Algorithm 1 Top-level Design of LS-Sampler.

Input:

F: SMT formula;

k: the number of samples;

Output:

A: the set of generated, valid samples;

$F_1 \leftarrow$ Simplify F;
$F_2 \leftarrow$ Boolean skeleton of F_1;
$A \leftarrow \{\}$;
Initialize(F_1, F_2)
for $iter \leftarrow 1$ to k **do**
  $\alpha \leftarrow$ pickSample$(F_1, prob, A)$;
  $A \leftarrow A \cup \alpha$;
  for each $b_i \in B(F_2)$ **do**
    prob$(b_i) \leftarrow$ updateProb(x_i, A);
  end for
end for
return A;

Algorithm 2 The Initialize Component.

Input:

F_1: Simplified SMT formula;
F_2: Boolean skeleton of F_1;

Output:

prob: vector of all skeleton variables' sampling probabilities;

for each integer variable $v_i \in V(F_1)$ **do**
  if $v_i \geq lb$ and $v_i \leq ub$ **then**
    v_i is assigned with a random value in $[lb, ub]$;
  else if $v_i \geq lb$ or $v_i \leq ub$ **then**
    v_i is assigned with lb or ub;
  else if v_i is unbounded **then**
    v_i is assigned with 0;
  end if
end for
for each $b_i \in B(F_2)$ **do**
  set prob(b_i) as 0.5;
end for

Algorithm 3 The PickSample Component.

Input:

F_1: Simplified SMT formula;
F_2: Boolean skeleton of F_1;
prob: vector of all skeleton variables' sampling probabilities;

Output:

α^*: the selected, valid sample;

```
C ← {};
for j ← 1 to λ do
    α_j ← generate a sample by sampling the value of each skeleton variable b_i according to prob(b_i);
    α*_j ← DAHS(F_1, F_2, α_j, prob);
    C ← α*_j;
    Calculate score(α*_j) for the valid sample α*_j;
end for
α* ← the valid sample in C with the largest score;
return α*
```

Algorithm 4 The UpdateProb Component.

Input:

b_i: skeleton variable in F_2;

A: The current set of valid samples;

Output:

$prob(b_i)$: Updated sampling probability for x_i;

$B \leftarrow$ the set of all b_i-false samples in A;

$prob(b_i) \leftarrow |B|/|A|$;

return $prob(b_i)$

Algorithm 5 The Diversity-aware Heuristic Search Component.

Input:

F_1: simplified SMT formula;
F_2: Boolean skeleton of F_1;
α: the input sample, possibly invalid;
$prob$: vector of all skeleton variables' sampling probabilities;

Output:

α^*: the modified, valid sample;

$\alpha^* \leftarrow \alpha$;
while $\alpha^* isinvalid$ **do**
  if $DV \neq$ **then**
    $x_i \leftarrow$ the variable with the largest $gain$ in DV, breaking ties by preferring the variable with the largest priority.
  else
    $c \leftarrow$ a random unsatisfied clause;
    $x_i \leftarrow$ the variable with the largest priority in c;
  end if
  if x_i is pure Boolean variable **then**
    $\alpha^* \leftarrow \alpha$ with x_i flipped;
  else if x_i is LIA atom **then**
    $a \leftarrow$ the theoretical atom corresponding to x_i;
    $op \leftarrow$ a $cm(v, a)$ operation with the greatest $cscore$;
  end if
end while
return α^*;

> $gain(x_i)$: the decrement in the number of unsitisfied clauses by flipping the current truth value of x_i.

> Decreasing variable: x_i is decreasing variable if $gain(x_i) > 0$.

> DV: the current set of all decreasing variables.

> $priority(x_i)$: $(1 - v_i) \times prob(x_i) + v_i \times (1 - prob(x_i))$