Object Normal Form, Fourth Normal Form and their Application to Database Security

Sebastian Link[0000−0002−1816−2863]

School of Computer Science
The University of Auckland, Auckland 1010, New Zealand
s.link@auckland.ac.nz

Abstract. An important question in database schema design concerns the effort required to maintain data consistency under updates. Similarly, an important question in database security concerns the effort required to maintain data confidentiality under inference attacks. Previous work has addressed these questions for the popular class of functional dependencies. In this paper, we will extend solutions to the more expressive class of multivalued dependencies. In particular, we will show that schemata in Fourth Normal Form with a unique minimal key require very little effort to maintain data consistency, and can guarantee confidentiality under inference attacks by access control only.

Keywords: Access Control, Boyce-Codd Normal Form, Confidentiality, Fourth Normal Form, Functional dependency, Independence, Inference Control, Key, Multivalued Dependency, Object Normal Form, Uniqueness

1 Introduction

The design of relational databases is a classical topic in database research. The overarching goal is to organize data in tables on which future update and query operations can be performed effectively and efficiently. In particular, schema normalization is concerned with minimizing the effort required to maintain consistency while processing update operations. Data redundancy, as caused by integrity constraints such as functional, multivalued and join dependencies (FDs, MVDs, JDs), slows down updates since redundant values need to be updated wherever they occur in the relation. Hence, during normalization tables are decomposed by transforming redundancy-causing data dependencies into keys that prevent data redundancy. For example, schemata in Boyce-Codd Normal Form (BCNF) only exhibit relations in which no data redundancy caused by FDs can ever occur [12, 33]. Fourth and Fifth Normal Form (4NF, 5NF) achieve the same but for MVDs and JDs, respectively [16,32]. Since JDs extend MVDs and MVDs extend FDs, schemata in 5NF are also in 4NF, and schemata in 4NF are also in BCNF, but not vice versa [17]. There is evidence that schemata in practice are often in BCNF, but not in higher normal forms, such as 4NF [37].

Biskup showed [2] that BCNF is equivalent to schemata in weak object normal form, where the left-hand side of any left-reduced FD of the schema forms a so-called weak object. The latter are attribute sets that are unique and weakly independent. That is, any relation has unique projections to weak objects, and inserting new combinations of values on attributes of the weak object ensures that these values can be completed with *some* values on the remaining attributes such that the updated relation will satisfy all constraints. Biskup also introduced objects as attribute subsets that are unique and strongly independent. Here, inserting new combinations of values on attributes of the object ensures that these values can be completed with *any* values on the remaining attributes such that the updated relation will satisfy all constraints. Object Normal Form (ONF) means that the left-hand sides of any left-reduced FD of the schema are objects, and Biskup showed that schemata are in ONF if and only if they are in BCNF with a unique minimal key [2] (there is only one key that is minimal with respect to set inclusion of attribute subsets). BCNF ensures that one will never need to worry about any non-key values when updating records in the database. However, one still needs to worry about the uniqueness of value combinations on all minimal keys. In ONF, one only needs to worry about the uniqueness of value combinations on *the* minimal key.

The theory of object normal forms has been limited to functional, inclusion and exclusion dependencies so far [2, 7]. As the following example illustrates, it would be interesting to extend the theory to more expressive dependencies, in particular tuple-generating dependencies such as multivalued dependencies.

Example 1. Consider the simple example where relation schema MEET collects information about project meetings, where members of projects meet on a date. An example relation over MEET is given as follows.

Project	*Date*	*Member*
Green Goddess	19/12/2021	Clyde
Green Goddess	19/12/2021	Bonnie

Since all project members should be present during all meetings of the same project, we specify the MVD $P \twoheadrightarrow M$. In addition, team members can only attend one project meeting on any given day, and therefore we have the FD $DM \rightarrow P$. It follows that $R = PDM$ with $P \twoheadrightarrow M$ and $DM \rightarrow P$ has the unique minimal key DM. This means that the schema is in BCNF and has only one minimal key. That is, the schema is in ONF for the given FDs. However, the schema is not in 4NF since $P \twoheadrightarrow M$ is a non-trivial MVD where P is not a key.

In fact, the left-hand side attribute P of the MVD $P \twoheadrightarrow M$ does not satisfy the uniqueness property: the relation r satisfies the constraints, but the two different tuples of r have the same value on P. While DM is strongly independent with respect to the FD, it is not strongly independent with respect to the MVD. Indeed, while tuple $t = (P\text{:Green Goddess}, D\text{:02/12/2021}, M\text{:Bonnie})$ has a projection on DM that does not occur in r, $r \cup \{t\}$ violates the MVD $P \twoheadrightarrow M$. □

Hence, our first objective is to generalize the concept of Object Normal Form from FDs to MVDs. We will show that 4NF is equivalent to Weak Object Normal Form, and Object Normal Form is equivalent to 4NF with a unique minimal key.

As a second contribution, we will show an application of our new results in database security. An important goal of security is *confidentiality*. In general, enforcing confidentiality requires costly dynamic inference control. In practice, security administrators often only use efficient access control based on static access rights. This, however, lays the burden on administrators to properly set access rights such that access to data must never allow users to infer information that is meant to be confidential. We illustrate the intrinsic difficulty of inference control on our example from before.

Example 2. Consider the relation r from Example 1. Suppose a user wants to keep the following combination of specific values confidential: (*D:*19/12/2021,*M:*Bonnie). That is, an answer to a query such as $(\exists P)\textsc{Meet}(P, \text{19/12/2021}, \text{Bonnie})$ must be refused since it would reveal the confidential combination of values.. This appears to be no problem at first glance. However, given the constraints from Example 1, a user may bypass access control by issuing the queries $(\exists M)\textsc{Meet}(\text{Green Goddess}, \text{19/12/2021}, M)$ and $(\exists D)\textsc{Meet}(\text{Green Goddess}, D, \text{Bonnie})$. None of the two queries reveals the confidential combination of values. However, applying the MVD $P \twoheadrightarrow M$ to the answers of the two queries results in the inferred tuple (*P:*Green Goddess, *D:*19/12/2021, *M:* Bonnie), which reveals the confidential combination of values. Such an inference is not possible with the given FD only. □

As we will show, combining schemata in Object Normal Form with a restriction of potential secrets to attribute sets that are so-called facts, ensures that costly inference control can be reduced to efficient access control while retaining confidentiality. Since facts are based on the constraints of the schema, an extension to MVDs ensures that administrators can declare a richer set of potential secrets as well.

