Privacy Impact Assessment (PIA) Report – COP-MODE

Carlos Ortega- up202303651

David Sá- up202303580

João Morais- up202307077

Index

1. Context

1.1 Overview

1.2 Data, processes and supporting assets

2. Fundamental Principles

2.1 Proportionality and necessity

2.2 Controls to protect the personal rights of data subjects

3. Risks

3.1 Planned or existing measures

- Illegitimate access to data
- Unwanted modification of data
- Data disappearance
- Risks overview
- Risk mapping with measures and without measures

4. Conclusion

1. Description of the COP-MODE System

1.1. Project Overview

COP-MODE is a project that aims to obtain rich datasets that map privacy preferences with user context and generate and release to the community privacy profiles that can be used for further research on privacy enhancing methodologies for mobile devices.**System Overview and Design**

1.2. Data, processes and supporting assets

1.2.1. Roles and Responsibilities

- **Data Protection and Privacy:**
  Only essential data is collected, such as email, consent date, and the list of installed applications, without ever accessing personal data contained within the applications.
  It is crucial to ensure that the information collected is used solely for the defined purposes (such as configuring the smartphones for the campaign) andHere is your risk matrix and impact analysis translated into English, following the CNIL methodology and using clear, formal language suitable for a report:
- **Informed Consent**:
  Ensure that participants are fully informed about what will be collected and how it will be used before consenting to participate.
  Obtain explicit consent for data collection and use, which is done through the data collection agreement participants sign when receiving their smartphones.
- **Data Security**:
  Implement appropriate measures to protect participants' data from unauthorized access, loss, or destruction.
  Ensure that data sent to the server is transmitted and stored securely.
- **Anonymization and Data Deletion**:
  Remove personal identifiers (such as email) at the end of the campaign to anonymize the remaining data.
  Ensure that personal data is not retained longer than necessary for the research purposes.
- **Transparency and Access**:
  Inform participants about how their data will be used, how they can access it, and how they can request its deletion.
  Allow participants to access information about which data is being collected and how it is processed.

Additionally, the COP-MODE team must commit to providing any necessary technical assistance during the campaign, ensuring continued protection and compliance with participants' rights.

This campaign was approved by the Ethics Committee of the Department of Computer Science and Technology at the University of Cambridge and by the Ethics Committee of the Faculty of Sciences at the University of Porto.

Data sharing by the COP-MODE team is restricted to academic partners and intended exclusively for research purposes. The shared data will be pre-processed to ensure it is sanitized and anonymized, guaranteeing user privacy. Only specific types of non-sensitive data, such as connectivity type, device context, and nearby devices, will be included in the shared dataset.

Researchers interested in accessing this dataset must sign an agreement that includes strict terms prohibiting further sharing, sale, or malicious use of the data. In addition, the data must be stored and handled according to best practices in security and privacy.

It is important to note that a participant may request the deletion of their data at any time, and this request must be honored by all data holders, who are required to erase the information as requested.

1.2.2Personal Data Collected

The following categories of personal data are processed in the COP-MODE campaign:

- **Contact Data**

  ***Data Type*:** Email address.

  ***Purpose*:** Main form of communication with participants.

  ***Retention Period*:** Until the "return" of the smartphone, after which it should be deleted unless there is consent to keep it longer for future communications or if legal requirements demand its retention for a longer period.

- **Device Data**

  ***Data Type*:** Installed applications and their permission settings.

  Type of connection (WiFi, etc.).

  Device context (inactive, in use, on a call, etc.).

  ***Purpose*:** Analysis of device usage and context for research purposes.

  ***Retention Period*:** Until the conclusion of data analysis for the project's purposes, after which it must be anonymized or deleted unless specific regulations require longer retention.

- **Location Data**

  ***Data Type*:** Geographic location.

  ***Purpose*:** Tracking movements and patterns for analysis within the project.

  ***Retention Period*:** Until the completion of relevant analyses, followed by anonymization or deletion.

- **Proximity Data**

  ***Data Type*:** Nearby devices.

  ***Purpose*:** Understanding social or environmental interactions.

  ***Retention Period*:** Should be limited to the period necessary for analyzing the collected data.

- **Application Interaction Data**

  ***Data Type*:** Information from applications at the time of permission prompts.

  User decisions regarding permissions.

