This doc contains two parts, one is result, the other is 'insight/analytics'.

Result:

Old algorithm:

Best CDF 90% 2.5m. Based on running random six continuous points selection, median error 1.75m, max error 3m.

New one:

With same inner threshold setting, strict hallway point outer threshold 62.89% accuracy, hallway tolerance 1m 78.35%, 2m 83.24%, 3m 86.21%, 4m 86.86%. For the whole figure check threshold_res.xlsx.

Analytics:

Old algorithm:

It requires pre-calculated function and parameters stored. Provided that 'n' is the number of input access points, runtime complexity is O(n) + O(1), O(1) is for function matching, and then finally is O(n). If take pre-calculation into consideration, FFT(Fast Fourier transform) is O(klog(k)), DFT(Discrete Fourier Transform) is $O(k^2)$, where k is the input size and $n>k\geq 0$, so the final RC is hard to determine, highly relies on k. k is the number of access points that cover the target hallway/area, can be calculated by O(n), one pass.

This algorithm has cons. Since wifi signal strength does not change at every centimeter (not continuous).The best accuracy can be expected. And if the data collected has a smaller interval, say half, it does NOT improve results, actually worse. Because this may cause over-fitting. While double the interval, the results does not get worse nor the performance get improved.

New One:

Footprint method is easy to implement and understand and it has RC O(n). The biggest cons is that it does not learn from training data, in the other words, the final results highly depends on the quality of training data.

From the threshold_res we can tell that for certain, if we have larger tolerance on outer threshold, the better accuracy we get. But unless we change outer threshold to all the floor, otherwise we will not get 100% since we have error at threshold 0. While for certain outer threshold, if we keep increasing inner tolerance, the result may drop instead. Because there are too many 'right' outputs.

Using slide window as measurement, (window size 1 is just the original algorithm), with 2 for both inner and outer threshold, we get 81.95% at size 2, 83.45% at size 3, 85.23 at size 4 and 86.68% at size 5.

After analyzing where error occurs, by just listing all the error points, we can tell our algorithm has low accuracy in open area with multi hallways (the elevator area) and hallways with short distance (less than 6m in our map, can be 7 or 8 or other numbers, just need simulate real situation). Re collection/better training data will NOT solve this problem since wifi signal strength does not change much under those circumstances and we are not able to obtain a unique 'signature' for certain location.

Comparison:

Old one is has higher accuracy and there is no significant accuracy drop in/at certain area/condition, there is in new one, but it performs well in small area ONLY, since the coverage of every wifi access point is limited and the further the more easily it can be disrupted. While new one fits situation regardless small or large area. And the new one can't deal with outliers.