Stam's Collision Resistance Conjecture

John Steinberger*

Institute of Theoretical Computer Science, Tsinghua University, Beijing
jpsteinb@gmail.com

Abstract. At CRYPTO 2008 Stam [7] made the following conjecture: 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. For example, a $2n$-bit to n-bit compression function making two calls to a random function of n-bit input cannot have collision security exceeding $2^{n/3}$. We prove this conjecture up to a constant multiplicative factor and under the condition $m' := (2m-n(r-1))/(r+1) \geq \log_2(17)$. This covers nearly all cases $r = 1$ of the conjecture and the aforementioned example of a $2n$-bit to n-bit compression function making two calls to a primitive of n-bit input.

1 Introduction

A popular paradigm for security proofs in the field of hash function design is to assume that some primitive used by the hash function, such as a blockcipher, is "ideal", namely perfectly random subject to the constraints of the type of primitive concerned, and then to bound the chance of success of some adversary given oracle access to this primitive in terms of the number of queries allowed to the adversary. In this "ideal primitive" model (or IPM, as we will call it) adversaries are usually information-theoretic: their only obstacle to achieving an 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.

In the same vein as the above argument, this paper pursues the task of determining the limits of IPM security. Specifically, we tackle the following question: given $m, n, r, s \geq 1$, what is the maximum collision security of a compression function $F : \{0,1\}^{m+s} \rightarrow$

* Supported by the National Natural Science Foundation of China Grant 60553001 and by the National Basic Research Program of China Grant 2007CB807900, 2007CB807901.

H. Gilbert (Ed.): EUROCRYPT 2010, LNCS 6110, pp. 597–615, 2010.

$\{0,1\}^s$ 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 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 $q = 2\sqrt{2}r2^{s/2}$ queries (cf. Proposition 1 Section 5). However other attacks may be more constraining than birthday attacks. In particular Stam [7] conjectured[2] 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 $\frac{1}{2}$. This bound becomes more constraining than the birthday attack when $s/2 > (nr-m)/(r+1)$. This 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 $2^{n/3}$ whereas a birthday attack caps the collision resistance at $2^{n/2}$. It is noteworthy that Stam's bound is independent of s. We explain later the intuition behind the exponent $(nr-m)/(r+1)$.

Stam's conjecture is particularly 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, to the best of current understanding, it seems that 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 in fact 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 [4,7,6].

So far, however, Stam's bound has not been proved for any case of interest (cases of "non-interest" being those for which $s/2 \leq (nr-m)/(r+1)$ or $nr-m \leq 0$; see Section 2). Here we try to remedy this situation. We show there is an absolute constant $C \geq 1$ such that if

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

then

$$q = Cr\lceil 2^{(nr-m)/(r+1)}\rceil \tag{3}$$

queries suffice in order to obtain a collision with probability at least $\frac{1}{2}$ (see Corollary 3 in Section 5 for a tighter statement). In other words, we prove Stam's conjecture up

[1] Immaterial to proving our upper bound; better upper bounds on security should be provable if f has sufficiently small range, see comments by Stam [7].

[2] Stam's wording is not quite as precise, as he omits the ceiling brackets, the '+1' term, and the fact that a collision can only be found with "sufficient" probability, but it is easy to see these changes are necessary for correctness of the conjecture.

to a constant multiplicative factor as long as (2) holds. To get a better handle on the restriction (2) note that it reduces to $m = m' \geq 4.09$ for $r = 1$ and to $\frac{2}{3}m - \frac{1}{3}n \geq 4.09$ for $r = 2$. For $r = 2$ setting $m = n$ reduces the condition to $n \geq 12.27$. Our result is partly based on the observation that Stam's conjecture reduces to the case $r = 1$ when $m' \geq 1$; see Section 4 for details.

We emphasize that our result holds for arbitrary primitives f. That is, if f has range $\{0,1\}^b$, then f may be sampled with any distribution from all functions of domain $\{0,1\}^n$ and range $\{0,1\}^b$. Thus our result covers not only perfectly random primitives but also random permutations and ideal ciphers[3]. Moreover, in the case where $r > 1$, F may call r distinct primitives (of potentially different distributions) rather than the same primitive r times.

PROBLEM HISTORY. The first authors to consider the limits of IPM security in the information-theoretic setting were Black, Cochran and Shrimpton [2], who showed that any iterated hash function using a $2n$-bit to n-bit compression function F making a single call to one of r different ideal n-bit permutations would have (unacceptably low) collision security of $r(n + \log(n))$ queries. Rogaway and Steinberger [5] generalized this result by showing that collisions could be found with probability 1 in $1.89s2^{n(1-\alpha)}$ queries for any permutation-based hash function of rate α and output length s (the rate being the number of n-bit message blocks processed per application of the n-bit primitive). The latter result is somewhat noteworthy because it does not make any assumption on the structure (iterated, etc) of the hash function, and does not even restrict the number of different independent permutations used by the hash function—moreover the result more generally holds (with the same proof) if the permutations are replaced by any primitives of domain $\{0,1\}^n$.

Rogaway and Steinberger also considered the IPM security of compression functions instantiated from n-bit random permutations (like above, their proofs in fact apply for any primitive of domain $\{0,1\}^n$). They showed that with $r(2^{n-m/r} + 1)$ queries an adversary could find a collision with probability 1 for any compression function $F : \{0,1\}^{m+s} \to \{0,1\}^s$ that makes r calls to an n-bit permutation. They also noted that, for compression functions F meeting a certain reasonable-looking heuristic assumption dubbed "collision uniformity", $r2^{n-(m+\frac{s}{2})/r}$ queries suffice for finding a collision with probability $\frac{1}{2}$. Stam [7] subsequently found examples of non-collision-uniform compression functions having higher collision security than $r2^{n-(m+\frac{s}{2})/r}$ and posited that collision security could not exceed $r\lceil 2^{(nr-m)/(r+1)} \rceil$ independently of any heuristic assumption. This is the bound we discuss in this paper.

ON 'OPTIMALITY'. Security upper bounds are useful as benchmarks for designers. In this area, though, the situation isn't so simple: when $2^{(nr-m)/(r+1)} < 2^{s/2}$ (namely when Stam's bound becomes more constraining than the birthday attack upper bound) then the only constructions which can achieve the best-possible collision security are non-uniform constructions, implying a questionable non-random behavior. The "better"

[3] A blockcipher of k-bit key and l-bit word is modeled as a primitive of $l + k$-bit input; note the absence of inverse queries does typically not affect the task of proving *upper bounds* on security, though if desired one may even emulate bidirectional blockcipher queries with an extra bit of input specifying forward or inverse queries.

construction may then be a uniform construction of lower collision security. On the other hand, some non-collision-uniform constructions have been proposed, for example the JHash compression function [9]. The non-uniformity of these compression functions is usually belied by the fact that many collisions are obtained whenever a single collision is obtained. (Uniformity is explained in more detail in Section 3.)

