Stam's Conjecture and Threshold Phenomena in Collision Resistance

John Steinberger[1⋆] Xiaoming Sun[2⋆⋆] and Zhe Yang[3]

[1] Institute of Theoretical Computer Science, Tsinghua University, Beijing. jpsteinb@gmail.com
[2] Institute of Computing Technology, China Academy of Sciences. xiaoming.sun@gmail.com
[3] Hulu Software, Beijing. yangzhe1990@gmail.com

Abstract. At CRYPTO 2008 Stam [8] conjectured that if an $(m+s)$-bit to s-bit compression function F makes r calls to a primitive f of n-bit input, then a collision for F can be obtained (with high probability) using $r2^{(nr-m)/(r+1)}$ queries to f, which is sometimes less than the birthday bound. Steinberger [9] proved Stam's conjecture up to a constant multiplicative factor for most cases in which $r = 1$ and for certain other cases that reduce to the case $r = 1$. In this paper we prove the general case of Stam's conjecture (also up to a constant multiplicative factor). Our result is qualitatively different from Steinberger's, moreover, as we show the following novel threshold phenomenon: that exponentially many (more exactly, $2^{s-2(m-n)/(r+1)}$) collisions are obtained with high probability after $O(1)r2^{(nr-m)/(r+1)}$ queries. This in particular shows that threshold phenomena observed in practical compression functions such as JH are, in fact, unavoidable for compression functions with those parameters. (This is the full version of the same-titled article that appeared at CRYPTO '12.)

1 Introduction

The ideal primitive model (IPM) is a popular paradigm in cryptographic security proofs. In this model one assumes that some primitive used by a construction, such as a blockcipher, is "ideal"—namely perfectly random subject to the constraints of the type of primitive under consideration—and one then bounds the chance of success of an adversary given oracle access to the ideal primitive, in some given security experiment, for some given number of queries. The adversary considered is almost always information-theoretic. As such, the adversary's only obstacle to achieving its attack is the randomness of the query responses.

Because the IPM considers information-theoretic adversaries certain limitations naturally arise as to what kind of security can be achieved for a certain functionality using a certain primitive a certain number of times. For example, consider the task of constructing a $2n$-bit to n-bit compression function F using a random n-bit to n-bit permutation f as a primitive. There are 2^{2n} inputs to F but only 2^n inputs to f. Thus each input to f corresponds on average to 2^n inputs to F, so with just two calls to f we can learn to evaluate F on at least $2 \cdot 2^n$ inputs. But this is more than the number of outputs of F, so a collision can be obtained with probability 1 in just two queries. Note that determining which two f-queries to make is no problem for an information-theoretic adversary, nor is "finding the collision" among the $2 \cdot 2^n$ mapped values. Thus it is not possible to design a compression function with these parameters that is collision resistant in the IPM.

This paper follows a line of work [2, 6, 8, 9] in the same vein as the above argument, seeking to establish the limits of provable security in the IPM model. Specifically, we focus the following question related to work of Stam [8] and, before that, of Rogaway and Steinberger [6]: given $m, n, r, s \geq 1$, what is the maximum collision security of a compression function $F : \{0,1\}^{m+s} \to \{0,1\}^s$ that makes r calls to an ideal primitive f of domain $\{0,1\}^n$? (The range of f is not specified because it turns out to

⋆ Supported by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174, and by NSF grant 0994380.
⋆⋆ Supported by the National Natural Science Foundation of China Grant 61170062, 61061130540, and the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301.

be immaterial[1].) Here "collision security" means the largest number of f-queries the best information-theoretic adversary can ask before achieving probability $\frac{1}{2}$ of obtaining a collision.

Since it costs at most r queries to evaluate any point in the domain, a birthday attack implies that collision security cannot exceed $q = O(1)r2^{s/2}$ queries. However, depending on the parameters, other attacks may be more effective than birthday attacks. In particular Stam [8] conjectured that

$$q = r\lceil 2^{(nr-m)/(r+1)} \rceil + 1 \tag{1}$$

queries should always suffice for finding a collision with probability at least $\frac{1}{2}$. (We restate Stam's conjecture as slightly modified by Steinberger [9].) Roughly speaking, this bound is less than a birthday attack when $s/2 > (nr - m)/(r + 1)$. The latter occurs for example when $(m, n, r, s) = (n, n, 2, n)$, the case of a $2n$-bit to n-bit compression function making two calls to a primitive of n-bit input, for which Stam's bound forecasts a maximum collision resistance of $2^{n/3}$, which is more restrictive than the birthday bound of $2^{n/2}$. As a second example, Stam's bound is even more restrictive when $(m, n, r, s) = (n, n, 1, n)$, for which it forecasts a maximum collision resistance of $O(1)$ queries; in fact this setting of parameters coincides with the first example discussed in the paper (regarding a compression function $F : \{0, 1\}^{2n} \to \{0, 1\}^n$ making a single call to an n-bit random permutation).

Stam's conjecture is appealing because it apparently constitutes the *optimal* upper bound on collision resistance for all cases for which it beats the birthday bound, while the birthday bound can apparently be achieved in all other cases. In other words, as far as currently understood, it seems like the maximum collision resistance of a compression function $F : \{0, 1\}^{m+s} \to \{0, 1\}^s$ making r calls to a random function f of n-bit input equals

$$\min(r2^{s/2}, r\lceil 2^{(nr-m)/(r+1)} \rceil)$$

up to possible lower order terms. This thesis is supported by a number of constructions [5, 7, 8].

Steinberger [9] obtained the only previous results on Stam's conjecture. He proved that when

$$(2m - n(r - 1))/(r + 1) \geq 4.09 \tag{2}$$

$O(1)r\lceil 2^{(nr-m)/(r+1)} \rceil$ queries suffice to find a collision for F with probability at least 0.5. The condition (2) is increasingly restrictive as r grows; for $r = 1$, it reduces to $m \geq 4.09$; for $r = 2$, it reduces to $\frac{2}{3}m - \frac{1}{3}n \geq 4.09$; for $r = 3$ it reduces to $\frac{1}{2}m - \frac{1}{2}n \geq 4.09$; and so on. If $m = n$ (a typical case in real-world constructions) then (2) is false for all $r \geq 3$. Thus Stam's conjecture was until now, and despite Steinberger's result, very much open in the general case.

Steinberger also made the observation that for certain parameters (m, n, r, s) Stam's conjecture can be reduced to parameters (m', n, r', s) such that $m' < m$ and $r' < r$. (To be precise, such a reduction can be effected whenever $mr \geq n$.) In fact, the core of Steinberger's result is a proof of Stam's conjecture for the case[2] $r = 1$ and $m \geq 4.09$. Other parameters (m, n, r, s) with $r > 1$ to which Steinberger's result applies are precisely those for which the reduced tuple (m', n, r', s) has $r' = 1$ and $m' \geq 4.09$ (inequality (2) is sufficient and necessary for both $r' = 1$ and $m' \geq 4.09$ to hold). Thus Steinberger's result is "really" about the case $r = 1$ of Stam's conjecture. (To be fair, the results of [9] nonetheless cover a large number of parameter settings of practical interest.)

In this paper we resolve the general case of Stam's conjecture. More precisely we show that if $F : \{0, 1\}^{m+s} \to \{0, 1\}^s$ is a compression function using r calls to a primitive f of n-bit input, where $m \geq 1$, then with high probability a collision can be found for F in at most

$$O(1)r \left\lceil 2^{\frac{nr-m}{r+1}} \right\rceil$$

[1] Immaterial to proving the upper bound under consideration in this paper; better upper bounds on security should be provable if f has sufficiently small range, see comments by Stam [8].

[2] When m is not an integer, our meaning is that F is a compression function of domain of size at least $\lceil 2^{s+m} \rceil$; the qualitative nature of the domain (be it bitstrings, or some other set) is not relevant. See Section 2 for a more precise statement.

queries to f, where the $O(1)$ term represents a constant independent of all other parameters. This constant, being in the vicinity of 16000, is large but not astronomical. Note that *some* lower bound must be imposed on m, since if $m = 0$ the domain has the same size as the range, and F may have no collisions at all (e.g., F may ignore its primitive f, and be the identity on $\{0,1\}^s$). In fact, Stam's conjecture doesn't hold under the sole assumption $m > 0$, as is easy to see. For example, this would allow the domain to have a single more point than the range, making a collision very hard to find[3].

At this point we emphasize that, like for Steinberger's theorem, the "primitive" f called by the compression function F can be *any* type of primitive of n-bit input, i.e., can be drawn from *any* distribution. Such a primitive can model, for example, a n-bit to n-bit permutation, but it can also model, say, a blockcipher[4], or essentially any type of function-like primitive.

On the other hand our result is qualitatively different from Steinberger's in that we show an interesting threshold phenomenon: for the range of parameters in which Stam's bound is less than the cost of a birthday attack (namely, when $s/2 > (nr-m)/(r+1)$), we show *many* collisions are obtained with high probability as soon as $O(1)r\lceil 2^{\frac{nr-m}{r+1}}\rceil$ queries are made; more precisely, one obtains at least

$$2^{s-\frac{2(nr-m)}{r+1}} \tag{3}$$

collisions with high probability, using at most $16000r\lceil 2^{\frac{nr-m}{r+1}}\rceil$ queries to f. In this regard, it is worth recalling that Stam's bound is conjecturally optimal (and for some settings of parameters provably optimal), implying that with only

$$o(1)r\lceil 2^{\frac{nr-m}{r+1}}\rceil$$

queries to f, *no* collisions for F are found with high probability (presuming an adequate F). Note the exponent $s-2(nr-m)/(r+1)$ in (3) is precisely twice the difference between the exponent in the cost of a birthday attack (a.k.a. $s/2$) and the exponent of Stam's conjecture (a.k.a. $(nr-m)/(r+1)$). Thus, the further Stam's bound is beneath the cost of a birthday attack, the sharper the threshold phenomenon.

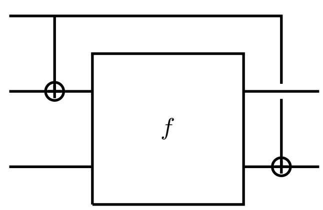

Fig. 1. The JH compression function $G : \{0,1\}^{1.5n} \to \{0,1\}^n$. All wires carry $n/2$-bit values.

As an example, we can consider the compression function G of JH [11], one of the finalists in NIST's SHA-3 competition, pictured in Figure 1. This is a compression function from $\{0,1\}^{1.5n}$ to $\{0,1\}^n$ using a single call to a primitive f of n-bit input (more precisely, f is a permutation). Thus the JH compression function G has parameters $(m,n,r,s) = (0.5n, n, 1, n)$. The cost of a birthday attack for G is $2^{n/2}$. Stam's bound, however, is

$$2^{\frac{nr-m}{r+1}} = 2^{\frac{n-0.5n}{2}} = 2^{n/4}.$$

[3] See Wiener [10] for details on the effectiveness of birthday attacks in functions were the size of the domain approaches the size of the range.

[4] A blockcipher with m-bit word and k-bit key can be modeled as a primitive of input length $n = m + k$, or of input length $n = m + k + 1$ if the construction also uses "inverse" blockcipher calls (in which case the extra bit indicates whether the call is forward or backward).

