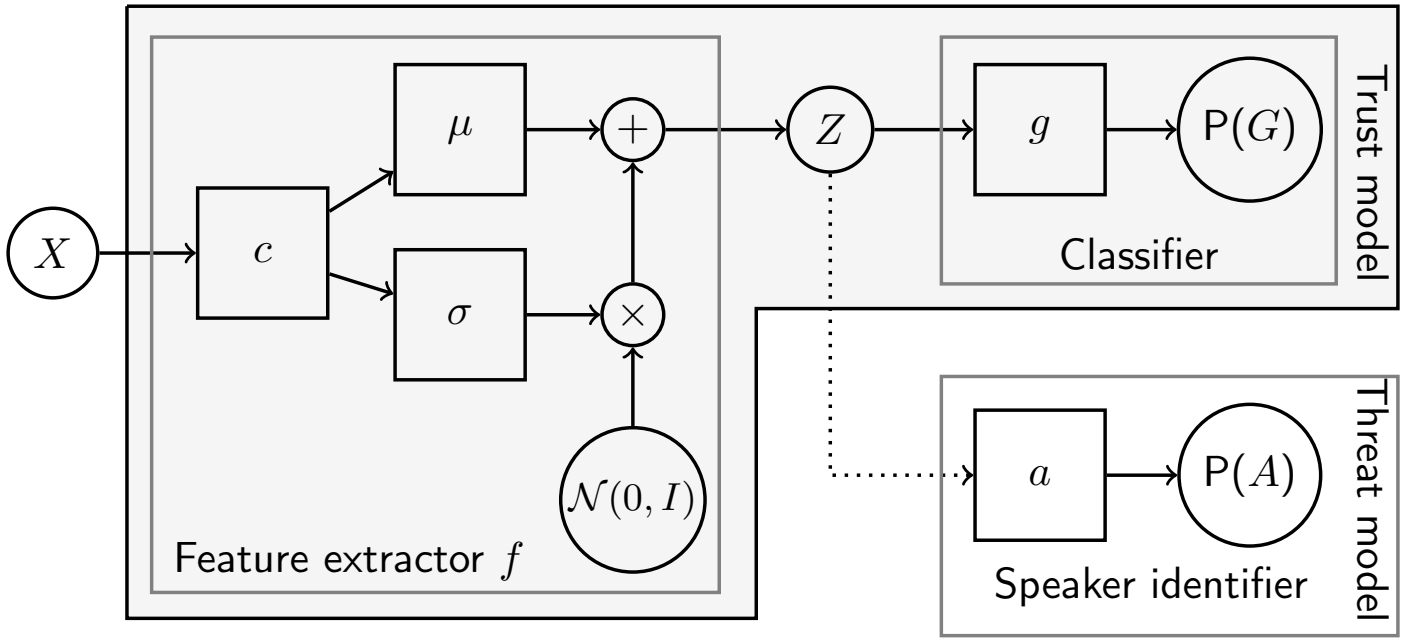
Gender Recognition vs. Speaker Identification



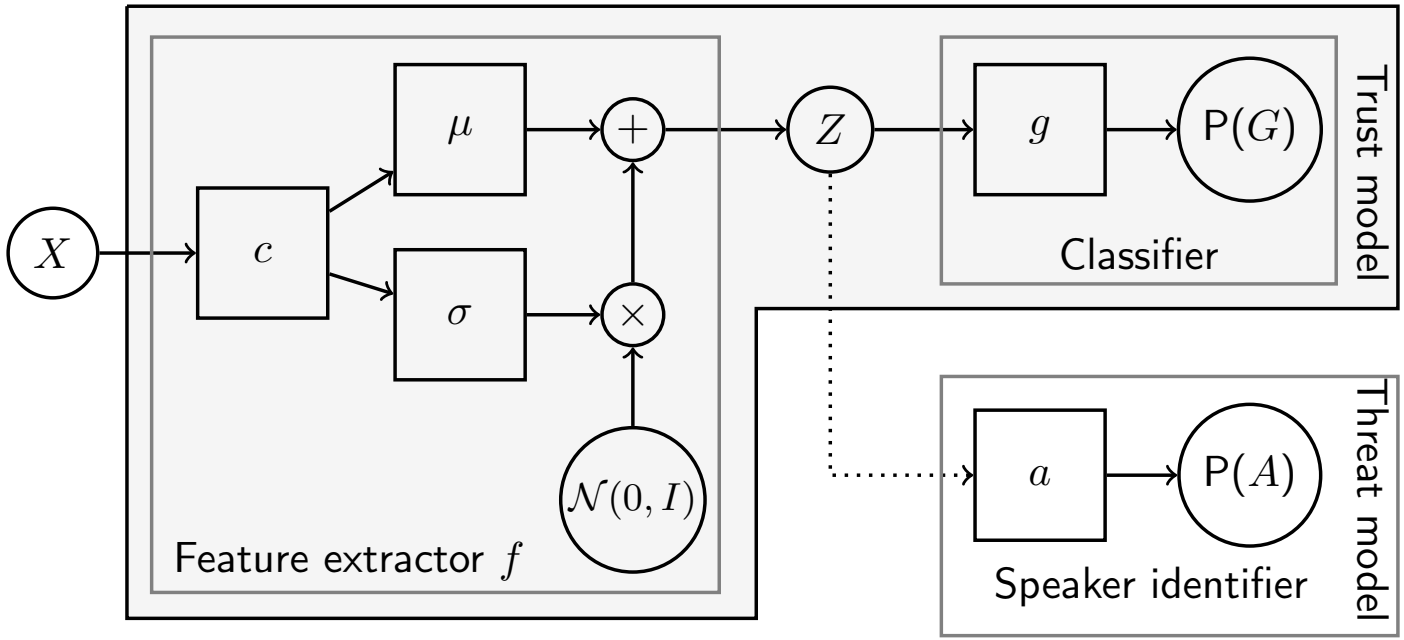
Gender Recognition vs. Speaker Identification



Variational Information Training

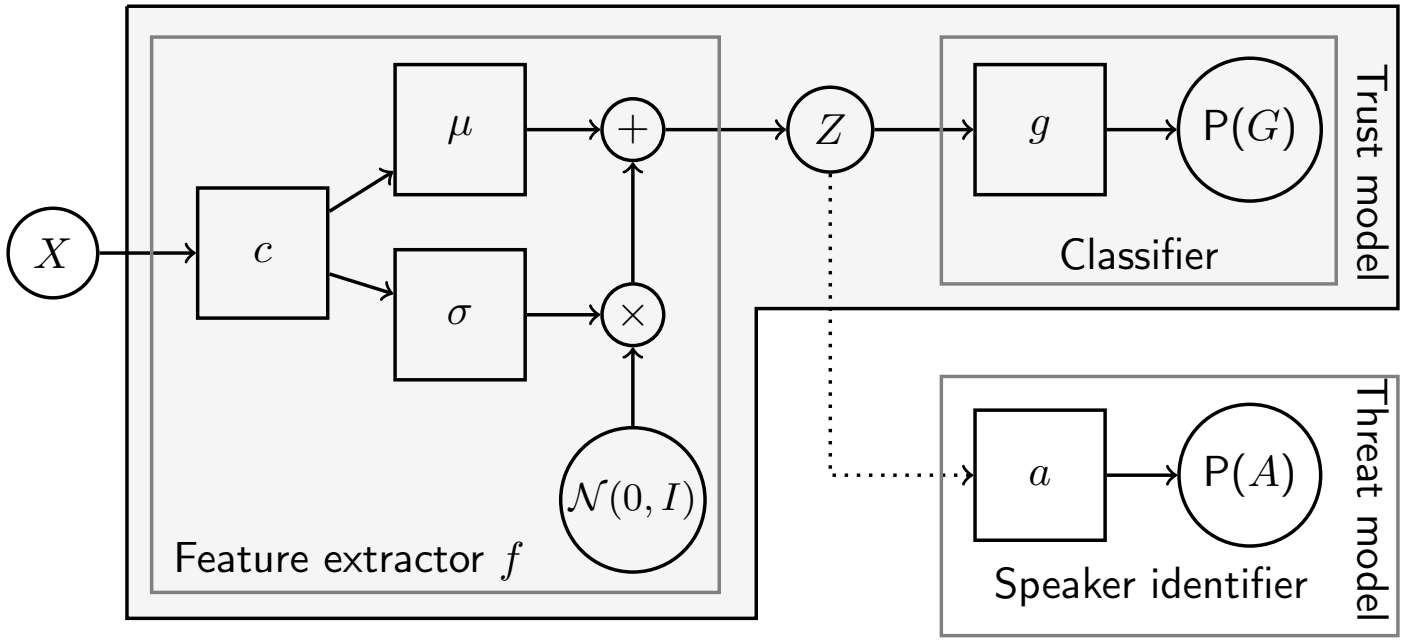


Variational Information Training



- Enhance trusted task (classifier) performance
- Expose a minimum amount of task-extraneous information

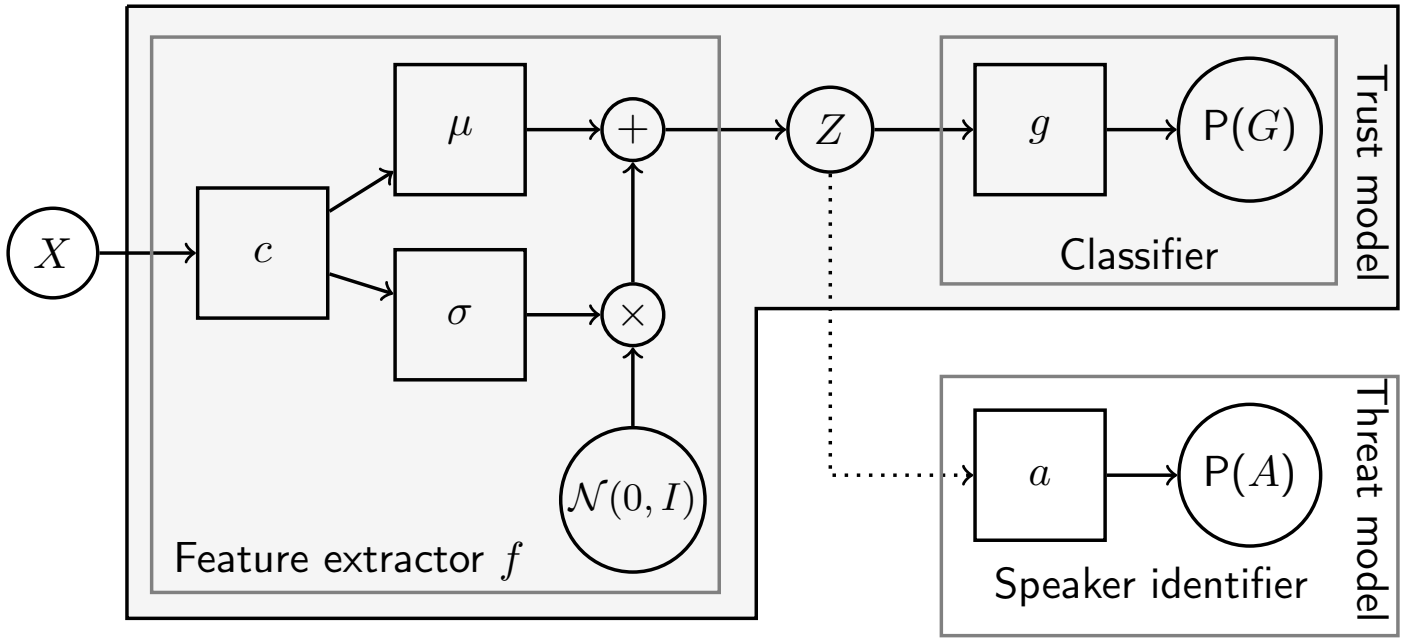
Variational Information Training



- Enhance trusted task (classifier) performance
- Expose a minimum amount of task-extraneous information

$$\min_{\Phi_c, \Phi_\mu, \Phi_\sigma, \Phi_g} \mathbb{E}_{G^t \sim p(G^t)}[-\log p(G)] + \beta I(X; Z)$$