Main Contributions. (1) We generalize object normal forms from the single class of FDs to the combined class of FDs and MVDs. For such constraints sets, we show that i) schemata are in weak ONF if and only if they are in 4NF, and ii) schemata are in ONF if and only if they are in 4NF and exhibit a unique minimal key. (2) For potential secrets that are defined over facts, we show that confidentiality can be guaranteed efficiently by access control whenever the underlying schemata are in ONF. Hence, we do not only provide insight on the effort required to retain data consistency under updates, but also on the effort required to guarantee confidentiality under inference attacks. Next we illustrate how the problems from Examples 1 and 2 are resolved by schemata in ONF.

Example 3. Consider the following relation over the schema MEET from Example 1.

Project	*Date*	*Member*
Green Goddess	19/12/2021	Clyde
Green Goddess	02/12/2021	Bonnie
Green Goddess	19/12/2021	Bonnie
Green Goddess	02/12/2021	Clyde

We may decompose MEET into the three schemata $(PD, \{PD\})$, $(PM, \{PM\})$ and $(DM, \{DM\})$ without loss of information. This is done following a 4NF decomposition with respect to the MVD $P \twoheadrightarrow M$ resulting in $(PD, \{PD\})$ and $(PM, \{PM\})$. The final schema $(DM, \{DM\})$ is added to preserve the minimal key *DM* and therefore the FD $DM \rightarrow P$. Indeed, each of the schemata is in Object Normal Form with respect to the input set of FDs and MVDs. Using the decomposition, the relation above is decomposed into the following relations as well.

Project	*Date*
Green Goddess	19/12/2021
Green Goddess	02/12/2021

Project	*Member*
Green Goddess	Clyde
Green Goddess	Bonnie

Date	*Member*
19/12/2021	Clyde
02/12/2021	Bonnie
19/12/2021	Bonnie
02/12/2021	Clyde

Potential secrets over these schemata are restricted to entire tuples in each of the three relations. Access control is sufficient to guarantee confidentiality under inferences. □

Outline. We discuss previous work in Section 2. Preliminary definitions are given in Section 3. Weak objects, weak ONF, and its equivalence to 4NF are discussed in Sections 4 and 5, respectively. Objects, ONF, and its equivalence to 4NF with a unique minimal key are established in Sections 6 and 7, respectively. The application to database security is detailed in Section 8. Section 9 concludes and discusses future work.

2 Related Work

Logical schema design for relational databases has been studied in great depth. Indeed, normal forms and database schema normalization are classical topics of introductory database textbooks [15]. Third [6], Boyce-Codd [12], and Fourth Normal Form [16] are well-known, and their achievements and limitations have been surveyed [3]. Biskup introduced Object Normal Forms for functional dependencies [2], and generalized them later to inclusion and exclusion dependencies [7]. Hence, the first contribution of the current paper addresses a shortcoming of relational database theory by extending Object Normal Form to MVDs. This is important since MVDs provide a sufficient and necessary condition for relations to be decomposable into two of its projections without loss of information, providing a strong basis for database normalization [16, 23]. Many database schemata in practice that are in BCNF actually violate 4NF [37].

Likewise, the guarantee of confidentiality is a fundamental topic in information security [18, 30]. Here, inference control is necessary to guarantee confidentiality but costly to implement, while access control is easy to implement but cannot guarantee confidentiality under inference attacks. Many textbooks on computer security [19] explain the main concepts of access control. Farkas and Jajodia give a general overview of the inference problem in databases [18]. Lunt et al. [26] propose a formal security model for mandatory access control in the context of relational databases. Subsequent work studies how to consistently declare the classifications of structured objects that are bound to constraints, in order to prevent unwanted inferences, e.g., see Olivier and von Solms [29], Cuppens and Gabillon [13], Dawson et al. [14], Wang and Liu [35], and Brodsky et al. [10]. Controlled query evaluation is rooted in the papers of Sicherman et al. [30] and Bonatti et al. [9]. Biskup and Bonatti propose a unified framework and analyze controlled query evaluation for closed queries in complete databases [4, 5]. Biskup et al. [8] have shown that inference control reduces to access control for schemata in Object Normal Form when potential secrets are specified on attribute sets that form left-hand side reduced functional dependencies with a singleton right-hand side attribute. Our second contribution extends these results to Object Normal Forms for multivalued dependencies, and even more expressive classes of constraints.

Schema design is essential to every data model, which means that the work on schema design in the context of relational databases influences schema design on any extensions of the relational model. It is therefore no surprise that schema design has been deeply investigated in other data models as well, including conceptual models [11], SQL data models [20, 22], nested data models [28, 34], object-relational models [31], temporal models [21], Web models such as XML and JSON [1,27], and models for uncertain data [24]. No matter which data model is used, the design of a database schema will always determine how well updates and queries can be processed on database instances [25]. In view of data quality, normalization reduces data redundancy which is the source of data inconsistency and update inefficiency. Hence, eliminating data redundancy can lead to better data quality and make data-driven decision making more effective. Recent work has also started to extend classical database normalization to design for data quality [36].

3 Preliminaries

In the real world we ascribe two properties to an object: 1) It is unique within its domain, and 2) It can emerge and exist independently of the current environment. We will formalize these properties for relational databases with different classes of data dependencies, including functional and multivalued dependencies.

A relation schema R is a finite set of attributes that denote the column names of a table. Every attribute $A \in R$ is associate with a domain $dom(A)$ that contains the set of possible values that may occur in column A. A tuple t over R assigns to every attribute $A \in r$ some value $t(A) \in dom(A)$. A relation r over R is a finite set of tuples over R. For an attribute subset $X \subseteq R$, we use $t(X)$ to denote the projection of tuple t over R onto the attribute set X.

