

Verifying Equivalence of Database-Driven Applications

YUEPENG WANG, University of Texas at Austin, USA

ISIL DILLIG, University of Texas at Austin, USA

SHUVENDU K. LAHIRI, Microsoft Research, USA

WILLIAM R. COOK, University of Texas at Austin, USA

This paper addresses the problem of verifying equivalence between a pair of programs that operate over databases with different schemas. This problem is particularly important in the context of web applications, which typically undergo database refactoring either for performance or maintainability reasons. While web applications should have the same externally observable behavior before and after schema migration, there are no existing tools for proving equivalence of such programs. This paper takes a first step towards solving this problem by formalizing the equivalence and refinement checking problems for database-driven applications. We also propose a proof methodology based on the notion of *bisimulation invariants* over *relational algebra with updates* and describe a technique for synthesizing such bisimulation invariants. We have implemented the proposed technique in a tool called MEDIATOR for verifying equivalence between database-driven applications written in our intermediate language and evaluate our tool on 21 benchmarks extracted from textbooks and real-world web applications. Our results show that the proposed methodology can successfully verify 20 of these benchmarks.

CCS Concepts: • **Theory of computation** → **Program verification**; • **Software and its engineering** → **Software verification**; **Formal software verification**;

Additional Key Words and Phrases: Program Verification, Equivalence Checking, Relational Databases.

ACM Reference Format:

Yuepeng Wang, Isil Dillig, Shuvendu K. Lahiri, and William R. Cook. 2018. Verifying Equivalence of Database-Driven Applications. *Proc. ACM Program. Lang.* 2, POPL, Article 56 (January 2018), 29 pages. <https://doi.org/10.1145/3158144>

1 INTRODUCTION

Due to its usefulness in several application domains, the problem of checking equivalence between programs has been the subject of decades of research in the programming languages community. For instance, in the context of translation validation, equivalence checking allows us to prove that the low-level code produced by an optimizing compiler is semantics-preserving with respect to the original program [Necula 2000; Pnueli et al. 1998; Rinard 1999; Zuck et al. 2003]. Other important applications of equivalence checking include *regression verification* [Godlin and Strichman 2009, 2013] and *semantic differencing* [Lahiri et al. 2012, 2013; Person et al. 2008].

While there has been significant progress in the area of **equivalence checking**, **existing techniques cannot be used to verify the equivalence of database-driven programs**. To see why verifying

Authors' addresses: Yuepeng Wang, University of Texas at Austin, USA, ypwang@cs.utexas.edu; Isil Dillig, University of Texas at Austin, USA, isil@cs.utexas.edu; Shuvendu K. Lahiri, Microsoft Research, USA, shuvendu.lahiri@microsoft.com; William R. Cook, University of Texas at Austin, USA, wcook@cs.utexas.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to the Association for Computing Machinery. 2475-1421/2018/1-ART56
<https://doi.org/10.1145/3158144>

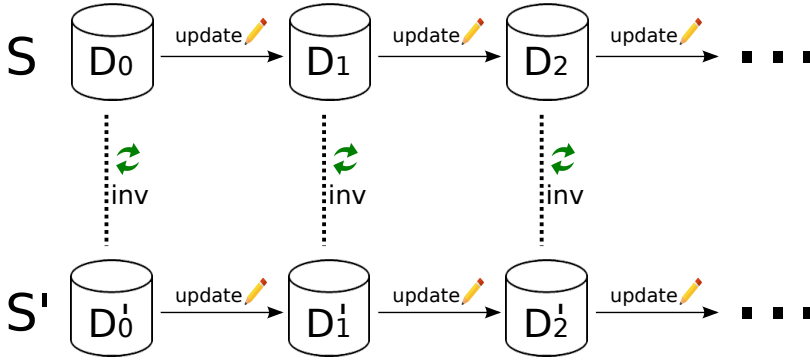


Fig. 1. Bisimulation invariant between two database-driven applications

equivalence is important in this context, consider the scenario in which a web application interacts with a relational database to dynamically render a web page, and suppose that the database schema needs to be changed either for performance or maintainability reasons. In this case, **the developer will need to migrate the database to the new schema and also re-implement the code that interacts with the database without changing the observable behavior of the application**. While this task of *database refactoring* arises very frequently during the life cycle of web applications, it is also known to be quite hard and error-prone [Ambler 2007; Wikimedia 2017], and several textbooks have been published on this topic [Ambler and Sadalage 2006; Faroult and L’Hermite 2008]. As pointed out by Faroult and L’Hermite, “*database applications are a difficult ground for refactoring*” because “*small changes are not always what they appear to be*” and “*testing the validity of a change may be difficult*” [Faroult and L’Hermite 2008].

Motivated by the prevalence of database-driven applications in the real world and the frequent need to perform schema migration, **this paper proposes a new technique for verifying the equivalence of database-driven applications**. Given a pair of programs P, P' that interact with two different databases, we would like to automatically construct a proof that P and P' are semantically equivalent. Unfortunately, the problem of equivalence checking for database-driven applications introduces several challenges that are not addressed by previous work: First, **it is unclear how to define equivalence in this context, particularly when the two database schemas are different**. Second, **database-driven applications typically use declarative query languages, such as SQL, but, to the best of our knowledge, there are no automated reasoning tools for a rich enough fragment of SQL that captures realistic use cases**.

In this paper, we formalize the equivalence checking problem for database-driven applications and propose a verification algorithm for proving equivalence. Suppose that we are given two programs P, P' that interact with databases D, D' . Let us also assume that each program comprises a set of database transactions (query or update) such that every transaction t in P has a corresponding transaction t' in P' . **Our goal is to prove that P and P' yield the same result on a pair of corresponding queries q, q' whenever we execute the same sequence of update operations on D and D' .**

To prove equivalence between a pair of database-driven applications, our approach infers a **so-called bisimulation invariant that relates states of the two programs**. In this context, program states correspond to database instances, so the bisimulation invariants relate a pair of database instances. As shown in Figure 1, our bisimulation invariants are preserved after each database transaction, and, in addition, they are strong enough to establish that any corresponding pair of queries must yield the same result.

In the context of software verification, program invariants are typically expressed in some first-order theory supported by modern SMT solvers. Unfortunately, since the bisimulation invariants that we require in this context relate a pair of databases, they typically involve non-trivial relational algebra operators, such as join and selection. To solve this difficulty, we consider the theory of relational algebra with updates, \mathcal{T}_{RA} , and present an SMT-friendly encoding of \mathcal{T}_{RA} into the theory of lists, which is supported by many SMT solvers.

Once we automate reasoning in relational algebra with updates, a remaining challenge is to automatically infer suitable bisimulation invariants. In this paper, we use the framework of monomial predicate abstraction to automatically synthesize conjunctive quantifier-free invariants that relate two database states [Ball et al. 2005; Das et al. 1999; Lahiri and Qadeer 2009]. Specifically, we identify a family F of predicates that are useful for proving equivalence and generate the strongest conjunctive bisimulation invariant over this universe. Towards this goal, we define a strongest post-condition semantics of database transactions and automatically generate verification conditions whose validity establishes the correctness of a candidate bisimulation invariant.

We have implemented the proposed approach in a tool called **MEDIATOR** for verifying equivalence between applications written in our intermediate representation (IR), which abstracts database-driven applications as a fixed set of queries and updates to the database. To evaluate our methodology, we consider 21 database-driven applications translated into our IR and show that **MEDIATOR** can successfully verify equivalence between benchmarks extracted from real-world web applications with up to hundreds of transactions. We also show that **MEDIATOR** can handle challenging textbook examples that illustrate a wide spectrum of structural changes to the database schema. Overall, our experiments show that the proposed method is useful and practical: **MEDIATOR** can successfully verify the desired property for 10 out of 11 real-world benchmarks in under 50 seconds on average.

To summarize, this paper makes the following key contributions:

- We introduce the equivalence and refinement checking problems for database-driven applications with different schemas and conduct an empirical study that demonstrates the practical relevance of this problem.
- We present a sound and relatively complete proof methodology for showing equivalence of database-driven applications.
- We show how to enable automated reasoning over relational algebra with updates (\mathcal{T}_{RA}) by presenting an SMT-friendly encoding of \mathcal{T}_{RA} into the theory of lists.
- We define a strongest postcondition semantics of database update operations and show how to automatically infer suitable simulation invariants for verifying equivalence and refinement.
- We implement our approach in a tool called **MEDIATOR** and experimentally evaluate it on 21 benchmarks taken from real-world applications and textbook examples. Our evaluation shows that **MEDIATOR** can verify the desired property for 20 out of 21 benchmarks.

2 MOTIVATING EXAMPLE

Consider a database-driven Connected Diagnostics Platform (cdx)¹ for notifying patients about the results of their medical tests and alerting community health workers about (anonymous) positive test results in their area. This application interacts with a database that stores information about subscribers, laboratories, institutions and so on. In an earlier version of the application found on Github, the underlying database contains 16 relations and 125 attributes, but, at some point, the developers see a need to change the database schema and migrate it to a new format containing 17 relations and 131 attributes. Specifically, the earlier version of the database contains a `Subscriber` relation with three attributes, namely `sid`, which corresponds to the subscriber id, `sname`, which is

¹<https://github.com/instded/cdx/commit/3006277ce4d1c8b097f1e7243ff205bf657ad3c0>

<pre> void createSub(int id, String name, String fltr) INSERT INTO Subscriber VALUES(id, name, fltr); void deleteSub(int id) DELETE FROM Subscriber WHERE sid=id; void updateSub(int id, String name, String fltr) UPDATE Subscriber SET filter=fltr WHERE sid=id; UPDATE Subscriber SET sname=name WHERE sid=id; List<Tuple> getSubName(int id) SELECT sname FROM Subscriber WHERE sid=id; List<Tuple> getSubFilter(int id) SELECT filter FROM Subscriber WHERE sid=id; </pre> <p>(a) Before Refactoring</p>	<pre> void createSub(int id, String name, String fltr) INSERT INTO Subscriber VALUES(id, name, UUID_x); INSERT INTO Filter VALUES(UUID_x, "Filter for " + name + " subscriber", fltr); void deleteSub(int id) DELETE FROM Filter WHERE fid IN (SELECT fid_fk FROM Subscriber WHERE sid=id); DELETE FROM Subscriber WHERE sid=id; void updateSub(int id, String name, String fltr) UPDATE Filter SET params=fltr WHERE fid IN (SELECT fid_fk FROM Subscriber WHERE sid=id); UPDATE Subscriber SET sname=name WHERE sid=id; List<Tuple> getSubName(int id) SELECT sname FROM Subscriber WHERE sid=id; List<Tuple> getSubFilter(int id) SELECT params FROM Filter JOIN Subscriber ON fid=fid_fk WHERE sid=id; </pre> <p>(b) After Refactoring</p>
---	--

Fig. 2. Sample Database Refactoring. UUID_x is a unique fid of Filter relation.

the name of the subscriber, and filter, which is used to filter out irrelevant subscriptions. In the updated version, the developers decide to refactor this information into two separate relations:

Subscriber(sid, sname, fid_fk) Filter(fid, fname, params)

The new database schema now contains an additional Filter relation, and the fid_fk attribute in Subscriber is now a foreign key referring to the corresponding filter in the Filter relation.²

After refactoring the database schema in this manner, the developers also re-implement the relevant parts of the code that interact with this database. In particular, Figure 2 shows the relevant functionality before and after the database refactoring. Both versions of the code contain three methods for updating the database, namely createSub, deleteSub, and updateSub, and two methods (getSubName and getSubFilter) for querying the database. **However, the underlying implementation of these methods has changed due to the migration of the schema to a new format.** Nonetheless, we would like to be able to show that the application returns the same query results before and after the schema migration. **This verification task is non-trivial because the database transactions in the two implementations are often structurally different and operate over different relations.**

Let us now see how MEDIATOR can be used to verify the equivalence of the two versions of the cdx application before and after database refactoring. As mentioned in Section 1, our method infers a *bisimulation invariant* that relates two versions of the database. In the remainder of this discussion, let us use primed variables to refer to database relations and attributes in the refactored database. For instance, Subscriber' refers to the version of the original Subscriber relation in the refactored database.

²The actual refactoring from the Github commit history involves many other changes to the database schema, but we only consider this single modification to simplify the example.

To come up with a suitable bisimulation invariant, MEDIATOR first generates a finite universe of atomic predicates that *could* be used to relate the two versions of the database. For example, one possible predicate is $\Pi_{sid,sname}(Subscriber) = \Pi_{sid',sname'}(Subscriber')$, which states that the sid and sname attributes in relation Subscriber correspond to sid' and sname' in Subscriber', respectively. Given such predicates, MEDIATOR then tries to find a conjunctive formula that is provably a valid bisimulation invariant.

As an example, let us consider the following candidate formula Φ in the theory of relational algebra, which we will describe later in this paper:

$$\begin{aligned} \Pi_{sid,sname}(Subscriber) &= \Pi_{sid',sname'}(Subscriber') \wedge \\ \Pi_{sid,sname,filter}(Subscriber) &= \Pi_{sid',sname',params'}(Subscriber' \bowtie Filter') \end{aligned}$$

Essentially, this formula states that the sid and sname attributes of the Subscriber relation are unchanged, and the Subscriber relation in the original database can be obtained by taking the natural join of Subscriber and Filter relations in the refactored database and then projecting the relevant attributes.

Using our verification methodology, we can automatically prove that the candidate formula Φ corresponds to a valid bisimulation invariant. In particular, assuming that Φ holds before each pair of update operations U, U' from the original and revised implementations, we can show that Φ still continues to hold after executing U and U' . In other words, this means that Φ is an *inductive* bisimulation invariant. To prove the inductiveness of Φ , we use a strongest postcondition semantics for database update operations as well an automated theorem prover for the theory of relational algebra with updates.