Variational Information Training

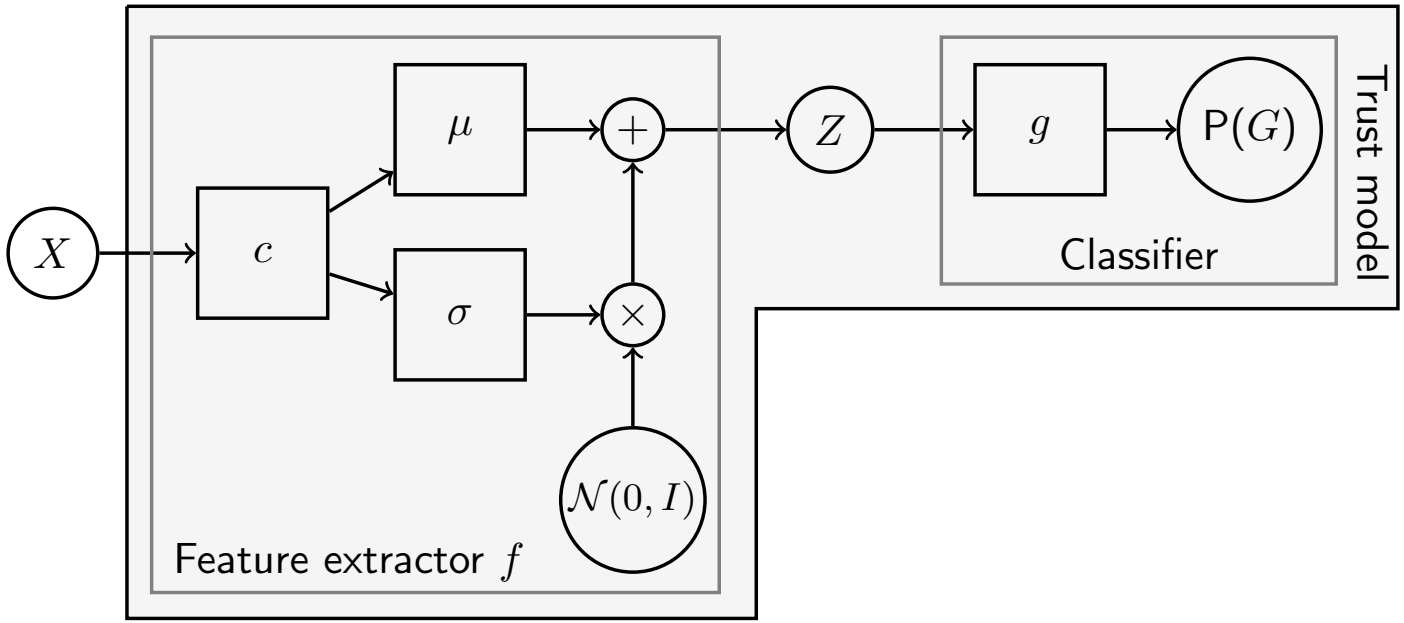


- Enhance trusted task (classifier) performance
- Expose a minimum amount of task-extraneous information

$$\min_{\Phi_c, \Phi_\mu, \Phi_\sigma, \Phi_g} \mathbb{E}_{G^t \sim p(G^t)}[-\log p(G)] + \beta I(X; Z)$$

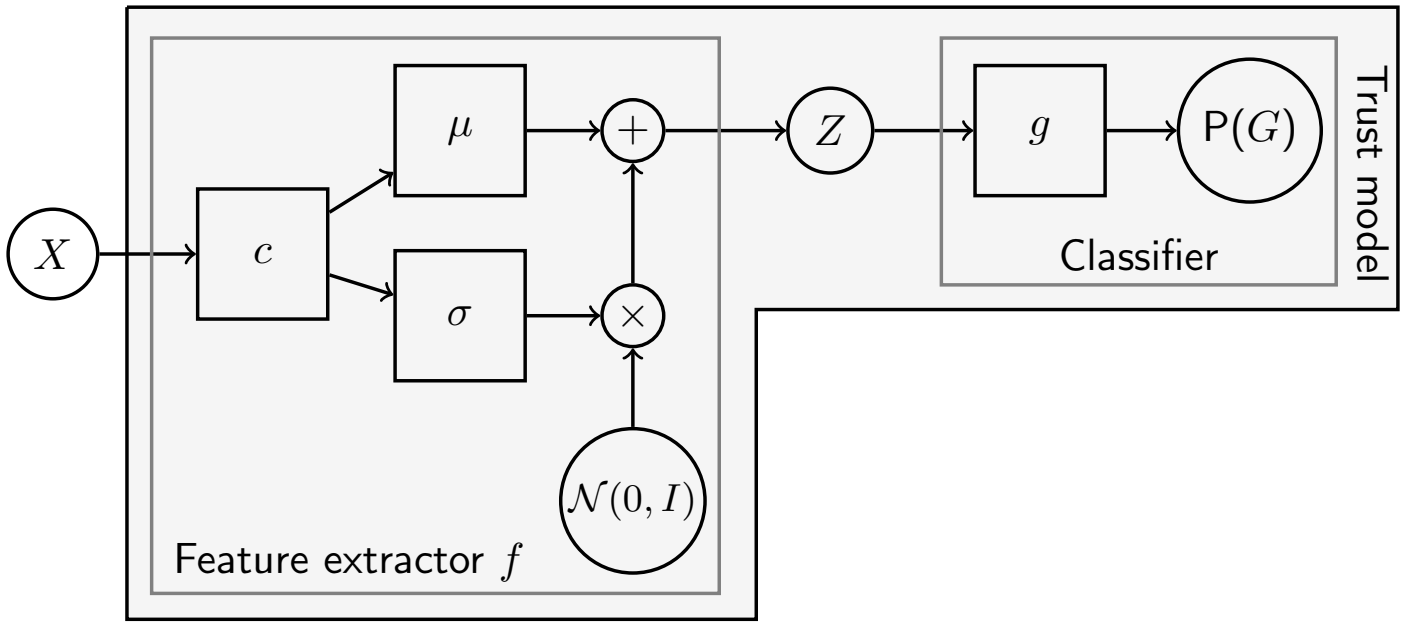
- Find $I_{max}(X; Z) \geq I(X; Z)$

Computing the Upper Bound $I_{max}(X;Z)$



- Introduce variational approximation for $p(z) \rightarrow q(z) = \mathcal{N}(0, I)$
- Use property of KL-divergence: $D\left(p(z)||q(z)\right) \geq 0$
- Obtain analytical formulation for upper bound:
  $I(X;Z) \leq D(p(z|x)||q(z)) = I_{max}(X;Z)$

Variational Information Training

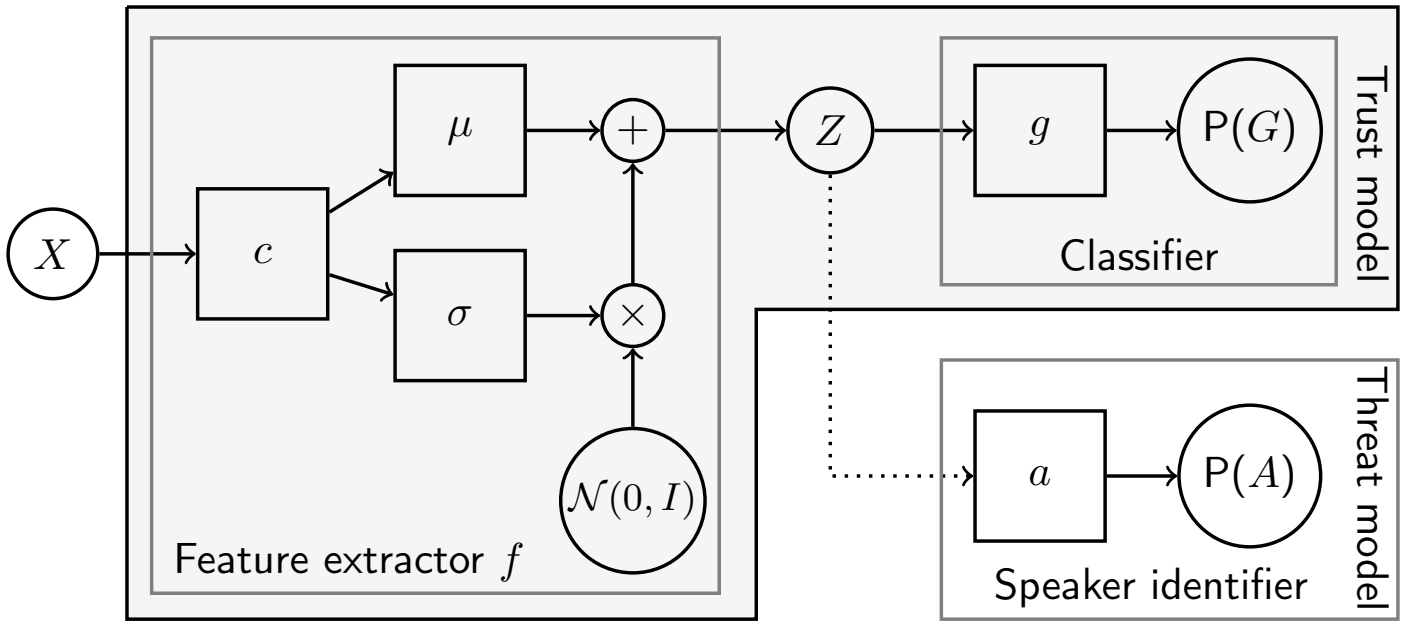


- Loss function to minimize:

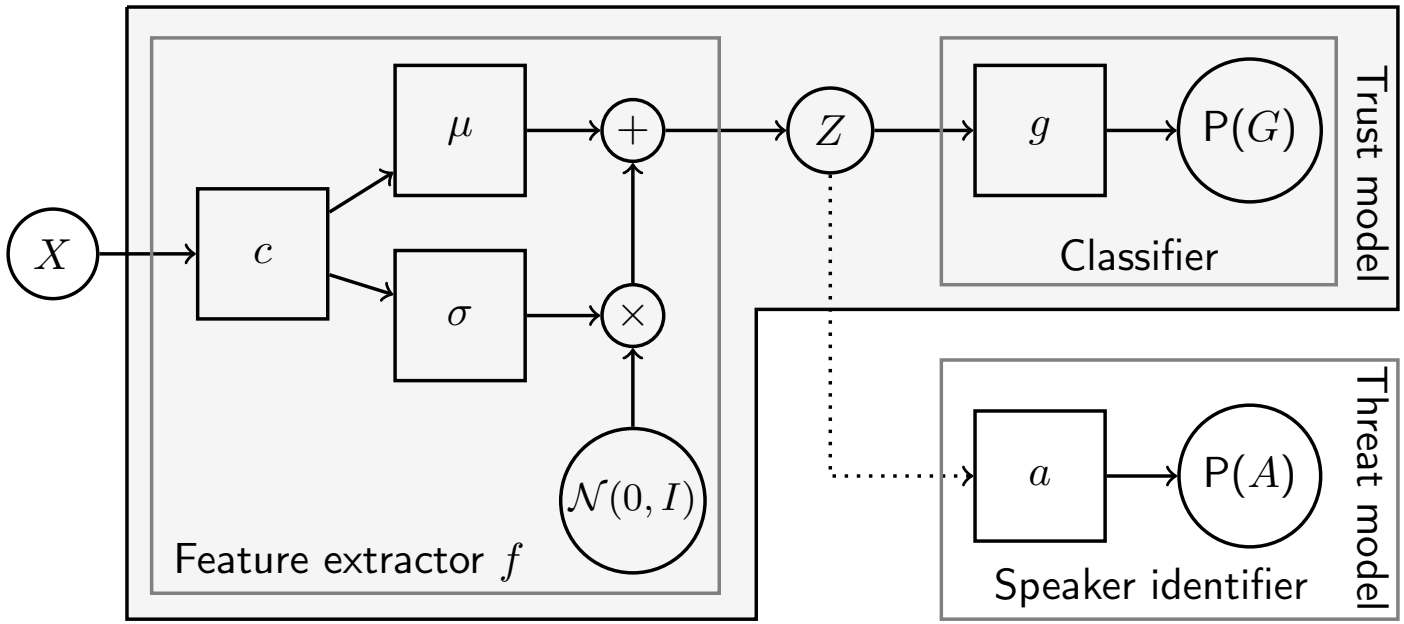
$$\min_{\Phi_c, \Phi_\mu, \Phi_\sigma, \Phi_g} \mathbb{E}_{G^t \sim p(G^t)}[-\log p(G)] + \beta I_{max}(X; Z)$$

- Control privacy vs. utility trade-off using budget scaling factor β

Speaker Identification Training



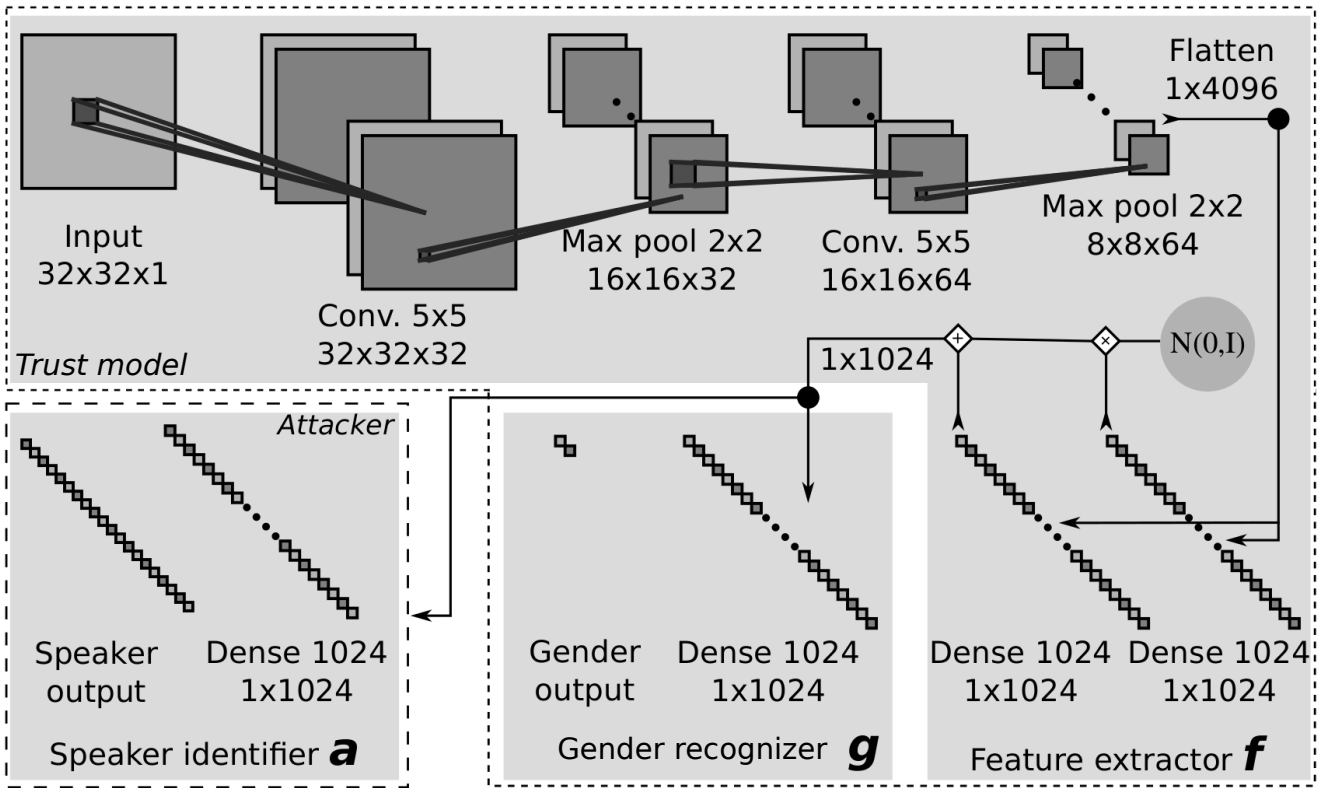
Speaker Identification Training



- Loss function to minimize:

