MODULE *BinarySearch*

This module defines a binary search algorithm for finding an item in a sorted sequence, and contains a *TLAPS*-checked proof of its safety property. We assume a sorted sequence *seq* with elements in some set Values of integers and a number *val* in *Values*, it sets the value *result* to either a number i with $seq[i] = val$, or to 0 if there is no such i.

It is surprisingly difficult to get such a binary search algorithm correct without making errors that have to be caught by debugging. I suggest trying to write a correct *PlusCal* binary search algorithm yourself before looking at this one.

This algorithm is one of the examples in Section 7.3 of "Proving Safety Properties", which is at

http:=*lamport:azurewebsites:net*/tla/proving-*safety:pdf*

EXTENDS *Integers*, *Sequences*, *TLAPS*

CONSTANT *Values*

ASSUME *ValAssump* $\triangleq$ *Values* $\subseteq$ *Int*

$$SortedSeqs \triangleq \{ss \in Seq(Values) : \forall i, j \in 1 \mathrel{..} Len(ss) : (i < j) \Rightarrow (ss[i] \leq ss[j])\}$$

****

--fair algorithm *BinarySearch* *f*
variables $seq \in SortedSeqs$, $val \in Values$,
$low = 1$, $high = Len(seq)$, $result = 0$ **;**
f *a*: **while** ($low \leq high \land result = 0$) *f*
with ($mid = (low + high) \div 2$, $mval = seq[mid]$) *f*
if ($mval = val$) *f* $result := mid$ *g*
else if ($val < mval$) *f* $high := mid - 1$ *g*
else *f* $low := mid + 1$ *g* *g* *g* *g* *g*

****

BEGIN TRANSLATION

VARIABLES *seq*, *val*, *low*, *high*, *result*, *pc*

$vars \triangleq \langle seq, val, low, high, result, pc\rangle$

Init $\triangleq$ Global variables
$\land\ seq \in SortedSeqs$
$\land\ val \in Values$
$\land\ low = 1$
$\land\ high = Len(seq)$
$\land\ result = 0$
$\land\ pc =$ "a"

a $\triangleq$ $\land\ pc =$ "a"
$\land$ IF $low \leq high \land result = 0$
THEN $\land$ LET $mid \triangleq (low + high) \div 2$ IN
LET $mval \triangleq seq[mid]$ IN

$$
\begin{array}{l}
\quad\quad\quad\quad\quad\quad\quad\quad \text{IF } mval = val \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{THEN } \land\ result' = mid \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \land\ \text{UNCHANGED } \langle low, high \rangle \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{ELSE } \land\ \text{IF } val < mval \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{THEN } \land\ high' = mid - 1 \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\ \ \land\ low' = low \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{ELSE } \land\ low' = mid + 1 \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\ \ \land\ high' = high \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad \land\ \text{UNCHANGED } result \\
\quad\quad\quad\quad\quad\quad \land\ pc' = \text{"a"} \\
\quad\quad\quad\quad \text{ELSE } \land\ pc' = \text{"Done"} \\
\quad\quad\quad\quad\quad\quad\quad \land\ \text{UNCHANGED } \langle low, high, result \rangle \\
\quad \land\ \text{UNCHANGED } \langle seq, val \rangle
\end{array}
$$

$$
\begin{array}{l}
Next \triangleq a \\
\quad\quad\quad \lor \ \text{Disjunct to prevent deadlock on termination} \\
\quad\quad\quad\quad (pc = \text{"Done"} \land \text{UNCHANGED } vars)
\end{array}
$$

$$
\begin{array}{l}
Spec \triangleq \land\ Init \land \Box[Next]_{vars} \\
\quad\quad\quad\ \ \land\ \mathrm{WF}_{vars}(Next)
\end{array}
$$

$$
Termination \triangleq \Diamond(pc = \text{"Done"})
$$

END TRANSLATION

Partial correctness of the algorithm is expressed by invariance of formula *resultCorrect*. To get *TLC* to check this property, we use a model that overrides the definition of *Seq* so $Seq(S)$ is the set of sequences of elements of S having at most some small length. For example,

$$
Seq(S) \triangleq \text{UNION } \{[1 \,..\, i \to S] : i \in 0 \,..\, 3\}
$$

is the set of such sequences with length at most 3.

$$
\begin{array}{l}
resultCorrect \triangleq \\
\quad (pc = \text{"Done"}) \Rightarrow \text{IF } \exists\, i \in 1 \,..\, Len(seq) : seq[i] = val \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{THEN } seq[result] = val \\
\quad\quad\quad\quad\quad\quad\quad\quad\quad \text{ELSE } \ result = 0
\end{array}
$$