Thus Stam's conjecture indicates that the JH compression function G must have significantly weaker than birthday collision resistance. It is indeed easy to see that, on average, only $2^{n/4}$ queries are required to find a collision for G, since all one needs is to find a collision on the top half of output (the bottom half can then be adjusted via the input wire). Steinberger's theorem (and our own result as well) shows that *any* compression function with parameters $(m, n, r, s) = (0.5n, n, 1, n)$, regardless of its design, will likewise have collision resistance at most $2^{n/4}$ (up to a small constant factor). Observe, also, that once a single collision is obtained for G, $2^{n/2}$ collisions are obtained at once, since we can replicate the collision with any value on the bottom output wire. Our own theorem, beyond showing that collision resistance cannot exceed $2^{n/4}$, predicts this threshold behavior as well. More precisely, we show that for *any* compression function with parameters $(m, n, r, s) = (0.5n, n, 1, n)$, an adversary making at most[5] $200 \cdot 2^{n/4}$ queries to f can obtain

$$2^{s-\frac{2(nr-m)}{r+1}} = 2^{n-\frac{2(n-0.5n)}{2}} = 2^{0.5n}$$

collisions[6] with high probability. On the other hand, we emphasize that *no* collisions are obtained for G with $\frac{1}{10} \cdot 2^{n/4}$ queries (for any adversary, w.h.p.). Thus we not only pinpoint the collision resistance up to a constant factor, but we also pinpoint the exact "payoff" that occurs once collision resistance is breached.

As a second example, in [8] Stam exhibits a compression function with parameters $(m, n, r, s) = (n, n, 2, n)$ of collision resistance $2^{n/3}$ (Stam's bound). Stam's compression function has the particularity that $n/3$ bits are simply forwarded untouched from the input the output, whereas the remaining $2n - n/3 = \frac{5}{3}n$ input bits are cryptographically processed into the remaining $n - n/3 = \frac{2}{3}n$ output bits. Obviously, for such a compression function, $2^{n/3}$ collisions are obtained once a single collision is obtained. Our own result shows this sudden jump (from no collisions to $2^{n/3}$ collisions) is essentially unavoidable, in the sense that with

$$16000 \cdot 2^{\frac{nr-m}{r+1}} = 16000 \cdot 2^{\frac{2n-n}{3}} = 16000 \cdot 2^{n/3}$$

queries an adversary can obtain

$$2^{s-\frac{2(nr-m)}{r+1}} = 2^{n-\frac{2(2n-n)}{3}} = 2^{n/3}$$

collisions with high probability, and this for *any* compression function with parameters $(m, n, r, s) = (n, n, 2, n)$.

ORGANIZATION. Section 2 contains relevant definitions and conventions. Section 3 states and briefly discusses our main result. Section 4 gives an overview of the proof, and briefly compares our proof techniques to those of Steinberger [9]. Section 5 finally discusses some of the interesting open problems that remain. The actual proof of our main theorem is in Appendix A, with supporting lemmas to be found in Appendix B.

2 Definitions and Preliminaries

COMPRESSION FUNCTIONS. Let $m \geq 0$ be a real number and let $s \geq 0$ be an integer. (Our results hold as stated even when $s \geq 0$ is a real number such that 2^s is an integer and, likewise, also when $n \geq 0$ is a real number such that 2^n is an integer. However, for notational and conceptual simplicity, we shall assume n, s are integers.) By "a function of domain $\{0,1\}^{s+m}$" we mean a function with a domain of size

[5] When $r = 1$ the multiplicative constant can be improved from 16000 to 200. See Section 3 for more details.

[6] Traditionally, the "number of collisions" means the "number of distinct pairs of inputs that collide". Note, however, that under this definition $2^{0.5n}$ "collisions" may be caused by only $2^{0.25n}$ inputs, all involved in one big multi-collision. We show, in fact, that the *number of different inputs involved in a collision* is at least $2^{0.5n}$, which constitutes an even stronger result.

$\lceil 2^{s+m} \rceil$—the exact nature of the domain will not matter for our results, but for notational convenience we still write the domain as $\{0,1\}^{s+m}$ (even though m is not necessarily an integer and, furthermore, even though 2^{s+m} is not necessarily an integer). Readers who feel uneasy about this convention may think of $\{0,1\}^{s+m}$ as being a shorthand for some fixed subset of $\{0,1\}^{\lceil s+m \rceil}$ of size $\lceil 2^{s+m} \rceil$.

Let now $m \geq 0$ be a real number and let $r \geq 1$, $n, s \geq 0$ be integers. We formalize the notion of a compression function $F : \{0,1\}^{s+m} \to \{0,1\}^s$ making r calls to a primitive f of domain $\{0,1\}^n$.

In fact we allow F to call potentially distinct primitives $f_1, \ldots, f_r$ in *fixed order mode*, meaning f_i is called before f_j for $i < j$. Let $f_1, \ldots, f_r$ be (not necessarily distinct) functions of domain $\{0,1\}^n$ and range $\{0,1\}^b$, where b is arbitrary. The compression function $F : \{0,1\}^{m+s} \to \{0,1\}^s$ is defined by r functions $g_1, \ldots, g_r$ where $g_i : \{0,1\}^{m+s} \times \{0,1\}^{b(i-1)} \to \{0,1\}^n$ and a function $h : \{0,1\}^{m+s} \times \{0,1\}^{br} \to \{0,1\}^s$. We then define $F(v) = h(v, y_1, \ldots, y_r)$ where $y_j = f_j(g_j(v, y_1, \ldots, y_{j-1}))$ for $j = 1 \ldots r$. We call the values $y_1, \ldots, y_r$ *intermediate chaining variables* and we refer to the functions $g_1, \ldots, g_r$ as the *intermediate processing functions*. We note that $g_1, \ldots, g_r$ are, for a given construction, fixed finite functions with a public description.

We say an adversary A with oracle access to $f_1, \ldots, f_r$ "knows the first k chaining variables" for some input $v \in \{0,1\}^{m+s}$ when A has made the queries $f_1(g_1(v)) = y_1$, $f_2(g_2(v, y_1)) = y_2$, $\ldots$, $f_k(g_k(v, y_1, \ldots, y_{k-1})) = y_k$, where $0 \leq k \leq r$. In this case, we also say A "knows the relevant queries to $f_1, \ldots, f_k$" for v.

When F is as defined above we call F an "(m, n, r, s) compression function". By default, the primitives called by such a compression function are always named $f_1, \ldots, f_r$ (in order).

Collision accounting. The following definition is somewhat nonstandard, but central to the paper:

Definition 1. *Let $F : D \to R$ be a function of domain D and range R. Let $S \subseteq D$. The set of* colliding inputs *in S (with respect to F) is the set*

$$\{x \in S : \exists y \in S, y \neq x, \ \text{s.t. } F(x) = F(y)\}.$$

Let F be an (m, n, r, s) compression function calling primitives $f_1, \ldots, f_r$. Let A be an adversary with oracle access to $f_1, \ldots, f_r$. The set of inputs *learned* by A is the set of inputs $S \subseteq \{0,1\}^{s+m}$ for which A has made the relevant queries to $f_1, \ldots, f_r$ at the end of its attack (and therefore, for which A knows the value of F). The set of *colliding inputs obtained by A* is the set $C \subseteq S$ of colliding inputs in S, with respect to F. We say A *obtains z colliding inputs* if $|C| \geq z$.

It is worth noting that $|C| \geq |S| - |R|$, given that only $|R|$ elements of S can occupy their "own" slots in the range. Thus an adversary that learns $|S|$ inputs for a compression function of range size $|R|$ automatically obtains at least $|S| - |R|$ colliding inputs.

Yield. The following basic observation is due to Rogaway and Steinberger [6]:

Lemma 1. *Let $F : \{0,1\}^{m+s} \to \{0,1\}^s$ be a compression function calling primitives $f_1, \ldots, f_r : \{0,1\}^n \to \{0,1\}^b$ in fixed-order mode. Then there exists an adversary that with at most q queries to each f_i can learn the first i intermediate chaining variables for at least*

$$2^{m+s}\left(\frac{q}{2^n}\right)^i$$

inputs, for $0 \leq i \leq r$.

In other words, there exists an adversary making at most q queries to each f_i, and for which

$$|S_i| \geq 2^{m+s}\left(\frac{q}{2^n}\right)^i$$

for $0 \leq i \leq r$, where $S_i \subseteq \{0,1\}^{m+s}$ is the set of inputs for which the relevant queries to $f_1, \ldots, f_i$ have been made. The adversary in question is very straightforward: it is a greedy adversary that starts by choosing its queries to f_1 such as to maximize the size of S_1, then, after making its queries to f_1, chooses its queries to f_2 such as to maximize the size of $S_2 \subseteq S_1$, and so on. For a full proof see any of [6], [8] or [9].

Setting $i = r$ in Lemma 1 we obtain the following corollary:

Corollary 1. *Let $F : \{0,1\}^{m+s} \to \{0,1\}^s$ be a compression function calling primitives $f_1, \dots, f_r : \{0,1\}^n \to \{0,1\}^b$ in fixed-order mode. Then with q queries to each f_i, an adversary can learn to evaluate F on at least*

$$2^{m+s}\left(\frac{q}{2^n}\right)^r$$

inputs.

3 Results

The following Theorem dispatches the "easy" cases of Stam's conjecture; similar results are already given in [6, 8, 9].

Theorem 1. (cf. [6, 8, 9]) *Let F be an (m, n, r, s) compression function with $m \geq 1$. Then: (i) if $s/2 \leq (nr - m)/(r + 1)$, a collision can be found for F with at most*

$$q = 2\sqrt{2} \cdot 2^{s/2} + 1 \leq 2\sqrt{2} \cdot 2^{\frac{nr-m}{r+1}} + 1$$

queries to each f_i, with probability at least 0.5; and (ii) if $m \geq nr$, a collision can be found for F with at most 2 queries to each f_i, with probability 1.

Proof. Statement (i) follows by a birthday attack and the fact that $m \geq 1$ (so that the domain of F has size at least twice the range); see [9, 10] for more details. Statement (ii) follows from by applying Corollary 1 with $q = 2$, and noting that when $m \geq nr$ we have

$$2^{m+s}\left(\frac{2}{2^n}\right)^r \geq 2 \cdot 2^s$$

so that, automatically, at least $2^{s+1} - 2^s = 2^s$ colliding inputs are obtained by the adversary. □

In light of Theorem 1, our remaining results are restricted to the case $(s/2 \geq (nr-m)/(r+1) \wedge m \leq nr)$. It is worth noting that $2^{(nr-m)/(r+1)} \geq 1$ when $m \leq nr$, since then $(nr - m)/(r + 1) \geq 0$.

To state and discuss our main result it will be convenient to define the function

$$\gamma(r, c) = 2e^{-c^2/5760} + \sum_{i=1}^{r-1} 2e^{-\frac{1}{32}(\frac{c}{80})^i}$$

where $r \geq 1$ is an integer and $c > 0$ is an arbitrary real number. We keep this definition of $\gamma(r, c)$ for the rest of the paper.

Our main result is the following:

Theorem 2. *Let F be an (m, n, r, s) compression function with $1 \leq m \leq nr$ and $s/2 \geq (nr-m)/(r+1)$. Let $c > 0$ be a real number such that $c2^{\frac{nr-m}{r+1}}$ is an integer. Then there exists an adversary making at most*

$$q = 2c2^{\frac{nr-m}{r+1}}$$

queries to each f_i and obtaining at least

$$2^{s-\frac{2(nr-m)}{r+1}}$$

colliding inputs, with probability at least $1 - \gamma(r, c)$.