After finding an inductive bisimulation invariant Φ , MEDIATOR still needs to ensure that Φ is strong enough to prove equivalence. For this purpose, it considers every pair of queries Q, Q' from the old and revised versions of the application and tries to prove that Q and Q' yield the same results. Using the axioms of theory of relational algebra with updates, it can be shown that $\Phi \Rightarrow Q = Q'$ is logically valid for both of the queries getSubName and getSubFilter in this application. Hence, MEDIATOR is able to prove equivalence between these two programs even though they use databases that operate over different schemas.

3 PROBLEM STATEMENT

In this section, we formalize the syntax and semantics of database-driven applications and precisely define the equivalence and refinement checking problems in this context.

3.1 Language Syntax

In the remainder of this paper, we represent a database-driven application P as a tuple $(S, \vec{T}_U, \vec{T}_Q)$, where S is the schema of the underlying database, \vec{T}_U is a vector of database update transactions, and \vec{T}_Q is a vector of database queries (see Figure 3). We collectively refer to any update or query in $\vec{T}_U \cup \vec{T}_Q$ as a *database transaction* and denote the i 'th transaction in \vec{T}_U (resp. in \vec{T}_Q) as U_i (resp. Q_i). Let us now take a closer look at the syntax of the language from Figure 3.

Database schema. The database schema S provides a logical view of how the database organizes its data. In particular, the schema describes all relations (i.e., tables) stored in the database as well as the typed attributes for each relation. More precisely, we represent the schema S as a mapping from table names R to their corresponding record types $\{a_1 : \tau_1; \dots; a_n : \tau_n\}$, which indicates that attribute a_i of table R has type τ_i . In the rest of the paper, we use the notation $dom(S)$ to denote the set of tables stored in the database.

<i>Program</i>	P	$:=$	$(S, \vec{T}_U, \vec{T}_Q)$
<i>Schema</i>	S	$:=$	$R \rightarrow \{a_1 : \tau_1; \dots; a_n : \tau_n\}$
<i>Transaction</i>	T	$:=$	$\lambda \vec{v}. U \mid \lambda \vec{v}. Q$
<i>Update</i>	U	$:=$	$\text{ins}(R, \{a_1 : v_1, \dots, a_n : v_n\}) \mid \text{del}(R, \phi) \mid \text{upd}(R, \phi, a, v) \mid U; U$
<i>Query</i>	Q	$:=$	$R \mid \Pi_\psi(Q) \mid \sigma_\phi(Q) \mid Q \bowtie_\phi Q \mid Q \cup Q \mid Q - Q$
<i>Attribute list</i>	ψ	$:=$	$a \mid \psi, \psi$
<i>Predicate</i>	ϕ	$:=$	$p \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg \phi$
<i>Atomic predicate</i>	p	$:=$	$a \odot a \mid a \odot v \mid a \in Q$
<i>Operator</i>	\odot	$:=$	$\leq \mid < \mid = \mid \neq \mid > \mid \geq$
	τ	\in	$\{Int, String, \dots\}$
	a	\in	<i>Attribute</i>
	R	\in	<i>Relation</i>
	v	\in	<i>Variable</i> \cup <i>Constant</i>

Fig. 3. Grammar of the language used in our formalization

Update transactions. An update transaction $\lambda \vec{v}. U \in \vec{T}_U$ contains a sequence of database update operations, including insertion, deletion, and modification. Specifically, the language construct $\text{ins}(R, t)$ models the insertion of tuple t into relation R , where $R \in \text{dom}(S)$ and tuple t is represented as a mapping from attributes to symbols (variable or constant). Similarly, the statement $\text{del}(R, \phi)$ removes all tuples satisfying predicate ϕ from R , and $\text{upd}(R, \phi, a, v)$ assigns value v to the a attribute of all tuples satisfying predicate ϕ in R . We assume that each database transaction occurs atomically (i.e., either all or none of the updates are committed).

Query transactions. In our language, query transactions $\lambda \vec{v}. Q \in \vec{T}_Q$ are expressed as relational algebra expressions involving projection (Π), selection (σ), join (\bowtie), union (\cup), and difference ($-$) operators. While our language allows general theta joins of the form $R_1 \bowtie_\phi R_2$, we abbreviate natural joins using the notation $R_1 \bowtie R_2$.

3.2 Language Semantics

To define the formal semantics of database-driven applications, we first need to define what we mean by an *input* to the programs defined in Figure 3. Since we consider a model in which the user interacts with the application by performing a sequence of updates and queries to the database, we consider a program input to be an *invocation sequence* ω of the form:

$$\omega = (i_1, \sigma_1); \dots; (i_{n-1}, \sigma_{n-1}); (i_n, \sigma_n)$$

where each i_j specifies an update transaction $\lambda \vec{v}. U_{i_j}$ for $j \in [1, n]$ and σ_j is its corresponding valuation, mapping values of formal parameters \vec{v} to their concrete values. The last element (i_n, σ_n) in the invocation sequence always corresponds to a *query transaction* with corresponding valuation σ_n .

Example 3.1. Consider the motivating example from Figure 2. This program can be expressed in our intermediate language as shown in Figure 4. Now, consider the following invocation sequence, assuming that the five transactions are indexed 1 – 5 from top to bottom:

$$\omega = (1, [id \mapsto 100, name \mapsto Alice, fltr \mapsto Filter1]); (4, [id \mapsto 100])$$

This sequence indicates that the user first invokes the update transaction $\text{createSub}(100, Alice, Filter1)$, followed by the query transaction $\text{getSubName}(100)$.

<pre> void createSub(int id, String name, String fltr) ins(Subscriber, (id, name, fltr)) void deleteSub(int id) del(Subscriber, sid=id) void updateSub(int id, String name, String fltr) upd(Subscriber, sid=id, filter, fltr) upd(Subscriber, sid=id, sname, name) List<Tuple> getSubName(int id) $\Pi_{\text{sname}}(\sigma_{\text{sid}=\text{id}}(\text{Subscriber}))$ List<Tuple> getSubFilter(int id) $\Pi_{\text{filter}}(\sigma_{\text{sid}=\text{id}}(\text{Subscriber}))$ </pre>	<pre> void createSub(int id, String name, String fltr) ins(Filter', (UUID_x, name, fltr)) ins(Subscriber', (id, name, UUID_x)) void deleteSub(int id) del(Filter', fid' $\in \Pi_{\text{fid_fk'}}(\sigma_{\text{sid}'=\text{id}}(\text{Subscriber'}))$) del(Subscriber', sid'=id) void updateSub(int id, String name, String fltr) upd(Filter', fid' $\in \Pi_{\text{fid_fk'}}(\sigma_{\text{sid}'=\text{id}}(\text{Subscriber'}))$, params', fltr) upd(Subscriber', sid'=id, sname', name) List<Tuple> getSubName(int id) $\Pi_{\text{sname'}}(\sigma_{\text{sid}'=\text{id}}(\text{Subscriber'}))$ List<Tuple> getSubFilter(int id) $\Pi_{\text{params'}}(\sigma_{\text{sid}'=\text{id}}(\text{Filter'} \bowtie \text{Subscriber'}))$ </pre>
(a) Before Refactoring	(b) After Refactoring

Fig. 4. Database Application in Intermediate Language.

Remark. The reader may wonder why our definition of *program input* requires all elements in an invocation sequence to be update transactions, except for the last one which is always a query. Our motivation here is to simplify the formalization of equivalence. Recall that we consider two programs to be equivalent if they yield the same answers to their corresponding queries under the same sequence of updates to the database. Since we can model invocation sequences that contain multiple queries as two different inputs, it suffices to consider inputs that contain a single query. Furthermore, we can disregard invocation sequences that do not contain any queries because the program does not “return” any output on such an input.

Figure 5 defines the denotational semantics for the language presented in Figure 3. We use the notation $\llbracket P \rrbracket_\omega$ to represent the result of the last query in ω after performing all updates on an *empty database*. Since any reachable database state can be modeled using a suitable sequence of insertions to an empty database, this assumption does not result in a loss of generality.

In our semantics, we model *database instances* Δ as a mapping from relation names to a list of tuples.³ Similarly, we model tuples as a mapping from attribute names to their corresponding values. Given a program P , an invocation sequence $\omega = \omega'; (n, \sigma)$, and a database instance Δ , we first obtain a new instance Δ' by running ω' on Δ and then evaluate the query Q_n on database instance Δ' and valuation σ . Observe that the result of a program is represented as a list of lists rather than as relations (list of maps). That is, our semantics disregards the names of attributes to enable meaningful comparison between programs over different schemas.

The semantics for update transactions in Figure 5 are described using the familiar list combinators, such as *append*, *filter*, *map*, and *fold*. In particular, $\llbracket U \rrbracket_{\sigma, \Delta}$ yields the database instance after executing update operation U with input σ on database Δ . For example, consider the semantics for $\text{ins}(R, t)$: To obtain the new database instance, we first evaluate t under valuation σ , where the notation $t[\sigma]$ denotes applying substitution σ to term t . The entry for relation R in the new database instance is

³We model relations as lists rather than bags because many libraries provide database interfaces based on ordered data structures.

$\llbracket P \rrbracket :: \text{Invocation Sequence } \omega \rightarrow \text{Instance } \Delta \rightarrow \text{List}$

$$\begin{aligned} \llbracket (S, \vec{T}_U, \vec{T}_Q) \rrbracket_{(n, \sigma), \Delta} &= \text{map}(\llbracket Q_n \rrbracket_{\sigma, \Delta}, \lambda x. \text{vals}(x)) \\ \llbracket (S, \vec{T}_U, \vec{T}_Q) \rrbracket_{(n, \sigma); \omega, \Delta} &= \llbracket (S, \vec{T}_U, \vec{T}_Q) \rrbracket_{\omega, \Delta'} \text{ where } \Delta' = \llbracket U_n \rrbracket_{\sigma, \Delta} \end{aligned}$$

$\llbracket U \rrbracket :: \text{Valuation } \sigma \rightarrow \text{Instance } \Delta \rightarrow \text{Instance}$

$$\begin{aligned} \llbracket U_1; U_2 \rrbracket_{\sigma, \Delta} &= \llbracket U_2 \rrbracket_{\sigma, \Delta'} \text{ where } \Delta' = \llbracket U_1 \rrbracket_{\sigma, \Delta} \\ \llbracket \text{ins}(R, t) \rrbracket_{\sigma, \Delta} &= \Delta[R \leftarrow \text{append}(\Delta(R), t[\sigma])] \\ \llbracket \text{del}(R, \phi) \rrbracket_{\sigma, \Delta} &= \Delta[R \leftarrow \text{filter}(\Delta(R), \lambda x. \neg \llbracket \phi \rrbracket_{\sigma, \Delta, x})] \\ \llbracket \text{upd}(R, \phi, a, v) \rrbracket_{\sigma, \Delta} &= \Delta \left[R \leftarrow \text{append} \left(\begin{array}{l} \text{filter}(\Delta(R), \lambda x. \neg \llbracket \phi \rrbracket_{\sigma, \Delta, x}), \\ \text{map}(\text{filter}(\Delta(R), \lambda x. \llbracket \phi \rrbracket_{\sigma, \Delta, x}), \lambda x. x[a \leftarrow v[\sigma]]) \end{array} \right) \right] \end{aligned}$$

$\llbracket Q \rrbracket :: \text{Valuation } \sigma \rightarrow \text{Instance } \Delta \rightarrow \text{Relation}$

$$\begin{aligned} \llbracket R \rrbracket_{\sigma, \Delta} &= \Delta(R) \\ \llbracket \Pi_\psi(Q) \rrbracket_{\sigma, \Delta} &= \text{map}(\llbracket Q \rrbracket_{\sigma, \Delta}, \lambda x. \text{filter}(x, \lambda y. \text{contains}(\text{first}(y), \psi))) \\ \llbracket \sigma_\phi(Q) \rrbracket_{\sigma, \Delta} &= \text{filter}(\llbracket Q \rrbracket_{\sigma, \Delta}, \lambda x. \llbracket \phi \rrbracket_{\sigma, \Delta, x}) \\ \llbracket Q_1 \times Q_2 \rrbracket_{\sigma, \Delta} &= \text{foldl}(\lambda ys. \lambda y. \text{append}(ys, \text{map}(\llbracket Q_2 \rrbracket_{\sigma, \Delta}, \lambda z. \text{merge}(y, z))), [], \llbracket Q_1 \rrbracket_{\sigma, \Delta}) \\ \llbracket Q_1 \bowtie_\phi Q_2 \rrbracket_{\sigma, \Delta} &= \llbracket \sigma_\phi(Q_1 \times Q_2) \rrbracket_{\sigma, \Delta} \\ \llbracket Q_1 \cup Q_2 \rrbracket_{\sigma, \Delta} &= \text{append}(\llbracket Q_1 \rrbracket_{\sigma, \Delta}, \llbracket Q_2 \rrbracket_{\sigma, \Delta}) \\ \llbracket Q_1 - Q_2 \rrbracket_{\sigma, \Delta} &= \text{foldl}(\lambda ys. \lambda y. \text{delete}(y, ys), \llbracket Q_1 \rrbracket_{\sigma, \Delta}, \llbracket Q_2 \rrbracket_{\sigma, \Delta}) \end{aligned}$$

$\llbracket \phi \rrbracket :: \text{Valuation } \sigma \rightarrow \text{Instance } \Delta \rightarrow \text{Tuple } x \rightarrow \text{Predicate}$

$$\begin{aligned} \llbracket a_1 \odot a_2 \rrbracket_{\sigma, \Delta, x} &= \text{lookup}(x, a_1) \odot \text{lookup}(x, a_2) \\ \llbracket a \odot v \rrbracket_{\sigma, \Delta, x} &= \text{lookup}(x, a) \odot v[\sigma] \\ \llbracket a \in Q \rrbracket_{\sigma, \Delta, x} &= \text{contains}(\text{lookup}(x, a), \text{map}(\llbracket Q \rrbracket_{\sigma, \Delta}, \lambda y. \text{head}(\text{vals}(y)))) \\ \llbracket \phi_1 \wedge \phi_2 \rrbracket_{\sigma, \Delta, x} &= \llbracket \phi_1 \rrbracket_{\sigma, \Delta, x} \wedge \llbracket \phi_2 \rrbracket_{\sigma, \Delta, x} \\ \llbracket \phi_1 \vee \phi_2 \rrbracket_{\sigma, \Delta, x} &= \llbracket \phi_1 \rrbracket_{\sigma, \Delta, x} \vee \llbracket \phi_2 \rrbracket_{\sigma, \Delta, x} \\ \llbracket \neg \phi \rrbracket_{\sigma, \Delta, x} &= \neg \llbracket \phi \rrbracket_{\sigma, \Delta, x} \end{aligned}$$

Fig. 5. Denotational semantics of database-driven applications. Our semantics are defined in terms of the standard list combinators *map*, *append*, *filter*, *foldl*, and *contains*. Given a map x , we write $\text{vals}(x)$ to denote the list of values stored in the map, and we think of a map as a list of (key, value) pairs. The function $\text{delete}(y, ys)$ removes the first occurrence of y in list ys . Given two maps y, z with disjoint keys, $\text{merge}(y, z)$ generates a new map that contains all key-value pairs in y and z .

obtained by appending the tuple $t[\sigma]$ to table $\Delta(R)$, which is represented as a list of tuples. Similarly, $\text{del}(R, \phi)$ filters from list R the set of all tuples that do not satisfy predicate ϕ . Finally, $\text{upd}(R, \phi, a, v)$ first obtains a new relation R_1 that contains all tuples in R that do not satisfy ϕ . It then also filters out all tuples of R that satisfy predicate ϕ , and updates the a attribute of each such tuple to a new value $v[\sigma]$. The new entry for R is then obtained by concatenating these two lists.

