1 Details of deriving the probability of collateral damage.

When the number (denoted as n) of onion services that are targeted for attack increases, we give an equation that quantifies the relationship between the probability (denoted as P) of collateral damage and the number (denoted as n) of onion services:

$$P = (1 - (2 - (\frac{N-8}{N})^{2n}) * (\frac{N-8}{N})^{2n})^2, \tag{1}$$

where n represents the number of onion services that are being targeted for attack, and N denotes the length of the DHT, i.e., the number of all HSDirs.

Next I will describe the derivation of Eq.(1) in detail.

Definition: When all 16 responsible HSDirs of a random and legitimate onion service are among the targets of HSDirSniper, it means that the onion service is collateral damage. As shown in Fig.1, in a DHT with a length of N, the region where the responsible HSDirs of the n attacked HSs falls is denoted as $T = \{T_0, .., T_m\}$, where m denotes the number of components in Region T and $m \leq 4 * n$. The unselected region in the DHT F is denoted as $F = \{F_0, .., F_m\}$. In short, the HSDirs contained in Region T are the targets of HSDirSniper's attacks, while the HSDirs contained in Region F will not be subject to HSDirSniper's attacks. We aim to solve for the probability (denoted as P) that a random and legitimate onion service is damaged under the above conditions. Considering that each onion service has two types of descriptors: the current and next descriptor, and the selection process of the responsible HSDirs for them is independent and identical. Therefore, we only need to consider only one type of descriptor(See section 2.1 of the manuscript). In this paper, we use the current descriptor as an example.

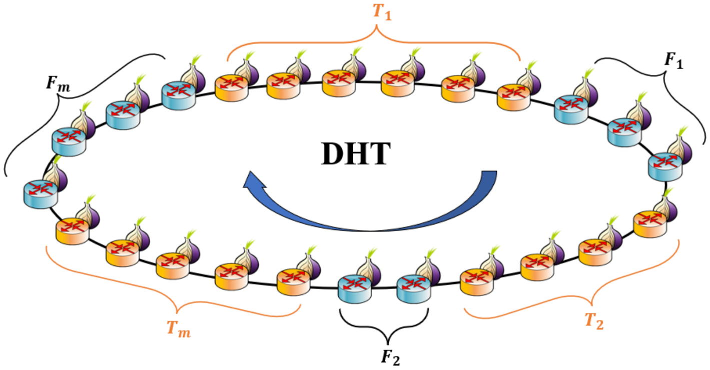

Figure 1: Diagrammatic representation of Region T and Region F. T is the region that is attacked, while F is the region that is not attacked.

Let event A indicate that all responsible HSDirs of a random and legitimate HS' current descriptor belong to Region T, and $p(A)$ denotes the probability $p(A)$ of event A occurring. Considering the complexity of solving $p(A)$ directly, we consider the opposing event B of event A, which indicates that at least one of the responsible HSDirs of the current descriptor belongs to Region F. Based on the nature of opposing events, we obtain

$$p(A) = 1 - p(B), \tag{2}$$

where $p(B)$ denotes the probability of event B occurring. Recall that the current descriptor has two replicas, and the Hidden Service calculates the descriptor ID for each replica using Eq.(1) in the manuscript. Subsequently, it determines which HSDir should be selected as the responsible HSDir for replica by comparing the descriptor ID to the index value (denoted as I, see line 310 in the manuscript) of the HSDir. Finally, each replica can be uploaded into four consecutive HSDirs whose indices follow the descriptor ID immediately.

Discrete random variables G_1 and G_2 indicate that the descriptor ID1 and ID2 (See section 2.1 and Fig. 8a of the manuscript) of the current descriptor fall in a range of the DHT, respectively. According to the law of total probability, we have

$$p(B) = p(G_1 = F) * p(B|G_1 = F) + p(G_1 = T) * p(B|G_1 = T) \tag{3}$$

Then, we illustrate how to derive the two condition probabilities one by one, i.e., $p(B|G_1 = T)$, and $p(B|G_1 = F)$. Recall that event B indicates that at least one responsible HSDir of the first or second descriptor replica belongs to Region F.