For $r = 1$ one can compute that $\gamma(1, 90) < 0.5$, whereas $\gamma(1, 100) < 0.36$ and $\gamma(1, 1000) < e^{-170}$. Thus $180\lceil 2^{\frac{nr-m}{r+1}} \rceil$ queries suffice to obtain a collision with probability at least 0.5. For $r > 1$, one can make the observation that

$$\gamma(r, c) \le 2e^{-c^2/5760} + \sum_{i=1}^{\infty} 2e^{-\frac{1}{32}(\frac{c}{80})^i}$$

where the right-hand side does not depend on r, and where the right-hand side is less than 0.5 for $c \ge 8000$. Thus $16000r\lceil 2^{\frac{nr-m}{r+1}} \rceil$ queries (in total to all f_i's) suffice to find a collision with probability at least 0.5 when $r > 1$.

The full proof of Theorem 2 is given in Appendix A, whereas the next section gives an overview of the proof's high-level structure.

4 Proof Overview

In this section we give an overview of the proof of Theorem 2. We emphasize that this section's contents constitute *intuition only* and have nil mathematical value. This material is included purely for pedagogical and readability purposes and the actual proof of Theorem 2, given in Appendix A, is entirely independent and self-contained.

A central ingredient in our proof is a lemma on collisions (Lemma 6 in Appendix B) that we start by paraphrasing here in order to facilitate the following discussion. Let $T_1, \dots, T_k$ be disjoint sets whose sizes are upper bounded by some constant M, let $F : T \to R$ be some function where $T = T_1 \cup \dots \cup T_k$, and let C be the total number of colliding inputs in T with respect to F. Note that if we select q of the k sets $T_1, \dots, T_k$ at random, and form a set T' as the union of the q selected sets, then each point of T has probability $p := q/k$ of ending up in T'. Since a colliding input $x_0 \in T$ has probability at least

$$\frac{q}{k}\frac{q-1}{k-1} \approx p^2$$

of winding up as a colliding input in T' (because x_0 must be selected for T' and also at least one of the other points[7] in T with which x_0 collides must be selected for T'), we can therefore expect T' to have approximately at least

$$p^2 C$$

colliding inputs. Roughly speaking, Lemma 6 states that as long as this expectation is a fair amount larger than M (the maximum size of the T_i's) then this intuition is borne out, and the number of colliding inputs in T' is not much less that p^2C with high probability. We point out that Lemma 6 does not "know" how to take advantage of multi-collisions: if a colliding input $x_0 \in T$ collides with very many other points in T, coming from many different T_j's, then x_0's chance of being a colliding input in T' will be significantly greater than p^2. Thus Lemma 6 does not give a sharp result in all situations. This lack plays a role in the proof sketch below (as well as in the proof itself).

For the proof sketch we start by reviewing some specific settings of the parameters and explain, in each case, how our collision-finding adversary operates, and why it can hope to find the desired number of collisions within the limits of Stam's bound. We call these "case studies". We later abstract more general observations from these case studies. (The first two case studies concern parameter settings that are already covered by Steinberger's [9] results. However, the point is to flesh out our line of attack, which is completely different from Steinberger's birthday-based approach.)

Conceptually we emphasize that, unlike for a typical collision resistance analysis in the provable security setting, it is more useful to view the primitives $f_1, \dots, f_r$ as being sampled (from whatever

[7] If such an other point comes from the same set T_i that contains x_0, this only helps us, in the sense that x_0 then has chance exactly p (the chance that T_i is selected) of becoming a colliding input in T'.

distribution) *before* the start of the collision resistance experiment, rather than as being lazy sampled. Thus one can think of the primitives $f_1, \ldots, f_r$ as functions that are "fixed but arbitrary", and to which the adversary has oracle access.

First Case Study: $(m, n, r, s) = (0.5n, n, 1, n)$. Let $F : \{0,1\}^{m+s} \to \{0,1\}^n$ be a compression function making a single call to an n-bit primitive f_1, where $(m, n, r, s) = (0.5n, n, 1, n)$. Thus $F : \{0,1\}^{1.5n} \to \{0,1\}^n$. We note this setting of parameters coincides, for example, with the parameters of the JH compression function (discussed in the introduction).

Stam's bound indicates that

$$q = 2^{\frac{nr-m}{r+1}} = 2^{\frac{n-0.5n}{2}} = 2^{n/4}$$

queries to f_1 should suffice to find collisions for F. Let $S_0 = \{0,1\}^{m+s}$ be F's domain, and write

$$S_0 = \bigcup_{y \in \{0,1\}^n} U_y^0$$

where

$$U_y^0 = \{x \in S_0 : g_1(x) = y\}$$

where g_1 is F's first and only intermediate processing function (see Section 2). We note that S_0 is the disjoint union of the sets U_y^0. Moreover, the collision adversary knows each set U_y^0, since g_1 is public. We also note that the *average size* of the sets $\{U_y^0 : y \in \{0,1\}^n\}$ is

$$\frac{|S_0|}{2^n} = \frac{2^{m+s}}{2^n} = \frac{2^{1.5n}}{2^n} = 2^{n/2}.$$

For simplicity we start by assuming that $|U_y^0| = 2^{n/2}$ for all $y \in \{0,1\}^n$. We will discuss later how to lift this assumption.

The adversary's most natural strategy is to make $2^{n/4}$ random queries to f_1. Let $\mathcal{B} \subseteq 2^{n/4}$ denote the set of values so queried to f_1, and set

$$S_1 = \bigcup_{y \in \mathcal{B}} U_y^0.$$

Then $S_1 \subseteq \{0,1\}^{m+s}$ is the set of inputs for which the relevant query to f_1 is known. Note that $|S_1| = 2^{n/4} \cdot 2^{n/2} = 2^{3n/4}$ by our assumption that each set U_y^0 has size $2^{n/2}$.

We could try, at this point, to estimate the number of colliding inputs in S_1 using Lemma 6, applied with $T = S_0$ and $T' = S_1$, where the sets $T_1, \ldots, T_k$ correspond to the family of sets $\{U_y^0 : y \in \{0,1\}^n\}$ (which form a disjoint partition of S_0), and where $M = 2^{n/2}$ is the upper bound on the size of the T_i's. Here $k = 2^n$ and, therefore, $p = q/k = 2^{n/3}/2^n = 2^{-2n/3}$. The number of colliding inputs C in $T = S_0$ is at least $|S_0| - 2^s = 2^{1.5n} - 2^n \approx 2^{1.5n}$. We therefore find that

$$p^2 C \approx 2^{-4n/3} 2^{1.5n} = 2^{n/6}.$$

Unfortunately, this number is not as large as $M = 2^{n/2}$ and, in such a case, Lemma 6 does not deliver anything meaningful. We are running up against the afore-mentioned shortcoming of Lemma 6, since we are in a case where the average colliding input in $T = S_0$ does not only collide with $O(1)$ other elements in T, but with very many other elements (or more precisely with $|S_0|/2^s = 2^{n/2}$ other elements).

We overcome this obstacle with a trick. We divide the adversary's querying process into two phases. In the first phase, the adversary selects (deterministically, say) a subset I of $\{0,1\}^n$ of size $2^{n/2+1}$. In the second phase, the adversary selects a set $\mathcal{B} \subseteq I$ of size $2^{n/4}$ uniformly from all such subsets of I, and queries the elements of $\mathcal{B}$ to f_1. We emphasize that the elements of I not in $\mathcal{B}$ are not queried to

f_1. Clearly, applying this two-step process is equivalent to directly selecting $2^{n/4}$ values $\mathcal{B}$ uniformly at random from $\{0,1\}^n$ and querying them to f_1.

Let

$$S_I = \bigcup_{y \in I} U_y^0.$$

Thus $S_0 \supseteq S_I \supseteq S_1$. Moreover $|S_I| = 2^{n/2+1} \cdot 2^{n/2} = 2^{n+1} = 2^{s+1}$, so S_I contains at least $2^{s+1} - 2^s = 2^s$ colliding inputs. (Note, crucially, that every colliding input in S_I *might* very well collide with only one other input in S_I, so that we are no longer in a case in which Lemma 6 is ignoring a key statistic.) We now apply Lemma 6 with $T = S_I$ and $T' = S_1$, where the sets $T_1, \ldots, T_k$ correspond, this time, to the family of sets $\{U_y^0 : y \in I\}$ (which form a partition of S_I). Thus $k = |I| = 2^{n/2+1}$. We have $q/k = 2^{n/4}/2^{n/2+1} \approx 2^{-n/4}$ and

$$p^2 C \approx 2^{-2n/4} 2^s = 2^{n/2}$$

where $C \geq 2^s$ is the number of colliding inputs in S_I. Thus, this time, p^2C is commensurate with the upper bound $M = 2^{n/2}$ on the size of the T_i's, and so Lemma 6 can be effectively applied. (To be a little more precise, by making, say, $200 \cdot 2^{n/2}$ queries to f_1 instead of $2^{n/2}$ queries to f_1, we can push C and p^2C to significantly higher than M, which remains capped at $2^{n/2}$. Moreover we can make $\frac{1}{20}p^2C \geq 2^{n/2}$ so that, by Lemma 6, we actually obtain $2^{n/2}$ colliding inputs with high probability.)

We emphasize that the above argument does not require f_1 to be "random" at all; f_1 can be *any* fixed function. The only randomness occurs in the selection of the set $\mathcal{B}$ of queries to f_1.

Finally, the "well-balancedness" assumption on the sets U_y^0 can be removed by using a common refinement of these sets. More precisely, by Lemma 2 in Appendix A, one can always refine the collection of sets $\{U_y^0 : y \in \{0,1\}^n\}$ into a collection of sets each of size at most $2^{n/2}$, at the cost of increasing the number of sets by a factor of at most 2. We then view the adversary as "querying" sets in this refinement (each such set is a subset of a particular U_y^0 and, therefore, associated to a particular value of y). This process may result in redundant queries to f_1 (when two or more subsets of the same U_y^0 are chosen to be queried), but this is harmless. In particular, we do not care about the fact that such redundant queries to f_1 produce dependent results—indeed, from the proof's standpoint, f_1 is anyway an arbitrary fixed function containing no entropy.

Second Case Study: $(m, n, r, s) = (n, n, 2, n)$. Let $F : \{0,1\}^{m+s} \to \{0,1\}^n$ be a compression function making calls to two n-bit primitives f_1 and f_2 in fixed-order mode, where $(m, n, r, s) = (n, n, 2, n)$. Thus $F : \{0,1\}^{2n} \to \{0,1\}^n$. As usual, let g_1 and g_2 be the intermediate processing functions for F.

Stam's bound forecasts a collision resistance of

$$q = 2^{\frac{nr-m}{r+1}} = 2^{\frac{2n-n}{3}} = 2^{n/3}$$

queries to each f_1 and f_2.

Let $S_0 = \{0,1\}^{m+s}$ and let

$$U_y^0 = \{x \in S_0 : g_1(x) = y\}.$$

The adversary starts by querying $f_1(y)$ for the q values y for which $|U_y^0|$ is largest. (Note this is a deterministic step.) Then

$$|S_1| \geq 2^{m+s}\left(\frac{q}{2^n}\right) = 2^{2n}\frac{2^{n/3}}{2^n} = 2^{4n/3}$$

where $S_1 \subseteq \{0,1\}^{m+s}$ is the set of inputs for which the relevant query to f_1 has been made. Note that $2^{4n/3} \gg 2^n = 2^s$, so S_1 contains many colliding inputs (more precisely, at least $2^{4n/3} - 2^n \approx 2^{4n/3}$) with probability 1. Moreover, depending on the structure of F and of f_1, there is no reason one could expect to beat this number of colliding inputs (in S_1) by using a randomized query strategy to f_1 instead of a greedy query strategy to f_1.