Regarding this issue, Stam has suggested that when $(nr - m)/(r + 1) < s/2$ one should consider lowering the state size s until $s/2 = (nr - m)/(r + 1)$, so that one may (at least theoretically) achieve the optimum collision resistance with a uniform construction, as opposed to achieving the same collision resistance with a non-uniform construction or a lower collision resistance with a uniform construction. This makes sense from the point of view of compression function design, though designers should bear in mind the hash function obtained by iterating the compression function will probably be weakened by lowering the state size at the same time the compression function is strengthened (the collision resistance of the hash function being typically higher than that of the compression function); for example, while a uniform $2n$-bit to n-bit compression function F_1 making two calls to an n-bit input random function f may have only $2^{n/4}$ collision security against $2^{n/3}$ collision security for a uniform $\frac{5}{3}n$-bit to $\frac{2}{3}n$-bit compression function F_2 also making two calls to a random n-bit input function f, the iteration of F_1 may have $2^{n/2}$ collision security[4] whereas the iteration of F_2 will be "stuck" at $2^{n/3}$ collision security.

Finally, the usual caveats regarding the ideal primitive model apply to this paper: as the IPM considers information-theoretic adversaries, our results do not imply security upper bounds with respect to real-world, computationally bounded adversaries.

ORGANIZATION. In the next section we give some background of results of Rogaway and Steinberger. Section 3 is an optional section giving some intuition about Stam's conjecture for $r > 1$. Section 4 examines the case $r = 1$ and how certain cases of Stam's conjecture with $r > 1$ reduce to the case $r = 1$. Section 5 contains the main proof and the formal statement of our result, which is summarized by Corollary 3. Appendix A discusses an alternate approach to our main result for the special case of random primitives.

2 Basic Results

We first formalize the notion of a compression function F making r calls to a primitive f. 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".

[4] This is indeed conjectured for a number of two-call constructions, such as the Grøstl compression function [3].

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 $0 \leq k \leq r$. We start with the following basic observation of Rogaway and Steinberger [5]:

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 and let $0 \leq k \leq r$. Then with at most q queries to each of the functions $f_1, \ldots, f_k$ an adversary can learn the first k chaining variables for at least*

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

inputs.

Proof. We proceed by induction on k, with the result obviously holding for $k = 0$. Now assume $1 \leq k \leq r$. By the induction hypothesis, the adversary can make q queries to each of $f_1, \ldots, f_{k-1}$ so that it knows the first $k-1$ chaining variables for at least

$$2^{m+s} \left(\frac{q}{2^n}\right)^{k-1}$$

inputs. Let X be the set of these inputs, and for each $z \in \{0,1\}^n$ let X_z be the set of inputs $v \in X$ such that $g_k(v, y_1, \ldots, y_{k-1}) = z$ where $y_1, \ldots, y_{k-1}$ are the first $k-1$ chaining variables for v. Because $\{X_z : z \in \{0,1\}^n\}$ are disjoint and have union X there exist distinct values $z_1, \ldots, z_q \in \{0,1\}^n$ such that $\sum_{i=1}^q |X_{z_i}| \geq q|X|/2^n$. By querying $f_k(z_1), \ldots, f_k(z_q)$ the adversary thus learns the first k intermediate variables for at least

$$q|X|/2^n \geq 2^{m+s} \left(\frac{q}{2^n}\right)^k$$

inputs. □

Rogaway and Steinberger originally stated this observation for primitives $f_1, \ldots, f_r : \{0,1\}^n \to \{0,1\}^n$, but the output length of the f_i's does not in fact play any role. Stam [7] subsequently generalized Lemma 1 to the case of compressing primitives $f_i : \{0,1\}^{n+c} \to \{0,1\}^n$, but this generalization is equivalent to Lemma 1 for the same reason (namely it can be obtained by substituting $n + c$ for n and n for b, the latter with no effect).

As a direct corollary of Lemma 1, we have the following:

Corollary 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 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.

In particular, if

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

then an adversary can obtain a collision for F with probability 1 in rq queries. Solving this inequality for q gives

$$q > 2^{n-m/r}$$

so that

$$r(\lfloor 2^{n-m/r} \rfloor + 1)$$

queries suffice to find a collision with probability 1 (when $n-m/r = 0$ one can improve this bound to $r+1$ queries). This proves Stam's conjecture for the case $nr - m \leq 0$. (In fact (1) is one more query than needed when $nr - m < 0$.)

3 Intuition for Stam's Bound: The Case $r > 1$

In this section we explain where Stam's bound "comes from". We assume $r > 1$; the case $r = 1$, which has certain peculiarities, is discussed in the next section. Our account of the intuition behind the conjecture gives a different viewpoint than Stam's own, so readers will find an additional perspective by consulting [7]. The rest of the paper does not rely on this section's discussion.

We keep the definitions of F, $f_1, \ldots, f_r$ as in Section 2. Let

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

Thus $\mathsf{Yield}(q)$ is a lower bound for the number of F-inputs an adversary can learn to evaluate with q queries to each primitive f_i (Corollary 1). However, $\mathsf{Yield}(q)$ may badly underestimate this number of inputs. For example an adversary can always learn to evaluate at least q inputs in q queries to each of the f_i's, whereas $\mathsf{Yield}(q)$ goes to zero for large r as long as (say) $q < 2^{n-1}$. A better (and in fact fairly accurate) lower bound is

$$\mathsf{BYield}(q) = \max\left(q, \mathsf{Yield}(q)\right)$$

where 'B' is for 'better'. Since

$$\mathsf{Yield}(q) \geq q \iff 2^{m+s}\left(\frac{q}{2^n}\right)^r \geq q \iff q \geq 2^{(nr-m-s)/(r-1)}$$

(where we use $r > 1$) we have more exactly that

$$\mathsf{BYield}(q) = \begin{cases} q & \text{if } q \leq 2^{(nr-m-s)/(r-1)}, \\ \mathsf{Yield}(q) & \text{if } q \geq 2^{(nr-m-s)/(r-1)}. \end{cases}$$

Notice[5] that as long as $q < 2^{(nr-m-s)/(r-1)}$ one may increase m or s without affecting $\mathsf{BYield}(q)$, whereas if $q \geq 2^{(nr-m-s)/(r-1)}$ increasing 2^{m+s} by a factor c increases

[5] It is also instructive to note that the threshold $q = 2^{(nr-m-s)/(r-1)}$ occurs when the adversary of Lemma 1 learns on average the value of exactly one input with each query it makes to f_r. Indeed,

$$2^{m+s}\left(\frac{q}{2^n}\right)^r = q \iff 2^{m+s}\left(\frac{q}{2^n}\right)^{r-1} = 2^n$$

meaning that with $q = 2^{(nr-m-s)/(r-1)}$ queries to $f_1, \ldots, f_{r-1}$ the adversary will have 2^n "surviving inputs" for which it knows the first $r-1$ intermediate chaining values, or on average one input for each point in the domain of f_r.

$\mathsf{BYield}(q)$ by that much; for example increasing s by 1, which doubles the size of the range, also doubles the size of $\mathsf{BYield}(q)$.

