数据安全（2023-2024学年2学期）

顶级会议论文阅读大作业

PRIVGUARD: Privacy Regulation Compliance Made Easier


展示人：付政烨


学号：2113203

目 录

CONTENTS

PART.01

文章简介

Article Introduction

文 章 简 介

PRIVGUARD核心组件

1. ANALYZER：基于抽象解释的静态分析器
2. 在数据生命周期内提供安全的保护组件

一个公司在从小型企业到商业巨头的过程中，持续遵守隐私法规（如GDPR和CCPA）已经成为一项沉重的负担。造成这一困境的主要原因是当前合规过程中对人工审核的高度依赖，这不仅成本高昂、速度缓慢，而且容易出错。为了解决这一问题，研究人员提出了一种新颖的系统设计——PRIVGUARD。

PART.02

PRIVGUARD 框架

PRIVUARD Framework

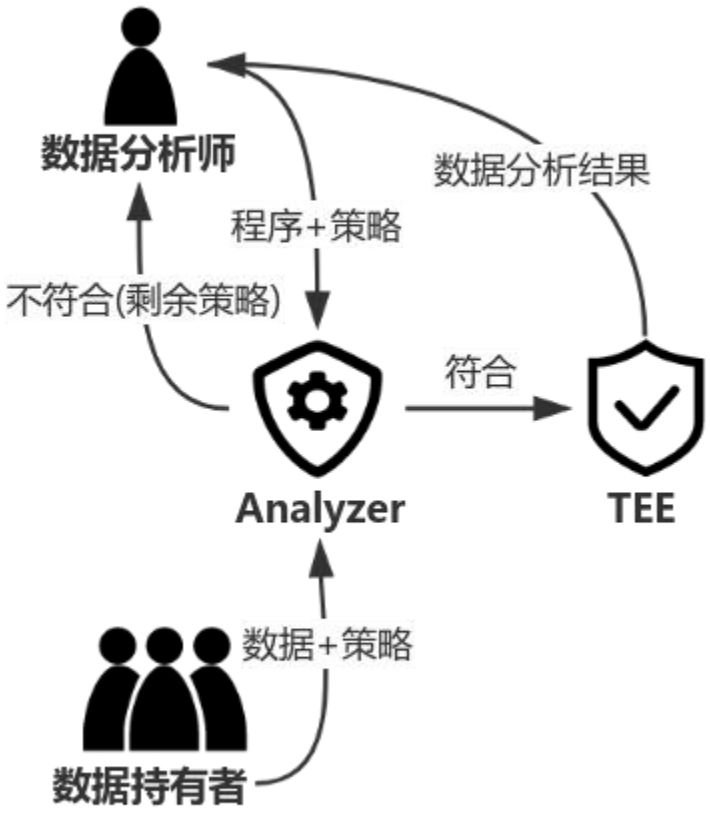
PRIVGUARD 框架——各方主体简述

主体：数据持有者

- 收集用户信息的企业。如用户的家庭住址、网页浏览偏好等；
- 向数据中心提交数据时，用户可指定隐私偏好（隐私策略），行使限制数据分析方处理其数据的权利。

政策编码

- 使用 LEGALEASE 策略语言对隐私策略和守护策略的形式化描述，便于 Analyzer 进行分析。



主体：数据分析方

- 利用用户数据来开发智能推荐算法的专业人员；
- 诸如抖音、今日头条等软件数据分析师，旨在为用户提供个性化的内容推荐；
- 需要向数据中心提供守护策略。

Analyzer合规检测器

- 输入：守护\隐私策略（编码）；
- Analyzer：比对检测两个策略；
- 输出：不符合 -> 返回剩余策略；
  符合 -> 传输到可信执行环境。

PRIVGUARD 框架——其他概念

可信执行环境(TEE)

