Improved Security Analyses for CBC MACs

(preliminary full version)

M. Bellare [*] K. Pietrzak [†] P. Rogaway [‡]

September 2005

Abstract

We present an improved bound on the advantage of any q-query adversary at distinguishing between the CBC MAC over a random n-bit permutation and a random function outputting n bits. The result assumes that no message queried is a prefix of any other, as is the case when all messages to be MACed have the same length. We go on to give an improved analysis of the encrypted CBC MAC, where there is no restriction on queried messages. Letting ℓ be the block length of the longest query, our bounds are about $\ell q^2/2^n$ for the basic CBC MAC and $\ell^{o(1)}q^2/2^n$ for the encrypted CBC MAC, improving prior bounds of $\ell^2q^2/2^n$. The new bounds translate into improved guarantees on the probability of forging these MACs.

Keywords: CBC MAC, message authentication, provable security.

[*] Dept. of Computer Science & Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093 USA. E-mail: mihir@cs.ucsd.edu; WWW: www.cse.ucsd.edu/users/mihir/. Supported by NSF grants ANR-0129617 and CCR-0208842, and by an IBM Faculty Partnership Development Award.

[†] Dept. of Computer Science, ETH Zürich, CH-8092 Zürich Switzerland, E-mail: pietrzak@inf.ethz.ch

[‡] Dept. of Computer Science, University of California, Davis, California, 95616, USA; and Dept. of Computer Science, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand. E-mail: rogaway@cs.ucdavis.edu; WWW: www.cs.ucdavis.edu/∼rogaway/. Most of this work carried out while hosted by the Department of Computer Science, Faculty of Science, Chiang Mai University, Thailand. Currently hosted by the School of Information Technology, Mae Fah Luang University, Thailand. Supported in part by NSF grant CCR-0208842 and a gift from Intel Corp.

Construct	atk	Previous bound	Our bound
CBC	pf	$\ell^2 q^2/2^n$ [2, 13, 15]	$\ell q^2/2^n \cdot (12 + 64\ell^3/2^n)$
ECBC	any	$2.5\,\ell^2 q^2/2^n$ [7]	$q^2/2^n \cdot (d'(\ell) + 32\ell^4/2^n)$

Figure 1: Bounds on $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell)$ and $\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q, n, \ell)$.

1 Introduction

Some definitions. The CBC function CBC_π associated to a key π: $\{0,1\}^n \to \{0,1\}^n$ takes as input a message $M = M^1 \cdots M^m$ that is a sequence of n-bit blocks and returns the n-bit string C^m computed by setting $C^i = \pi(C^{i-1} \oplus M^i)$ for each $i \in [1..m]$, where $C^0 = 0^n$. Consider three types of attacks for an adversary given an oracle: $\mathsf{atk} = \mathsf{eq}$ means all queries are exactly ℓ blocks long; $\mathsf{atk} = \mathsf{pf}$ means they have at most ℓ blocks and no query is a prefix of any another; $\mathsf{atk} = \mathsf{any}$ means the queries are arbitrary distinct strings of at most ℓ blocks. Let $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{CBC}}(q, n, \ell)$ denote the maximum advantage attainable by any q-query adversary, mounting an atk attack, in distinguishing whether its oracle is CBC^π_n for a random permutation π on n bits, or a random function that outputs n bits. We aim to upper bound this quantity as a function of n, ℓ, q.

Past work and our results on CBC. Bellare, Kilian and Rogaway [2] showed that $\mathbf{Adv}^{\mathsf{eq}}_{\mathrm{CBC}}(q, n, \ell) \leq 2\ell^2 q^2/2^n$. Maurer reduced the constant 2 to 1 and provided a substantially different proof [13]. Petrank and Rackoff [15] showed that the same bounds hold (up to a constant) for $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell)$. In this paper we show that $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell) \leq 20\ell q^2/2^n$ for $\ell \leq 2^{n/3}$. (The result is actually a little stronger. See Figure 1.) This implies the same bound holds for $\mathbf{Adv}^{\mathsf{eq}}_{\mathrm{CBC}}(q, n, \ell)$.

Context and discussion. When $\pi = E(K, \cdot)$, where $K \in \mathcal{K}$ is a random key for blockcipher E: $\mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$, the function CBC_π is a popular message authentication code (MAC). Assuming E is a good pseudorandom permutation (PRP), the dominant term in a bound on the probability of forgery in an atk-type chosen-message attack is $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{CBC}}(q, n, \ell)$, where q is the sum of the number of MAC-generation and MAC-verification queries made by the adversary (cf. [1]). Thus the quality of guarantee we get on the security of the MAC is a function of how good an upper bound we can prove on $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{CBC}}(q, n, \ell)$.

It is well known that the CBC MAC is insecure when the messages MACed have varying lengths (specifically, it is forgeable under an any-attack that uses just one MAC-generation and one MAC-verification query, each of at most two blocks) so the case $\mathsf{atk} = \mathsf{any}$ is not of interest for CBC. The case where all messages MACed have the same length ($\mathsf{atk} = \mathsf{eq}$) is the most basic one, and where positive results were first obtained [2]. The case $\mathsf{atk} = \mathsf{pf}$ is interesting because one way to get a secure MAC for varying-length inputs is to apply a prefix-free encoding to the data before MACing it. The most common such encoding is to include in the first block of each message an encoding of its length.

We emphasize that our results are about CBC_π for a random permutation π: $\{0,1\}^n \to \{0,1\}^n$, and not about CBC_ρ for a random function ρ: $\{0,1\}^n \to \{0,1\}^n$. Since our bounds are better than the cost to convert between a random n-bit function and a random n-bit permutation using the switching lemma [2], the distinction is significant. Indeed for the prefix-free case, applying CBC over a random function on n bits is known to admit an attack more effective than that which is ruled out by our bound [6].

Encrypted CBC. The ECBC function $\mathrm{ECBC}_{\pi_1,\pi_2}$ associated to permutations π_1, π_2 on n bits takes a message M that is a multiple of n bits and returns $\pi_2(\mathrm{CBC}_{\pi_1}(M))$. Define $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{ECBC}}(q,n,\ell)$ analogously to the CBC case above ($\mathsf{atk} \in \{\mathsf{any},\mathsf{eq},\mathsf{pf}\}$). Petrank and Rackoff [15] showed that $\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q,n,\ell) \le 2.5\,\ell^2 q^2/2^n$. A better bound, $\mathbf{Adv}^{\mathsf{eq}}_{\mathrm{ECBC}}(q,n,\ell) \le q^2/2^n \cdot (1 + c\ell^2/2^n + c\ell^6/2^{2n})$ for some constant c, is possible for the $\mathsf{atk} = \mathsf{eq}$ case based on a lemma of Dodis *et al.* [9], but the point of the ECBC construction is to achieve any-security. We improve on the result of Petrank and Rackoff to show that $\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q,n,\ell) \le q^2/2^n \cdot (d'(\ell) + 4\ell^4/2^n)$ where $d'(\ell)$ is the maximum, over all $\ell' \le \ell$, of the number of divisors of ℓ'. (Once again see Figure 1.) Note that the function $d'(\ell) \approx \ell^{1/\ln\ln(\ell)}$ grows slowly.

The MAC corresponding to ECBC (namely $\mathrm{ECBC}_{\pi_1,\pi_2}$ when $\pi_1 = E(K_1,\cdot)$ and $\pi_2 = E(K_2,\cdot)$ for random keys $K_1, K_2 \in \mathcal{K}$ of a blockcipher $E\colon \mathcal{K} \times \{0,1\}^n \to \{0,1\}^n$) was developed by the RACE project [5]. This MAC is interesting as a natural and practical variant of the CBC MAC that correctly handles messages of varying lengths. A variant of ECBC called CMAC was recently adopted as a NIST-recommended mode of operation [14]. As with the CBC MAC, our results imply improved guarantees on the forgery probability of the ECBC MAC under a chosen-message attack, but this time of type any rather than merely pf, and with the improvement being numerically more substantial.

MORE DEFINITIONS. The collision-probability $\mathbf{CP}^{\mathsf{atk}}_{n,\ell}$ of the CBC MAC is the maximum, over all pairs of messages (M_1, M_2) in an appropriate atk-dependent range, of the probability, over random π, that $\mathrm{CBC}_\pi(M_1) = \mathrm{CBC}_\pi(M_2)$. For $\mathsf{atk} = \mathsf{any}$ the range is any pair of distinct strings of length a positive multiple of n but at most ℓn; for $\mathsf{atk} = \mathsf{pf}$ it is any such pair where neither string is a prefix of the other; and for $\mathsf{atk} = \mathsf{eq}$ it is any pair of distinct strings of exactly ℓn bits. The *full collision probability* $\mathbf{FCP}^{\mathsf{atk}}_{n,\ell}$ is similar except that the probability is of the event $C_2^{m_2} \in \{C_1^1, \ldots, C_1^{m_1}, C_2^1, \ldots, C_2^{m_2-1}\}$ where, for each $b \in \{1,2\}$, we have $C_b^i = \pi(C_b^{i-1} \oplus M_b^i)$ for $m_b = |M_b|/n$ and $i \in [1..m_b]$ and $C_b^0 = 0^n$. Note that these definitions do not involve an adversary and in this sense are simpler than the advantage functions considered above.

REDUCTIONS TO FCP AND CP. By viewing ECBC as an instance of the Carter-Wegman paradigm [18], one can reduce bounding $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{ECBC}}(q,n,\ell)$ (for $\mathsf{atk} \in \{\mathsf{any},\mathsf{eq},\mathsf{pf}\}$) to bounding $\mathbf{CP}^{\mathsf{atk}}_{n,\ell}$ (see [7], stated here as Lemma 4). This simplifies the analysis because one is now faced with a combinatorial problem rather than consideration of a dynamic, adaptive adversary.

The first step in our analysis of the CBC MAC is to provide an analogous reduction (Lemma 1) that reduces bounding $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q,n,\ell)$ to bounding $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell}$. Unlike the case of ECBC, the reduction is not immediate and does not rely on the Carter-Wegman paradigm. Rather it is proved directly using the game-playing approach [4, 16].

BOUNDS ON FCP AND CP. Black and Rogaway [7] show that $\mathbf{CP}^{\mathsf{any}}_{n,\ell} \le 2(\ell^2+\ell)/2^n$. Dodis, Gennaro, Håstad, Krawczyk, and Rabin [9] show that $\mathbf{CP}^{\mathsf{eq}}_{n,\ell} \le 2^{-n} + c\ell^2/2^{2n} + c\ell^3/2^{3n}$ for some absolute constant c. (The above-mentioned bound on $\mathbf{Adv}^{\mathsf{eq}}_{\mathrm{ECBC}}(q,n,\ell)$ is obtained via this.) We build on their techniques to show (cf. Lemma 5) that $\mathbf{CP}^{\mathsf{any}}_{n,\ell} \le 2d'(\ell)/2^n + 64\ell^4/2^{2n}$. Our bound on $\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q,n,\ell)$ then follows. We also show that $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell} \le 8\ell/2^n + 64\ell^4/2^{2n}$. Our bound on $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q,n,\ell)$ then follows.

We remark that the security proof of RMAC [11] had stated and used a claim that implies $\mathbf{CP}^{\mathsf{any}}_{n,\ell} \le 12\ell/2^n$, but the published proof was wrong. Our Lemma 5 both fixes and improves that result.

FURTHER RELATED WORK. Other approaches to the analysis of the CBC MAC and the encrypted CBC MAC include those of Maurer [13] and Vaudenay [17], but they only obtain bounds of $\ell^2 q^2/2^n$.

2 Definitions

Notation. The empty string is denoted ε. If x is a string then $|x|$ denotes its length. We let $B_n = \{0,1\}^n$. If $x \in B_n^*$ then $|x|_n = |x|/n$ denotes the number of n-bit blocks in it. If $X \subseteq \{0,1\}^*$ then $X^{\leq m}$ denotes the set of all non-empty strings formed by concatenating m or fewer strings from X and X^+ denotes the set of all strings formed by concatenating one or more strings from X. If $M \in B_n^*$ then M^i denotes its i-th n-bit block and $M^{i\to j}$ denotes the string $M^i \,\|\, \cdots \,\|\, M^j$, for $1 \leq i \leq j \leq |M|_n$. If S is a set equipped with some probability distribution then $s \xleftarrow{\$} S$ denotes the operation of picking s from S according to this distribution. If no distribution is explicitly specified, it is understood to be uniform.

We denote by $\mathrm{Perm}(n)$ the set of all permutations over $\{0,1\}^n$, and by $\mathrm{Func}(n)$ the set of all functions mapping $\{0,1\}^*$ to $\{0,1\}^n$. (Both these sets are viewed as equipped with the uniform distribution.) A blockcipher E (with blocklength n and key-space $\mathcal{K}$) is identified with the set of permutations $\{E_K\colon K \in \mathcal{K}\}$ where $E_K\colon \{0,1\}^n \to \{0,1\}^n$ denotes the map specified by key $K \in \mathcal{K}$. The distribution is that induced by a random choice of K from $\mathcal{K}$, so $f \xleftarrow{\$} E$ is the same as $K \xleftarrow{\$} \mathcal{K},\ f \leftarrow E_K$.

