Certified Connection Tableaux Proofs
for HOL Light and TPTP

Cezary Kaliszyk
University of Innsbruck
cezary.kaliszyk@uibk.ac.at

Josef Urban
Radboud University Nijmegen
josef.urban@gmail.com

Jiří Vyskočil
Czech Technical University in Prague
jiri.vyskocil@gmail.com

Abstract
In the recent years, the Metis prover based on ordered paramodulation
and model elimination has replaced the earlier built-in methods
for general-purpose proof automation in HOL 4 and Isabelle/HOL.
In the annual CASC competition, the leanCoP system based on
connection tableaux has however performed better than Metis. In
this paper we show how the leanCoP’s core algorithm can be im-
plemented inside HOL Light. leanCoP’s flagship feature, namely
its minimalistic core, results in a very simple proof system. This
plays a crucial role in extending the MESON proof reconstruction
mechanism to connection tableaux proofs, providing an implementa-
tion of leanCoP that certifies its proofs. We discuss the differ-
ences between our direct implementation using an explicit Pro-
log stack, to the continuation passing implementation of
leanCoP. We compare its performance against the
resolution based
HOL Light
goals. The resulting prover can be also used as a gen-
eral purpose TPTP prover. We compare its performance against
the resoluted tables of Metis on TPTP and other interesting datasets.

Categories and Subject Descriptors CR-number [subcategory]: third-level

Keywords keyword1, keyword2

1. Introduction and Related Work
The leanCoP [19] automated theorem prover (ATP) has an unusu-
ally good ratio of performance to its implementation size. While its
core algorithm fits on some twenty lines of Prolog, starting with
CASC-21 [27] it has regularly beaten Otter [17] and Metis [5]
in the FOF division of the CASC ATP competitions. In 2014,
leanCoP solved 158 FOF problems in CASC-J7 while Prover9
solved 95 problems. On the large-theory (chainy) division of
the MPTP Challenge benchmark leanCoP’s goal-directed calculus
beats also SPASS 2.2 [31], and its further AI-style strengthening
by integrating into leanCoP’s simple core learning-based guidance
trained on such larger ITP corpora is an interesting possibility [34].

Compact ATP calculi such as leanTAP [3] and MESON [16]
have been used for some time in Isabelle [20, 21] and HOLs [4]
as general first-order automation tactics for discharging goals that are
already simple enough. With the arrival of large-theory “hammer”
linkups [2] [13] [23] between ITPs, state-of-the-art ATPs such as
Vampire [12] and E [24], and premise selection methods [14],
such tactics also became used as a relatively cheap method for
reconstructing the (minimized) proofs found by the stronger ATPs.
In particular, Hurd’s Metis has been adopted as the main proof
reconstruction tool used by Isabelle’s Sledgehammer linkup [3]
[23], while Harrison’s version of MESON could reconstruct in 1
second about 80% of the minimized proofs found by E in the first
experiments with the HOL’ Hammer linkup [4].

Since HOL Light already contains a lot of the necessary in-
frastructure for Prolog-style proof search and its reconstruction,
integrating leanCoP into HOL Light in a similar way as MESON
should not be too costly, while it could lead to interesting strength-
ening of HOL Light’s first-order automation and proof reconstruc-
tion methods. In this paper we describe how this was done, resulting
in an OCaml implementation of leanCoP and a general leanCoP
first-order tactic in HOL Light. We compare their performance with
MESON, Metis and the Prolog version of leanCoP in several sce-
narios, showing quite significant improvements over MESON and
Metis.

2. leanCoP and Its Calculus

leanCoP is an automated theorem prover for classical first-order
logic based on a compact Prolog implementation of the clausal
connection (tableaux) calculus [12] [19] with several simple strate-
gies that significantly reduce the search space on many problems.
In contrast to saturation-based calculi used in most of the state-
of-the-art ATPs (E, Vampire, etc.), connection calculus implements
an open literal on the current path to a new literal with the same predicate
symbol but different polarity. The formal definition (derived from
Otten [18]) of the particular connection calculus relevant in lean-
CoP is as follows:

\textbf{Definition 1.} \textbf{[Connection calculus]} The axiom and rules of the
connection calculus are given in Figure 1. The words of the cal-
culus are tuples \”C, M, Path\” where the clause \(C\) is the \textit{open}
subgoal, \(M\) is a set of clauses in disjunctive normal form (DNF)
transformed from axioms \& conjecture with added nullary predic-
te \(\exists\) to all positive clause \(\exists\) and the \textit{active path Path} is a subset
of a path through \(M\). In the rules of the calculus \(C, C'\) and \(C''\) are
clauses, \(\sigma\) is a term substitution, and \(L_1, L_2\) is a \textit{connection} with

\begin{itemize}
  \item We suppose that \(\exists\) is a new predicate that does not occur anywhere in
  axioms and conjecture
  \item Thus by default all positive clauses are used as possible start clauses.
\end{itemize}
\( \sigma(L_1) = \sigma(L_2) \). The rules of the calculus are applied in an analytic (i.e., bottom-up) way. The term substitution \( \sigma \) is applied to the whole derivation.