For simplicity, we will assume that S_1 has size exactly $2^{4n/3}$ (anyway the adversary could choose to "throw out" or ignore elements of S_1 to reduce the effective size of S_1 to $2^{4n/3}$, if desired).

At this point, before queries to f_2 are made, note that we are essentially reduced to attacking a compression function F' with paramaters $(m', n', r', s') = (n/3, n, 1, n)$ whose domain is S_1, where $|S_1| = 2^{m'+s'} = 2^{4n/3}$. For such a compression function, Stam's bound quotes a collision resistance of

$$2^{\frac{n'r'-m'}{r'+1}} = 2^{\frac{n-n/3}{2}} = 2^{n/3}$$

queries, which is exactly our budget query for f_2. We have thus reduced the parameter setting $(m, n, r, s) = (n, n, 2, n)$ of Stam's conjecture to the parameter setting $(n/3, n, 1, n)$, namely to a case of Stam's conjecture where $r = 1$. (This type of reduction was first brought to attention by Steinberger [9].) What follows is therefore fairly similar to the first case study for the parameters $(m, n, r, s) = (0.5n, n, 1, n)$.

Let

$$U_y^1 = \{x \in S_1 : g_2(x, y_1) = y\}$$

for each $y \in \{0,1\}^n$, where, above, $y_1 = f_1(g_1(x))$ is the first intermediate chaining variable for x (implicitly dependent on x). For simplicity, we can assume that $|U_y^1| = |S_1|/2^n = 2^{n/3}$ for all $y \in \{0,1\}^n$ (this assumption can be lifted by using a refinement of the U_y^1's, as in the previous case study). To make his queries to f_2, the adversary starts by (deterministically) selecting a set $I \subseteq \{0,1\}^n$ of size $2^{2n/3+1}$. Let

$$S_I = \bigcup_{y \in I} U_y^1$$

so that $|S_I| = 2^{2n/3+1} \cdot 2^{n/3} = 2^{n+1} = 2^{s+1}$. Thus S_I contains at least 2^s colliding inputs. The adversary then selects a random subset $\mathcal{B}$ of I of size $q = 2^{n/3}$, and queries f_2 at all the points in $\mathcal{B}$. Let

$$S_2 = \bigcup_{y \in \mathcal{B}} U_y^1$$

so that $S_2 \subseteq S_1$ is the set of inputs for which the relevant queries to f_1 and f_2 are both known. Applying Lemma 6 with $T = S_1$, $T' = S_2$, and with sets $T_1, \ldots, T_k$ corresponding to $\{U_y^1 : y \in I\}$, where $k = |I| = 2^{2n/3+1}$ and $|T_i| \le M := 2^{n/3}$ for all i, we find that $p = q/k = 2^{n/3}/2^{2n/3+1} \approx 2^{-n/3}$ and

$$p^2 C \approx 2^{-2n/3} 2^s = 2^{n/3}$$

where $C \ge 2^s$ is the number of colliding inputs in S_I. Since the latter quantity is commensurate with M, we can effectively apply Lemma 6 to conclude that we will obtain $2^{n/3}$ colliding inputs in S_2 with high probability (by making some constant factor more queries than $2^{n/3}$).

Third Case Study: $(m, n, r, s) = (1.25n, n, 4, 2n)$. Let $F : \{0,1\}^{m+s} \to \{0,1\}^n$ be a compression function making calls to four n-bit primitives $f_1, \ldots, f_4$ in fixed-order mode, where $(m, n, r, s) = (1.25n, n, 4, 2n)$. In this case, therefore, $F : \{0,1\}^{3.25n} \to \{0,1\}^{2n}$.

Stam's bound places collision resistance at

$$q = 2^{\frac{nr-m}{r+1}} = 2^{\frac{4n-1.25n}{5}} = 2^{0.55n}$$

queries to each of the primitives f_1, f_2, f_3 and f_4 (which, we note, is less than the cost of a birthday attack).

Let $S_0 = \{0,1\}^{m+s}$ and let $U_y^0 = \{x \in S_0 : g_1(x) = y\}$ for all $y \in \{0,1\}^n$. The adversary starts by querying $f_1(y)$ for the q values y for which $|U_y^0|$ is largest. Then

$$|S_1| \ge 2^{m+s} \left(\frac{q}{2^n} \right) = 2^{3.25n} \frac{2^{0.55n}}{2^n} = 2^{2.8n}$$

where S_1 is the set of inputs for which queries to f_1 have been made.

After the queries to f_1 are completed, let $U_y^1 = \{x \in S_1 : g_2(x, y_1) = y\}$ where y_1 is the first intermediate chaining variable for x. For f_2 the adversary *again* makes greedy queries, i.e. queries $f_2(y)$ for the q values y for which $|U_y^1|$ is largest. Then

$$|S_2| \geq |S_1| \left(\frac{q}{2^n}\right) = 2^{2.8n} \frac{2^{0.55n}}{2^n} = 2^{2.35n}$$

where $S_2 \subseteq S_1$ is the set of inputs for which queries to both f_1 and f_2 have been made. Note S_2, like S_1 is still larger than $2^s = 2^{2n}$; thus we are "automatically" assured the presence of colliding inputs in S_1 and S_2 by virtue of the size of these sets, which accounts for the sufficiency of the greedy approach.

After the queries to f_2 are completed, let $U_y^2 = \{x \in S_2 : g_3(x, y_1, y_2) = y\}$, where y_1, y_2 are the first two intermediate chaining variables for x. At this point, if the adversary were again to apply a greedy strategy for f_3, we would find a lower bound of

$$|S_2| \left(\frac{q}{2^n}\right) = 2^{2.35n} \frac{2^{0.55n}}{2^n} = 2^{1.9n}$$

on the size of S_3, which is no longer larger than 2^s. Applying a (deterministic) greedy strategy would therefore be a very bad idea, since one could easily set up F and its primitives $f_1, \ldots, f_4$ so that S_3 contains no colliding inputs with probability 1, and the adversary finds collisions with probability 0.

Instead, at this stage we revert to using Lemma 6 and the two-step "trick" involving the set I. Assume for simplicity (and in fact without loss of generality) that $|U_y^2| = 2^{2.35n}/2^n = 2^{1.35n}$ for all $y \in \{0,1\}^n$. The adversary starts by deterministically selecting a set $I \subseteq \{0,1\}^n$ of size $2^{0.65n+1}$. Let

$$S_I = \bigcup_{y \in I} U_y^2.$$

Thus $|S_I| = 2^{0.65n+1} 2^{1.35n} = 2^{2n+1} = 2^{s+1}$. (We note the adversary *has been deterministic up to now*—namely the adversary remains deterministic as long as the underlying set of known inputs contains colliding inputs simply by virtue of its size. Now the adversary is about to switch to, and stick with, a randomized strategy.) The adversary then randomly selects a set $\mathcal{B} \subseteq I$ of size $q = 2^{0.55n}$, and queries these values to f_3. We set

$$S_3 = \bigcup_{y \in \mathcal{B}} U_y^2.$$

We apply Lemma 6 with $T = S_I$, $T' = S_3$, $\{T_i : 1 \leq i \leq k\} = \{U_y^2 : y \in I\}$, $k = |I| = 2^{0.65n+1}$, $M = 2^{1.35n}$, $p = q/k = 2^{0.55n}/2^{0.65n+1} \approx 2^{-0.1n}$ and $C \geq |S_I| - 2^s \geq 2^s = 2^{2n}$, so that

$$p^2 C \geq 2^{-0.2n} 2^{2n} = 2^{1.8n}.$$

In particular, $p^2 C \gg M$, so Lemma 6 can be effectively applied to show that the number of colliding inputs in S_3 is not much less than $p^2 C = 2^{1.8n}$. For simplicity, we will assume the number of colliding inputs in S_3 is exactly $2^{1.8n}$. Moreover, note that $|S_3| = 2^{1.35n} 2^{0.55n} = 2^{1.9n}$ by virtue of our assumption that $|U_y^2| = 2^{1.35n}$ for each $y \in \{0,1\}^n$.

For queries to f_4, the adversary directly continues with a randomized strategy and an application of Lemma 6—no need for a preliminary selection of inputs I, here, because the number of colliding inputs in S_3 is already less than 2^s.

More precisely, let $U_y^3 = \{x \in S_3 : g_4(x, y_1, y_2, y_3) = y\}$ for all $y \in \{0,1\}^n$, where y_1, y_2, y_3 are the intermediate chaining variables for x. Assume for simplicity that $|U_y^3| = |S_3|/2^n = 2^{1.9n}/2^n = 2^{0.9n}$ for all y. The adversary selects a random set $\mathcal{B} \subseteq \{0,1\}^n$ of size $q = 2^{0.55n}$, and queries these values to f_4. Set

$$S_4 = \bigcup_{y \in \mathcal{B}} U_y^3.$$

We apply Lemma 6 with $T = S_3$, $T' = S_4$, $\{T_i : 1 \le i \le k\} = \{U_y^3 : y \in \{0,1\}^n\}$, $k = |I| = 2^n$, $M = 2^{0.9n}$, $p = q/k = 2^{0.55n}/2^n = 2^{-0.45n}$ and $C = 2^{1.8n}$, where the latter equality comes from our simplifying assumption that S_3 contains exactly $2^{1.8n}$ colliding inputs. Then

$$p^2C = 2^{-0.9n}2^{1.8n} = 2^{0.9n}$$

so that p^2C is commensurate with $M = 2^{0.9n}$, and Lemma 6 can be effectively applied (after, potentially, multiplying the number of queries by some small constant) to show that at least

$$p^2C = 2^{0.9n} = 2^{2n-1.1n} = 2^{s-\frac{2(nr-m)}{r+1}}$$

colliding inputs can be obtained with good probability.

DIGEST. The last case study exhibits more or less all the features of the general case. In the general case, the adversary's querying strategy has two phases. The first phase is a "deterministic" phase where the adversary makes greedy queries to maximize the yield. This phase lasts as long the next set S_i obtained is guaranteed to be larger than 2^s. This phase also "spills over" into the (still deterministic) selection of the set I. The second phase then commences, consisting of purely random queries. (First q random queries selected from I and then, for subsequent f_i's, q random queries selected from $\{0,1\}^n$.) It so turns out that the "phase change" occurs exactly when it is time to make queries to f_{r_0+1} where

$$r_0 = \left\lfloor \frac{m(r+1)}{m+n} \right\rfloor.$$

Thus, in the general case, the two-phase strategy determines a sequence of sets

$$S_0 \supseteq S_1 \supseteq \cdots \supseteq S_{r_0} \supseteq S_I \supseteq S_{r_0+1} \supseteq \cdots \supseteq S_r$$

where $S_0 = \{0,1\}^{m+s}$ is F's domain and S_i, $i \ge 1$, is the set of inputs for which the queries to $f_1, \ldots, f_i$ have been made. (When $\frac{m(r+1)}{m+n}$ happens to be an integer—which does not occur in any of the case studies above—then $r_0 = \frac{m(r+1)}{m+n} \ge 1$ and, by adding a constant factor to the number of queries, one finds $|S_{r_0}| \ge 2^{s+1}$ instead of $|S_{r_0}| = 2^s$, so that there is "still room" for S_I to be selected.)

