Privacy Threat Modeling

Engin Bozdag
IWPE 2021

Uber

Introduction

Engin Bozdag

Sr. Privacy Architect

Sr. Researcher

Member ISO/PC 317

PhD in Ethics of Technology
(Algorithmic Bias in Personalization)

Uber

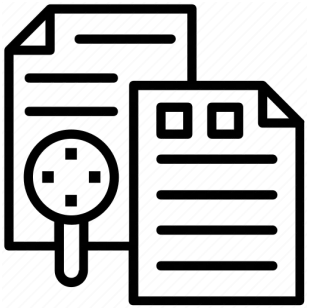
4TU.
Centre for Ethics and Technology

ISO

TU Delft

Technical Privacy Consulting

Privacy Reviews

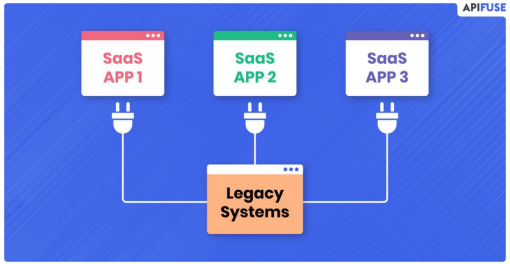


Mitigation Tracking

Technical Privacy Trainings

3rd Party Assessments

Legacy Platforms



Guidelines

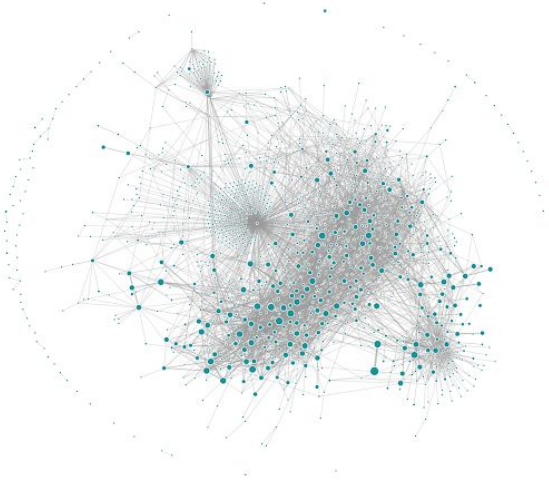
Some Numbers

40+

Reviews of new feature or activity per week

4000+

Microservices



LINDDUN

LINDDUN

- Most known threat modeling framework
- Some threat domains only applicable in very specific settings (e.g. non-repudiation,linkability)
- Some threats contain very important technical sub-threats that deserve separate categories (e.g. Non-compliance, unawareness)

Linkability

Be able to link two datasets

LINDDUN Examples:

- Insufficient anonymization
- Profiling
  - Credentials
  - search & session
  - Ip address
  - browsing patterns

Remarks:

- Is this always an undesired property? Counterexamples:
  - Fraud detection
  - Consent + purpose
  - Transaction data <-> profile
- Linkable to what?
  - Public and future datasets? How do we do the re-identification risk analysis?
- Do we always need anonymity?
  - True anonymity is difficult
  - Alternative: Link on a need to know basis?

Identifiability

Identify a user within a dataset

LINDDUN Examples:

- Insufficient anonymization
  - Quasi identifiers
  - Unique behavior
  - Pseudonyms are re-relinkable
  - credentials

Remarks

- Do we always need anonymity?
- 100% anonymity is not always possible and expensive
- For many cases, identifiability is needed (e.g. KYC, safety, etc.)

Detectability and Non-Repudiation

- User cannot deny being part of a dataset or an action
- User can be detected in a dataset (no access to data itself)

LINDDUN Examples:

- Whistleblower / voting
- Data breach with company email
- Celebrity in a health record
- Address/user already exists
- Person is a user of the service: adult site, health forum)

Remarks

- Real problem, but applicable in some domains only (health, sexual orientation, race, etc.)
  - E.g. sensitive sub-product
    - Social media site offers dating service

Unawareness

User is not aware of consequences of sharing too much information.

LINDDUN Examples:

- No access to personal data
- Opaque privacy policy (no notice, missing purpose/retention, not informed notice)
- Unfriendly UX
- Default settings not privacy friendly
- Consent
  - Not given
  - Cannot revoke
  - Data not deleted after withdrawal
- Insufficient erasure workflow (scope)
- Insufficient correction workflow

Remarks

- Transparency, control, erasure/correction flows categorized under “Unawareness”
- Erasure/correction flow go beyond unawareness
- User can be aware, but not in control
- Different sub-teams/project deal with these issues, why are they all categorized under “Unawareness” bucket?
- Erasure workflow goes beyond scope and awareness.
- What about other privacy rights (e.g. restriction)?