The connection calculus is correct and complete in the following sense: A first-order formula \( M \) in clausal form is valid iff there is a connection proof for \( \neg \exists \neg M, \{ \} \), i.e., a derivation for \( \neg \neg \exists \neg M, \{ \} \) in the connection calculus so that all leaves are axioms. The Prolog predicate prove/5 implements the axiom, the reduction and the extension rule of the basic connection calculus in leanCoP:

\[
\begin{align*}
1 & \text{prove}\left([\text{Lit}|\text{Cla}], \text{Path}, \text{PathLim}, \text{Lem}, \text{Set}\right) \ :- \\
2 & \quad \text{\%} \\
3 & \quad \text{\%} \\
4 & \quad \text{\%} \\
5 & \quad \text{\%} \\
6 & \quad \text{\%} \\
7 & \quad \text{\%} \\
8 & \quad \text{\%} \\
9 & \quad \text{\%} \\
10 & \quad \text{\%} \\
11 & \quad \text{\%} \\
12 & \quad \text{\%} \\
13 & \quad \text{\%} \\
14 & \quad \text{\%} \\
15 & \quad \text{\%} \\
16 & \quad \text{\%} \\
17 & \quad \text{\%} \\
18 & \text{prove}(\text{Set}).
\end{align*}
\]

The tuple \( C, M, \text{Path} \) in connection calculus is here represented as follows:

- \( C \) representing the open subgoal is a Prolog list \( \text{Cla} \);
- the active path \( \text{Path} \) is a Prolog list \( \text{Path} \);
- \( M \) is written into Prolog’s database before the actual proof search starts in a such way that for every clause \( C \in M \) and for every literal \( \text{Lit} \in \text{Indexing}_L \text{L} \) and \( \text{Grnd} \) is stored, where \( \text{Cl}=C(\text{L}) \) and \( \text{Grnd} \) is if \( C \) is ground, otherwise \( \text{Grnd} \) is a list of options, \( \text{Prolog} \) is used in the following sense starting at the database with \( \text{Cla} \), the DNF representation of the problem stored in prolog’s database using the literal predicate, and a path with \( \text{Path} \) \( \text{Path} \) where \( \text{PathLim} \) is the maximum size of the active \( \text{Path} \). The predicate works as follows:

Line 18 implements the axiom, line 4 calculates the complement of the first literal \( \text{Lit} \) in \( \text{Cla} \), which is used as the principal literal for the next reduction or extension step. The reduction rule is implemented in lines 7, 8 and 17. At line 7 and 8 it is tested whether the active path \( \text{Path} \) contains a literal \( \text{NegLit} \) that unifies with the complement \( \text{NegLit} \) of the principal literal \( \text{Lit} \). In this case the alternative lines after the semicolon are skipped and the proof search for the premise of the reduction rule is invoked in line 17. The extension rule is implemented in lines 10, 11, 14 and 17. Lines 10 and 11 are used to find a clause that contains the complement \( \text{NegLit} \) of the principal literal \( \text{Lit} \) \( \text{Cla} \) is then the remaining set of literals of the selected clause and the new open subgoal of the left premise. The proof search for the left premise of the extension rule, in which the active path \( \text{Path} \) is extended by the principal literal \( \text{Lit} \), is invoked in line 14, and if successful, we again continue on line 17.

Compared to standard tableau or sequent calculi, connection calculi are not confluence. To achieve completeness, an extensive use of backtracking is required. leanCoP uses two simple incomplete strategies (namely options cut and cut) for restricting backtracking that significantly reduces the search space without affecting the ability to find proofs in most tested cases (see Section 4).

Another major problem in connection calculi is the integration of equality. The paramodulation method that is widely used in saturation-based ATPs is not complete for goal-oriented approach of connection calculi. Therefore equality in leanCoP and similar ATPs is usually managed by adding the axioms of equality (reflexivity, symmetry, transitivity and substitutivity). To obtain the clausal form, leanCoP uses its own implementation of classifier introducing definitions (the def option), which seems to perform better with the leanCoP’s core prover than other standard classifiers (TPTP2X using the option \( \neg \text{clausify:tptp} \), FLOTTER and E) or direct transformation into clausal form (nodef option in leanCoP). In the following subsections, we summarize several further methods used by leanCoP that improve its performance.