One can point out that the number of colliding inputs in the sets $S_0, \ldots, S_r$ evolves differently during the first and second phases. During the first phase, each colliding input in S_i collides on average with a very large number of other points, so that the key factor determining whether a colliding input makes it from S_i to S_{i+1} (assuming $i+1 \le r_0$) is just whether that particular point makes it to S_{i+1} (since it is very likely that at least one of the myriad other points it collides with has made it to S_{i+1} as well). The "rate of attrition" of colliding inputs is therefore $p = q/2^n$ in going from S_i to S_{i+1}, for $i+1 \le r_0$, and, similarly, is $|I|/2^n$ in going from S_{r_0} to S_I. During the second phase, on the other hand, both a colliding input *and the (on average unique) other input it collides with* must simultaneously survive the selection process, so that the rate of attrition of colliding inputs in going from S_I to S_{r_0+1} is $(q/|I|)^2$ whereas the rate of attrition in going from S_i to S_{i+1} is $(q/2^n)^2$ for $r_0 + 1 \le i \le r-1$. It is possible to compute that these rates of attrition lead to a final expected number of colliding inputs equal to $2^s - \frac{2(nr-m)}{r+1}$. The latter also equals, by no coincidence, $|S_{r-1}|/2^n$.

COMPARISON WITH [9]. The proof of Lemma 6—our paper's "key lemma"—uses ideas from Steinberger's "MECMAC lemma" [9] (a lemma which is actually unused in the main result of [9]), and more precisely recycles the nice idea of using a bipartition of sets to overcome dependencies between collision events. Our work also uses Steinberger's parameter reduction idea (as discussed in the second case study). However, these are essentially the only similarities with [9]. In particular, our proof does not consist in a generalization of Steinberger's techniques, since our proof, as restricted to $r = 1$, does not reduce to a birthday attack, but instead uses Lemma 6 which itself relies on Martingale concentration results. Moreover, the key idea of focusing on the number of colliding inputs (as opposed to the more usual "number of colliding pairs of inputs") as the correct metric for measuring the progress of an attack is an original contribution of this paper.

5 Future Work

Many related interesting open problems remain. One of the basic questions that remains is to show Stam's bound is tight. This would require exhibiting an infinite class of compression functions (parameterized by m, n, r, s, where m, r and s are linear functions of n) whose collision resistance is provably in the vicinity of

$$\min(2^{s/2}, \lceil 2^{\frac{nr-m}{r+1}} \rceil).$$

Another remaining open question concerns parallelism. Could better attacks be found for compression functions that call their primitives in parallel? So far, rather amazingly, we are not aware of any provable separation between the power of parallel and sequential compression functions. A third type of question concerns adapting results like those in this paper to compression functions with primitives of not-all-equal input lengths and, maybe more interestingly, to primitives with small output lengths. Indeed, primitives with small output lengths constitute a vulnerability, as pointed out by Stam [8], though a classification and quantification of such vulnerabilities still awaits.

References

1. Mihir Bellare and Tadayoshi Kohno, *Hash function imbalance and its impact on birthday attacks*, EUROCRYPT 2004, LNCS 3027, pp. 401–418.
2. John Black, Martin Cochran, Thomas Shrimpton, *On the Impossibility of Highly-Efficient Blockcipher-Based Hash Functions*, EUROCRYPT 2005, LNCS 3494, pp. 526–541.
3. Fan Chung and Linyuan Lu, *Concentration Inequalities and Martingale Inequalities: A Survey.* Internet Mathematics Vol. 3, No. 1, pp. 79–127.
4. C. McDiarmid. "Concentration." In *Probabilistic Methods for Algorithmic Discrete Mathematics,* pp. 195–248, edited by M. Habib, C. McDiarmid, J. Ramier-Alfonsin and B. Reed, Algorithms and Combinatorics 16. Springer, 1998.
5. Phillip Rogaway and John Steinberger, *Constructing cryptographic hash functions from fixed-key blockciphers*, CRYPTO 2008, LNCS 5157, pp. 433–450.
6. Phillip Rogaway and John Steinberger, *Security/Efficiency Tradeoffs for Permutation-Based Hashing*, in EUROCRYPT 2008, LNCS 4965, pp. 220–236.
7. Thomas Shrimpton and Martijn Stam, *Building a Collision-Resistant Compression Function from Non-Compressing Primitives*, ICALP 2008, pp. 643–654. Also available at the Cryptology ePrint Archive: Report 2007/409.
8. Martijn Stam, *Beyond uniformity: Better Security/Efficiency Tradeoffs for Compression Functions*, CRYPTO 2008, LNCS 5157, pp. 397–412.
9. John Steinberger, *Stam's collision resistance conjecture*, EUROCRYPT 2010, LNCS 6610, pp. 597–615.
10. Michael Wiener, *Bounds on birthday attack times*, Cryptology ePrint archive, 2005.
11. Hongjun Wu, *The JH hash function*, NIST SHA-3 competition submission, October 2008.

A Proof of Theorem 2

We start with a quick definition and lemma.

Definition 2. *Let $U_1, \ldots, U_\ell$ be a collection of finite disjoint sets. Then another collection $T_1, \ldots, T_k$ of finite disjoint sets is a* refinement *of $U_1, \ldots, U_\ell$ if $\bigcup_{i=1}^k T_i = \bigcup_{i=1}^\ell U_i$ and if either $T_i \subseteq U_j$ or $T_i \cap U_j = \emptyset$ for all $1 \le i \le k$, $1 \le j \le \ell$.*

Lemma 2. *Let $U_1, \ldots, U_\ell$ be disjoint finite sets. Let $M \ge 1$ be a positive integer upper bounding the average size of the U_i's. (That is, $M \ge (\sum_i |U_i|)/\ell$.) Then there exists a refinement $T_1, \ldots, T_k$ of the sets $U_1, \ldots, U_\ell$ such that $|T_i| \le M$ for all i and such that $k \le 2\ell$.*

Proof. We can refine each set U_i into at most $\lceil \frac{|U_i|}{M} \rceil$ sets of size at most M each[8]. Thus we can find a refinement $T_1, \ldots, T_k$ of $U_1, \ldots, U_\ell$ where $|T_i| \le M$ for all $1 \le i \le k$ and where

$$k \le \sum_{i=1}^{\ell} \left\lceil \frac{|U_i|}{M} \right\rceil \le \sum_{i=1}^{\ell} \left(\frac{|U_i|}{M} + 1 \right) \le 2\ell.$$

[8] Note this actually requires M to be an integer.

□

The following theorem is actually the main "engine" behind Theorem 2. It essentially picks up the proof of Theorem 2 in the middle of the attack, starting with queries to f_{r_0+1} (refering to the proof description of Section 4). Further down, we prove Theorem 2 from Theorem 3.

Theorem 3. *Let F be an (m, n, r, s) compression function with $m \geq 1$ such that either $m = 1$ or $mr \leq n$, and such that $s/2 \geq (nr - m)/(r + 1)$. Let c be such that $c2^{\frac{nr-m}{r+1}}$ is an integer. Then there exists an adversary that with at most*

$$q = 2c2^{\frac{nr-m}{r+1}}$$

queries to each f_i obtains at least

$$2^{s-\frac{2(nr-m)}{r+1}}$$

colliding inputs in $\{0, 1\}^{s+m}$ with probability at least $1 - \gamma(r, c)$.

Proof. We start by describing high-level characteristics of the proof and of the collision adversary. As preamble we recall that, by opposition to security proofs in which it is typically convenient to implement the ideal primitives $f_1, \ldots, f_r$ via lazy sampling, in the current context it is more convenient to view $f_1, \ldots, f_r$ as being sampled at the start of the experiment (from whatever distribution). Thus, when the attack starts, $f_1, \ldots, f_r$ are fixed.

The adversary first makes all its queries to f_1; then all its queries to f_2; etc, until it makes its queries to f_r. In other words, if $j > i$, the adversary never makes a query to f_i after making a query to f_j.

For $1 \leq i \leq r$ we let $S'_i \subseteq \{0, 1\}^{s+m}$ be the set of inputs for which the adversary knows the relevant queries to $f_1, \ldots, f_i$ after completing its queries to f_i. We will also define a set $S_i \subseteq S'_i$ for $1 \leq i \leq r$; the elements of S_i, are, roughly speaking, the only elements of S'_i that the adversary cares about. We emphasize that, after making the queries to f_i, the adversary knows S'_i and S_i. By construction, moreover, the proof maintains the invariant

$$|S_i| \leq 2^i \left(\frac{q}{2^n}\right)^i 2^{s+m} = (4c)^i 2^{s+\frac{m(r-i+1)-ni}{r+1}} \tag{4}$$

for $1 \leq i \leq r$.

For $1 \leq i \leq r$ we let $C^i \subseteq S_i$ be set of colliding inputs in S_i. That is,

$$C^i = \{x \in S_i : \exists y \in S_i, y \neq x, \text{ s.t. } F(x) = F(y)\}. \tag{5}$$

Thus, C^i is determined as soon as the adversary has finished its queries to f_i, but naturally, the adversary does not know C^i. We note that $|C^r|$ lower bounds the total number of colliding inputs at the end of the attack.

For $1 \leq i \leq r$, we define the boolean predicate ('SuffColl' stands for "sufficient number of collisions"):

$$\mathsf{SuffColl}_i = \left(|C^i| \geq \left(\frac{c^2}{20}\right)^{i-1} 2^{s+\frac{2m(r-i+1)-2ni}{r+1}}\right).$$

The truth value of $\mathsf{SuffColl}_i$ is determined once the adversary has made its queries to f_i (but the truth value of $\mathsf{SuffColl}_i$ is not known to the adversary). Note that if $\mathsf{SuffColl}_r$ holds then $|C^r| \geq 2^{s-\frac{2(nr-m)}{r+1}}$, which is the desired outcome for the adversary. Therefore it suffices to prove that $\mathsf{SuffColl}_r$ holds with probability at least $1 - \gamma(r, c)$. In fact, we show more strongly that

$$\Pr\left[\bigwedge_{i=1}^{r} \mathsf{SuffColl}_i\right] \geq 1 - \gamma(r, c). \tag{6}$$

Note that if the event $\bigwedge_{i=1}^{r} \mathsf{SuffColl}_i$ does not occur, then either

$$\neg\mathsf{SuffColl}_1$$

occurs, or

$$\neg\mathsf{SuffColl}_{i+1} \wedge \mathsf{SuffColl}_i$$

occurs for some i, $1 \le i \le r-1$. In order to show (6) we will show on the one hand that

$$\Pr[\neg\mathsf{SuffColl}_1] \le 2e^{-c^2/5760} \tag{7}$$

and on the other hand that

$$\Pr[\neg\mathsf{SuffColl}_{i+1} \wedge \mathsf{SuffColl}_i] \le 2e^{-\frac{1}{32}(\frac{c}{80})^i} \tag{8}$$

for $1 \le i \le r-1$. Combining (7), (8) and the definition of $\gamma(r,c)$ we obtain (6). Thus it suffices to establish (7) and (8) to prove the theorem.