Empirically, one might estimate that the chance of finding a collision for a given value of q is lower bounded by

$$\mathsf{BYield}^2(q)/2^s$$

since a birthday attack which learns t outputs in a range of size 2^s has chance approximately $t^2/2^s$ of yielding a collision. This is correct when $\mathsf{BYield}(q) = q$, since then the adversary can independently sample each input point for which it chooses to learn the output, but when $\mathsf{BYield}(q) > q$ the inputs for which the adversary learns the output are not independently sampled, and, hence, it is not clear the attack works (indeed it is in fact easy to construct an artificial compression function F that will fool the deterministic adversary of Lemma 1 in this regard). Roughly speaking, Rogaway and Steinberger [5] say that a compression function F is *collision uniform* if learning to evaluate F on any t inputs gives chance $\approx t^2/2^s$ of obtaining a collision. Since a random F has this property, they argue that so should most cryptographically good constructions (i.e. constructions of interest). So far this thesis seems to bear out for all real-world constructions with $r > 1$. The $1.5n$-bit to n-bit JHash compression function (Fig. 1) is a nice example of a non-collision-uniform compression function with $r = 1$: a single query to the underlying permutation already allows the evaluation of $t = 2^{n/2}$ inputs, but one must actually make $q = 2^{n/4}$ queries on average to the permutation before finding a collision (at which point $2^{n/2}$ different collisions are found at once). One can also note the JHash compression function is quite "non-random", as $2^{n/2}$ input-output pairs can be deduced from any single input-output pair.

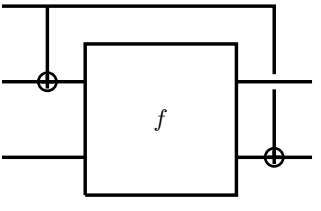

Fig. 1. The JH compression function from $\{0,1\}^{1.5n}$ to $\{0,1\}^n$. All wires carry $n/2$-bit values.

In any case, let us momentarily (and heuristically) assume that adversaries have chance $\mathsf{BYield}^2(q)/2^s$ of obtaining a collision in q queries. If so, the collision resistance of F will be (r times) the least q such that $\mathsf{BYield}(q) = 2^{s/2}$. If $2^{s/2} \leq 2^{(nr-m-s)/(r-1)}$ this is $2^{s/2}$, otherwise it is the solution to

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

which is $q = 2^{(nr-m-\frac{s}{2})/r}$. Thus, noting $\mathsf{HeuristicSec}(m,n,r,s)$ this "heuristic maximum collision security", we have

$$\mathsf{HeuristicSec}(m,n,r,s) = \begin{cases} r2^{s/2} & \text{if } 2^{s/2} \leq 2^{(nr-m-s)/(r-1)}, \\ r2^{(nr-m-\frac{s}{2})/r} & \text{if } 2^{s/2} \geq 2^{(nr-m-s)/(r-1)}. \end{cases}$$

Now consider m, n, r as fixed and s as variable. Note that for sufficiently large s we will be in the second case, $2^{s/2} \geq 2^{(nr-m-s)/(r-1)}$. Also note that if we increase s while in the second case, $\mathsf{HeuristicSec}$ *decreases*[6]. However, as noted by Stam, increasing the state size s should never decrease the best-possible collision security of a compression function, as additional input bits can always be forwarded to the output as the identity without affecting collision security. This shows that $\mathsf{HeuristicSec}$ is *provably not* the correct maximum collision security for the range $2^{s/2} \geq 2^{(nr-m-s)/(r-1)}$.

This leaves us with the question of determining the "real" collision security when $2^{s/2} \geq 2^{(nr-m-s)/(r-1)}$. Still thinking of m, n, r as fixed and s as variable, Stam conjectured that as s increases collision security simply "tops off" when $2^{s/2}$ reaches $2^{(nr-m-s)/(r-1)}$ and remains constant afterwards. We have $2^{(nr-m-s)/(r-1)} = 2^{s/2}$ when $s = s_0 = 2(nr-m)/(r+1)$, meaning that collision security can never exceed $r2^{(nr-m)/(r+1)}$ according to this conjecture (or more precisely, since q must be kept integer, that collision security can never exceed $r\lceil 2^{(nr-m)/(r+1)} \rceil$). Succinctly put, while the heuristic attack gives an incorrect bound, it still manages to "freeze" collision security at the point where the attack comes into effect.

Summarizing, Stam's conjecture for $r > 1$ stipulates the "true maximum collision security" $\mathsf{TrueSec}(m, n, r, s)$ is

$$\begin{aligned}\mathsf{TrueSec}(m, n, r, s) &= \begin{cases} r2^{s/2} & \text{if } 2^{s/2} \leq 2^{(nr-m-s)/(r-1)} \\ r\lceil 2^{(nr-m)/(r+1)} \rceil & \text{if } 2^{s/2} \geq 2^{(nr-m-s)/(r-1)} \end{cases} \\ &= \min(r2^{s/2}, r\lceil 2^{(nr-m)/(r+1)} \rceil)\end{aligned}$$

up to some small multiplicative constant. Since $r2^{s/2}$ queries obviously do suffice for finding a collision with probability $\frac{1}{2}$ (up to said small multiplicative constant), the problem reduces to showing that $r\lceil 2^{(nr-m)/(r+1)} \rceil$ queries also always suffice.

4 Intuition for $r = 1$ and Reduction to $r = 1$

For $r = 1$ the conjectured maximum collision security is again

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

but a separate explanation is required. Note that when $r = 1$ an adversary can learn to evaluate F on at least $2^{m+s-n}q$ inputs in $q \leq 2^n$ queries to the (unique) primitive f_1. If $m \geq n$ this gives a 2-query attack, so we may assume $m \leq n$. If $n \geq m + s$ then $2^{(n-m)/2} \geq 2^{s/2}$ is more than the cost of a birthday attack, so we may also assume $n \leq m + s$.

We now argue the bound of $2^{(n-m)/2}$ queries "by example" for the case $m \leq n \leq m + s$ by showing a construction collision secure up to that many queries. As each input to f_1 corresponds on average to 2^{m+s-n} inputs from the domain $\{0,1\}^{m+s}$, it is natural to write the domain as $\{0,1\}^{m+s-n} \times \{0,1\}^n$, and to have $g_1(x||y) = y$ for

[6] This can be seen as a consequence of the fact that $\mathsf{BYield}(q)$ is proportional to 2^s when $q \geq 2^{(nr-m-s)/(r-1)}$, and that the chance of obtaining a collision is estimated as $\mathsf{BYield}^2(q)/2^s$, so that increasing s actually increases this ratio.

any $x \in \{0,1\}^{m+s-n}$ and $y \in \{0,1\}^n$ (this at least "balances" g_1 across the domain). Since we do not want the adversary to obtain a collision from a single query $f_1(y)$, we "reserve" $m+s-n$ output bits for the portion of the domain which does not affect y; namely we set $F(x||y) = x||z$ where z is the truncation to $s-(m+s-n) = n-m$ bits of $f_1(y)$, where we can assume f_1 has output length $b \geq n-m$. To find a collision the adversary only needs to find a collision in the last $n-m$ bits of output (and can then adjust the first $m+s-n$ bits as it wants), leading to collision resistance of $2^{(n-m)/2}$.