Security. An adversary is a randomized algorithm that always halts. Let $\mathcal{A}^{\mathsf{atk}}_{q,n,\ell}$ denote the class of adversaries that make at most q oracle queries, where if $\mathsf{atk} = \mathsf{eq}$, then each query is in B_n^ℓ; if $\mathsf{atk} = \mathsf{pf}$, then each query is in $B_n^{\leq \ell}$ and no query is a prefix of another; and if $\mathsf{atk} = \mathsf{any}$ then each query is in $B_n^{\leq \ell}$. We remark that the adversaries considered here are computationally unbounded. In this paper we always consider deterministic, stateless oracles and thus we will assume that an adversary never repeats an oracle query. We also assume that an adversary never asks a query outside of the implicitly understood domain of interest.

Let $F\colon D \to \{0,1\}^n$ be a set of functions and let $A \in \mathcal{A}^{\mathsf{atk}}_{q,n,\ell}$ be an adversary, where $\mathsf{atk} \in \{\mathsf{eq}, \mathsf{pf}, \mathsf{any}\}$. By "$A^f \Rightarrow 1$" we denote the event that A outputs 1 with oracle f. The advantage of A (in distinguishing an instance of F from a random function outputting n bits) and the advantage of F are defined, respectively, as

$$\begin{aligned}\mathbf{Adv}_F(A) &= \Pr[f \xleftarrow{\$} F :\ A^f \Rightarrow 1] - \Pr[f \xleftarrow{\$} \mathrm{Func}(n) :\ A^f \Rightarrow 1] \quad \text{and} \\ \mathbf{Adv}^{\mathsf{atk}}_F(q,n,\ell) &= \max_{A \in \mathcal{A}^{\mathsf{atk}}_{q,n,\ell}} \{\, \mathbf{Adv}_F(A) \,\}\ .\end{aligned}$$

Note that since $\mathcal{A}^{\mathsf{eq}}_{q,n,\ell} \subseteq \mathcal{A}^{\mathsf{pf}}_{q,n,\ell} \subseteq \mathcal{A}^{\mathsf{any}}_{q,n,\ell}$, we have

$$\mathbf{Adv}^{\mathsf{eq}}_F(q,n,\ell) \ \leq\ \mathbf{Adv}^{\mathsf{pf}}_F(q,n,\ell) \ \leq\ \mathbf{Adv}^{\mathsf{any}}_F(q,n,\ell)\ . \tag{1}$$

Cbc and Ecbc. Fix $n \geq 1$. For $M \in B_n^m$ and $\pi\colon B_n \to B_n$ then define $\mathrm{CBC}^M_\pi[i]$ inductively for $i \in [0..m]$ via $\mathrm{CBC}^M_\pi[0] = 0^n$ and $\mathrm{CBC}^M_\pi[i] = \pi(\mathrm{CBC}^M_\pi \oplus M^i)$ for $i \in [1..m]$. We associate to π the CBC MAC function $\mathrm{CBC}_\pi\colon B_n^+ \to B_n$ defined by $\mathrm{CBC}_\pi(M) = \mathrm{CBC}^M_\pi[m]$ where $m = |M|_n$. We let $\mathrm{CBC} = \{\mathrm{CBC}_\pi\colon \pi \in \mathrm{Perm}(n)\}$. This set of functions has the distribution induced by picking π uniformly from $\mathrm{Perm}(n)$.

To functions $\pi_1, \pi_2\colon B_n \to B_n$ we associate the encrypted CBC MAC function $\mathrm{ECBC}_{\pi_1,\pi_2}\colon B_n^+ \to B_n$ defined by

$$\mathrm{ECBC}_{\pi_1,\pi_2}(M) \ = \ \pi_2(\mathrm{CBC}_{\pi_1}(M))$$

for all $M \in B_n^+$. We let $\mathrm{ECBC} = \{\mathrm{ECBC}_{\pi_1,\pi_2}\colon \pi_1, \pi_2 \in \mathrm{Perm}(n)\}$. This set of functions has the distribution induced by picking π_1, π_2 independently and uniformly at random from $\mathrm{Perm}(n)$.

Collisions. For $M_1, M_2 \in B_n^*$ we define the *prefix predicate* $\mathsf{pf}(M_1, M_2)$ to be true if either M_1 is a prefix of M_2 or M_2 is a prefix of M_1, and false otherwise. Note that $\mathsf{pf}(M, M) = \mathsf{true}$ for any

$M \in B_n^*$. Let

$$\begin{aligned}
\mathcal{M}^{\mathsf{eq}}_{n,\ell} &= \{(M_1, M_2) \in B_n^\ell \times B_n^\ell : \ M_1 \neq M_2\}, \\
\mathcal{M}^{\mathsf{pf}}_{n,\ell} &= \{(M_1, M_2) \in B_n^{\leq \ell} \times B_n^{\leq \ell} : \ \mathsf{pf}(M_1, M_2) = \mathsf{false}\}, \quad \text{and} \\
\mathcal{M}^{\mathsf{any}}_{n,\ell} &= \{(M_1, M_2) \in B_n^{\leq \ell} \times B_n^{\leq \ell} : \ M_1 \neq M_2\} \ .
\end{aligned}$$

For $M_1, M_2 \in B_n^+$ and $\mathsf{atk} \in \{\mathsf{eq}, \mathsf{pf}, \mathsf{any}\}$ we then let

$$\begin{aligned}
\mathbf{CP}_n(M_1, M_2) &= \Pr[\pi \xleftarrow{\$} \mathrm{Perm}(n) : \ \mathrm{CBC}_\pi(M_1) = \mathrm{CBC}_\pi(M_2)] \\
\mathbf{CP}^{\mathsf{atk}}_{n,\ell} &= \max_{(M_1,M_2)\in\mathcal{M}^{\mathsf{atk}}_{n,\ell}} \{ \ \mathbf{CP}_n(M_1, M_2) \ \} \ .
\end{aligned}$$

For $M_1, M_2 \in B_n^+$ we let $\mathbf{FCP}_n(M_1, M_2)$ (the full collision probability) be the probability, over $\pi \xleftarrow{\$} \mathrm{Perm}(n)$, that $\mathrm{CBC}_\pi(M_2)$ is in the set

$$\{\mathrm{CBC}_\pi^{M_1}[1], \ldots, \mathrm{CBC}_\pi^{M_1}[m_1], \mathrm{CBC}_\pi^{M_2}[1], \ldots, \mathrm{CBC}_\pi^{M_2}[m_2 - 1]\}$$

where $m_b = |M_b|_n$ for $b = 1, 2$. For $\mathsf{atk} \in \{\mathsf{eq}, \mathsf{pf}, \mathsf{any}\}$ we then let

$$\mathbf{FCP}^{\mathsf{atk}}_{n,\ell} = \max_{(M_1,M_2)\in\mathcal{M}^{\mathsf{atk}}_{n,\ell}} \{ \ \mathbf{FCP}_n(M_1, M_2) \ \} \ .$$

3 Results on the CBC MAC

We state results only for the $\mathsf{atk} = \mathsf{pf}$ case; results for $\mathsf{atk} = \mathsf{eq}$ follow due to (1). To bound $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell)$ we must consider a dynamic adversary that adaptively queries its oracle. Our first lemma reduces this problem to that of bounding a more "static" quantity whose definition does not involve an adversary, namely the full collision probability of the CBC MAC. The proof is in Section 5.

Lemma 1 *For any n, ℓ, q*

$$\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell) \ \leq \ q^2 \cdot \mathbf{FCP}^{\mathsf{pf}}_{n,\ell} + \frac{4\ell q^2}{2^n} \ . \quad \blacksquare$$

The next lemma bounds the full collision probability of the CBC MAC. The proof is given in Section 8.

Lemma 2 *For any n, ℓ*

$$\mathbf{FCP}^{\mathsf{pf}}_{n,\ell} \ \leq \ \frac{8\ell}{2^n} + \frac{64\ell^4}{2^{2n}} \ . \quad \blacksquare$$

Combining the above two lemmas we bound $\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell)$:

Theorem 3 *For any n, ℓ, q*

$$\mathbf{Adv}^{\mathsf{pf}}_{\mathrm{CBC}}(q, n, \ell) \ \leq \ \frac{\ell q^2}{2^n} \cdot \left(12 + \frac{64\ell^3}{2^n}\right) \ . \quad \blacksquare$$

4 Results on the Encrypted CBC MAC

Following [7], we view ECBC as an instance of the Carter-Wegman paradigm [18]. This enables us to reduce the problem of bounding $\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{ECBC}}(q,n,\ell)$ to bounding the collision probability of the CBC MAC, as stated in the next lemma. A proof of the following is provided in Appendix A.

Lemma 4 *For any n,ℓ,q and any $\mathsf{atk} \in \{\mathsf{eq},\mathsf{pf},\mathsf{any}\}$,*

$$\mathbf{Adv}^{\mathsf{atk}}_{\mathrm{ECBC}}(q,n,\ell) \;\leq\; \frac{q(q-1)}{2} \cdot \left(\mathbf{CP}^{\mathsf{atk}}_{n,\ell} + \frac{1}{2^n}\right) . \quad \blacksquare$$

Petrank and Rackoff [15] show that

$$\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q,n,\ell) \;\leq\; 2.5\,\ell^2 q^2/2^n . \tag{2}$$

Dodis *et al.* [9] show that $\mathbf{CP}^{\mathsf{eq}}_{n,\ell} \leq 2^{-n} + c\ell^2 \cdot 2^{-2n} + c\ell^6 \cdot 2^{-3n}$ for some absolute constant c. Combining this with Lemma 4 leads to

$$\mathbf{Adv}^{\mathsf{eq}}_{\mathrm{ECBC}}(q,n,\ell) \;\leq\; \frac{q^2}{2^n} \cdot \left(1 + \frac{c\ell^2}{2^n} + \frac{c\ell^6}{2^{2n}}\right) .$$

However, the case of $\mathsf{atk} = \mathsf{eq}$ is not interesting here, since the point of ECBC is to gain security even for $\mathsf{atk} = \mathsf{any}$. To obtain an improvement for this, we show the following, whose proof is in Section 7:

Lemma 5 *For any n,ℓ*

$$\mathbf{CP}^{\mathsf{any}}_{n,\ell} \;\leq\; \frac{2d'(\ell)}{2^n} + \frac{64\ell^4}{2^{2n}}$$

where $d'(\ell)$ is the maximum, over all $\ell' \leq \ell$, of the number of positive numbers that divide ℓ'. $\blacksquare$

The function $d'(\ell)$ grows slowly; in particular, $d'(\ell) < \ell^{0.7/\ln\ln(\ell)}$ for all sufficiently large ℓ [10, Theorem 317]. We have verified that $d'(\ell) \leq \ell^{1.07/\ln\ln\ell}$ for all $\ell \leq 2^{64}$ (and we assume for all ℓ), and also that $d'(\ell) \leq \lg^2 \ell$ for all $\ell \leq 2^{25}$.

Combining the above with Lemma 4 leads to the following:

Theorem 6 *For any n,ℓ,q*

$$\mathbf{Adv}^{\mathsf{any}}_{\mathrm{ECBC}}(q,n,\ell) \;\leq\; \frac{q^2}{2^n} \cdot \left(d'(\ell) + \frac{32\ell^4}{2^n}\right) . \quad \blacksquare$$

5 Bounding FCP Bounds CBC (Proof of Lemma 1)

The proof is by the game-playing technique [2, 4]. Let A be an adversary that asks exactly q queries, $M_1,\ldots,M_q \in B_n^{\leq \ell}$, where no queries M_r and M_s, for $r \neq s$, share a prefix in B_n^+. We must show that $\mathbf{Adv}_{\mathrm{CBC}}(A) \leq q^2 \cdot \mathbf{FCP}^{\mathsf{pf}}_{n,\ell} + 4\ell q^2/2^n$.