For the remainder, it will be practical to define a boolean predicate to denote condition (4). We define, for $1 \le i \le r$,

$$\mathsf{SmallS}_i = \left(|S_i| \le 2^i \left(\frac{q}{2^n}\right)^i 2^{s+m} = (4c)^i 2^{s+\frac{m(r-i+1)-ni}{r+1}} \right).$$

By contrast with $\mathsf{SuffColl}_i$, SmallS_i holds with probability 1 for all i, as will become clear from our description of the adversary. Effectively, (7) and (8) will be established by showing the following inequalities:

(A) $\Pr[\mathsf{SmallS}_1] = 1$

(B) $\Pr[\mathsf{SmallS}_{i+1} \,|\, \mathsf{SmallS}_i] = 1$ for $1 \le i \le r-1$

(C) $\Pr[\neg\mathsf{SuffColl}_1] \le 2e^{-c^2/5760}$

(D) $\Pr[\neg\mathsf{SuffColl}_{i+1} \,|\, \mathsf{SuffColl}_i \wedge \mathsf{SmallS}_i] \le 2e^{-\frac{1}{32}(\frac{c}{80})^i}$ for $1 \le i \le r-1$

Here (C) is a restatement of (7), and (D) implies (8) because

$$\begin{aligned}\Pr[\neg\mathsf{SuffColl}_{i+1} \wedge \mathsf{SuffColl}_i] &= \Pr[\neg\mathsf{SuffColl}_{i+1} \wedge \mathsf{SuffColl}_i \wedge \mathsf{SmallS}_i] \\ &\le \Pr[\neg\mathsf{SuffColl}_{i+1} \,|\, \mathsf{SuffColl}_i \wedge \mathsf{SmallS}_i]\end{aligned}$$

(Here the equality follows from $\Pr[\mathsf{SmallS}_i] = 1$. The second inequality is just an application of the general principle that $\Pr[A \wedge B] \le \Pr[A \,|\, B]$ for any events A, B.)

Structurally our proof has two parts. In the first part we describe the adversary's f_1-query strategy, define $S_1 \subseteq S_1'$ and prove (A) and (C). In the second part we fix a value i, $1 \le i \le r-1$, assume the adversary has completed its queries to $f_1, \ldots, f_i$ and that $S_i \subseteq S_i'$ has been defined such that SmallS_i holds, describe the adversary's f_{i+1}-query strategy, define S_{i+1}, and prove (B) and (D). (We note that the adversary's f_1-query strategy is, indeed, slightly different than its f_i-query strategy for $i \ge 2$.) Together the two parts imply (A)–(D), which implies (7), (8), and hence the theorem.

As a preliminary, we note that the condition $s/2 \ge (nr-m)/(r+1)$ implies

$$n \le \frac{r+1}{2r}s + \frac{1}{r}m. \tag{9}$$

Also note that if $2^n \le q = 2c2^{\frac{nr-m}{r+1}}$ then the adversary can apply a trivial strategy consisting of querying every f_i at every point. In this case the output of each input is learned, resulting in at least $2^{m+s} - 2^s \ge 2^s$ colliding inputs (using $m \ge 1$ for the latter inequality). We can therefore assume that $2^n > q$. Finally, we can assume $c \ge 80$, since otherwise $1 - \gamma(r,c) \le 0$ and the theorem holds trivially.

We now delve into the proof's first part, i.e., the description of the adversary's f_1-query strategy, the definition of S_1, and the proof of (A) and (C). The adversary starts by selecting a set $I \subseteq \{0,1\}^n$ of $\lceil 2^{1+n-m} \rceil$ inputs to f_1 such as to maximize $|S'_I|$ where $S'_I = \{x \in \{0,1\}^{s+m} : g_1(x) \in I\}$ (but the adversary does not query these inputs). We then have

$$|S'_I| \geq 2^{s+m}\Big(\frac{2^{1+n-m}}{2^n}\Big) = 2^{s+1}.$$

The adversary creates a set S_I by arbitrarily removing points from S'_I until $|S_I| = 2^{s+1}$. Let $C^I = \{x \in S_I : \exists y \in S_I \text{ s.t. } F(x) = F(y), x \neq y\}$ be the set of colliding inputs in S_I. Note that $|C^I| \geq |S_I| - 2^s = 2^s$, since F has range of size 2^s.

We consider two cases according to whether $q \geq |I|$ or $q < |I|$ (where $q = 2c2^{\frac{nr-m}{r+1}}$). Assume first that $q \geq |I|$. Then

$$q \geq |I| \geq 2^{1+n-m}$$

and if $m = 1$ one obtains $q \geq 2^n$, a contradiction. Therefore $m > 1$ and (by the assumption in the theorem's statement), $mr \leq n$. The adversary queries f_1 at all points in I, resulting in $|I| \leq q$ queries to f_1. The set of inputs learned at the next iteration is $S_1 := S_I$, so that $|S_1| = 2^{s+1}$. For SmallS_1 to hold we need $|S_1| \leq 2c \cdot 2^{s+\frac{mr-n}{r+1}}$. Note, however, that

$$\begin{aligned} 2^{s+1} \leq 2c \cdot 2^{s+\frac{mr-n}{r+1}} &\iff 1 \leq c2^{\frac{mr-n}{r+1}} \\ &\iff 1 \leq c2^{\frac{nr-m}{r+1}}/2^{n-m} \\ &\iff 2^{n-m} \leq c2^{\frac{nr-m}{r+1}} \\ &\iff 2^{n-m} \leq \frac{1}{2}q \end{aligned}$$

and the latter inequality follows because $q \geq |I| \geq 2^{1+n-m}$. Hence SmallS_1 holds. Finally, for $\mathsf{SuffColl}_1$ to hold we need $|C^1| \geq 2^{s+\frac{2(mr-n)}{r+1}}$. But $|C^1| \geq 2^s$ and $mr \leq n$, so $\mathsf{SuffColl}_1$ holds with probability 1. Thus (A) and (C) are established for the case $q \geq |I|$.

Assume now that $q < |I|$. Let $U_y = \{x \in S_I : g_1(x) = y\}$ for all $y \in I$. The average size of the U_y's is at most M_0 where

$$M_0 := \left\lceil \frac{|S_I|}{|I|} \right\rceil = \left\lceil \frac{2^{s+1}}{\lceil 2^{1+n-m} \rceil} \right\rceil \leq \lceil 2^{s+m-n} \rceil \leq 2 \cdot 2^{s+m-n}$$

where the last inequality follows since (9) implies $s + m - n \geq 0$. By Lemma 2 there exists a refinement $T_1, \ldots, T_k$ of $\{U_y : y \in I\}$ such that $k = 2|I| = 2\lceil 2^{1+n-m} \rceil$ and $|T_j| \leq M_0$ for $1 \leq j \leq k$ (potentially some T_j's are empty). Fix such a refinement. Note that for each T_j there exists a unique $y_j \in I$ such that $T_j \subseteq U_{y_j}$.

The adversary chooses uniformly at random a subset $S \subseteq [k]$ of size q, and queries $f_1(y_j)$ for all $j \in S$. (Possibly, this results in redundant queries to f_1, which does not matter.)

Let $S_1 := \bigcup_{j \in S} T_j$. (We note that S_1 is indeed a subset of S'_1, the set of inputs in $\{0,1\}^{s+m}$ for which the adversary knows the queries to f_1.) We have

$$\begin{aligned} |S_1| &\leq qM_0 \leq 2c2^{\frac{nr-m}{r+1}} \cdot 2 \cdot 2^{s+m-n} \\ &= 4c2^{s+\frac{mr-n}{r+1}} \end{aligned}$$

which establishes (A).

Let $p = q/k$. Since $|I| = \lceil 2^{1+n-m} \rceil \leq 1.5 \cdot 2^{1+n-m}$ (using $mr \leq n \implies n - m \geq 0$), we have

$$p^2 = \frac{q^2}{4|I|^2} \geq \frac{4c^2 2^{\frac{2(nr-m)}{r+1}}}{4 \cdot 9 \cdot 2^{2(n-m)}} = \frac{1}{9}c^2 2^{\frac{2(mr-n)}{r+1}}.$$

Let C^1 be defined as in (5) and let $t = \frac{1}{20}p^2|C^I|$. Applying Lemma 6 with $M = M_0$, we obtain

$$\Pr\left[|C^1| < t\right] \le 2e^{-\frac{t}{16M_0}}. \tag{10}$$

Since

$$\begin{aligned}
\frac{t}{16M_0} &\ge \frac{p^2|C^I|}{20 \cdot 32 \cdot 2^{s+m-n}} \\
&\ge \frac{c^2 2^{\frac{2(mr-n)}{r+1}} 2^s}{9 \cdot 640 \cdot 2^{s+m-n}} \\
&= \frac{1}{5760} c^2 2^{(m+n)\frac{(r-1)}{r+1}} \\
&\ge c^2/5760
\end{aligned}$$

and since, using $c \ge 80 \ge \sqrt{180}$,

$$t = \frac{1}{20}p^2|C^I| \ge \frac{1}{180}c^2 2^{s+\frac{2(mr-n)}{r+1}} \ge 2^{s+\frac{2(mr-n)}{r+1}}$$

(10) implies

$$\Pr\left[|C^1| < 2^{s+\frac{2(mr-n)}{r+1}}\right] \le 2e^{-c^2/5760}$$

which is exactly (C). Thus we have proved (A) and (C) also for the case $q < |I|$, which completes the proof's first part (i.e., the definition of S_1 and the proofs of (A) and (C)).

For the second part, fix a value i, $1 \le i \le r-1$, and assume that the adversary has made its queries to $f_1, \ldots, f_i$, and that the set $S_i \subseteq S'_i$ has been defined such that SmallS_i holds. We will define the adversary's f_{i+1}-query strategy, define S_{i+1}, and prove (B) and (D). Note that C^i is determined by S_i.

Let

$$U_y = \{x \in S_i : g_{i+1}(x, y_1, \ldots, y_i) = y\}$$

for all $y \in \{0,1\}^n$, where $y_1, \ldots, y_i$ are the intermediate chaining variables for x (implicitly depending on x, in this notation). The average size of the U_y's is at most M_i where

$$\begin{aligned}
M_i := \left\lceil \frac{|S_i|}{2^n} \right\rceil \le \left\lceil \frac{(4c)^i 2^{s+\frac{m(r-i+1)-ni}{r+1}}}{2^n} \right\rceil &= \left\lceil (4c)^i 2^{s+\frac{m(r-i+1)-n(r+i+1)}{r+1}} \right\rceil \\
&\le 2(4c)^i 2^{s+\frac{m(r-i+1)-n(r+i+1)}{r+1}}
\end{aligned}$$

where the last inequality follows because

$$\begin{aligned}
s + \frac{m(r-i+1)-n(r+i+1)}{r+1} \ge 0 \iff\!\!\!\!\Longleftarrow\; &s + \frac{2m-2rn}{r+1} \ge 0 \\
\iff\; &s/2 \ge (nr-m)/(r+1) \qquad (11)
\end{aligned}$$

(using $i \le r-1$ for the first implication), where (11) holds by assumption.

By Lemma 2 there exists a refinement $T_1, \ldots, T_k$ of $\{U_y : y \in \{0,1\}^n\}$ such that $k = 2^{n+1}$ and $|T_j| \le M_i$ for $1 \le j \le k$. Fix such a refinement. The adversary chooses uniformly at random a subset $S \subseteq [k]$ of size q, and queries $f_{i+1}(y_j)$ for all $j \in S$. We put $S_{i+1} := \bigcup_{j \in S} T_j$, and note that $S_{i+1} \subseteq S'_{i+1}$. Moreover,

$$|S_{i+1}| \le qM_i = 2c2^{\frac{nr-m}{r+1}} 2(4c)^i 2^{s+\frac{m(r-i+1)-n(r+i+1)}{r+1}} = (4c)^{i+1} 2^{\frac{m(r-(i+1)+1)-n(i+1)}{r+1}}$$

which establishes SmallS_{i+1}, i.e., which establishes (B).