A relation schema R typically comes with a set Σ of data dependencies from a given class, such as the class of functional dependencies or multivalued dependencies. A functional dependency (FD) is a statement $X \to Y$ with $X, Y \subseteq R$, and a multivalued dependency (MVD) is a statement $X \twoheadrightarrow Y$ with $X, Y \subseteq R$. The FD $X \to Y$ is satisfied by a relation r over R whenever every pair of tuples with matching values on all the attributes of X have also matching values on all the attributes of Y. The MVD $X \twoheadrightarrow Y$ is satisfied by a relation r over R whenever for every pair of tuples in r that has matching values on all the attributes of X there is some tuple in r that has matching values with the first tuple on all the attributes in XY and matching values with the second tuple on all the attributes in $R - XY$, respectively. For a given class of data dependencies, and for a given set Σ of data dependencies from that class, we denote by Σ^+ the set of dependencies from that class implied by Σ, that is, the set of all dependencies from that class that are satisfied by every relation that already satisfies all the dependencies in Σ. For example, every FD $X \to Y$ implies the MVD $X \twoheadrightarrow Y$, but not vice versa. Trivial dependencies are those satisfied by every relation. For instance, an FD $X \to Y$ is trivial iff $Y \subseteq X$, and an MVD $X \twoheadrightarrow Y$ is trivial iff $Y \subseteq X$ or $XY = R$.

We say that (R, Σ) is in *Boyce-Codd Normal Form* (BCNF) if and only if for every non-trivial FD $X \to Y \in \Sigma^+$, $X \to R \in \Sigma^+$ holds. That is, every left-hand side X of a non-trivial FD functionally determines all the attributes of the schema. In other words,

X is a key. That means no relation that satisfies a key X can have different tuples with matching values on all the attributes of X.

Similarly, we say that (R, Σ) is in *Fourth Normal Form* (4NF) if and only if for every non-trivial MVD $X \twoheadrightarrow Y \in \Sigma^+$, $X \to R \in \Sigma^+$ holds.

Example 1 shows a relation that satisfies the set $\Sigma = \{P \twoheadrightarrow M, DM \to P\}$. While (MEET, Σ) is in BCNF, it is not in 4NF. For example, the given relation violates the FD $P \to D$, which means that P is not a key. Consequently, for the MVD $P \twoheadrightarrow M \in \Sigma^+$, $P \to R \notin \Sigma^+$.

4 Weak Objects

The aim of the next two sections is to recall the concepts of weak objects and weak ONF, and to show that schemata are in weak ONF if and only if they are in 4NF.

We recall the definition of weak objects. While Biskup [2] only considered FDs, we assume Σ may contain other types of data dependencies as well, such as MVDs.

Definition 1 (weak object). *Let (R, Σ) denote a relation schema R together with a set Σ of data dependencies over R. An attribute subset X of R is said to be a* weak object *if and only if the following hold:*

- *(Uniqueness) For all relations r over R that satisfy Σ, for all $t \in r$, $t(X)$ is unique. That is, there is no $t' \in r - \{t\}$ such that $t(X) = t'(X)$.*
- *(Weak independence) For all R-relations r that satisfy Σ, for all $\mu \in dom(X)$ such that $\mu \notin r(X)$, there is some $\nu \in dom(R - X)$ such that $r \cup \{\mu\nu\}$ satisfies Σ.* □

In Example 1, the attribute set DM is a weak object while P is not. The first property shows that keys satisfying weak independence are actually minimal keys.

Proposition 1. *Let X be a key over (R, Σ). If weak independence holds for X, then $Y \to R \notin \Sigma^+$ for all $Y \subset X$.*

Proof. Let t denote a tuple over R. The relation $r = \{t\}$ satisfies Σ. Now, let t' be a tuple over R such that $t'(Y) = t(Y)$ and $t'(X - Y) \neq t(X - Y)$. It follows that $\mu := t'(X) \notin r(X)$. Due to weak independence there is some $\nu \in dom(R - X)$ such that $r' = r \cup \{\mu\nu\}$ satisfies Σ. It follows that r' does not satisfy $Y \to R$. □

The second property shows that minimal keys of schemata in 4NF satisfy weak independence.

Proposition 2. *Let X be a key over (R, Σ). If X is a minimal key and (R, Σ) is in Fourth Normal Form, then the weak independence property holds for X.*

Proof. Let r denote a relation over R that satisfies Σ. Let $\mu \in dom(X)$ such that $\mu \notin r(X)$. As there are infinitely many constants, we can find a tuple $\nu \in dom(R-X)$ such that $\nu(A) \notin r(A)$ for all $A \in R-X$, that is, ν is composed of values not occurring in the projection of r onto $R - X$. We claim that $r \cup \{\mu_i\nu\}$ is a relation over R that satisfies Σ. For consider $Y \twoheadrightarrow Z \in \Sigma^+$ with $Z \not\subseteq Y$ and $Z \not\subseteq R - Y$ and assume $t' := \mu\nu$ and some $t \in r$ violates $Y \twoheadrightarrow Z$. In particular, $t'(Y) = t(Y)$ must hold and, by the construction of ν, $Y \subset X$ (Y is a proper subset of X). Due to the minimality of the key X for R, we have $Y \rightarrow R \notin \Sigma^+$. This, however, violates the hypothesis that (R, Σ) is in 4NF. Consequently, our assumption that $r \cup \{\mu\nu\}$ violates $Y \twoheadrightarrow Z$ must have been wrong. This shows that the weak independence property for X. □

5 Weak Object Normal Form

We will now define when a schema is in weak ONF, and prove that weak ONF and 4NF are equivalent. Here we extend the previous definition [2] from FDs to MVDs.

Definition 2. *Let (R, Σ) be a relation schema and*

$$\begin{aligned} LHS = \{X \subseteq R \mid \exists Y \subseteq R (Y \nsubseteq X \wedge XY \neq R \wedge (X \twoheadrightarrow Y \in \Sigma^+)) \\ \textit{with minimal } X \textit{ or minimal key } X \textit{ for } (R, \Sigma)\}. \end{aligned}$$

The relation schema (R, Σ) is in Weak Object Normal Form *if and only if for every left-hand side $X \in$* LHS*, X is a weak object over (R, Σ).* □

In Example 1, $P \in LHS$ but P is not a weak object. Consequently, (Meet, Σ) is not in weak ONF. Evidently, it is also not in 4NF.

