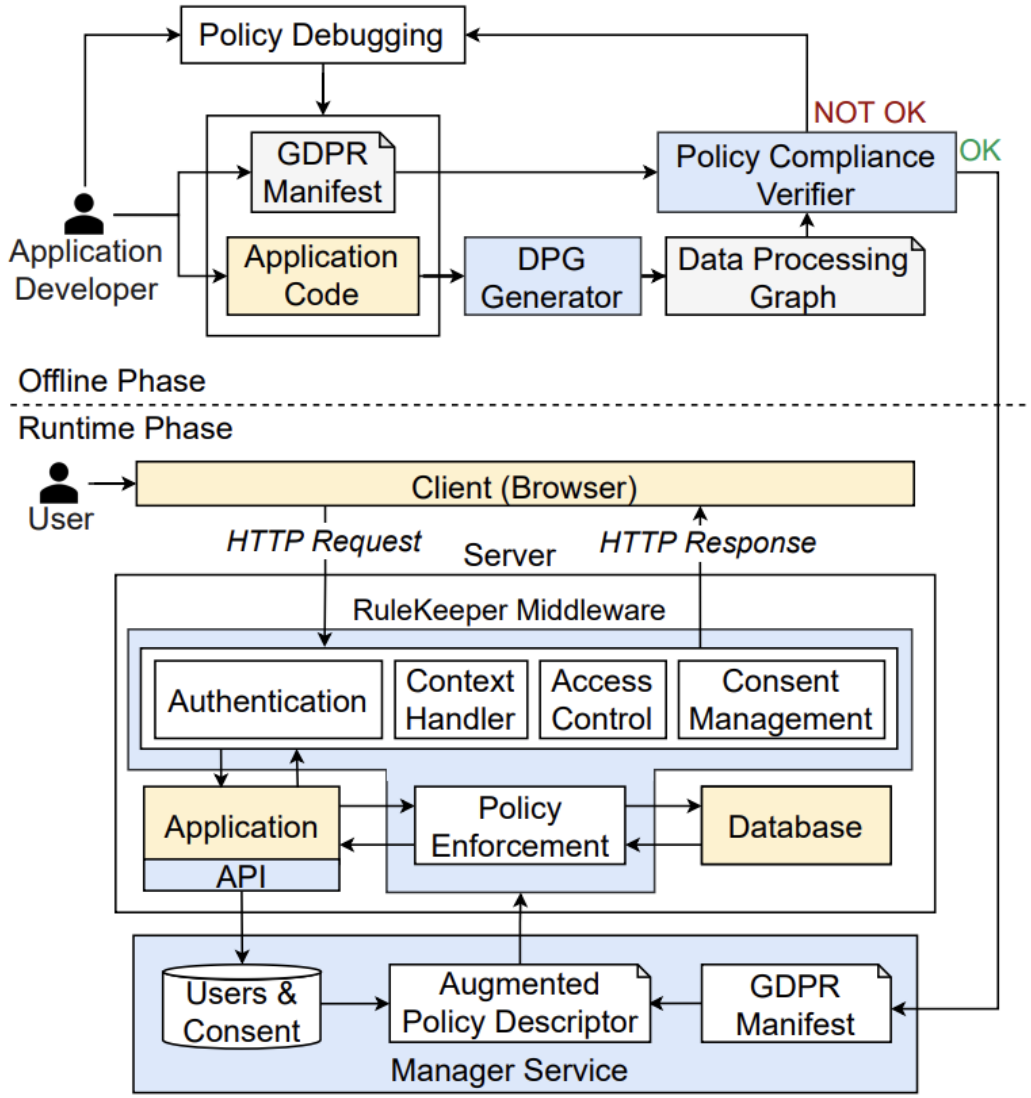
RuleKeeper

- Difficult to guarantee GDPR enforcement – bugs, vulnerabilities.
- No system-level support for enforcing GDPR policies.
- RuleKeeper:
  - Represent policy as GDPR Manifest using DSL
  - Static and dynamic code analysis for detecting violations
  - Proof-of-Concept for MERN stack