To prove (D) we assume $\mathsf{SuffColl}_i$ holds, or else (D) is void. Let $p = q/k$. We have

$$p^2 = \frac{q^2}{k^2} = \frac{4c^2 2^{\frac{2(nr-m)}{r+1}}}{2^{2(n+1)}} = c^2 2^{-\frac{2(n+m)}{r+1}}.$$

Let $t = \frac{1}{20}p^2|C^i|$. Applying Lemma 6 with $M = M_i$, we find

$$\Pr\left[|C^{i+1}| < t\right] \leq 2e^{-\frac{t}{16M_i}}. \tag{12}$$

Since $|C^i| \geq \left(\frac{c^2}{20}\right)^{i-1} 2^{s+\frac{2m(r-i+1)-2ni}{r+1}}$ by $\mathsf{SuffColl}_i$,

$$\begin{aligned}
\frac{t}{16M_i} &= \frac{\frac{1}{20}p^2|C^i|}{16M_i} \\
&\geq \frac{\frac{1}{20}c^2 2^{-\frac{2(n+m)}{r+1}} \left(\frac{c^2}{20}\right)^{i-1} 2^{s+\frac{2m(r-i+1)-2ni}{r+1}}}{16 \cdot 2(4c)^i 2^{s+\frac{m(r-i+1)-n(r+i+1)}{r+1}}} \\
&= \frac{\left(\frac{c^2}{20}\right)^i 2^{s+\frac{2m(r-i)-2n(i+1)}{r+1}}}{16 \cdot 2(4c)^i 2^{s+\frac{m(r-i+1)-n(r+i+1)}{r+1}}} \\
&= \frac{\left(\frac{c}{80}\right)^i}{32} 2^{(m+n)\frac{r-i-1}{r+1}} \\
&\geq \frac{1}{32}\left(\frac{c}{80}\right)^i
\end{aligned}$$

and since

$$t = \frac{1}{20}p^2|C^i| \geq \left(\frac{c^2}{20}\right)^i 2^{s+\frac{2m(r-i)-2n(i+1)}{r+1}}$$

(12) implies

$$\Pr\left[|C^{i+1}| < \left(\frac{c^2}{20}\right)^i 2^{s+\frac{2m(r-i)-2n(i+1)}{r+1}}\right] \leq e^{-\frac{1}{32}\left(\frac{c}{80}\right)^i}$$

which completes the proof of (D), and the proof of the theorem. □

Proof of Theorem 2. If $2^n \leq q = 2c2^{\frac{nr-m}{r+1}}$ then the adversary can apply a trivial strategy consisting of querying every f_i at every point. In this case the output of each input is learned, resulting in at least $2^{m+s} - 2^s \geq 2^s$ colliding inputs (using $m \geq 1$ for the latter inequality). We can therefore assume that $2^n > q$.

Let $r_0 = \lfloor m(r+1)/(m+n) \rfloor$ and let $\alpha = m(r+1)/(m+n) - r_0$ be the fractional part of $m(r+1)/(m+n)$. We note that since $m < nr$, $0 \leq r_0 \leq r-1$. If $r_0 = 0$ then $m(r+1) < m+n$ and $mr < n$. In this case, we can thus directly apply Theorem 3. We therefore assume $r_0 \geq 1$.

The adversary starts by making

$$\left\lceil 2 \cdot 2^{\frac{nr-m}{r+1}} \right\rceil \leq \left\lceil 2c2^{\frac{nr-m}{r+1}} \right\rceil = q$$

queries to each of $f_1, \ldots, f_{r_0}$ to maximize its yield. Then the set of inputs $S'_{r_0} \subseteq \{0,1\}^{s+m}$ for which the adversary knows the queries to $f_1, \ldots, f_{r_0}$ has size at least

$$\begin{aligned}
2^{s+m}\left(\frac{2 \cdot 2^{\frac{nr-m}{r+1}}}{2^n}\right)^{r_0} &= 2^{s+m} 2^{r_0} \left(2^{-\frac{n+m}{r+1}}\right)^{r_0} \\
&= 2^{s+m} 2^{r_0} \left(2^{-\frac{n+m}{r+1}}\right)^{\frac{m(r+1)}{m+n}-\alpha} \\
&= 2^s 2^{r_0} 2^{\alpha\frac{n+m}{r+1}}.
\end{aligned}$$

In particular, $|S'_{r_0}| \geq 2^{s+1}$ since $r_0 \geq 1$. We now consider two cases, according to whether $\alpha\frac{n+m}{r+1} \geq 1$ or $\alpha\frac{n+m}{r+1} < 1$.

Assume first $\alpha\frac{n+m}{r+1} \geq 1$. Then the adversary constructs a subset S_{r_0} of S'_{r_0} of size $\lceil 2^{s+\alpha\frac{n+m}{r+1}} \rceil$. Consider the residual compression function $F' : S_{r_0} \to \{0,1\}^s$ induced by F. Then F' is a (m', n, r', s) compression function[9] where $m' := \alpha\frac{n+m}{r+1} \geq 1$ and $r' := r - r_0$. Then

$$m' = \alpha\frac{n+m}{r+1} = \left(\frac{m(r+1)}{m+n} - r_0\right)\frac{n+m}{r+1} = m - \frac{n+m}{r+1}r_0$$

so

$$\begin{aligned} nr' - m' - (r'+1)\frac{nr-m}{r+1} &= nr' - m' - (r - r_0 + 1)\frac{nr-m}{r+1} \\ &= nr' - m' - (nr - m) + r_0\frac{nr-m}{r+1} \\ &= -nr_0 + m - m' + \frac{nr-m}{r+1}r_0 \\ &= m - m' - \frac{n+m}{r+1}r_0 \ = \ 0 \end{aligned}$$

namely

$$\frac{nr' - m'}{r'+1} = \frac{nr-m}{r+1}.$$

In particular, $s/2 \geq (nr' - m')/(r'+1)$ (since $s/2 \geq (nr-m)/(r+1)$ by assumption). Moreover,

$$m'r' = \alpha\frac{n+m}{r+1}(r - r_0) \leq \frac{n+m}{r+1}\big(r - \frac{m(r+1)}{m+n} + 1\big) = n.$$

Therefore we can apply Theorem 3 to the (m', n, r', s) compression function F' with the same constant c. Hence, since $2^{\frac{nr'-m'}{r'+1}} = 2^{\frac{nr-m}{r+1}}$, we find that with at most q queries to each of $f_{r_0+1}, \ldots, f_r$ the adversary can obtain at least

$$2^{s - \frac{2(nr'-m')}{r'+1}} = 2^{s - \frac{2(nr-m)}{r+1}}$$

colliding inputs with probability at least $1 - \gamma(r, c)$.

Assume now that $\alpha\frac{n+m}{r+1} < 1$. Then the adversary constructs a subset S_{r_0} of S'_{r_0} of such that $|S_{r_0}| = 2^{s+1}$. Consider the residual compression function $F' : S_{r_0} \to \{0,1\}^s$ induced by F. Then F' is a (m'', n, r', s) compression function where $m'' = 1$ and $r' = r_0 - r$. Let $m' = \alpha\frac{n+m}{r+1}$ be as defined above; we then note that, as above, $\frac{nr'-m'}{r'+1} = \frac{nr-m}{r+1}$. By assumption, $m' < 1 = m''$. Therefore

$$2^{\frac{nr-m}{r+1}} = 2^{\frac{nr'-m'}{r'+1}} > 2^{\frac{nr'-m''}{r+1}}.$$

Let $c' = \lceil c2^{\frac{nr'-m''}{r'+1}} \rceil / 2^{\frac{nr'-m'}{r'+1}} \geq c$. Since $s/2 \geq (nr-m)/(r+1) = (nr'-m')/(r'+1)$, the case $m = 1$ of Theorem 3 implies that with at most

$$2c'2^{\frac{nr'-m''}{r'+1}} = 2\lceil c2^{\frac{nr'-m''}{r'+1}} \rceil \leq 2\lceil c2^{\frac{nr'-m'}{r'+1}} \rceil = 2c2^{\frac{nr-m}{r+1}}$$

queries to each of $f_{r_0+1}, \ldots, f_r$ the adversary can obtain at least

$$2^{s - \frac{2(nr'-m'')}{r'+1}} \geq 2^{s - \frac{2(nr'-m')}{r'+1}} = 2^{s - \frac{2(nr-m)}{r+1}}$$

colliding inputs with probability at least $1 - \gamma(r', c') \geq 1 - \gamma(r, c)$. $\square$

[9] Recall our convention that a (m', n, r', s) compression function has a domain of size $\lceil 2^{m'+s} \rceil$.

B Supporting Lemmas

We recall that for random variables X, Y, a notation such as $\mathrm{Var}(X|Y=s)$ means the variance of X conditioned on the event $Y=s$, whereas $\mathrm{Var}(X|Y)$ is a *function* from the range of Y to $\mathbb{R}$, that assigns $\mathrm{Var}(X|Y=s)$ to each s in the range of Y (or more precisely, to each s such that $\Pr[Y=s]$ is nonzero). The notation "$\mathrm{Var}(X|Y)\le c$" indicates this function is upper bounded by c: $\mathrm{Var}(X|Y=s)\le c$ for all s such that $\Pr[Y=s]>0$.

We use, as a starting point, the following concentration result for Martingales. See Theorem 6.1 of [3] for a proof. (We note that our notation is slightly modified from standard in order to avoid discussion of filters.)

Lemma 3 (Folklore [3,4]). *Let $Y_1,\ldots,Y_n$ be a sequence of random variables of range R, $f: R^n\to\mathbb{R}$ be a function and let $Y=f(Y_1,\ldots,Y_n)$. Let*

$$X_i = E[Y|Y_1,\ldots,Y_i]$$

for $0\le i\le n$. Then if

1. $\mathrm{Var}(X_i|Y_1,\ldots,Y_{i-1})\le\sigma_i^2$ *for $1\le i\le n$, and*
2. $|X_i-X_{i-1}|\le M$, *for every $1\le i\le n$,*

we have

$$\Pr[Y\le E[Y]-\lambda]\le e^{-\frac{\lambda^2}{2(\sum_{i=1}^n\sigma_i^2+M\lambda/3)}}$$

for any $\lambda\ge 0$.

Lemma 4. *Let k,q be integers such that $1\le q\le k$. Let $\mathcal{B}$ be random a subset of $[k]=\{1,\ldots,k\}$ of size q. Let M and $c_1,\ldots,c_k$ be nonnegative constants such that $M\ge c_i$ for $1\le i\le k$. Put $Y=\sum_{i\in\mathcal{B}}c_i$. Then*

$$\Pr[Y\le E[Y]-t]\le e^{-\frac{t^2}{2M(3E[Y]+t/3)}}$$

for all $t\ge 0$.

Proof. Note that if $q=k$ the lemma is obviously true, and so we can assume $q<k$.

We view the elements of $\mathcal{B}$ as being selected sequentially, with the i-th element of $\mathcal{B}$ coming uniformly at random from a set of size $k-i+1$ (the complement of the currently selected elements). Let s_i be the i-th chosen element, and define $f:[k]^q\to\mathbb{R}$ by $f(s_1,\ldots,s_q)=\sum_{i=1}^q c_{s_i}$. Note $Y=f(s_1,\ldots,s_q)$. In view of applying Lemma 3 (with $Y_i=s_i$), we define

$$X_i = E[Y|s_1,\ldots,s_i]$$

for $0\le i\le q$. Thus, X_i is the expected "value" of $\mathcal{B}$ after the first i elements have been chosen.

Note that for any values $t_1,\ldots,t_q\in[k]$ and $t_i'\in[k]$,

$$|f(t_1,\ldots,t_q)-f(t_1,\ldots,t_{i-1},t_i',t_{i+1},\ldots,t_q)|\le M.$$

That is, changing the i-th input of f (i.e. the i-th element chosen) can only change f's output by M, at most. It follows (by a short but standard argument) that $|X_i-X_{i-1}|\le M$ for $1\le i\le q$.

