

A new bound for l -wise almost universal hash functions

L.H. Nguyen and A.W. Roscoe
Oxford University Computing Laboratory

Abstract. Using the pigeon-hole principle, we derive a new bound for the key length in a l -wise almost universal hash function where the multicollision or l -collision probability is bounded above by $\epsilon \in [0, 1]$. The important features of this bound are (1) it decreases very slowly as l increases, and (2) the key length grows at least linearly with the logarithm of the message length. To our knowledge, this is the first almost universal hash bound for any integer $l \geq 2$. This work arises from the use of l -wise almost universal hash functions in manual authentication protocols.

1 Introduction

An almost universal family of hash functions AU with parameters (K, M, b) was introduced by Carter and Wegman [3]. A universal family consists of 2^K hash functions, each of which maps a M -bit message from $\{0, 1\}^M$ into $\{0, 1\}^b$ or a b -bit output. In this paper, we will derive a AU_l -bound whose l -collision probability is $\epsilon \in [0, 1]$ for any $l \geq 2$.

Definition 1. A l -wise almost universal hash function ϵ - AU_l satisfies that for any l distinct and equal-length messages m_1, \dots, m_l and as key k is selected randomly from $\{0, 1\}^K$:

$$\text{Prob}_{\{0 \leq k < 2^K\}}[h_k(m_1) = h_k(m_2) = \dots = h_k(m_l)] \leq \epsilon$$

This specification arises from the use of AU_l in a number of group protocols in the manual authentication technology, including the schemes of Laur and Pasini [7], and the authors [8, 9]. In these protocols, parties have to manually compare a universal hash value of some key and public data that they seek to agree on. An attacker therefore will attempt to fool multiple parties into accepting more than two different versions of that piece of data. For this reason, it is desirable that we constrain the chance of a multicollision attack as specified in Definition 1. Moreover, to build such an ϵ - AU_l , we want to understand the theoretical constraints among the values of different l -wise universal hash parameters, which is the bound we derive here.

We note that there have been bounds for a l -wise *strongly* universal hash function, which is a stronger version of AU_l , due to Kurosawa et al. [6], sadly these are only proved for the perfect case where $\epsilon = 2^{-b}$. What also relevant to our work is the equivalence between error-correcting codes (ECC) and pairwise universal hash functions, and thus several ECC-bounds have been transformed into bounds for universal hash functions [4, 11]. This strategy however cannot be used to derive AU_l -bound for $l > 2$ because the minimum Hamming distance among pairs of codewords corresponds to the pairwise-collision property in AU_2 . ECC-parameters therefore do not give enough information to analyse l -collision in AU_l .

Even though multicollision attacks in AU_l and cryptographic hash functions are not the same, it might be worth to mention that the idea of multicollision has been encountered in cryptographic hash design such as the cascaded or Merkle-Damgård structure [5] and NIST's specification for SHA-3 candidates [1] (Section 2.B.1). Multicollision resistance in cryptographic hash functions is also required in several identification and signature schemes [2, 5, 10]. The intuitive reason is because constructing l messages with the same hash value should be much harder than constructing only two of these.

2 A new bound for almost universal hash functions

The following AU_l -bound tells us the lower bound for the key bitlength in terms of the l -collision probability ϵ and the bitlengths of message and hash output. Although K, M , and b are often integers, this bound applies to both integer and non-integer bitlengths. For simplicity the subscript in \log_2 is omitted as logarithms in all formulae are taken to base 2.

Theorem 1. For any integer $l \geq 2$, if there exist a l -wise almost universal hash function ϵ - AU_l with parameters (K, M, b) , then the below conditions apply. In this bound, we define an integer x such that $2^{xb} < l \leq 2^{(x+1)b}$ and $M = bt + b'$ where t is an integer and $0 \leq b' < b$.