Let us now turn our attention to the semantics of query transactions (third part of Figure 5). The semantics are defined inductively, with the first rule for R being the base case. Since R corresponds to the name of a database table, we obtain the query result by simply looking up R in Δ . As another example, consider the semantics of the selection (σ) operator. Given a query of the form $\sigma_\phi(Q)$, we first recursively evaluate Q and obtain the query result $T = \llbracket Q \rrbracket_{\sigma, \Delta}$. We then evaluate the predicate ϕ under σ and obtain a *symbolic* predicate $p = \llbracket \phi \rrbracket_{\sigma, \Delta, x}$. In particular, predicate p is symbolic in the sense that it refers to a variable x , which ranges over tuples in T . We obtain the final query result by filtering out those rows of T that do not satisfy predicate $\lambda x. p$.

The final part of Figure 5 describes predicate semantics inductively. For instance, consider a predicate of the form $a \odot v$, where a is an attribute, \odot is a (logical) binary operator, and v is a symbol (variable or constant). Since the predicate takes as input a tuple x , we evaluate attribute a by looking up a in x , which is represented as a mapping from attributes to values. Thus, the evaluation of the predicate is given by $\text{lookup}(x, a) \odot v[\sigma]$, where the notation $v[\sigma]$ applies substitution σ to symbol v .

3.3 Equivalence and Refinement

Having defined the semantics of database-driven applications, we are now ready to precisely state the notion of semantic equivalence in this context:

Definition 3.2. (Program equivalence) A database-driven program P' is said to be semantically equivalent to another program P , denoted $P' \simeq P$, if and only if executing ω on P' yields the same result as executing ω on P for any invocation sequence ω , i.e.,

$$P' \simeq P \triangleq \forall \omega. \llbracket P' \rrbracket_\omega = \llbracket P \rrbracket_\omega$$

In the above definition, we assume that P and P' have the same number of query and update transactions.⁴ If this condition does not hold, we can immediately conclude that P and P' are not equivalent because some inputs that are valid for P are not valid for P' or vice versa. We also assume that P and P' have transactions that are supposed to be functionally equivalent at the same index; otherwise, the transactions can be syntactically re-arranged. Under these assumptions, our definition effectively states that programs P and P' are equivalent whenever their corresponding queries yield the same result under the same sequence of update transactions to the database.

While there are many real-world scenarios in which we would like to prove equivalence, there are also some cases where one application *refines* the other. For instance, consider a situation in which a web application developer changes the database schema for performance reasons, but also decides to add some new piece of information to the underlying database such that query results also include this new information. In this scenario, the updated version of the application will *not* be semantically equivalent to its prior version, but we would still like to verify that adding new features does not break existing functionality. Towards this goal, we also formally define what it means for a database-driven program P' to *refine* another program P .

Definition 3.3. (Valuation refinement) Consider two valuations σ and σ' . We say that σ' is a refinement of σ , denoted $\sigma' \leq \sigma$, if and only if σ' maps the variables that occur in σ to the same values as in σ . In other words,

$$\sigma' \leq \sigma \triangleq \forall x \in \text{dom}(\sigma), \sigma'(x) = \sigma(x)$$

⁴Note that this assumption is not a fundamental limitation. If the number of transactions in P and P' are different, we can either allow users to specify the correspondence between transactions in P and P' , or alternatively, we can enumerate all possible correspondences and see if we can prove equivalence under some (possibly one-to-many or many-to-one) mapping between transactions in P and P' .

Definition 3.4. (Input refinement) Given invocation sequences $\omega = (i_1, \sigma_1)(i_2, \sigma_2) \dots (i_n, \sigma_n)$ and $\omega' = (i'_1, \sigma'_1)(i'_2, \sigma'_2) \dots (i'_n, \sigma'_n)$, we say that ω' refines ω , denoted $\omega' \leq \omega$, if and only if ω' has the same index sequence as ω and the valuations in ω' refine the corresponding valuations in ω . That is,

$$\omega' \leq \omega \triangleq \forall k \in [1, n]. i_k = i'_k \wedge \sigma'_k \leq \sigma_k$$

Using these definitions, we can now also state what it means for an application to refine another one:

Definition 3.5. (Program refinement) Program P' is said to refine another program P , denoted $P' \leq P$, if and only if, for any invocation sequences ω', ω satisfying $\omega' \leq \omega$, executing ω on P yields a relation that is a projection of executing ω' on P' , i.e.,

$$P' \leq P \triangleq \forall \omega, \omega'. \omega' \leq \omega \rightarrow \exists L. \llbracket \Pi_L(P') \rrbracket_{\omega'} = \llbracket P \rrbracket_{\omega}$$

where $\Pi_L((S, \vec{T}_U, \vec{T}_Q)) = (S, \vec{T}_U, \vec{T}_Q')$ such that $Q'_i = \Pi_L(Q_i)$ for all Q_i in \vec{T}_Q .

As in Definition 3.2, we require that ω' and ω are valid inputs for P' and P respectively. However, unlike in Definition 3.2, we do not assume that P' contains the same number of update and query transactions in P . Our definition simply disregards the new transactions that are added by P' and only considers invocation sequences that are valid for both. Thus, intuitively, if an application P' refines P , the query results of P can be obtained by applying a projection to the corresponding query results in P' .

4 PROOF METHODOLOGY

Having defined the semantic equivalence and refinement properties for database-driven applications, let us now turn our attention to the proof methodology for showing these properties.

4.1 Proving Equivalence

A standard methodology for proving equivalence between any two systems A, B is to find a *bisimulation relation* that relates states in A with those in B [Cleaveland and Hennessy 1993]. In our case, these systems are database-driven applications, and the states that we need to relate are database instances. Our approach does not directly infer an explicit mapping between database instances, but instead finds a *bisimulation invariant* that (a) is satisfied by pairs of database instances from the two systems, and (b) is strong enough to prove equivalence.

In this paper, we prove that a bisimulation invariant Φ is valid by showing that it is inductive. That is, Φ must hold initially, and assuming that it holds for a pair of databases Δ, Δ' , it must continue to hold after executing any pair of corresponding update operations $\lambda \vec{x}.U$ and $\lambda \vec{y}.U'$.

Definition 4.1. (Inductive bisimulation invariant) Consider programs $P = (S, \vec{T}_U, \vec{T}_Q)$ and $P' = (S', \vec{T}'_U, \vec{T}'_Q)$ and suppose that P, P' contain a disjoint set of variables (which can be enforced using α -renaming if necessary). A bisimulation invariant Φ is said to be *inductive* with respect to programs P and P' if (a) Φ is satisfied by a pair of empty databases, and (b) the following Hoare triple is valid for all $\lambda \vec{x}.U_i \in \vec{T}_U$ and $\lambda \vec{y}.U'_i \in \vec{T}'_U$:

$$\{\Phi \wedge \vec{x} = \vec{y}\} U_i \parallel U'_i \{\Phi\}$$

In the above definition, the notation $U \parallel U'$ denotes the parallel execution of updates $\lambda \vec{x}.U$ and $\lambda \vec{y}.U'$. However, since programs P, P' contain a disjoint set of variables, $U \parallel U'$ is semantically equivalent to the sequential composition $U; U'$. Thus, to prove inductiveness, we need to show the validity of the Hoare triple

$$\{\Phi \wedge \vec{x} = \vec{y}\} U_i ; U'_i \{\Phi\}$$

for every pair of updates $\lambda\vec{x}.U_i$ and $\lambda\vec{y}.U'_i$ in P and P' . Also, observe that Φ must hold for a pair of empty databases in the base case because we assume that the databases are initially empty (recall Section 3).⁵

While there are many possible inductive bisimulation invariants (including *true*, for example), we need a bisimulation invariant that is strong enough to prove equivalence. According to Definition 3.2, two programs are equivalent if they yield the same result for every pair of corresponding queries $\lambda\vec{x}.Q$ and $\lambda\vec{y}.Q'$ given the same input. Thus, we can define what it means for a bisimulation invariant to be sufficient in the following way:

Definition 4.2. (Sufficiency) A formula Φ is said to be *sufficient* with respect to programs $P = (S, \vec{T}_U, \vec{T}_Q)$ and $P' = (S', \vec{T}'_U, \vec{T}'_Q)$ if, for all $\lambda\vec{x}.Q_i \in \vec{T}_Q$ and $\lambda\vec{y}.Q'_i \in \vec{T}'_Q$, we have:

$$(\Phi \wedge \vec{x} = \vec{y}) \models Q_i = Q'_i$$

Our general proof methodology for proving equivalence is to find a *sufficient, inductive bisimulation invariant* between the given pair of programs. If we can find such an invariant Φ , we know that Φ holds after executing any invocation sequence ω on P, P' , so Φ must also hold before issuing any database query. Furthermore, since Φ is a sufficient bisimulation invariant, it implies that any pair of queries yield the same result. Thus, the existence of such a bisimulation invariant Φ implies that P, P' are semantically equivalent.

THEOREM 4.3. (Soundness) Given database applications P, P' , the existence of a sufficient, inductive bisimulation invariant Φ implies $P \simeq P'$.

PROOF. We show that $\llbracket P \rrbracket_\omega = \llbracket P' \rrbracket_\omega$ by performing induction on the length of the invocation sequence ω . However, we consider the following strengthened inductive hypothesis: “If Φ is satisfied by database instances Δ, Δ' , then we have $\llbracket P \rrbracket_{\omega, \Delta} = \llbracket P' \rrbracket_{\omega, \Delta'}$ ”. Since $\llbracket P \rrbracket_\omega$ is the same as $\llbracket P \rrbracket_{\omega, \emptyset}$, and Φ is satisfied by a pair of empty databases according to Definition 4.1, this implies $\llbracket P \rrbracket_\omega = \llbracket P' \rrbracket_\omega$.

For the base case, we have $\omega = (n, \sigma)$. By assumption, $\Delta \cup \Delta'$ is a model of Φ ; thus, Φ holds initially. Since we evaluate queries Q_n and Q'_n on the same valuation σ , we also have $\vec{x} = \vec{y}$. Because Φ is a *sufficient* bisimulation invariant, this implies $Q_i = Q'_i$; hence we have $\llbracket P \rrbracket_{\omega, \Delta} = \llbracket P' \rrbracket_{\omega, \Delta'}$ for the base case. For the inductive step, suppose the invocation sequence is of the form $(n, \sigma); \omega'$, where ω' is non-empty. In this case, we have $\llbracket P \rrbracket_{\omega, \Delta} = \llbracket P \rrbracket_{\omega', \Delta_1}$ and $\llbracket P' \rrbracket_{\omega, \Delta'} = \llbracket P' \rrbracket_{\omega', \Delta_2}$ where $\Delta_1 = \llbracket U_n \rrbracket_{\sigma, \Delta}$ and $\Delta_2 = \llbracket U'_n \rrbracket_{\sigma, \Delta'}$. Since we evaluate U_n and U'_n on the same valuation σ , and since $\Delta \cup \Delta'$ is a model of Φ , we have $\Phi \wedge \vec{x} = \vec{y}$. Using Definition 4.1, we therefore know that $\Delta_1 \cup \Delta_2$ is also a model of Φ . Thus, the theorem follows immediately from the inductive hypothesis. \square

THEOREM 4.4. (Relative Completeness) Suppose we have an oracle for proving any valid Hoare triple and logical entailment. If $P \simeq P'$, then there always exists a sufficient, inductive bisimulation invariant Φ for programs P, P' .

PROOF SKETCH. Recall that we have $P \simeq P'$ iff $\llbracket P \rrbracket_\omega = \llbracket P' \rrbracket_\omega$ for an *arbitrary* invocation sequence ω . Because any valid invocation sequence starts with arbitrarily many update transactions, followed by a single query, we have $\llbracket P \rrbracket_\omega = \llbracket P' \rrbracket_\omega$ iff the imperative program shown in Figure 6 is safe. Essentially, the program in Figure 6 picks a random number of update transactions (together with a randomly chosen valuation), followed by a single, but arbitrary query transaction (and its corresponding randomly chosen valuation).

⁵This assumption is realistic in situations where database migration is performed by calling the new update methods in the application. Otherwise, the base case needs to establish that Φ holds for the initial databases.

