A More Elaborate Computation Model for Analyzing Balloon Hashing

Cui Hongrui, Sha Jinrui

November 25, 2017

1 An improved big-spread lemma

Room for improvement The *big spread lemma* plays an important role in the security proof of balloon hashing algorithm [1], however, the theorem require that the subset being considered, $V\prime$ have the same size of the pebbles available by the adversary. This is also reflected in Lemma 30, which bound the probability of the graph being too "dense". Recall in Lemma 30, for $\delta = 3$, every sandwich graph is an (m, m, n/16)-consecutively-avoiding graph, for $n_0 > 16$ and all $n_0 \leq m < n/6$ except with probability $P_{consec}(n, d, n_0) \leq 2 * 8 \cdot d \cdot n \cdot 2^{-n_0/2}$; And for $\delta = 7$, such stack of sandwich graph is (n/64, n/32)-everywhere-avoiding graph when $n > 2^{11}$, except with probability $P_{every} \leq 128 \cdot d \cdot 2^{-n/50}$. Note that in the later part of the lemma, the size of V' is defined as $m = \frac{n}{2^{\omega+1}}$, and thus m grows with n.

This limitation that the subset of V begin fixed by the number of pebbles of the adversary is somehow inconvenient, in that to make the probability of bad event small, one need the number of pebbles available grow, which is the contrary of natural instinct, as the more pebble the adversary acquires, the easier the attack (re-computation) should be. Although the result in [1] is absolutely correct, as we can view the growth of m as the result of the growth of n, an if the subset V' being considered too small, the variables sampled independently uniform at random from the last layer would have a bigger probability to being not so well-spread. Still, we want to separate the size of subset (m) and the number of pebbles ($|V'|$).

For the sake of clarity, and also because of the nature of an internal manuscript, I will replicate the content and the proof of big spread lemma.

Lemma 1 (Big Spread Lemma Variation)**.** *For all positive integer* $\delta \geq 3$*,* $\omega \geq 2$*,* n_0*, and* n*, and for all positive integers* m *such that* $2^\omega < m < 2^{2-\omega+\frac{2\omega}{\delta e}}$*, a list of* δm *elements sampled independently and uniformly at random from* $1, \ldots, n$ *is an* $(n_0,\ n/2^\omega)$*-well-spread set with probability at least* $1-2^{(1-\omega)\frac{\delta}{2}m} \cdot (2^{\omega m}+2^{\omega+1}2^{\omega n_0})$

Proof. The Strategy to bound the bad event that such property does not exist is the same as that used in [1]. Let $R = (R_1, \ldots, R_{\delta m})$ be integers sampled independently and uniformly at random from $1, \ldots, n$, we want to prove that for all subset $S \subseteq R$ of size at most n_0, $spread_S(R) \geq n/2^\omega$. To do so, we first define a bad event B, and then show that bounding Pr[B] is enough to prove the lemma, then bound Pr[B] to complete the lemma.

The Bad Event B Write the integers in R in non-decreasing order as $X_1, \ldots, X_{\delta m}$, then define $X_0 = 0$, $X_{\delta m+1} = n$. Let bad event B be the event that there exists a set $S' \subseteq X_1, \ldots, X_{\delta m+1}$ of size at most (n_0+1), such that $\sum_{X_i \in S'} X_i - X_{i-1} \geq (1-2^{-\omega})n$.

Whenever there exists a set $S \subseteq R$ of size at most n_0 that cause $spread_S(R) < n/2^\omega$, then bad event B must occur. Assuming that such a bad set S exists, construct a set $S' = S \bigcup X_{\delta m+1}$ of size at most $n_0 + 1$, Then we compute

$$\sum_{X_i \in S'} (X_i - X_{i-1}) = n - X_{\delta m} + \sum_{X_i \in S} (X_i - X_{i-1}) = n - \sum_{X_i \notin S} (X_i - X_{i-1}) \quad (1)$$

The last inequality holds because $X_{\delta m} = \sum_{X_i \in R} (X_i - X_{i-1})$. Now that $Spread_S(R) < n/2^\omega$, so we have bad event B occurs. Thus, $\Pr[Spread_S(R)] \leq \Pr[\text{B}]$, and therefore bounding a upper bound of bad event B is sufficient to prove the lemma.

