Birthday Paradox for Multi-collisions

Kazuhiro Suzuki[1], Dongvu Tonien[2], Kaoru Kurosawa[3], and Koji Toyota[3]

[1] Venture Business Laboratory, Ibaraki University 4–12–1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan
[2] School of Information Technology and Computer Science, University of Wollongong, Wollongong 2522, Australia
[3] Department of Computer and Information Sciences, Ibaraki University 4–12–1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan

Abstract. In this paper, we study multi-collision probability. For a hash function $H : D \to R$ with $|R| = n$, it has been believed that we can find an s-collision by hashing $Q = n^{(s-1)/s}$ times. We first show that this probability is at most $1/s!$ which is very small for large s. We next show that by hashing $(s!)^{1/s} \times Q$ times, an s-collision is found with probability approximately 0.5 for sufficiently large n. Note that if $s = 2$, it coincides with the usual birthday paradox. Hence it is a generalization of the birthday paradox to multi-collisions.

Keywords: hash function, birthday paradox, multi-collision, collision resistant.

1 Introduction

Let $H : D \to R$ be a hash function, where D is the domain and R is the range such that $|R| = n$. A collision for H is a distinct pair $x_1, x_2 \in D$ such that $H(x_1) = H(x_2)$. We usually require that H is collision resistant, which means that it is hard to find a collision. This security notion is used in many cryptographic applications such as digital signatures. All hash functions, however, suffer from the so-called birthday paradox which is a generic collision-finding attack. In this attack, we choose $x_1, \cdots, x_q \in D$ independently at random and compute $y_i = H(x_i)$ for $i = 1, \cdots, q$. We succeed if there is a pair i, j such that $H(x_i) = H(x_j)$. It is then well known that if $q = O(\sqrt{n})$, then we succeed with non-negligible probability (say, 0.5). Bellare, Kilian and Rogaway derived a nice upper bound and a lower bound on this success probability [1, Appendix].

Multi-collisions, on the other hand, are also an important notion of hash functions. An s-collision for H is s distinct points $x_1, \cdots, x_s \in D$ such that $H(x_1) = \cdots = H(x_s)$. As a negative side, Joux [5] showed a multi-collision attack on iterated hash functions at Crypto'04. As a positive side, the notion of multi-collisions was used for indentification schemes by Girault and Stern [4], for signature schemes by Brickell *et al.* [3] and for the micropayment scheme of Rivest and Shamir [6]. These schemes made use of an intuition such that finding an s-collision would be much harder than finding a usual collsion if s is large. Indeed, as a generalization of the birthday paradox, it has been believed that

M.S. Rhee and B. Lee (Eds.): ICISC 2006, LNCS 4296, pp. 29–40, 2006.

"We can find an s-collision by hashing $q = n^{(s-1)/s}$ x-values"

as written in [6, Sec.4] [5, Sec.2].

In this paper, we first present a negative result which shows that the above sentence is wrong. More precisely, we prove that by hashing $Q = n^{(s-1)/s}$ x-values, an s-collision is found with probability at most $1/s!$. Note that this probability is very small if s is large. Hence the above sentence is wrong for large s.

We next show a positive result such that by hashing $q = (s!)^{1/s} \times Q$ x-values[1], an s-collision is found with probability approximately at least 0.5 for sufficiently large n. Note that if $s = 2$, it coincides with the usual birthday paradox. Hence we can consider that it is a generalization of the birthday paradox to multi-collisions.

Throughout this paper, we suppose that each image $y \in R$ has the same number of preimages, that is, $|H^{-1}(y)| = |D|/|R|$ for all $y \in R$. In Sec. 2, we present a recursive formula which expresses the exact probability of finding an s-collision. In Sec. 3, we present a general lower and an upper bound of the probability of finding an s-collision. In Sec. 4, we show a more tight lower and an upper bound which agree within a constant factor for $q \leq n^{(s-1)/s}$. In Sec. 5, we show our main (negative and positive) results for $q = O(n^{(s-1)/s})$.