If the assertion of program in Figure 6 is valid, the Hoare triple

$$\{true\} S_0; \text{while}(\ast)S_1; S_2 \{R = R'\}$$

must be provable (by relative completeness of Hoare logic, and assuming an oracle for proving Hoare triples that involve database transactions). Again, using relative completeness of Hoare logic, this means the following three Hoare triples must be valid, where I is an inductive invariant of the while loop:

- (1) $\{true\} S_0 \{I\}$
- (2) $\{I\} S_1 \{I\}$
- (3) $\{I\} S_2 \{R = R'\}$

(1) implies that I is satisfied by a pair of empty databases, since S_0 just initializes Δ and Δ' to be empty. (2) implies Hoare triple $\{I \wedge \vec{x} = \vec{y}\} U_i; U'_i \{I\}$ is valid since the loop body S_1 executes U_i and U'_i on the same valuation and \vec{x}, \vec{y} refer to the parameters of U_i, U'_i , respectively.

(3) implies Hoare triple $\{I \wedge \vec{x} = \vec{y}\} Q_i; Q'_i \{R = R'\}$ is valid since we again execute Q_i, Q'_i on the same valuation. Thus, we have $I \wedge \vec{x} = \vec{y} \models Q_i = Q'_i$ because R, R' are the return values of Q_i, Q'_i . Therefore, we have shown that I is a sufficient, inductive bisimulation invariant of P and P' . \square

4.2 Proving Refinement

Since our notion of refinement is a generalization of equivalence (recall Definition 3.5), our proof methodology for showing program refinement closely follows that for verifying equivalence. In particular, rather than finding a one-to-one mapping between database states as in the case of equivalence, it suffices to find a one-to-many mapping for showing refinement. Hence, our proof methodology relies on finding a *simulation invariant* rather than a stronger bisimulation invariant:

Definition 4.5. (Inductive simulation invariant) Consider programs $P = (S, \vec{T}_U, \vec{T}_Q)$ and $P' = (S', \vec{T}'_U, \vec{T}'_Q)$ and suppose that P, P' contain a disjoint set of variables. A *simulation invariant* Φ is said to be *inductive* with respect to programs P and P' if (a) Φ is satisfied by a pair of empty databases, and (b) the following Hoare triple is valid for all $\lambda\vec{x}. U_i \in \vec{T}_U$ and $\lambda\vec{y}. U'_i \in \vec{T}'_U$ where $i \in [1, |\vec{T}_U|]$:

$$\left\{ \Phi \wedge \bigwedge_{x_j \in \vec{x}} x_j = y_j \right\} U_i; U'_i \left\{ \Phi \right\}$$

Recall from Definition 3.5 that we allow the second program P' to contain more transactions than P , but the notion of refinement only talks about invocation sequences that use shared transactions from P and P' . Therefore, in Definition 4.5, we only require Φ to be preserved by pairs of update transactions that are both present in P and P' . Furthermore, since transactions in P' can take additional arguments not present in their counterparts in P , our precondition states that the arguments are pairwise equal for only the “shared” variables.

As in the equivalence scenario, finding an inductive simulation invariant Φ between P and P' is *not* sufficient for proving that P' refines P , as Φ may not be strong enough to show refinement. Hence, we also need to define what it means for an inductive simulation invariant to be *sufficient*

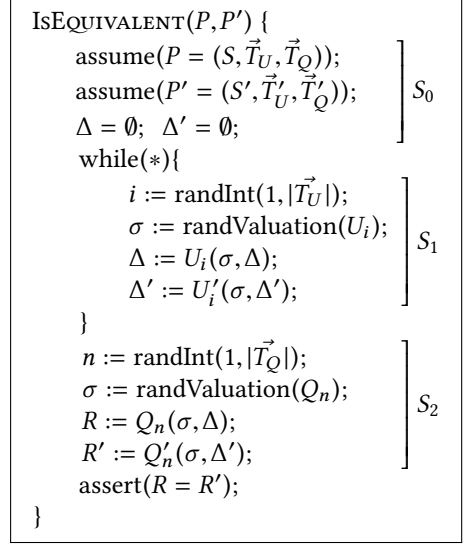


Fig. 6. Program that is safe iff P, P' are equivalent.

Formula	F	$:=$	$\text{true} \mid \text{false} \mid t = t \mid F \wedge F \mid F \vee F \mid \neg F \mid F \rightarrow F \mid \exists x.F \mid \forall x.F$
Term	t	$:=$	$x \mid T \mid \Pi_{a_1, \dots, a_n}(t) \mid \sigma_\phi(t) \mid t \times t \mid t \cup t \mid t - t \mid t \langle a_i \triangleleft v \rangle$
Predicate	ϕ	$:=$	$a_i \odot a_j \mid a_i \odot v \mid a_i \in t \mid \phi \wedge \phi \mid \phi \vee \phi \mid \neg \phi$
BinOp	\odot	$:=$	$\leq \mid < \mid = \mid \neq \mid > \mid \geq$
	T	\in	Table
	v	\in	Constant \cup Variable
	a_i	\in	Attribute
	x	\in	Variable

Fig. 7. Formula in Theory of Relational Algebra with Updates. The notation a_i represents the i 'th attribute in a relation.

for showing refinement. However, since the notion of refinement is weaker than equivalence, we also weaken our corresponding notion of sufficiency as follows:

Definition 4.6. (Projective sufficiency) A formula Φ is said to be *projectively sufficient* with respect to programs $P = (S, \vec{T}_U, \vec{T}_Q)$ and $P' = (S', \vec{T}'_U, \vec{T}'_Q)$ if for all $\lambda \vec{x}. Q_i \in \vec{T}_Q$ and $\lambda \vec{y}. Q'_i \in \vec{T}'_Q$ where $i \in [1, |\vec{T}_Q|]$, we have:

$$\left(\Phi \wedge \bigwedge_{x_j \in \vec{x}} x_j = y_j \right) \models \exists L. Q_i = \Pi_L(Q'_i)$$

Observe that the notion of projective sufficiency is weaker than Definition 4.2, as we do not require Q_i and Q'_i to yield exactly the same relation and allow the result of Q'_i to contain attributes not present in the result of Q_i . Our general proof methodology for proving refinement then relies on finding a simulation invariant that is both inductive and projectively sufficient.

THEOREM 4.7. (Soundness) Given database applications P, P' , the existence of a projectively sufficient and inductive simulation invariant Φ implies $P' \leq P$.

THEOREM 4.8. (Relative Completeness) Suppose we have an oracle for proving any valid Hoare triple and logical entailment. If $P' \leq P$, then there always exists a projectively sufficient and inductive simulation invariant Φ for programs P, P' .

The proofs of these theorems are very similar to those of Theorems 4.3 and 4.4 for equivalence.

5 SMT ENCODING OF RELATIONAL ALGEBRA WITH UPDATES

In the previous section, we defined what it means for simulation and bisimulation invariants to be inductive, but we have not fixed a logical theory over which we express these invariants. In this section, we discuss the theory of relational algebra with updates, \mathcal{T}_{RA} , and show how to enable reasoning in \mathcal{T}_{RA} using existing SMT solvers.

Figure 7 gives the syntax of the theory of relational algebra with updates, \mathcal{T}_{RA} , which we use to express simulation and bisimulation invariants. Atomic formulas in \mathcal{T}_{RA} are of the form $t = t$ where t is a term representing a relation. Basic terms include variables x and concrete tables T , and more complex terms can be formed using the relational algebra operators Π (projection), σ (selection), \times (Cartesian product), \cup (union), and $-$ (difference). In addition to these standard relational algebra operators, \mathcal{T}_{RA} also includes an *update* operator, denoted as $t \langle a_i \triangleleft v \rangle$, which represents the new relation after changing the i 'th attribute of all tuples in t to v . Observe that the theta join operator \bowtie_ϕ is expressible in this logic as $\sigma_\phi(t_1 \times t_2)$. We also write $t_1 \bowtie t_2$ as syntactic sugar for $\sigma_\phi(t_1 \times t_2)$ where ϕ is a predicate stating that the shared attributes of t_1 and t_2 are equal.

Since we view tables as lists of tuples, we axiomatize \mathcal{T}_{RA} using the theory of lists [Barrett et al. 2010]. Our axiomatization is presented in Figure 8 in the form of inference rules, where we view

$$\begin{array}{c}
\boxed{\text{get } \text{get}(i, l)} \\
\frac{l = h :: t \quad i = 0}{\text{get}(i, l) = h} \quad \frac{l = h :: t \quad i \neq 0}{\text{get}(i, l) = \text{get}(i-1, t)} \\
\\
\boxed{\text{projection } \Pi_l(t)} \\
\frac{t = []}{\Pi_l(t) = []} \quad \frac{t = h :: t_1}{\Pi_l(t) = \Pi'_l(h) :: \Pi_l(t_1)} \\
\frac{l = []}{\Pi'_l(h) = []} \quad \frac{l = a_i :: l_1}{\Pi'_l(h) = \text{get}(i, h) :: \Pi'_{l_1}(h)} \\
\\
\boxed{\text{union } t_1 \cup t_2} \\
\frac{t_1 = []}{t_1 \cup t_2 = t_2} \quad \frac{t_1 = h :: t}{t_1 \cup t_2 = h :: (t \cup t_2)} \\
\\
\boxed{\text{minus } t_1 - t_2} \\
\frac{t_2 = []}{t_1 - t_2 = t_1} \quad \frac{t_2 = h_2 :: t}{t_1 - t_2 = (t_1 -' h_2) - t} \\
\frac{t_1 = []}{t_1 -' h_2 = []} \quad \frac{t_1 = h_1 :: t_3 \quad h_1 = h_2}{t_1 -' h_2 = t_3} \\
\frac{t_1 = h_1 :: t_3 \quad h_1 \neq h_2}{t_1 -' h_2 = h_1 :: (t_3 -' h_2)} \\
\\
\boxed{\text{selection } \sigma_\phi(t)} \\
\frac{t = []}{\sigma_\phi(t) = []} \quad \frac{\phi(h) \quad t = h :: t_1}{\sigma_\phi(t) = h :: \sigma_\phi(t_1)} \\
\frac{\neg\phi(h) \quad t = h :: t_1}{\sigma_\phi(t) = \sigma_\phi(t_1)} \\
\\
\boxed{\text{Cartesian product } t_1 \times t_2} \\
\frac{t_1 = []}{t_1 \times t_2 = []} \quad \frac{t_1 = h_1 :: t}{t_1 \times t_2 = (h_1 \times' t_2) \cup (t \times t_2)} \\
\frac{t_2 = []}{h_1 \times' t_2 = []} \quad \frac{t_2 = h_2 :: t_3}{h_1 \times' t_2 = \text{cat}(h_1, h_2) :: h_1 \times' t_3} \\
\frac{h_1 = []}{\text{cat}(h_1, h_2) = h_2} \quad \frac{h_1 = c :: h}{\text{cat}(h_1, h_2) = c :: \text{cat}(h, h_2)} \\
\\
\boxed{\text{update } t \langle a_i \triangleleft v \rangle} \\
\frac{t = []}{t \langle a_i \triangleleft v \rangle = []} \quad \frac{t = h :: t_1}{t \langle a_i \triangleleft v \rangle = \text{upd}(h, i, v) :: t_1 \langle a_i \triangleleft v \rangle} \\
\frac{h = []}{\text{upd}(h, i, v) = []} \quad \frac{h = c :: h_1 \quad i = 0}{\text{upd}(h, i, v) = v :: h_1} \\
\frac{h = c :: h_1 \quad i \neq 0}{\text{upd}(h, i, v) = c :: \text{upd}(h_1, i-1, v)}
\end{array}$$

Fig. 8. Axioms in Theory of Relation Algebra with Updates. $[]$ represents an empty list nil . The binary operator $::$ denotes the list constructor cons . \times' and cat are auxiliary functions for axiomatizing Cartesian product. Π' , $-'$, and upd are auxiliary functions for axiomatizing projection, minus, and update, respectively.

$$\begin{aligned}
(a_i \odot a_j)(h) &= \text{get}(i, h) \odot \text{get}(j, h) \\
(a_i \odot v)(h) &= \text{get}(i, h) \odot v \\
(a_i \in t)(h) &= \exists j. \text{get}(0, \text{get}(j, t)) = \text{get}(i, h) \\
(\phi_1 \wedge \phi_2)(h) &= \phi_1(h) \wedge \phi_2(h) \\
(\phi_1 \vee \phi_2)(h) &= \phi_1(h) \vee \phi_2(h) \\
(\neg\phi)(h) &= \neg\phi(h)
\end{aligned}$$

Fig. 9. Auxiliary functions for selection axiom schema $\phi(h)$

tuples as *lists* of values and relations as *lists* of tuples. An attribute a_i of a tuple is simply an index into the list representing that tuple. For example, consider the axioms for *projection* $\Pi_l(t)$, which projects term t given attribute list l . We first define an auxiliary function $\Pi'_l(h)$ that projects a single tuple h given l . In particular, if the attribute list l is empty, then $\Pi'_l(h)$ yields $[]$. Otherwise, if l consists of head a_i and tail l_1 , $\Pi'_l(h)$ composes the i 'th value of h (i.e., $\text{get}(i, h)$) and the projection of tail $\Pi'_{l_1}(h)$. Similarly, $\Pi_l(t)$ is also recursively defined. If $t = []$, then $\Pi_l(t) = []$. Otherwise if $t = h :: t_1$, then $\Pi_l(t)$ composes the projection $\Pi'_l(h)$ of head h and the projection $\Pi_l(t_1)$ of tail t_1 . Also, please observe that the inference rules for selection in Figure 8 actually correspond to *axiom*

schemata rather than axioms: Because the selection operator is parameterized over a predicate ϕ , this schema needs to be instantiated for each predicate that occurs in the formula.