Crucially to the results of this paper, certain cases $r > 1$ of the conjecture reduce to the case $r = 1$. Assume $r > 1$. By Lemma 1, an adversary making $q = 2^{(nr-m)/(r+1)}$ queries to each $f_1, \ldots, f_{r-1}$ can learn the first $r-1$ chaining variables for at least

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

inputs to F. Let A be the set of these inputs. Consider the compression function $F' : A \to \{0,1\}^s$ defined by $F'(v) = F(v)$. Let $m' = (2m-n(r-1))/(r+1)$. If $m' \geq 1$ then we may view F' as a compression function from $\{0,1\}^{m'+s}$ bits to $\{0,1\}^s$ making a single call to a primitive of n-bit input, namely f_r (when m' is non-integral we simply mean that F' has domain of size at least $2^{m'+s}$). According to the case $r = 1$ of Stam's conjecture, $2^{(n-m')/2}$ queries to f_r should suffice for finding a collision in F'. However,

$$2^{(n-m')/2} = 2^{(n-\frac{2m-n(r-1)}{r+1})/2} = 2^{(nr-m)/(r+1)} = q,$$

the number of queries allotted to $f_1, \ldots, f_{r-1}$. Thus if Stam's conjecture holds for $r = 1$ and for non-integral $m \geq 1$ (to allow non-integral m') then it more generally holds whenever $(2m-n(r-1))/(r+1) \geq 1$. We make this idea more formal in the next section.

5 Main Result

We first prove Stam's conjecture for $r = 1$ and $m \geq \log_2(17) \approx 4.09$. The more general result will follow as a corollary via the reduction outlined at the end of the previous section.

Clearly the fact that the compression function F manipulates bit strings is unimportant: the determining factors are the size of the domain, the size of the range, and the size of f's domain. We let the size of F's domain and range be MS and S, respectively, where S is a positive integer and $M \geq 2$. If MS is non-integral then our meaning is that F has domain of size *at least* MS (so $\lceil MS \rceil$ or more). The size of f's domain will be N. Thus under our original notation, $M = 2^m$, $S = 2^s$ and $N = 2^n$. For $r = 1$, the object is to show that $\approx 2^{(n-m)/2} = \sqrt{N/M}$ queries to f suffice for finding a collision in F.

Our collision attack ultimately reduces to a birthday attack. To make fully precise what we mean by a "birthday attack" let $B : D_B \to R_B$ be any fixed function of finite domain D_B and finite range R_B. Then performing a q-query birthday attack on B

means evaluating B at q points of D_B sampled uniformly without replacement, halting when a collision is found. We use the following proposition due to Wiener [8] lower bounding the probability of success of a birthday attack:

Proposition 1. (cf. [8] Theorem 7) *Let $B : D_B \to R_B$ such that D_B, R_B are finite and $D_B \geq 2R_B$. Then a q-query birthday attack on B has chance at least $1 - 3e^{-2} > 0.5$ of success when $q \geq 2\sqrt{2R_B} + 1$.*

We can now state and prove our main technical result:

Theorem 1. *Let S, N be positive integers and let $M \geq 17$ be a real number such that $N/M \geq 128$. Let F be a compression function of domain of size at least MS and range of size S making a single call to a primitive f of domain of size N. Then a collision can be found for F with probability at least 0.5 in $q = \lceil 4\sqrt{8N/M} \rceil$ queries to f.*

Proof. Let $w = MS/2N$ and let $b = \lceil 4S/w \rceil = \lceil 8N/M \rceil$.

Let D_F, R_F denote the domain and range of F and let D_f denote the domain of f. For each $x \in D_f$ let $T_x = \{y \in D_F : g_1(y) = x\}$ (namely T_x is the set of F-inputs that can be evaluated once f is queried at x). Let $W = \{x \in D_f : |T_x| \geq w\}$. Note the adversary can compute W.

For each $x \in W$ the adversary divides T_x into sets $T_x^1, \ldots, T_x^{j_x}$ such that each $w \leq |T_x^i| < 2w$ for $i = 1 \ldots j_x$. Let U be the set of all these sets, namely $U : \{T_x^i : x \in W, 1 \leq i \leq j_x\}$. The adversary's attack will consist in repeatedly choosing without replacement a random element T_x^i from U uniformly among the elements of U that have not yet been chosen and querying f at x if f has not yet been queried at that point, until either q queries have been made or until no elements are left in U.

We lower bound the adversary's chance of finding a collision with this attack. In fact, we will only give the adversary credit if it finds a collision for inputs that belong to sets that it chose from U, so we more precisely lower bound the probability of the latter event happening.

Let $U_1 = \{T_x^i \in U : |F(T_x^i)| = |T_x^i|\}$ and let $U_2 = U \backslash U_1 = \{T_x^i \in U : |F(T_x^i)| < |T_x^i|\}$. Thus U is the disjoint union of U_1 and U_2. For $T_x^i \in U_1$ consider the event $A_{T_x^i}$ that b random elements of R_F chosen uniformly with replacement do not intersect $F(T_x^i)$. Since $|F(T_x^i)| = |T_x^i| \geq w$ and $|R_F| = S$, we have

$$\Pr[A_{T_x^i}] \leq \left(1 - \frac{w}{S}\right)^b \leq e^{-4} \leq 0.02.$$

Thus there exists some set of b values $\{r_1, \ldots, r_b\} \subseteq R_F$ such that at least 0.98 of the sets in U_1 contain one of the values $r_1, \ldots, r_b$.

Let $D'_F = \bigcup_{x \in W} T_x$. Since $\sum_{x \notin W} |T_x| \leq Nw$ we have $|D'_F| \geq MS - Nw$. Since each element of U is a set of size at most $2w$ and since $D'_F = \bigcup_{T_x^i \in U} T_x^i$, we have

$$|U| \geq \frac{|D'_F|}{2w} \geq \frac{MS - Nw}{2w} = \frac{1}{2}\left(\frac{MS}{w} - N\right)$$

and so $0.98|U| \geq 2b$, since

$$\begin{aligned} 0.98|U| \geq 2b &\impliedby \frac{0.98}{2}\left(\frac{MS}{w} - N\right) \geq 4S/w + 2 \\ &\iff 0.49N \geq 8N/M + 2 \\ &\iff 0.49M \geq 8 + 2M/N \\ &\iff M(0.98 - \frac{4}{N}) \geq 16 \\ &\impliedby M \geq 17 \end{aligned}$$

using $N \geq 128M \geq 128 \cdot 17$ for the last implication.