Threats to Compliance

- **Data Processed for Incompatible Purposes**
  - Updates to functionality accessing personal data granted for other purposes.
  - **Example**: Trip history being used for marketing
- **Reflective Compliance Bug**
  - Updates to functionality now include personal data
  - **Example**: Updated SQL schema including more personal data
- **Purpose Escalation Attack**
  - Leaking personal data through vulnerable function
  - **Example**: SQL injections
- **Defective Consent Management**
  - User cookie decision does not reflect on server side
  - **Example**: Trip history being used for marketing

```
# Application plane
DATA-ITEMS: ticket buyer name, ticket destination, ticket date,
↪  ticket buyer credit card, trip destination, trip date, email.

OPERATIONS: see schedules, buy ticket, see purchase history,
↪  subscribe to newsletter.

# GDPR plane
PERSONAL-DATA: ticket buyer name, ticket destination, ticket date,
↪  ticket buyer credit card, email.

PURPOSES: ticket management, marketing.

DATA-COLLECTION:
  ticket buyer name, ticket destination, ticket date, ticket buyer
  ↪  credit card ARE COLLECTED FOR ticket management purposes.
  email IS COLLECTED FOR marketing purposes.

LAWFULNESS-BASE:
PURPOSE ticket management HAS LAWFULNESS BASE consent.
PURPOSE marketing HAS LAWFULNESS BASE consent.

EXECUTED-FOR:
  buy ticket, see purchase history ARE EXECUTED FOR ticket
  ↪  management purposes.
  subscribe to newsletter IS EXECUTED FOR marketing.

# Mapping planes
DATA-MAPPING:
ticket buyer name IS IN COLUMN name OF TABLE tickets.
ticket destination IS IN COLUMN destination OF TABLE tickets.
ticket date IS IN COLUMN date OF TABLE tickets.
ticket buyer credit card IS IN COLUMN credit_card OF TABLE tickets.
trip destination IS IN COLUMN destination OF TABLE schedules.
trip date IS IN COLUMN date OF TABLE schedules.
email IS IN COLUMN e_mail OF TABLE newsletter.

OPERATION-MAPPING:
see schedules IS MAPPED TO ENDPOINT GET /schedules.
buy ticket IS MAPPED TO ENDPOINT POST /buy_ticket.
see purchase history IS MAPPED TO ENDPOINT POST /purchase_history.
subscribe to newsletter IS MAPPED TO ENDPOINT POST /subscribe.

DATA-OWNERSHIP:
OWNER IN TABLE tickets IS IN COLUMN name.
OWNER IN TABLE newsletters IS IN COLUMN e_mail.
```



Results

Application	DSL Statements				Lines of Code
	Personal Data	Purposes	Operations	Total	
LEB	11	2	10	66	8
Habitica	14	3	21	168	10
Amazona	13	3	15	161	7
Blog	11	2	18	127	6

App	Server Latency			Client Latency			Throughput			Efficiency	
	5th	avg.	95th	5th	avg.	95th	min.	avg.	max.	cpu	mem.
L	1.08×	1.26×	1.34×	1.06×	1.20×	1.25×	0.66×	0.74×	0.94×	7.40%	4.23%
H	1.14×	1.16×	1.19×	1.09×	1.10×	1.12×	0.88×	0.91×	0.94×	3.67%	7.15%
A	1.07×	1.14×	1.21×	1.04×	1.14×	1.26×	0.70×	0.85×	0.96×	2.30%	1.32%
B	1.25×	1.26×	1.27×	1.07×	1.15×	1.22×	0.70×	0.74×	0.77×	6.00%	4.58%

Table 4: Summary of RuleKeeper's experimental evaluation.