Example 5.1. Consider the formula $\sigma_{a_1 \geq 2}(x) = \sigma_{a_2 > 1}(y)$ in the theory of relational algebra with updates. We generate the following axioms using the axiom schemata for selection:

$$\begin{aligned}
 (1a) \quad & \forall x. (x = []) \rightarrow \sigma_{a_1 \geq 2}(x) = [] \\
 (1b) \quad & \forall x, h, t. \left(x = h :: t \rightarrow \left(\begin{array}{l} (get(1, h) \geq 2 \rightarrow \sigma_{a_1 \geq 2}(x) = h :: \sigma_{a_1 \geq 2}(t)) \wedge \\ (\neg(get(1, h) \geq 2) \rightarrow \sigma_{a_1 \geq 2}(x) = \sigma_{a_1 \geq 2}(t)) \end{array} \right) \right) \\
 (2a) \quad & \forall y. (y = []) \rightarrow \sigma_{a_2 > 1}(y) = [] \\
 (2b) \quad & \forall y, h, t. \left(y = h :: t \rightarrow \left(\begin{array}{l} (get(2, h) > 1 \rightarrow \sigma_{a_2 > 1}(y) = h :: \sigma_{a_2 > 1}(t)) \wedge \\ (\neg(get(2, h) > 1) \rightarrow \sigma_{a_2 > 1}(y) = \sigma_{a_2 > 1}(t)) \end{array} \right) \right)
 \end{aligned}$$

Remark. Since the problem of checking equivalence between a pair of relational algebra expressions is known to be undecidable [Trakhtenbrot 1950], our theory of relational algebra with updates is also undecidable. However, with the aid of some optimizations that we discuss in Section 7, we are able to determine the validity of most \mathcal{T}_{RA} formulas that we encounter in practice.

6 AUTOMATED VERIFICATION

So far, we have explained our general proof methodology and introduced a first-order theory in which we will express our bisimulation invariants. However, we have not yet explained how to *automatically* prove equivalence between programs. In this section, we discuss our strategy for proof automation. Specifically, we first discuss how to automatically prove equivalence assuming that an oracle provides bisimulation invariants (Section 6.1), and then explain how we infer them automatically (Section 6.2). Because the automation of refinement checking is very similar, this section only addresses equivalence.

6.1 Automation for Bisimulation Invariant Inductiveness

Consider two database-driven programs $P = (S, \vec{T}_U, \vec{T}_Q)$ and $P' = (S', \vec{T}'_U, \vec{T}'_Q)$, and suppose that an oracle provides a bisimulation invariant Φ between P and P' . Based on the proof methodology we outlined in Section 4, we can prove that P, P' are equivalent by showing that Φ satisfies the following conditions for any pair of updates $\lambda \vec{x}. U_i, \lambda \vec{y}. U'_i$ and any pair of queries $\lambda \vec{x}. Q_i, \lambda \vec{y}. Q'_i$:

- (1) $\Phi \wedge \vec{x} = \vec{y} \models Q_i = Q'_i$ (*Sufficiency*)
- (2) $\{\Phi \wedge \vec{x} = \vec{y} \mid U_i; U'_i \mid \Phi\}$ (*Inductiveness*)

The first condition (sufficiency) is easy to prove since we have already defined a logical theory that allows us to write terms of the form $Q_i = Q'_i$. The only small technical hiccup is that \mathcal{T}_{RA} uses the syntax a_i to denote the i 'th attribute in a relation, whereas attributes in the queries Q_i, Q'_i are *names* of attributes. To solve this difficulty, we assume a function ς which replaces attribute names s in constructs from Figure 3 with a_i , where i is the index of s . Thus, we can check whether formula Φ satisfies condition (1) by querying whether the following formula is valid modulo \mathcal{T}_{RA} :

$$(\Phi \wedge \vec{x} = \vec{y}) \rightarrow \varsigma(Q_i) = \varsigma(Q'_i)$$

However, to prove the second condition (i.e., inductiveness), we need a way to prove Hoare triples for update statements U from Figure 3. Towards this goal, we define a *strongest post-condition* semantics for update statements. Given a formula Φ over \mathcal{T}_{RA} and an update statement U , Figure 10 describes the computation of $sp(\Phi, U)$, which represents the strongest post-condition of Φ with respect to statement U .

$$\begin{aligned}
sp(\Phi, \text{ins}(R, \{a_1 : v_1, \dots, a_n : v_n\})) &= \exists x. (R = x \cup [r]) \wedge \Phi[x/R] \text{ where } r = [v_1, \dots, v_n] \\
sp(\Phi, \text{del}(R, \phi)) &= \exists x. (R = \sigma_{\neg\phi}(x)) \wedge \Phi[x/R] \\
sp(\Phi, \text{upd}(R, \phi, a, v)) &= \exists x. (R = \sigma_{\neg\phi}(x) \cup \sigma_{\phi}(x) \langle \zeta(a) \triangleleft v \rangle) \wedge \Phi[x/R] \\
sp(\Phi, U_1; U_2) &= sp(sp(\Phi, U_1), U_2)
\end{aligned}$$

Fig. 10. Strongest Postcondition for Update Transactions

To compute the strongest postcondition of Φ with respect to $\text{ins}(R, \{a_1 : v_1, \dots, a_n : v_n\})$, we think of the insertion as the assignment $R := \text{append}(R, [r])$ where r is the tuple (list) $[v_1, \dots, v_n]$. Since the union operator \cup in \mathcal{T}_{RA} corresponds to list concatenation, the new value of R after the assignment is given by $x \cup [r]$, where the existentially quantified variable x represents the old value of R .

To understand the strongest postcondition semantics of deletion, recall that $\text{del}(R, \phi)$ removes all rows in R that satisfy ϕ . Hence, we can model this statement using the assignment $R := \sigma_{\neg\phi}(R)$. Thus, when we compute the strongest postcondition of ϕ with respect to $\text{del}(R, \phi)$, the new value of R is given by $\sigma_{\neg\phi}(x)$ where the existentially quantified variable x again represents the old value of R and $\zeta(\phi)$ replaces attribute names in ϕ with their corresponding indices.

Finally, let us consider the strongest postcondition for update statements of the form $\text{upd}(R, \phi, a, v)$. Recall that this statement assigns value v to the a attribute of all tuples in R that satisfy ϕ . Specifically, according to the denotational semantics from Figure 5, we can model $\text{upd}(R, \phi, a, v)$ using the assignment statement:

$$R := (\sigma_{\neg\phi}(R)) \cup (\sigma_{\phi}(R)) \langle a \triangleleft v \rangle$$

Hence, we obtain the strongest postcondition of Φ with respect to $\text{upd}(R, \phi, a, v)$ by computing the strongest postcondition of the above assignment, where the right-hand side is a term in \mathcal{T}_{RA} (modulo changing attribute names to indices).

Definition 6.1. (Agreement) Consider database instance Δ and valuation σ , and let $\zeta(\Delta)$ denote the representation of Δ where each tuple $\{a_1 : v_1, \dots, a_n : v_n\}$ is represented as the list $[v_1, \dots, v_n]$. We say that (Δ, σ) agrees with \mathcal{T}_{RA} -formula Φ , written $(\Delta, \sigma) \sim \Phi$, iff $\zeta(\Delta) \uplus \sigma \models \Phi$.

THEOREM 6.2. (Soundness of sp) Suppose that $\llbracket U \rrbracket_{\sigma, \Delta} = \Delta'$, and let Φ be a \mathcal{T}_{RA} formula. If $(\Delta, \sigma) \sim \Phi$, then we also have $(\Delta', \sigma) \sim sp(\Phi, U)$.

PROOF. See Appendix A of the extended version of this paper [Wang et al. 2017]. \square

Now that we have defined a strongest post-condition semantics for update statements in our language, it is easy to check the correctness of the Hoare triple $\{\Phi \wedge \vec{x} = \vec{y}\} U_i; U'_i \{\Phi\}$ by simply querying the validity of the following formula modulo \mathcal{T}_{RA} :

$$sp(\Phi \wedge \vec{x} = \vec{y}, U_i; U'_i) \rightarrow \Phi$$

6.2 Bisimulation Invariant Synthesis

So far, we have discussed how to automate the proof that Φ is an inductive and sufficient bisimulation invariant. However, since we do not want users to manually provide such bisimulation invariants, our verification algorithm automatically infers them using *monomial predicate abstraction* [Ball et al. 2005; Das et al. 1999; Lahiri and Qadeer 2009].

Our technique for inferring suitable bisimulation invariants is shown in Algorithm 1. Given two programs P, P' with corresponding schemas S, S' , the **VERIFY** procedure first generates the universe of all predicates that may be used in the bisimulation invariant (line 4). We generate all such predicates by instantiating the following database of predefined templates:

Algorithm 1 Verification Algorithm for Database Application Equivalence

```

1: procedure VERIFY( $P, P'$ )
2:   Input: program  $P = (S, \vec{T}_U, \vec{T}_Q)$  and  $P' = (S', \vec{T}'_U, \vec{T}'_Q)$ 
3:   Output: Bisimulation invariant  $\Phi$  establishing equivalence or  $\perp$  to indicate failure
4:    $\mathcal{P} := \text{GetAllPredicates}(S, S')$ 
5:    $\Phi := \bigwedge_{\varphi \in \mathcal{P}} \varphi$ 
6:   while CheckSufficiency( $\Phi, \vec{T}_Q, \vec{T}'_Q$ ) do
7:      $\text{ind} := \text{true}$ 
8:     for each  $\varphi \in \mathcal{P}$  do
9:       if  $\neg \text{CheckInductiveness}(\Phi, \vec{T}_U, \vec{T}'_U, \varphi)$  then
10:         $\text{ind} := \text{false}; \mathcal{P} := \mathcal{P} \setminus \{\varphi\}; \Phi := \bigwedge_{\varphi \in \mathcal{P}} \varphi;$ 
11:        break
12:     if  $\text{ind}$  then return  $\Phi;$ 
13:   return  $\perp;$ 

14: procedure CHECKSUFFICIENCY( $\Phi, \vec{T}_Q, \vec{T}'_Q$ )
15:   for each  $\lambda \vec{x}. Q_i \in \vec{T}_Q$  and  $\lambda \vec{y}. Q'_i \in \vec{T}'_Q$  do
16:     if  $\mathcal{T}_{RA} \not\models (\Phi \wedge \vec{x} = \vec{y} \rightarrow \varsigma(Q_i) = \varsigma(Q'_i))$  then return false;
17:   return true;

18: procedure CHECKINDUCTIVENESS( $\Phi, \vec{T}_U, \vec{T}'_U, \varphi$ )
19:   for each  $\lambda \vec{x}. U_i \in \vec{T}_U$  and  $\lambda \vec{y}. U'_i \in \vec{T}'_U$  do
20:     if  $\mathcal{T}_{RA} \not\models (sp(\Phi \wedge \vec{x} = \vec{y}, U_i; U'_i) \rightarrow \varphi)$  then return false;
21:   return true;

```

- (1) $\Pi_L(R) = \Pi_{L'}(R')$
- (2) $\Pi_L(R_1 \bowtie R_2) = \Pi_{L'}(R'_1)$
- (3) $\Pi_L(R) = \Pi_{L'}(R'_1 \bowtie R'_2)$
- (4) $\Pi_L(R_1 \bowtie R_2) = \Pi_{L'}(R'_1 \bowtie R'_2)$

In these templates, L and R represent an attribute list and a relation under schema S , while L' and R' represent the attribute list and relation under schema S' . Please note that we only consider templates with at most one join operator on each side. Any predicate containing a longer join chain can be decomposed into several predicates of these forms.

Once we generate the universe \mathcal{P} of all predicates that may be used in the invariant, we perform a fixed point computation in which we iteratively *weaken* the candidate bisimulation invariant. Initially, the candidate bisimulation invariant Φ starts out as the conjunction of all predicates in our universe. During the fixed point computation (lines 6–12 in Algorithm 1), the candidate invariant Φ is always stronger than the actual bisimulation invariant. Hence, if we get to a point where Φ is not strong enough to show equivalence, we conclude that the program cannot be verified using conjunctive formulas over our templates (line 13). On the other hand, assuming that Φ is strong enough to prove equivalence, we then proceed to check whether Φ is inductive (lines 7–12). If it is, the strongest postcondition of $\Phi \wedge \vec{x} = \vec{y}$ must logically imply φ for every predicate φ used in Φ . If some predicate φ is not preserved by a pair of updates (i.e., call to `CheckInductiveness` returns false), we then remove φ from both Φ and our universe of predicates \mathcal{P} . We continue this process

1. $A \cup [] = A$ (\cup nil)
2. $A \bowtie [] = []$ (\bowtie nil)
3. $A - A = []$ ($-$ nil)
4. $\Pi_L(A \cup B) = \Pi_L(A) \cup \Pi_L(B)$ ($\Pi \cup$ distributivity)
5. $\sigma_\phi(A \cup B) = \sigma_\phi(A) \cup \sigma_\phi(B)$ ($\sigma \cup$ distributivity)
6. $(A \cup B) \bowtie C = (A \bowtie C) \cup (B \bowtie C)$ ($\bowtie \cup$ distributivity)
7. $\sigma_\phi(A) \bowtie B = \sigma_\phi(A \bowtie B)$ ($\sigma \bowtie$ associativity1)
8. $A \bowtie \sigma_\phi(B) = \sigma_\phi(A \bowtie B)$ ($\sigma \bowtie$ associativity2)
9. $\sigma_\phi(\sigma_\phi(A)) = \sigma_\phi(A)$ (σ idempotence)
10. $(\Pi_L(A) = \Pi_{L'}(B) \wedge \phi \leftrightarrow \phi'[L/L']) \rightarrow \Pi_L(\sigma_\phi(A)) = \Pi_{L'}(\sigma_{\phi'}(B))$ ($\Pi \sigma$ introduction)
11. $\Pi_L(A) = \Pi_{L'}(B) \rightarrow \Pi_L(A \langle L_i \triangleleft v \rangle) = \Pi_{L'}(B \langle L'_i \triangleleft v \rangle)$ ($\Pi \langle \triangleleft \rangle$ introduction)

Fig. 11. List of additional (redundant) axioms used for proving validity.

of weakening the invariant until it becomes an inductive bisimulation invariant, or we prove that no such invariant exists over our universe of predicates.