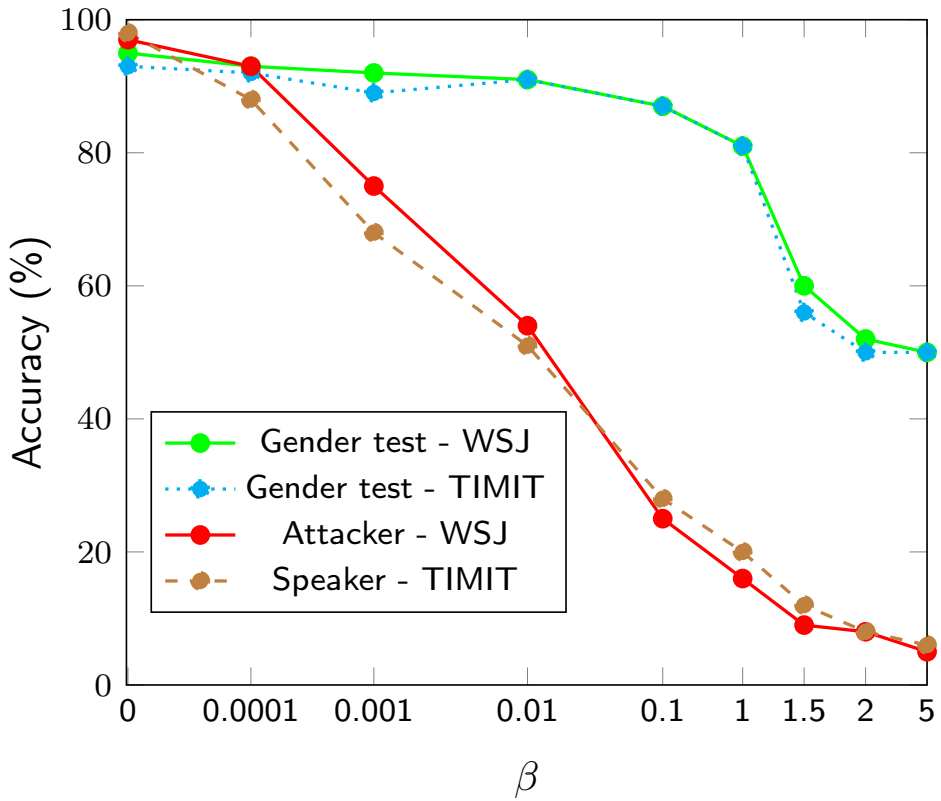
$$\min_{\Phi_a} \mathbb{E}_{A^t \sim p(A^t)}[-\log p(A)]$$

Network Architecture

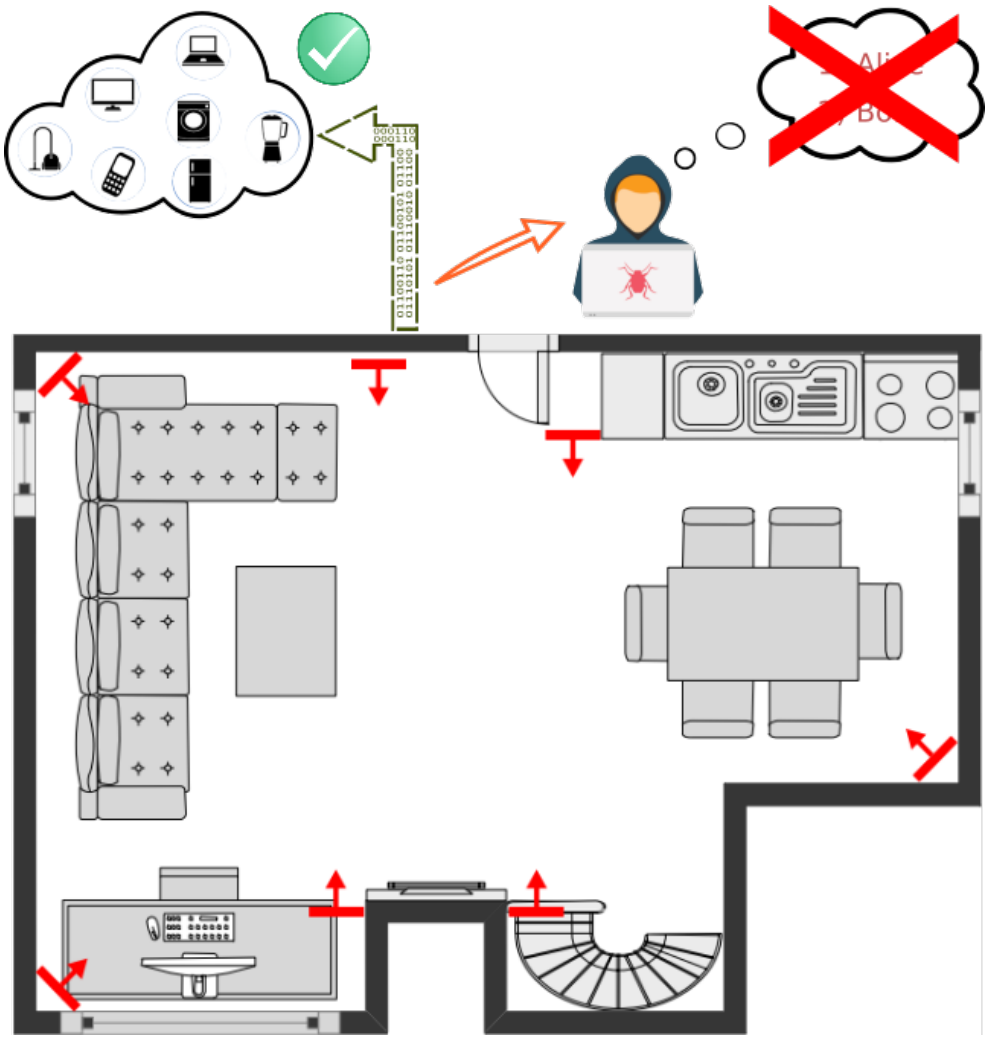


Results

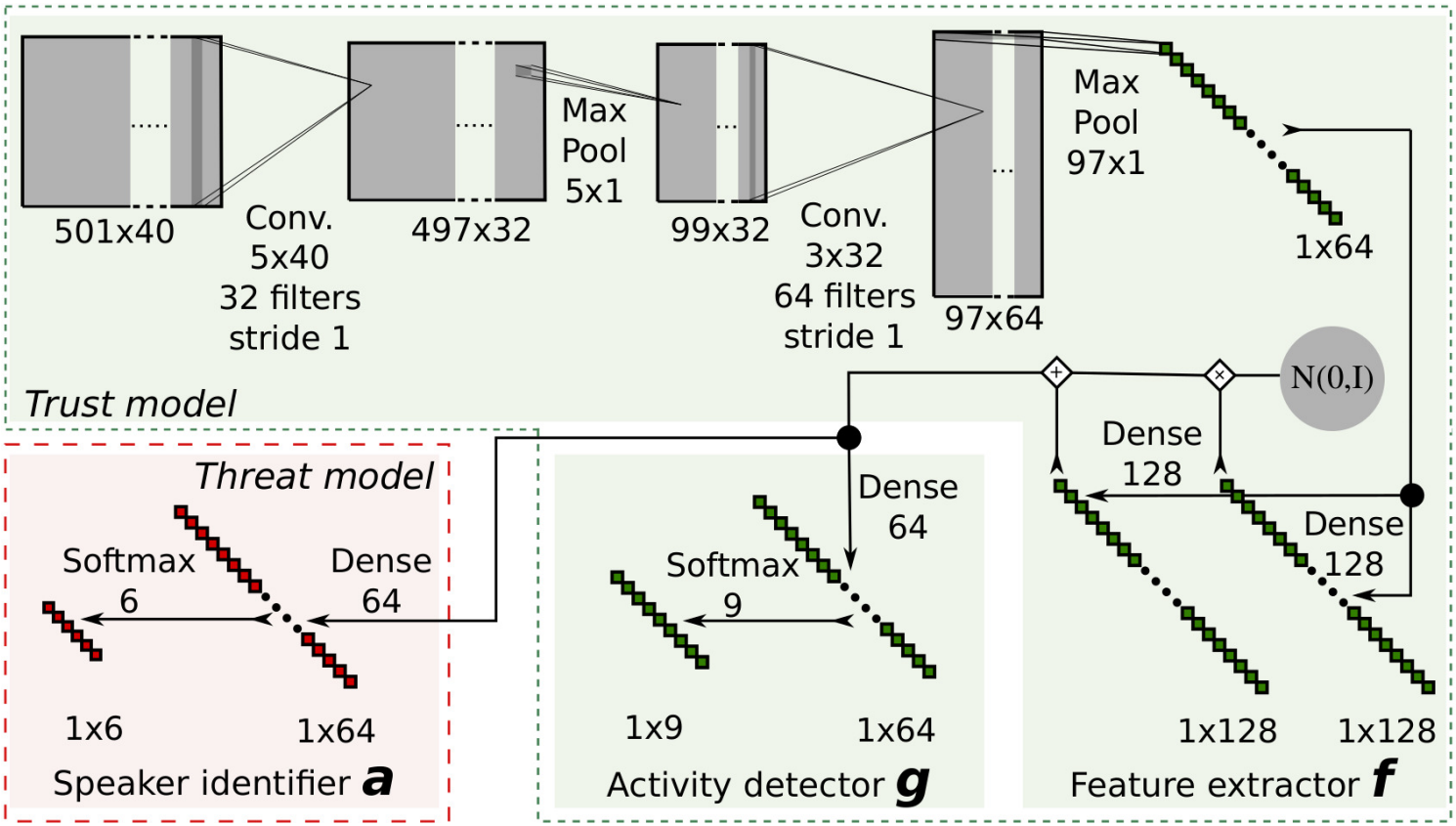
- Gender recognizer:
  - train on Wall Street Journal (WSJ)
  - test on WSJ and TIMIT
- Attacker:
  - train and test on WSJ and TIMIT
  - use 20 speaker groups
- Results aggregated over 10 cross-validations



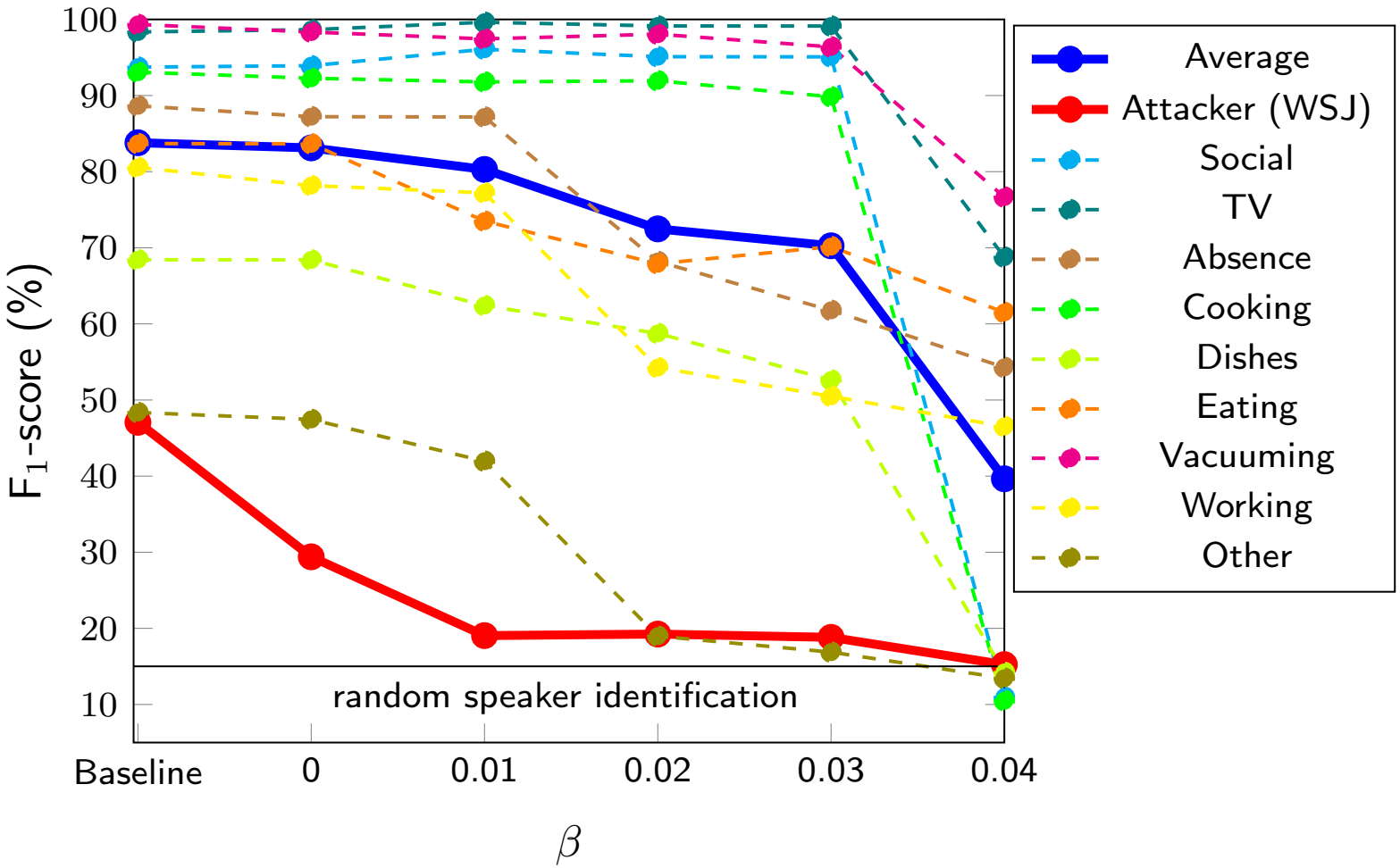
Domestic Activity Detection vs. Speaker Identification