Proving the invariance of *resultCorrect* requires finding an inductive invariant that implies it. A suitable inductive invariant *Inv* is defined here. You can use *TLC* to check that *Inv* is an inductive invariant.

$$
\begin{array}{l}
TypeOK \triangleq \land\ seq \in SortedSeqs \\
\quad\quad\quad\quad\quad \land\ val \in Values \\
\quad\quad\quad\quad\quad \land\ low \in 1 \,..\, (Len(seq) + 1) \\
\quad\quad\quad\quad\quad \land\ high \in 0 \,..\, Len(seq) \\
\quad\quad\quad\quad\quad \land\ result \in 0 \,..\, Len(seq) \\
\quad\quad\quad\quad\quad \land\ pc \in \{\text{"a"}, \text{"Done"}\}
\end{array}
$$

$$
\begin{array}{l}
Inv \triangleq \land\ TypeOK \\
\quad\quad\quad \land\ (result \neq 0) \Rightarrow (Len(seq) > 0) \land (seq[result] = val)
\end{array}
$$

$\land\ (pc = \text{"a"}) \Rightarrow$
$\quad$ IF $\exists\, i \in 1 \,..\, Len(seq) : seq[i] = val$
$\quad\quad$ THEN $\exists\, i \in low \,..\, high : seq[i] = val$
$\quad\quad$ ELSE $result = 0$
$\land\ (pc = \text{"Done"}) \Rightarrow (result \neq 0) \lor (\forall\, i \in 1 \,..\, Len(seq) : seq[i] \neq val)$

Here is the invariance proof.

THEOREM $Spec \Rightarrow \Box resultCorrect$
$\langle 1 \rangle 1.\ Init \Rightarrow Inv$
$\quad$ BY DEF $Init, Inv, TypeOK$
$\langle 1 \rangle 2.\ Inv \land [Next]_{vars} \Rightarrow Inv'$
$\quad \langle 2 \rangle$ SUFFICES ASSUME $Inv,$
$\quad\quad\quad\quad\quad\quad\quad\quad\quad [Next]_{vars}$
$\quad\quad\quad\quad\quad\quad$ PROVE Inv'
$\quad\quad$ OBVIOUS
$\quad \langle 2 \rangle 1.$ CASE a
$\quad\quad \langle 3 \rangle 1.$ CASE $low \leq high \land result = 0$
$\quad\quad\quad \langle 4 \rangle$ DEFINE $mid \triangleq (low + high) \div 2$
$\quad\quad\quad\quad\quad\quad\quad mval \triangleq seq[mid]$
$\quad\quad\quad \langle 4 \rangle\ (low \leq mid) \land (mid \leq high) \land (mid \in 1 \,..\, Len(seq))$
$\quad\quad\quad\quad$ BY $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK$
$\quad\quad\quad \langle 4 \rangle 1.\ TypeOK'$
$\quad\quad\quad\quad \langle 5 \rangle 1.\ seq' \in SortedSeqs$
$\quad\quad\quad\quad\quad$ BY $\langle 2 \rangle 1$ DEF $a, Inv, TypeOK$
$\quad\quad\quad\quad \langle 5 \rangle 2.\ val' \in Values$
$\quad\quad\quad\quad\quad$ BY $\langle 2 \rangle 1$ DEF $a, Inv, TypeOK$
$\quad\quad\quad\quad \langle 5 \rangle 3.\ (low \in 1 \,..\, (Len(seq) + 1))'$
$\quad\quad\quad\quad\quad \langle 6 \rangle 1.$ CASE $seq[mid] = val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 1$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$
$\quad\quad\quad\quad\quad \langle 6 \rangle 2.$ CASE $seq[mid] \neq val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 2$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$
$\quad\quad\quad\quad\quad \langle 6 \rangle 3.$ QED
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 1$, $\langle 6 \rangle 2$
$\quad\quad\quad\quad \langle 5 \rangle 4.\ (high \in 0 \,..\, Len(seq))'$
$\quad\quad\quad\quad\quad \langle 6 \rangle 1.$ CASE $seq[mid] = val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 1$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$
$\quad\quad\quad\quad\quad \langle 6 \rangle 2.$ CASE $seq[mid] \neq val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 2$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$
$\quad\quad\quad\quad\quad \langle 6 \rangle 3.$ QED
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 1$, $\langle 6 \rangle 2$
$\quad\quad\quad\quad \langle 5 \rangle 5.\ (result \in 0 \,..\, Len(seq))'$
$\quad\quad\quad\quad\quad \langle 6 \rangle 1.$ CASE $seq[mid] = val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 1$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$
$\quad\quad\quad\quad\quad \langle 6 \rangle 2.$ CASE $seq[mid] \neq val$
$\quad\quad\quad\quad\quad\quad$ BY $\langle 6 \rangle 2$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$, $Z3$ DEF $Inv, TypeOK, a$

