Extraction and checking of LaTeX mathematics

“Client”: Dominique Unruh, Cryptography Group

Quick summary: Develop a tool that parses LaTeX documents for formulas and definitions of mathematical symbols. Formulas and definitions are extracted in a parsed, semantic format (e.g., Content MathML), and checked for correctness (i.e., syntax checking).

Rationale: LaTeX documents often contain a large amount of formulas. These are specified in a presentation oriented format (i.e., the layout of the formula is described, but not its meaning). There is no way to automatically check for errors in the syntax of formulas. (Compare this, e.g., with writing programs – imagine if the compiler would not tell you if you have a syntax error.) A tool that parses the mathematics in a document will have the following advantages:

- Syntactically wrong formulas are detected early on. (Catches many typos.)
- It can be checked whether all mathematical symbols used in a paper are actually declared.
- Formulas can be type checked for further checking of correctness
- Refactoring of papers can be supported (e.g., I want to systematically write $a_x$ instead of $a(x)$).

Type checking and refactoring are outside the scope of this project. But the output of this project will strongly simplify those tasks.