Non-Compliance

Non-compliant with legislation, regulations and corporate policies or data protection principles

Examples:

- Tampering with the policy data store and consents being effected
- Disproportionate storage/collection
  - not needed, but might be useful
  - collection without purpose
  - too much PII in logs
- Disproportionate processing
  - PII in testing
  - secondary use of access logs,
  - location data for profiling
- Unlawful processing (no legal basis)
- Automated decision making

Remark:

- Many unique challenges are all categorized under Non-compliance
- Typically different teams own these subcategories.
- Misses technical abilities to execute privacy rights, etc.
- Not focused on data lifecycle
- Consent => not always
- Automated decision making => not always

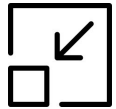
Threat Modeling Based on Data Lifecycle

Collection

- Insufficient Consent
  - Consent on everything
  - Tampered records
  - No records
  - Consent is not respected (e.g. data not deleted)
  - No refresh
- Inaccurate data imported
- Overcollection
  - Not needed
  - Too granular
    - linkable to internal tables while there is no need
    - Precise information while aggregate is enough
- Data not labeled (cannot be found by internal tools/services)
- Unlawful Processing (collection without purpose or purpose unclear)

AGGREGATION

PSEUDONYMIZATION



SUPPRESSION

PIA

DATA TAGGING

AUDITABLE CONSENT

Storage

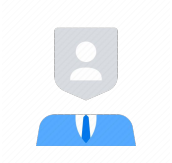
- Inappropriate storage
  - Test environment
  - Not secure for sensitive data
- Privacy Rights Technical Capabilities
  - Insufficient scope
  - Insufficient support for deletion, export or restriction
  - Scalability issues
  - Soft deletion capabilities
  - Insufficient anonymization

SECURE STORAGE

SCALABLE PRIVACY RIGHTS WORKFLOWS

Handling

- Secondary use without lawful basis and controls
- Too granular/identifiable (e.g. analytics with identifiable data)
- Insufficient logging
- No correction (data quality)
- Automated decisions
  - Not explainable
  - No human in the loop
  - No correction
  - No legal ground
- Privacy rights
  - Insufficient process
  - Insufficient security protections in data export
  - Export creates risk for other users



PSEUDONYMIZATION

LOGGING

SCALABLE PRIVACY RIGHTS WORKFLOWS

TRANSPARENCY

HUMAN REVIEW

PIA

Sharing

- Insecure Data Transfer to vendor / 3rd party
  - Insufficient sharing controls
  - Insufficient security or privacy controls at the vendor
  - Insufficient contracts
- Insufficient anonymization
  - Dataset too small
  - Large number of attributes
  - Keys exposed
  - Algorithm can be reversed
  - Quasi identifiers can lead to re-identification
- Data cannot be anonymized, but no controls are in place in 3rd party (TTL, API's for privacy rights propagation, secure storage)

DATA SHARING CONTROLS

ANONYMIZATION

PSEUDONYMIZATION

TIME TO LIVE

PIA

AGREEMENTS

SECURE STORAGE

Deletion (End of Life)

- Inactive users
- Backups restoring deleted data
- Retention policy not implemented in all downstream services, caches, backups, log files, employee devices, snapshots/blobs on cloud
- Too long TTL just in case
- TTL does not map to the retention policy or the retention policy's scope is insufficient
- Soft deletes or insufficient anonymization instead of deletion

