Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Journal Track

A False Sense of Security (Extended Abstract)*

Piero A. Bonatti
Università degli Studi di Napoli Federico II
pab@unina.it

Abstract

The growing literature on confidentiality in knowledge representation and reasoning sometimes may cause a false sense of security, due to lack of details about the attacker, and some misconceptions about security-related concepts. This note analyzes the vulnerabilities of some recent knowledge protection methods to increase the awareness about their actual effectiveness and their mutual differences.

Knowledge bases may contain confidential or even sensitive information that should be protected by concealing axioms, individuals, and query answers. The most important aspect in the design of a confidentiality-preserving system is the underlying *confidentiality model*, that is, the criterion that specifies under which conditions confidential information is considered to be protected.

The first confidentiality model that comes to one's mind is the *simple confidentiality model* (SCM). Informally, it requires that: *no confidential piece of information should be a logical consequence of the visible axioms and query answers*.

In [Bonatti and Sauro, 2013], the weaknesses of this confidentiality model have been analyzed by introducing several attacks to the main access control methods for knowledge bases available in 2013, that were all based on the SCM, and operated mostly by publishing a subset of the knowledge base. The solution proposed by B. and Sauro [2013] consisted in adopting a stronger, *indistinguishability-based* (IB) confidentiality model, according to which: *the observable behavior of a knowledge base should be indistinguishable from that of a knowledge base that contains no secrets*. The IB model has been introduced and extensively studied in the literature on *controlled query evaluation* (CQE) for complete and incomplete databases, see for example [Biskup, 2000; Biskup and Bonatti, 2001; Biskup and Bonatti, 2004; Biskup and Bonatti, 2007; Biskup *et al.*, 2010].

Despite the above analysis, some later approaches were again based on the SCM [Cuenca Grau *et al.*, 2015; Lembo *et al.*, 2019; Cuenca Grau and Kostylev, 2019], and inherit its vulnerabilities. A couple of very recent approaches adopt the IB model, instead [Benedikt *et al.*, 2018; Cima *et al.*, 2020]. However, due to some simplifications to the confidentiality model, they are vulnerable to some of the attacks introduced in [Bonatti and Sauro, 2013].

The rigorous formal confidentiality proofs contained in these papers may give a false sense of security to prospective adopters. Thus it is paramount to increase the awareness about the different effectiveness of the two confidentiality models, and about the importance of making the hypotheses on attackers fully explicit. In particular, the full paper:

- discusses several simple attacks that illustrate the main vulnerabilities of the new SCM-based approaches;
- it illustrates why the IB criterion is preferable to the SCM.
- it analyzes the risks introduced by simplifying the general IB models introduced in [Bonatti and Sauro, 2013; Studer and Werner, 2014];
- it discusses a few misconceptions, and the risks caused by opaque assumptions about the attackers.

Unfortunately, in this abstract, space is not sufficient to illustrate the attacks to the approaches based on the SCM. We only mention that the mappings from the original knowledge bases to their secure views are – roughly speaking – frequently injective; therefore, an attacker may invert the mapping, and reconstruct the information that has been concealed. Next, we summarize how the full paper addresses the other points.

Concerning the reasons for preferring the IB model to the SCM model and the opaqueness of the assumptions about attackers, we note that the papers based on the SCM do not assume explicitly that attackers should *not* know the protection algorithm – or that at least they should not know the TBox. The full paper shows that: (i) without such hypotheses (which are typically false in the real world) the methods based on the SCM do not protect confidentiality, and (ii) *IB approaches are secure even if these two hypotheses are violated*. This difference has not been sufficiently highlighted in the literature.

As a consequence of the above discussion, *the assumptions on the attacker should always be explicit*, so as to prevent a false sense of security by warning potential adopters about the prerequisites of the access control mechanisms. Moreover, by evaluating such prerequisites, appropriate protection methods can be chosen.

*The full paper has been published in *Artificial Intelligence* [Bonatti, 2022].

The general IB model by B. and Sauro [2013] has been simplified in some works by removing the attackers' meta-knowledge. This simplification enables attacks where structural properties of the knowledge base are leveraged to reconstruct concealed facts.

Finally, concerning misconceptions, in the literature some IB approaches are erroneously classified as SCM models, and some comparisons of SCM-based and IB-based approaches fail to recognize the different robustness of the two methods.

References

[Benedikt *et al.*, 2018] Michael Benedikt, Bernardo Cuenca Grau, and Egor V. Kostylev. Logical foundations of information disclosure in ontology-based data integration. *Artif. Intell.*, 262:52–95, 2018.

[Biskup and Bonatti, 2001] Joachim Biskup and Piero A. Bonatti. Lying versus refusal for known potential secrets. *Data Knowl. Eng.*, 38(2):199–222, 2001.

[Biskup and Bonatti, 2004] Joachim Biskup and Piero A. Bonatti. Controlled query evaluation for enforcing confidentiality in complete information systems. *Int. J. Inf. Sec.*, 3(1):14–27, 2004.

[Biskup and Bonatti, 2007] Joachim Biskup and Piero A. Bonatti. Controlled query evaluation with open queries for a decidable relational submodel. *Ann. Math. Artif. Intell.*, 50(1-2):39–77, 2007.

[Biskup *et al.*, 2010] Joachim Biskup, Cornelia Tadros, and Lena Wiese. Towards controlled query evaluation for incomplete first-order databases. In Sebastian Link and Henri Prade, editors, *Foundations of Information and Knowledge Systems, 6th International Symposium, FoIKS 2010, Sofia, Bulgaria, February 15-19, 2010. Proceedings*, volume 5956 of *Lecture Notes in Computer Science*, pages 230–247. Springer, 2010.

[Biskup, 2000] Joachim Biskup. For unknown secrecies refusal is better than lying. *Data Knowl. Eng.*, 33(1):1–23, 2000.

[Bonatti and Sauro, 2013] Piero A. Bonatti and Luigi Sauro. A confidentiality model for ontologies. In Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha F. Noy, Chris Welty, and Krzysztof Janowicz, editors, *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, volume 8218 of *Lecture Notes in Computer Science*, pages 17–32. Springer, 2013.

[Bonatti, 2022] Piero A. Bonatti. A false sense of security. *Artif. Intell.*, 310:103741, 2022.

[Cima *et al.*, 2020] Gianluca Cima, Domenico Lembo, Riccardo Rosati, and Domenico Fabio Savo. Controlled query evaluation in description logics through instance indistinguishability. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1791–1797. ijcai.org, 2020.

[Cuenca Grau and Kostylev, 2019] Bernardo Cuenca Grau and Egor V. Kostylev. Logical foundations of linked data anonymisation. *J. Artif. Intell. Res.*, 64:253–314, 2019.

[Cuenca Grau *et al.*, 2015] Bernardo Cuenca Grau, Evgeny Kharlamov, Egor V. Kostylev, and Dmitriy Zheleznyakov. Controlled query evaluation for datalog and OWL 2 profile ontologies. In Qiang Yang and Michael J. Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2883–2889. AAAI Press, 2015.

[Lembo *et al.*, 2019] Domenico Lembo, Riccardo Rosati, and Domenico Fabio Savo. Revisiting controlled query evaluation in description logics. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1786–1792. ijcai.org, 2019.

[Studer and Werner, 2014] Thomas Studer and Johannes Werner. Censors for boolean description logic. *Trans. Data Priv.*, 7(3):223–252, 2014.