2 Exact Probability of s-Collision

In this section, we present a recursive formula for the probability of s-collision. We will use this formula to find the exact value and to derive bounds for the probability.

Let $2 \leq s \leq q \leq n$ and consider the following experiment. Suppose that there are q balls and n buckets. We throw the balls one by one at random into the buckets. Let $C(n, q, s)$ denote the event (called *s-collision*) that there exists at least one bucket that contains at least s balls.

The above experiment mimics the generic hashing attack as follows. We call n elements of the set R buckets. The q random x-values $x_1, \ldots, x_q$ are called balls. Each time we calculate the hash value $H(x_i)$, we imagine that the ball x_i is thrown into the bucket $H(x_i)$. If a bucket r contains at least s balls, say $x_{i_1}, \ldots, x_{i_s}$, then we have found an s-collision $H(x_{i_1}) = \ldots = H(x_{i_s}) = r$. Thus, the probability $Pr[C(n, q, s)]$ models the s-collision probability.

We now present a recursive formula of $Pr[C(n, q, s)]$.

Theorem 1

$$Pr[C(n,q,s)] = \frac{1}{n^{s-1}} \sum_{i=s}^{q} \binom{i-1}{s-1} \left(1 - \frac{1}{n}\right)^{i-s} (1 - Pr[C(n-1, i-s, s)]).$$

Proof. In the experiment of throwing q balls one by one at random into n buckets, for each $s \leq i \leq q$, let $C(n, q, s, i)$ denote the event that the i^{th} ball causes the

[1] Approximately, $q \approx (s/2.71) \times n^{(s-1)/s}$ from Stirling formula.

first s-collision, that is, s-collision does not occur until the i^{th} ball but does when the ball is thrown. Then

$$Pr[C(n,q,s)] = \sum_{i=s}^{q} Pr[C(n,q,s,i)].$$

We can find $Pr[C(n,q,s,i)]$ as follows:

1. One bucket (denoted by B), where the first s-collision occurs, can be selected from n buckets in n ways;
2. $s-1$ balls, which are put into B can be selected from the previous $i-1$ balls in $\binom{i-1}{s-1}$ ways;
3. The probability that the s selected balls land in the one selected bucket is $1/n^s$;
4. The probability that for the s selected balls and the one selected bucket B, none of the other $i-s$ balls land in B and cause an s-collision is $(1 - 1/n)^{i-s} \times (1 - Pr[C(n-1, i-s, s)])$.

Thus we have

$$\begin{aligned} Pr[C(n,q,s,i)] &= n \times \binom{i-1}{s-1} \times \frac{1}{n^s} \times \left(1 - \frac{1}{n}\right)^{i-s} \times (1 - Pr[C(n-1, i-s, s)]) \\ &= \frac{1}{n^{s-1}} \binom{i-1}{s-1} \left(1 - \frac{1}{n}\right)^{i-s} (1 - Pr[C(n-1, i-s, s)]). \end{aligned}$$

Therefore,

$$Pr[C(n,q,s)] = \frac{1}{n^{s-1}} \sum_{i=s}^{q} \binom{i-1}{s-1} \left(1 - \frac{1}{n}\right)^{i-s} (1 - Pr[C(n-1, i-s, s)]). \quad \blacksquare$$

We will use this recursive formula to calculate the exact value of the s-collision probability and derive its bounds in the next sections. Before doing that we need some auxiliary results. The proofs are shown in the Appendix.