Refer to games D0–D7 as defined in Figure 2. Sets $\mathrm{Dom}(\pi)$ and $\mathrm{Ran}(\pi)$ start off as empty and automatically grow as points are added to the domain and range of the partial function π. Sets $\overline{\mathrm{Dom}}(\pi)$ and $\overline{\mathrm{Ran}}(\pi)$ are the complements of these sets relative to $\{0,1\}^n$. They automatically shrink as points join the domain and range of π. We write boolean values as 0 (false) and 1 (true),

On the s^{th} query $F(M_s)$ Game D1

$m_s \leftarrow |M_s|_n,\quad C_s^0 \leftarrow 0^n$
for $i \leftarrow 1$ **to** $m_s - 1$ **do**
  $X_s^i \leftarrow C_s^{i-1} \oplus M_s^i$
  if $X_s^i \in \mathrm{Dom}(\pi)$ **then** $C_s^i \leftarrow \pi(X_s^i)$
    else $\pi(X_s^i) \leftarrow C_s^i \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$
$X_s^{m_s} \leftarrow C_s^{m_s-1} \oplus M_s^{m_s}$
$\widehat{C}_s^{m_s} \leftarrow C_s^{m_s} \xleftarrow{\$} \{0,1\}^n$
if $C_s^{m_s} \in \mathrm{Ran}(\pi)$: $bad \leftarrow 1$, $C_s^{m_s} \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$
if $X_s^{m_s} \in \mathrm{Dom}(\pi)$: $bad \leftarrow 1$, $C_s^{m_s} \leftarrow \pi(X_s^{m_s})$
$\pi(X_s^{m_s}) \leftarrow C_s^{m_s}$
if bad **then return** $C_s^{m_s}$
return $\widehat{C}_s^{m_s}$

On the s^{th} query $F(M_s)$ Game D2

$m_s \leftarrow |M_s|_n,\quad C_s^0 \leftarrow 0^n$
for $i \leftarrow 1$ **to** $m_s - 1$ **do**
  $X_s^i \leftarrow C_s^{i-1} \oplus M_s^i$
  if $X_s^i \in \mathrm{Dom}(\pi)$ **then** $C_s^i \leftarrow \pi(X_s^i)$
    else $\pi(X_s^i) \leftarrow C_s^i \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$
$X_s^{m_s} \leftarrow C_s^{m_s-1} \oplus M_s^{m_s}$
$C_s^{m_s} \xleftarrow{\$} \{0,1\}^n$
if $X_s^{m_s} \in \mathrm{Dom}(\pi) \vee C_s^{m_s} \in \mathrm{Ran}(\pi)$
  then $bad \leftarrow 1$
$\pi(X_s^{m_s}) \leftarrow C_s^{m_s}$
return $C_s^{m_s}$

On the s^{th} query $F(M_s)$ Game D3

$m_s \leftarrow |M_s|_n,\quad C_s^0 \leftarrow 0^n$
for $i \leftarrow 1$ **to** $m_s - 1$ **do**
  $X_s^i \leftarrow C_s^{i-1} \oplus M_s^i$
  if $(\exists r < s)(X_s^i = X_r^{m_r})$: $bad \leftarrow 1$
  if $X_s^i \in \mathrm{Dom}(\pi)$ **then** $C_s^i \leftarrow \pi(X_s^i)$
    else $\pi(X_s^i) \leftarrow C_s^i \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$,
      if $(\exists r < s)(C_s^i = C_r^{m_r})$: $bad \leftarrow 1$
$X_s^{m_s} \leftarrow C_s^{m_s-1} \oplus M_s^{m_s}$
$C_s^{m_s} \xleftarrow{\$} \{0,1\}^n$
if $X_s^{m_s} \in \mathrm{Dom}(\pi) \vee C_s^{m_s} \in \mathrm{Ran}(\pi) \vee$
  $(\exists r < s)(X_s^{m_s} = X_r^{m_r} \vee C_s^{m_s} = C_r^{m_r})$
  then $bad \leftarrow 1$
return $C_s^{m_s}$

On the s^{th} query $F(M_s)$ Game D4

$m_s \leftarrow |M_s|_n,\quad C_s^0 \leftarrow 0^n$
for $i \leftarrow 1$ **to** $m_s - 1$ **do**
  $X_s^i \leftarrow C_s^{i-1} \oplus M_s^i$
  if $(\exists r < s)(X_s^i = X_r^{m_r})$: $bad \leftarrow 1$
  if $X_s^i \in \mathrm{Dom}(\pi)$ **then** $C_s^i \leftarrow \pi(X_s^i)$
    else $\pi(X_s^i) \leftarrow C_s^i \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$
$X_s^{m_s} \leftarrow C_s^{m_s-1} \oplus M_s^{m_s}$
if $X_s^{m_s} \in \mathrm{Dom}(\pi) \vee$
  $(\exists r < s)(X_s^{m_s} = X_r^{m_r})$ **then** $bad \leftarrow 1$
$C_s^{m_s} \xleftarrow{\$} \{0,1\}^n$
return $C_s^{m_s}$

Game D5

for $s \leftarrow 1$ **to** q **do**
  $C_s^0 \leftarrow 0^n$
  for $i \leftarrow 1$ **to** $\mathtt{m}_s - 1$ **do**
    $X_s^i \leftarrow C_s^{i-1} \oplus \mathtt{M}_s^i$
    if $(\exists r < s)(X_s^i = X_r^{\mathtt{m}_r})$: $bad \leftarrow 1$
    if $X_s^i \in \mathrm{Dom}(\pi)$ **then** $C_s^i \leftarrow \pi(X_s^i)$
      else $\pi(X_s^i) \leftarrow C_s^i \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi)$
  $X_s^{\mathtt{m}_s} \leftarrow C_s^{\mathtt{m}_s-1} \oplus \mathtt{M}_s^{\mathtt{m}_s}$
  if $(\exists r < s)$ $(X_s^{\mathtt{m}_s} \in \mathrm{Dom}(\pi) \vee$
    $X_s^{\mathtt{m}_s} = X_r^{\mathtt{m}_r})$ **then** $bad \leftarrow 1$

Game D6

$\pi \xleftarrow{\$} \mathrm{Perm}(n)$
for $s \in [1\,..\,q]$ **do**
  $C_s^0 \leftarrow 0^n$
  for $i \leftarrow 1$ **to** $\mathtt{m}_s - 1$ **do**
    $X_s^i \leftarrow C_s^{i-1} \oplus \mathtt{M}_s^i$
    $C_s^i \leftarrow \pi(X_s^i)$
  $X_s^{\mathtt{m}_s} \leftarrow C_s^{\mathtt{m}_s-1} \oplus \mathtt{M}_s^{\mathtt{m}_s}$
$bad \leftarrow (\exists (r,i) \neq (s, \mathtt{m}_s))\ [X_r^i = X_s^{\mathtt{m}_s}]$

Game D7

$\pi \xleftarrow{\$} \mathrm{Perm}(n)$
$C_1^0 \leftarrow C_2^0 \leftarrow 0^n$
for $i \leftarrow 1$ **to** $\mathtt{m}_1$ **do**
  $X_1^i \leftarrow C_1^{i-1} \oplus \mathtt{M}_1^i$, $C_1^i \leftarrow \pi(X_1^i)$
for $i \leftarrow 1$ **to** $\mathtt{m}_2$ **do**
  $X_2^i \leftarrow C_2^{i-1} \oplus \mathtt{M}_2^i$, $C_2^i \leftarrow \pi(X_2^i)$
$bad \leftarrow X_2^{\mathtt{m}_2} \in \{X_1^1, \ldots, X_1^{\mathtt{m}_1},$
  $X_2^1, \ldots, X_2^{\mathtt{m}_2-1}\}$

Figure 2: Games D0–D7 used in the proof of Lemma 1.

and we sometimes write **then** as a colon. The flag *bad* is initialized to 0 and the map π is initialized as everywhere undefined. We now briefly explain the sequence.

D1: Game D1 faithfully simulates the CBC MAC construction. Instead of choosing a random permutation π up front, we fill in its values as-needed, so as to not to create a conflict. Observe that if $bad = 0$ following lines 107–108 then $\widehat{C}_s^{m_s} = C_s^{m_s}$ and so game D1 always returns $C_s^{m_s}$, regardless of *bad*. This makes clear that $\Pr[A^{\mathrm{D1}} \Rightarrow 1] = \Pr[\pi \xleftarrow{\$} \mathrm{Perm}(n) : A^{\mathrm{CBC}_\pi} \Rightarrow 1]$. **D0:** Game D0 is obtained from game D1 by omitting line 110 and the statements that immediately follow the setting of *bad* at lines 107 and 108. Thus this game returns the random n-bit string $C_s^{m_s} = \widehat{C}_s^{m_s}$ in response to each query M_s, so $\Pr[A^{\mathrm{D0}} \Rightarrow 1] = \Pr[\rho \xleftarrow{\$} \mathrm{Func}(n) : A^\rho \Rightarrow 1]$. Now games D1 and D0 have been defined so as to be syntactically identical except on statements that immediately follow the setting of *bad* to true or the checking if *bad* is true, so the fundamental lemma of game-playing [4] says us that $\Pr[A^{\mathrm{D1}} \Rightarrow 1] - \Pr[A^{\mathrm{D0}} \Rightarrow 1] \leq \Pr[A^{\mathrm{D0}} \text{ sets } bad\,]$. As $\mathbf{Adv}_{\mathrm{CBC}}(A) = \Pr[A^{\mathrm{CBC}_\pi} \Rightarrow 1] - \Pr[A^\rho \Rightarrow 1] = \Pr[A^{\mathrm{D1}} \Rightarrow 1] - \Pr[A^{\mathrm{D0}} \Rightarrow 1]$, the rest of the proof bounds $\mathbf{Adv}_{\mathrm{CBC}}(A)$ by bounding $\Pr[A^{\mathrm{D0}} \text{ sets } bad\,]$.

D0→D2: We rewrite game D0 as game D2 by dropping the variable $\widehat{C}_s^{m_s}$ and using variable $C_s^{m_s}$ in its place, as these are always equal. We have that $\Pr[A^{\mathrm{D0}} \text{ sets } bad\,] = \Pr[A^{\mathrm{D2}} \text{ sets } bad\,]$. **D2→D3**: Next we eliminate line 209 and then, to compensate, we set *bad* any time the value $X_s^{m_s}$ or $C_s^{m_s}$ would have been accessed. This accounts for the new line 303 and the new disjunct on lines 310. To compensate for the removal of line 209 we must also set *bad* whenever C_s^i, chosen at line 204, happens to be a prior value $C_r^{m_r}$. This is done at line 306. We have that $\Pr[A^{\mathrm{D2}} \text{ sets } bad\,] \leq \Pr[A^{\mathrm{D3}} \text{ sets } bad\,]$. **D3→D4**: Next we remove the test $(\exists r < s)(C_s^i = C_r^{m_r})$ at line 306, the test if $C_s^{m_s} \in \mathrm{Ran}(\pi)$ at line 309, and the test for $C_s^{m_s} = C_r^{m_r}$ at line 310, bounding the probability that *bad* gets set due to any of these three tests. To bound the probability of *bad* getting set at line 306: A total of at most ℓq times we select at line 305 a random sample C_s^i from a set of size at least $2^n - \ell q \geq 2^{n-1}$. (We may assume that $\ell q \leq 2^{n-1}$ since the probability bound given by our lemma exceeds 1 if $\ell q > 2^{n-1}$.) The chance that one of these points is equal to any of the at most q points $C_r^{m_r}$ is thus at most $2\ell q^2/2^n$. To bound the probability of *bad* getting set by the $C_s^{m_s} \in \mathrm{Ran}(\pi)$ test at line 309: easily seen to be at most $\ell q^2/2^n$. To bound the probability of *bad* getting set by the $C_s^{m_s} = C_r^{m_r}$ test at line 310: easily seen to be at most $q^2/2^n$. Overall then, $\Pr[A^{\mathrm{D3}} \text{ sets } bad\,] \leq \Pr[A^{\mathrm{D4}} \text{ sets } bad\,] + 4\ell q^2/2^n$.

D4→D5: The value $C_s^{m_s}$ returned to the adversary in response to a query in game D4 is never referred to again in the code and has no influence on the game and the setting of *bad*. Accordingly, we may think of these values as being chosen up-front by the adversary who, correspondingly, makes an optimal choice of message queries $\mathtt{M}_1, \ldots, \mathtt{M}_q$ so as to maximize the probability that *bad* gets set in game D4. Queries $\mathtt{M}_1, \ldots, \mathtt{M}_q \in B_n^{\leq m}$ are prefix-free (meaning that no two strings from this list share a prefix $P \in B_n^+$) and the strings have block lengths of $\mathtt{m}_1, \ldots, \mathtt{m}_q$, respectively, where each $\mathtt{m}_i \leq m$. We fix such an optimal vector of messages and message lengths in passing to game D5, so that $\Pr[A^{\mathrm{D4}} \text{ sets } bad\,] \leq \Pr[\mathrm{D5} \text{ sets } bad\,]$. The adversary has effectively been eliminated at this point.