2.1 Iterative deepening

Prolog uses a simple incomplete depth-first search strategy to explore the search space. This kind of incompleteness would result in a calculus that hardly proves any formula. In order to obtain a complete proof search in the connection calculus, iterative deepening on the proof depth, i.e. the size of the active path, is performed. It is achieved by inserting the following lines into the code:

\[
\begin{align*}
(12) & \text{ Grnd} \text{eq} \rightarrow \text{ true } ; \\
(13) & \text{ pathlim} \rightarrow \text{ true } ; \\
(14) & \text{ pathlim} \rightarrow \text{ assert(pathlim), fail },
\end{align*}
\]

When the extension rule is applied and the new clause is not ground, i.e. it does not contain any variable, it is checked whether the size \( K \) of the active path exceeds the current path limit \( \text{PathLim} \).
(line 12). In this case the dynamic predicate \texttt{pathlim/0} is written into the Prolog’s database (line 13) indicating the need to increase the path limit if the proof search with the current path limit fails. If the proof search fails and the predicate \texttt{pathlim} can be found in the database, then \texttt{PathLim} is increased by one and the proof search starts again.

2.2 Regularity Condition Optimization

Definition 2. A connection proof is \textit{regular} iff no literal occurs more than once in the active path.

Since the active path corresponds to the set of literals in a branch in the connection tableau representation, a connection tableau proof is regular if in the current branch no literal occurs more than once. The regularity condition is integrated into the connection calculus in Figure \cite{1} by imposing the following restriction on the reduction and extension rule: \( \forall L' \in C \cup \{L\} : \sigma(L') \notin \sigma(Path) \)

Lemma 2.1. A formula \( M \) in clausal form described above is valid iff there is a regular connection proof for \( \lnot \exists y. M. \{\} \)

Regularity is correct, since it only imposes a restriction on the applicability of the reduction and extension rules. The completeness proof can be found in \cite{18,19}.

The Prolog predicate \( \\backslash + \text{Goal} \) succeeds only if \textit{Goal} cannot be proven. In line 3 the corresponding \textit{Goal} succeeds if the open subgoal [\texttt{Lit|Cla}] contains a literal \texttt{LitC} that is syntactically identical (built-in predicate \( \texttt{==/2} \) in Prolog) with a literal \texttt{LitP} in the active path \texttt{Path}. The built-in predicate \texttt{member/2} is used to enumerate all elements of a list.

2.3 Lemma optimization

The set of lemmata is represented by the list \texttt{Lem}. The lemma rule is implemented by inserting the following lines:

\begin{verbatim}
(3) \(+ \) (member(LitC,[Lit|Cla]), member(LitP,Path), LitC==LitP),
\end{verbatim}

The \texttt{Prolog} predicate \( \\backslash + \text{Goal} \) succeeds only if \textit{Goal} cannot be proven. In line 3 the corresponding \textit{Goal} succeeds if the open subgoal [\texttt{Lit|Cla}] contains a literal \texttt{LitC} that is syntactically identical (built-in predicate \( \texttt{==/2} \) in Prolog) with a literal \texttt{LitP} in the active path \texttt{Path}. The built-in predicate \texttt{member/2} is used to enumerate all elements of a list.

2.4 Restricted backtracking

In Prolog the cut (!) is used to cut off alternative solutions when Prolog tries to prove a goal. The Prolog cut is a built-in predicate, which succeeds immediately when first encountered as a goal. Any attempt to re-satisfy the cut fails for the parent goal, i.e. other alternative choices are discarded that have been made from the point when the parent goal was invoked. Consequently, \textit{restricted backtracking} is achieved by inserting a Prolog cut after the lemma, reduction, or extension rule is applied. It is implemented by inserting the following line into the code:

\begin{verbatim}
(16) \{ member(cut,Set) -> ! ; true \},
\end{verbatim}

Restricted backtracking is switched on if the list \texttt{Set} contains the option \texttt{cut}. The \textit{restricted start step} is used if the list \texttt{Set} includes the option \texttt{scut}. In this case only the first matching clause to starting \( \lnot \exists \) literal is used.