Given an event $G_1 = F$, i.e., descriptor ID1 of the first replica falls into Region F. Note that this replica can be uploaded into four consecutive HSDirs whose indices follow the descriptor ID1 immediately. In this case, event B always holds regardless of where the second set of descriptor replicas falls in the DHT, i.e., $p(B|G_1 = F) = 1$. Then we have $p(G_1 = F) * p(B|G_1 = F) = p(G_1 = F)$.

Given an event $G_1 = T$, i.e., descriptor ID1 of the first replica falls into Region T. According to the law of total probability, we can have

$$\begin{aligned} p(B|G_1 = T) &= p(G_2 = F) * p(B|G_1 = T, G_2 = F) \\ &+ p(G_2 = T) * p(B|G_1 = T, G_2 = T) \end{aligned} \tag{4}$$

To simplify Eq.(4), we need to derive two conditional probabilities, i.e., $p(B|G_1 = T, G_2 = F)$, and $p(B|G_1 = T, G_2 = T)$.

① Given an event $G_1 = T, G_2 = F$, i.e., descriptor ID1 of the first replica falls into Region T and descriptor ID2 of the second replica falls into Region F. This case satisfies the requirement that at least one responsible HSDir falls into Region F, as the descriptor ID2 of the second replica falls into Region F, i.e., event B holds. Therefore, we get $p(B|G_1 = T, G_2 = F) = 1$. From this, it can be inferred that $p(G_2 = F) * p(B|G_1 = T, G_2 = F) = p(G_2 = F)$.

② Given an event $G_1 = T, G_2 = T$, i.e., descriptor ID1 of the first replica and descriptor ID2 of the second replica both fall into Region T. In this case, neither replica's responsible HSDir belongs to Region F, i.e., event B does not hold. Then we know that $p(B|G_1 = T, G_2 = T) = 0$. Thus, we have $p(G_2 = T) * p(B|G_1 = T, G_2 = T) = 0$. By substituting the analysis results of ① and ② into Eq.(4), we get

$$\begin{aligned} p(B|G_1 = T) &= p(G_2 = F) * p(B|G_1 = T, G_2 = F) \\ &+ p(G_2 = T) * p(B|G_1 = T, G_2 = T) \\ &= p(G_2 = F) * 1 + p(G_2 = T) * 0 \\ &= p(G_2 = F) \end{aligned} \tag{5}$$

Substituting Eq.(5) into Eq.(3), we will get a more concise expression:

$$\begin{aligned} p(B) &= p(G_1 = F) * p(B|G_1 = F) + p(G_1 = T) * p(B|G_1 = T) \\ &= p(G_1 = F) * 1 + p(G_1 = T) * p(G_2 = F) \end{aligned} \tag{6}$$

According to Eq.(14) in the manuscript, the expected value of the number of responsible HSDirs for n onion services in a DHT is $N - N * (\frac{N-8}{N})^{2n}$, where n represents the number of onion services are being targeted for attack, and N denotes the length of the DHT, i.e., the number of all HSDirs. In short, the size of Region T is $N - N * (\frac{N-8}{N})^{2n}$, while the size of Region F is $N * (\frac{N-8}{N})^{2n}$. Thus, we can obtain the values of $p(G_1 = F)$, $p(G_1 = T)$ and $p(G_2 = F)$:

$$\begin{cases} p(G_1 = F) = p(G_2 = F) = \dfrac{N * (\frac{N-8}{N})^{2n}}{N} = (\dfrac{N-8}{N})^{2n}, \\ p(G_1 = T) = \dfrac{N - N * (\frac{N-8}{N})^{2n}}{N} = 1 - (\dfrac{N-8}{N})^{2n} \end{cases} \tag{7}$$

Substituting Eq.(6) and Eq.(7) into Eq.(2), we get

$$p(A) = 1 - p(B) = 1 - (2 - (\frac{N-8}{N})^{2n}) * (\frac{N-8}{N})^{2n} \tag{8}$$

Due to the independent nature of the current and next descriptor, we can deduce that

$$P = (p(A))^2 = (1 - (2 - (\frac{N-8}{N})^{2n}) * (\frac{N-8}{N})^{2n})^2, \tag{9}$$

where n represents the number of onion services are being targeted for attack, and N denotes the length of the DHT, i.e., the number of all HSDirs.