Proof of Theorem 5 from [RST07]

July 6, 2007

In our setup there are \( K = 60k \) arrays and we choose \( 40k \) of them at random to group into pairs, leaving \( 20k \) untouched arrays.

Theorem 5 states the following:

**Theorem 5** Consider the following compound event:

- The proofs of value-consistency succeed for all \( 20k \) pairs of arrays \( T^{(i_s)} \) and \( T^{(j_s)} \), \( 1 \leq s \leq 20k \), that the Verifier obtains from the EP; but
- fewer than \( 18k \) of the remaining \( 20k \) arrays are pairwise value-consistent.

The probability of this event is at most \( 2^{-k} \). (Note that this probability is taken over the random choice of \( 40k \) subscripts as described above and over the randomized verifications of value consistency for each of the resulting \( 20k \) pairs.)

1 Proof of Theorem 5

Let us view the \( 60k \) translations as being grouped into equivalence classes, where all translations in a given equivalence class are pairwise value-consistent. We may think of one class of translations as being “white”, another class as being “blue,” and so on. Let us say that “white” is the color of the largest equivalence class.

To prove the theorem it suffices to establish the following: for any partition of the \( 60k \) translations into colors (where white is the color of the largest equivalence class), we have

\[
\Pr[(\text{more than } 2k \text{ of the untouched } 20k \text{ translations are not white}) \& \text{ (verification succeeds)}] \leq \frac{1}{2^k}.
\]

Let us refer to the compound event inside the \( \Pr[] \) above as event \( E \).

Given any partition of the \( 60k \) translations into colors (white and other colors), it is not difficult to see that the probability of event \( E \) only increases if we “collapse” all the non-white colors into one color, say black. (This collapsing does not change whether any given translations are non-white, and the collapsing only makes it easier for verification to succeed.) Let \( Bk \) denote the number of black translations among the \( 60k \). We want to show that for all possible values of \( B \) we have \( \Pr[E] \leq 1/2^k \).

**Claim 1** If \( B \leq 3 \) then \( \Pr[\text{more than } 2k \text{ of the untouched } 20k \text{ are black}] < 1/2^k \).

**Proof:** It is not difficult to see that if \( B \leq 3 \) then, assuming that more than \( 2k \) of the untouched \( 20k \) are black, the most likely outcome is that in fact exactly \( 2k \) of the untouched \( 20k \) are black. Thus we have

\[
\Pr[\text{more than } 2k \text{ of the untouched } 20k \text{ are black}] \leq k \cdot \Pr[\text{exactly } 2k \text{ of the untouched } 20k \text{ are black}].
\]

The probability on the RHS is

\[
\frac{\binom{Bk}{2k} \cdot \binom{(60-B)k}{18k}}{\binom{60k}{20k}}.
\]
Using $B \leq 3$ and the identity
\[
\left( \binom{n}{\alpha m} \right) = 2^{n \cdot (H(\alpha) \pm o(1))}
\] which holds for constant $0 < \alpha < 1/2$, (1)

it can be verified that the above fraction is at most $2^{-1.05 \cdot k}$. □

Claim 2 If $B \geq 30$ then $\Pr[\text{verification succeeds}] \leq 1/2^k$.

**Proof:** Suppose that $B \geq 30$. Recalling that white is the color of the largest equivalence class, we have that in the initial setup each array was value-consistent with at most $30k$ other arrays. Now consider the choice of the first $10k$ pairs of arrays (i.e. the choice of the first $20k$ many arrays selected from the $60k$). For each of these first $10k$ pairs, no matter which array was chosen as the first element of the pair and which arrays were chosen for the previously selected pairs, the probability that the second element of the pair is value-consistent with the first is at most $30k/40k = 3/4$. In other words, the probability that the second element of the pair is not value-consistent with the first is at least $1/4$, and hence the probability that this pair causes verification to fail is at least $1/8$ (since each pair that is not value-consistent is found out with probability at least $1/2$). Consequently, the probability that all of the first $10k$ pairs pass the verification is at most $(1 - 1/8)^{10k} \ll 1/2^k$. □

Thus we may assume that $3 \leq B \leq 30$. We now forget about the condition that more than $2k$ of the untouched $20k$ are not white, and just focus on upper bounding the probability that verification succeeds. We will show that this is at most $1/2^k$, and thus complete the proof.