$\langle 6 \rangle 3.$ QED
BY $\langle 6 \rangle 1$, $\langle 6 \rangle 2$
$\langle 5 \rangle 6.$ $(pc \in \{\text{“a”}, \text{“Done”}\})'$
BY $\langle 2 \rangle 1$, $\langle 3 \rangle 1$ DEF Inv, $TypeOK$, a
$\langle 5 \rangle 7.$ QED
BY $\langle 5 \rangle 1$, $\langle 5 \rangle 2$, $\langle 5 \rangle 3$, $\langle 5 \rangle 4$, $\langle 5 \rangle 5$, $\langle 5 \rangle 6$ DEF $TypeOK$
$\langle 4 \rangle 2.$ $((result \neq 0) \Rightarrow (Len(seq) > 0) \land (seq[result] = val))'$
$\langle 5 \rangle 1.$CASE $seq[mid] = val$
BY $\langle 5 \rangle 1$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$ DEF Inv, $TypeOK$, a
$\langle 5 \rangle 2.$CASE $seq[mid] \neq val$
BY $\langle 5 \rangle 2$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$ DEF Inv, $TypeOK$, a
$\langle 5 \rangle 3.$ QED
BY $\langle 5 \rangle 1$, $\langle 5 \rangle 2$
$\langle 4 \rangle 3.$ $((pc = \text{“a”}) \Rightarrow$
IF $\exists i \in 1 \,..\, Len(seq) : seq[i] = val$
THEN $\exists i \in low \,..\, high : seq[i] = val$
ELSE $result = 0)'$
$\langle 5 \rangle 1.$CASE $seq[mid] = val$
BY $\langle 5 \rangle 1$, $\langle 2 \rangle 1$, $\langle 3 \rangle 1$ DEF Inv, $TypeOK$, a
$\langle 5 \rangle 2.$CASE $seq[mid] \neq val$
$\langle 6 \rangle 1.$ $\land (low \leq mid) \land (mid \leq high) \land (mid \in 1 \,..\, Len(seq))$
$\land Len(seq) > 0 \land Len(seq) \in Nat$
$\land low \in 1 \,..\, Len(seq)$
$\land high \in 1 \,..\, Len(seq)$
BY $ValAssump$ DEF Inv, $TypeOK$
$\langle 6 \rangle 2.$CASE $\exists i \in 1 \,..\, Len(seq) : seq[i] = val$
$\langle 7 \rangle 1.$ PICK $i \in low \,..\, high : seq[i] = val$
BY $\langle 6 \rangle 2$, $\langle 2 \rangle 1$ DEF a, Inv
$\langle 7 \rangle 2.$ $\land (low \leq mid) \land (mid \leq high) \land (mid \in 1 \,..\, Len(seq))$
$\land Len(seq) > 0 \land Len(seq) \in Nat$
$\land low \in 1 \,..\, Len(seq)$
$\land high \in 1 \,..\, Len(seq)$
$\land seq[i] = val$
BY $ValAssump$, $\langle 6 \rangle 2$, $\langle 7 \rangle 1$ DEF Inv, $TypeOK$
$\langle 7 \rangle 3.$ $\forall j \in 1 \,..\, Len(seq) : seq[j] \in Int$
$\langle 8 \rangle 1.$ $seq \in Seq(Values)$
BY DEF Inv, $TypeOK$, $SortedSeqs$
$\langle 8 \rangle 2.$ $seq \in Seq(Int)$
BY $\langle 8 \rangle 1$, $ValAssump$
$\langle 8 \rangle 3.$ QED
BY $\langle 8 \rangle 2$ DEF Inv, $TypeOK$, $SortedSeqs$
$\langle 7 \rangle 4.$ $\forall j, k \in 1 \,..\, Len(seq) : j < k \Rightarrow seq[j] \leq seq[k]$
BY DEF Inv, $TypeOK$, $SortedSeqs$
$\langle 7 \rangle 5.$CASE $val < seq[mid]$
$\langle 8 \rangle 1.$ $seq[i] < seq[mid]$

BY $\langle 7\rangle 2$, $\langle 7\rangle 5$, $\langle 8\rangle 5$

$\langle 8\rangle 2$. $i < mid$

BY $ValAssump$, $\langle 7\rangle 2$, $\langle 8\rangle 1$, $\langle 7\rangle 4$, $\langle 7\rangle 3$, $Z3$