We next want to upper bound $\mathrm{Var}(X_{i+1}|s_1,\ldots,s_i)$ independently of $s_1,\ldots,s_i$. We have:

$$\begin{aligned}
X_{i+1} &= c_{s_{i+1}}+\frac{q-i}{k-i-1}\sum_{h\notin\{s_1,\ldots,s_{i+1}\}}c_h+\sum_{j=1}^{i}c_{s_j}\\
&= c_{s_{i+1}}\Big(1-\frac{q-i}{k-i-1}\Big)+\frac{q-i}{k-i-1}\sum_{h\notin\{s_1,\ldots,s_i\}}c_h+\sum_{j=1}^{i}c_{s_j}\\
&= c_{s_{i+1}}\Big(1-\frac{q-i}{k-i-1}\Big)+K
\end{aligned}$$

where K is a constant depending only on $s_1, \ldots, s_i$. Therefore,

$$\begin{aligned}
\mathrm{Var}(X_{i+1}|s_1, \ldots, s_i) &= \left(1 - \frac{q-i}{k-i-1}\right)^2 \cdot \mathrm{Var}(c_{s_j}|s_1, \ldots s_i) \\
&\leq \left(1 - \frac{q-i}{k-i-1}\right)^2 \cdot \frac{1}{k-i} \sum_{j \notin \{s_1, \ldots, s_i\}} c_j^2 \\
&\leq \frac{(k-q)^2}{(k-i)(k-i-1)^2} \sum_{j=1}^{k} c_j^2.
\end{aligned}$$

We set σ_{i+1}^2 to this last expression, $0 \leq i < q$, so that $\mathrm{Var}(X_{i+1}|s_1, \ldots, s_i) \leq \sigma_{i+1}^2$.

Let $p = q/k$. Note that $p < 1$ since we are assuming $q < k$ and that $X_0 = E[Y] = p \sum_{i=1}^{k} c_i$. We have

$$\begin{aligned}
\sum_{i=1}^{q} \sigma_i^2 &= \sum_{j=1}^{k} c_j^2 \cdot (k-q)^2 \sum_{i=0}^{q-1} \frac{1}{(k-i)(k-i-1)^2} \\
&\leq \sum_{j=1}^{k} c_j^2 \cdot (k-q)^2 \sum_{i=1}^{q} \frac{1}{(k-i)^3} \\
&= \sum_{j=1}^{k} c_j^2 \cdot (k-q)^2 \left(\sum_{i=0}^{q-1} \frac{1}{(k-i)^3} + \frac{1}{(k-q)^3} - \frac{1}{k^3} \right) \\
&\leq \sum_{j=1}^{k} c_j^2 \cdot (k-q)^2 \left(\int_0^q \frac{1}{(k-x)^3} dx + \frac{1}{(k-q)^3} - \frac{1}{k^3} \right) \\
&= \sum_{j=1}^{k} c_j^2 \cdot \left((k-q)^2 \cdot \frac{1}{2} \left(\frac{1}{(k-q)^2} - \frac{1}{k^2} \right) + \frac{1}{k-q} - \frac{(k-q)^2}{k^3} \right) \\
&= \left(\frac{1}{2} \left(1 - \frac{(k-q)^2}{k^2} \right) + \frac{1}{k-q} - \frac{(k-q)^2}{k^3} \right) \sum_{j=1}^{k} c_j^2 \\
&= \left(\frac{1}{2} \left(1 - (1-p)^2 \right) + \frac{1}{k(1-p)} - \frac{(1-p)^2}{k} \right) \sum_{j=1}^{k} c_j^2 \\
&\leq \left(\frac{1}{2}(2p - p^2) + \frac{1}{k(1-p)} \right) \sum_{j=1}^{k} c_j^2 \\
&= p \left(\frac{1}{2}(2-p) + \frac{1}{q(1-p)} \right) \sum_{j=1}^{k} c_j^2 \\
&= p \left(\frac{1}{2}(2-p) + \frac{1}{q} + \frac{1}{k(1-p)} \right) \sum_{j=1}^{k} c_j^2 \\
&\leq p \left(\frac{1}{2}(2-p) + \frac{1}{q} + \frac{1}{k\frac{1}{k}} \right) \sum_{j=1}^{k} c_j^2 \\
&\leq 3p \sum_{j=1}^{k} c_j^2
\end{aligned}$$

$$
\begin{aligned}
&\leq 3p\sum_{j=1}^{k} c_j M \\
&= 3E[Y]M
\end{aligned}
$$

Then by Lemma 3, we have

$$\Pr[Y < E[Y] - t] \leq e^{-\frac{t^2}{2(\sum_{i=1}^{q}\sigma_i^2 + Mt/3)}} \leq e^{-\frac{t^2}{2(3E[Y]M + Mt/3)}} = e^{-\frac{t^2}{2M(3E[Y]+t/3)}}.$$

□

Lemma 5. *Let k, q be integers such that $1 \leq q \leq k$. Let $\mathcal{B}$ be random a subset of $[k] = \{1, \ldots, k\}$ of size q. Let M and $c_1, \ldots, c_k$ be nonnegative constants such that $M \geq c_i$ for $1 \leq i \leq k$. Put $Y = \sum_{i \in \mathcal{B}} c_i$. Then*

$$\Pr[Y < \phi - t] \leq e^{-\frac{t^2}{2M(3\phi + t/3)}} \tag{13}$$

for any t, ϕ such that $0 \leq t \leq \phi \leq E[Y]$.

Proof. Let $u = E[Y] - \phi$. Then by Lemma 4,

$$
\begin{aligned}
\Pr[Y < \phi - t] &= \Pr[Y < E[Y] - u - t] \\
&\leq e^{-\frac{(t+u)^2}{2M(3(u+\phi)+(t+u)/3)}}
\end{aligned} \tag{14}
$$

Let $f(u) = (t+u)^2$, $g(u) = 3(u+\phi) + (t+u)/3$, we find that

$$
\begin{aligned}
\left(\frac{f(u)}{g(u)}\right)' \geq 0 &\iff f'(u)g(u) \geq g'(u)f(u) \\
&\iff 2g(u) \geq g'(u)(t+u) \\
&\iff 2g(u) \geq (3 + 1/3)(t+u) \\
&\Longleftarrow g(u) \geq (3 + 1/3)(t+u) \\
&\iff 3(u+\phi) + (t+u)/3 \geq (3+1/3)(t+u) \\
&\Longleftarrow 3(u+t) + (t+u)/3 \geq (3+1/3)(t+u)
\end{aligned}
$$

where we use $\phi \geq t$ for the last implication. Thus (14), considered as a function of u and restricted to $u \geq 0$, takes its maximum at $u = 0$, which establishes (13). □

Lemma 6. *Let k, q be integers such that $1 \leq q \leq k$ and such that q is even. Let $M > 0$ be a constant and let T be the disjoint union of sets $T_1, \ldots, T_k$ such that $|T_i| \leq M$ for $1 \leq i \leq k$. Let $F : T \to U$ be some function and let*

$$C_i = |\{x \in T_i : \exists y \in T, y \neq x, \text{ s.t. } F(x) = F(y)\}|$$

Let $C = C_1 + \cdots + C_k$. Let $\mathcal{B}$ be a random subset of $[k]$ of size q. Let

$$\overline{C_i} = |\{x \in T_i : \exists y \in T_j, j \in \mathcal{B}, y \neq x, F(x) = F(y)\}|$$

and let

$$\overline{C} = \sum_{i : i \in \mathcal{B}} \overline{C_i}$$

then

$$\Pr[\overline{C} < t] \leq 2e^{-\frac{t}{16M}}$$

where $t = \frac{1}{20}p^2 C$ and $p = q/k$.

Proof. We use the following equivalent selection process for $\mathcal{B}$: we first select, independently and uniformly at random, two subsets L and R of $[k]$ of size $q/2$ each, then select an additional set H of size $q - |L \cup R|$ uniformly at random from $[k]$, and finally set $\mathcal{B} = L \cup R \cup H$. Clearly, this process yields a set $\mathcal{B}$ of size q that is uniformly distributed at random among all subsets of $[k]$ of size q.

Define random variables $Y_1, \dots, Y_k$ by putting $Y_i = C_i$ if $i \in L$, $Y_i = 0$ otherwise. Let $Y = \sum_{i=1}^k Y_i$. We have $|Y_i| \le |T_i| \le M$. Note that $E[Y] = \frac{q/2}{k} C = \frac{1}{2} pC$. Let $t_0 = \frac{1}{4} pC = \frac{1}{2} E[Y]$. By Lemma 4,

$$\begin{aligned}\Pr[Y < E[Y] - t_0] &\le e^{-\frac{t_0^2}{2M(3E[Y]+t_0/3)}} \\ &= e^{-\frac{p^2C^2/16}{2M(3pC/2+pC/12)}} \\ &= e^{-\frac{pC}{32M(3/2+1/12)}} \\ &\le e^{-\frac{pC}{51M}}.\end{aligned}$$

For the rest of the proof we assume that $Y \ge E[Y] - t_0 = \frac{1}{4} pC$. For $1 \le i \le k$, let

$$C_i^L = |\{x \in T_i : \exists y \in T_j, j \in L, y \ne x, \text{ s.t. } F(x) = F(y)\}|.$$

Recall that $t = \frac{1}{20} p^2 C$. It is not difficult to see that if $\sum_{i \in [k]} C_i^L \le \sum_{i \in L} C_i - t$, then

$$\sum_{i \in L} |\{x \in T_i : \exists y \in T_j, j \in L, y \ne x, \text{ s.t. } F(x) = F(y)\}| \ge t + 1,$$

implying that $\overline{C} \ge t$. We can therefore assume that $\sum_{i \in [k]} C_i^L \ge \sum_{i \in L} C_i - t \ge \frac{1}{4} pC - t$.

Define random variables $Z_1, \dots, Z_k$ by putting $Z_i = C_i^L$ if $i \in R$, $Z_i = 0$ otherwise, and let $Z = \sum_{i=1}^k Z_i$. Then $|Z_i| \le |T_i| \le M$ for all i. Let $\phi = (p/2)(\frac{1}{4} pC - t) \le E[Z]$ (the latter equality follows from the fact that each set is added to R with probability $p/2$, and from the fact that $\sum_{i \in [k]} C_i^L \ge \frac{1}{4} pC - t$). Then

$$\phi = \frac{1}{8} p^2 C - \frac{1}{2} pt = (2.5 - \frac{1}{2} p) t \ge 2t.$$

Since $0 \le t \le \phi \le E[Z]$, $t \le \phi - t$, and $\phi \le 2.5t$, we have by Lemma 5 that

$$\begin{aligned}\Pr[Z < t] &\le \Pr[Z < \phi - t] \\ &\le e^{-\frac{t^2}{2M(3\phi+t/3)}} \\ &\le e^{-\frac{t^2}{2M(3\cdot 2.5t+t/3)}} \\ &= e^{-\frac{t}{M(15+2/3)}} \\ &\le e^{-\frac{t}{16M}}\end{aligned}$$

Since $\overline{C} \ge Z$ and since $e^{-\frac{t}{16M}} = e^{-\frac{p^2C}{20\cdot 16M}} \ge e^{-\frac{p^2C}{51M}} \ge e^{-\frac{pC}{51M}}$, a sum bound on the two bad events (these being the event that either $Y < E[Y] - t_0$, or that $Z < t$) concludes the lemma. $\square$