Restricted backtracking and restricted start step lead to an incomplete proof search. In order to regain completeness, these strategies can be switched off when the search reaches a certain path limit. If the list \texttt{Set} contains the option \texttt{comp(Limit)}, where \texttt{Limit} is a natural number, the proof search is stopped and started again without using incomplete search strategies.

3. OCaml Implementation

In this section, we first discuss our implementation of \textit{leanCoP} in OCaml and its integration in HOL Light: the transformation of the higher-order goal to first order and the proof reconstruction. After that we compare our implementation to Harrison’s implementation of MESON.

3.1 \textit{leanCoP} in OCaml

Otten’s implementation of \textit{leanCoP} uses the Prolog search, backtracking, and indexing mechanisms to implement the connection tableau proof search. This is a variation of the general idea of using the “Prolog technology theorem prover” (PTTP) proposed by Stickel \cite{25}, in which connection tableaux takes a number of advantages from its similarity to Prolog.

In order to implement an equivalent program in a functional programming language, one needs to use either an explicit stack for keeping track of the current proof state (including the trail of variable bindings), or the continuation passing style. We choose to do the former, namely we add an explicit \texttt{todo} (stack), \texttt{subst} (trail) and \texttt{off} (offset in the trail) arguments to the main \texttt{prove} function. The stack keeps a list of tuples that are given as arguments to the recursive invocations of \texttt{prove}, whose full OCaml declaration (taking the open subgoal as its last argument) then looks as follows:

\begin{verbatim}
let rec lprove off subst path limit lemmas todo = function
  | [] -> begin ... end
  | ((lit1 :: rest_clause) as clause) -> ... ;
\end{verbatim}

The function performs the proof search to the given depth, and if a proof has not been found, it returns the unit. It takes special attention to traverse the tree in the same order as the Prolog version. In particular, when the global option "cut" (restricting backtracking) is off, it performs all the backtracking explicitly, while if "cut" is on, the parts of backtracking avoided in Prolog are also omitted. When a proof is found, the exception 'Solved' is raised: no further search is performed and the function exits with this exception.

The OCaml version and the Prolog version (simplified and with symbols renamed for clarity of comparison) are displayed together in Fig \cite{2}. The algorithm proceeds as follows:

1. If nonempty, decompose the current open subgoal into the first literal \texttt{lit} and the rest \texttt{cla}.
2. Check for intersection between the current open subgoal and \texttt{path}.
3. Compute the negation of \texttt{lit}.

\footnote{Available online at \url{http://cl-informatik.uibk.ac.at/users/cek/cppl5}}
Let rec prove path lim lem stack = function (lit :: cla) ->
prove([lit|Cla],Path,PathLim,Lem,Set) :-
if not (exists (fun litp -> exists (subseq litp) path))
(lit :: cla) then (prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

prove([lit|Cla],Path,PathLim,Lem,Set) :-
\+ (member(LitC, [Lit|Cla]), member(LitP, Path),
LitC==LitP),
(-NegLit=Lit; Lit=NegLit) -> {
member(LitL,Lem), Lit==LitL
}

let rec prove path lim lem stack = function (lit :: cla) ->

prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

prove([lit|Cla],Path,PathLim,Lem,Set) :-
if not (exists (fun litp -> exists (subseq litp) path))
(lit :: cla) then (prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

prove([lit|Cla],Path,PathLim,Lem,Set) :-
if not (exists (fun litp -> exists (subseq litp) path))
(lit :: cla) then (prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

prove([lit|Cla],Path,PathLim,Lem,Set) :-
if not (exists (fun litp -> exists (subseq litp) path))
(lit :: cla) then (prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

prove([lit|Cla],Path,PathLim,Lem,Set) :-
if not (exists (fun litp -> exists (subseq litp) path))
(lit :: cla) then (prove path lim lem stack cla; cut) :: stack
if not (fold_left (fun sf plit -> if sf then true
else try (unify_lit neglit plit; prove path lim (lit :: lem)
stack cla; cut) with Unify -> sf) false path) then (

let iter_fun (lit2, cla2, ground) = if lim > 0 ||
| ground then try let clal = unify_rename (snd lit) (lit2, cla2) in
| prove (lit :: path) (lim - 1) lem ((if cut then lim else -1)),
| path, lim, lit :: lem, cla) :: stack) cla1 with Unify -> () in
| try iter iter_fun (try assoc neglit lits with Not_found -> [])
| with Cut n -> if n = lim then () else raise Cut n))
| | [] -> match stack with
| | (ct, path, lim, lem, cla) :: t ->
| | prove path lim lem t cla; if ct > 0 raise (Cut ct)
| | [] -> raise Solved;

Figure 2. The simplified OCaml and Prolog code side by side. The explicit trail argument and the computation of the resulting proof have been omitted for clarity, and some symbols were renamed to better correspond to each other. White-space and order of clauses has been modified to exemplify corresponding parts of the two implementations. Function subseq checks equality under the current context of variable bindings. Note that the last-but-one line in the Prolog code was merged into each of the three cases in the OCaml code. See the function lprove in file leancop.ml on our web page for the actual (non-simplified) OCaml code.