D5→D6: Next we postpone the evaluation of *bad* and undo the "lazy defining" of π to arrive at game D6. We have $\Pr[\mathrm{D5} \text{ sets } bad\,] \leq \Pr[\mathrm{D6} \text{ sets } bad\,]$. **D6→D7**: Next we observe that in game D6, some pair r, s must contribute at least an average amount to the probability that *bad* gets set. Namely, for any $r, s \in [1 \mathinner{..} q]$ where $r \neq s$ define $bad_{r,s}$ as

$$(X_s^{\mathtt{m}_s} = X_r^i \text{ for some } i \in [1 \mathinner{..} \mathtt{m}_r]) \vee (X_s^{\mathtt{m}_s} = X_s^i \text{ for some } i \in [1 \mathinner{..} \mathtt{m}_s - 1])$$

and note that *bad* is set at line 607 iff $bad_{r,s} = 1$ for some $r \neq s$, and so there must be an $r \neq s$ such that $\Pr[\mathrm{D6} \text{ sets } bad_{r,s}] \geq (1/q(q-1))\ \Pr[\mathrm{D6} \text{ sets } bad\,]$. Fixing such an r, s and renaming $\mathtt{M}_1 = \mathtt{M}_r$,

$\mathtt{M}_2 = \mathtt{M}_s$, $\mathtt{m}_1 = \mathtt{m}_r$, and $\mathtt{m}_2 = \mathtt{m}_s$, we arrive at game D7 knowing that

$$\Pr[\text{D6 sets } bad\,] \;\leq\; q^2 \cdot \Pr[\text{D7 sets } bad\,]\,. \tag{3}$$

Now $\Pr[\text{D7 sets } bad\,] = \mathbf{FCP}_n(\mathtt{M}_1, \mathtt{M}_2) \leq \mathbf{FCP}^{\mathsf{pf}}_{n,m}$ by the definition of FCP and the fact that π is a permutation. Putting all the above together we are done.

6 A Graph-Based Representation of CBC

In this section we describe a graph-based view of CBC computations and provide some lemmas that will then allow us to reduce the problem of upper bounding the collision probabilities $\mathbf{CP}^{\mathsf{any}}_{n,\ell}$ and $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell}$ to combinatorial counting problems. We fix for the rest of this section a blocklength $n \geq 1$, the number of messages $t \geq 1$ and t distinct messages $M_1, \ldots, M_t$, where for $1 \leq i \leq t$ we denote with $m_i \geq 1$ the length (in blocks) of the i'th message $M_i = M_i^1 \cdots M_i^{m_i} \in B_n^{m_i}$.[1] Let $\mathcal{M} = \{M_1, \ldots, M_t\}$ be the ordered set of all messages, for $1 \leq j \leq t$ let $m^j = \sum_{i=1}^{j} m_i$ be the length of the first j messages. It is convenient to set $m^0 = 0$ and $m = m^t$ to be the total length. Let $M = M_1 \| M_2 \| \cdots \| M_t = M^1 \cdots M^m$ denote the concatenation of all messages.

Structure graphs. To $\mathcal{M}$ and any $\pi \in \mathrm{Perm}(n)$ we associate the *structure graph* $G^{\mathcal{M}}_\pi$, which is a directed graph (V, E) where $V \subseteq \{0, \ldots, m\}$ together with a edge labelling function $L : E \to \{M^1, \ldots, M^m\}$.

> The structure graph $G^{\mathcal{M}}_\pi = G = (V, E, L)$ is defined as follows: We set $C_0 = 0^n$ and for $i = 1, \ldots, m$ we define
>
> $$C_i = \begin{cases} \pi(C_{i-1} \oplus M_i) & \text{if } i \notin \{m_0 + 1, \ldots, m_{t-1} + 1\} \\ \pi(M_i) & \text{otherwise} \end{cases}$$
>
> From this C_i's we define the mapping $[.]_G : \{0, \ldots, m\} \to \{0, \ldots, m\}$ as $[i]_G = \min\{j : C_j = C_i\}$. It is convenient to define a mapping $[.]'_G$ as $[i]'_G = [i]_G$ if $i \notin \{m^0, \ldots, m^{t-1}\}$ and $[i]'_G = 0$ otherwise. Now the structure graph $G^{\mathcal{M}}_\pi = G = (V, E, L)$ is given by
>
> $$V = \{[i]_G : 1 \leq i \leq m\} \qquad E = \{([i-1]'_G, [i]_G) : 1 \leq i \leq m\} \qquad L(([i-1]'_G, [i]_G)) = M_i$$

From this definition it is clear that the mapping $[.]_G$ defines G uniquely and vice versa. Throughout we will refer to $([i-1]'_G, [i]_G)$ as the i'th edge of G.

If the C_i's are all distinct, then G is simply a tree with t paths leaving the root 0, the i'th path being $0 \to m^{i-1} \to m^{i-1} + 1 \to \ldots \to m^{i-1} + m_i = m^i$. In general G is the graph one gets by starting with the tree just described and doing the following while possible: if there are two vertices i, j where $i \neq j$ and $C_i = C_j$ then collapse i and j into one vertex and label it $\min\{i, j\}$.

Let $\mathcal{G}(\mathcal{M}) = \{G_\pi : \pi \in \mathrm{Perm}(n)\}$ denote the set of all structure graphs associated to messages $\mathcal{M}$. This set has the probability distribution induced by picking π at random from $\mathrm{Perm}(n)$. For $G \in \mathcal{G}(\mathcal{M})$, $G = (V, E, L)$ we denote with $G_i = (V_i, E_i, L_i)$ the subgraph of G given by the i first edges, i.e. $V_i = \{v \in V : v \leq i\}, E_i = \{(u, v) \in E : u, v \in V_i\}$ and L_i is L with the domain restricted to E_i.

Collisions. Suppose a structure graph $G = G^{\mathcal{M}}_\pi \in \mathcal{G}(\mathcal{M})$ is exposed edge by edge (i.e. in step i the value $[i]_G$ is shown to us). We say that G has a *collision* is step i if the edge exposed in step i

[1] To bound $\mathbf{CP}^{\mathsf{any}}_{n,\ell}$ and $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell}$ it is sufficient to only consider the case $t = 2$, but as this restriction does not simplify things we prove all our lemmas for general t.

points to a vertex which is already in the graph. With $\mathsf{Col}(G)$ we denote all collisions, i.e. all pairs (i,j) where in step i there was a collision which hit the vertex computed in step $j < i$:

$$\mathsf{Col}(G) = \{(i, [i]_G) : [i]_G \neq i\}$$

or equivalently for $G_{i-1} = (V_{i-1}, E_{i-1}, L_{i-1})$

$$\mathsf{Col}(G) = \{(i, [i]_G) : [i]_G \in V_{i-1}\} \tag{4}$$

We distinguish two types of collisions, *induced collisions* and *accidents*. Informally, an induced collision in step i is a collision which is implied by the collisions in the first $i-1$ steps, whereas an accident is a "surprising" collision.

Induced Collisions. Assume that after step $i-1$ we see that for some $a < i$ the a'th edge $([a-1]'_G, [a]_G)$ has the same label ($M_a = M_i$) and the same tail ($[a-1]'_G = [i-1]'_G$) as the next (i'th) edge to be exposed. Then we know that the head of the i'th edge must also be $[a]_G$ as $[a-1]_G = [i-1]_G$ means $C_{[a-1]_G} = C_{[i-1]_G}$, and as π must produce the same output on the same input also $C_{[a]_G} = \pi(C_{[a-1]'_G} \oplus M_a) = \pi(C_{[i-1]'_G} \oplus M_i) = C_{[i]_G}$. More generally one can show that G has an induced collision is step i if the edge added in step i (or, that would be added if it was not already there) closes a cycle with alternating edge directions, moreover then the labels of all the edges of that cycle XOR to 0^n (note that two parallel edges as considered before are exactly such a cycle of length two and of course the labels M_i, M_a of the edges on that cycle XOR to $0^n = M_i \oplus M_a$ as we saw that $M_i = M_a$). It's not easy to see that indeed all induced collisions are of his type, this will follow from the proof of Lemma 8.

To make that more formal, we define a function AltCyc, which takes as input a partial structure graph $G_i = (V_i, E_i, L_i)$, a vertex v and a label X as follows

$$\mathsf{AltCyc}(G_i = (V_i, E_i, L_i), v, X) = \begin{cases} j = v_{2k} & \text{if } \exists k \geq 1, \{v_1, \ldots, v_{2k}\} \in V_i, \{e_1, \ldots, e_{2k}\} \in E_i \text{ where} \\ & e_i = (v_i, v_{i+1}) \text{ for odd, and } e_i = (v_i, v_{i+1}) \text{ for even } i, \\ & \text{and } v_1 = v, \\ & \text{and } X \oplus L_i((u_1, v_1)) \oplus \ldots L_i((u_{2k}, v_{2k})) = 0^n. \\ \perp & \text{otherwise} \end{cases}$$

Now the induced collisions are are the collisions (i,j) where the i'th edge $([i-1]'_G, j)$ can (and thus must) be added to G_{i-1} such that we close a cycle with alternating edge directions where the labels on the cycle XOR to 0^n, i.e.

$$\mathsf{IndCol}(G) = \{(i,j) : 1 \leq i \leq m, j = \mathsf{AltCyc}(G_{i-1}, [i-1]'_G, M_i) \text{ and } j \neq \perp\}$$

Accidents. The accidents are all the non-induced collisions:

$$\mathsf{Acc}(G) = \mathsf{Col}(G) \setminus \mathsf{IndCol}(G) \tag{5}$$

Lemma 7 *$G \in \mathcal{G}(\mathcal{M})$ is uniquely determined by $\mathsf{Acc}(G)$ and $\mathcal{M}$ alone.*

Proof: We leave the reader to verify that the algorithm given in figure 6 outputs $G = (V, E, L)$ on input $\mathsf{Acc}(G)$ and $\mathcal{M}$. The labelling function L is given by a set of pairs where $(e,j) \in L$ means $L(e) = j$. ∎

algorithm $\mathsf{Acc2Graph}(A, \mathcal{M})$ //$A = \{(i_1, j_1), \ldots, (i_t, j_t)\}$
$V \leftarrow \{0\}, E \leftarrow \emptyset, L \leftarrow \emptyset, tail \leftarrow 0$
for $i \leftarrow 1$ **to** m **do**
  if $i \in \{m^1, \ldots, m^{t-1}\}$ **then** $tail \leftarrow 0$ //postcondition: $tail = [i-1]'_G$, i.e. is the tail of the i'th edge.
  if $\exists j$ s.t. $(i, j) \in A$ **then**
    $e \leftarrow (tail, j), tail \leftarrow j$ //Accident
  else if $\mathsf{AltCyc}((V, E, L), M_i) = j \neq \perp$ **then**
    $e \leftarrow (tail, j), tail \leftarrow j$ //Induced Collision
  else $V \leftarrow V \cup i, e \leftarrow (tail, i), tail \leftarrow i$ // No collision, add vertex i
  $E \leftarrow E \cup e, L \leftarrow L \cup (e, M_i)$ //Add edge e and define label for e.
return $G \leftarrow (V, E, L)$

Figure 3:

Let $\mathcal{G}^a(\mathcal{M}) = \{G : G \in \mathcal{G}(\mathcal{M}), |\mathsf{Acc}(G)| = a)$ denote all structure graphs with exactly a accidents. By the previous lemma every $G \in \mathcal{G}^a(\mathcal{M})$ is determined by its accidents, i.e. a tuples (i, j) where $0 \leq j < i \leq m$, thus

$$|\mathcal{G}^a(\mathcal{M})| \leq \left(\frac{(m+1)m}{2}\right)^a \tag{6}$$

The following lemma states that the probability that a randomly sampled structure graph will be some particular graph H is exponentially small in $\mathsf{Acc}(H)$.

Lemma 8 *Let $n \geq 1, t \geq 1, \mathcal{M} = \{M_1, \ldots, M_t\}$ where $M_i \in B_n^{m_i}$ and $m = m_1 + \ldots + m_t$. Then for any structure graph $H \in \mathcal{G}(\mathcal{M})$:*

$$\Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ G = H] \leq (2^n - m)^{-|\mathsf{Acc}(H)|}$$

The lemma builds on an unpublished technique from [8, 9]. A proof is given in Appendix B.