Strategy to bound Pr[B] Let D be a random variable denoting the number of distinct integers in the list of random integers R. For any fixed integer $D^* \in \{1, \ldots, n\}$, we can write:

$$\begin{aligned} Pr[B] &= Pr[B|D < d^*] \cdot Pr[D < d^*] + Pr[B|D \geq d^*] \cdot Pr[D \geq d^*] && (2) \\ &\leq Pr[D < d^*] + Pr[B, D \geq d^*] && (3) \end{aligned}$$

In the following proof, we take $d^* = \delta m/2$.

Bounding $Pr[D < d^*]$ The probability that this event occurs is at most the probability when we throw δm balls into n bins, all the balls fall into a set of $\delta m/2$ bins.

$$Pr[D < d^*] \leq \binom{n}{d^*} (\frac{d^*}{n})^{\delta m} \leq (\frac{n \cdot e}{d^*})^{d^*} (\frac{d^*}{n})^{\delta m} = (\frac{d^*}{n})^{\delta m - d^*} e^{d^*} \quad (4)$$

By the hypothesis of the lemma, $m < 2^{(1-\omega+\frac{2\omega}{\delta})} \frac{2n}{\delta e}$. Therefore, we have

$$Pr[D < d^*] \leq (\frac{\delta m e}{2n})^{\delta m/2} \leq 2^{[(1-\omega)\frac{\delta}{2}+\omega]m} \quad (5)$$

Bounding $Pr[B, D \geq d^*]$ First I would like to re state the lemma 8 in the original work [1], because this lemma is useful in the proof of big spread lemma.

Lemma 2. *Let $(R_1, \ldots, R_d)$ be random variables respecting integers sampled uniformly, but without replacement, from $\{1, \ldots, n\}$. Write the Rs in ascending order as $(Y_1, \ldots, Y_d)$, and let $Y_0 = 0$, and $Y_{d+1} = n$. Next define $L = (L_1, \ldots, L_{d+1})$, where $L_i = Y_i - Y_{i-1}$. Then for all functions $f : \mathbb{Z}^{d+1} \to \{0, 1\}$, and for all permutations π on $d+1$ elements,*

$$Pr[f(L_1, \ldots, Ld+1) = 1] = Pr[f(L_{\pi(1)}, \ldots, L_{\pi(d+1)}) = 1]. \quad (6)$$

Particularly, this lemma shows that if we define a function that maps a list of segments (separated from $\{1, \ldots, n\}$ by the δm random integers) in to $\{0, 1\}$, such that if bad event B happens, then it is one. Using this lemma, we can define a permutation on those integers that if it appears in a bad set, then we shift it with the first n_0+1 integers in the list, and thus we can only focus on the first few elements on the list without changing the result. As lemma 2 require the list to be mutually distinct, we have to transform the random list in to a distinct one first.

Step 1: Bad event B occurs whenever there is a subset $S' \subseteq X = (X_1, \dots, X_{\delta m+1})$ of size at most (n_0+1), such that $\sum_{X_i \in S'} (X_i - X_{i-1}) \geq (1-2^{-\omega})n$. Whenever bad event B occurs, there is also a subset S'_{dist} of distinct integers that also satisfy the sum equation, and also has the size (n_0+1). (This is because if duplicate exists, we can always delete duplicates, without changing the sum, and then add new integers to make the sum even bigger.) thus we can use lemma 2 to simplify calculation.