7 IMPLEMENTATION

We have implemented the proposed verification technique in a new tool called **MEDIATOR**, which consists of approximately 10,500 lines of Java code in total. **MEDIATOR** utilizes the Z3 SMT solver [de Moura and Bjørner 2008] to automate reasoning over the theory of relational algebra with updates. In particular, we decide the validity of a \mathcal{T}_{RA} formula ϕ by asking Z3 whether the \mathcal{T}_{RA} axioms logically imply ϕ . All queries to the solver are configured to have a time budget of 2 seconds, and we assume that the answer to any query exceeding the time budget is “invalid”. In the remainder of this section, we describe several important optimizations that we have found necessary for making **MEDIATOR** practical.

Redundant axioms. During the development of **MEDIATOR**, we have found that many validity queries cannot be resolved due to Z3’s limited capabilities for performing inductive reasoning. In particular, some of the \mathcal{T}_{RA} theorems needed for proving equivalence require performing structural induction over lists; but Z3 times out in most of these cases. In our implementation, we address this issue by providing a redundant set of additional axioms, which are logically implied by the \mathcal{T}_{RA} axioms. Figure 11 shows a representative subset of the additional theorems that we use when issuing validity queries to Z3. Because these axioms alleviate the need for performing induction, many queries that would otherwise time out can now be successfully proven using Z3. We would like to emphasize that we came up with these redundant axioms during tool development and did not have to add any additional axioms while evaluating **MEDIATOR** on real-world examples in our experiments.

Conjunctive queries. While the full theory of relational algebra with updates is undecidable, we have identified a class of formulas for which we can come up with an optimization to check validity. Let us call a query *conjunctive* if it uses only projection, selection, and equi-join and all predicates are conjunctions of equalities. As pointed out in prior work, two conjunctive queries are equivalent under bag semantics if (and only if) they are syntactically isomorphic [Cohen et al. 1999; Green 2009]. Inspired by their work, we optimize the validity checking of formulas of the form $\Phi \rightarrow Q_i = Q'_i$, which arise when checking sufficiency of a candidate bisimulation invariant Φ . If Q_i, Q'_i are conjunctive queries, we can use the schema mapping induced by Φ to rewrite the query Q_i to another query Q''_i such that Q'_i, Q''_i refer to the same schema elements. If they are

syntactically the same modulo reordering of equalities, we can conclude the original formula is valid; otherwise, we check its \mathcal{T}_{RA} -validity using an SMT solver.

Invariant synthesis. Recall from Algorithm 1 that our verification procedure looks for conjunctive invariants over a universe of predicates \mathcal{P} . While these predicates are constructed from a small set of pre-defined templates, the number of possible instantiations of these templates grows quickly in the number of attributes and relations in the database schema. To prevent a blow-up in the size of the universe \mathcal{P} , we use the implementation of insertion transactions to rule out infeasible predicates. For example, we only generate a predicate $\Pi_{[a_1, a_2, \dots, a_n]}(A) = \Pi_{[b_1, b_2, \dots, b_n]}(B)$ if there are two corresponding insertion transactions U, U' such that U inserts its argument x_i to attribute a_i of relation A , whereas U' inserts x_i into attribute b_i of relation B . Similarly, we only generate a predicate of the form $\Pi_{[a_1, a_2, \dots, a_n]}(A) = \Pi_{[b_1, b_2, \dots, b_n]}(B \bowtie C)$ if B and C can be joined and there are two corresponding transactions U, U' such that U inserts its argument x_i to attribute a_i of relation A , whereas U' inserts the same argument into attribute b_i of relation B or C . We have found these heuristics work quite well in that they do not lead to a loss of completeness in practice but significantly reduce the number of predicates considered by the invariant synthesis algorithm.

Proving refinement. Recall that proving refinement requires showing that the inferred simulation invariant Φ is projectively sufficient, i.e.,

$$\left(\Phi \wedge \bigwedge_{x_j \in \vec{x}} x_j = y_j \right) \models \exists L. Q_i = \Pi_L(Q'_i)$$

In our implementation, we determine the \mathcal{T}_{RA} -validity of this formula by instantiating the existentially quantified variables L with attributes in the database schema and check validity for each possible instantiation. In particular, suppose that Q_i, Q'_i can contain attributes A, A' respectively. Each instantiation of L essentially corresponds to a mapping M such that $M(A) \subseteq A'$. Our implementation rank-orders candidate mappings based on similarity metrics between attribute names and tries more likely instantiations first.

8 EVALUATION

In this section, we evaluate the ideas proposed in this paper by performing two experiments. The goal of our first experiment is to investigate the practical relevance of the class of equivalence checking problems that we address in this paper. In our second experiment, we evaluate the practicality and usefulness of the MEDIATOR tool by using it to verify equivalence/refinement between different versions of database-driven web applications.

8.1 Study of Real-World Web Applications

To the best of our knowledge, there is no prior work on proving equivalence between database-driven applications, so one may wonder whether this problem is important in practice. In other words, do developers need to re-implement parts of a web application due to changes to the underlying database schema? While there are many articles and on-line posts that attest to the frequency and difficulty of this task, we nonetheless perform an empirical study that aims to answer precisely this question.

In our study, we collect 100 database-driven web applications that (a) are written in Ruby-on-Rails and (b) have at least 400 commits on Github. We choose to study applications written in Ruby-on-Rails because the Rails framework provides a convenient way to view changes to the database schema. Among the applications we have studied, the underlying database schema had been changed *at least* once in *all* 100 applications, with the average application exhibiting around 28 schema changes over their lifetime. Table 1 categorizes the schema changes that we

Table 1. Summary of categorization

Addition	Introduce new relations and/or attributes.
Deletion	Remove existing relations and/or attributes.
Type	Change the type of attributes in the schema.
Rename	Rename existing relations and/or attributes in the schema.
Structure	Change the structure of the schema, including but not limited to moving attributes from one relation to another, splitting a relation into multiple ones, merging multiple relations into one.
Index	Add, delete, or modify indices of the database system.
Constraint	Add, delete, or modify constraints of the database system.
Other	Bulk schema change of several other categories.

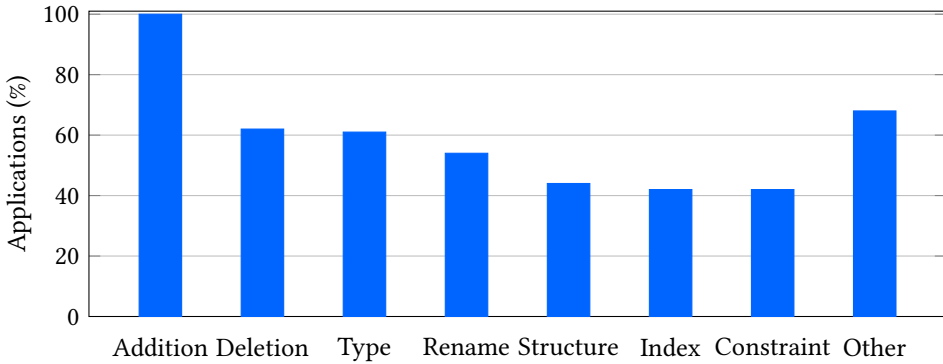


Fig. 12. Empirical Study. Percentage of applications that have different categories of database schema changes.

have encountered, and Figure 12 shows the percentage of applications that have undergone schema change of a certain type (addition, deletion etc.). For example, according to Figure 12, some attributes or relations were deleted from 62% of these applications at least once in the commit history.

Of course, not all schema change categories listed in Table 1 require significant rewriting of parts of the application code. For example, 100% of the applications in our dataset exhibit a schema change induced by the addition of relations or attributes, but such additions typically require only minor changes to the application code. Therefore, many of the schema change categories listed in Table 1 (e.g., addition, renaming) are not particularly interesting from a verification perspective. However, one of these categories, namely *structure*, requires significant rewriting of the application code. In our definition, structural schema changes include splitting a relation into multiple ones, moving attributes from one relation to another, merging multiple relations into a single one etc. As shown in Figure 12, 44% of the applications in our dataset have undergone a structural schema change at least once during their life time. Therefore, this empirical study demonstrates that structural schema changes are quite common and that web application developers can benefit from equivalence checking techniques that can handle non-trivial changes to the database schema.

8.2 Evaluation of MEDIATOR

In the previous subsection, we demonstrated the importance of the problem that we introduce in this paper; however, we have not yet evaluated the feasibility of our solution. Our goal in this

Table 2. Benchmark source and description. The textbook refers to [Ambler and Sadalage 2006] and the Oracle tutorial refers to [Oracle 2005].

	ID	Source	Description
textbook bench	1	Oracle tutorial	Merge relations
	2	Oracle tutorial	Split relations
	3	Textbook	Split relations
	4	Textbook	Merge relations
	5	Textbook	Move attributes
	6	Textbook	Rename attributes
	7	Textbook	Introduce associative relations
	8	Textbook	Replace the surrogate key with natural key
	9	Textbook	Introduce new attributes
	10	Textbook	Denormalization
real-world bench	11	cdx	Rename attributes and split relations
	12	coachup	Split relations
	13	2030Club	Split relations
	14	rails-ecomm	Split relations and introduce new attributes
	15	royk	Introduce and move attributes
	16	MathHotSpot	Rename relations and move attributes
	17	gallery	Split relations
	18	DeeJBase	Rename attributes and split relations
	19	visible-closet-1	Split relations
	20	visible-closet-2	Move attributes to a polymorphic relation
	21	probable-engine	Merge relations

section is to evaluate our *solution* by using *MEDIATOR* to prove equivalence/refinement between different versions of 21 database-driven web applications containing over 1000 transactions in total.

Benchmarks. To perform this evaluation, we collect benchmarks from two different sources, namely challenging refactoring examples from textbooks/tutorials and different versions of web applications collected from Github. The textbook examples are useful for evaluating *MEDIATOR*, as they illustrate challenging database refactoring tasks that require non-trivial changes to the application code. The remaining half of the benchmarks are real-world web applications taken from the dataset used in our empirical evaluation. Specifically, we evaluate *MEDIATOR* on two different versions A, B of the application such that (a) A, B are consecutive versions in the commit history, (b) B is obtained from A by performing a structural schema change that requires rewriting parts of the application code, and (c) B is meant to be equivalent to (or a refinement of) A . Table 2 describes the source of each benchmark and changes to the database schema between the two versions. Since our current implementation requires manually translating the web application to our IR representation, we only used the first 10 real-world applications that meet the afore-mentioned criteria. ⁶

Results. Table 3 summarizes the results of our evaluation of *MEDIATOR* on these benchmarks. For each benchmark, the column labeled *Type* shows whether we used *MEDIATOR* to check refinement (\leq) or equivalence (\approx), and *Trans* shows the number of transactions in each application. The next

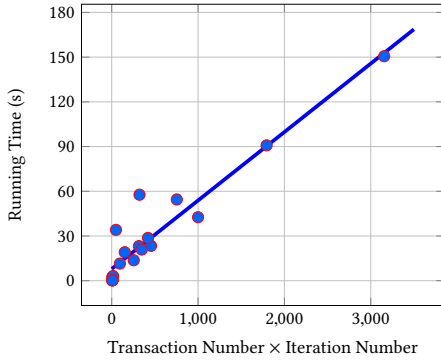
⁶It took us approximately three days of manual effort to translate all of these applications to our IR. We plan to automate this transformation in future work.

Table 3. Summary of experimental results. All experiments are performed on a computer with Intel Xeon(R) E5-1620 v3 CPU and 32GB of memory, running Ubuntu 14.04 operating system.

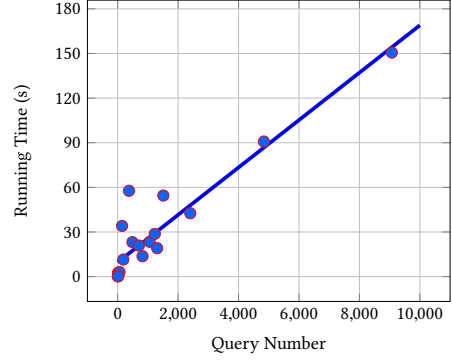
	ID	Type	Trans	Source Schema		Target Schema		Status	Time (s)	Iters	Queries
				Rels	Attrs	Rels	Attrs				
textbook bench	1	\approx	4	2	8	1	6	✓	2.4	2	10
	2	\approx	19	3	17	7	25	✓	11.5	5	188
	3	\approx	10	1	6	2	7	✓	2.5	2	32
	4	\approx	10	2	7	1	6	✓	0.3	1	16
	5	\approx	7	2	5	2	5	✓	34.1	7	147
	6	\approx	5	1	2	1	2	✓	0.2	1	5
	7	\approx	8	2	5	3	6	✓	3.1	2	61
	8	\approx	10	2	9	2	8	✓	0.4	1	18
	9	\preceq	8	2	7	2	8	✓	0.3	1	14
	10	\approx	14	3	10	3	13	✓	57.7	23	374
real-world bench	11	\approx	138	16	125	17	131	✓	90.8	13	4840
	12	\approx	45	4	51	5	55	✓	23.2	7	489
	13	\approx	125	15	155	16	159	✓	42.6	8	2403
	14	\preceq	65	8	69	9	75	✓	23.4	7	1059
	15	\preceq	151	19	152	19	155	✓	19.1	1	1307
	16	\approx	54	7	38	8	42	✓	20.9	6	701
	17	\approx	58	7	52	8	57	✓	54.5	13	1512
	18	\approx	70	10	92	11	97	✓	28.7	6	1228
	19	\approx	263	26	248	27	252	✓	150.6	12	9072
	20	\approx	267	28	262	29	261	×	-	-	-
	21	\approx	85	12	83	11	78	✓	13.7	3	823

two columns provide information about the number of relations and total number of attributes in the source and target schema, respectively. The last four columns provide information about MEDIATOR results: *Status* shows whether MEDIATOR was able to verify the desired property (i.e., equivalence or refinement), and *Time* provides total running time in seconds. The column labeled *Iters* shows the number of iterations that MEDIATOR takes to find an inductive simulation (or bisimulation) invariant. Finally, the last column labeled *Queries* shows the number of validity modulo \mathcal{T}_{RA} checks issued by MEDIATOR.