4. Check if lit is among the lemmas lem, if so try to prove cla.
If cut is set, no other options are tried.
5. For each literal on the path, if neglit unifies with it, try to prove cla. If the unification succeeded and cut is set, no other options are tried.
6. For each clause in the matrix, try to find a literal that unifies with neglit, and then try to prove the rest of the newly created subgoal and the rest of the current open subgoal. If the unification and the first proof succeeded and cut is set, no other options are tried.
7. When the current open subgoal is empty, the subproof is finished (the axiom rule).

In Otten’s implementation, the behaviour of the program with cut set is enabled by the use of the Prolog cut (!). Implementing it in OCaml amounts to a different mechanism in each of the three cases. In point 4 in the enumeration above, given that a single lemma has been found, there is no need to check for other lemmas. Therefore a simple List.exists call is sufficient to emulate this behaviour in OCaml. No backtracking over other possible occurrences of the lemma is needed here, and it is not necessary to add in this case the literal again into the list of lemmas as is done in the Prolog code (last-but-one line).

In point 5, multiple literals on the path may unify with different substitutions. We therefore use a list fold mechanism which changes the value whenever the unification is successful and cut is set. In point 6, we need to change our behaviour in between two successive calls to prove. As the first call takes the arguments to the second call on the stack, we additionally add a cut marker on the stack and handle an exception that can be raised by the call on the stack.

Whenever the clause becomes empty, all the tuples in the stack list need to be processed. For each tuple, the first component is the cut marker: if it is set, the Cut exception is raised with a depth level. The exception is handled only at the appropriate level.

This directly corresponds to the semantics of the cut operator in Prolog [26].

What remains to be implemented is efficient equality checking and unification. Since we want to integrate our mechanism in HOL Light, we reuse the first order logic representation used in the implementation of HOL Light’s MESON procedure: the substitutions are implemented as association lists, and applications of substitutions are delayed until an actual equality check or a unification step.

3.2 leanCoP in HOL Light

In order to enable the OCaml version of leanCoP as a proof tactic and procedure in HOL Light, we first need to transform a HOL goal to a leanCoP problem and when a proof has been found we replay the proof in higher-order logic. In order to transform a problem in higher-order logic to first-order logic without equality, we mostly reuse the steps of the transformation already used by MESON, additionally ensuring that the conjecture is separated from the axioms to preserve leanCoP’s goal-directed approach. The transformation starts by assuming the duplicated copies of polymorphic theorems to match the existing assumptions. Next the goal axioms → conjecture is transformed to (axioms ∧ ∀) → (conjecture ∧ ∀) with the help of a special symbol, which we define in HOL as: ∀ = ⊤. Since the conjecture is refuted and the problem is converted to CNF, the only positive occurrence of ∀ is present in a singleton clause, and the negative occurrences of ∀ are present in every clause originating from the conjecture. The CNF clauses directly correspond to the DNF used by the Prolog version of leanCoP. Since no steps distinguish between the positivity of literals, the two can be used interchangeably in the proof procedure. We start the FOL algorithm by trying to prove ¬∀.

Since the final leanCoP proof may include references to lemmas, the reconstruction cannot be performed the same way as it is done in MESON. There, a tree structure is used for finished proofs. Each subgoal either closes the branch (the literal is a negation of a literal already present on the path) or is a branch extension with
We define a type of proof steps, one for each proof step in the HOL theorem as follows:

```ocaml
let rec mexpand rules ancestors g cont (env,n,k) = 
  if n < 0 then failwith "too deep" else 
  try tryfind (fun a -> cont (unify_literals 
    env (g,negate a),n,k)) ancestors 
  with Failure -> tryfind (fun rule -> 
    let (asm,c),k = rename_rule k rule in 
    itlist (mexpand rules (g:ancestors)) asm cont 
    (unify_literals env (g,c),n - length asm, k )) rules;;

let puremeson fm = 
  let cls = simpf(specialize(pnf fm)) in 
  let rules = itlist ((@) ** contrapositives) cls [] in 
  deepen (fun n -> mexpand rules [] False (fun x -> x) 
    (undefined,n,0); n) 0;;
```

The top-level `puremeson` function proceeds by turning the input formula into a clausal form, making contrapositives (rules) from the clauses, and then repeatedly calling the `mexpand` function with these rules using iterative deepening over the number of nodes permitted in the proof tree.