Lemma 9 *With $\mathcal{M}, m$ as in the previous lemma*

$$\Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ |\mathsf{Acc}(G)| \geq 2] \leq \frac{4m^4}{2^{2n}}$$

Proof:

$$\begin{aligned}
\Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ |\mathsf{Acc}(G)| \geq 2] &= \sum_{i=2}^{\infty} \Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ |\mathsf{Acc}(G)| = i] \\
&\leq \sum_{i=2}^{\infty} \sum_{H \in \mathcal{G}^i(\mathcal{M})} \Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ G = H] \\
&\leq \sum_{i=2}^{\infty} \frac{|\mathcal{G}^i(\mathcal{M})|}{(2^n - m)^i} && (7) \\
&\leq \sum_{i=2}^{\infty} \left(\frac{(m+1)m}{2(2^n - m)}\right)^i && (8) \\
&\leq \frac{4m^4}{2^{2n}} && (9)
\end{aligned}$$

We used Lemma 8 for (7) and eq.(6) for (8). For (9) we assumed that $n > 1$ and $m < 2^{n/2}$. We can do so as if either $n = 1$ or $m \geq 2^{n/2}$ the lemma trivially holds as then the $\frac{4m^4}{2^{2n}} \geq 1$ and 1 is a trivial bound for any probability. ▮

Conventions for Section 7 and 8. In the next two sections we will only consider the two accidents case $\mathcal{M} = \{M_1, M_2\}$. A $G \in \mathcal{G}(M_1, M_2)$ consists of two paths, the "M_1-path" which passes through the vertices $0, [1]_G, \ldots, [m_1]_G$ and the "M_2-path" $0, [m_1 + 1]_G, \ldots, [m_1 + m_2]_G$. With $V_j^i(G)$ we denote the i'th vertex on the M_j-path, i.e. for $1 \leq i \leq m_1 : V_1^i(G) \stackrel{\text{def}}{=} [i]_G$, for $1 \leq i \leq m_2 : V_2^i(G) \stackrel{\text{def}}{=} [i + m_1]_G$ and $V_1^0(G) = V_2^0(G) = 0$. $\ell \geq m_1, m_2$ will always denote an upper bound on the message length.

Further if P is a predicate on structure graphs. Then $\phi_{M_1,M_2}[P]$ will denote the set of structure graphs G having exactly one accident and satisfying the predicate P:

$$\phi_{M_1,M_2}[P] = \{G \in \mathcal{G}^1(M_1, M_2) : \ G \text{ satisfies } P\} \ .$$

For example, predicate P might be $V_1^{m_1}(\cdot) = V_2^{m_2}(\cdot)$ and in that case $\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]$ is $\{G \in \mathcal{G}^1(M_1, M_2) : \ V_1^{m_1}(G) = V_2^{m_2}(G)\}$.

Recall that in our graph view an accident corresponds to a collision which does not close a cycle with alternating edge directions. This is not a very convenient definition to work with. But as in the following two sections we will only consider the two message case and structure graphs with at most one accident we can take on a simpler view, for this we define $\mathsf{trueCol}(G)$ to denote all collisions in G except those which are due to parallel edges. Or equivalently, the true collisions are the collisions which increase the indegree of some vertex. For $G_{i-1} = (V_{i-1}, E_{i-1}, L_{i-1})$ (compare this to eq. (4)) the true collisions are

$$\mathsf{trueCol}(G) = \{(i, [i]_G) : [i]_G \in V_{i-1} \wedge ([i-1]'_G, [i]_G) \notin E_{i-1}\}$$

Clearly $\mathsf{trueCol}(G) \subseteq \mathsf{Col}(G)$ and $\mathsf{acc}(G) \subseteq \mathsf{trueCol}(G)$, but what makes this concept useful for us is that in the two message case the first two true collisions are always accidents

Lemma 10 *For $G \in \mathcal{G}(M_1, M_2)$*

$$\textbf{if} \ \ |\mathsf{trueCol}(G_i)| \leq 2 \ \ \textbf{then} \ \ \mathsf{Acc}(G_i) = \mathsf{trueCol}(G_i)$$

▮

As corollaries we get that the $G \in \mathcal{G}(M_1, M_2)$ with exactly one accident also have exactly one true collision

$$|\mathsf{Acc}(G)| = 1 \iff |\mathsf{trueCol}(G)| = 1$$

and further that whenever we have two or more true collisions we also have at least tow accidents

$$|\mathsf{trueCol}(G)| \geq 2 \Rightarrow |\mathsf{Acc}(G)| \geq 2.$$

So to exclude structure graphs which have two or more accidents it is sufficient to exclude those graphs which have more that two true collisions. To see that the lemma holds it is enough to show that there's no way to draw two paths, both starting at the same point such that the resulting graph has only one true collision and a cycle with alternating edge directions of length at least four, we leave the verification of that to the reader. Lemma 10 is tight in the sense that we can't replace the 2 with a 3 there. In figure 4 a structure graph $G \in \mathcal{G}(A\|B\|C, D\|E\|F)$ is shown with three true collisions $(4, 3), (5, 2), (6, 1)$ (the edges D, E, F account for those collisions) but only two

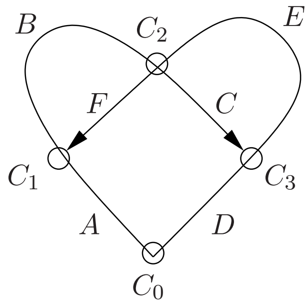

Figure 4: A $G \in \mathcal{G}(A\|B\|C, D\|E\|F)$ with three true collisions but only two accidents.

accidents as the collision $(6,1)$ is induced: the F-edge closes a cycle $C_0 \xrightarrow{A} C_1 \xleftarrow{F} C_2 \xrightarrow{C} C_3 \xleftarrow{D} C_0$ with alternating edge directions. Like on every cycle with alternating edge directions the labels of the edges must XOR to 0^n, so here $A \oplus F \oplus C \oplus D = 0^n$ must hold. This can be derived from the following equalities implied by the graph: $C_0 \oplus A = C_2 \oplus F$, $C_0 \oplus D = C_2 \oplus C$.

7 Bounding $\mathbf{CP}^{\mathsf{any}}_{n,m}$ (Proof of Lemma 5)

In this section we prove Lemma 5, showing that $\mathbf{CP}^{\mathsf{any}}_{n,\ell} \leq 2d'(\ell)/2^n + 64\ell^4/2^{2n}$ for any n, ℓ, thereby proving Lemma 5.

Lemma 11 *Let $n \geq 1$ and $1 \leq m_1, m_2 \leq \ell$. Let $M_1 \in B_n^{m_1}$ and $M_2 \in B_n^{m_2}$ be distinct messages. Then*

$$\mathbf{CP}^{\mathsf{any}}_{n,\ell}(M_1, M_2) \leq \frac{2 \cdot |\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]|}{2^n} + \frac{64\ell^4}{2^{2n}} . \blacksquare$$

Proof: With the probability over $G \xleftarrow{\$} \mathcal{G}(M_1, M_2)$, we have:

$$\begin{aligned}
&\mathbf{CP}_n(M_1, M_2) \\
&\quad = \Pr[V_1^{m_1} = V_2^{m_2}] \\
&\quad = \Pr[V_1^{m_1} = V_2^{m_2} \wedge \mathsf{Acc}(G) = 1] + \Pr[V_1^{m_1} = V_2^{m_2} \wedge \mathsf{Acc}(G) \geq 2] &(10)\\
&\quad \leq \frac{|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]|}{2^n - m_1 - m_2} + \frac{4(m_1+m_2)^4}{2^{2n}} &(11)\\
&\quad \leq \frac{|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]|}{2^n - 2\ell} + \frac{64\ell^4}{2^{2n}} \\
&\quad \leq \frac{2 \cdot |\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]|}{2^n} + \frac{64\ell^4}{2^{2n}} . &(12)
\end{aligned}$$

In (10) above we used that $\Pr[V_1^{m_1} = V_2^{m_2} \wedge \mathsf{Acc}(G) = 0] = 0$ as $V_1^{m_1} = V_2^{m_2}$ with $M_1 \neq M_2$ implies that there is at least one accident. In (11) we first used Lemma 8, and then used Lemma 9. In (12) we assumed that $\ell \leq 2^{n/2-1.5}$, which we can do as otherwise $64\ell^4/2^{2n} \geq 1$. $\blacksquare$

Next we bound the size of the set that arises above:

Lemma 12 *Let $n, \ell \geq 1$ and $1 \leq m_2 \leq m_1 \leq \ell$. Let $M_1 \in B_n^{m_1}$ and $M_2 \in B_n^{m_2}$ be distinct messages. Then*

$$|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]| \leq d'(\ell) . \quad \blacksquare$$

Putting together Lemmas 11 and 12 completes the proof of Lemma 5.

Proof of Lemma 12: Let $k \geq 0$ be the largest integer such that M_1, M_2 have a common suffix of k blocks. Note that $V_1^{m_1} = V_2^{m_2}$ iff $V_1^{m_1-k} = V_2^{m_2-k}$. Thus, we may consider M_1 to be replaced by $M_1^{1\to m_1-k}$ and M_2 to be replaced by $M_2^{1\to m_2-k}$, with m_1, m_2 correspondingly replaced by m_1-k, m_2-k respectively. We now have distinct messages M_1, M_2 of at most ℓ blocks each such that either $m_2 = 0$ or $M_1^{m_1} \neq M_2^{m_2}$. (Note that now m_2 could be 0, which was not true before our transformation.) Now consider three cases. The first is that $m_2 \geq 1$ and M_2 is a prefix of M_1. This case is covered by Lemma 13. (Note in this case it must be that $m_1 > m_2$ since M_1, M_2 are distinct and their last blocks are different.) The second case is that $m_2 = 0$ and is covered by Lemma 14. (In this case, $m_1 \geq 1$ since M_1, M_2 are distinct.) The third case is that $m_2 \geq 1$ and M_2 is not a prefix of M_1. This case is covered by Lemma 15. ▮

Lemma 13 *Let $n \geq 1$ and $1 \leq m_2 < m_1 \leq \ell$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$. Assume M_2 is a prefix of M_1 and $M_1^{m_1} \neq M_2^{m_2}$. Then $|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]| \leq d'(\ell)$.* ▮

Proof: Because M_2 is a prefix of M_1 we have that $V_2^{m_2} = V_1^{m_2}$, and thus $|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]| = |\phi_{M_1,M_2}[V_1^{m_2} = V_1^{m_1}]|$. We now bound the latter.

Let $G \in \mathcal{G}^1(M_1, M_2)$. Then we claim that $V_1^{m_1}(G) = V_1^{m_2}(G)$ implies that there exists a $t > m_2$ such that

1. The (only) accident in G is a $(t, V_1^{m_2})$-accident.
2. $t - m_2$ divides $m_1 - m_2$.

Before we prove the two points we'll show why they imply the lemma. First note that here $V_1^{m_2} \stackrel{\text{def}}{=} [m_2]_G = m_2$ as the first accident happens after step m_2 and thus $[m_2]_G = m_2$. Further, by Lemma 7 every $G \in \phi_{M_1,M_2}[V_1^{m_2} = V_1^{m_1}]$ corresponds to a pair $(t, V_1^{m_2}) = (t, m_2)$ where t satisfies the second point. Now there are exactly $d(m_1 - m_2) \leq d'(\ell)$ different t's which satisfy the second point.

To see the first point we observe that $\mathsf{in}_G(V_1^{m_2}(G)) = 2$ as by $V_1^{m_1}(G) = V_1^{m_2}(G)$ the vertex $V_1^{m_2}(G)$ has ingoing edges with distinct labels $M_1^{m_1} \neq M_1^{m_2}$ and edges with distinct labels cannot be parallel.[2] So there was an accident $(t, V_1^{m_2}(G))$ for some $t \geq m_2$.

We now prove the second point. We just saw that G_t (the subgraph of G build by the t first edges) is a ρ-shaped graph where the cycle has length $t - m_2$. Now either $t = m_1$ (then the second point is satisfied), or the remaining $m_1 - t > 0$ edges of the M_1 path must be drawn such that we come back to the vertex m_2. As we already used up or only accident, the only possibility is to go along the cycle (note that if we leave the ρ then there's no way to come back without a second accident[3]). But then we'll only end up in m_2 (and thus satisfy $V_1^{m_1}(G) = V_1^{m_2}(G)$) if the remaining $m_1 - m_2$ edges are a multiple of the cycle length $t - m_2$.

Lemma 14 *Let $n \geq 1$ and $1 \leq m_1 \leq \ell$. Let $M_1 \in B_n^{m_1}$, let $M_2 = \varepsilon$ and let $m_2 = 0$. Then $|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]| \leq d'(\ell)$.* ▮

[2]As this would mean that the permutation which generated the structure-graph produces the same output on different inputs. This is one of the arguments where it is crucial that we only consider permutations and not general functions.

[3]Recall that here accidents are equivalent to true collisions.

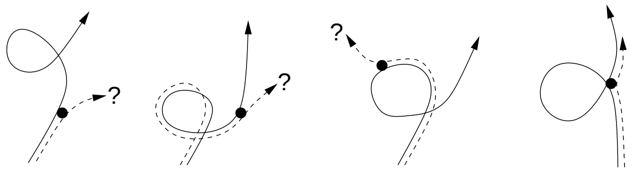

Figure 5: Some shapes where the M_1-path (solid line) makes a loop. In the first three cases the M_1-path passes only once through V_1^p (the dot), and we see that we cannot draw the M_2-path such that $V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\}$ without a second accident in any of those cases. In the last graph $V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\}$, but there also $V_1^p \in \{V_1^0, \ldots, V_1^{p-1}, V_1^{p+1}, \ldots, V_1^{m_1}\}$.