As we can see from Table 3, MEDIATOR is able to successfully verify the desired property for 20 out of 21 benchmarks. The running time of the tool ranges between 0.2 seconds for small textbook examples with a few transactions to 150 seconds for large, real-world benchmarks with hundreds of transactions. As expected, the running time of MEDIATOR on real-world benchmarks is typically much longer (46.8 seconds on average) than on textbook examples (11.3 seconds on average). However, some textbook examples (namely benchmarks 5 and 10) take longer than some of the real-world examples because many iterations are required to find an inductive bisimulation invariant. As shown in Figure 13(b), the running time of MEDIATOR is roughly linear in the number of validity queries to the SMT solver. Because the number of validity checks depends on the number of transactions in the application as well as the number of iterations required for finding an inductive bisimulation invariant, Figure 13(a) also shows that system's running time is roughly linear with respect to $\#transactions \times \#iterations$.



(a) Running time with respect to number of transactions and iterations



(b) Running time with respect to number of \mathcal{T}_{RA} validity queries

Fig. 13. Relationship between running time, transaction number, iteration number and query number

Cause of false positives. As shown in Table 3, MEDIATOR fails to verify equivalence for benchmark 20, where the schema change involves moving the shared attributes of two relations into a new polymorphic relation. Upon further inspection, we determined this warning to be a false positive that is caused by a shortcoming of our inference engine for synthesizing bisimulation invariants. In particular, proving equivalence of this benchmark requires a bisimulation invariant of the form $\Pi_L(R) = \Pi_{L'}(\sigma_\phi(R'))$, which is currently not supported in our implementation (recall Section 6.2). While it is possible to extend our templates to include predicates of this form, this modification would significantly increase the search space, as the selection predicate ϕ can be instantiated in *many* different ways.

9 RELATED WORK

The research problem that we address in this paper is related to a long line of work from the programming languages and databases communities. In what follows, we survey papers that are most relevant and explain how they differ from our approach.

Translation Validation. One of the most well-known applications of equivalence checking is *translation validation*, where the goal is to prove that the compiled version of the code is equivalent to the original one. [Necula 2000; Pnueli et al. 1998; Rinard 1999; Stepp et al. 2011; Zaks and Pnueli 2008; Zuck et al. 2003]. More recent work extends translation validation to *parameterized equivalence checking* (PEC), which aims to prove equivalence between templated programs representing many concrete programs [Kundu et al. 2009]. Most of the work in this area focuses on imperative programs and proves equivalence by inferring some form of bisimulation relation. Another common technique for proving equivalence is to generate a so-called *product program* [Barthe et al. 2011; Zaks and Pnueli 2008] and reduce the equivalence checking problem to the safety verification of a single program. Rather than validating the correctness of compiler optimizations, our goal in this work is to show equivalence between database-driven programs before and after changes to the database schema. Our bisimulation invariants relate database states rather than program variables and are expressed in the theory of relational algebra with updates instead of standard first-order theories directly supported by SMT solvers.

Contextual Equivalence. There has also been a significant body of work on verifying *contextual equivalence*, where the goal is to determine whether two expressions are equivalent under any program context. One important application of contextual equivalence is to identify compiler optimization opportunities, particularly in the context of functional programming languages. For example, Sumii and Pierce define an untyped call-by-value lambda calculus with sealing and present a bisimulation-based approach to prove contextual equivalence [Sumii and Pierce 2004]. They later present another sound and complete proof methodology based on bisimulation relations, but apply it to a lambda calculus with full universal, existential, and recursive types [Sumii and Pierce 2005]. Koutavas and Wand extend this line of work to prove contextual equivalence in an untyped lambda calculus with an explicit store by introducing a new notion of bisimulation. Their method enables constructive proofs in the presence of higher-order functions [Koutavas and Wand 2006b]. They also extend the same proof technique to the imperative untyped object calculus [Koutavas and Wand 2006a]. Sangiorgi et al. step further and develop a notion of *environmental bisimulation* for higher-order languages. Their technique does not require induction on evaluation derivations and is applicable to different calculi, ranging from pure lambda calculus to higher-order π -calculus [Sangiorgi et al. 2007, 2011]. Existing techniques for proving contextual equivalence offer a limited degree of automation and do not address database-driven applications.

Regression Verification. Regression verification is concerned with checking equivalence across different versions of an evolving program [Felsing et al. 2014; Godlin and Strichman 2009; Lahiri et al. 2012; Person et al. 2008]. Unlike translation validation where the two programs are expressed at different levels of abstraction and transformations are intra-procedural, regression verification deals with two programs across a refactoring, feature addition or bug-fix with changes possibly spread across procedures. Approaches range from the use of differential symbolic execution [Person et al. 2008] for loop-free and recursion-free programs, techniques based on uninterpreted functions to deal with mutual recursion [Godlin and Strichman 2009; Lahiri et al. 2012], use of mutual summaries for checking heap-manipulating programs [Hawblitzel et al. 2013; Wood et al. 2017], to the use of standard invariant inference for inferring intermediate relational invariants [Felsing et al. 2014; Lahiri et al. 2013]. Because the invariants and assertions relate program variables, the resulting verification conditions are expressible in first-order theories supported by SMT solvers. Existing regression verification techniques do not handle database-driven applications and therefore cannot reason about changes to the database schema.

Relational Program Logics. The problem of verifying equivalence between a pair of programs can be viewed as a special kind of *relational verification* problem in which the goal is to prove relational Hoare triples of the form $\{P\} S_1 \sim S_2 \{Q\}$. Here, P is a relational pre-condition that relates inputs to programs S_1, S_2 , and Q is a relational post-condition that relates their outputs. In the context of equivalence checking, the pre-condition simply assumes equality between program inputs, and the post-condition asserts equality between their outputs. Prior work has presented program logics, such as Relational Hoare Logic and Cartesian Hoare logic, for showing such relational correctness properties [Benton 2004; Sousa and Dillig 2016; Yang 2007]. As mentioned earlier, another common technique for proving relational correctness properties is to construct a product program $S_1 \times S_2$, which is semantically equivalent to $S_1; S_2$ but that is somehow easier to verify [Barthe et al. 2011, 2013]. In contrast to existing relational verification techniques that work on imperative programs, our approach addresses database-driven programs that work over different schema. Furthermore, while the main focus of this paper is to verify equivalence/refinement between programs, we believe our technique can be easily extended for proving other relational correctness properties.

Testing and Verification of Database-Driven Applications. There has been a significant body of prior work on testing and analyzing database-driven applications [Artzi et al. 2008; Benedikt et al. 1998; Chays et al. 2000; Deutsch et al. 2005, 2007; Emmi et al. 2007; Gligoric and Majumdar 2013; Near and Jackson 2012; Vianu 2009; Wassermann et al. 2008]. Some of these techniques aim to verify integrity constraints, find shallow bugs, or identify security vulnerabilities, while others attempt to uncover violations of functional correctness properties. For example, [Benedikt et al. 1998] statically verifies that database transactions preserve integrity constraints by means of computing weakest preconditions. As another example, the Agenda framework generates test cases by randomly populating the database with tuples that satisfy schema constraints [Chays et al. 2000; Deng et al. 2005], and several papers use concolic testing to find crashes or SQL injection vulnerabilities [Artzi et al. 2008; Wassermann et al. 2008]. Gligoric and Majumdar describe an explicit state model checking technique for database-driven applications and use this technique to find concurrency bugs [Gligoric and Majumdar 2013].

There have also been proposals for checking functional correctness of database-driven applications. For example, Near and Jackson present a bounded verifier that uses symbolic execution to check functional correctness properties specified using an extension of the RSpec specification language [Near and Jackson 2012]. As another example, the WAVE project allows users to specify functional correctness properties using LTL formulas and model checks a given database-driven application against this specification [Deutsch et al. 2005, 2007; Vianu 2009]. The WAVE framework can also be used to automatically synthesize web applications from declarative specifications [Deutsch et al. 2007]. However, we are not aware of any existing work on verifying relational correctness properties of database-driven applications.

Transformation of Database-Driven Applications. There has also been some prior work on optimizing database-driven applications [Cheung et al. 2013; Delaware et al. 2015]. For example, the QBS system detects performance bugs in web applications and repairs them by replacing inefficient Java code with SQL queries [Cheung et al. 2013]. Similar to our work, QBS also models database relations using lists and defines a *theory of ordered relations* for reasoning about loop invariants. However, the theory of ordered relations proposed by Cheung et al. does not allow reasoning about updates to database relations.

Another related work in this space is the FIAT system for synthesizing abstract datatypes (ADTs) with SQL-like operations [Delaware et al. 2015]. Given a user-provided specification, FIAT uses domain-specific refinement theorems and automation tactics to transform the given specification into an (ideally efficient) implementation. However, unlike our approach, FIAT requires different implementations to use the same schema.

Query Equivalence. Query equivalence has been a long-standing problem in the database community. For example, Chandra and Merlin study the equivalence of conjunctive queries under set semantics and show that every conjunctive query has a unique representation up to isomorphism [Chandra and Merlin 1977]. Aho et al. use tableaux to represent conjunctive queries and present a polynomial time algorithm for deciding equivalence between certain kinds of conjunctive queries [Aho et al. 1979]. Sagiv and Yannakis generalize this tableau approach to union and difference operators, but place limitations on the use of difference [Sagiv and Yannakis 1980]. In more recent work, Green studies equivalence between conjunctive queries and gives an algorithm for deciding their equivalence [Green 2009]. As mentioned in Section 7, our implementation leverages insights from this work when checking validity of certain classes of \mathcal{TR}_A formulas.

In the past year, there has been significant activity on proving query equivalence using interactive theorem provers and constraint solvers. Specifically, Chu et al. define a new semantics for SQL

based on K-Relations and homotopy type theory and develop a Coq library to interactively prove equivalence between SQL queries [Chu et al. 2017b]. Another recent work by Chu et al. provides a greater degree of automation by incorporating constraint solvers [Chu et al. 2017a]. Specifically, their tool, COSETTE, first translates each SQL query into a corresponding logical formula and uses an SMT solver to find a counterexample that shows the queries are *not* equivalent. If COSETTE fails to find a counterexample, it then tries to prove equivalence using the Coq theorem prover augmented with a library of domain-specific tactics. While tools like COSETTE could be useful for deciding validity of some of our \mathcal{T}_{RA} formulas, existing tools do not support reasoning about updates to the database.

Schema Evolution. In the database community, *schema evolution* refers to the problem of evolving database schemas to adapt to requirement changes. Schema evolution also typically requires data migration and synchronization of legacy transactions, and many papers from the database community address the problem of facilitating schema evolution [Caruccio et al. 2016; Curino et al. 2013; Fagin et al. 2011; Rahm and Bernstein 2006; Visser 2008]. For example, the PRISM and PRISM++ projects [Curino et al. 2013] investigate techniques for automatically evolving transactions using a given set of SQL-based schema modification operators.

As another example, Fagin et al. study the schema mapping adaptation problem using composition and inversion operators [Fagin et al. 2011] and work by Visser utilizes strategy combinators and point-free program calculation to handle coupled transformation of schemas, data instances, queries, and constraints [Visser 2008]. In contrast to our method, existing work on schema evolution cannot be used to prove equivalence between any pair of database-driven applications.

Schema Equivalence. There has also been some work on proving schema equivalence under various different definitions of equivalence [Beeri et al. 1979; Miller et al. 1993; Rosenthal and Reiner 1994]. Under one definition, two schemas are considered equivalent if there is a bijection from the set of database instances from one schema to that of another [Miller et al. 1993; Rosenthal and Reiner 1994]. According to another definition, two schemas S_1, S_2 are equivalent if there is a query mapping from S_1 to S_2 such that its inverse is also a valid mapping from S_2 to S_1 [Albert et al. 1999; Atzeni et al. 1982; Hull 1986]. While many of these papers provide algorithms for deciding schema equivalence according to these definitions, they do not address the problem of verifying equivalence between applications that operate over databases with different schemas.

10 LIMITATIONS

As a research prototype, our current implementation of MEDIATOR has a number of limitations: First, MEDIATOR analyzes programs that are written in the database transaction language given in Figure 3. Hence, the applicability of MEDIATOR relies on translating the original database-driven application to our IR, which abstracts programs as a fixed set of queries and updates to the database. Thus, programs that use dynamically generated transactions or control-flow constructs cannot be translated into our IR. Second, MEDIATOR synthesizes simulation and bisimulation invariants by finding the strongest conjunctive formula over a given class of predicates. However, as exemplified by the false positive from our evaluation, MEDIATOR may not be able to prove equivalence if the bisimulation requires additional predicates (or boolean connectives) beyond the ones we consider. Third, MEDIATOR axiomatizes \mathcal{T}_{RA} using the theory of lists, which is also undecidable. Therefore, the SMT solver may time-out when checking validity queries over the theory of lists. Fourth, MEDIATOR can only be used to prove equivalence but not *disequivalence*. In particular, MEDIATOR cannot provide witnesses to prove that two applications are indeed not equivalent. Finally, our verification technique proves equivalence under *list semantics*. Thus, if a web application uses

set/bag semantics to represent the results of database queries, MEDIATOR may end up reporting false positives. However, despite these limitations, our evaluation shows that MEDIATOR is still practical and that it can verify equivalence between different versions of many real-world web applications.