The `mexpand` function takes as its arguments the rules, an (initially empty) list of goal ancestors, the goal `g` to prove (initially `False`, which was also added to all-negative clauses when creating contrapositives), a continuation function `cont` for solving the rest of the subgoals (initially the identity), and a tuple consisting of the current trail `env`, the number `n` of additional nodes in the proof tree permitted, and a counter `k` for variable renaming.

If the allowed node count is not negative, `mexpand` first tries to unify the current goal with a negated ancestor, followed by calling the current continuation (trying to solve the remaining goals) with the extended trail. If all such unification/continuation attempts fail (i.e., they throw `Failure`), an extension step is tried with all rules. This means that the head of a (renamed) rule is unified with the goal `g`, the goal is appended to the ancestors and the `mexpand` is called (again using list folding with the subsequently modified trail and continuation) for all the assumptions of the rule, decreasing the allowed node count for the recursive calls.

The full HOL Light version of MESON additionally uses a smarter (divide-and-conquer) policy for the size limit, checks the goal for being already among the ancestors, caches continuations, and uses simple indexing. Below we enumerate some of the most important differences between the leanCoP algorithm and MESON and their implementations in HOL Light. Their practical effect is measured in Section[4]

- **leanCoP** computes and uses lemmas. The literals that correspond to closed branches are stored in a list. Each call to the main `prove` function additionally looks for the first literal in the list of lemmas. This can cost a linear number of equality checks if no parts of the proof are reused, but it saves computations if there are repetitions.

- Both algorithms use iterative deepening; however the depth and termination conditions are computed differently.

- MESON is implemented in the continuation-passing-style, so it can use as an additional optimization caching of the continuations. If any continuations are repeated (at the same depth level), the subproof is not retried. Otten’s leanCoP uses a direct Prolog implementation which cannot (without further tricks) do such repetition elimination. The implementation of leanCoP in OCaml behaves the same.

- **leanCoP** may use the cut after the lemma step, path step or successful branch closing in the extension step. Implementing this behaviour in OCaml exactly requires multiple `cut` exceptions – one for each depth of the proof.

- The checking for repetitions is done in a coarser way in MESON than in leanCoP, allowing leanCoP to skip some work done by MESON.

- The search is started differently in leanCoP and in MESON. leanCoP starts with a conjecture clause, which likely contributes to its relatively good performance on larger problems.

4. Experimental Setup and Results

For the experiments we use HOL Light SVN version 199 (September 2014), Metis 2.3, and leanCoP 2.1. Unless noted otherwise, the systems are run on a 48-core server with AMD Opteron 6174 2.2 GHz CPUs, 320 GB RAM, and 0.5 MB L2 cache per CPU. Each problem is always assigned one CPU.

The systems are compared on several benchmarks, corresponding to different modes of use: goals coming from HOL Light itself,
the general set of problems from the TPTP library, and the large-
theory problems extracted from Mizar [28]. The first set of prob-
lems is important, however since these problems come from HOL
Light itself, they are likely naturally biased towards MESON. The
Mizar problems come from the two MPTP-based benchmarks: the
MPTP Challenge and MPTP2078 [1]. These are large-theory prob-
lems coming from a different ITP, hence they do not introduce the
implicit bias as the HOL Light problems, while coming from a
more relevant application domain than the general TPTP problems.

For HOL Light, we evaluate (with 5 second time limit) on two sets of problems. First, we look at 872 MESON-solved HOL
Light goals that were made harder by removing splitting. In this
scenario the tactic is applied to a subgoal of a proof, which is
a bit similar to the Judgement-Day [3] evaluation used for
Isabelle/Sledgehammer, where the goals are however restricted to
the solvable ones. Table 1 shows the results. Second, we eval-
uate on the top-level goals (with their dependencies already minim-
ized) that have been solved with the HOL4Hammer system [1], i.e.,
by using the strongest available ATPs. This set is important because
tactics such as MESON, Metis and now also leanCoP can be tried
as a first cheap method for reconstructing the proofs found by the
stronger ATPs. The results are shown in Table 2. In both cases,
the OCaml implementation of leanCoP performs best, improving the
MESON’s performance in the first case by about 11%, and improv-
ing on Metis on the second set of problems by about 45%.