  Semantic location (user input).

  ***Purpose*:** Understanding user behavior and privacy decisions. No internal data from the applications will be collected.

  ***Retention Period*:** Until the evaluation of user behavior is complete, after which it will be anonymized or deleted.

1.2.3 Lifecycle of personal data and the related processes

Personal data is collected when participants install the COP-MODE Apps Retriever (CM-AR) application on their smartphones and provide consent. The app gathers device information, installed applications and their permissions, location data, nearby devices, and user decisions regarding app permissions.

Collected data is transmitted securely from the participant's smartphone to the COP-MODE server using encrypted channels such as HTTPS or VPN to prevent interception.

Data is stored in a secure database accessible only to authorized COP-MODE team members. Processing involves analyzing user behavior, application permissions, and device context through automated scripts and manual review.

Access to personal data is restricted by role-based access controls, with audit logs maintained to monitor data access and modifications.

At the conclusion of the research campaign or upon participant request, all personal data is either anonymized or permanently deleted in accordance with data protection policies and participant consent

1.2.4 What are the data supporting assets?

The COP-MODE project relies on a combination of hardware, software, and communication tools to collect, transmit, store, and process personal data securely.

- **Smartphones:**
  Data collection occurs on participants' personal smartphones via the COP-MODE Apps Retriever (CM-AR) app and on dedicated test devices equipped with the COP-MODE Naive

Permission Manager (CM-NPM) app. The personal device is used for initial data gathering and consent, while the test device monitors permission requests during the study.

- **Applications:**

  CM-AR collects device information, installed apps, and user consent.

  CM-NPM records user responses to permission prompts throughout the campaign.

- **Server and Database:**
  Data from both applications is transmitted over encrypted channels (HTTPS/VPN) to a secure COP-MODE server. The server hosts a protected database where data is stored, processed, and analyzed. Access is limited to authorized personnel under strict security policies.

- **Communication and Documentation:**
  Participant communication is managed via email for instructions, consent, and support. Occasionally, signed consent forms are handled in physical format and securely stored.

- **Security Measures:**
  All systems enforce encryption, role-based access controls, audit logging, and regular backups to maintain data confidentiality and integrity throughout the project lifecycle.

2. Fundamental Principles

2.1 Proportionality and Necessity

The purpose of data processing in the COP-MODE project is clear, well-defined, and justified.
It is specific because it is solely aimed at analyzing users' privacy choices on mobile devices during controlled data collection campaigns.
It is explicit since participants are clearly informed about the types of data being collected (e.g., app usage patterns, device context, and consent information) and the purposes of the collection.
It is legitimate because the data is processed for scientific research purposes, following ethical guidelines and with approval from academic ethics committees.

- **What is the legal basis for processing personal data?**

  The legal basis for processing personal data in the COP-MODE project is the **explicit consent** of participants.
  Data is only collected after participants are fully informed and have signed a consent form confirming their voluntary, informed, and specific agreement, in accordance with ethical standards and applicable regulations such as the GDPR.

- **Are the personal data collected adequate, relevant, and limited to what is necessary in relation to the purpose (data minimization principle)?**
  Yes, the personal data collected complies with the principle of data minimization, being

strictly limited to what is necessary for the study.
Only essential information is gathered, such as email for contact, installed applications, permission settings, and device context, avoiding the collection of any excessive or irrelevant data.

- **Are the personal data accurate and kept up to date?**
  Yes, the data collected is accurate and up to date, as it is either provided directly by participants or collected automatically during the smartphone usage period.
  Ongoing communication with participants (via email) also allows corrections or updates when necessary.