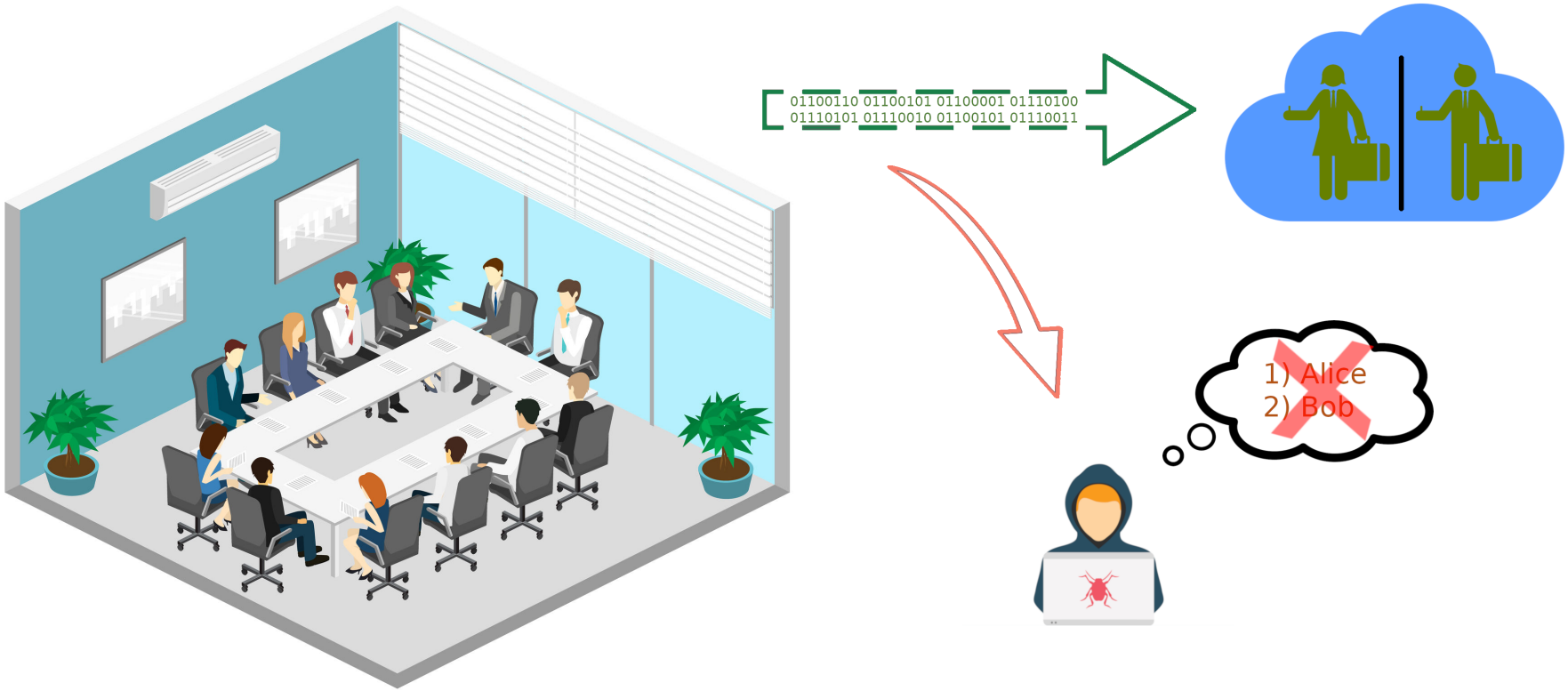
Network Architecture



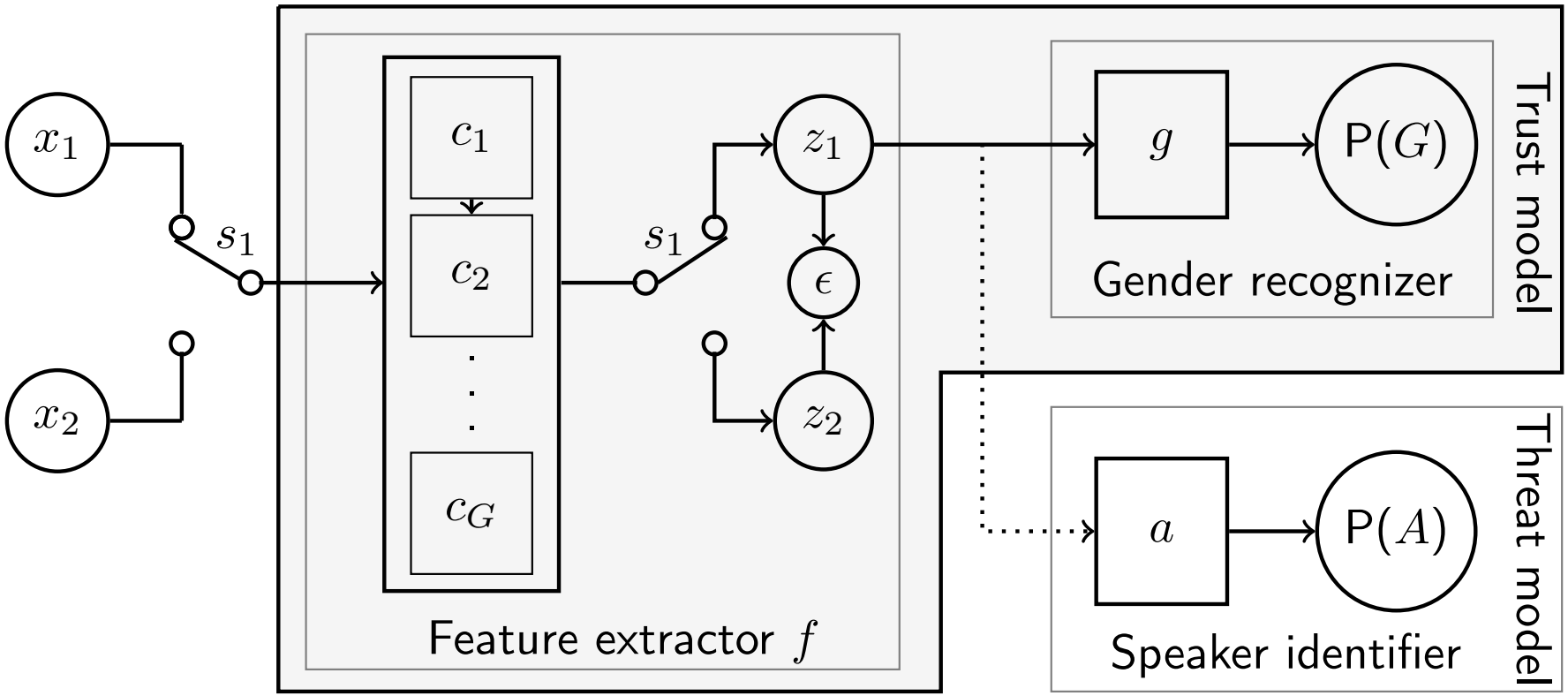
Results


100
90
80
70
60
50
40
30
20
10
F_1-score (%)
random speaker identification
Baseline
0
0.01
0.02
0.03
0.04
β
Average
Attacker (WSJ)
Social
TV
Absence
Cooking
Dishes
Eating
Vacuuming
Working
Other

Gender Recognition vs. Speaker Identification using Siamese Training

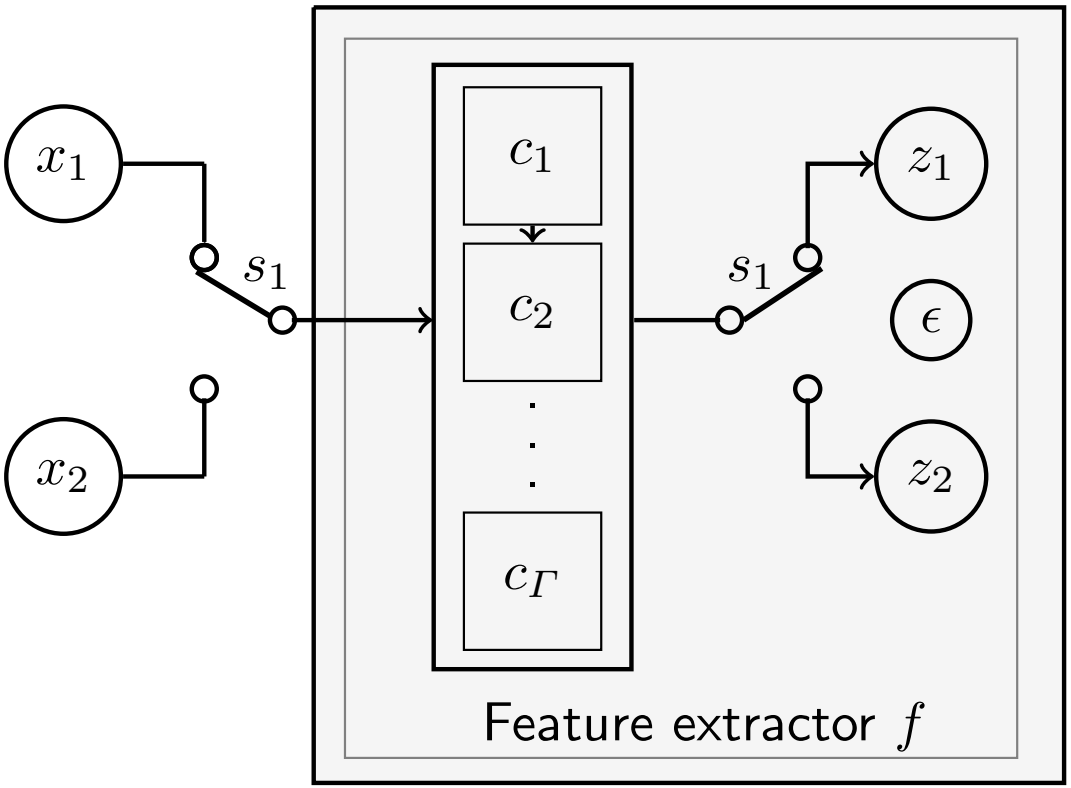


Trust Model vs. Threat Model



- Utility: accuracy gain of gender recognition
- Privacy: accuracy loss of speaker identification

Stage I: Training the Feature Extractor

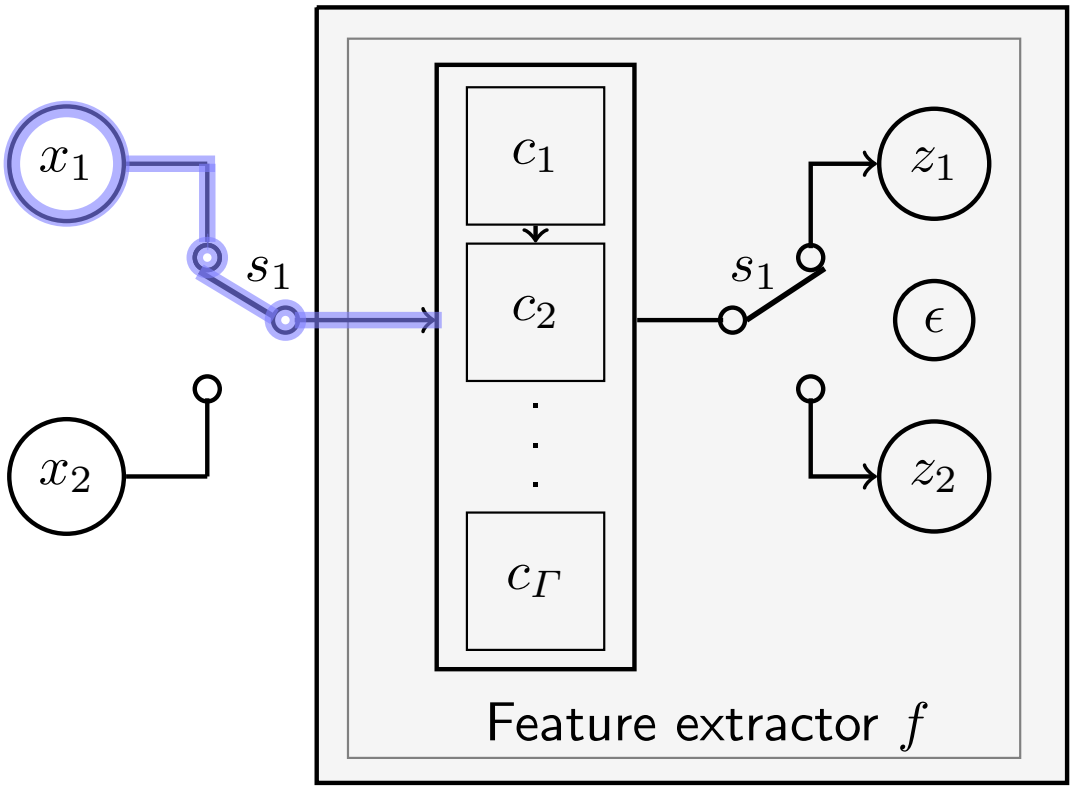


- Group input vectors into similar and dissimilar pairs

$$\begin{cases} (x_1, x_2) \in \{(m, m), (f, f)\} \\ (x_1, x_2) \in \{(m, f), (f, m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

