UMR IRISA
INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALEATOIRES
Campus Universitaire de Beaulieu • 263 Avenue du Général Leclerc - 35042 Rennes Cedex France
Tél. : +33 2 99 84 71 00 • Télécopie : +33 2 99 84 75 22 • Internet : www.irisa.fr

Privacy-Compliant REST APIs

Equipe : DiverSE
Site web de l'équipe : http://www.irisa.fr/diverse/
Date de début de thèse : 1er septembre 2026
Lieu : Rennes
Unité de recherche : IRISA - UMR 6074

Description du sujet de la thèse :

REST APIs have become ubiquitous for exposing web services across diverse domains, from industry and cloud computing to government services. Despite their popularity, their development remains complex and error-prone: inconsistencies between business requirements, code, and documentation are frequent. For example, a healthcare API may document a parameter as optional while it is actually required. Such inconsistencies not only cause integration failures but also pose serious privacy risks, potentially exposing sensitive data or enabling unauthorized access.
The objective of this thesis is to **identify, prevent, and mitigate privacy risks in REST APIs that arise from insufficiently specified or poorly enforced privacy requirements.** The thesis seeks to ensure that REST APIs comply with privacy regulations such as the GDPR and avoid unintended exposure or misuse of sensitive information. This will be done by focusing on the explicit representation of privacy-related data and regulatory constraints at the API specification level, which is the formal contracts exposed to clients.

Objectives of this doctorate
The objectives for this doctoral thesis are formulated as the following research questions:

- **RQ1:** How can privacy risks in REST APIs be systematically identified at the specification level, particularly with respect to the exposure and processing of personal data defined by data protection regulations such as the GDPR?
- **RQ2:** How can personal data and regulatory constraints, such as data minimization, purpose limitation and access restrictions, be explicitly represented in REST API specifications to prevent unintended data exposure or misuse?
- **RQ3:** How effective is specification-level analysis in detecting and reducing GDPR-related privacy violations in REST APIs before deployment?

The originality of this work lies in its integrated approach combining Model-Driven Engineering (MDE) with regulatory compliance verification. This approach places OpenAPI specifications at the center of the analysis, treating them as a key abstraction for reasoning about privacy and regulatory compliance. By grounding the approach in empirical observations of real-world APIs, the proposed methodology supports the systematic identification, validation, and continuous monitoring of privacy risks and compliance issues throughout the REST API lifecycle.

Methodology

To address **RQ1**, the PhD candidate will conduct an **empirical study of thousands of REST APIs** collected from APIs.guru, RapidAPI, and GitHub in order to identify recurring privacy-related inconsistencies and risk patterns. These observations will support the **design of a dedicated modeling language,** based on OpenAPI, for REST APIs that explicitly captures personal data and associated privacy rules. To address **RQ2**, the proposed modeling language and its associated tools will be used to **generate API specifications that explicitly embed privacy properties and regulatory constraints.** These enriched specifications aim to make personal data exposure and compliance requirements explicit, supporting the automated identification of privacy risks in REST APIs, in particular, the exposure and processing of personal data defined by data protection regulations such as the GDPR. The enriched specifications also provide a concrete basis for advanced verification and testing mechanisms such as automated privacy compliance checks and privacy-oriented regression testing. To address **RQ3**, the PhD candidate will integrate the verification and testing tools developed in previous steps into **continuous integration and deployment (CI/CD) pipelines**, enabling ongoing privacy checks and regression testing throughout the API lifecycle. This integration aims to prevent unintended personal data disclosure, unauthorized access, and privacy misconfigurations in evolving REST APIs.

The expected outcomes include a **formal modeling framework, automated detection and verification tools, enriched API specifications, and practical strategies for continuous privacy compliance.** Experimental validation on both open-source and industrial APIs will assess the consistency, robustness, and effectiveness of these mechanisms in identifying and mitigating privacy risks.

Skills

- Master's degree in software engineering or a closely related field
- Knowledge in Web Development, OpenAPI and/or Model-Driven Engineering (MDE) and/or Web Privacy
- Good programming skills
- Ability to work autonomously
- Willingness to learn about new technologies/techniques
- Good writing and communication skills
- Ability to propose, present, and discuss new ideas

The PhD candidate will be encouraged to publish in top-tier conferences such as ICSE, MODELS, WebConf, PETS, and S&P, and to release the developed tools as open-source software (GPL or Apache), facilitating technology transfer to industry. Experiments on real-world cases will demonstrate the effectiveness of the proposed tools.

As is a common practice in the DiverSE research team, all source code will be open-sourced using either the GPL or Apache Licenses. It is expected that the doctoral student participate in related open-source communities. This should also assist in the technological transfer from academic prototypes to industrial-ready tools. Experimentations to demonstrate the effectiveness of developed tools on real-world issues are actively encouraged and expected.

References

1. A. Cavoukian, 2009. Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario, Canada.
2. T. Antignac et al., 2018. Privacy Compliance Via Model Transformations. IEEE European Symposium on Security and Privacy Workshops (EuroS&P Workshops).
3. S. Krstić et al., 2024. Model-Driven Privacy. PETS.
4. B.O. Emeka et al., 2023. A Practical Model Driven Approach for Designing Security Aware RESTful Web APIs Using SOFL. IEICE TRANSACTIONS on Information and Systems.
5. **S. Challita** et al., 2018. A Precise Model for Google Cloud Platform. IEEE IC2E.
6. H. Cao et al., 2017. Automated Generation of REST API Specification from Plain HTML Documentation. ICSOC.
7. E. Grünewald et al., 2021. TIRA: an OpenAPI Extension and Toolbox for GDPR Transparency in RESTful Architectures. IEEE European Symposium on Security and Privacy Workshops (EuroS&PW).
8. A. El fraihi et al., 2024. Client-side and Server-side Tracking on Meta: Effectiveness and Accuracy. PETS.

Liste des encadrants et encadrantes de thèse :

Stéphanie Challita
Type d'encadrement : Encadrante de thèse
Unité de recherche : IRISA
Département : D4 - Langage et génie logiciel
Equipe : DiverSE

Olivier Barais
Type d'encadrement : Directeur de thèse
Unité de recherche : IRISA
Département : D4 - Langage et génie logiciel
Equipe : DiverSE

Contact :

Nom : Stéphanie Challita
Email : stephanie.challita@irisa.fr
Mots-clés : REST APIs, OpenAPI, Model-Driven Engineering, Privacy