Theorem 1. *Let Σ denote a set of FDs over relation schema R. Then (R, Σ) is in Weak Object Normal Form if and only if (R, Σ) is in Fourth Normal Form.*

Proof. (If). Let $X \in LHS$. Since (R, Σ) is in 4NF, X must satisfy the uniqueness property. Proposition 2 implies the weak independence property. Consequently, X is a weak object, and (R, Σ) is in Weak Object Normal Form.

(Only if). Let $Y \twoheadrightarrow V \in \Sigma^+$ such that $V \nsubseteq Y$ and $YV \neq R$. Consider any minimal $Z \subseteq Y$ such that $Z \twoheadrightarrow U \in \Sigma^+$ where $U \nsubseteq V$ and $ZU \neq R$. Then $Z \in LHS$, and thus, by assumption, Z is a weak object. It follows that $Z \rightarrow R \in \Sigma^+$, and, in particular, $Y \rightarrow R \in \Sigma^+$. Hence, (R, Σ) is in Fourth Normal Form. □

6 Objects

We will now recall the concepts of objects and ONF, and prove that ONF is equivalent to 4NF with a unique minimal key.

Definition 3 (object). *Let (R, Σ) denote a relation schema R together with a set Σ of FDs and MVDs over R. An attribute subset X of R is an object iff the following hold:*

- *(Uniqueness) For all relations r over R that satisfy Σ, for all $t \in r$, $t(X)$ is unique. That is, there is no $t' \in r - \{t\}$ such that $t(X) = t'(X)$.*
- *(Independence) For all relations r over R that satisfy Σ, for all $\mu \in dom(X)$ such that $\mu \notin r(X)$, for all $\nu \in dom(R - X)$, $r \cup \{\mu\nu\}$ satisfies Σ.* □

Our running example illustrates the definition of independence.

Example 4. Let $\Sigma = \{P \twoheadrightarrow D, DM \rightarrow P\}$ be a set of FDs and MVDs over $R = PDM$. In particular, (R,Σ) is not in 4NF since the only minimal key is DM. For instance, consider the following relation $r = \{t, t'\}$.

Project	*Date*	*Member*
Green Goddess	19/12/2021	Clyde
Green Goddess	19/12/2021	Bonnie

P is not an object since the relation satisfies Σ but violates uniqueness on P since $t(P) = t'(P)$. Indeed, the attribute subset DM is also not an object since it does not satisfy independence as we show now. Let $t'' :=$ (Green Goddess, 02/12/2021, Bonnie) where (02/12/2021, Bonnie) $\notin r[DM]$. Then $r \cup \{t''\}$ does not satisfy $P \twoheadrightarrow D$. Hence, DM is not an object. □

Example 5. Let $\Sigma = \{P \twoheadrightarrow D, PD \rightarrow L, PL \rightarrow D\}$ be a set of FDs and MVDs over MEET-LEAD $= PDL$ where L denotes the lead for the project on the day. In particular, (R,Σ) is in 4NF with the unique minimal key P. For instance, consider the following relation $r = \{t\}$.

Project	*Date*	*Lead*
Green Goddess	19/12/2021	Bonnie

Let $t' :=$ (Passion Pop, 19/12/2021, Bonnie) where *Passion Pop* $\notin r(P)$. Then $r \cup \{t'\}$ satisfies Σ. In fact, P is an object. □

The next result shows that a key satisfies the independence property if and only if the underlying schema is in 4NF and the key is the only minimal key.

Proposition 3. *Let X denote some key over (R,Σ). Then X satisfies the independence property if and only if all of the following hold:*

1. *X is a minimal key for (R,Σ).*
2. *If Z denotes some minimal key for (R,Σ), then $Z = X$.*
3. *(R,Σ) is in Fourth Normal Form.*

Proof. We show first that the three conditions are sufficient for the independence property for X to hold.

Let r denote a relation over R that satisfies Σ, let $\mu \in dom(X)$ such that $\mu \notin r(X)$. Let $\nu \in dom(R - X)$. We claim that $r' = r \cup \{t'\}$ satisfies Σ for $t' = \mu\nu$. For consider $Y \twoheadrightarrow V \in \Sigma^+$ such that $V \not\subseteq Y$ and $YV \neq R$, and assume that r' violates $Y \twoheadrightarrow V$. In particular, $t(Y) = t'(Y)$ for some $t \in r$. The 4NF condition (3) implies that $Y \rightarrow R \in \Sigma^+$. Hence, Y is a key for (R,Σ). If Y is even a minimal key, then by (1) and (2), $Y = X$. If Y is not a minimal key, then Y is a superset of some minimal key, and by (1) and (2), $X \subseteq Y$. Consequently, $t'(X) = t(X) \in r[X]$, which is a contradiction. That means our assumption that r' violates $Y \twoheadrightarrow V$ must have been wrong, which means that r' satisfies Σ. This proves the independence property for X.

We show now that the independence property for X is sufficient for (1), (2), and (3) to hold.

Firstly, since the independence property for X implies the weak independence property for X, Proposition 1 implies (1).

We are now going to show that (2) holds as well. Assume there is some other minimal key Z, different from X. We then know that $X - Z \neq \emptyset$ and $Z - X \neq \emptyset$ holds. Let $r := \{t\}$ with any tuple t over R. It follows that r satisfies Σ. We now define a tuple t' over R such that $t'(X - Z) \neq t(X - Z)$ and $t'(Z - X) = t(Z - X)$, and $t'(R - X) = t(R - X)$. In particular, we have $t'(X) \notin r(X)$. Due to the independence property for X it follows that $r' := r \cup \{t'\}$ satisfies Σ. However, it follows that $t'(Z) = t(Z)$, which means that r does not satisfy $Z \rightarrow R \in \Sigma^+$, a contradiction to the assumption that Z is another minimal key. Consequently, (2) must hold.

It remains to show (3). For consider $Y \twoheadrightarrow V \in \Sigma^+$ such that $V \nsubseteq Y$ and $YV \neq R$. We distinguish between two cases.