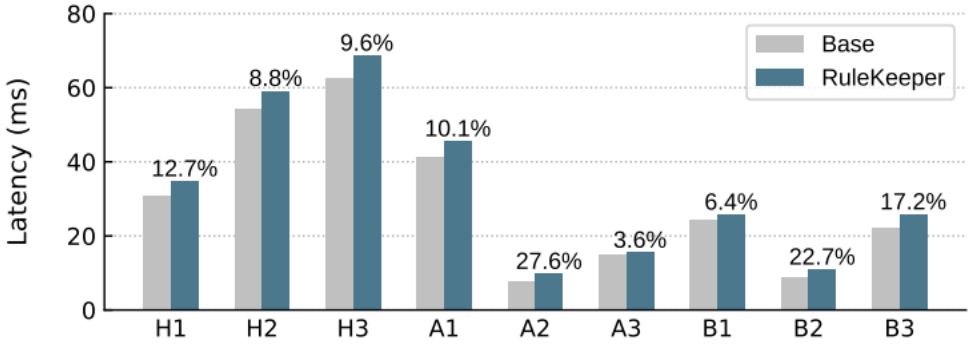
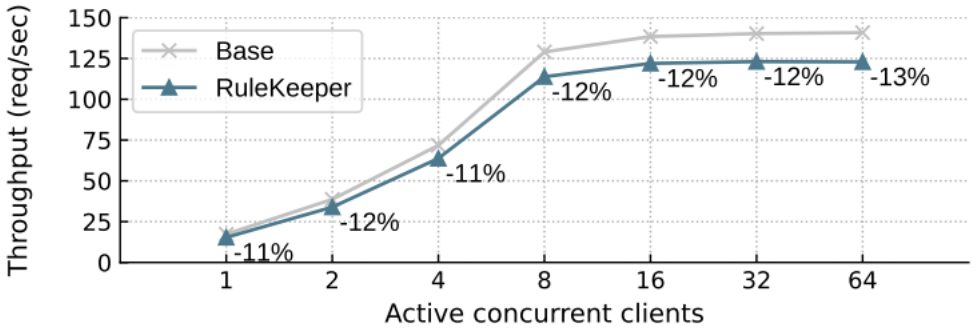


Figure 9: Average client-perceived latency in legacy application tasks. Labels show the relative overhead, in percentage.



More results

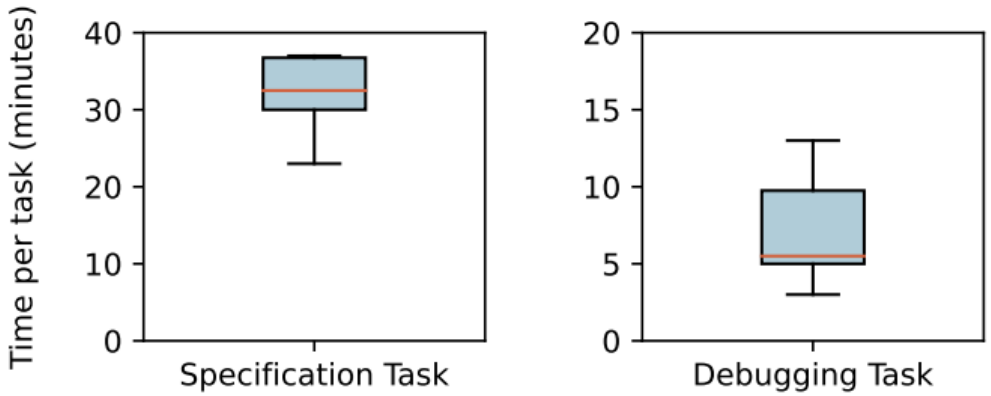


Figure 11: Time for participants to perform the tasks. The box extends from the 1st to the 3rd quartile; the median line is shown.