Lemma 1. *The following statements must hold*

1. *For any positive integers k, s, and $i \geq (k+1)s$,*

$$\sum_{i=s}^{q} \binom{i-1}{s-1} = \binom{q}{s}.$$

2. *For any positive integers k, s, and $i \geq (k+1)s$,*

$$\binom{i-1}{ks-1}\binom{i-ks}{s} = \binom{(k+1)s-1}{s}\binom{i-1}{(k+1)s-1}.$$

3. *For any positive integers $k \geq 2$, s, and $q \geq ks$,*

$$\binom{ks-1}{s}\binom{q}{ks} = \frac{k-1}{k}\binom{q}{s}\binom{q-s}{(k-1)s}.$$

4. *For any integers* $n, s \geq 2$,
$$(n-1)^{s-1} > (n^{\frac{s-1}{s}} - 1)^s.$$
5. *For any* $1 < a \leq b$,
$$\frac{a-1}{b-1} \leq \frac{a}{b}$$
6. *Let* $e_k = (1-1/k)^{-k}$ *then* $\{e_k\}_{k=2}^{\infty}$ *is a decreasing sequence and* $\lim_{k\to\infty} e_k = e \approx 2.7$ *– the Euler constant. For any* $0 < x < 1$, *we have*
$$e_k^{-x} > 1 - x \ln e_k \approx 1 - x.$$
7. *For any integer* $s \geq 2$,
$$(s!)^{-1/s}(s+1)/2 > 1.$$

3 Bounds on the Probability of s-Collision

In this section, we present the following bounds on the probability of s-collision.

Theorem 2
$$Pr[C(n,q,s)] \leq \frac{1}{n^{s-1}}\binom{q}{s},$$
and
$$Pr[C(n,q,s)] \geq \frac{1}{n^{s-1}}\binom{q}{s}\left(1-\frac{1}{n}\right)^{q-s}\left\{1 - \frac{1}{2(n-1)^{s-1}}\binom{q-s}{s}\right\}.$$

Proof. By Theorem 1 and Lemma 1(1), we obtain the upper bound
$$\begin{aligned}Pr[C(n,q,s)] &= \frac{1}{n^{s-1}}\sum_{i=s}^{q}\binom{i-1}{s-1}\left(1-\frac{1}{n}\right)^{i-s}(1 - Pr[C(n-1,i-s,s)])\\ &\leq \frac{1}{n^{s-1}}\sum_{i=s}^{q}\binom{i-1}{s-1} = \frac{1}{n^{s-1}}\binom{q}{s}.\end{aligned}$$

We have
$$\begin{aligned}Pr[C(n,q,s)] &= \frac{1}{n^{s-1}}\sum_{i=s}^{q}\binom{i-1}{s-1}\left(1-\frac{1}{n}\right)^{i-s}(1 - Pr[C(n-1,i-s,s)])\\ &\geq \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\sum_{i=s}^{q}\binom{i-1}{s-1}(1 - Pr[C(n-1,i-s,s)])\\ &= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\sum_{i=s}^{q}\binom{i-1}{s-1} - \sum_{i=s}^{q}\binom{i-1}{s-1}Pr[C(n-1,i-s,s)]\right]\\ &= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \sum_{i=2s}^{q}\binom{i-1}{s-1}Pr[C(n-1,i-s,s)]\right]\end{aligned}$$

where the last equality follows from the fact that $Pr[C(n-1, i-s, s)] = 0$ for $i \le 2s-1$ and Lemma 1(1).

Now using the above upper bound, we derive the lower bound,

$$Pr[C(n,q,s)] \ge \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \sum_{i=2s}^{q}\binom{i-1}{s-1}\frac{1}{(n-1)^{s-1}}\binom{i-s}{s}\right]$$
$$= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \frac{1}{(n-1)^{s-1}}\sum_{i=2s}^{q}\binom{i-1}{s-1}\binom{i-s}{s}\right].$$

By Lemma 1(2),

$$= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \frac{1}{(n-1)^{s-1}}\binom{2s-1}{s}\sum_{i=2s}^{q}\binom{i-1}{2s-1}\right].$$

By Lemma 1(1),

$$= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \frac{1}{(n-1)^{s-1}}\binom{2s-1}{s}\binom{q}{2s}\right].$$

By Lemma 1(3),

$$= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\left[\binom{q}{s} - \frac{1}{2(n-1)^{s-1}}\binom{q}{s}\binom{q-s}{s}\right]$$
$$= \frac{1}{n^{s-1}}\left(1-\frac{1}{n}\right)^{q-s}\binom{q}{s}\left\{1 - \frac{1}{2(n-1)^{s-1}}\binom{q-s}{s}\right\}. \qquad \blacksquare$$