Case 1: $X \subseteq Y$. Since X is a key for (R, Σ) we have $Y \rightarrow R \in \Sigma^+$.

Case 2: There is some $A \in X - Y$. We will show that this case cannot occur. Indeed, let $r := \{t\}$ be some relation over R. Then r satisfies Σ. Define a tuple t' over R such that, $t'(Y \cap X) := t(Y \cap X)$, $t'(Y - X) := t(Y - X)$, for all $B \in X - Y$, $t'(B) \neq t(B)$, and for all $B \in R - Y$, $t'(B) \neq t(B)$. It follows that $t'(X) \notin r(X)$. By the independence property for X we conclude that $r' := r \cup \{t'\}$ satisfies Σ. However, by construction of t', we have $t'(Y) = t(Y)$ and, since $V - Y \neq \emptyset$ and $R - VY \neq \emptyset$, r' does not satisfy $Y \twoheadrightarrow V$ since there is no $t'' \in r$ such that $t''(YV) = t(YV)$ and $t''(R - YV) = t'(R - YV)$. This is a contradiction, and Case 2 cannot occur. □

7 Object Normal Form

We can now recall the definition of ONF and show that it is equivalent to 4NF with a unique minimal key.

Definition 4. *The relation schema (R, Σ) is in* object normal form *if and only if for every left-hand side $X \in$ LHS, X is an object over (R, Σ).* □

Theorem 2. *Let Σ denote a set of FDs and MVDs over relation schema R. (R, Σ) is in Object Normal Form if and only if (R, Σ) is in Fourth Normal Form and there is one minimal key.*

Proof. (If). Let $X \in LHS$. Since (R, Σ) is in 4NF, X must satisfy the uniqueness property. Due to Proposition 3, X satisfies the independence property and is, therefore, an object. Consequently, (R, Σ) is in Object Normal Form.

(Only if). Let $X \twoheadrightarrow Y \in \Sigma^+$ such that $Y \nsubseteq X$ and $XY \neq R$. Since (R, Σ) is in object normal form, it follows that there is some minimal key Z for (R, Σ) such that $Z \subseteq X$ and Z is an object. Proposition 3 then yields the assertion that (R, Σ) is in Fourth Normal Form and Z is the only minimal key. □

Our running example illustrates the difference between 4NF and ONF.

Example 6. Let $\Sigma = \{P \twoheadrightarrow D, D \twoheadrightarrow P, PD \rightarrow L, PL \rightarrow D, DL \rightarrow P\}$ be a set of FDs and MVDs over $R = PDL$. In particular, (R, Σ) is in 4NF with minimal key P and minimal key D. For instance, consider the following relation $r = \{t\}$.

Project	*Date*	*Lead*
Green Goddess	19/12/2021	Bonnie

Let $t' := (\mathit{Passion\ Pop}, \mathit{19/12/2021}, \mathit{Clyde})$ where *Passion Pop* $\notin r[P]$. Then $r \cup \{t'\}$ violates $D \twoheadrightarrow P$. Hence, P is not an object. However, for $\bar{t}' := (\mathit{Passion\ Pop}, \mathit{02/12/2021}, \mathit{Clyde})$ where *Passion Pop* $\notin r[P]$, $r \cup \{\bar{t}'\}$ does satisfy Σ. Indeed, P is a weak object.

Similarly, let $t'' := (\mathit{Green\ Goddess}, \mathit{02/12/2021}, \mathit{Clyde})$ where *02/12/2021* $\notin r[D]$. Then $r \cup \{t''\}$ violates $P \twoheadrightarrow D$. Hence, D is not an object. However, for $\bar{t}' := (\mathit{Passion\ Pop}, \mathit{02/12/2021}, \mathit{Clyde})$ where *02/12/2021* $\notin r[D]$, $r \cup \{\bar{t}'\}$ does satisfy Σ. Indeed, D is a weak object.

We conclude that the schema is in weak ONF but not in ONF. □

8 Application to Database Security

We report on a showcase of our new results in the area of database security, in particular the ability to reduce inference control to access control for schemata that are in object normal form.

8.1 Motivating Example

An important goal of security in information systems is *confidentiality*. In general, enforcing confidentiality requires costly dynamic inference control. In practice, security administrators often only use efficient access control based on static access rights. This, however, lays the burden on the administrator to properly set access rights so that permitted data accesses may never allow users to infer information to be kept secret.

Illustrating the difficulties arising in this context, imagine an application that associates the fee for some insurance policy and the name of some beneficiary.

Policy	*Name*	*Fee*
JF759	James Muller	125.60

Now assume the beneficiary wants to keep the combination of his name with the fee confidential. Consequently, a security administrator will need to protect the combination *(James Muller, 125.60)* by specifying access rights appropriately. Without further consideration, protecting only the critical information itself suffices to preserve unwanted disclosures. However, suppose that the data is governed by the functional dependencies *Policy* $\rightarrow$ *Name*, *Policy* $\rightarrow$ *Fee*, *Name* $\rightarrow$ *Policy* and *Name* $\rightarrow$ *Fee*. In this case, the protection can be bypassed by querying the combinations *(P:JFY759,F:125.60)* and *(P:JFY759,N:James Muller)*, both unprotected, and joining the results to associate *James Muller* with the fee *125.60*.

In response, Biskup et al. [8] identified a common situation guaranteed to meet, for any discretionary assignment of access rights, the goals of inference control. Essentially, this situation is given by a relational database schema in Object Normal Form with respect to a given set of FDs, and restricting the confidentiality policy to the protection of certain parts of a tuple, called facts. Hence, when a security administrator can declare access rights in a content-dependent way, the system can guarantee confidentiality and easily perform inspections by a simple lookup of access rights.