- 安全数据分析环境，能够执行符合隐私策略的分析程序;
- 是数据胶囊“溶解”并发挥效用的场所;
- 提供隔离机制，确保数据分析不会侵犯隐私。

数 据 胶 囊

- 将用户数据和隐私策略的政策编码打包在一起，形成数据胶囊;
- 对于符合策略的程序，才可以将胶囊传输至TEE“溶解”，供分析程序使用。

合 规 检 测

- 是数据胶囊“溶解”前关键步骤（静态分析）;
- 旨在确保数据分析方的程序符合严格符合数据持有者提供的隐私政策;
- 分为策略检查和程序检查(Analyzer 实现)。

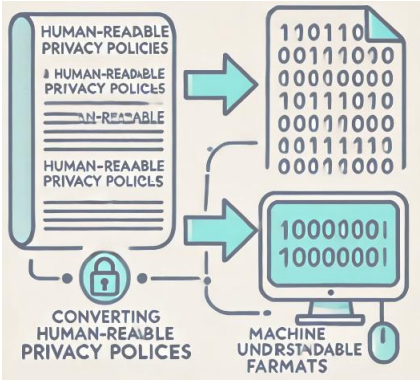
PART.03

合 规 检 测

Compliance testing

参考源码(wanglun1996)：https://github.com/sunblaze-ucb/privguard-artifact

合规检测——策略检查



技术背景

在现实应用中，隐私政策是依据特定的法律法规编写，并采用易于人类理解的语言进行表述。然而，这些以人类可理解的文字形式（如英语）编写的政策，机器却难以解析。因此，在进行隐私合规性检查之前，必须使用特定的政策语言对隐私政策进行编码，将其转换为机器可理解的格式。

LEGALEASE 策略语言

本文采用的政策语言是 LEGALEASE，一种利用各类属性来表达政策的编码框架。通过使用LEGALEASE提供的一部分属性，并扩展其他属性，以完成政策的编码。编码后的政策具有特定的语法结构（详见下图）：

```
Policy → ('AlLOW' Clause)+
Clause → Attribute | Clause 'AND' Attribute | Clause 'OR' Attribute
Attribute → Filter_Attribute | Redact_Attribute | Schema_Attribute | Privacy_Attribute | Role_Attribute | Purpose_Attribute
```

合规检测——策略检查

属性	属性描述	策略示例	含义
Filter	规定能对数据表中哪些数据记录进行访问，其他数据记录被排除	ALLOW FILTER score >= 80	允许访问得分不低于80的记录
Schema	规定能对数据表中哪些字段进行访问，其他字段要被排除	ALLOW SCHEMA age, cost	仅允许访问age和cost字段
Redact	规定在对数据进行访问时只能访问数据内容一部分，剩下的部分需要进行脱敏	ALLOW REDACT address(x:)	仅允许查看地址信息的高x位（县级及以上）
Privacy	规定使用何种隐私保护技术控制数据的使用	ALLOW PRIVACY DP()	使用差分隐私技术保护数据
Role	规定了哪些个体或群体能访问或查看数据	ALLOW ROLE administrators	仅允许管理员访问数据表
Purpose	限制数据访问的目的	ALLOW PURPOSE goodsale	仅允许基于增加销售量的目的访问数据