Application	CRG Size		Execution Time (s)			Accuracy
	Nodes	Edges	CRG	DPG Queries (N,C)		EM-pairs
LEB (257 LoC)	1047	1710	0.201	27.814	21.213	10/11
Amazona (570 LoC)	2238	4508	0.357	41.154	34.520	16/16
Blog (1075 LoC)	4189	8987	0.637	589.669	540.623	31/34

What People Loved / Wanted Improved

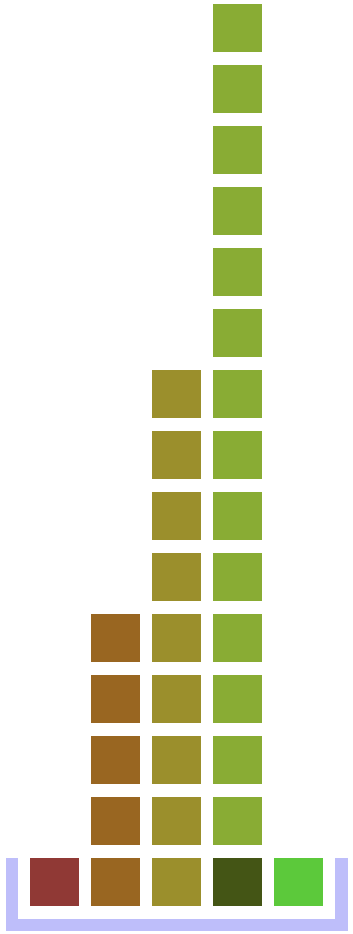
- Identified/solved gaps in SOTA – system-level compliance.
- Sticky banners
- Dynamic policy enforcement.
- Various real-world applications
- Combined both static and dynamic analysis.
- MERN is easily adaptable.
- Other policy frameworks / principles? (CCPA/LGPD, retention, subject rights)
- More complex data access?
- Static analysis accuracy?
- Better UI – visual drag/drop?
- Expand beyond MERN?

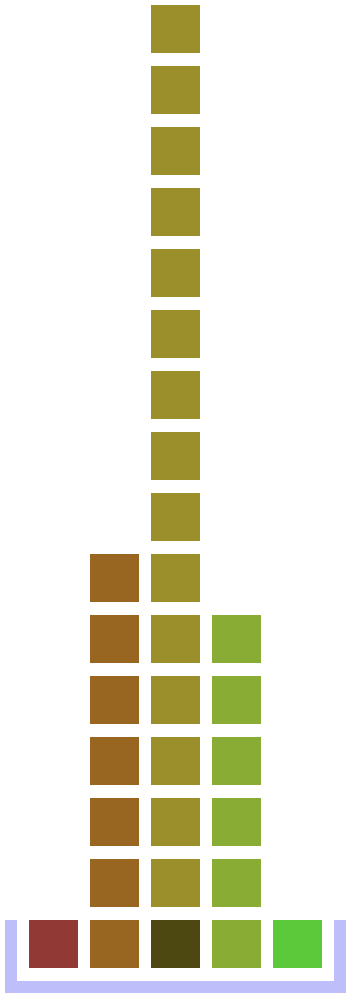
What People Hated

- **13% increase in client-perceived latency**
  - Handwaved as small, likely not noticeable – really?
  - Seemed to increase with user count, how effective at scale?
- **Usability Study**
  - 10 students, not developers?
  - Same university, and not randomly selected??
  - 45 minutes, and no monetary compensation???

Discussion

- **Should we assume good intent?**
  - Are companies trying to **not** violate GDPR *accidentally*?
  - How do we guarantee compliance?
- **Who's responsible (if things go wrong)?**
  - What do lawyers think? Is RuleKeeper (middleware) compliant?
- **Is it scalable?**
  - How automatable is spec generation? What other challenges?
  - Static analysis for large code base? Is +13% latency really NBD?
- **How's the UX?**
  - What did users see when things were blocked?
  - How will user respond to the "sticky banners"?





Scoring