- **What is the data retention period?**
  Data is retained only for the duration of the data collection campaign.
  Personal information such as email addresses is deleted once the participant returns the device or upon request.
  All other data is anonymized after the necessary analysis for the research is completed

2.2. Controls to protect the personal rights of data subjects

The COP-MODE project is designed to study individual privacy decisions regarding app permissions on smartphones. All participants are thoroughly informed about the nature and purpose of data processing before joining the study. Information is provided both via email and in person, detailing what data will be collected-such as email address, installed applications, permissions, device context, and location data-how it will be used, and the robust privacy and security measures implemented. Participation is entirely voluntary, and explicit, informed consent is obtained through a signed data collection agreement when the participant receives the test device.

Throughout the project, participants are empowered to exercise their data protection rights at any time. They may request access to their personal data or data portability by contacting the COP-MODE team, who will provide the requested data in a structured, machine-readable format. Should participants wish to rectify inaccuracies or request erasure of their data, the team will promptly update or delete all relevant information and confirm the action. Participants also have the right to restrict processing or object to certain uses of their data; any such request will be respected, and processing will be limited or ceased as appropriate.

All personal data is processed exclusively for research purposes and only for the duration of the campaign. Data is anonymized at the end of the study, with personal identifiers such as email addresses removed to ensure privacy. The retention of data never exceeds what is necessary for the research objectives, and participants can request deletion of their data at any stage.

The obligations of all data processors involved in COP-MODE are clearly defined and governed by formal contracts. These contracts specify the scope, duration, and purpose of processing, as well as security and confidentiality requirements. Data sharing is strictly limited to academic partners and only involves anonymized datasets. Any external researcher accessing the data must sign a non-disclosure agreement and adhere to strict data protection standards.

All data is processed and stored within the European Union. In the event that data needs to be transferred outside the EU, COP-MODE will ensure that such transfers only occur to countries recognized by the European Commission as providing an adequate level of data protection, or will implement appropriate safeguards, such as Standard Contractual Clauses, to maintain the highest standards of privacy and security.

In summary, COP-MODE is committed to upholding the highest standards of data protection, ensuring that participants are fully informed, their rights are respected, and their data is handled with transparency, security, and integrity throughout the entire research process.

Feel free to copy and paste this text directly into your form. If you need it shortened or tailored further, just let me know!

3.Risks

3.1. Planned or existing measures

Utilization of HTTPS:

All data transmitted between smartphones and the server is encrypted using HTTPS, which employs TLS (Transport Layer Security) to protect data in transit. This ensures that any sensitive information exchanged between the user and the system is safeguarded from potential interception by unauthorized third parties. By using HTTPS, the system ensures that the data remains secure during transmission, preventing attackers from eavesdropping or tampering with the data as it moves across the network.

Utilization of VPNs:

A Virtual Private Network (VPN) is implemented for all data transmissions, adding an extra layer of encryption and security to protect data while in transit. The VPN creates a secure tunnel between the devices and the server, further preventing unauthorized access and ensuring that the data remains confidential. This measure enhances the security of data transfers, especially in situations where devices may be using public or untrusted networks, thus offering an additional level of protection.

Certificate Pinning:

To protect against man-in-the-middle (MITM) attacks, certificate pinning is used within the mobile application. This technique ensures that the app communicates only with the server identified by a specific SSL/TLS certificate, preventing attackers from intercepting or altering the communication between the app and the server. By verifying that the server’s certificate matches the one embedded

in the app, certificate pinning adds an important layer of defense, preventing malicious intermediaries from tampering with the data.

Access Logging:

A comprehensive access logging system is implemented to track all access and modifications to the data. This system allows for the detection of any unauthorized or inappropriate access to sensitive data. By maintaining detailed logs of all activities, administrators can quickly identify suspicious actions, such as unauthorized data access or potential tampering, and take immediate corrective measures. This logging system is essential for ensuring accountability and transparency in how data is handled, ensuring that any breach or misuse can be traced back to its source.

Backup and Recovery:

Regular backup policies are in place to ensure that data is securely backed up at regular intervals. This allows for quick recovery of data in the event of accidental or malicious modification. Additionally, disaster recovery testing is conducted to verify that backup procedures are effective and that data can be restored promptly in case of a system failure or security breach. This measure is crucial in maintaining data integrity and ensuring that the system can quickly recover from unforeseen incidents, minimizing downtime and reducing the impact of data loss.