Step 2: we consider the case there is explicitly d distinct integers in the list R, and then use union bound to deduce the final probability. Write the d distinct integers in R is ascending order as $Y = (Y_1, \dots, Y_d)$. We want to bound the probability that some subset Y'_{dist} of size n_0 is bad. (Note that the only change from the original lemma is that I changed m into n_0.) Then define L_i as in lemma 2. Let the set of indices $I \subseteq \{1, \dots, d+1\}$ of size n_0+1 be the set of index of bad subset. That is, $S'_{dist} = \{Y_i | i \in I\}$. Then define the permutation $\pi : \mathbb{Z}^{d+1} \to \mathbb{Z}^{d+1}$, such that if $i \in I$, then L_i appears in the first n_0+1 elements of $(L_{\pi(1)}, \dots, L_{\pi(d+1)})$. Define function $f : \mathbb{Z}^{d+1} \to \{0,1\}$ such that if its first n_0+1 arguments sums to at least $(1-2^{-\omega})n$, returns 1 and it returns 0 otherwise. Therefore, we have:

$$Pr[\sum_{i \in I} L_i \geq (1-2^{-\omega})n | D = d] = Pr[f(L_{\pi(1)}, \dots, L_{\pi(d+1)}) = 1 | D = d] \tag{7}$$

$$= Pr[f(L_1, \dots, Ld+1) = 1 | D = d] \tag{8}$$

$$= Pr[(L_1 + \dots + L_{n_0+1}) \geq (1-2^{-\omega})n | D = d] \tag{9}$$

Step 3: the event defined above can only happen when the other $d-(n_0+1)$ integers falls into the right most $2^{-\omega}n$ bins, and then we have:

$$Pr[(L_1 + \dots + L_{n_0+1}) \geq (1-2^{-\omega})n | D = d] \leq (\frac{1}{2})^{\omega(d-n_0-1)} \tag{10}$$

Above is the probability that a single set is bad, we then sum it over all $\binom{d+1}{n_0+1}$ possible size n_0+1 subsets, to get:

$$Pr[B | D = d] \leq \binom{d+1}{n_0+1} (\frac{1}{2})^{\omega(d-n_0-1)} \tag{11}$$

$$\leq 2^{(1-\omega)d + \omega n_0 + \omega + 1} \tag{12}$$

Equipped with $Pr[B|D=d]$, we can then proceed to calculate $Pr[B, D \geq d^*]$. Note when $\omega > 1$, $Pr[B|D=d]$ is non-increasing in d. And therefore we have:

$$Pr[B, D \geq d^*] \leq 2^{(1-\omega)d^* + \omega n_0 + \omega + 1} \cdot \sum_{d=d^*}^{\delta m+1} Pr[D = d] \tag{13}$$

$$\leq 2^{1+\omega} \cdot 2^{(1-\omega)\frac{\delta m}{2} + \omega n_0} \tag{14}$$

Completing the proof By the calculation above, we have

$$Pr[B] \leq Pr[D < d^*] + Pr[B, D \geq d^*] \tag{15}$$

$$\leq 2^{(1-\omega)\frac{\delta m}{2}} \cdot (2^{\omega m} + 2^{1+(n_0+1)\omega}) \tag{16}$$

□

2 Everywhere avoiding property

An analysis of original strategy In the original work [1], the author proves a *consecutive avoiding property* over all possible size of m, and then for m of fixed size ($n/64$), proves a tighter *everywhere avoiding property.* The reason of combining these two property is that the first one allows the number pebbles to be arbitrary within a given range. In particular, one can consider m to have the same number of the pebbles the adversary owns. In order to prove a tighter bound on $S \cdot T$ (like in part b of Lemma 31), one needs to first apply consecutive avoiding property and then use the everywhere avoiding property. What we want to do is to use the derived big spread lemma 1 to sum up over the choice of n_0, while keeping the amount of m as we want throughout the proof.

3 A bound on the pebbling moves of random sandwich graph

4 A comparison on re-computation and DRAM fetching

5 Simulation on common parameter settings

References

[1] Henry Corrigan-Gibbs, Dan Boneh, and Stuart E Schechter. Balloon hashing: Provably space-hard hash functions with data-independent access patterns. *IACR Cryptology ePrint Archive*, 2016:27, 2016.