- (i) If $b' \leq \log(l-1) - xb$ then $K \geq \log(\epsilon^{-1}(\lfloor M/b \rfloor - x - 1))$
- (ii) If $b' > \log(l-1) - xb$ then $K \geq \log(\epsilon^{-1}(\lfloor M/b \rfloor - x))$

Proof. The *pigeon-hole* principle states that given two positive integers n and m , if n items are put into m holes then at least one hole must contain more than or equal to $\lceil n/m \rceil$ items.

For any key k_1 , there exists a hash value h_1 such that there are at least $\lceil 2^{M-b} \rceil$ distinct messages forming a set S_1 all hashing to h_1 under the same key k_1 , thanks to the pigeon-hole principle. For any choice of k_2 other than k_1 , there will also be a collection of at least $\lceil 2^{M-2b} \rceil$ different messages from set S_1 mapping to some hash value h_2 under k_2 . We note that the value of h_1 can be either different from or equal to h_2 .

Since we defined $2^{xb} < l \leq 2^{(x+1)b}$, we can always repeat this process $t-x-1$ times and obtain at least $v = \lceil 2^{M-(t-x-1)b} \rceil = \lceil 2^{(x+1)b+b'} \rceil$ distinct messages m_1, \dots, m_v where $v \geq l$, and $t-x-1$ different keys k_1, \dots, k_{t-x-1} such that

$$\forall k \in \{k_1, \dots, k_{t-x-1}\} : h_k(m_1) = h_k(m_2) = \dots = h_k(m_v)$$

This leads to two possibilities:

- (i) If $b' \leq \log(l-1) - xb$ then $l \leq \lceil 2^{(x+1)b+b'} \rceil \leq 2^b(l-1)$. We *cannot* repeat the above process further because at least l distinct messages must be left to have a l -collision. Thus to bound the l -collision probability above by ϵ , we arrive at:

$$\begin{aligned} \epsilon 2^K &\geq t-x-1 = \lfloor M/b \rfloor - x - 1 \\ K &\geq \log(\epsilon^{-1}(\lfloor M/b \rfloor - x - 1)) \end{aligned}$$

- (ii) If $b' > \log(l-1) - xb$ then $\lceil 2^{(x+1)b+b'} \rceil \geq 2^b(l-1) + 1$. Repeating the above process for one more random key k_{t-x} will end up with at least l distinct messages that map to the same values under $t-x = \lfloor M/b \rfloor - x$ keys. We therefore have $\epsilon 2^K \geq \lfloor M/b \rfloor - x$, which means that $K \geq \log(\epsilon^{-1}(\lfloor M/b \rfloor - x))$.

We cannot repeat the above process $t-x+1$ times because the number of different messages we would end up with is $\lceil 2^{(x-1)b+b'} \rceil \leq 2^{xb} < l$, which is insufficient to form a l -collision. \square

3 Interpretations of the new bound

We observe that the bound decreases very slowly as we increase l , which is not surprising since the bigger l the more unlikely a l -collision can be formed, and so fewer keys are required. Moreover, if (ϵ, l, b) are fixed then as M increases K grows at least in proportion to $\log M$.

For $l = 2$, our AU_2 -bound is satisfied with equality by the well-studied polynomial hashing scheme over finite field of Johansson et al. [4] where $x = x_1 \dots x_t \in \{0, 1\}^{tb}$ for any integer $t \in [2, 2^b]$, $k \in \{0, 1\}^b$ and $h_k(x) = x_1 + x_2 k + \dots + x_t k^{t-1}$, because $\epsilon = (t-1)/2^b \in [2^{-b}, 1)$, and so $K = \log(\epsilon^{-1}(M/b - 1))$. Theorem 1(i) is satisfied with equality.

For $l > 2$, we give two constructions for AU_3 and AU_5 that meet the bound with equality.

$(\epsilon = 1/2)$ - AU_3	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
k_1	0	1	2	3	0	1	2	3	0
k_2	3	2	1	0	2	3	2	1	0

Table 1. An AU_3 having $2^M = 9$, $2^b = 4$, and $\epsilon = 1/2$ requires $2^K \geq \epsilon^{-1} \lfloor M/b \rfloor = 2$, due to Theorem 1(ii).