Table 3 shows the results of the evaluation on all 7036 FOF
problems coming from TPTP 6.0.0, using 10 second time limit.
Here the difference to Metis is not so significant, probably be-
cause Metis implements ordered paramodulation, which is useful
for many TPTP problems containing equality. The improvement
over MESON is about 17%. Table 4 and Table 5 show the results
on the small (heuristically minimized) and large MPTP Challenge
problems. The best version of the OCaml implementation of lean-
CoP improves by 54% on Metis and by 90% on MESON on the
small problems, and by 88% on Metis and 100% on MESON on the
large problems. Here the goal directedness of leanCoP is prob-
ably the main factor.

Finally, to get a comparison also with the best ATPs on a larger
ITP-oriented benchmark (using different hardware), we have done
a 10s evaluation of several systems on the newer MPTP2078 bench-
mark (used in the 2012 CASC@Turing competition), see Table 6
and Table 7. The difference to Metis and MESON on the small
problems is still quite significant (40% improvement over ME-
SON), while on the large problems the goal-directedness again shows even more (about 90% improvement). While Vampire’s
(version 2.6) SnE heuristic [6] helps a lot on the larger prob-
lems [29], the difference there between E (1.8) and our version of
leanCoP is not so great as one could imagine giving the several or-
ders of magnitude difference in the size of their implementations.

<table>
<thead>
<tr>
<th>Prover</th>
<th>Theorem (%)</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>m lleancop-cut-comp</td>
<td>759 (87.04)</td>
<td>2</td>
</tr>
<tr>
<td>m lleancop-nocut</td>
<td>759 (87.04)</td>
<td>2</td>
</tr>
<tr>
<td>p lleancop-cut</td>
<td>752 (86.23)</td>
<td>0</td>
</tr>
<tr>
<td>p lleancop-nocut</td>
<td>751 (86.12)</td>
<td>0</td>
</tr>
<tr>
<td>mesit-23</td>
<td>708 (81.19)</td>
<td>26</td>
</tr>
<tr>
<td>meson</td>
<td>683 (78.32)</td>
<td>4</td>
</tr>
<tr>
<td>any</td>
<td>832 (95.41)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Core HOL Light MESON calls without splitting (872 goals), 5sec per goal

<table>
<thead>
<tr>
<th>Prover</th>
<th>Theorem (%)</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>m lleancop-cut-comp</td>
<td>1178 (75.70)</td>
<td>12</td>
</tr>
<tr>
<td>m lleancop-nocut</td>
<td>1162 (74.67)</td>
<td>0</td>
</tr>
<tr>
<td>meson</td>
<td>1110 (71.33)</td>
<td>39</td>
</tr>
<tr>
<td>p lleancop-nocut</td>
<td>1085 (69.73)</td>
<td>0</td>
</tr>
<tr>
<td>p lleancop-cut</td>
<td>1084 (69.66)</td>
<td>0</td>
</tr>
<tr>
<td>mesit-23</td>
<td>814 (52.31)</td>
<td>16</td>
</tr>
<tr>
<td>any</td>
<td>1260 (80.97)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. HOL Light dependencies (1556 goals, 5sec)

<table>
<thead>
<tr>
<th>Prover</th>
<th>Theorem (%)</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>plleancop-cut-conj</td>
<td>1669 (23.72)</td>
<td>73</td>
</tr>
<tr>
<td>plleancop-cut-conj</td>
<td>1648 (23.42)</td>
<td>21</td>
</tr>
<tr>
<td>plleancop-cut</td>
<td>1622 (23.05)</td>
<td>34</td>
</tr>
<tr>
<td>plleancop-cut-conj</td>
<td>1571 (22.32)</td>
<td>9</td>
</tr>
<tr>
<td>mesit-23</td>
<td>1562 (22.20)</td>
<td>261</td>
</tr>
<tr>
<td>meson</td>
<td>1430 (20.32)</td>
<td>28</td>
</tr>
<tr>
<td>plleancop-nocut</td>
<td>1358 (19.30)</td>
<td>25</td>
</tr>
<tr>
<td>m lleancop-nocut</td>
<td>1158 (16.45)</td>
<td>3</td>
</tr>
<tr>
<td>any</td>
<td>2433 (34.57)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. TPTP (7036 goals with at least one conjecture, 10sec)