From now on, we use the following notation,

$$f(n) = \left(1-\frac{1}{n}\right)^{q-s} \quad \text{and} \quad g(n) = \frac{1}{2(n-1)^{s-1}}\binom{q-s}{s}.$$

Theorem 2 can be rewritten as

$$f(n)(1-g(n))\,\frac{1}{n^{s-1}}\binom{q}{s} \le Pr[C(n,q,s)] \le \frac{1}{n^{s-1}}\binom{q}{s}. \tag{1}$$

4 Bounds for $q = \Theta(n^{\epsilon})$ Where $\epsilon < (s-1)/s$

The graph in Figure 1 demonstrates the upper bound and the lower bound in Theorem 2 and the exact probability of $Pr[C(n,q,s)]$ for $n = 365$ and $s = 3$. From this figure, we can see that for $q < n^{(s-1)/s} \approx 52$, the difference between values of these three graphs is small. We will show that when $q = n^{\epsilon}$ with $\epsilon < \frac{s-1}{s}$, the upper bound and the lower bound are indeed very close to each other. We also show that in this case, the upper bound asymptotically tends to zero.

Theorem 3. *Let ϵ be a positive number such that $\epsilon < \frac{s-1}{s}$. Then for any positive number $c < 1$, there exists a positive number n_0 such that*

$$c \times \frac{1}{n^{s-1}} \binom{q}{s} < Pr[C(n,q,s)] \le \frac{1}{n^{s-1}} \binom{q}{s},$$

for any $n > n_0$ and $2 \le s \le q = n^{\epsilon}$.

Proof. The theorem follows from the following two claims.
Claim 1.

$$g(n) < \frac{1}{2\ s!} \frac{q^s}{n^{s-1}} = \frac{1}{2\ s!\ n^{s-1-s\epsilon}},$$

thus, $g(n) \to 0$ when $n \to \infty$.
Proof. We have

$$\binom{q-s}{s} = \frac{(q-s)(q-s-1)\dots(q-2s+1)}{s!} < \frac{(q-1)^s}{s!},$$

By Lemma 1(4),

$$(n-1)^{s-1} > (n^{\frac{s-1}{s}} - 1)^s.$$

Thus,

$$g(n) = \frac{1}{2(n-1)^{s-1}} \binom{q-s}{s} < \frac{1}{2(n^{\frac{s-1}{s}} - 1)^s} \frac{(q-1)^s}{s!} = \frac{1}{2\ s!} \left(\frac{q-1}{n^{\frac{s-1}{s}} - 1} \right)^s$$

By Lemma 1(5),

$$g(n) < \frac{1}{2\ s!} \left(\frac{q}{n^{\frac{s-1}{s}}} \right)^s = \frac{1}{2\ s!} \frac{q^s}{n^{s-1}} = \frac{1}{2\ s!\ n^{s-1-s\epsilon}}.$$

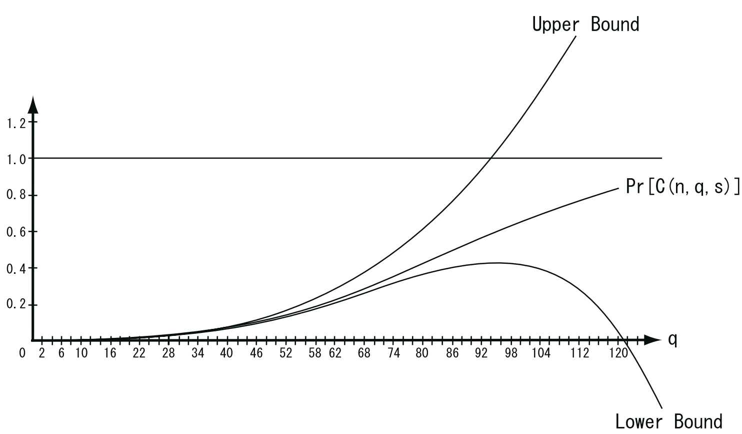

Fig. 1. The upper bound and the lower bound of Theorem 2 and the exact probability of $Pr[C(n,q,s)]$ for $n = 365$ and $s = 3$. We use the recursive formula in Section 2 to calculate the exact probability $Pr[C(n,q,s)]$.