User Awareness and Transparency:

Participants receive clear information about data handling and their rights, with transparent consent processes.

Data Minimization and Retention:

The principle of data minimization is strictly followed, ensuring that only the minimum amount of personal data necessary for research purposes is collected. This helps reduce the exposure of sensitive data by limiting the amount that is stored or processed. Additionally, clear data retention policies are implemented to ensure that personal data is deleted or anonymized after it is no longer needed for its intended purpose. By restricting the storage duration and minimizing the data collected, the system reduces the potential risks associated with data breaches and unauthorized access.

3.1.1. Illegitimate access to data

What could be the main impacts on the data subjects if the risk were to occur?

- Privacy Violation
- Identity Theft
- Data Integrity Loss

What are the main threats that could lead to the risk?

- Credential Leak
- Unauthorized Access to Server/Data

- Interception of Communications Between Smartphones and the Project Server
- Malware
- Identity Fraud

What are the risk sources?

- External attackers (hackers)
- Internal staff with data access
- Third-party service providers
- Technical vulnerabilities
- Unsecured participant devices

Which of the identified planned controls contribute to addressing the risk?

- Utilization of HTTPS:
- Utilization of VPNs:
- Access Logging:
- Data Minimization and Retention:

3.1.2.Unwanted modification of data

What could be the main impacts on the data subjects if the risk were to occur?

- Incorrect decisions may be made based on altered data, directly affecting the privacy, integrity, and rights of the data subjects.
- Modified data could generate false profiles, incorrect analyses, or misleading conclusions about participants.
- There could be reputational damage, discrimination, or legal actions based on incorrect information.

What are the main threats that could lead to the risk?

- Unauthorized access with the ability to edit or overwrite data.
- Human errors during data entry, processing, or export.
- System synchronization failures between devices or databases.
- Malicious attacks (e.g., SQL injection) that intentionally alter data.

What are the risk sources?

- Internal users with excessive permissions or inadequate training.
- External attackers (hackers)
- Technical failures (e.g., bugs in code or file corruption).
- Third-party service providers

Which of the identified planned controls contribute to addressing the risk?

- Access Logging
- Backup and Recovery
- Utilization of HTTPS
- Utilization of VPNs
- Certificate Pinning
- User Awareness and Transparency
- Data Minimization and Retention

3.1.3. Data disapperance

What could be the main impacts on the data subjects if the risk were to occur?

- Compromised Data Integrity
- Risk of Exposure and Misuse of Personal Data
- Loss of Control over Data

What are the main threats that could lead to the risk?

- Malware
- Credential Leak
- Unauthorized access with the ability to edit or overwrite data.
- Interception of Communications Between Smartphones and the Project Server
- Poor Management of Backup Systems