We say that a set T_x^i chosen by the adversary during its attack (as described above) is "lost" if $T_x^i \in U_1$ and $T_x^i \cap \{r_1, \ldots, r_b\} = \emptyset$. Since $|U| \geq 2b$ and $q \leq 4\sqrt{b} + 1$, any set chosen by the adversary has probability at most

$$\begin{aligned} \frac{0.02|U|}{|U| - q} &\leq \frac{0.02(2b)}{2b - 4\sqrt{b} - 1} \\ &= \frac{0.04}{2 - 4/\sqrt{b} - 1/b} \\ &= \frac{0.04}{2 - 4/32 - 1/1024} \\ &\leq 0.0214 \end{aligned}$$

of being lost independently of the result of previous choices, using $b \geq 8N/M \geq 1024$. By a multiplicative Chernoff bound, the probability that total number of non-lost sets is less than $0.8(1 - 0.0214)q = 0.8 \cdot 0.9786q = 0.78288q$ is therefore at most

$$e^{-\frac{0.9786q0.2^2}{2}} \leq e^{-2.505}$$

using $q \geq 4\sqrt{8N/M} \geq 128$. Thus with chance at least $1 - e^{-2.505} \geq 0.918$, the adversary chooses at least $0.78288q \geq 3\sqrt{b}$ non-lost sets.

The theorem follows by ascribing to each non-lost element of U_1 an element of $\{r_1, \ldots, r_b\}$ that it contains and to each element of U_2 an arbitrary element of $\{r_1, \ldots, r_b\}$, and noting that the adversary wins if it ever chooses two (non-lost) elements of U that are ascribed the same element of $\{r_1, \ldots, r_b\}$. (Indeed, if the adversary ever chooses an element of U_2, it finds a collision automatically.) Thus the adversary's attack becomes a birthday attack on a function of domain at least $0.98|U| \geq 2b$ and range b, in which the adversary queries at least $3\sqrt{b} \geq 2\sqrt{2b} + 1$ independent domain points of the function with probability at least 0.918. By Proposition 1 the latter number of queries is sufficient to find a collision with probability at least $1 - 3e^{-2} \geq 0.5/0.918$, thus concluding the proof. □

Corollary 2. *Let S, N be positive integers and let $M \geq 17$. Let F be a compression function of domain of size at least MS and range of size S making a single call to a primitive f of domain of size N. Then a collision can be found for F with probability at least 0.5 in*

$$q = \begin{cases} 2175 & \text{if } N/17 < 128 \\ 128 & \text{if } N/17 \geq 128 \text{ and } N/M < 128 \\ \lceil 4\sqrt{8N/M} \rceil & \text{if } N/M \geq 128 \end{cases}$$

queries to f.

Proof. The last case is Theorem 1 and the first case is obvious since $N < 17 \cdot 128 = 2176$ when $N/17 < 128$, and f has domain of size N. For the second case, it suffices to observe that we can apply Theorem 1 to a restricted version F' of F, where F' is the restriction of F to a domain $D'_{F'} \subseteq D_F$, $|D'_{F'}| = M'S$ where $M' = N/128 \geq 17$. In the latter case, the cost of the Theorem 1 attack on F' is $q = 4\lceil\sqrt{8N/M'}\rceil = 128$. □

The next corollary is the paper's main result:

Corollary 3. *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 if* $m' = (2m - n(r-1))/(r+1) \geq \log_2(17)$*, an adversary making*

$$q = (r-1)\lceil 2^{(nr-m)/(r+1)}\rceil + \begin{cases} 2175 & \text{if } 2^n/17 < 128 \\ 128 & \text{if } 2^n/17 \geq 128 \text{ and } n - m' < 7 \\ \lceil 8\sqrt{2} \cdot 2^{(nr-m)/(r+1)}\rceil & \text{if } n - m' \geq 7 \end{cases}$$

queries to the f_i*'s can find a collision for* F *with probability* > 0.5.

Proof. As shown at the end of section 4, an adversary making $q_0 = \lceil 2^{(nr-m)/(r+1)}\rceil$ queries to each of the functions $f_1, \ldots, f_{r-1}$ can learn the intermediate chaining values $y_1, \ldots, y_{r-1}$ for at least $2^{s+m'}$ inputs. We then consider the restriction F' of F to those inputs as a single-call compression function. F' has a domain of size MS and a range of size S where $S = 2^s$, $M = 2^{m'} \geq 17$, and uses a primitive of domain $N = 2^n$. The result then follows from Corollary 2 by noting that $\sqrt{N/M} = 2^{(n-m')/2} = 2^{(nr-m)/(r+1)}$. □

Acknowledgements

The author would like to thank the referees for their careful read and Greg Kuperberg for helpful discussions.

References

1. Bellare, M., Kohno, T.: Hash function imbalance and its impact on birthday attacks. In: Cachin, C., Camenisch, J.L. (eds.) EUROCRYPT 2004. LNCS, vol. 3027, pp. 401–418. Springer, Heidelberg (2004)
2. Black, J., Cochran, M., Shrimpton, T.: On the Impossibility of Highly-Efficient Blockcipher-Based Hash Functions. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 526–541. Springer, Heidelberg (2005)
3. Gauravaram, P., Knudsen, L., Matusiewicz, K., Mendel, F., Rechberger, C., Schläffer, M., Thomsen, S.: Grøstl, a SHA-3 candidate, NIST SHA-3 competition submission (October 2008)

4. Rogaway, P., Steinberger, J.: Constructing cryptographic hash functions from fixed-key blockciphers. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 433–450. Springer, Heidelberg (2008)
5. Rogaway, P., Steinberger, J.: Security/Efficiency Tradeoffs for Permutation-Based Hashing. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 220–236. Springer, Heidelberg (2008)
6. Shrimpton, T., Stam, M.: Building a Collision-Resistant Compression Function from Non-Compressing Primitives. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) ICALP 2008, Part II. LNCS, vol. 5126, pp. 643–654. Springer, Heidelberg (2008), Cryptology ePrint Archive: Report 2007/409
7. Stam, M.: Beyond uniformity: Better Security/Efficiency Tradeoffs for Compression Functions. In: Wagner, D. (ed.) CRYPTO 2008. LNCS, vol. 5157, pp. 397–412. Springer, Heidelberg (2008)
8. Wiener, M.: Bounds on birthday attack times. Cryptology ePrint archive (2005)
9. Wu, H.: The JH hash function, NIST SHA-3 competition submission (October 2008)

Appendix

A An Alternate Approach for Random Primitives

In this section we give an alternate proof of (a version of) Theorem 1 when the primitive $f = f_1$ of the compression function F is random, or more exactly when its outputs are independently distributed from each other (though not necessarily uniformly distributed across the range of f). This alternate version implies corollaries similar to corollaries 2 and 3, which we do not list. We present this alternate proof partly because some may find it more intuitive than the proof of Theorem 1 and partly because of the intrinsic interest of a supporting lemma, whose content and proof technique are of independent interest from the rest of the paper.

We start by stating this lemma, which we dub the 'MECMAC' lemma for 'Many Expected Collisions Means A Collision'.