Since $s-1-s\epsilon>0$, we have $g(n)\to 0$ when $n\to\infty$.
Claim 2. With the notation in Lemma 1(6), for any $n>k$,

$$f(n)>e_k^{-q/n}=e_k^{-n^{\epsilon-1}},$$

where $e_k\approx e$, thus, $f(n)\to 1$ when $n\to\infty$.
Proof. We have

$$f(n)=\left(1-\frac{1}{n}\right)^{q-s}>\left(1-\frac{1}{n}\right)^{q}=\left[\left(1-\frac{1}{n}\right)^{-n}\right]^{-q/n}=e_n^{-q/n}.$$

Since $n>k$, by Lemma 1(6), $e_n<e_k$, thus,

$$f(n)>e_k^{-q/n}=e_k^{-n^{\epsilon-1}}.$$

Since $\epsilon<1$, $n^{\epsilon-1}\to 0$ and $f(n)\to 1$ as $n\to\infty$.

From *Claim 1* and *Claim 2*, we have $f(n)(1-g(n))\to 1$, thus, the theorem follows. ■

Example. Let $s=4$, $\epsilon=\frac{1}{2}<\frac{s-1}{s}=\frac{3}{4}$, and $n>100$ then

$$g(n)<\frac{n^{s\epsilon-(s-1)}}{2\ s!}=\frac{n^{-1}}{48}<\frac{1}{4800}=0.000208333,$$

$$f(n)>e_{100}^{-n^{\epsilon-1}}=e_{100}^{-n^{-1/2}}>e_{100}^{-100^{-1/2}}=\left[(1-1/100)^{-100}\right]^{-100^{-1/2}}>.9$$

Thus $f(n)(1-g(n))>.8998$, and we have

$$.8998\times\frac{1}{n^{s-1}}\binom{q}{s}<Pr[C(n,q,s)]\leq\frac{1}{n^{s-1}}\binom{q}{s}.$$ ■

Even though Theorem 3 shows that the upper bound and the lower bound are very closed to each other, the following lemma shows that these bounds asymptotically tend to zero.

Lemma 2. *Let ϵ be a positive number such that $\epsilon<\frac{s-1}{s}$ and $q=n^{\epsilon}$, then*

$$\frac{1}{n^{s-1}}\binom{q}{s}\to 0\qquad \text{when } n\to\infty.$$

Proof. We have

$$\frac{1}{n^{s-1}}\binom{q}{s}<\frac{1}{n^{s-1}}\frac{q^s}{s!}=\frac{1}{s!\ n^{s-1-s\epsilon}}.$$

Since $s-1-s\epsilon>0$, we have

$$\frac{1}{n^{s-1}}\binom{q}{s}\to 0.$$ ■

5 Bounds for $q = \Theta(n^{(s-1)/s})$

In this section, we consider the case $q = \Theta(n^{(s-1)/s})$. We prove two main theorems. Theorem 4 shows that if $q \approx n^{(s-1)/s}$ and n is sufficiently large then $Pr[C(n, q, s)] \approx 1/s!$, and Theorem 5 shows that if $q \approx (s!)^{1/s} n^{(s-1)/s}$ and n is sufficiently large then $Pr[C(n, q, s)] \gtrapprox 1/2$.