Stage I: Training the Feature Extractor

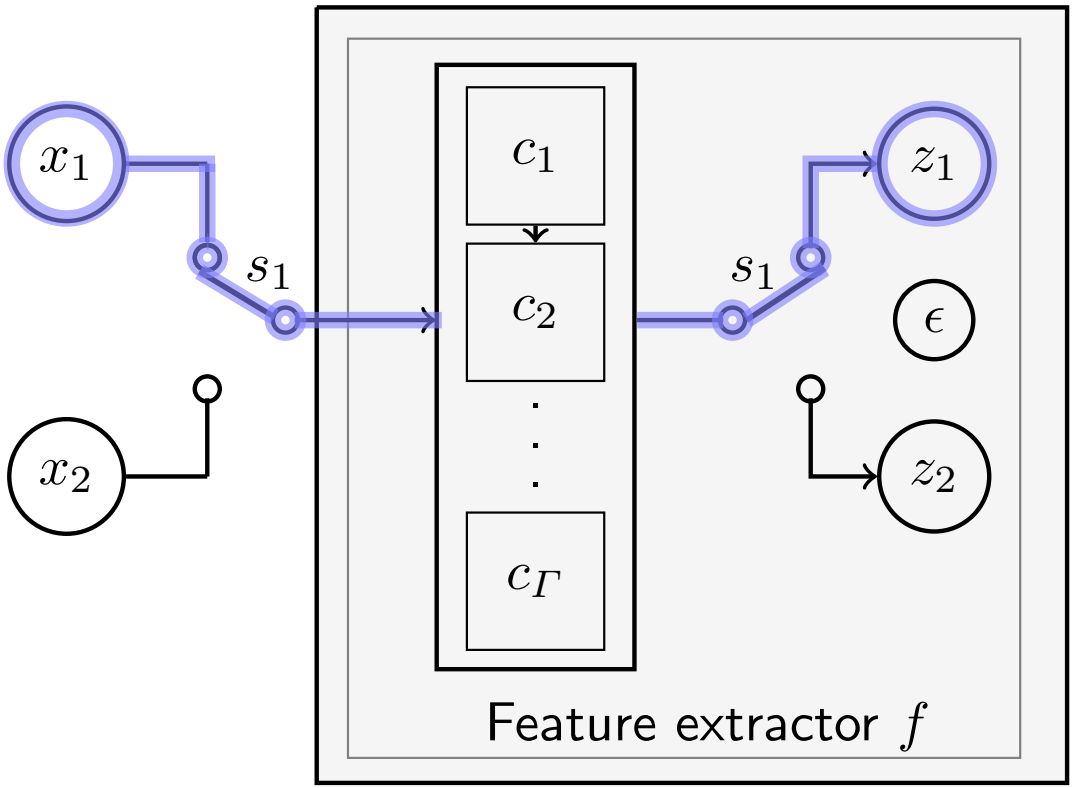


- Group input vectors into similar and dissimilar pairs

$$\begin{cases} (x_1, x_2) \in \{(m, m), (f, f)\} \\ (x_1, x_2) \in \{(m, f), (f, m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

Stage I: Training the Feature Extractor

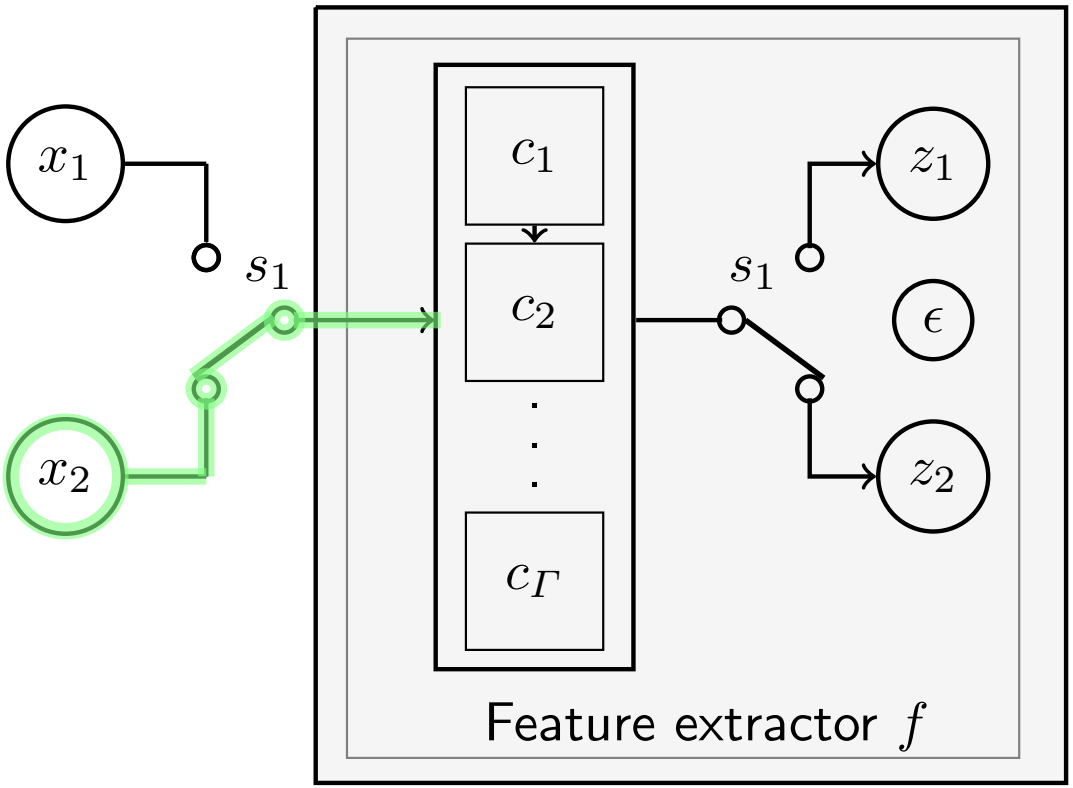


- Group input vectors into similar and dissimilar pairs

$$\begin{cases} (x_1, x_2) \in \{(m, m), (f, f)\} \\ (x_1, x_2) \in \{(m, f), (f, m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

Stage I: Training the Feature Extractor

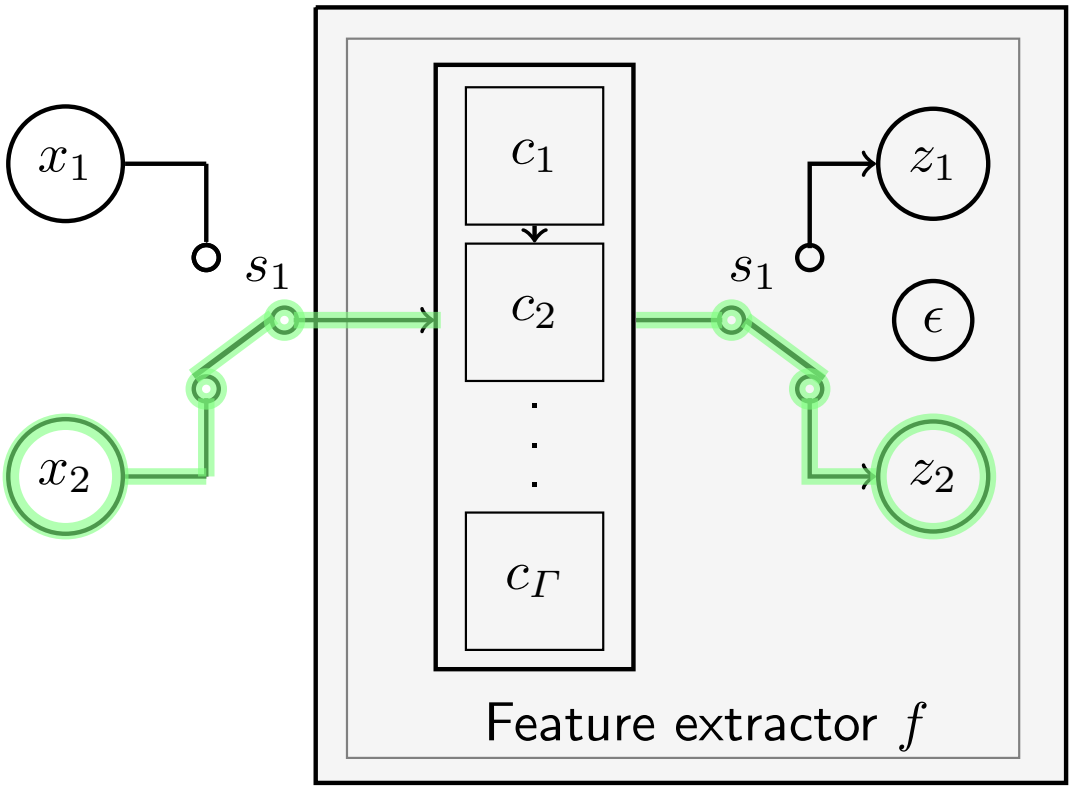


- Group input vectors into similar and dissimilar pairs

$$\begin{cases} (x_1, x_2) \in \{(m, m), (f, f)\} \\ (x_1, x_2) \in \{(m, f), (f, m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

Stage I: Training the Feature Extractor

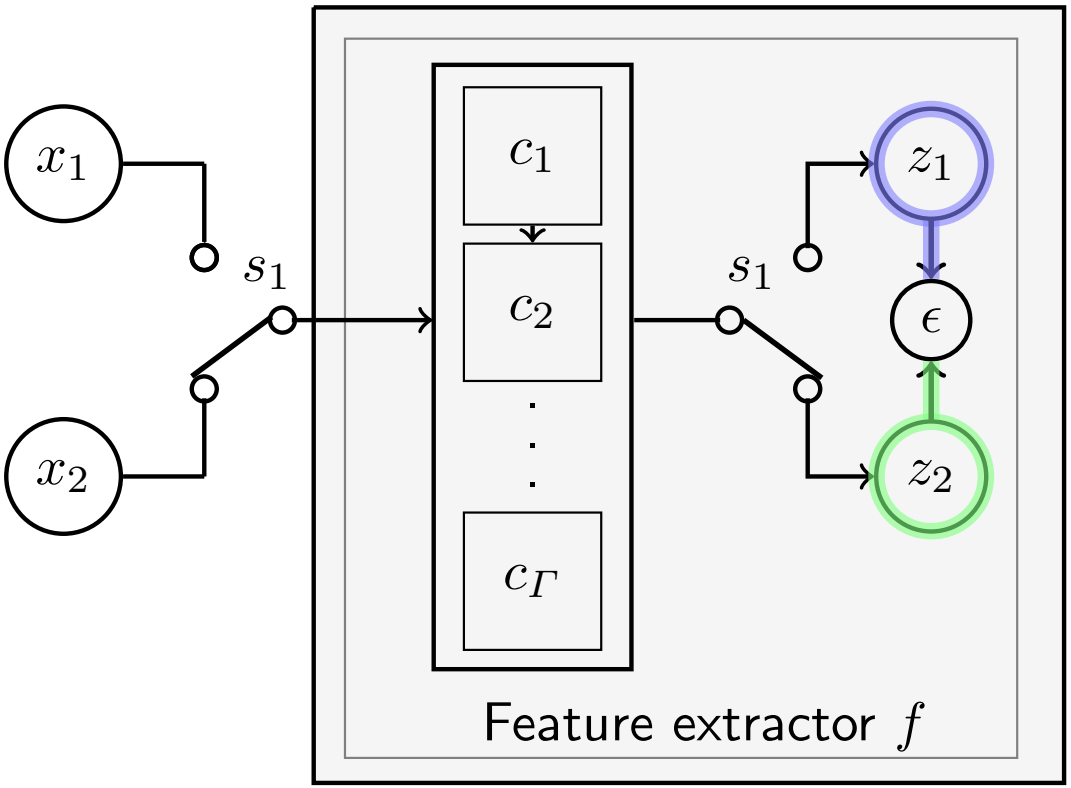


- Group input vectors into similar and dissimilar pairs

$$\begin{cases} (x_1, x_2) \in \{(m, m), (f, f)\} \\ (x_1, x_2) \in \{(m, f), (f, m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