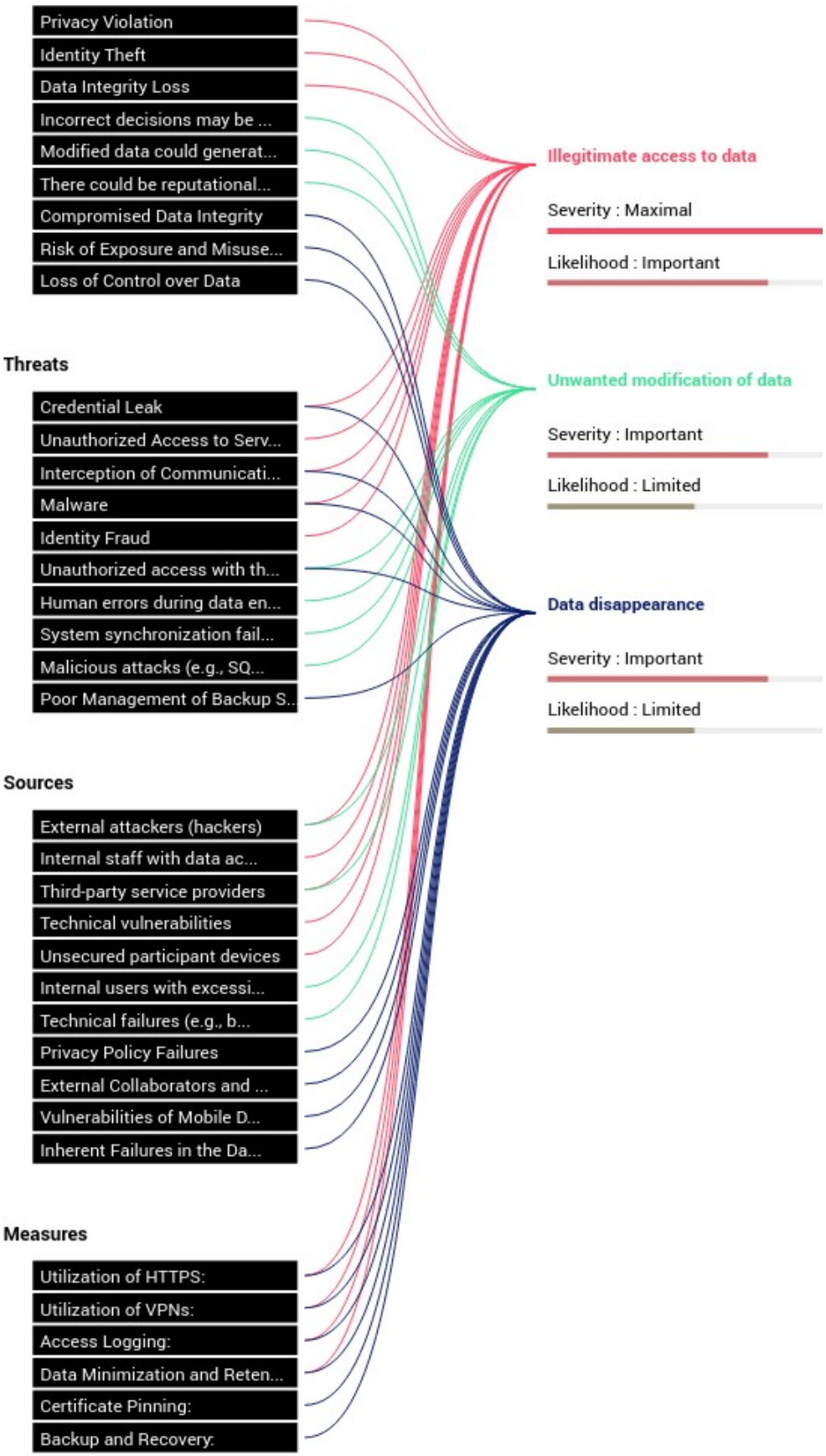
What are the risk sources?

- Privacy Policy Failures
- External Collaborators and Partners
- Vulnerabilities of Mobile Devices
- Inherent Failures in the Data Storage System