It implies the following *generalized birthday paradox*

> *For a hash function $H : D \to R$ with $|R| = n$, if n is sufficiently large then by $n^{(s-1)/s}$ number of hashings, an s-collision can be found with probability $\approx 1/s!$, and by $(s!)^{1/s}\, n^{(s-1)/s}$ number of hashings an s-collision can be found with probability $\gtrapprox 1/2$.*

Theorem 4. *We suppose that $q = \alpha\, n^{(s-1)/s}$, $q - s = \alpha'\, n^{(s-1)/s}$, where $0 < \alpha' < \alpha < 1$. If $2 \le s \le q$ then*

$$Pr[C(n, q, s)] \le \frac{1}{n^{s-1}} \binom{q}{s} < \frac{\alpha^s}{s!} < \frac{1}{s!} \tag{2}$$

and

$$Pr[C(n, q, s)] > \frac{\alpha'^s}{s!} - \left(\frac{\alpha'^{s+1} \ln e_n}{s!\, n^{1/s}} + \frac{(\alpha\alpha')^s}{2(s!)^2} \right)$$

where $e_n = (1 - 1/n)^{-n} \approx e$. In particular, if n is sufficiently large so that $1/n^{1/s} \approx 0$, and $\alpha' \lessapprox \alpha \lessapprox 1$, then we have

$$Pr[C(n, q, s)] > \frac{\alpha'^s}{s!} - \left(\frac{\alpha'^{s+1} \ln e_n}{s!\, n^{1/s}} + \frac{(\alpha\alpha')^s}{2(s!)^2} \right) \approx \frac{1}{s!} - \frac{1}{2(s!)^2}$$

Proof. We have

$$\frac{1}{n^{s-1}} \binom{q}{s} = \frac{1}{n^{s-1}} \frac{q(q-1)\dots(q-s+1)}{s!} < \frac{1}{n^{s-1}} \frac{q^s}{s!} = \frac{\alpha^s}{s!}$$

thus

$$Pr[C(n, q, s)] \le \frac{1}{n^{s-1}} \binom{q}{s} < \frac{\alpha^s}{s!} < \frac{1}{s!}.$$

We have

$$\frac{1}{n^{s-1}} \binom{q}{s} = \frac{1}{n^{s-1}} \frac{q(q-1)\dots(q-s+1)}{s!} > \frac{1}{n^{s-1}} \frac{(q-s)^s}{s!} = \frac{\alpha'^s}{s!}.$$

As in the proof of Theorem 3, we have

$$g(n) < \frac{1}{2\ s!} \frac{q^s}{n^{s-1}} = \frac{\alpha^s}{2\ s!}.$$

and by Lemma 1(6),

$$f(n) = e_n^{-(q-s)/n} > 1 - \frac{q-s}{n} \ln e_n = 1 - \frac{\alpha' \ln e_n}{n^{1/s}}.$$

Thus,

$$f(n)(1-g(n)) \geq f(n) - g(n) > 1 - \frac{\alpha' \ln e_n}{n^{1/s}} - \frac{\alpha^s}{2\ s!}.$$

Therefore,

$$\begin{aligned} Pr[C(n,q,s)] &\geq f(n)(1-g(n))\,\frac{1}{n^{s-1}}\binom{q}{s} \\ &> \left(1 - \frac{\alpha' \ln e_n}{n^{1/s}} - \frac{\alpha^s}{2\ s!}\right)\frac{\alpha'^s}{s!} \\ &= \frac{\alpha'^s}{s!} - \left(\frac{\alpha'^{s+1} \ln e_n}{s!\, n^{1/s}} + \frac{(\alpha\alpha')^s}{2(s!)^2}\right). \end{aligned}$$

■

Theorem 5. *If* $2 \leq s \leq q$, *and* $q = (s!)^{1/s} n^{(s-1)/s} + s - 1 (< n)$, *then we have*

$$Pr[C(n,q,s)] > \frac{1}{2} - \left(\frac{s!}{n}\right)^{1/s} \ln e_n.$$

In particular, if n *is sufficiently large so that* $(s!/n)^{1/s} \approx 0$, *then we have*

$$Pr[C(n,q,s)] > \frac{1}{2} - \left(\frac{s!}{n}\right)^{1/s} \ln e_n \approx \frac{1}{2}. \tag{3}$$