Proof: Use an argument similar to that of Lemma 13, noting that $V_{m_1}^0(G) = V_1^0(G)$ implies that $\mathsf{in}_G(V_1^0(G)) \geq 1$. ▮

Lemma 15 *Let $n \geq 1$ and $1 \leq m_2 \leq m_1 \leq \ell$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$. Assume M_2 is not a prefix of M_1 and $M_1^{m_1} \neq M_2^{m_2}$. Then $|\phi_{M_1,M_2}[V_1^{m_1} = V_2^{m_2}]| \leq 1$.* ▮

Proof: Let $p \in [0..m_2 - 1]$ be the largest integer such that $M_1^{1 \to i} = M_2^{1 \to i}$ for all $i \in [1..p]$. Then $V_1^i = V_2^i$ for $i \in [1..p]$ and $V_1^{p+1} \neq V_2^{p+1}$. Now to have $V_1^{m_1} = V_2^{m_2}$ we need an accident. Since $M_1^{m_1} \neq M_2^{m_2}$ and there is only one accident, the only possibility is that this is a $(m_1, m_1 + m_2)$-accident. Thus, there is only one way to draw the graph. ▮

8 Bounding $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell}$ (Proof of Lemma 2)

In this section we show that $\mathbf{FCP}^{\mathsf{pf}}_{n,\ell} \leq 8\ell/2^n + 64\ell^4/2^{2n}$ thereby proving Lemma 2. Recall that $\mathsf{pf}(M_1, M_2) = \mathsf{false}$ iff M_1 is not a prefix of M_2 and M_2 is not a prefix of M_1. The proof of the following is similar to the proof of Lemma 11 and is omitted.

Lemma 16 *Let $n \geq 1$ and $1 \leq m_1, m_2 \leq \ell$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$ with $\mathsf{pf}(M_1, M_2) = \mathsf{false}$. Then*

$$\mathbf{FCP}^{\mathsf{pf}}_{n,\ell}(M_1, M_2) \leq \frac{2 \cdot \left|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^1, \ldots, V_1^{m_1}, V_2^1, \ldots, V_2^{m_2-1}\}]\right|}{2^n} + \frac{64\ell^4}{2^{2n}} . ▮$$

Next we bound the size of the set that arises above:

Lemma 17 *Let $n, \ell \geq 1$ and $1 \leq m_1, m_2 \leq \ell$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$ with $\mathsf{pf}(M_1, M_2) = \mathsf{false}$. Then*

$$\left|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^1, \ldots, V_1^{m_1}, V_2^1, \ldots, V_2^{m_2-1}\}]\right| \leq 4\ell . ▮$$

Putting together Lemmas 16 and 17 completes the proof of Lemma 2.

We denote by $\mathsf{cpl}(M_1, M_2)$ the number of blocks in the longest common block-prefix of M_1, M_2. That is, $\mathsf{cpl}(M_1, M_2)$ is the largest integer p such that $M_1^i = M_2^i$ for all $i \in [1..p]$. Define the predicate $\mathsf{NoLoop}(G)$ to be true for structure graph $G \in \mathcal{G}_2^1(M_1, M_2)$ iff $V_1^0(G), \ldots, V_1^{m_1}(G)$ are all distinct and also $V_2^0(G), \ldots, V_2^{m_2}(G)$ are all distinct. Let Loop be the negation of NoLoop.

Proof of Lemma 17: Let $p = \mathsf{cpl}(M_1, M_2)$. Since $\mathsf{pf}(M_1, M_2) = \mathsf{false}$, it must be that $p < m_1, m_2$ and $M_1^{p+1} \neq M_2^{p+1}$. Note then that $V_1^i = V_2^i$ for all $i \in [0..p]$ but $V_1^{p+1} \neq V_2^{p+1}$. Now we break up

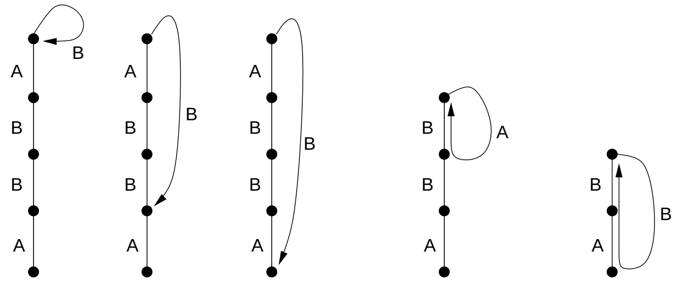

Figure 6: An example for the proof of Lemma 18 with $m_1 = 5$ and $M_1 = A\|B\|B\|A\|B$ for distinct $A, B \in \{0,1\}^n$. Here we have $N_5 = 5 - \mu_1(M_1^5) + 1 = 5 - \mu_1(B) + 1 = 5 - 3 + 1 = 3$ and $N_4 = \mu_1(M_1^5) - \mu_1(M_1^{4\to 5}) = \mu_1(B) - \mu_1(A\|B) = 3 - 2 = 1$ and $N_3 = \mu_1(M_1^{4\to 5}) - \mu_1(M_1^{3\to 5}) = \mu_1(A\|B) - \mu_1(B\|A\|B) = 2 - 1 = 1$ and $N_2 = N_1 = 0$. The first three graphs show the N_5 cases, the fourth and the fifth graph show the single cases for N_4 and N_3.

the set in which we are interested as

$$\begin{aligned}
&\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^1, \ldots, V_1^{m_1}, V_2^1, \ldots, V_2^{m_2-1}\}] \\
&\quad = \phi_{M_1,M_2}[V_2^{m_2} \in \{V_2^1, \ldots, V_2^{m_2-1}\}] \cup \phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\}] \ .
\end{aligned}$$

Lemma 18 implies that $|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_2^1, \ldots, V_2^{m_2-1}\}]| \leq m_2$ and Lemma 20 says that $|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\} \wedge \mathsf{NoLoop}]| \leq m_1$. It remains to bound $|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\} \wedge \mathsf{Loop}]|$. We use a case analysis, which is illustrated in Figure 5. The condition Loop means that either the M_1- or the M_2-path (or both) must make a loop. If the M_1-path makes a loop then we can only draw the M_2-path such that $V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\}$ if the loop goes twice through V_1^p. The same argument works if only the M_2-path makes a loop. Thus

$$\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\} \wedge \mathsf{Loop}] \subseteq \mathcal{S}_1 \cup \mathcal{S}_2$$

where

$$\begin{aligned}
\mathcal{S}_1 &= \phi_{M_1,M_2}[V_1^p \in \{V_1^0, \ldots, V_1^{p-1}, V_1^{p+1}, \ldots, V_1^{m_1}\}] \\
\mathcal{S}_2 &= \phi_{M_1,M_2}[V_2^p \in \{V_2^0, \ldots, V_2^{p-1}, V_2^{p+1}, \ldots, V_2^{m_2}\}] \ .
\end{aligned}$$

Lemma 19 says that $|\mathcal{S}_1| \leq m_1$ and $|\mathcal{S}_2| \leq m_2$. Putting everything together, the lemma follows as $2(m_1 + m_2) \leq 4\ell$. ▮

Lemma 18 *Let $n, m_1, m_2 \geq 1$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$ then for $b \in \{1,2\}$,*

$$\left|\phi_{M_1,M_2}[V_b^{m_b} \in \{V_b^0, V_b^1, \ldots, V_b^{m_b-1}\}]\right| = m_b \quad ▮$$

Proof: We prove only the claim for $b = 1$, and then briefly discuss how to extend the proof to $b = 2$.

If $V_1^{m_1} \in \{V_1^0, \ldots, V_1^{m_1-1}\}$ then there must be a (j, V_1^i)-accident for some $i \in [0..m_1 - 1]$ and $j \in [i + 1..m_1]$ and then induced collisions in steps $j + 1$ to m_1. Thus $V_1^{j+k} = V_1^{i+k}$ for all $k \in [0..m_1 - j]$. For $j \in [1..m_1]$ let N_j be the number of structure graphs $G \in \mathcal{G}^1(M_1, M_2)$ such that $V_1^{m_1}(G) \in \{V_1^0(G), \ldots, V_1^{m_1-1}(G)\}$ and there is a $(V_1^i(G), j)$-accident for some $i \in [0..j - 1]$. Then

$$\left|\phi_{M_1,M_2}[V_1^{m_1} \in \{V_1^0, \ldots, V_1^{m_1-1}\}]\right| = \sum_{j=1}^{m_1} N_j \ .$$

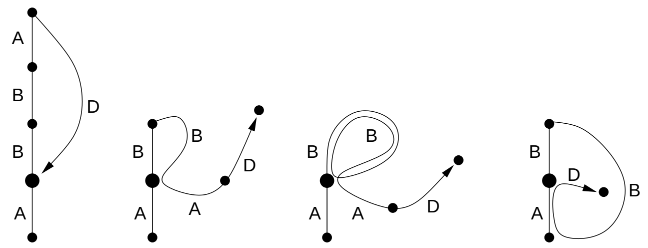

Figure 7: An example for the proof of Lemma 19 with $m_1 = 5, M_1 = A\|B\|B\|A\|D$ and $r = 1$, where $A, B, D \in \{0,1\}^n$ are distinct. (The large dot is $V_1^r = V_1^1$.) Here we have $N_r = m - r = \mu_2(M_1^1) = N_1 = m_1 - 1 - \mu_2(M_1^1) = 5 - 1 - \mu_2(A) = 5 - 1 - 1 = 3$. Those cases correspond to the first three graphs in the figure. The fourth graph corresponds to $N_{r-1} = N_0 = \mu_2(\star \parallel M_1^{1\to r}) = \mu_2(\star \parallel A) = 1$.

Let $\mu_1(S)$ denote the number of block-aligned occurrences of the substring S in M_1. (For example, $\mu_1(A \parallel B) = 2$ if $M_1 = A \parallel B \parallel B \parallel A \parallel B$ for some distinct $A, B \in \{0,1\}^n$.) It is possible to have a (m_1, V_1^i)-accident for any $i \in [0..m_1 - 1]$ for which $M_1^i \neq M_1^{m_1}$ (cf. Figure 6) and thus $N_{m_1} = m_1 - \mu_1(M_1^{m_1}) + 1$. It is possible to have a $(V_1^i, m_1 - 1)$-accident and also have $V_1^{m_1} \in \{V_1^0, \ldots, V_1^{m_1-1}\}$ for any $i \in [0..m_1 - 2]$ for which $M_1^i \neq M_1^{m_1-1}$ and $M_1^{i+1} = M_1^{m_1}$ and thus $N_{m_1-1} = \mu_1(M_1^{m_1}) - \mu_1(M_1^{m_1-1\to m_1})$. In general for $j \in [1..m_1 - 1]$ we have $N_j = \mu_1(M_1^{j+1\to m_1}) - \mu_1(M_1^{j\to m_1})$. Using cancellation of terms in the sum we have

$$\sum_{j=1}^{m_1} N_j = m_1 + 1 - \mu_1(M_1^{1\to m_1}) = m_1$$

which proves the lemma for the case $b = 1$. The case for $b = 2$ follows by symmetry, more precisely

$$\left|\phi_{M_1,M_2}[V_b^{m_b} \in \{V_b^0, V_b^1, \ldots, V_b^{m_b-1}\}]\right|$$

is invariant under exchanging M_1 with M_2 and changing b (from 1 to 2 or 2 to 1) simultaneously. In fact, all predicates which do not make use of the representation of the vertices have this property (so for example $V_1^5 = 5$ is not symmetric in the above sense but $V_1^5 \notin \{V_1^0, \ldots, V_1^4\}$ is). ∎

Next we have a generalization of Lemma 18.