Stage I: Training the Feature Extractor



- Group input vectors into similar and dissimilar pairs

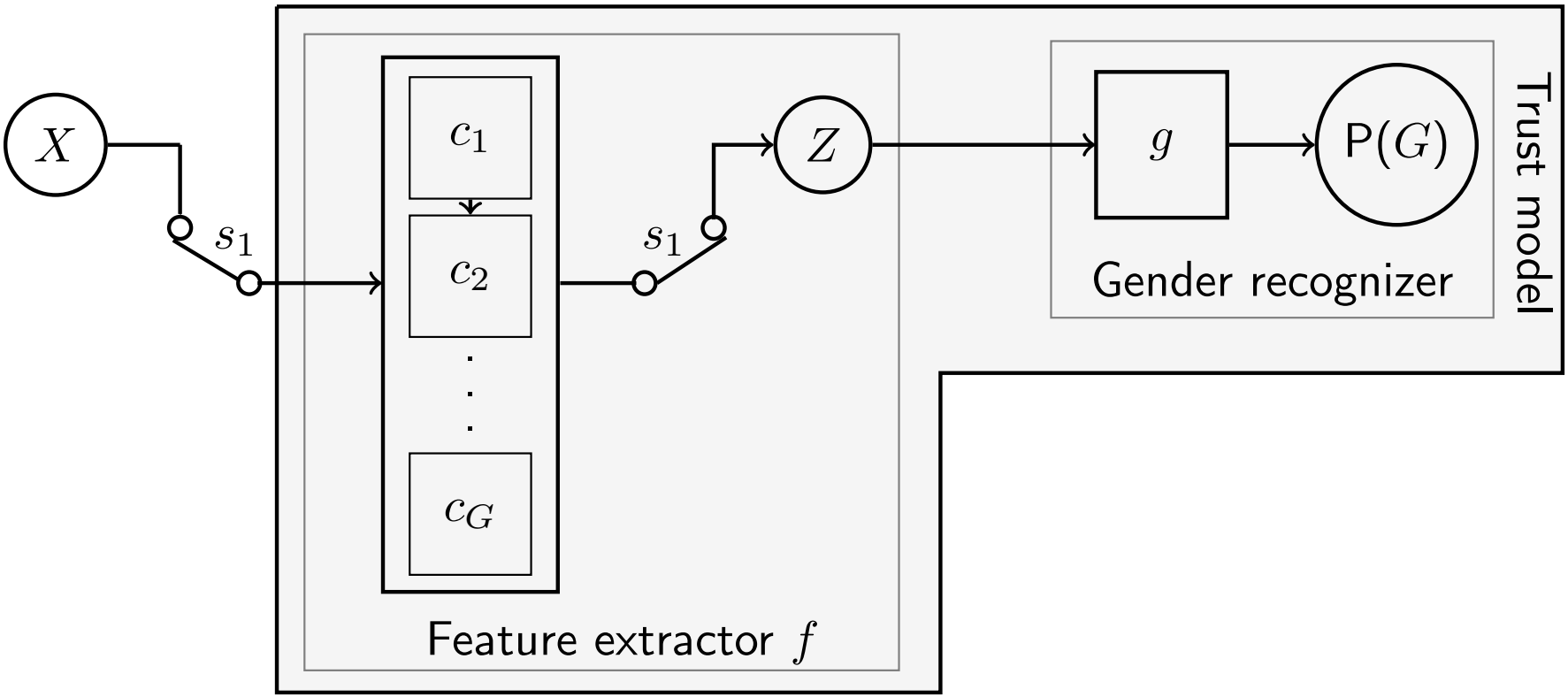
$$\begin{cases} (x_1, x_2) \in \{(m,m),(f,f)\} \\ (x_1, x_2) \in \{(m,f),(f,m)\} \end{cases}$$

- Map $x_1 \rightarrow z_1$ & $x_2 \rightarrow z_2$
- Compute $\epsilon = \|z_1 - z_2\|_2$

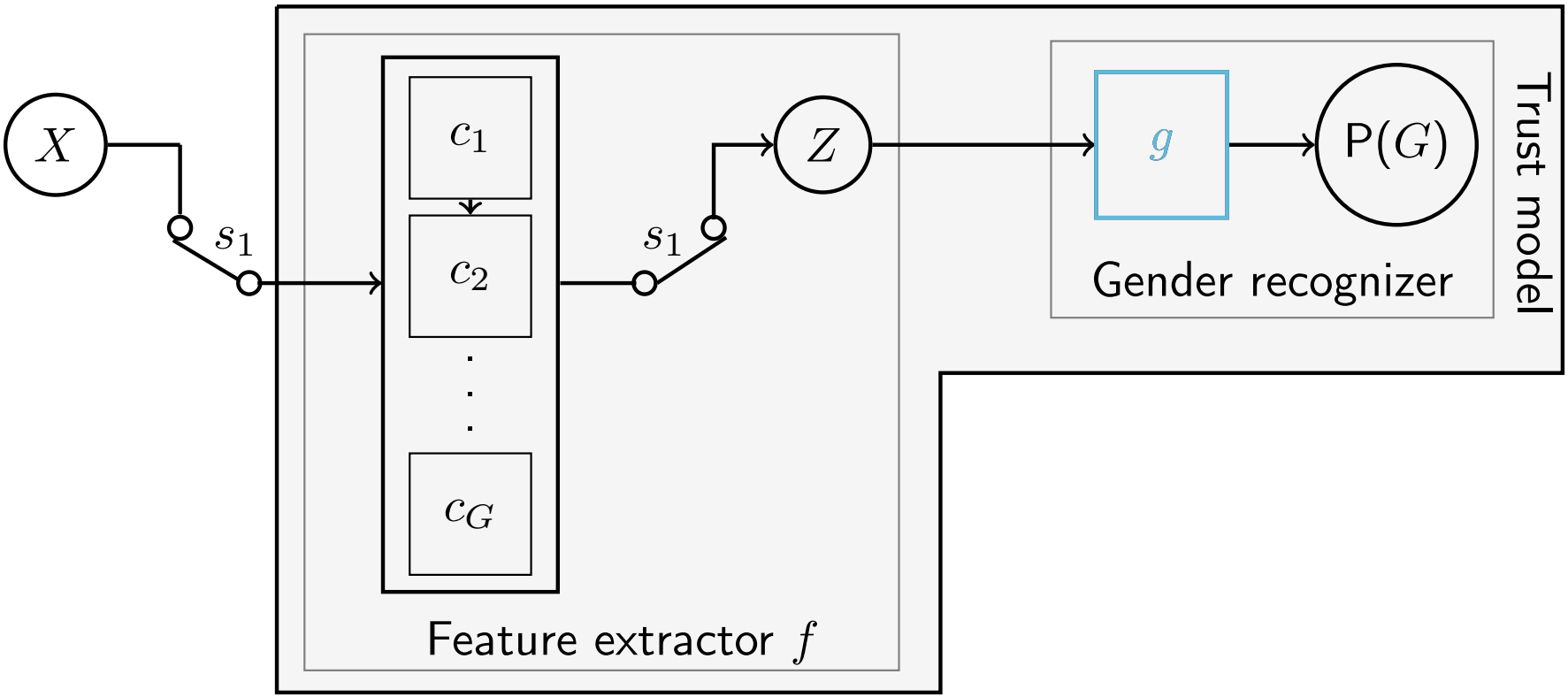
- $\min_{\Phi_c} \mathbb{E}_{Y \sim p(Y)} \left[(1-Y)\epsilon^2 + Y(\max(0, marg - \epsilon)^2) \right]$ where

$$Y = \begin{cases} 0, (z_1, z_2) \in \{(m,m),(f,f)\} \\ 1, (z_1, z_2) \in \{(m,f),(f,m)\} \end{cases}$$