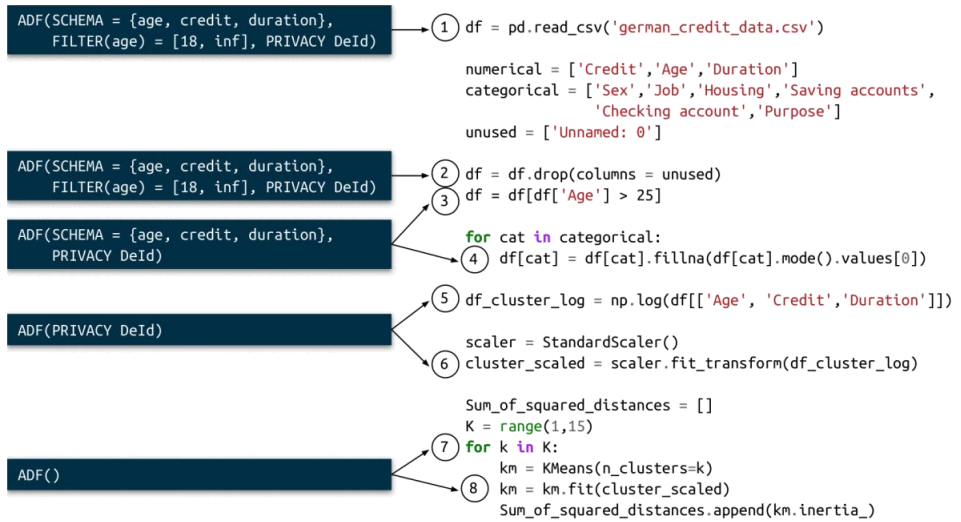
表1：属性含义介绍

合规检测——策略检查

语法分析技术

- 隐私政策编码时，各属性和操作被整合成一段由关键字和属性值构成的字符串文本；
- 对文本解析，这一过程需要应用语法分析技术；
- 本文采用了上下文无关文法来形式化定义这种编码语言。

Policy → Policy | ClauseClause → 'ALLOW' Content
Content → Attribute | Content 'AND' Attribute | Content 'OR' Attribute
Attribute → Filter_Attribute | Redact_Attribute | Schema_Attribute | Privacy_Attribute | Role_Attribute | Purpose_Attribute
Filter_Attribute → 'FILTER' Column Comparator Data
Redact_Attribute → 'REDACT' Column '(' Integer ':' Integer ')'
Schema_Attribute → 'SCHEMA' List
Privacy_Attribute → 'PRIVACY' Method
Method → 'Anonymization' | 'Aggregation' | 'k-anonymity' Integer | 'l-diversity' Integer | 't-closeness' Integer | DP '(' Float ':' Float ')
Role → 'ROLE' Alphanums
Purpose → 'PURPOSE' Alphanums
Data → Integer | String
Comparator → '>' | '<' | '>=' | '<=' | '==' | '!='
List → List ',' Column | Column
Column → Alphanums



CFG

使用上下文无关文法，并借助语法分析工具，可以对字符串进行语法分析。完成语法分析后，设计适当的翻译模式，将隐私政策的编码结果转换成语法树并存储在内存中。随后，隐私合规性检测的过程就转变为对语法树和代码进行遍历的过程。

合规检测——策略检查

- ✓ 对编码后的政策进行语法分析后，将得到一个类似于语法树的结果；
- ✓ 对于以 ALLOW 开头的政策，语法树为二叉树，叶子节点保存属性，非叶子节点保存运算符（AND或OR）。

- ✓ 隐私合规检测的原理是对属性运算构成的集合进行操作；
- ✓ 若属性满足，则从集合中删除；
- ✓ 若集合为空，则已满足隐私政策。

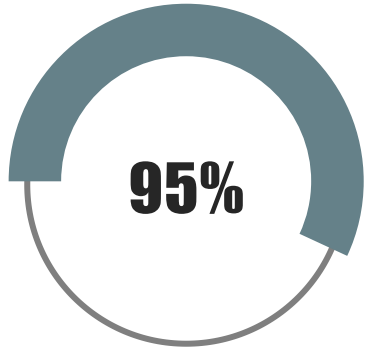
- ✓ 为了便于后续的隐私合规检测计算，需要将语法树转换为析取范式（DNF）形式，即若干个属性的合取后再析取。

- ✓ 转换的过程也就是自顶向下遍历二叉树的递归过程；
- ✓ 分为：叶节点；非叶AND节点;非叶OR节点三种情况。

合 规 检 测 — — 程 序 分 析

主要原理

- ✓ 判断代码中执行的操作是否满足预定义的政策属性要求，来确保数据处理的合规性；
- ✓ 大部分Python程序依赖于第三方库来读取数据表中的数据；
- ✓ 通过重写第三方库函数，使其功能不再是直接操作数据，而是检查和验证政策属性。




```
1  import pandas as pd
2  c1 = pd.read_csv("1.csv")
3  c2 = pd.read_csv("2.csv")
4  ehr = pd.merge(c1, c2, left_on="ID",
5                 right_on="ID")
6  ehr = ehr[ehr.score >= 80]
```