<table>
<thead>
<tr>
<th>Prover</th>
<th>Theorem (%)</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>plleancop-cut-conj</td>
<td>103 (40.87302)</td>
<td>2</td>
</tr>
<tr>
<td>plleancop-cut</td>
<td>99 (39.28571)</td>
<td>8</td>
</tr>
<tr>
<td>m lleancop-cut-conj</td>
<td>91 (36.11111)</td>
<td>2</td>
</tr>
<tr>
<td>m lleancop-cut</td>
<td>79 (31.34921)</td>
<td>0</td>
</tr>
<tr>
<td>m lleancop-nocut</td>
<td>76 (30.15873)</td>
<td>0</td>
</tr>
<tr>
<td>plleancop-nocut</td>
<td>62 (24.60317)</td>
<td>1</td>
</tr>
<tr>
<td>mesit-23</td>
<td>59 (23.41270)</td>
<td>3</td>
</tr>
<tr>
<td>meson-infer</td>
<td>48 (19.04762)</td>
<td>0</td>
</tr>
<tr>
<td>any</td>
<td>124 (49.20635)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Bushy (small) MPTP Challenge problems (252 in total), 10sec)

<table>
<thead>
<tr>
<th>Prover</th>
<th>Theorem (%)</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>plleancop-cut-conj</td>
<td>61 (24.20635)</td>
<td>5</td>
</tr>
<tr>
<td>plleancop-cut-conj</td>
<td>60 (23.80952)</td>
<td>9</td>
</tr>
<tr>
<td>plleancop-cut</td>
<td>57 (22.61905)</td>
<td>4</td>
</tr>
<tr>
<td>m lleancop-nocut</td>
<td>47 (18.65079)</td>
<td>0</td>
</tr>
<tr>
<td>m lleancop-cut</td>
<td>47 (18.65079)</td>
<td>0</td>
</tr>
<tr>
<td>mesit-23</td>
<td>32 (12.69841)</td>
<td>3</td>
</tr>
<tr>
<td>meson-infer</td>
<td>30 (11.90476)</td>
<td>0</td>
</tr>
<tr>
<td>plleancop-nocut</td>
<td>26 (10.31746)</td>
<td>0</td>
</tr>
<tr>
<td>any</td>
<td>83 (32.93651)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Chainy (large) MPTP Challenge problems (252 in total), 10sec)
**Prover** | **Theorem (%)**  
---|---  
Vampire | 1198 (57.65)  
e18 | 1022 (49.18)  
mileanCoP-cut-conj | 613 (29.49)  
pllean-cut-conj | 597 (28.72)  
metis-23 | 564 (27.14)  
mileanCoP-cut | 559 (26.90)  
pllean-cut | 544 (26.17)  
pllean-comp7 | 539 (25.93)  
mileanCoP-nocut | 521 (25.07)  
pllean-nc | 454 (21.84)  
meson-infer | 438 (21.07)  
any | 1277 (61.45)  

Table 6. Bushy (small) MPTP2078 problems (2078 in total), 10sec

**Prover** | **Theorem (%)**  
---|---  
Vampire | 634 (30.51)  
e18 | 317 (15.25)  
mileanCoP-cut-conj | 243 (11.69)  
pllean-cut-conj | 196 (9.43)  
pllean-cut | 170 (8.18)  
pllean-comp7 | 159 (7.65)  
mileanCoP-nocut | 150 (7.21)  
mileanCoP-cut | 146 (7.02)  
meson-infer | 145 (6.97)  
metis-23 | 138 (6.64)  
pllean-nc | 126 (6.06)  
any | 693 (33.34)  

Table 7. Chainy (large) MPTP2078 problems (2078 in total), 10sec

## 5. Conclusion

We have implemented an OCaml version of the leanCoP compact connection prover, and the reconstruction of its proofs inside HOL Light. This proof-reconstruction functionality can be also used to certify in HOL Light an arbitrary TPTP proof produced by leanCoP, thus turning leanCoP into one of the few ATPs whose proofs enjoy LCF-style verification in one of the safest LCF-based systems. The performance of the OCaml version on the benchmarks is comparable to the Prolog version, while it always outperforms Metis and MESON, sometimes very significantly on the relevant ITP-related benchmarks.

We provide a HOL Light interface that is identical to the one offered by MESON, namely we provide two tactics and a rule. LEANCOP_TAC and ASM_LEANCOP_TAC are a list of helpert heorems, and then try to solve the given goal together (or the given goal and assumptions, respectively). The LEANCOP rule, given a list of helper theorems acts as a conversion, i.e., given a term statement it tries to prove a theorem whose conclusion is identical to that of the term. The benchmarks show that these are likely the strongest single-step proof-reconstructing first-order tactics available today in any ITP system.

## Acknowledgments

Supported by the Austrian Science Fund (FWF): P26201.