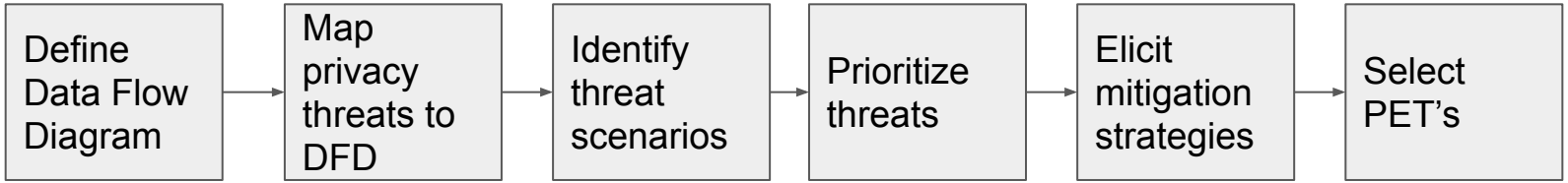
INACTIVE USER DELETION

TIME TO LIVE

ANONYMIZATION

The Process

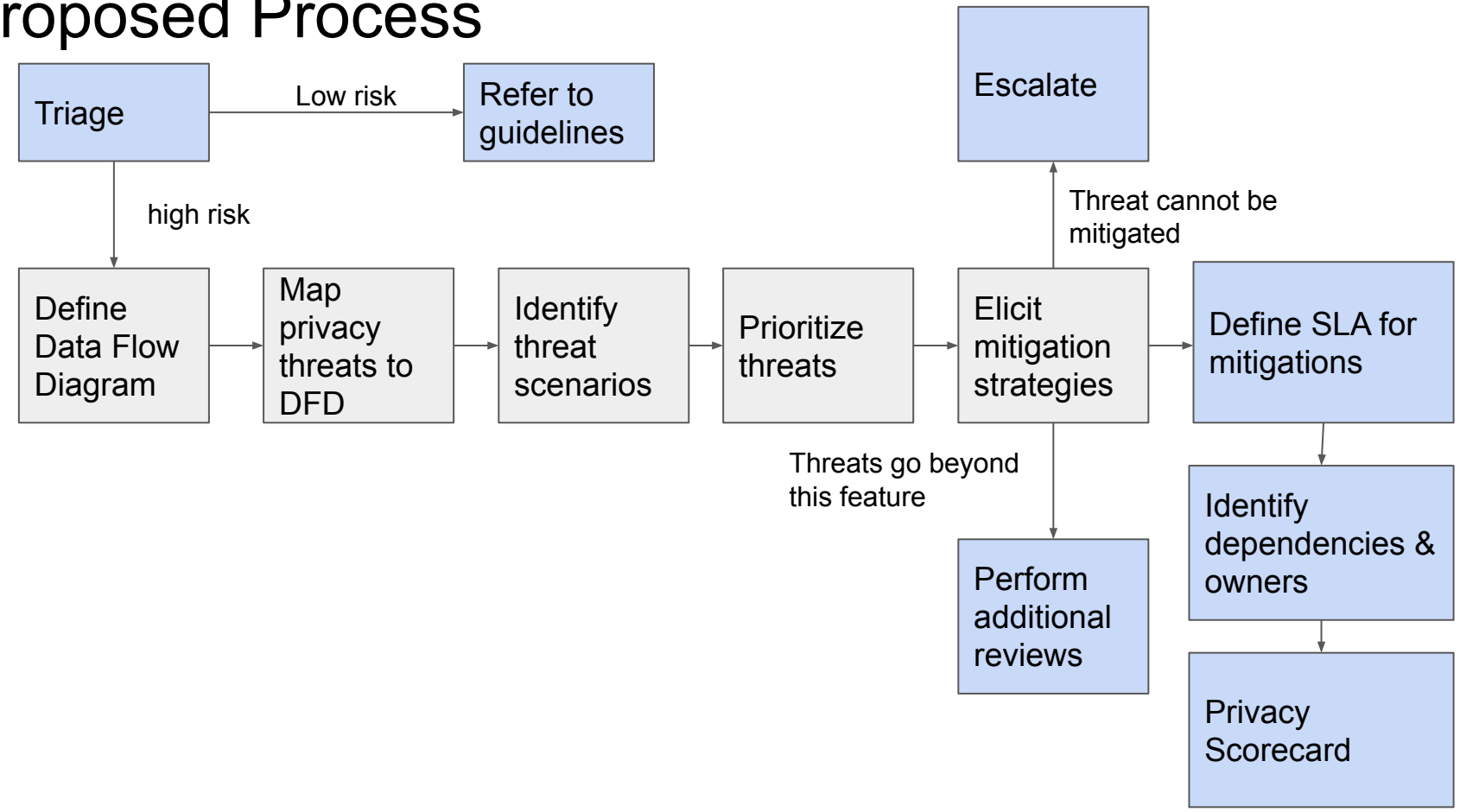
LINDDUN Process



Challenges

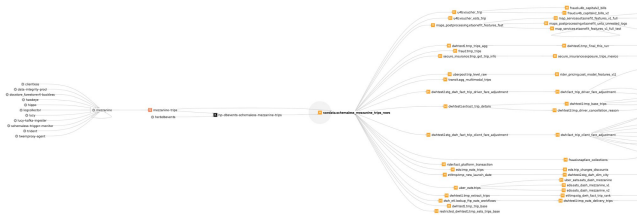
- High number of reviews: cannot follow LINDDUN for all reviews
- Complex data flow diagrams
- The identified threat might go beyond this feature
- Datasets not connected to a previous threat analysis
- Dependencies for mitigations
- No ownership
- What if threat cannot be mitigated?

Proposed Process



Supporting Resources

- Data Lineage



- Privacy champs
  - You need experts in specific teams to identify dependencies find optimum controls
- Privacy Scorecard
  - You may need to escalate a list of items to org leaders to get buy-in for remediations

Beyond Reviews

- Reviews cannot cover all activities
- Continuous data discovery/tagging is needed
- Future:
  - Connect data classification to CI/CD
  - Privacy policy enforcement engines
  - More review automation
  - See Privacy is an afterthought in the software lifecycle