Lemma 2 (MECMAC). *Let S be a set and let $c \leq |S|$ be a positive integer. Let $X_1, \ldots, X_n$ be independent random variables whose values are subsets of S of size at most c. Let $X = \sum_{i<j} |X_i \cap X_j|$ and let $\mu = E[X]$. Then*

$$\Pr[X = 0] \leq e^{-\mu/4c} + e^{-\sqrt{\mu/2c}} + \sqrt{\mu/2c}\, e^{1-\frac{3}{28}\sqrt{\mu/2c}}.$$

We do not believe the bound of Lemma 2 is sharp; we expect the optimal upper bound for $\Pr[X = 0]$ to be closer to $(1 + \sqrt{2\mu/c})e^{-\sqrt{2\mu/c}}$, but we could not achieve this bound with our current proof technique. Note Lemma 2 has a statement of the form: "Let $X_1, \ldots, X_n$ be independent random variables, and let $\mu = \sum_{i<j} f_{ij}(X_i, X_j)$ where $f_{ij} : Range(X_i) \times Range(X_j) \rightarrow [0, c]$. Then if μ/c is large, $\Pr[\sum f_{ij}(X_i, X_j) = 0]$ is small". However, this more general type of statement is not true, as can be seen from easily-constructed counterexamples. Thus Lemma 2 crucially relies on structural properties of set intersections (and in particular on the fact that if many sets intersect a single one, these are also likely to intersect each other).

Our alternate version of Theorem 1 for random primitives is the following:

Theorem 2. *Let S, N be positive integers and let $M \geq 16$ be a real number. Let F be a compression function of domain of size at least MS and range of size S making a single call to a primitive f of domain of size N whose outputs are independently distributed. Let $E \geq 16$ be such that $q = 1 + E\sqrt{N/M}$ is an integer and let $\psi = E/4$. Then if $1 + \lceil \log\log N \rceil < \frac{3}{8}M$ a collision can be found for F with probability at least $1 - g(\psi)$ where*

$$g(\psi) = e^{-\psi/4} + e^{-\sqrt{\psi/2}} + \sqrt{\psi/2}e^{1-\frac{3}{28}\sqrt{\psi/2}}$$

by using q queries to f.

Note: The constraint $\lceil \log\log N \rceil < \frac{3}{8}M$ does not correspond to any constraint in Theorem 1. In practice N is around 2^{128}, say, in which case $\lceil \log\log N \rceil < \frac{3}{8}M$ becomes $M > 64/3$, which is not much more restrictive than $M \geq 16$.

Proof of Theorem 2. Let D_F, R_F be the domain and range of F, and let g, h be the deterministic functions such that $F(v) = h(v, f(g(v)))$. Also let D_f be the domain of f. For each $x \in D_f$ let $T_x \subseteq T$ be the set of inputs $v \in D_F$ such that $g(v) = x$. Thus if the adversary makes the query $f(x)$ it learns to evaluate $F(v)$ for all $v \in T_x$.

For each $x \in D_f$ we let $\overline{X}_x = F(T_x)$. Note $\overline{X}_x$ is a random variable that depends on $f(x)$ and whose value is a subset of R_F. Then $\{\overline{X}_x : x \in D_f\}$ is an independent set of random variables. Let Coll_x be the event that a collision occurs among the inputs in T_x, namely that $|\overline{X}_x| < |F(T_x)|$. If $\Pr[\mathsf{Coll}_x] = 1$ for some x the adversary can simply query $f(x)$, so we may assume $\Pr[\mathsf{Coll}_x] < 1$ for all $x \in D_f$. (This poses the question of how the adversary "knows" the existence of such an x; however since the adversary is chosen after the parameters m, n, r, s and the distribution for f is fixed, the value of x may be hardcoded. Similar remarks apply to further points in the proof.) Let $X_x = \overline{X}_x | \neg\mathsf{Coll}_x$ be the modified random variable whose distribution is conditioned on the event $\neg\mathsf{Coll}_x$. Thus $|X_x| = |F(T_x)|$ and $\{X_x : x \in D_f\}$ is an independent set of random variables. We will exhibit a set $Z \subseteq D_f$ of size q such that

$$\Pr[X_x \cap X_y = \emptyset \text{ for all } x, y \in Z, x \neq y] \leq g(\psi).$$

This will prove the theorem since the adversary can query f at all the points in Z, and since the adversary obtains a collision anyway if Coll_x occurs for some $x \in Z$.

Define a sequence $\beta_0, \beta_1, \beta_2, \ldots$ by

$$\beta_k = E^{2^{k-1}-1}MS/N$$

for $k \geq 0$. Note that $\beta_{k+1} = \beta_k^2 EN/MS$. Let $U_k = \{x \in D_f : \beta_k < |T_x| \leq \beta_{k+1}\}$ and let $\Sigma_k = \sum_{x \in U_k} |T_x|$ for all $k \geq 0$.

Let $t \geq 0$ be the least integer such that $\beta_{t+1} \geq S$. Then

$$\begin{aligned} t &\leq \lceil \log(1 + \log(N/M)/\log(E)) \rceil \\ &\leq \lceil \log\log N \rceil \end{aligned}$$

using $M, E \geq 16$. Note we cannot have $|U_k| > 0$ for $k > t$, or else $\Pr[\mathsf{Coll}_x] = 1$ for $x \in U_k$. If $|\Sigma_k| \leq 2S$ for all $k \geq 0$ then because $E^{-\frac{1}{2}} \leq \frac{1}{4}$ and $1 + \lceil \log\log N \rceil < \frac{3}{8}M$,

$$\begin{aligned}
|T| &= \sum_{x \in D_f} |T_x| \\
&\leq N\beta_0 + \sum_{k=0}^{t} |\Sigma_k| \\
&\leq N\beta_0 + (t+1)2S \\
&\leq E^{-\frac{1}{2}} MS + 2(1 + \lceil \log\log N \rceil)S \\
&< S\left(\frac{1}{4}M + \frac{3}{4}M\right) \\
&= MS
\end{aligned}$$

a contradiction. Thus there must exist a value k_0 such that $\Sigma_{k_0} \geq 2S$.

If $|U_{k_0}| \leq q$ then the adversary can query f at all points in U_{k_0} and obtain a collision with probability 1, so we may assume $|U_{k_0}| \geq q$. Now consider the following two experiments:

(1) query f at all points in U_{k_0}, resulting in values of X_x for $x \in U_{k_0}$, then select q distinct sets $X_{x_1}, \ldots, X_{x_q}$ uniformly at random from $\{X_x : x \in U_{k_0}\}$, and remove the other sets
(2) query f at q distinct random points $x_1, \ldots, x_q$ in U_{k_0}, resulting in q known sets $X_{x_1}, \ldots, X_{x_q}$

Clearly these two experiments have identical outcomes. For each experiment, let a "collision" be a triple (i, j, t) with $i < j$ such that $t \in X_{x_i} \cap X_{x_j}$. We will show that in experiment (1) the expected number of collisions is at least $\psi\beta_{k_0+1}$ and hence that there exists some set Z of q distinct values $x_1, \ldots, x_q \in U_{k_0}$ such that the expected number of collisions among $X_{x_1}, \ldots, X_{x_q}$ is at least $\psi\beta_{k_0+1}$.