示例介绍

- ✓ 如果属性集合中存在FILTER score >= 70，该条件比score >= 80要弱，则该属性满足，可以从集合中删除；
- ✓ 反之，则该属性不满足，会返回UNSAT。

PART.04

结 果 与 展 望

Results and Outlook

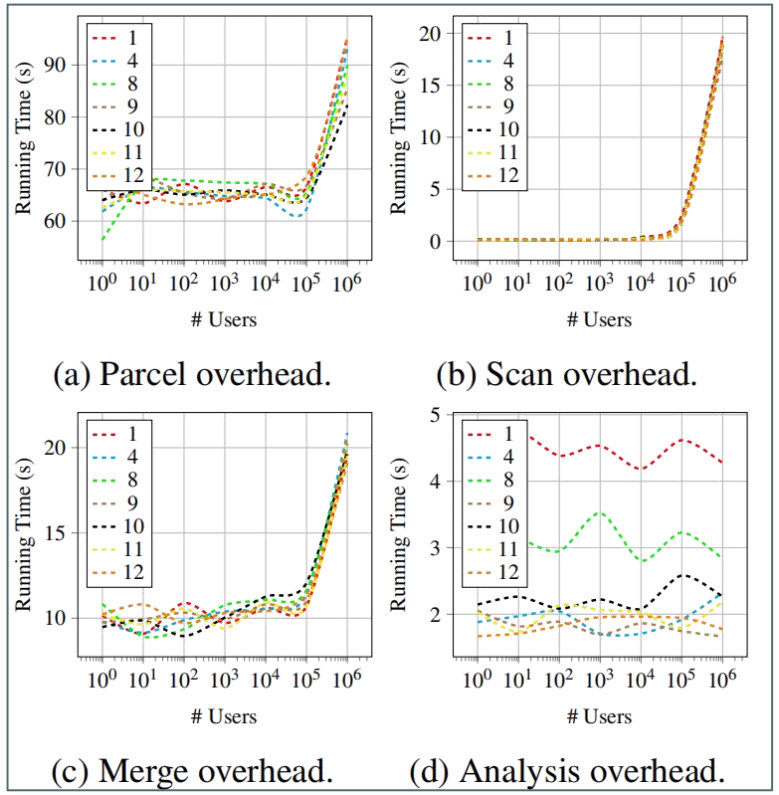
基于ANALYZER的实验结果

实验环境

1. 存储层：Inter Planetary File System (IPFS)
2. 数据加密：AES-256-GCM
3. 可信执行环境：AMD SEV

实验结果

1. PRIVANALYZER性能开销小；
2. 控制流构造的影响；
3. PRIVGUARD的可扩展性。



实验过程

1. 从Kaggle收集了23个不同任务的分析程序，为每个程序设计了 LEGALEASE 策略；
2. 手动检查 PRVANALYZER 输出的结果是否正确，以验证其准确性和可靠性。

ANALYZER 限制

1. 容易受到内部攻击；
2. 许多 PURPOSE 不能自动执行；
3. 依赖于 TEE，如AMD SEV。

结 果 与 展 望

- ✓ 对数据的分析与使用已经成为互联网公司发展过程中不可或缺的环节；
- ✓ 然而，数据不合规使用、隐私数据泄漏等问题也日益凸显。

- ✓ 本文针对隐私合规检测，设计了基于数据胶囊的隐私合规自动化检测系统；
- ✓ 用策略保护数据，确保了数据流动过程遵循既定的隐私政策。

- ✓ 确保数据胶囊中的数据使用如同精准医疗中的靶向给药，数据释放恰到好处；
- ✓ 提高了隐私合规检测的效率与准确性，降低了合规检测成本。

顶 级 会 议 论 文 阅 读 大 作 业

感谢各位老师指导

The autumn wind, cool breeze, flowers and trees, like a loving mother is humming a lullaby to put their children into the sweet dreams.

展示人：付政烨　　学号：2113203