Lemma 19 *Let $n, m_1, m_2 \geq 1$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$ then for $b \in \{1,2\}$ and any $r \in [0..m_b]$,*

$$\left|\phi_{M_1,M_2}[V_b^r \in \{V_b^0, \ldots, V_b^{r-1}, V_b^{r+1}, \ldots, V_b^{m_b}\}]\right| \leq m_b\ . \quad ∎$$

Proof: We prove it for the case $b = 1$. (The case $b = 2$ is analogous.) By Lemma 18 we have $|\phi_{M_1,M_2}[V_1^r \in \{V_1^0, \ldots, V_1^{r-1}\})| = r$. It remains to show that

$$\left|\phi_{M_1,M_2}[V_1^r \in \{V_1^{r+1}, \ldots, V_1^{m_1}\} \wedge V_1^r \notin \{V_1^0, \ldots, V_1^r\}]\right| \leq m_1 - r\ .$$

We may assume that $V_1^i \neq V_1^j$ for all $0 \leq i < j \leq r-1$, as otherwise we have already used up our accident and there's no way to get $V_1^r \in \{V_1^{r+1}, \ldots, V_1^{m_1}\}$ any more. If $V_r^{\in}\{V_1^{r+1}, \ldots, V_1^{m_1}\}$ then there is a (i, V_1^j)-accident for some $0 \leq j \leq r < i$. For $j \in [0..r]$ let N_j be the number of structure graphs $G \in \mathcal{G}^1(M_1, M_2)$ such that $V_1^r(G) \in \{V_1^{r+1}(G), \ldots, V_1^{m_1}(G)\}$, $V_1^r(G) \notin \{V_1^0(G), \ldots, V_1^r(G)\}$ and there is a (i, V_1^j)-accident for some $i \in [r+1..m_1]$. Then

$$\left|\phi_{M_1,M_2}[V_1^r \in \{V_1^{r+1}, \ldots, V_1^{m_1}\} \wedge V_1^r \notin \{V_1^0, \ldots, V_1^r\}]\right| = \sum_{j=0}^{r} N_j\ .$$

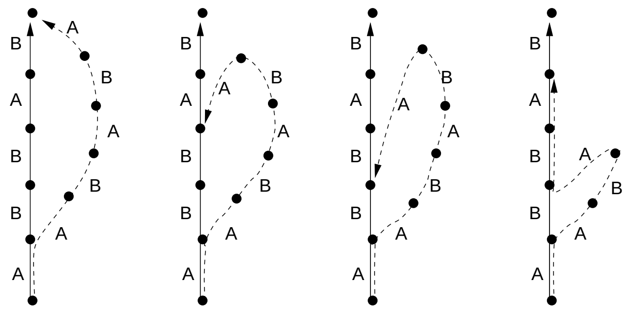

Figure 8: An example for the proof of Lemma 20 with $M_1 = A\|B\|B\|A\|B$ and $M_2 = A\|A\|B\|A\|B\|A$, thus $m_1 = 5, m_2 = 6, p = 1$. Here $\mu_3(S)$ is the number of block-aligned occurrences of S in $M_1^{p+1\to m} = B\|B\|A\|B$. The solid line corresponds to M_1 and the dotted line to M_2. We get $N_{m_2} = N_6 = m_1 - p - \mu_3(M_2^{m_2}) = 5 - 1 - \mu_3(A) = 5 - 1 - 1 = 3$, the three cases correspond to the first three graphs in the figure. Furthermore, $N_5 = \mu_3(\star \parallel A) - \mu_3(B\|A) = 1 - 1 = 0$ and $N_4 = \mu_3(\star \parallel B \parallel A) - \mu_3(A \parallel B \parallel A) = 1 - 0 = 1$, this last case corresponding to the last graph in the figure.

Let $\mu_2(S)$ be the number of block-aligned occurrences of the substring S in $M_1^{r+1\to m_1}$, and adopt the convention that $\mu_2(M_1^0) = 0$. Since we can only have an (j, V_1^r)-accident when $M_1^j \neq M_1^r$ we have $N_r = m - r - \mu_2(M_1^r)$. For $i > r$, a (i, V_1^r)-accident is possible and will result in $V_1^r \in \{V_1^{r+1}, \ldots, V_1^{m_1}\}$ only if $M_1^{i\to i+1} = X\|M_r$ for some $X \neq M_1^{r-1}$. Now with $\star$ being a wildcard standing for an arbitrary block we have $N_{r-1} = \mu_2(\star \parallel M_1^r) - \mu_2(M_1^{r-1\to r})$. In general, for $j \in [1..r-1]$ we have $N_j = \mu_2(\star \parallel M_1^{j+1\to r}) - \mu_2(M_1^{j\to r})$ and $N_0 = \mu_2(\star \parallel M_1^{1\to r})$. Now, as $\mu_2(\star \parallel S) \leq \mu_2(S)$ for any S, we get

$$\sum_{j=0}^{r} N_j \leq m_1 - r \ . \quad \blacksquare$$

Lemma 20 *Let $n, m_1, m_2 \geq 1$. Let $M_1 \in B_n^{m_1}, M_2 \in B_n^{m_2}$ with $\mathsf{pf}(M_1, M_2) = \mathsf{false}$. Let $p = \mathsf{cpl}(M_1, M_2)$. Then*

$$\left|\phi_{M_1,M_2}[V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\} \wedge \mathsf{NoLoop}]\right| \leq m_1 \ . \quad \blacksquare$$

Proof of Lemma 20: The condition that NoLoop is true means that the neither the M_1- nor the M_2-path makes a loop. Thus the only possibility to get $V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\}$ here is to have an (j, i)-accident for some $p < i \leq m_1$ and $m_1 + p < j$ where $M_1^{i\to i+m-j} = X \parallel M_2^{j+1\to m_2}$ for an $X \neq M_2^j$. Let N_j denote the number of those cases. Let $\mu_3(S)$ denote the number of block-aligned occurrences of the substring S in $M_1^{p+1\to m}$. Then for $j = m_2$ we have

$$N_{m_2} = m - p - \mu_3(M_2^{m_2})$$

and for $j = m_2 - 1$ we get

$$N_{m-1} = \mu_3(\star \parallel M_2^{m_2}) - \mu_3(M_2^{m_2-1\to m_2})$$

and in general for $p + 1 \leq j < m_2$

$$N_j = \mu_3(\star \parallel M_2^{j+1\to m_2}) - \mu_3(M_2^{j\to m_2})$$

Now as $\mu_3(\star \parallel S) \leq \mu_3(S)$, we get

$$|\phi_{M_1,M_2}(V_2^{m_2} \in \{V_1^{p+1}, \ldots, V_1^{m_1}\} \wedge \mathsf{NoLoop})| = \sum_{j=p+1}^{m_2} N_j \leq m - p \,. \quad \Box$$

Acknowledgments

Bart Preneel was the first we heard to ask, back in 1994, if the m^2 term can be improved in the CBC MAC bound of $m^2q^2/2^n$.

References

[1] M. Bellare, O. Goldreich, and A. Mityagin. The power of verification queries in message authentication and authenticated encryption. Cryptology ePrint Archive: Report 2004/309.

[2] M. Bellare, J. Kilian, and P. Rogaway. The security of the cipher block chaining message authentication code. *Journal of Computer and System Sciences* (JCSS), vol. 61, no. 3, pp. 362–399, 2000. Earlier version in *Crypto '94*.

[3] M. Bellare, K. Pietrzak, and P. Rogaway. Improved security analyses for CBC MACs. Preliminary version of this paper, *Advances in Cryptology – CRYPTO '05*, Lecture Notes in Computer Science Vol. , V. Shoup ed., Springer-Verlag, 2005.

[4] M. Bellare and P. Rogaway. The game-playing technique. Cryptology ePrint Archive: Report 2004/331.

[5] A. Berendschot, B. den Boer, J. Boly, A. Bosselaers, J. Brandt, D. Chaum, I. Damgård, M. Dichtl, W. Fumy, M. van der Ham, C. Jansen, P. Landrock, B. Preneel, G. Roelofsen, P. de Rooij, and J. Vandewalle. *Final Report of Race Integrity Primitives*. Lecture Notes in Computer Science, vol. 1007, Springer-Verlag, 1995

[6] R. Berke. On the security of iterated MACs. Diploma Thesis, ETH Zürich, August 2003.

[7] J. Black and P. Rogaway. CBC MACs for arbitrary-length messages: the three-key constructions. *Advances in Cryptology – CRYPTO '00*, Lecture Notes in Computer Science Vol. 1880, M. Bellare ed., Springer-Verlag, 2000.

[8] Y. Dodis. Personal communication to K. Pietrzak. 2004.

[9] Y. Dodis, R. Gennaro, J. Håstad, H. Krawczyk, and T. Rabin. Randomness extraction and key derivation using the CBC, Cascade, and HMAC modes. *Advances in Cryptology – CRYPTO '04*, Lecture Notes in Computer Science Vol. 3152, M. Franklin ed., Springer-Verlag, 2004.

[10] G. Hardy and E. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 1980.

[11] E. Jaulmes, A. Joux, and F. Valette. On the security of randomized CBC-MAC beyond the birthday paradox limit: a new construction. *Fast Software Encryption '02*, Lecture Notes in Computer Science Vol. 2365 , J. Daemen, V. Rijmen ed., Springer-Verlag, 2002.

[12] J. Kilian and P. Rogaway. How to protect DES against exhaustive key search (an analysis of DESX). *Journal of Cryptology*, vol. 14, no. 1, pp. 17–35, 2001. Earlier version in *Crypto '96*.

[13] U. Maurer. Indistinguishability of random systems. *Advances in Cryptology – EUROCRYPT '02*, Lecture Notes in Computer Science Vol. 2332, L. Knudsen ed., Springer-Verlag, 2002.

[14] National Institute of Standards and Technology, U.S. Department of Commerce, M Dworkin, author. Recommendation for block cipher modes of operation: the CMAC mode for authentication. NIST Special Publication 800-38B, May 2005.

[15] E. Petrank and C. Rackoff. CBC MAC for real-time data sources. *Journal of Cryptology*, vol. 13, no. 3, pp. 315–338, 2000.

[16] V. Shoup. Sequences of games: a tool for taming complexity in security proofs. Cryptology ePrint report 2004/332, 2004.

[17] S. Vaudenay. Decorrelation over infinite domains: the encrypted CBC-MAC case. *Communications in Information and Systems* (CIS), vol. 1, pp. 75–85, 2001.

[18] M. Wegman and L. Carter. New classes and applications of hash functions. *Symposium on Foundations of Computer Science* (FOCS), pp. 175–182, 1979.

A Proof of Lemma 4

Let $A \in \mathcal{A}^{\mathsf{atk}}_{q,n,m}$. We wish to bound $\mathbf{Adv}_{\mathrm{ECBC}}(A) = \Pr[A^{\mathrm{ECBC}_{\pi_1,\pi_2}} \Rightarrow 1] - \Pr[A^f \Rightarrow 1]$ where $\pi_1, \pi_2 \xleftarrow{\$} \mathrm{Perm}(n)$ and $f \xleftarrow{\$} \mathrm{Func}(n)$. We realize ECBC by the game W1 that follows this paragraph: $\Pr[A^{\mathrm{ECBC}_{\pi_1,\pi_2}} \Rightarrow 1] = \Pr[A^{\mathrm{W1}} \Rightarrow 1]$. Define game W0 from game W1 by dropping the (shaded) statements that immediate follow the setting of *bad* at lines 12 and 13. Observe that $\Pr[A^f \Rightarrow 1] = \Pr[A^{\mathrm{W0}} \Rightarrow 1]$. Thus $\mathbf{Adv}_{\mathrm{ECBC}}(A) = \Pr[A^{\mathrm{W1}} \Rightarrow 1] - \Pr[A^{\mathrm{W0}} \Rightarrow 0]$ and, by the fundamental lemma of game playing, $\mathbf{Adv}_{\mathrm{ECBC}}(A) \leq \Pr[A^{\mathrm{W0}} \text{ sets } bad\,]$. We now bound this quantity.

Initialize Game W1 (as written) and W0 (omit shaded statements)
00 $\pi_1 \xleftarrow{\$} \mathrm{Perm}(n)$, $\pi_2(X) \leftarrow \mathsf{undefined}$ for all $X \in \{0,1\}^n$, $bad \leftarrow \mathsf{false}$
On query M
10 $X \leftarrow \mathrm{CBC}_{\pi_1}(M)$
11 $Y \xleftarrow{\$} \{0,1\}^n$
12 **if** $Y \in \mathrm{Ran}(\pi_2)$ **then** $bad \leftarrow \mathsf{true}$, $Y \xleftarrow{\$} \overline{\mathrm{Ran}}(\pi_2)$
13 **if** $X \in \mathrm{Dom}(\pi_2)$ **then** $bad \leftarrow \mathsf{true}$, $Y \leftarrow \pi_2(X)$
14 **return** $\pi_2(X) \leftarrow Y$

The probability that *bad* gets set at line 12 of game W0 is at most $(1 + 2 + \cdots + (q-1))/2^n = 0.5q(q-1)/2^n$. The probability that *bad* gets set at line 13 of game W0 is at most $\binom{q}{2}\delta$ where δ is the maximum of the probability that some two queries M and M' collide under CBC_{π_1}. Note that game W0 provides to the adversary *no* information about π_1, and so $\delta \leq \mathbf{CP}^{\mathsf{atk}}_{n,m}$. Concluding, we have that $\Pr[A^{\mathrm{W2}} \text{ sets } bad\,] \leq 0.5q(q-1)/2^n + 0.5q(q-1) \cdot \mathbf{CP}^{\mathsf{atk}}_{n,m}$ and the lemma follows.