Proof. By Cauchy's inequality,

$$\begin{aligned} \binom{q-s}{s} &= \frac{(q-s)(q-s-1)\ldots(q-2s+1)}{s!} \\ &< \frac{1}{s!}\left(\frac{(q-s)+(q-s-1)+\ldots+(q-2s+1)}{s}\right)^s \\ &= \frac{[q-(3s-1)/2]^s}{s!} = \frac{[(s!)^{1/s}\, n^{(s-1)/s} - (s+1)/2]^s}{s!} \\ &= [n^{(s-1)/s} - (s!)^{-1/s}(s+1)/2]^s \end{aligned}$$

By Lemma 1(7), $(s!)^{-1/s}(s+1)/2 > 1$, thus,

$$\binom{q-s}{s} < (n^{(s-1)/s} - 1)^s.$$

By Lemma 1(4),

$$(n-1)^{s-1} > (n^{(s-1)/s} - 1)^s,$$

thus,

$$g(n) = \frac{1}{2(n-1)^{s-1}}\binom{q-s}{s} < \frac{1}{2}. \tag{4}$$

We have

$$f(n) = \left(1 - \frac{1}{n}\right)^{q-s} > \left(1 - \frac{1}{n}\right)^{q-s+1} = e_n^{-(q-s+1)/n},$$

thus, by Lemma 1(6),

$$f(n) > e_n^{-(q-s+1)/n} > 1 - \frac{q-s+1}{n} \ln e_n = 1 - \left(\frac{s!}{n}\right)^{1/s} \ln e_n. \tag{5}$$

From (4) and (5), we have

$$f(n)(1-g(n)) \geq f(n) - g(n) > \frac{1}{2} - \left(\frac{s!}{n}\right)^{1/s} \ln e_n. \tag{6}$$

We have

$$\frac{1}{n^{s-1}}\binom{q}{s} > \frac{(q-s+1)^s}{s!\; n^{s-1}} = \frac{((s!)^{1/s} n^{(s-1)/s})^s}{s!\; n^{s-1}} = 1. \tag{7}$$

Combining (6) and (7) gives

$$Pr[C(n,q,s)] \geq f(n)(1-g(n))\frac{1}{n^{s-1}}\binom{q}{s} > \frac{1}{2} - \left(\frac{s!}{n}\right)^{1/s} \ln e_n. \qquad \blacksquare$$

Example. If $s \geq 2$, $n > s!\,32^s (\geq 2048)$ and $q = (s!)^{1/s} n^{(s-1)/s} + s - 1 (< n)$ then

$$\left(\frac{s!}{n}\right)^{1/s} < \frac{1}{32} \quad \text{and } \ln e_{2048} < 1.00025,$$

thus

$$Pr[C(n,q,s)] > \frac{1}{2} - \frac{1}{32} \times 1.00025 > .4687 \qquad \blacksquare$$

6 Conclusion

In this paper, we have studied multi-collision probabilities for regular hash functions $H : D \to R$, where "regular" means that each image $y \in R$ has the same number of preimages. Suppose that that $|R| = n$. Then our main results are summarized as follows.

– By hashing about $n^{(s-1)/s}$ times, an s-collision is found with probability at most $1/s!$ (see eq.(2)). Since it is very small for large s, this disproves the folklore which has been believed so far.

– By hashing about $(s!)^{1/s} n^{(s-1)/s}$ times, an s-collision is found with probability approximately 1/2 or more if n is large enough so that $(s!/n)^{1/s} \approx 0$ (see eq.(3)). Hence this is a true generalization of the birthday paradox to multicollisions.

Bellare and Kohno genralized the birthday paradox (for $s = 2$) to non-regular hash functions [2]. It will be a furter work to generalize our result on multicollision to non-regular hash functions.

Acknowledgement

The approximation of the footnote of Sec.1 was pointed out by an anonymous reviewer.