Let $\Sigma_{k_0} = aS$ where $a \geq 2$. After the first stage of experiment (1) it is easy to see (even when a is not an integer) that there are at least $\frac{a(a-1)}{2}S \geq a^2S/4$ collisions among the sets $\{X_x : x \in U_{k_0}\}$. When selecting q distinct sets at random from the set of $|U_{k_0}|$ sets, each collision remains selected with probability at least $\frac{q(q-1)}{|U_{k_0}|(|U_{k_0}|-1)} \geq (q-1)^2/|U_{k_0}|^2$, so by linearity of expectation the expected number of collisions in experiment (1) is at least $a^2S(q-1)^2/4|U_{k_0}|^2$. Since $|U_{k_0}|\beta_{k_0} \leq \Sigma_{k_0} = aS$, we have $|U_{k_0}| \leq aS/\beta_{k_0}$, so we have

$$\begin{aligned}
\frac{a^2S(q-1)^2}{4|U_{k_0}|^2} &\geq \frac{a^2S(q-1)^2}{4a^2S^2/\beta_{k_0}^2} \\
&= \frac{\beta_{k_0}^2(q-1)^2}{4S} \\
&= \frac{\beta_{k_0}^2 E^2 N}{4MS} \\
&= \psi\beta_{k_0+1}
\end{aligned}$$

where we used $\beta_{k+1} = \beta_k^2 EN/MS$ and $\psi = E/4$.

By the probabilistic argument outlined earlier, there therefore exist a set Z of q distinct points $x_1, \ldots, x_q$ such that the expected number of collisions among

$X_{x_1}, \ldots, X_{x_q}$ is at least $\psi\beta_{k_0+1}$. However by the definition of U_{k_0} we have $|X_{x_i}| \leq \beta_{k_0+1}$ for $i = 1 \ldots q$, so, because $X_{x_1}, \ldots, X_{x_q}$ are independent, Lemma 2 applied with $\mu = \psi\beta_{k_0+1}$ and $c = \beta_{k_0+1}$ implies the probability of no collisions among them is at most $g(\psi)$, as desired. □

Proof of the MECMAC Lemma. Because the bound is void for $\mu \leq 2c$ we can assume $\mu \geq 2c$. For any partition $\mathcal{C}, \mathcal{D}$ of $[n] = \{1, 2, \ldots, n\}$ let

$$X_{\mathcal{C},\mathcal{D}} = |\{(i, j, s) : s \in X_i \cap X_j \text{ and } (i, j) \in (\mathcal{C} \times \mathcal{D}) \cup (\mathcal{D} \times \mathcal{C})\}|$$

and let $\mu_{\mathcal{C},\mathcal{D}} = E[X_{\mathcal{C},\mathcal{D}}]$. If $\mathcal{C}, \mathcal{D}$ are selected at random by independently placing each element of $[n]$ in $\mathcal{C}$ or $\mathcal{D}$ with probability $\frac{1}{2}$ then

$$E[\mu_{\mathcal{C},\mathcal{D}}] = \frac{1}{2}\mu$$

since for each triplet (i, j, s) such that $s \in X_i \cap X_j$ and $i \neq j$ there is chance $\frac{1}{2}$ that $(i, j) \in (\mathcal{C} \times \mathcal{D}) \cup (\mathcal{D} \times \mathcal{C})$. Therefore there must exist a partition $\mathcal{A}, \mathcal{B}$ of $[n]$ such that $\mu_{\mathcal{A},\mathcal{B}} \geq \frac{1}{2}\mu$.

Let $k = |\mathcal{A}|$, $\ell = |\mathcal{B}|$. We rename $X_1, \ldots, X_n$ as two lists $A_1, \ldots, A_k$ and $B_1, \ldots, B_\ell$ such that $\{A_1, \ldots, A_k\} = \{X_i : i \in \mathcal{A}\}$ and $\{B_1, \ldots, B_\ell\} = \{X_j : j \in \mathcal{B}\}$. For $1 \leq i \leq k$ let $Y_i = |\{(j, s) : s \in A_i \cap B_j\}|$ and let $\mu_i = E[Y_i]$. Then

$$\sum_{i=1}^{k} \mu_i \;=\; \mu_{\mathcal{A},\mathcal{B}}.$$

For all $U \subseteq S$ let

$$\beta_U = \sum_{j=1}^{\ell} E\big[|B_j \cap U|\big].$$

We have

$$\mu_i = \sum_{U \subseteq S} \beta_U \Pr[A_i = U] = E[\beta_{A_i}].$$

Let $M = \sqrt{\mu_{\mathcal{A},\mathcal{B}}/c} \geq \sqrt{\mu/2c} \geq 1$. Assume first there is some $s \in S$ such that $\beta_s > M$. Then letting $\alpha_j = \Pr[B_j = s] = E[|B_j \cap \{s\}|]$ we have $\alpha_1 + \cdots + \alpha_\ell > M$ and

$$\begin{aligned}
\Pr[X = 0] &\leq \prod_{j=1}^{\ell}(1 - \alpha_j) + \sum_{j=1}^{\ell} \alpha_j \prod_{h=1, h\neq j}^{\ell} (1 - \alpha_h) \\
&\leq e^{-\alpha_1 - \ldots - \alpha_\ell} + \sum_{j=1}^{\ell} \alpha_j e^{\alpha_j - \alpha_1 - \ldots - \alpha_\ell} \\
&\leq e^{-M} + e^{1-\alpha_1-\ldots-\alpha_\ell} \sum_{j=1}^{\ell} \alpha_j \\
&\leq e^{-M} + Me^{1-M} \\
&\leq e^{-\sqrt{\mu/2c}} + \sqrt{\mu/2c}\, e^{1-\sqrt{\mu/2c}}
\end{aligned}$$

where the last two inequalities use the fact that ye^{-y} is a decreasing function of y for $y \geq 1$.