11 CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced the problem of verifying equivalence between applications that operate over databases with different schemas. As shown by our empirical study from Section 8.1, this problem is practically relevant because many web applications undergo a structural schema change that requires re-implementation of parts of the application code. Our technique for verifying equivalence relies on finding an inductive bisimulation invariant over the theory of relational algebra with updates and automatically discharges verification conditions that are needed to prove equivalence. We have implemented this technique in a new tool called MEDIATOR and evaluate it on different versions of database-driven web applications containing up to hundreds of transactions. Our evaluation shows that MEDIATOR can successfully verify equivalence/refinement between 95% of the web applications it was evaluated on and that involve structural changes to the underlying database schema.

We see this work as a first step towards verifying equivalence of database-driven applications. In future work, we plan to address the limitations discussed in Section 10 to increase the applicability of the tool. First, we plan to develop a richer intermediate representation that can model conditional updates and other interactions between the application and the database. While this extension would likely not necessitate fundamental changes to the proposed verification methodology, it would require performing reasoning over a richer logical theory (e.g., combined theory of Presburger arithmetic, relational algebra, arrays, etc). Second, we plan to automate the translation of web applications written in realistic languages (e.g., Ruby, PHP, etc) into our intermediate language. Third, we are interested in adding counterexample-generation capabilities to MEDIATOR so that it can prove that two applications are definitely not equivalent.

ACKNOWLEDGMENTS

We would like to thank James Cheney, Kostas Ferles, Xinyu Wang, Yu Feng, Jacob Van Geffen, and the anonymous POPL'18 reviewers for their thorough and helpful comments on an earlier version of this paper. This material is based on research sponsored by NSF Awards #1712060 and #1453386.

REFERENCES

- Alfred V. Aho, Yehoshua Sagiv, and Jeffrey D. Ullman. 1979. Equivalences Among Relational Expressions. *SIAM J. Comput.* 8, 2 (1979), 218–246.
- Joseph Albert, Yannis E. Ioannidis, and Raghu Ramakrishnan. 1999. Equivalence of Keyed Relational Schemas by Conjunctive Queries. *J. Comput. Syst. Sci.* 58, 3 (1999), 512–534.
- Scott W Ambler. 2007. Test-driven development of relational databases. *Ieee Software* 24, 3 (2007).
- Scott W Ambler and Pramod J Sadalage. 2006. *Refactoring databases: Evolutionary database design*. Pearson Education.
- Shay Artzi, Adam Kiezun, Julian Dolby, Frank Tip, Danny Dig, Amit Paradkar, and Michael D Ernst. 2008. Finding bugs in dynamic web applications. In *Proc. of ISSTA*. 261–272.
- Paolo Atzeni, Giorgio Ausiello, Carlo Batini, and Marina Moscarini. 1982. Inclusion and Equivalence between Relational Database Schemata. *Theor. Comput. Sci.* 19 (1982), 267–285.
- Thomas Ball, Todd Millstein, and Sriram K. Rajamani. 2005. Polymorphic Predicate Abstraction. *ACM Trans. Program. Lang. Syst.* 27, 2 (2005), 314–343.
- Clark Barrett, Aaron Stump, and Cesare Tinelli. 2010. The satisfiability modulo theories library (SMT-LIB). *www. SMT-LIB.org* 15 (2010), 18–52.
- Gilles Barthe, Juan Manuel Crespo, and César Kunz. 2011. Relational verification using product programs. In *Proc. of FM*. 200–214.

- Gilles Barthe, Juan Manuel Crespo, and César Kunz. 2013. Beyond 2-Safety: Asymmetric Product Programs for Relational Program Verification. In *Proc. of LFCS*. 29–43.
- Catriel Beeri, Alberto O. Mendelzon, Yehoshua Sagiv, and Jeffrey D. Ullman. 1979. Equivalence of Relational Database Schemes. In *Proc. of STOC*. 319–329.
- Michael Benedikt, Timothy Griffin, and Leonid Libkin. 1998. Verifiable Properties of Database Transactions. *Inf. Comput.* 147, 1 (1998), 57–88.
- Nick Benton. 2004. Simple Relational Correctness Proofs for Static Analyses and Program Transformations. In *Proc. of POPL*. 14–25.
- Loredana Caruccio, Giuseppe Polese, and Genoveffa Tortora. 2016. Synchronization of Queries and Views Upon Schema Evolutions: A Survey. *Proc. of TODS* 41, 2 (2016), 9:1–9:41.
- Ashok K. Chandra and Philip M. Merlin. 1977. Optimal Implementation of Conjunctive Queries in Relational Data Bases. In *Proc. of STOC*. 77–90.
- David Chays, Saikat Dan, Phyllis G. Frankl, Filippos I. Vokolos, and Elaine J. Weyuker. 2000. A Framework for Testing Database Applications. In *Proc. of ISSTA*. 147–157.
- Alvin Cheung, Armando Solar-Lezama, and Samuel Madden. 2013. Optimizing database-backed applications with query synthesis. In *Proc. of PLDI*. 3–14.
- Shumo Chu, Chenglong Wang, Konstantin Weitz, and Alvin Cheung. 2017a. Cosette: An Automated Prover for SQL. In *Proc. of CIDR*.
- Shumo Chu, Konstantin Weitz, Alvin Cheung, and Dan Suciu. 2017b. HoTTSQL: Proving query rewrites with univalent SQL semantics. In *Proc. of PLDI*. 510–524.
- Rance Cleaveland and Matthew Hennessy. 1993. Testing equivalence as a bisimulation equivalence. *Formal Aspects of Computing* 5, 1 (1993), 1–20.
- Sara Cohen, Werner Nutt, and Alexander Serebrenik. 1999. Rewriting Aggregate Queries Using Views. In *Proc. of PODS*. 155–166.
- Carlo Curino, Hyun Jin Moon, Alin Deutsch, and Carlo Zaniolo. 2013. Automating the database schema evolution process. *VLDB J.* 22, 1 (2013), 73–98.
- Satyaki Das, David Dill, and Seungjoon Park. 1999. Experience with predicate abstraction. In *Proc. of CAV*. 687–687.
- Leonardo Mendonça de Moura and Nikolaj Bjørner. 2008. Z3: An Efficient SMT Solver. In *Proc. of TACAS*. 337–340.
- Benjamin Delaware, Clément Pit-Claudel, Jason Gross, and Adam Chlipala. 2015. Fiat: Deductive Synthesis of Abstract Data Types in a Proof Assistant. In *Proc. of POPL*. 689–700.
- Yuetang Deng, Phyllis Frankl, and David Chays. 2005. Testing Database Transactions with AGENDA. In *Proc. of ICSE*. 78–87.
- Alin Deutsch, Monica Marcus, Liying Sui, Victor Vianu, and Dayou Zhou. 2005. A Verifier for Interactive, Data-Driven Web Applications. In *Proc. of SIGMOD*. 539–550.
- Alin Deutsch, Liying Sui, and Victor Vianu. 2007. Specification and verification of data-driven Web applications. *J. Comput. Syst. Sci.* 73, 3 (2007), 442–474.
- Michael Emmi, Rupak Majumdar, and Koushik Sen. 2007. Dynamic Test Input Generation for Database Applications. In *Proc. of ISSTA*. 151–162.
- Ronald Fagin, Phokion G Kolaitis, Lucian Popa, and Wang-Chiew Tan. 2011. Schema mapping evolution through composition and inversion. In *Schema matching and mapping*. Springer, 191–222.
- Stéphane Faroult and Pascal L'Hermite. 2008. *Refactoring SQL applications*. O'Reilly Media.
- Dennis Felsing, Sarah Grebing, Vladimir Klebanov, Philipp Rümmer, and Matthias Ulbrich. 2014. Automating regression verification. In *Proc. of ASE*. 349–360.
- Milos Gligoric and Rupak Majumdar. 2013. Model Checking Database Applications. In *Proc. of TACAS*. 549–564.
- Benny Godlin and Ofer Strichman. 2009. Regression verification. In *Proc. of DAC*. 466–471.
- Benny Godlin and Ofer Strichman. 2013. Regression verification: proving the equivalence of similar programs. *Software Testing, Verification and Reliability* 23, 3 (2013), 241–258.
- Todd J. Green. 2009. Containment of conjunctive queries on annotated relations. In *Proc. of ICDT*. 296–309.
- Chris Hawblitzel, Ming Kawaguchi, Shuvendu K. Lahiri, and Henrique Rebêlo. 2013. Towards Modularly Comparing Programs Using Automated Theorem Provers. In *Proc. of CADE*. 282–299.
- Richard Hull. 1986. Relative Information Capacity of Simple Relational Database Schemata. *SIAM J. Comput.* 15, 3 (1986), 856–886.
- Vasileios Koutavas and Mitchell Wand. 2006a. Bisimulations for Untyped Imperative Objects. In *Proc. of ESOP*. 146–161.
- Vasileios Koutavas and Mitchell Wand. 2006b. Small bisimulations for reasoning about higher-order imperative programs. In *Proc. of POPL*. 141–152.
- Sudipta Kundu, Zachary Tatlock, and Sorin Lerner. 2009. Proving optimizations correct using parameterized program equivalence. In *Proc. of PLDI*. 327–337.

- Shuvendu K Lahiri, Chris Hawblitzel, Ming Kawaguchi, and Henrique Rebêlo. 2012. Symdiff: A language-agnostic semantic diff tool for imperative programs. In *Proc. of CAV*. 712–717.
- Shuvendu K Lahiri, Kenneth L McMillan, Rahul Sharma, and Chris Hawblitzel. 2013. Differential assertion checking. In *Proc. of ESEC/FSE*. 345–355.
- Shuvendu K Lahiri and Shaz Qadeer. 2009. Complexity and algorithms for monomial and clausal predicate abstraction. In *Proc. of CADE*. 214–229.
- Renée J. Miller, Yannis E. Ioannidis, and Raghu Ramakrishnan. 1993. The Use of Information Capacity in Schema Integration and Translation. In *Proc. of VLDB*. 120–133.
- Joseph P. Near and Daniel Jackson. 2012. Rubicon: bounded verification of web applications. In *Proc. of FSE*. 60.
- George C. Necula. 2000. Translation validation for an optimizing compiler. In *Proc. of PLDI*. 83–94.
- Oracle. 2005. Oracle Schema Optimization Guide. https://docs.oracle.com/cd/B14099_19/web.1012/b15901/tuning007.htm. (2005).
- Suzette Person, Matthew B. Dwyer, Sebastian G. Elbaum, and Corina S. Pasareanu. 2008. Differential symbolic execution. In *Proc. of FSE*. 226–237.
- Amir Pnueli, Michael Siegel, and Eli Singerman. 1998. Translation Validation. In *Proc. of TACAS*. 151–166.
- Erhard Rahm and Philip A. Bernstein. 2006. An online bibliography on schema evolution. *SIGMOD Record* 35, 4 (2006), 30–31.
- Martin Rinard. 1999. Credible Compilation. In *MIT TechReport*. MIT-LCS-TR-776.
- Arnon Rosenthal and David S. Reiner. 1994. Tools and Transformations - Rigorous and Otherwise - for Practical Database Design. *ACM Trans. Database Syst.* 19, 2 (1994), 167–211.
- Yehoshua Sagiv and Mihalis Yannakakis. 1980. Equivalences Among Relational Expressions with the Union and Difference Operators. *J. ACM* 27, 4 (1980), 633–655.
- Davide Sangiorgi, Naoki Kobayashi, and Eijiro Sumii. 2007. Environmental Bisimulations for Higher-Order Languages. In *Proc. of LICS*. 293–302.
- Davide Sangiorgi, Naoki Kobayashi, and Eijiro Sumii. 2011. Environmental bisimulations for higher-order languages. *ACM Trans. Program. Lang. Syst.* 33, 1 (2011), 5:1–5:69.
- Marcelo Sousa and Isil Dillig. 2016. Cartesian Hoare logic for verifying k-safety properties. In *Proc. of PLDI*. 57–69.
- Michael Stepp, Ross Tate, and Sorin Lerner. 2011. Equality-based translation validator for LLVM. In *Proc. of CAV*. 737–742.
- Eijiro Sumii and Benjamin C. Pierce. 2004. A bisimulation for dynamic sealing. In *Proc. of POPL*. 161–172.
- Eijiro Sumii and Benjamin C. Pierce. 2005. A bisimulation for type abstraction and recursion. In *Proc. of POPL*. 63–74.
- Boris A Trakhtenbrot. 1950. Impossibility of an algorithm for the decision problem in finite classes. *Doklady Akademii Nauk SSSR* 70 (1950), 569–572.
- Victor Vianu. 2009. Automatic verification of database-driven systems: a new frontier. In *Proc. of ICDT*. 1–13.
- Joost Visser. 2008. Coupled Transformation of Schemas, Documents, Queries, and Constraints. *Electr. Notes Theor. Comput. Sci.* 200, 3 (2008), 3–23.
- Yuepeng Wang, Isil Dillig, Shuvendu K. Lahiri, and William R. Cook. 2017. Verifying Equivalence of Database-Driven Applications. <http://arxiv.org/abs/1710.07660>. (2017). arXiv:1710.07660
- Gary Wassermann, Dachuan Yu, Ajay Chander, Dinakar Dhurjati, Hiroshi Inamura, and Zhendong Su. 2008. Dynamic test input generation for web applications. In *Proc. of ISSSTA*. 249–260.
- Wikimedia. 2017. Schema changes. https://wikitech.wikimedia.org/wiki/Schema_changes. (2017).
- Tim Wood, Sophia Drossopoulou, Shuvendu K. Lahiri, and Susan Eisenbach. 2017. Modular Verification of Procedure Equivalence in the Presence of Memory Allocation. In *Proc. of ESOP*. 937–963.
- Hongseok Yang. 2007. Relational separation logic. *Theoretical Computer Science* 375, 1-3 (2007), 308–334.
- Anna Zaks and Amir Pnueli. 2008. Covac: Compiler validation by program analysis of the cross-product. In *Proc. of FM*. 35–51.
- Lenore D. Zuck, Amir Pnueli, and Benjamin Goldberg. 2003. VOC: A Methodology for the Translation Validation of Optimizing Compilers. *J. UCS* 9, 3 (2003), 223–247.