B Proof of Lemma 8

Like in section 6 we let $n \geq 1, t \geq 1$, $\mathcal{M} = \{M_1, \ldots, M_t\}$, each $M_i = M_i^1 \cdots M_i^{m_i} \in B_n^{m_i}$. For $1 \leq j \leq t : m^j = \sum_{i=1}^{j} m_i$ and $m = m^t$. The definitions of structure graphs $\mathcal{G}(\mathcal{M})$ and accidents

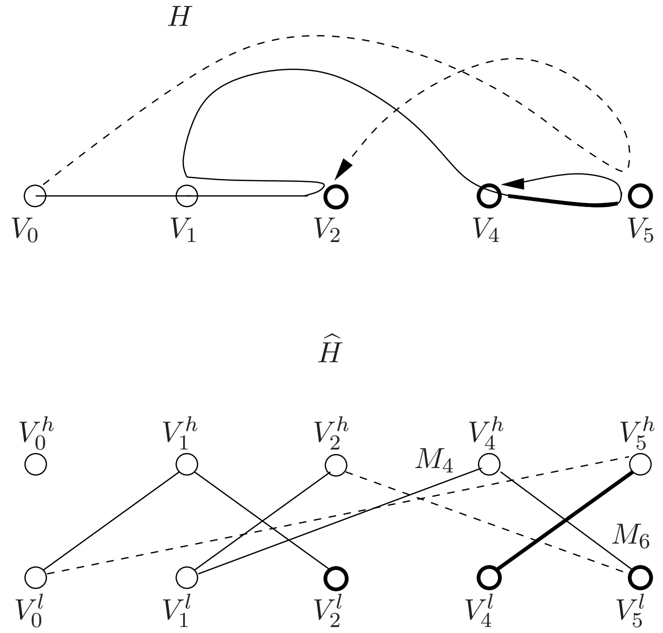

Figure 9: A structure graph $H \in \mathcal{G}(M_1 \| M_2 \cdots M_6, M_7 \| M_8)$ and the corresponding $\widehat{H}$ (the second message $M_7 \| M_8$ is drawn by the dotted line). To illustrate the proof of the lemma, note that here we have $I_H = \{0, 1, 2, 4, 5\}$ and $\mathsf{CoCo}(\widehat{H}) = 3$. Those components are given by the subgraphs induced by the vertex sets $\{V_0^h\}, \{V_0^l, V_1^h, V_2^l, V_5^h, V_4^l\}$ and $\{V_1^l, V_2^h, V_5^l, V_4^h\}$. Let P_1 and P_2 denote the latter two components respectively, then $I_1 = \{0, 2, 4\}, I_2 = \{1, 5\}$ and $I_1^- = \{2, 4\}, I_2^- = \{5\}$, so $I^- = \{2, 4, 5\}$. For example in step $5 \in I^-$ (indicated by the think line) the value C_5 (corresponding to the vertex V_5^l) is revealed to us, and there is only one possibility for C_5 to stay consistent with $\widehat{H}$ (i.e. it must satisfy $C_1 \oplus M_4 = C_5 \oplus M_6$).

acc are in section 6.

We need some new definitions for this section. From $G \in \mathcal{G}(\mathcal{M})$ we get the *undirected* bipartite graph $\widehat{G}$ by splitting each $i \in V(G)$ into two vertices V_i^l and V_i^h such that the vertex V_i^l keeps the outgoing edges and V_i^h keeps the ingoing edges (cf. figure B). So the directed edge $(V_i, V_j) \in E(G)$ maps to the undirected edge $(V_i^l, V_j^h) \in E(\widehat{G})$.

For a graph G we denote with $\mathsf{CoCo}(G)$ the number of connected components of G. The number of accidents of a graph $G \in \mathcal{G}(\mathcal{M})$, denoted $\mathsf{acc}(G)$, is

$$\mathsf{acc}(G) = |V(G)| - \mathsf{CoCo}(\widehat{G}) + 1 \tag{13}$$

We'll later show that this is indeed the number of accidens i.e.

Claim 21 *For any $G \in \mathcal{G}(\mathcal{M})$, acc as defined above and Acc as defined in eq.(5)*

$$\mathsf{acc}(G) = |\mathsf{Acc}(G)|$$

With the above Claim we can state Lemma 8 as

$$\Pr[G \stackrel{\$}{\leftarrow} \mathcal{G}(\mathcal{M}) : \ G = H] \leq (2^n - m)^{-\mathsf{acc}(H)} \tag{14}$$

Proof of Lemma 8: Let $V^l = \{V_i^l : i \in I_H\}$ and $V^h = \{V_i^h : i \in I_H\}$ denote the two partitions of the bipartite graph $\widehat{H}$.

The vertex $V_0 \in V(H)$ is somewhat special, and we must consider the two cases where $\mathsf{in}_H(V_0)$, the indegree of V_0, is 0 and when it is > 0 separately. We first prove the case where $\mathsf{in}_H(V_0) = 0$. At the end of the proof we will describe how to adapt the proof for H's where $\mathsf{in}_H(V_0) > 0$.

Let $c = \mathsf{CoCo}(\widehat{H})-1$ and $P_1, \ldots, P_c$ denote the connected components of $\widehat{H}$ without the component build by the isolated vertex V_0^h (V_0^h is isolated as $\mathsf{in}_H(V_0) = 0$). Let $I_i = \{j : V_j^l \in P_i\}$ (so the $I_1, \ldots, I_c$ are a partition of I_H) and $I_i^- = I_i \setminus \min\{i, i \in I_i\}$.

We now upper bound the probability that $\Pr[G \xleftarrow{\$} \mathcal{G}(\mathcal{M}) : \ G = H]$, or equivalently $\Pr[\pi \xleftarrow{\$} \mathrm{Perm}(n) : \ G_\pi^{\mathcal{M}} = H]$ which again is equivalent to $\Pr[\pi \xleftarrow{\$} \mathrm{Perm}(n) : \ \widehat{G}_\pi^{M} = \widehat{H}]$.

Assume a $\pi \xleftarrow{\$} \mathrm{Perm}(n)$ is sampled (but not shown to us). Now in m steps the values $C_1, \ldots, C_m$ are revealed to us, we say that we are consistent after step i if

$$\Pr[\widehat{G}_\pi^{\mathcal{M}} = \widehat{H} | C_0, \ldots, C_i] > 0$$

i.e. given the values $C_0, \ldots, C_i$ there is a non-zero probability that finally $\widehat{G}_\pi^{M_1,\ldots,M_t} = \widehat{H}$.

Now we claim the following, for each $i \in I^-$ the following is true: the probability that we are consistent after step i, conditioned on being consistent after step $i-1$, is at most $(2^n - m)^{-1}$. The lemma then follows as

$$|I^-| = |I_1^- \cup \cdots \cup I_c^-| = |I_H| - c = |V(H)| - (\mathsf{CoCo}(\widehat{H}) - 1) = \mathsf{acc}(H). \tag{15}$$

We prove the above claim. Consider any $i \in I_j^- \subseteq I^-$ and $C_0, \ldots, C_{i-1}$ which are consistent with $\widehat{H}$. Assume that π is indeed such that $\widehat{G}_\pi^{\mathcal{M}} = \widehat{H}$, then $\widehat{G}_\pi^{\mathcal{M}}$ has a connected component P_j like $\widehat{H}$. In step i at least one C_k where $k \in I_j$ (namely $\{k\} = I_j \setminus I_j^-$) is already determined, and this C_k uniquely determines all other C_x where $x \in I_j$ (and in particular $i = x$). So we know the value $C_i = \pi(C'_{i-1} \oplus M_i)$ which will be exposed to us in this ith step, let X denote this value. But without the assumption that $\widehat{G}_\pi^{\mathcal{M}} = \widehat{H}$ we have two possibilities

1. If $C'_{i-1} \oplus M_i = C'_{j-1} \oplus M_j$ for some $j < i$ then also $C_i = C_j$. But this means that $[i]_{\widehat{G}_\pi^{\mathcal{M}}} = [j]_{\widehat{G}_\pi^{\mathcal{M}}}$ but as as $i \in I_H$ we have $[i]_H = i > j \geq [j]_H$ and thus $\widehat{G}_\pi^{\mathcal{M}} \neq \widehat{H}$.
2. If for all $j < i$ we have $C'_{i-1} \oplus M_i \neq C'_{j-1} \oplus M_j$ then the value $\pi(C'_{i-1} \oplus M_i)$ is just uniformly random over $\{0,1\}^n \setminus \{C_1, \ldots, C_{i-1}\}$, and the probability that $C_i = \pi(C'_{i-1} \oplus M_i) = X$, which as we just showed is the only possibility not to become inconsistent, is at most $(2^n - i)^{-1}$.

This concludes the proof for the case $\mathsf{in}_H(V_0) = 0$, we now describe how to adapt the proof for the case $\mathsf{in}_H(V_0) > 0$. Now let $c = \mathsf{CoCo}(\widehat{H})$ (not $\mathsf{CoCo}(\widehat{H}) - 1$ as before) and $P_1, \ldots, P_c$ denote all the connected components of $\widehat{H}$. Let the I_i, I_i^- and I^- be defined as before except that we add a value z to I^- where $z = min\{j : j > 0, V_j = V_0\}$ is the first step in which we "come back" to the vertex V_0 (note that (15), i.e. $|I^-| = \mathsf{acc}(H)$, is still satisfied). We now show that like for all other values in I^- also for z it is true that assuming that we are consistent after step $z-1$, we will be consistent after step z (which holds iff $\pi(C'_{z-1} \oplus M_z) = C_0 = 0^n$) with probability at most $(2^n - z)^{-1}$. The reason for this is that although after the step $z - 1$ the vertex V_0 (which has always the value $C_0 = 0^n$) is already there, it has no ingoing edge, so $\pi^{-1}(C_0)$ is not determined (i.e. it is uniformly random over $\{0,1\}^n \setminus \{\pi^{-1}(C_1), \ldots, \pi^{-1}(C_{z-1})\}$), thus $\Pr[\pi(C'_{z-1} \oplus M_z) = 0^n] < (2^n - z)^{-1}$.

Proof of Claim 21: Recall that for $G \in \mathcal{G}(\mathcal{M})$ we denote by G_i the subgraph of G containing the i first edges. We show $\mathsf{acc}(G) = |\mathsf{Acc}(G)|$ by induction on i for $\mathsf{acc}(G_i) = |\mathsf{Acc}(G_i)|$ (recall that $G_m = G$). We leave it to the reader to verify the anchor $\mathsf{acc}(G_1) = |\mathsf{Acc}(G_1)|$. Now assume $\mathsf{acc}(G_{i-1}) = |\mathsf{Acc}(G_{i-1})|$ and recall that $\mathsf{acc}(G_i) = |V(G_i)| - \mathsf{CoCo}(\widehat{G}_i) + 1$.

- If in step i we have no collision, i.e. $\forall j < i : (i, j) \notin \mathsf{Col}(G)$, then $|\mathsf{Acc}(G_{i-1})| = |\mathsf{Acc}(G_i)|$ by definition. If there's no collision then the number of verticies increases by one, $V(G_i) = V(G_{i-1})+1$, but also the number of connected components increases, $\mathsf{CoCo}(\widehat{G}_i) = \mathsf{CoCo}(\widehat{G}_{i-1})+1$, so $|\mathsf{acc}(G_{i-1})| = |\mathsf{acc}(G_i)|$. This shows $\mathsf{acc}(G_i) = |\mathsf{Acc}(G_i)|$.

- If in step i we an have an induced collision, i.e. for some $j, (i, j) \in \mathsf{IndCol}(G)$ then again $|\mathsf{Acc}(G_{i-1})| = |\mathsf{Acc}(G_i)|$ by definition. As we have a collision we get no new vertex, $V(G_i) = V(G_{i-1})$, but also we don't decrease the nuber of connected components $\mathsf{CoCo}(\widehat{G}_i) = \mathsf{CoCo}(\widehat{G}_{i-1})+1$, to see this note that and induced collision means that the i'th edge closes a cycle with alternating edge directions in G_{i-1}, but such a cycle is also a cycle in the bipartite graph $\widehat{G}_{i-1}$, so the edge connets two verticies which already are in the same component in $\widehat{G}_{i-1}$. Again $|\mathsf{acc}(G_{i-1})| = |\mathsf{acc}(G_i)|$ and thus $\mathsf{acc}(G_i) = |\mathsf{Acc}(G_i)|$ follows.

- If in step i we have an accident, i.e. for some $j, (i, j) \in \mathsf{Col}(G)$ but $(i, j) \notin \mathsf{IndCol}(G)$ then $|\mathsf{Acc}(G_{i-1})|+1 = |\mathsf{Acc}(G_i)|$ by definition. As we have a collision we get no new vertex, $V(G_i) = V(G_{i-1})$, but the nuber of connencted componets decreases, $\mathsf{CoCo}(\widehat{G}_i) = \mathsf{CoCo}(\widehat{G}_{i-1})-1$. To see this consider what happens if we add the i'th edge (u, v) to $\widehat{G}_{i-1}$: there cannot already be a path from u to v in $\widehat{G}_{i-1}$ because as we have seen in the previous case this would mean that the collision is induced. So u and v do not lie in the same connected component. So $|\mathsf{acc}(G_{i-1})| + 1 = |\mathsf{acc}(G_i)|$ and $\mathsf{acc}(G_i) = |\mathsf{Acc}(G_i)|$ follows.