8.2 Reducing Inference to Access Control on Schemata in ONF

In the context of [8], queries are expressed in a fragment of relational calculus. Let *Var* denote a set of variables. The query language $\mathcal{L}_Q$ is the set of all closed formulae of the form $(\exists X_1)\cdots(\exists X_l)R(v_1,\ldots,v_n)$ with $0 \leq l \leq n$, $X_i \in \mathit{Var}$, $v_i \in \mathit{Const} \cup \mathit{Var}$, $\{X_1,\ldots,X_n\} \subseteq \{v_1,\ldots,v_n\}$, and $v_i, v_j \in \mathit{Var}$ and $i \neq j$. Let $\Phi \in \mathcal{L}_Q$ be a query and r a relation over R. The *ordinary evaluation* of closed queries is defined by

$$\mathit{eval}(\Phi)(r) := \texttt{if } \models_r \Phi \texttt{ then } \Phi \texttt{ else } \neg\Phi.$$

Biskup and Bonatti [4] developed *controlled query evaluation* (CQE). We briefly outline their approach in terms of our framework. A *potential secret* Ψ is a formula of a given language. If $\not\models_r \Psi$ for a relation r, the database user may learn that Ψ is false in the relation; however, if $\models_r \Psi$, it needs to be kept secret that Ψ is actually true. The set $\mathit{pot\_sec} \subseteq \mathcal{L}_Q$ denotes a confidentiality policy being *known* to the database user, and $\mathit{log}_0 \subseteq \mathcal{L}_Q \cup \Sigma$ is the a priori user knowledge with $\models_r \mathit{log}_0$ and $\mathit{log}_0 \not\models \Psi$ for every $\Psi \in \mathit{pot\_sec}$, that is, log_0 is actually true and none of the potential secrets is known to the user in advance. Finally, let a query sequence be given by $\mathcal{Q} = \langle \Phi_1, \Phi_2, \ldots \rangle$ with $\phi_i \in \mathcal{L}_Q$. The CQE for known potential secrets enforced by refusal (that is, the answer is refused by returning the constant *mum*) is defined by

$$\mathit{cqe}(Q, \mathit{log}_0)(r, \mathit{pot\_sec}) := \langle (\mathit{ans}_1, \mathit{log}_1), (\mathit{ans}_2, \mathit{log}_2), \ldots \rangle.$$

The values of the returned answers ans_i and the representation of the current user knowledge log_i are determined subject to a *censor function* [**?**, 5]:

$$\begin{aligned}
\mathit{censor}(\mathit{pot\_sec}, \mathit{log}, \Phi) &:= (\exists\Psi)[\Psi \in \mathit{pot\_sec} \wedge \\
&\quad ((\mathit{log} \cup \{\Phi\} \models \Psi) \vee (\mathit{log} \cup \{\neg\Phi\} \models \Psi))] \\
\mathit{ans}_i &:= \texttt{if } \mathit{log}_i \models \mathit{eval}(\Phi_i)(r) \texttt{ then } \mathit{eval}(\Phi_i)(r) \texttt{ else} \\
&\quad \texttt{if } \mathit{censor}(\mathit{pot\_sec}, \mathit{log}_{i-1}, \Phi_i) \texttt{ then mum else} \\
&\quad \mathit{eval}(\Phi_i)(r) \\
\mathit{log}_i &:= \texttt{if } \mathit{censor}(\mathit{pot\_sec}, \mathit{log}_{i-1}, \Phi_i) \texttt{ then } \mathit{log}_{i-1} \texttt{ else} \\
&\quad \mathit{log}_{i-1} \cup \{\mathit{ans}_i\}.
\end{aligned}$$

A CQE is *secure for pot_sec* if for every finite prefix $\mathcal{Q}'$ of $\mathcal{Q}$ the following holds: For every $\Psi \in \mathit{pot\_sec}$, for every relation r_1, and for every log_0 with $\models_{r_1} \mathit{log}_0$ there is some relation r_2 with $\models_{r_1} \mathit{log}_0$ and (1) $\mathit{cqe}(\mathcal{Q}', \mathit{log}_0)(r_1, \mathit{pot\_sec}) = \mathit{cqe}(\mathcal{Q}', \mathit{log}_0)(r_2, \mathit{pot\_sec})$ and (2) $\mathit{eval}(\Psi)(r_2) = \neg\Psi$. A CQE is *secure* if it is secure for all possible confidentiality policies, and *cqe* is secure in the sense of this definition.

Biskup et al. [8] showed how schemata in ONF with respect to FDs alone, in combination with restricting the confidentiality policy *pot_sec* to so-called facts avoids nontrivial inferences. Let $\mathit{lhs}(\sigma)$ denote the set of attributes that appears on the left-hand side of an FD $\sigma \in \Sigma$, which we have assumed to be a minimal cover. Given a schema in Boyce-Codd Normal Form, we are interested in those attribute sets that might be considered as domains of basic meaningful subtuples. We call these *facts*. Formally, we define the set of facts of a BCNF schema by

$$\mathit{fact}(R) := \{X \subseteq R \mid \exists\sigma \in \Sigma(\mathit{lhs}(\sigma) = X) \cup \{XA \mid A \in R \wedge \exists\sigma \in \Sigma(\mathit{lhs}(\sigma) = X)\}.$$

For example, when $R = PNF$ and $\Sigma = \{P \to N, P \to F, N \to P, N \to F\}$, then *fact*$(R) = \{P, N, PF, PN, NF\}$. A confidentiality policy *pot_sec* is restricted to the facts of R if for every potential secret $\Psi \in$ *pot_sec* the set of attributes instantiated with some constants in Ψ is an element of *fact*(R).

Biskup et al. [8][Theorem 3] were then able to show the following.

Theorem 3. *Let (R, Σ) denote a relation schema that is in Object Normal Form for the given set of FDs over R and pot_sec be restricted to the facts of R. Consider a query Φ_i and the user knowledge log_{i-1} and assume that $log_{i-1} \not\models eval(\Phi_i)(r)$[1]. Then the censor of cqe returns* `true` *if and only if Φ directly implies a potential secret Ψ, that is, if $\Phi_i \models \Psi$.*

Consequently, the answer needs to be refused only if the user asks for a formula directly implying a potential secret. In this case, however, there is no more need for inference control, but access control is sufficient to preserve confidentiality by protecting exactly the elements from *pot_sec*. This leads to access control mechanisms generating a single label per potential secret (for each of them there can be at most one tuple).

8.3 Tigthness of the Conditions

Note that the conditions in Theorem 3 cannot be easily generalized.