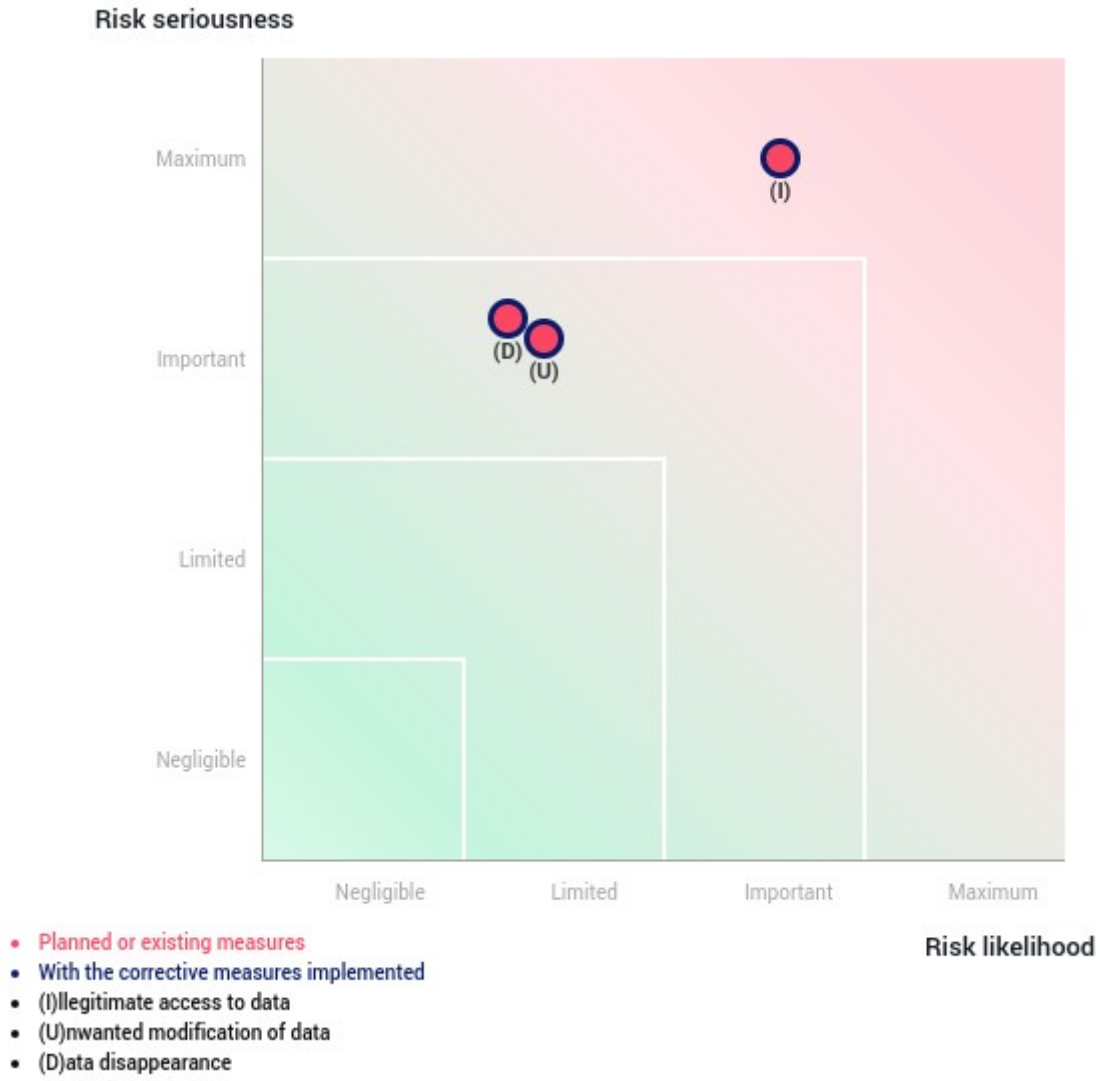
Which of the identified planned controls contribute to addressing the risk?

- Utilization of HTTPS
- Utilization of VPNs
- Certificate Pinning
- Access Logging
- Data Minimization and Retention
- Backup and Recovery

3.1.4. Overview of Risks With the Referred Security Measures



3.1.5. Risk Mapping With the Referred Mitigation Processes



1. Illegitimate Access to Data

- **Likelihood: Important**

  COP-MODE deals with sensitive data related to citizens, law enforcement entities, ongoing investigations, or operations. Without strong access controls, multi-factor authentication, and permission segmentation, there is a significant risk of unauthorized access, whether by external attackers or malicious insiders.

- **Risk Severity: Maximum**

  Unauthorized access to COP-MODE data can have severe consequences: compromise of police operations, risks to officers' lives, violation of citizens' rights, and institutional damage. As it involves public safety, the severity of impact reaches the maximum level.

2. Unwanted Modification of Data

- **Likelihood: Limited**

  Unauthorized data modification depends on improper permissions, validation failures, or lack of integrity mechanisms. In COP-MODE, this risk may be limited if there is some level of monitoring or audit logging — but without robust controls, the possibility remains.