Stage II: Training the Gender Recognizer

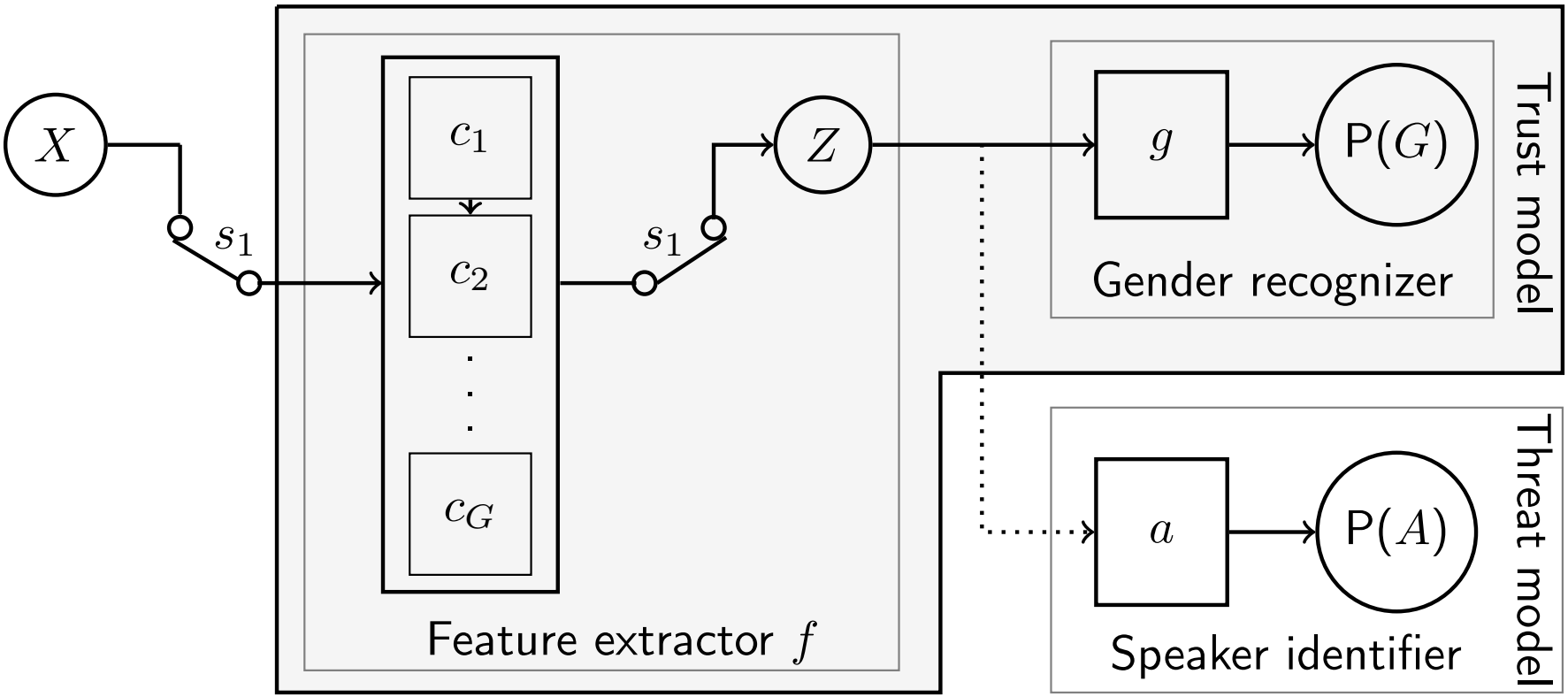


Stage II: Training the Gender Recognizer

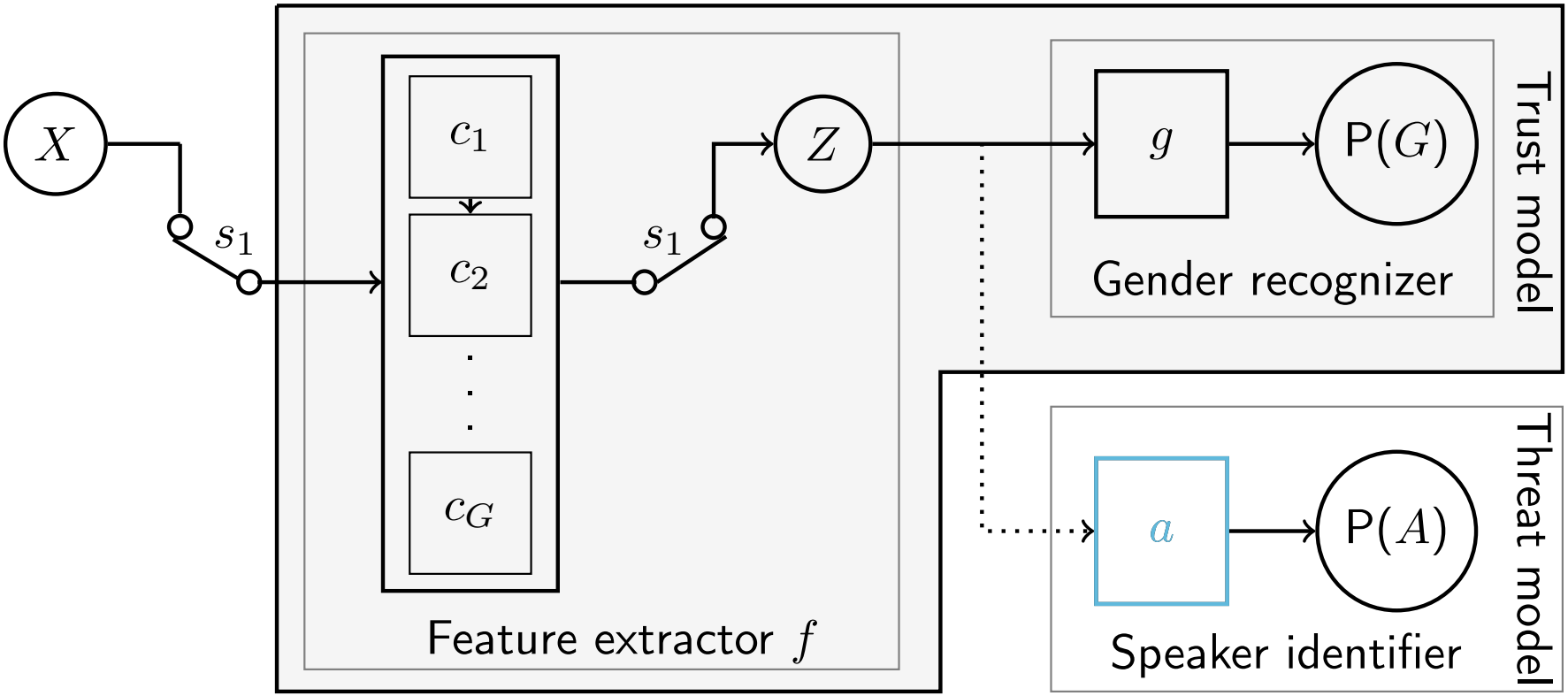


- $\min_{\Phi_g} \mathbb{E}_{G^t \sim p(G^t)}[-\log p(G)]$

Training the Threat Model

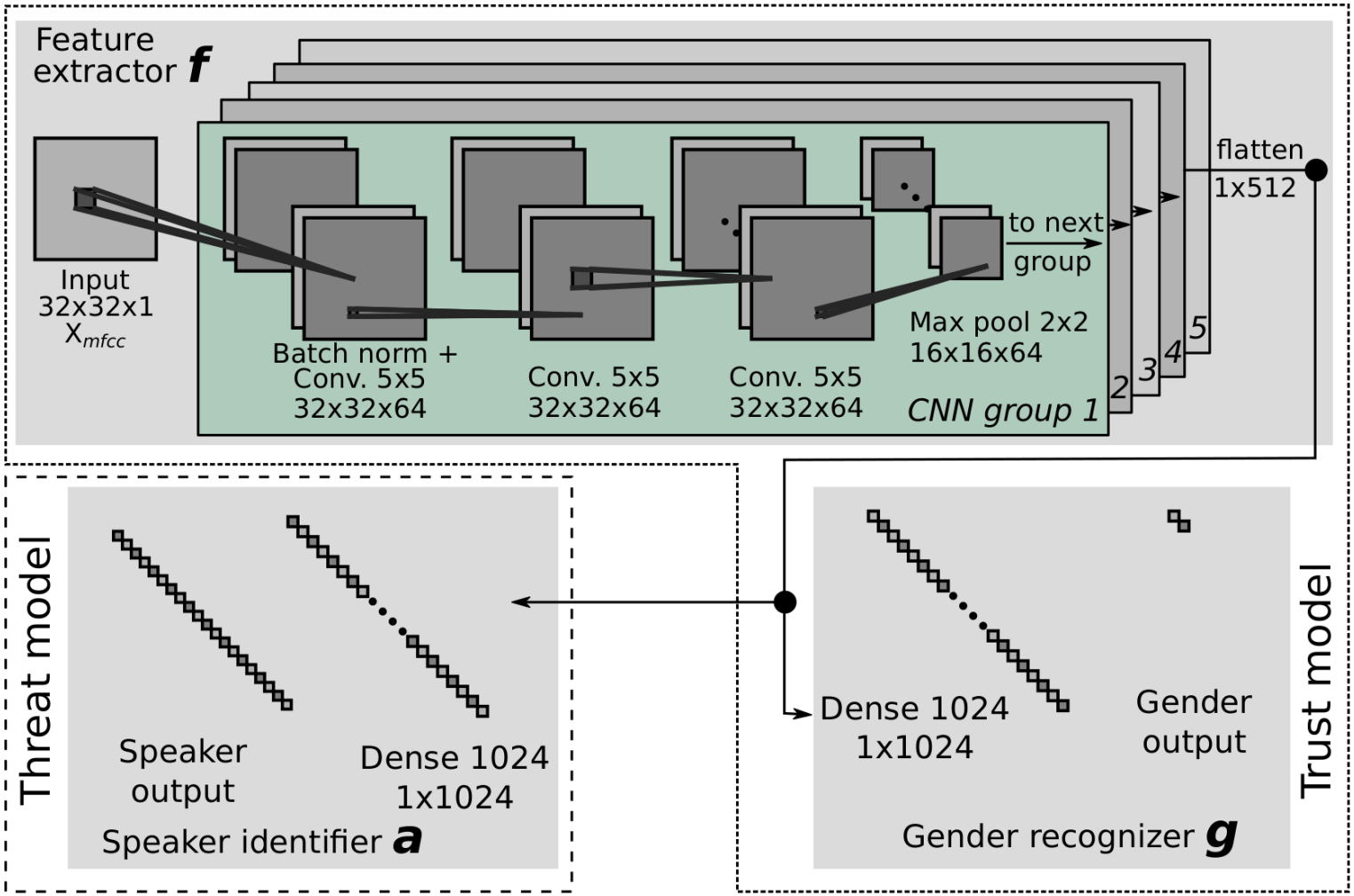


Training the Threat Model

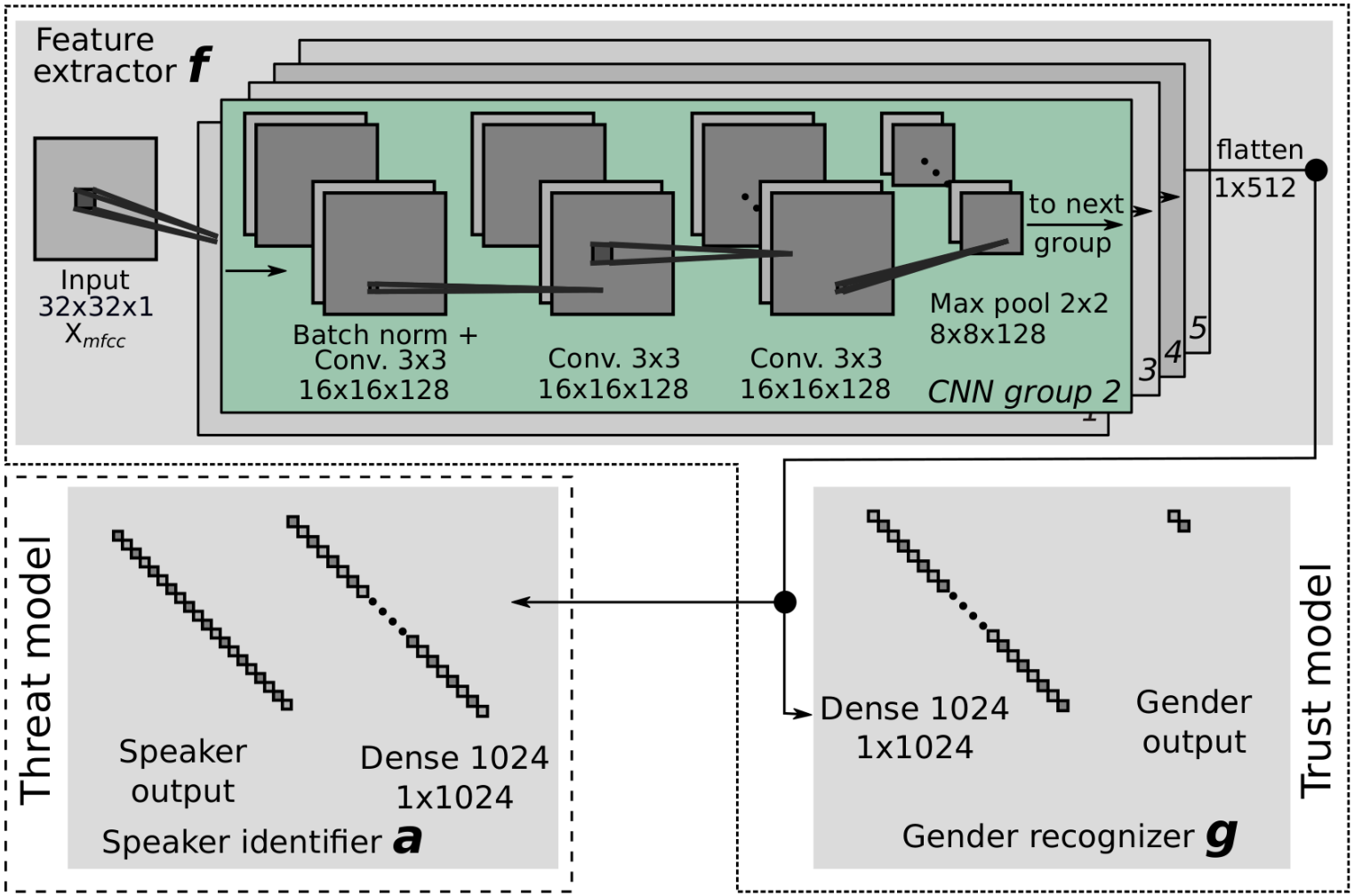


- $\min_{\Phi_a} \mathbb{E}_{A^t \sim p(A^t)}[-\log p(A)]$

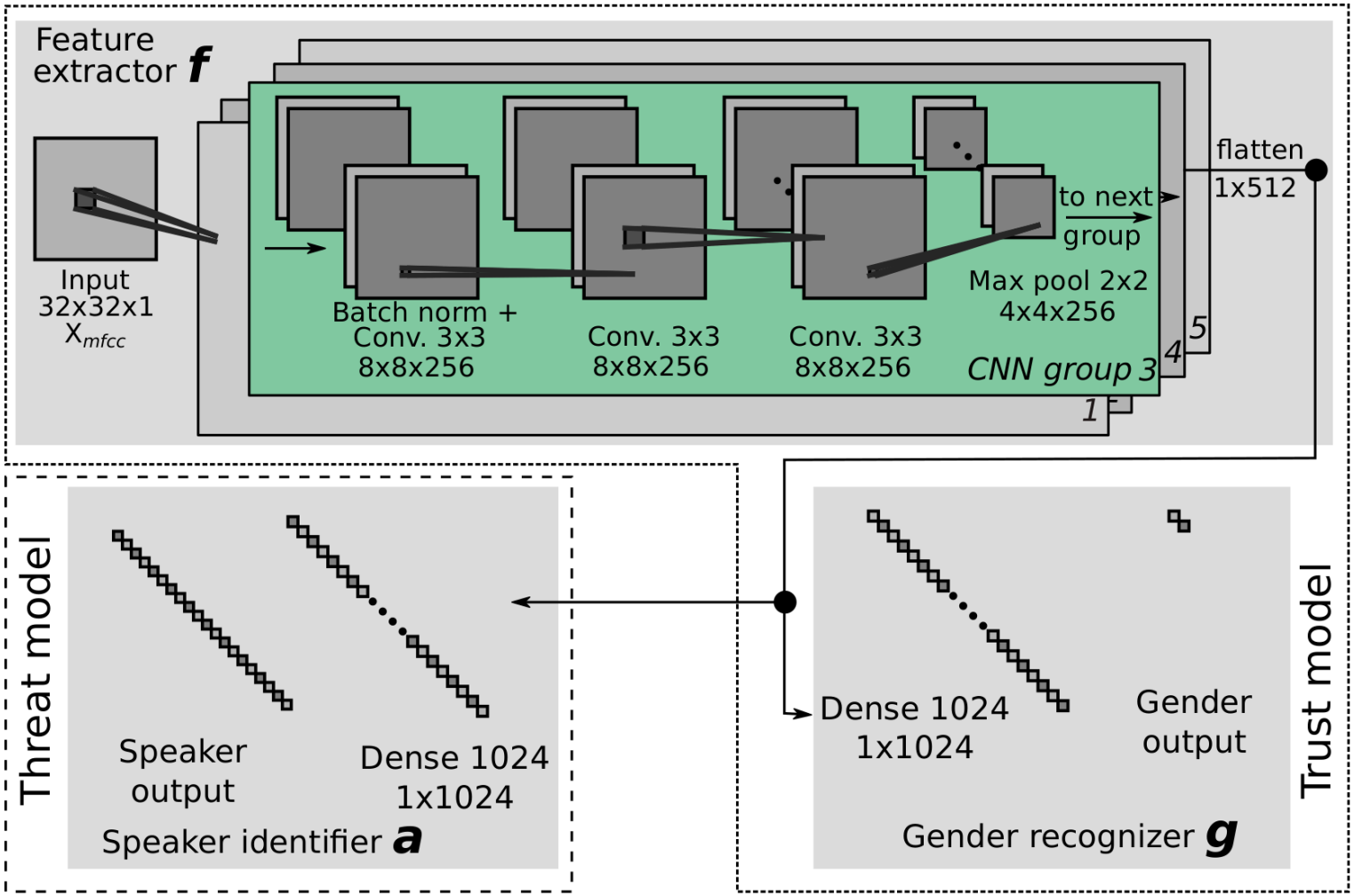
Network Architecture



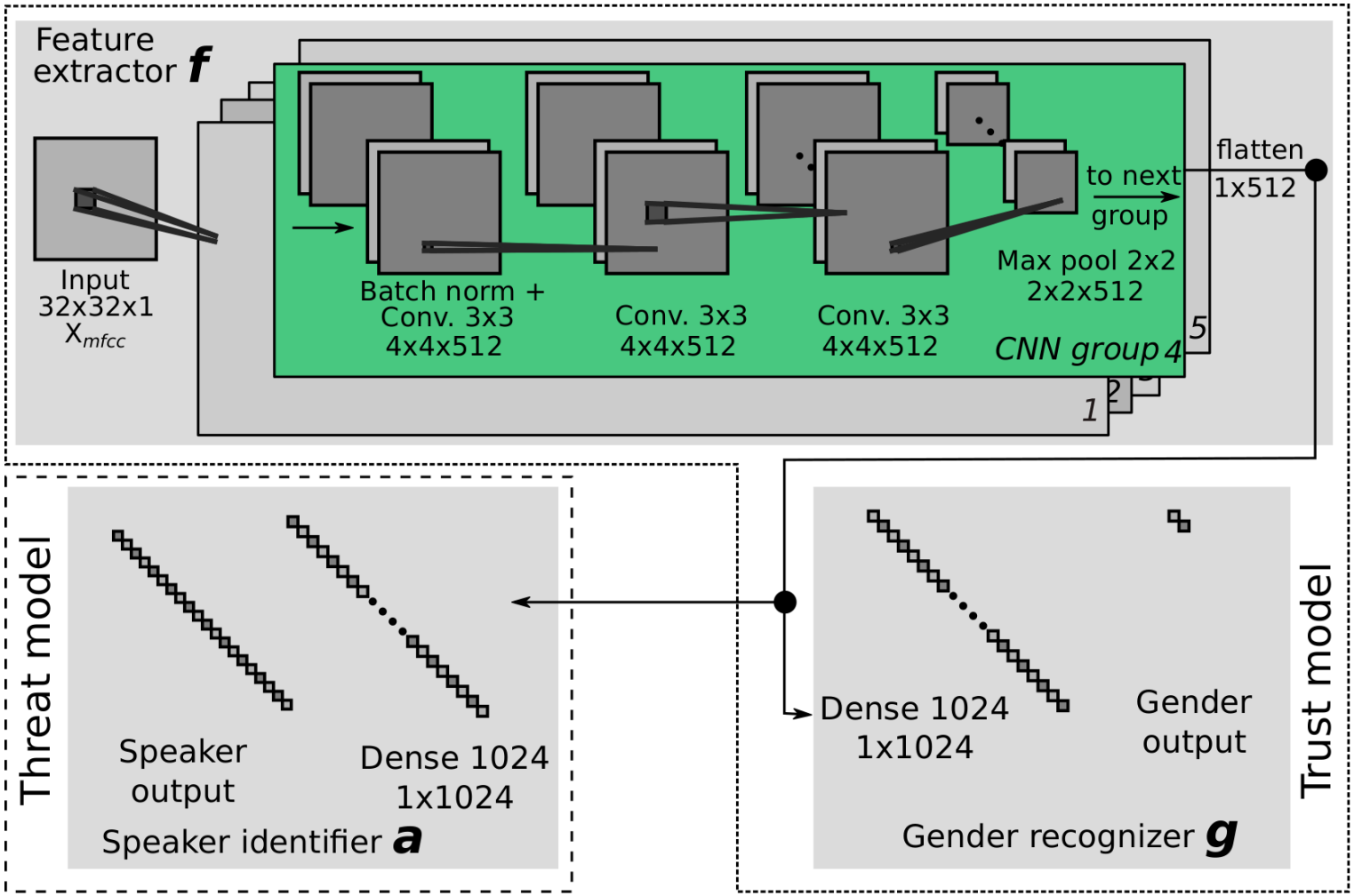
Network Architecture



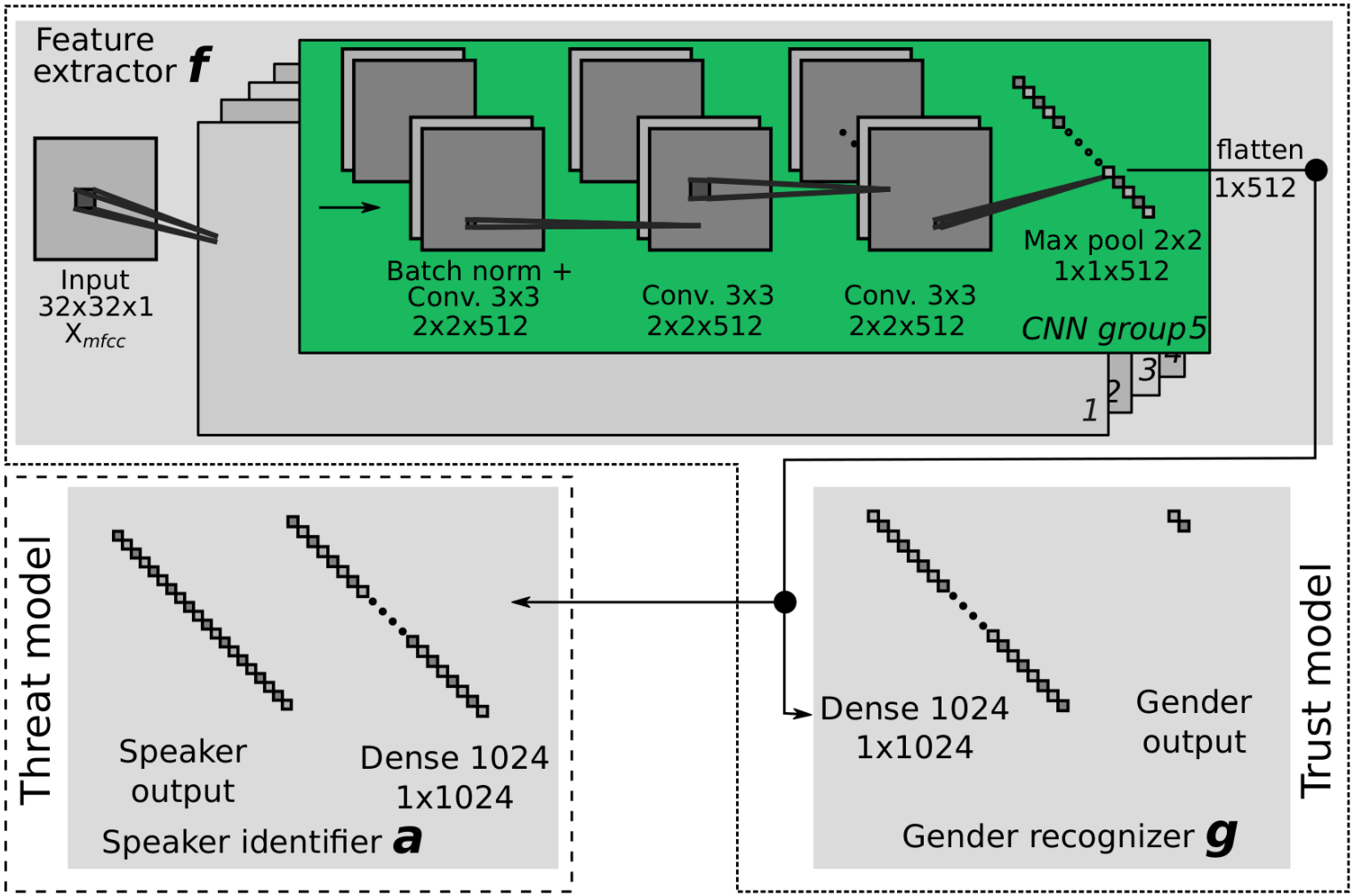
Network Architecture



Network Architecture

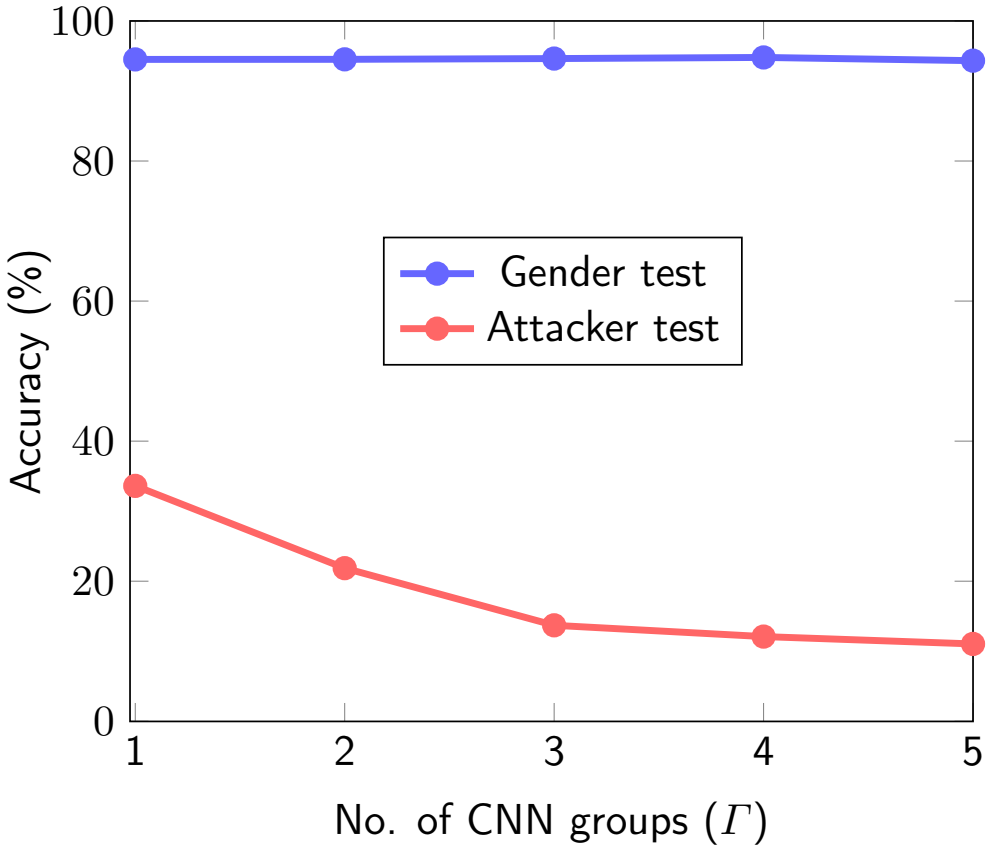


Network Architecture



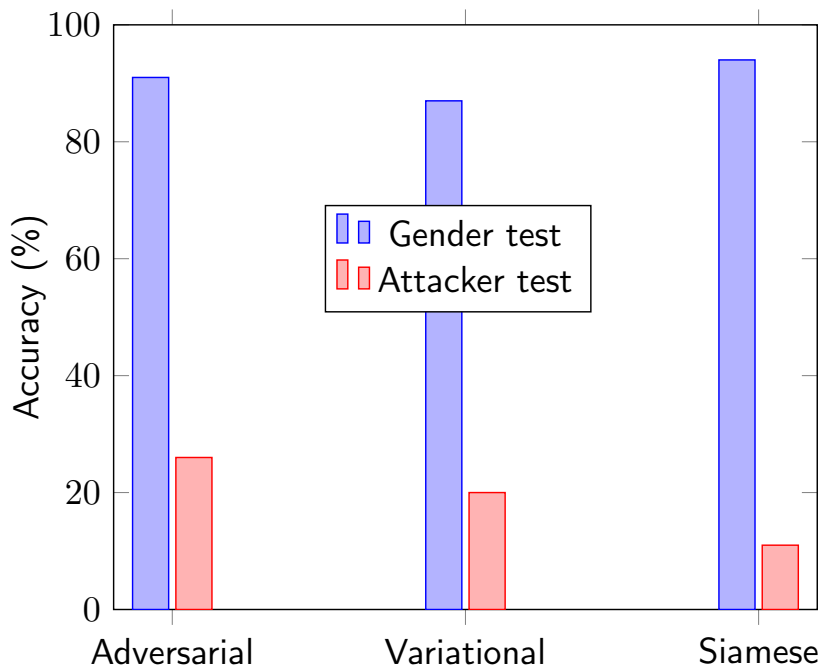
Results

- Gender recognizer:
  - train and test on TIMIT
- Attacker:
  - train and test on TIMIT
  - use 20 speaker groups
- Results aggregated over 10 cross-validations

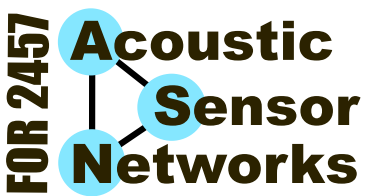


Conclusions

- Proposed a privacy-aware feature extraction framework:
  - variational information, siamese and adversarial
- Siamese approach offers better privacy vs. utility trade-off:
  - direct action on feature-level euclidean distances
  - higher level of feature specialization
- Preference for variational information approach:
  - versatile across ASN applications
  - robust against various attackers
  - upper bound guarantee on the mutual information level between high-level and raw feature representation



Thank you!



Acknowledgment
This work has been supported by the German Research Foundation (DFG) - Project Number 282835863.

References

1. Andreas Nautsch, Abelino Jiménez, Amos Treiber, Jascha Kolberg, Catherine Jasserand, Els Kindt, Héctor Delgado, Massimiliano Todisco, Mohamed Amine Hmani, Aymen Mtibaa, Mohammed Ahmed Abdelraheem, Alberto Abad, Francisco Teixeira, Driss Matrouf, Marta Gomez-Barrero, Dijana Petrovska-Delacrétaz, Gérard Chollet, Nicholas W. D. Evans, Christoph Busch. *Preserving privacy in speaker and speech characterisation. Comput. Speech Lang. 58: 441-480 (2019)*
