How do we know that our system is correct?

Hana Chockler

King's College London, UK

Abstract. A negative answer from the model-checking procedure is accompanied by a counterexample – a trace demonstrating what went wrong. On the other hand, when the answer from the model-checker is positive, usually no further information is given. The issue of "suspecting the positive answer" first arose in industry, where positive answers from model-checkers often concealed serious bugs in hardware designs. In this talk, I discuss some reasons why the positive answer from the model-checker may require further investigation and briefly and in broad terms describe algorithms for such investigations, called *sanity checks*.
The talk also (briefly) introduces the theory of causality and counterfactual reasoning and its applications to model-checking, mostly in the context of the subject of this talk, including some recent complexity results and applications of structure-based causality.
The talk then attempts to define the main goal of the sanity checks, explanations, and related algorithms, or at least provide some food for thought regarding the question of the mail goal.
I conclude the talk with outlining some promising future directions.
The talk is based on many papers written by many people, and is not limited to my own research. It is reasonably self-contained.