- **Risk Severity: Important**

  Modified data (such as suspect identities, operation schedules, or crime statistics) can lead to operational or legal errors. The impact is important, though lower than full data access or loss, as there may be ways to reverse or correct the alteration.

3. Data Disappearance

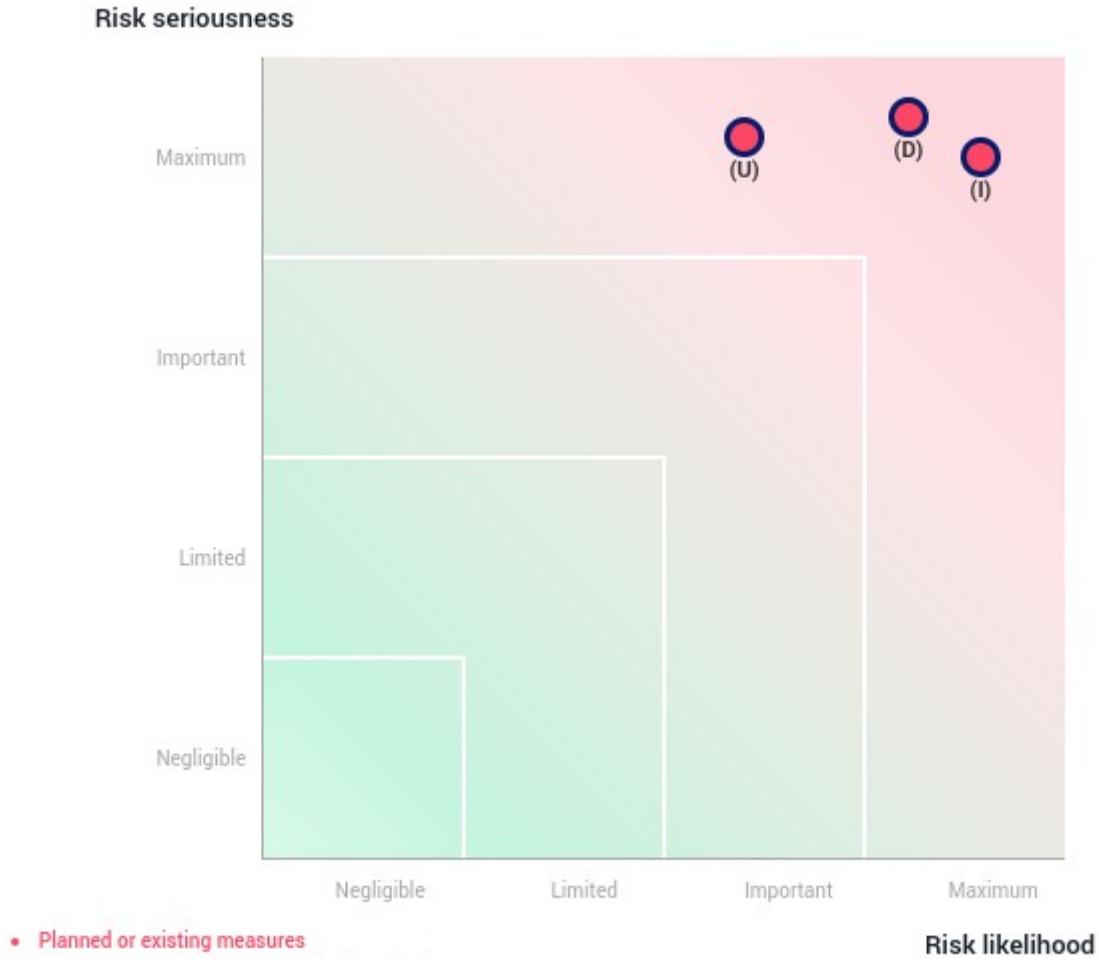
- **Likelihood: Limited**

With regular backups and recovery plans (as mentioned in the image), the likelihood of total data loss is limited. However, if COP-MODE lacks redundancy or offsite backups, the risk could increase.