2. Alexandru Nelus, Sebastian Gergen, Jalal Taghia, and Rainer Martin. *Towards opaque audio features for privacy in acoustic sensor networks.* In Speech Communication; Proceedings of 12. ITG Symposium, pages 1-5. VDE, 2016.
3. Alexandru Nelus, Sebastian Gergen, and Rainer Martin. *Analysis of temporal aggregation and dimensionality reduction on feature sets for speaker identification in wireless acoustic sensor networks.* In Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on, pages 1-6. IEEE, 2017.
4. Alexandru Nelus and Rainer Martin. *Privacy-aware feature extraction for gender discrimination versus speaker identification.* In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2019.

References

5. Alexandru Nelus, Janek Ebbers, Reinhold Haeb-Umbach, and Rainer Martin. *Privacy-preserving variational information feature extraction for domestic activity monitoring versus speaker identification.* In Proceedings of Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, Graz, Austria, 15-19 September 2019.
6. Alexandru Nelus and Rainer Martin. *Gender discrimination versus speaker identification through privacy-aware adversarial feature extraction.* In Speech Communication; Proceedings of 13. ITG Symposium. VDE, 2018.
7. Sebastian Gergen, Rainer Martin, and Nilesh Madhu. *Source separation by feature-based clustering of microphones in ad hoc arrays.* In 16th International Workshop on Acoustic Signal Enhancement (IWAENC), pages 530-534, 2018.
8. Alexandru Nelus, Silas Rech, Timm Koppelmann, Henrik Biermann, and Rainer Martin. *Privacy-preserving siamese feature extraction for gender recognition versus speaker identification.* In Proceedings of Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, Graz, Austria, 15-19 September 2019.