$\langle 8\rangle 3$. $i \in low \,..\, mid - 1$

BY ONLY $\langle 7\rangle 2$, $\langle 8\rangle 1$, $\langle 8\rangle 2$, $Z3$

$\langle 8\rangle 4$. $\land\, (pc' = \text{"a"}) \land (low' = low) \land (high' = mid - 1)$
$\land\, \exists\, j \in 1 \,..\, Len(seq) : seq[j] = val$

BY $\langle 2\rangle 1$, $\langle 3\rangle 1$, $\langle 5\rangle 2$, $\langle 6\rangle 2$, $\langle 7\rangle 5$ DEF a, mid

$\langle 8\rangle 5$. QED

BY ONLY $\langle 7\rangle 2$, $\langle 8\rangle 4$, $\langle 8\rangle 3$, $\langle 8\rangle 5$

$\langle 7\rangle 6$.CASE $\neg(val < seq[mid])$

$\langle 8\rangle$ HIDE DEF mid

$\langle 8\rangle 1$. $seq[mid] < seq[i]$

BY $ValAssump$, $\langle 7\rangle 2$, $\langle 7\rangle 6$, $\langle 5\rangle 2$, $\langle 7\rangle 3$, $Z3$

$\langle 8\rangle 2$. $mid < i$

BY $ValAssump$, $\langle 7\rangle 2$, $\langle 8\rangle 1$, $\langle 8\rangle$a, $\langle 9\rangle 1$,$\langle 7\rangle 3$, $\langle 7\rangle 4$, $Z3$

$\langle 8\rangle 3$. $i \in mid + 1 \,..\, high$

BY $\langle 7\rangle 2$, $\langle 8\rangle 1$, $\langle 8\rangle 2$, $Z3$

$\langle 8\rangle 4$. $\land\, (pc' = \text{"a"}) \land (low' = mid + 1) \land (high' = high)$
$\land\, \exists\, j \in 1 \,..\, Len(seq) : seq[j] = val$

BY $\langle 2\rangle 1$, $\langle 3\rangle 1$, $\langle 5\rangle 2$, $\langle 6\rangle 2$, $\langle 7\rangle 6$ DEF a, mid

$\langle 8\rangle 5$. QED

BY ONLY $\langle 7\rangle 2$, $\langle 8\rangle 4$, $\langle 8\rangle 3$, $\langle 8\rangle 5$

$\langle 7\rangle 7$. QED

BY $\langle 7\rangle 5$, $\langle 7\rangle 6$

$\langle 6\rangle 3$.CASE $\neg\exists\, i \in 1 \,..\, Len(seq) : seq[i] = val$

BY $\langle 6\rangle 3$, $\langle 5\rangle 2$, $\langle 2\rangle 1$, $\langle 3\rangle 1$ DEF Inv, $TypeOK$, a

$\langle 6\rangle 4$. QED

BY $\langle 6\rangle 2$, $\langle 6\rangle 3$

$\langle 5\rangle 3$. QED

BY $\langle 5\rangle 1$, $\langle 5\rangle 2$

$\langle 4\rangle 4$. $((pc = \text{"Done"}) \Rightarrow (result \neq 0) \lor (\forall\, i \in 1 \,..\, Len(seq) : seq[i] \neq val))'$

BY $\langle 3\rangle 1$, $\langle 2\rangle 1$ DEF Inv, $TypeOK$, a

$\langle 4\rangle 5$. QED

BY $\langle 4\rangle 1$, $\langle 4\rangle 2$, $\langle 4\rangle 3$, $\langle 4\rangle 4$ DEF Inv

$\langle 3\rangle 2$.CASE $\neg(low \leq high \land result = 0)$

BY $\langle 3\rangle 2$, $\langle 2\rangle 1$ DEF Inv, $TypeOK$, a

$\langle 3\rangle 3$. QED

BY $\langle 3\rangle 1$, $\langle 3\rangle 2$

$\langle 2\rangle 2$.CASE UNCHANGED $vars$

BY $\langle 2\rangle 2$ DEF Inv, $TypeOK$, $vars$

$\langle 2\rangle 3$. QED

BY $\langle 2\rangle 1$, $\langle 2\rangle 2$ DEF $Next$

$\langle 1\rangle 3$. $Inv \Rightarrow resultCorrect$

BY DEF $resultCorrect$, Inv, $TypeOK$

$\langle 1 \rangle 4.$ QED

BY $\langle 1 \rangle 1$, $\langle 1 \rangle 2$, $\langle 1 \rangle 3$, *PTL* DEF *Spec*

* Modification History

* Last modified *Fri* May 03 16:28:58 *PDT* 2019 by *lamport*

* Created *Wed Apr* 17 15:15:12 *PDT* 2019 by *lamport*