- **Risk Severity: Important**

  The loss of critical data can compromise investigations, ongoing operations, or legal evidence. Although contingency plans may exist, the impact on security and operational continuity remains important.

3.1.6. Risk Mapping Without the Referred Mitigation Processes



- (I)llegitimate access to data
- (U)nwanted modification of data
- (D)ata disappearance

1. Illegitimate Access to Data

- **Likelihood: Maximum**
  Without any access controls, authentication, or data isolation, the risk of unauthorized access is extremely high, specially if data is stored in plain text or the server is publicly exposed.
- **Risk Severity: Maximum**
  The impact remains critical: unauthorized access could expose operational data, compromise investigations, violate participants' privacy, and result in serious legal and institutional consequences.

2. Unwanted Modification of Data

- **Likelihood: Important to Maximum**
  If there are no integrity checks, permissions, logs, or validation mechanisms, the likelihood of accidental or malicious data changes increases significantly, either through human error or intentional tampering.
- **Risk Severity: Important**
  Data modification can lead to incorrect actions, flawed reports, and legal issues. While

serious, this is not considered "maximum" since such changes could theoretically be manually corrected, though with difficulty.

3. Data Disappearance

- **Likelihood: Maximum**
  Without backups, redundancy, or failure monitoring, data loss due to errors, hardware failures, or attacks becomes very likely.
- **Risk Severity: Maximum**
  Permanent loss of sensitive data (such as operational logs or participant contact details) would critically impact project continuity, public trust, and legal compliance.

3.1.7. Comparison of Risks With and Without Mitigation Measures

The comparative analysis between scenarios with and without mitigation measures highlights a significant difference in risk management within the COP-MODE system. Without implementing controls such as VPNs, HTTPS, certificate pinning, and access logging, risks ,such as unauthorized data access or sensitive information loss , reach maximum levels of severity and likelihood. In contrast, when these security measures are in place, there is a clear reduction in both the probability and impact of the risks, resulting in a safer and more controlled environment. This comparison reinforces the effectiveness of the implemented controls and emphasizes the importance of maintaining and regularly updating them.

3.1.8. Additional Improvement Suggestions

Suggested Measure	Justification
Multi-Factor Authentication (MFA)	Adds an extra security layer to prevent unauthorized access, especially for administrators.
Data Encryption at Rest	Ensures that even if storage is compromised, the data remains unreadable to attackers.
Real-Time Monitoring and Alerts	Enables quick detection and response to suspicious activities or potential breaches.
Regular Penetration Testing	Identifies system vulnerabilities before they can be exploited by malicious actors.
Continuous Staff Training	Keeps all team members updated on best practices in security and privacy protection.

Conclusion

Conducting this Privacy Impact Assessment (PIA) for the COP-MODE system allowed us to critically examine the system's entire approach to personal data handling, from collection to deletion. Throughout this work, we observed that COP-MODE manages sensitive information, such as location data, app usage, and user decisions,with a clear awareness of the associated risks and with solid mitigation strategies already in place.

By applying PIA methodologies, we were able not only to map data flows and assess threats, but also to measure the real effectiveness of the current technical and organizational controls. Our analysis showed that without these controls, the system would be exposed to severe risks, particularly in areas like unauthorized access and data loss. However, thanks to the use of encryption (HTTPS, VPNs), access logging, and strict data minimization policies, the overall risk profile is significantly reduced.

That said, we believe there is room for improvement. Measures such as multi-factor authentication, encryption of data at rest, and real-time anomaly detection would further strengthen the system's resilience. The risk matrices we developed show that even limited improvements can have a meaningful impact on both likelihood and severity, especially in a research context dealing with human participants.

In our opinion, COP-MODE demonstrates a sincere and well-structured commitment to privacy by design. The system is transparent, consent-based, and aligned with GDPR principles. Yet, continuous monitoring, adaptation to new threats, and investment in user awareness are key to ensuring that this commitment holds in the long term.

In summary, COP-MODE is on the right path, and with ongoing effort, it can serve as a strong example of how to conduct responsible, privacy-conscious data research in mobile environment