$(\epsilon = 1/3)$ - AU_5	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9
k_1	0	1	0	1	0	1	0	1	0
k_2	1	0	1	1	1	0	0	0	1
k_3	0	1	0	0	1	1	0	1	1

Table 2. An AU_5 having $2^M = 9$, $2^b = 2$, and $\epsilon = 1/3$ requires $2^K \geq \epsilon^{-1} (\lfloor M/b \rfloor - 2) = 3$, due to Theorem 1(ii).

4 Comparison against other AU -bound

To our knowledge, the only other AU -bound is due to Stinson [11] but it works with $l = 2$. It is very different from our AU_2 -bound that can be rounded up to $2^K \geq \frac{M}{\epsilon b}$ for comparison.

$$\text{Stinson's } AU_2\text{-bound: } 2^K \geq \frac{2^M(2^b - 1)}{2^M(\epsilon 2^b - 1) + 2^{2b}(1 - \epsilon)}$$

For $\epsilon = 2^{-b}$, Stinson's bound is stronger than ours for then it gives $K \geq M - b$, which means that K grows at least linearly with M . We stress that although our AU_2 -bound can also be met with equality when $\epsilon = 2^{-b}$, it does so with a very limited range of values of (K, M, b) , i.e. $M = 2K = 2b$ as in the polynomial hashing construction. In contrast, if $\epsilon > 2^{-b}$ and $M \gg 2b$, Stinson's bound significantly underestimates K because it can never prove stronger a bound than $2^K \geq 2^b / (\epsilon 2^b - 1)$. In particular, when $\epsilon > 2^{-b} M / (M - b)$ that is only slightly greater than 2^{-b} , Stinson's bound becomes weaker than our AU_2 -bound.

Our AU_2 -bound and Stinson's bound therefore represent two spectrums of the asymptotic behaviour of any AU_2 [4]: when ϵ only slightly exceeds 2^{-b} the key length grows in proportion to the logarithm of message length, but if $\epsilon = 2^{-b}$ it will grow at least linearly.

References

1. http://csrc.nist.gov/groups/ST/hash/documents/FR_Notice_Nov07.pdf
2. E. Brickell, D. Pointcheval, S. Vaudenay, and M. Yung. *Design validation for discrete logarithm based signature schemes*. In PKC 2000, LNCS vol. 1751, pp. 276-292.
3. J. Carter and M. Wegman. *Universal Classes of Hash Functions*. Computer & System Sciences, 18(1979), 143-154.
4. T. Johansson, G.A. Kabatianskii and B. Smeets. *On the relation between A-Codes and Codes correcting independent errors*. Eurocrypt 1993, LNCS vol. 765, pp. 1-11.
5. A. Joux. *Multicollisions in Iterated Hash Functions*. CRYPTO 2004, LNCS vol. 3152, pp. 306-316, 2004.
6. K. Kurosawa, K. Okada, H. Saido, and D.R. Stinson. *New combinatorial bounds for authentication codes and key predistribution schemes*. Designs, Codes and Cryptography, 15 (1998), 87-100.
7. S. Laur and S. Pasini. *SAS-Based Group Authentication and Key Agreement Protocols*. Public Key Cryptography, PKC, 197-213 (2008).
8. L.H. Nguyen and A.W. Roscoe. *Authenticating ad hoc networks by comparison of short digests*. Information and Computation 206 (2008), 250-271.
9. L.H. Nguyen and A.W. Roscoe. *Authentication protocols based on low-bandwidth unspoofable channels: a comparative survey*. Journal of Computer Security (to appear).
10. R. Rivest, A. Shamir. *PayWord and MicroMint two simple micropayment schemes*. CryptoBytes, 2(1):7-11, 1996.
11. D.R. Stinson. *On the Connections Between Universal Hashing, Combinatorial Designs and Error-Correcting Codes*. Congressus Numerantium, vol. 114 (1996), 7-27.