Example 7. In our motivating example, the relation schema R consists of the attributes P, N, and F, and the set Σ of FDs consists of $P \to N$, $P \to F$, $N \to P$ and $N \to F$. Consequently, (R, Σ) is in BCNF but not in ONF since it has two minimal keys P and N. Indeed, access to the potential secret $(\exists P)R(P,$ James Muller, $125.60)$ can be bypassed by the queries

$$(\exists N)R(\text{JFY759}, N, 125.60) \text{ and } (\exists F)R(\text{JFY759}, \text{James Muller}, F).$$

Note that the potential secret is limited to the fact $\{N, F\}$. □

Example 8. Similarly, suppose that $\Sigma' = \{P \to N, P \to F\}$. Then (R, Σ') is in Object Normal Form. Suppose we want to protect the potential secret

$$(\exists P)R(P, \text{James Muller}, 125.60),$$

which is not a fact of (R, Σ'). We can issue the queries $(\exists N)R(\text{JFY759}, N, 125.60)$ and $(\exists F)R(\text{JFY759}, \text{James Muller}, F)$, allowing inference of (JFY279, James Muller, 125.60). □

Example 9. Finally, suppose that $\Sigma'' = \{P \to N, N \to F\}$ such that R has the unique minimal key P but is not in BCNF. Suppose we want to protect the potential secret $(\exists P)R(P, \text{James Muller}, 125.60)$, which is a fact of (R, Σ''). Here, we can issue the queries $(\exists N)R(\text{JFY759}, N, 125.60)$ and $(\exists F)R(\text{JFY759}, \text{James Muller}, F)$, allowing inference of (JFY279, James Muller, 125.60) which uncovers the potential secret. □

[1] Note that the censor function is not computed at all whenever $log_{i-1} \models eval(\Phi)(r)$

8.4 Extension to Multivalued Dependencies

Multivalued dependencies occur frequently in database practice. Indeed, FDs $X \to Y$ that are satisfied by a relation r, provide a sufficient condition for r to be the lossless join of $r[XY]$ and $r[X(R-Y)]$. However, MVDs $X \twoheadrightarrow Y$ that are satisfied by a relation r, provide a sufficient and necessary condition for r to be the lossless join of $r[XY]$ and $r[X(R-Y)]$. Wu [37] conducted a practical study identifying that approximately 20% of database schemata in practice satisfy BCNF, but not 4NF. Hence, it is important to consider MVDs.

Theorem 3 can be generalized to schemata that are in ONF for a given set of FDs and MVDs. The reason is that every FD and MVD in a given set over a schema in 4NF are implied by some minimal key.

Proposition 4. *For a schema (R, Σ) that is in 4NF there is an FD set Σ' that is a minimal cover of Σ and where (R, Σ') is in BCNF.*

Proof. Since (R, Σ) is in 4NF, every constraint implied by Σ is implied by a minimal key. Let $K_1, \ldots, K_n$ denote the set of minimal keys for (R, Σ). Then the set $\Sigma' = \{K_i \to R - K_i\}_{i=1}^{n}$ forms a minimal cover of Σ, which is obviously in BCNF. □

Consequently, we can guarantee confidentiality by access control for schemata in ONF, even if the constraint set includes multivalued dependencies.

Theorem 4. *In a relation schema in ONF, controlled query evaluation with pot_sec protecting only facts can be replaced by access control generating a single label per element of pot_sec. The resulting system is still secure.*

Proof. For a given schema (R, Σ) that is in ONF for a set of FDs and MVDs, there is a minimal cover Σ' of Σ that consists of FDs only. Consequently, (R, Σ') is in ONF for the FD set Σ'. The theorem follows from Theorem 3. □

Similar to Theorem 3, the conditions of Theorem 4 cannot be easily generalized. In Example 7 the minimal FD cover Σ might be represented in the form $\Sigma_m = \{P \twoheadrightarrow N, N \twoheadrightarrow P, PN \to F, PF \to N, NF \to P\}$. Then (PNF, Σ_m) is in 4NF and has the minimal keys P and N. Indeed, $P \twoheadrightarrow N$ implies $P \twoheadrightarrow PN$, and $P \twoheadrightarrow PN$ and $PN \to F$ imply $P \to F$. Similarly, $P \twoheadrightarrow N$ implies $P \twoheadrightarrow F$ and$P \twoheadrightarrow PF$, and $P \twoheadrightarrow PF$ and $PF \to N$ imply $P \to N$. Similarly, we can infer $N \to P$ and $N \to F$.Vice versa, all elements of Σ_m are implied by Σ. Consequently, Example 7 shows how access control based on Σ_m can be bypassed.

In Example 8 the minimal FD cover Σ' might be represented in the form $\Sigma'_m = \{P \twoheadrightarrow N, PN \to F, PF \to N\}$. Then (PNF, Σ'_m) is in 4NF and has the minimal key P. Consequently, Example 8 shows how access control based on Σ'_m can be bypassed.

In Example 9 the minimal FD cover Σ'' might be represented in the form $\Sigma''_m = \{P \twoheadrightarrow N, N \twoheadrightarrow F, PN \to F, PF \to N\}$. Then (PNF, Σ''_m) has the minimal key P, but is not in 4NF. Consequently, Example 9 shows how access control based on Σ''_m can be bypassed.

9 Conclusion and Future Work

We have provided further insight into the effort required to i) maintain data consistency under updates, and ii) guarantee data confidentiality under inference attacks. Firstly, we found that updates over schemata in 4NF with a unique minimal key only require attention to values on attributes of the key when inserts happen. Secondly, we found that schemata in 4NF with a unique minimal key can guarantee confidentiality by simple means of access control, as long as potential secrets are declared over facts.

In future work it would be interesting to understand the effort required for data consistency and confidentiality when schemata are in less restrictive normal forms, such as 3NF, BCNF, or 4NF with a fixed number of keys. The properties of uniqueness and independence appear to be interesting subjects of study in richer data models, including incomplete, inaccurate, and uncertain data.