References

1. M.Bellare, J.Kilian and P.Rogaway: The Security of the Cipher Block Chaining Message Authentication Code. J. Comput. Syst. Sci. 61(3): pp.362–399 (2000)
2. M.Bellare and T.Kohno: Hash Function Balance and its Impact on Birthday Attacks. EUROCRYPT 2004, pp.401–418 (2004)
3. E.Brickell, D.Pointcheval, S.Vaudenay and M.Yung: Design Validations for Discrete Logarithm Based Signature Schemes. Public Key Cryptography 2000: pp.276–292 (2000)
4. M.Girault and J.Stern: On the Length of Cryptographic Hash-Values Used in Identification Schemes. CRYPTO 1994: pp.202–215 (1994)
5. Antoine Joux: Multicollisions in Iterated Hash Functions. Application to Cascaded Constructions. CRYPTO 2004: 306-316 (2004)
6. R.Rivest and A.Shamir: PayWord and MicroMint: Two Simple Micropayment Schemes. Security Protocols Workshop 1996: pp.69–87 (1996)

Appendix: Proofs of Lemma 1

Proof. (1) Since

$$\binom{i}{s} = \binom{i-1}{s} + \binom{i-1}{s-1},$$

we have

$$\sum_{i=s}^{q} \binom{i-1}{s-1} = 1 + \sum_{i=s+1}^{q} \binom{i-1}{s-1} = 1 + \sum_{i=s+1}^{q} \left[\binom{i}{s} - \binom{i-1}{s} \right] = 1 + \binom{q}{s} - \binom{s}{s} = \binom{q}{s}.$$

(2) We have

$$\begin{aligned} \binom{i-1}{ks-1}\binom{i-ks}{s} &= \frac{(i-1)!}{(ks-1)!(i-ks)!} \times \frac{(i-ks)!}{s!(i-(k+1)s)!} \\ &= \frac{((k+1)s-1)!}{s!(ks-1)!} \times \frac{(i-1)!}{((k+1)s-1)!(i-(k+1)s)!} \\ &= \binom{(k+1)s-1}{s}\binom{i-1}{(k+1)s-1}. \end{aligned}$$

(3) We have

$$\binom{ks-1}{s}\binom{q}{ks} = \frac{(ks-1)!}{s!((k-1)s-1)!} \times \frac{q!}{(ks)!(q-ks)!}$$

$$= \frac{q!}{(ks)\ s!\ ((k-1)s-1)!\ (q-ks)!}$$

$$= \frac{k-1}{k} \times \frac{q!}{s!(q-s)!} \times \frac{(q-s)!}{((k-1)s)!(q-ks)!}$$

$$= \frac{k-1}{k}\binom{q}{s}\binom{q-s}{(k-1)s}.$$

(4) Let $0 < t = \frac{s-1}{s} < 1$ and consider the function $a(n) = (n-1)^t - n^t + 1$. We have $a'(n) = t[(n-1)^{t-1} - n^{t-1}] > 0$. Thus, $a(n) \geq a(2) = 2 - 2^t > 0$. Therefore,

$$(n-1)^{\frac{s-1}{s}} > n^{\frac{s-1}{s}} - 1,$$

and thus,

$$(n-1)^{s-1} > (n^{\frac{s-1}{s}} - 1)^s.$$

(5) We have

$$\frac{a-1}{b-1} \leq \frac{a}{b} \leftrightarrow b(a-1) \leq a(b-1) \leftrightarrow a \leq b.$$

(6) It is a basic result that the sequence $\{e_k\}_{k=2}^{\infty}$ is a decreasing sequence, $e_k > e$, and $\lim_{k\to\infty} e_k = e$, the proof of this result can be found in any calculus textbook. We have

$$e^{-x} > 1 - x,$$

thus,

$$e_k^{-x} = (e^{-x})^{\ln e_k} > (1-x)^{\ln e_k}, \text{ and by Bernoulli's inequality,}$$
$$> 1 - x\,\ln e_k.$$

(7) By Cauchy's inequality,

$$s! < \left(\frac{1+2+\ldots+s}{s}\right)^s = \left(\frac{s+1}{2}\right)^s,$$

thus, $(s+1)/2 > (s!)^{1/s}$. It follows that $(s!)^{-1/s}\,(s+1)/2 > 1$. ■