Now assume instead that $\beta_s \leq M$ for all $s \in S$. Since

$$\beta_{A_i} = \sum_{s \in A_i} \beta_s \leq Mc,$$

β_{A_i} is a nonnegative r.v. bounded by Mc of mean μ_i for $1 \leq i \leq k$, so

$$\mathrm{Var}(\beta_{A_i}) \leq \mu_i(Mc - \mu_i)$$

for $1 \leq i \leq k$ and

$$\begin{aligned}\sum_{i=1}^{k} \mathrm{Var}(\beta_{A_i}) &\leq \sum_{i=1}^{k} \mu_i(Mc - \mu_i) \\ &\leq Mc \sum_{i=1}^{k} \mu_i \\ &= c^{\frac{1}{2}} \mu_{\mathcal{A},\mathcal{B}}^{\frac{3}{2}}.\end{aligned}$$

Because $\beta_{A_1}, \ldots, \beta_{A_k}$ are independent and uniformly bounded by Mc, Bernstein's inequality (see notes at bottom) then implies

$$\begin{aligned}\Pr\left[\sum_{i=1}^{k} \beta_{A_i} \leq \mu_{\mathcal{A},\mathcal{B}}/2\right] &\leq \exp\left(-\frac{(\mu_{\mathcal{A},\mathcal{B}}/2)^2/2}{\sum_{i=1}^{k} \mathrm{Var}(\beta_{A_i}) + Mc\mu_{\mathcal{A},\mathcal{B}}/6}\right) \\ &\leq \exp\left(-\frac{\mu_{\mathcal{A},\mathcal{B}}^2/8}{c^{\frac{1}{2}}\mu_{\mathcal{A},\mathcal{B}}^{\frac{3}{2}} + c^{\frac{1}{2}}\mu_{\mathcal{A},\mathcal{B}}^{\frac{3}{2}}/6}\right) \\ &\leq e^{-\frac{3}{28}(\mu_{\mathcal{A},\mathcal{B}}/c)^{\frac{1}{2}}} \\ &\leq e^{-\frac{3}{28}\sqrt{\mu/c}}.\end{aligned}$$

Let "$\Sigma_{\geq}$" be the event that $\sum_{i=1}^{k} \beta_{A_i} \geq \mu_{\mathcal{A},\mathcal{B}}/2$ and let "$A_{\neq}$" be the event that $A_i \cap A_j = \emptyset$ for $i \neq j$. We have

$$\begin{aligned}\Pr[X = 0] &= \Pr[X = 0 \mid \Sigma_{\geq}] \Pr[\Sigma_{\geq}] + \Pr[X = 0 \mid \neg\Sigma_{\geq}] \Pr[\neg\Sigma_{\geq}] \\ &\leq \Pr[X = 0 \mid \Sigma_{\geq}] + \Pr[\neg\Sigma_{\geq}] \\ &\leq \Pr[X = 0 \mid \Sigma_{\geq}] + e^{-\frac{3}{28}\sqrt{\mu/c}}\end{aligned}$$

and, since $\neg A_{\neq} \implies X \geq 1$,

$$\begin{aligned}\Pr[X = 0 \mid \Sigma_{\geq}] &= \Pr[X = 0 \mid \Sigma_{\geq} \wedge A_{\neq}] \Pr[A_{\neq} \mid \Sigma_{\geq}] + \\ &\quad \Pr[X = 0 \mid \Sigma_{\geq} \wedge \neg A_{\neq}] \Pr[\neg A_{\neq} \mid \Sigma_{\geq}] \\ &= \Pr[X = 0 \mid \Sigma_{\geq} \wedge A_{\neq}] \Pr[A_{\neq} \mid \Sigma_{\geq}] \\ &\leq \Pr[X = 0 \mid \Sigma_{\geq} \wedge A_{\neq}].\end{aligned}$$

Moreover

$$\Pr[X = 0 \mid \Sigma_{\geq} \wedge A_{\neq}] \leq \prod_{j=1}^{\ell} \Pr[B_j \cap (A_1 \cup \cdots \cup A_k) = \emptyset \mid \Sigma_{\geq} \wedge A_{\neq}].$$

To upper bound the latter probability, fix any values of $A_1, \ldots, A_k$ such that $\Sigma_{\geq} \wedge A_{\neq}$. For $1 \leq j \leq \ell$ let B'_j be a new random variable that selects uniformly at random an element from B_j. Then

$$\begin{aligned}
\prod_{j=1}^{\ell} \Pr[B_j \cap (A_1 \cup \cdots \cup A_k) = \emptyset] &\leq \prod_{j=1}^{\ell} \Pr[B'_j \notin A_1 \cup \cdots \cup A_k] \\
&= \prod_{j=1}^{\ell} \left(1 - \sum_{i=1}^{k} \Pr[B'_j \in A_i]\right) \\
&\leq e^{-\sum_{j=1}^{\ell} \sum_{i=1}^{k} \Pr[B'_j \in A_i]} \\
&= e^{-\sum_{j=1}^{\ell} \sum_{i=1}^{k} E[|B'_j \cap A_i|]} \\
&\leq e^{-\sum_{i=1}^{k} \beta_{A_i}/c} \\
&\leq e^{-\mu_{\mathcal{A},\mathcal{B}}/2c} \\
&\leq e^{-\mu/4c}
\end{aligned}$$

where $A_{\neq}$ is used going to the second line and $\Sigma_{\geq}$ is used in the next-to-last inequality. Thus

$$\Pr[X = 0 \mid \Sigma_{\geq} \wedge A_{\neq}] \leq e^{-\mu/4c}.$$

Combining these results we have

$$\Pr[X = 0] \;\leq\; e^{-\mu/4c} + e^{-\frac{3}{28}\sqrt{\mu/c}}$$

if $\beta_s \leq M$ for all $s \in S$, and

$$\Pr[X = 0] \;\leq\; e^{-\sqrt{\mu/2c}} + \sqrt{\mu/2c}\, e^{1-\sqrt{\mu/2c}}$$

if $\beta_s \geq M$ for some $s \in S$, so we can conclude that

$$\Pr[X = 0] \;\leq\; e^{-\mu/4c} + e^{-\sqrt{\mu/2c}} + \sqrt{\mu/2c}\, e^{1-\frac{3}{28}\sqrt{\mu/2c}}$$

in all cases. □

Bernstein's Inequality

Let $Z_1, \ldots, Z_n$ be independent random variables of mean zero such that $|Z_i| \leq M$ almost surely for $1 \leq i \leq n$. Bernstein's inequality states that

$$\Pr\left[\sum_{i=1}^{n} Z_i \geq t\right] \leq \exp\Big(-\frac{t^2/2}{\sum_{i=1}^{n} E[Z_i^2] + Mt/3}\Big).$$

for all $t > 0$. Now let $T_1, \ldots, T_n$ be independent random variables of nonzero mean such that $T_i \in [0, M]$ almost surely, and let $\mu = E[T_1 + \cdots + T_n]$. By Bernstein's inequality applied to $Z_1 = -(T_1 - E[T_1])$, …, $Z_n = -(T_n - E[T_n])$ (so $|Z_i| \le M$ a.s.) we have

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} T_i \le \mu/2\right] &= \Pr\left[\sum_{i=1}^{n} (T_i - E[T_i]) \le -\mu/2\right] \\
&= \Pr\left[\sum_{i=1}^{n} Z_i \ge \mu/2\right] \\
&\le \exp\Big(-\frac{(\mu/2)^2/2}{\sum_{i=1}^{n} E[Z_i^2] + M\mu/6}\Big) \\
&= \exp\Big(-\frac{(\mu/2)^2/2}{\sum_{i=1}^{n} \mathrm{Var}(T_i) + M\mu/6}\Big).
\end{aligned}$$

This is the form used in the proof of the MECMAC lemma.