References

1. Arenas, M.: Normalization theory for XML. SIGMOD Record **35**(4), 57–64 (2006)
2. Biskup, J.: Boyce-Codd normal form and Object Normal Forms. Inf. Process. Lett. **32**(1), 29–33 (1989)
3. Biskup, J.: Achievements of relational database schema design theory revisited. In: Semantics in Databases, Selected Papers from a Workshop, Prague, Czech Republic, 1995. pp. 29–54 (1995)
4. Biskup, J., Bonatti, P.A.: Controlled query evaluation for enforcing confidentiality in complete information systems. Int. J. Inf. Sec. **3**(1), 14–27 (2004)
5. Biskup, J., Bonatti, P.A.: Controlled query evaluation for known policies by combining lying and refusal. Ann. Math. Artif. Intell. **40**(1-2), 37–62 (2004)
6. Biskup, J., Dayal, U., Bernstein, P.A.: Synthesizing independent database schemas. In: Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data, Boston, Massachusetts, USA, May 30 - June 1. pp. 143–151 (1979)
7. Biskup, J., Dublish, P.: Objects in relational database schemes with functional, inclusion, and exclusion dependencies. RAIRO Theor. Informatics Appl. **27**(3), 183–219 (1993)
8. Biskup, J., Embley, D.W., Lochner, J.: Reducing inference control to access control for normalized database schemas. Inf. Process. Lett. **106**(1), 8–12 (2008)
9. Bonatti, P.A., Kraus, S., Subrahmanian, V.S.: Foundations of secure deductive databases. IEEE Trans. Knowl. Data Eng. **7**(3), 406–422 (1995)
10. Brodsky, A., Farkas, C., Jajodia, S.: Secure databases: Constraints, inference channels, and monitoring disclosures. IEEE Trans. Knowl. Data Eng. **12**(6), 900–919 (2000)
11. Chen, P.P.: The entity-relationship model - toward a unified view of data. ACM Trans. Database Syst. **1**(1), 9–36 (1976)
12. Codd, E.F.: Recent investigations in relational data base systems. In: Information Processing, Proceedings of the 6th IFIP Congress 1974, Stockholm, Sweden, August 5-10, 1974. pp. 1017–1021 (1974)
13. Cuppens, F., Gabillon, A.: Logical foundations of multilevel databases. Data Knowl. Eng. **29**(3), 259–291 (1999)
14. Dawson, S., di Vimercati, S.D.C., Lincoln, P., Samarati, P.: Minimal data upgrading to prevent inference and association. In: Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA. pp. 114–125 (1999)

15. Elmasri, R., Navathe, S.B.: Fundamentals of Database Systems, 3rd Edition. Addison-Wesley-Longman (2000)
16. Fagin, R.: Multivalued dependencies and a new normal form for relational databases. ACM Trans. Database Syst. **2**(3), 262–278 (1977)
17. Fagin, R.: A normal form for relational databases that is based on domains and keys. ACM Trans. Database Syst. **6**(3), 387–415 (1981)
18. Farkas, C., Jajodia, S.: The inference problem: A survey. SIGKDD Explor. **4**(2), 6–11 (2002)
19. Gollmann, D.: Computer Security (3. ed.). Wiley (2011)
20. Hartmann, S., Link, S.: The implication problem of data dependencies over SQL table definitions: Axiomatic, algorithmic and logical characterizations. ACM Trans. Database Syst. **37**(2), 13:1–13:40 (2012)
21. Jensen, C.S., Snodgrass, R.T., Soo, M.D.: Extending existing dependency theory to temporal databases. IEEE Trans. Knowl. Data Eng. **8**(4), 563–582 (1996)
22. Köhler, H., Link, S.: SQL schema design: Foundations, normal forms, and normalization. Inf. Syst. **76**, 88–113 (2018)
23. Link, S.: Characterisations of multivalued dependency implication over undetermined universes. J. Comput. Syst. Sci. **78**(4), 1026–1044 (2012)
24. Link, S., Prade, H.: Relational database schema design for uncertain data. Inf. Syst. **84**, 88–110 (2019)
25. Link, S., Wei, Z.: Logical schema design that quantifies update inefficiency and join efficiency. In: SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021. pp. 1169–1181 (2021)
26. Lunt, T.F., Denning, D.E., Schell, R.R., Heckman, M.R., Shockley, W.R.: The seaview security model. IEEE Trans. Software Eng. **16**(6), 593–607 (1990)
27. Mok, W.Y.: Utilizing nested normal form to design redundancy free JSON schemas. iJES **4**(4), 21–25 (2016)
28. Mok, W.Y., Ng, Y., Embley, D.W.: A normal form for precisely characterizing redundancy in nested relations. ACM Trans. Database Syst. **21**(1), 77–106 (1996)
29. Olivier, M.S., von Solms, S.H.: A taxonomy for secure object-oriented databases. ACM Trans. Database Syst. **19**(1), 3–46 (1994)
30. Sicherman, G.L., de Jonge, W., van de Riet, R.P.: Answering queries without revealing secrets. ACM Trans. Database Syst. **8**(1), 41–59 (1983)
31. Tari, Z., Stokes, J., Spaccapietra, S.: Object normal forms and dependency constraints for object-oriented schemata. ACM Trans. Database Syst. **22**(4), 513–569 (1997)
32. Vincent, M.W.: A corrected 5nf definition for relational database design. Theor. Comput. Sci. **185**(2), 379–391 (1997)
33. Vincent, M.W.: Semantic foundations of 4nf in relational database design. Acta Informatica **36**(3), 173–213 (1999)
34. Vincent, M.W., Levene, M.: Restructuring partitioned normal form relations without information loss. SIAM J. Comput. **29**(5), 1550–1567 (2000)
35. Wang, W.H., Liu, R.: Privacy-preserving publishing data with full functional dependencies. In: Database Systems for Advanced Applications, 15th International Conference, DASFAA 2010, Tsukuba, Japan, April 1-4, 2010, Proceedings, Part II. pp. 176–183 (2010)
36. Wei, Z., Link, S.: Embedded functional dependencies and data-completeness tailored database design. ACM Trans. Database Syst. **46**(2), 7:1–7:46 (2021)
37. Wu, M.S.: The practical need for fourth normal form. In: Proceedings of the 23rd SIGCSE Technical Symposium on Computer Science Education, SIGCSE 1992, Kansas City, Missouri, USA, March 5-6, 1992. pp. 19–23 (1992)