Let $c \cdot k$ be the number of black arrays that are chosen among the $40k$ arrays that are chosen for verification (so $0 \leq c \leq B$, and we expect $c$ to be about $2B/3$ since $40k$ arrays are chosen for verification out of $60k$). Note that the probability that $ck$ are chosen among the $40k$ is
\[
\left( \binom{Bk}{ck} \cdot \binom{(60-B)k}{(40-c)k} \right) \left( \binom{40k}{60k} \right)
\]

We can think of the choice of $20k$ pairs for verification as being done in the following way: given the $40k$ arrays that will be used for verification, first randomly split the $40k$ arrays into two groups of $20k$ each, which we call the “left group” and the “right group.” Then for each group, assign the $20k$ members to “seats” $1, \ldots, 20k$ in a random order. The resulting $20k$ pairs are

(left group member who occupies seat 1, right group member who occupies seat 1)
(left group member who occupies seat 2, right group member who occupies seat 2)
\[\ldots\]
(left group member who occupies seat 20k, right group member who occupies seat 20k)

Suppose that exactly $ck/2$ of the black arrays are assigned to the left group and $ck/2$ are assigned to the black group (one can show that if this does not happen and there is an unbalanced allocation to the left and right groups, it only causes the probability that verification succeeds to decrease). We may assume w.l.o.g. that the $ck/2$ black arrays in the left group are assigned to seats $1, \ldots, ck/2$. Now for any value $\ell$, the probability that $\ell k$ of the $ck/2$ black arrays in the right group are assigned to seats $1, \ldots, ck/2$ is
\[
\left( \binom{ck/2}{\ell k} \cdot \binom{(20-c/2)k}{(c/2-\ell)k} \right) \left( \binom{20k}{ck/2} \right)
\]

If $\ell k$ of the $ck/2$ black arrays in the right group are assigned to seats $1, \ldots, ck/2$ then this means there are $(c - 2\ell)k$ seat indices which are such that exactly one of the two seats with that index (either left or right but not both) is black. This means that the probability that verification passes is at most $2^{-(c-2\ell)k}$, since each such seat passes with probability at most $1/2$.

So putting the pieces together, we have that the probability that

- $ck$ black arrays are chosen among the $40k$ arrays selected for verification; and
- there are $\ell k$ black-black pairs among the $20k$ pairs; and
verification passes

is at most

\[ X \cdot Y \cdot Z, \quad \text{where} \quad X(B, c) = \frac{(Bk) \cdot (20 - Bc)k}{(40 - Bc)k}, \quad Y(c, \ell) = \frac{(ck/2)^k \cdot (20 - c/2\ell)k}{(c/2 - \ell)k}, \quad Z(c, \ell) = \frac{1}{2^{(c-2\ell)k}}. \]  

Note that \( X(B, c), Y(c, \ell), \) and \( Z(c, \ell) \) are all probabilities hence are each at most 1. Recall that we have \( 3 \leq B \leq 30, \) \( 0 \leq c \leq B, \) and \( 0 \leq \ell \leq c/2. \) We will show that for each possible settings of \( B, c, \ell \) the product \( X \cdot Y \cdot Z \) is at most \( 1/2^{1.05k}. \) Then adding up over all (at most \( \text{poly}(k) \) many – recall that \( Bk, ck \) and \( \ell k \) are all integers) possible settings of \( B, c, \ell, \) we will have that the total probability that verification passes is at most \( 1/2^k. \)

We first note that if \( c \leq B/3 \) then, as can be verified using (1) and the fact that \( B \geq 3, \) the value of \( X \) is at most \( 2^{-1.05k}, \) and hence (2) is at most \( 2^{-1.05k}. \) So we can assume \( c \geq B/3; \) in particular since \( B \geq 3 \) this means \( c \geq 1. \)

Using (1) one can verify that for any value \( 1.2 \leq c \leq 30 \) and any value \( 0 \leq \ell \leq c/2, \) the value of \( Y(c, \ell) \cdot Z(c, \ell) \) is at most \( 1/2^{1.1k}. \) So it remains only to consider the range \( 1 \leq c \leq 1.2. \)

Using (1) again, one can verify that for any value \( 1 \leq c \leq 1.2 \) and any value \( 0 \leq \ell \leq c/2, \) the value of \( Y(c, \ell) \cdot Z(c, \ell) \) is at most \( 1/2^{-0.94k}. \) But it is also easy to check using (1) that for any value \( B \geq 3 \) and any \( 1 \leq c \leq 1.2, \) the value of \( X(B, c) \) is at most \( 1/2^{-0.6k}. \) Consequently \( X \cdot Y \cdot Z \) is indeed always at most \( 1/2^{1.05k}, \) and we are done. \( \square \)