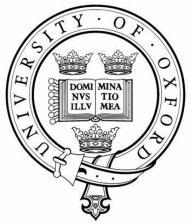
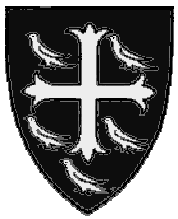
Bound for almost universal hash functions

Long Nguyen and Bill Roscoe

Oxford University Computing Laboratory
University College





ε-almost universal hash functions

- An ε-almost universal hash function is denoted ε-*AU*.

- For any pair of distinct messages (*m*,*m*'), we always have:

$$\Pr_{k \in \{1 \ldots 2^K\}} \left[Hash_k(m) = Hash_k(m') \right] \leq \varepsilon$$

- (*K*,*M*,*H*) are bitlengths of key *k*, input message *m*, and hash output *h*.

Equivalence between *AU* and error-correcting codes

- If there exists an $\mathcal{E}$-*AU* (*K*,*M*,*H*) then there exists a code whose parameters can be derived from ($\mathcal{E}$,*K*,*M*,*H*) , and vice versa.

Equivalence between *AU* and error-correcting codes

- If there exists an $\mathcal{E}$-*AU* (*K*,*M*,*H*) then there exists a code whose parameters can be derived from ($\mathcal{E}$,*K*,*M*,*H*) , and vice versa.

- Any bound in coding theory corresponds to an *AU*-bound, and vice versa.

Equivalence between *AU* and error-correcting codes

- If there exists an ε-*AU* (*K*,*M*,*H*) then there exists a code whose parameters can be derived from (ε,*K*,*M*,*H*) , and vice versa.
- Any bound in coding theory corresponds to an *AU*-bound, and vice versa.
- The Singleton bound in coding theory matches the following *AU*-bound:

$$K \geq \log_2 \left\lceil \varepsilon^{-1} \left(\frac{M}{H} - 1 \right) \right\rceil$$

The pigeon-hole principle

- 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 $\left\lceil \frac{n}{m} \right\rceil$ items.

A new bound for an $\mathcal{E}$-AU (K,M,H)

- Assume that $M = H * t + H'$

A new bound for an ε-AU (K,M,H)

- Assume that $M = H * t + H'$

- For any key k_1, there exists a hash-value $h_1 \in \{1 \ldots 2^H\}$ such that

$$\left\lceil \frac{2^M}{2^H} \right\rceil \text{ or more messages } \xrightarrow{Hash_{k_1}(\ldots)} h_1$$

A new bound for an ε-AU (K,M,H)

- Assume that $M = H * t + H'$

- For any key k_1, there exists a hash-value $h_1 \in \{1 \ldots 2^H\}$ such that

$$\left\lceil \frac{2^M}{2^H} \right\rceil \text{ or more messages} \xrightarrow{Hash_{k_1}(...)} h_1$$

- For any key $k_2 \neq k_1$, there exists a hash-value $h_2 \in \{1 \ldots 2^H\}$ such that

$$\left\lceil \frac{2^M}{2^{2H}} \right\rceil \text{ or more messages} \xrightarrow{Hash_{k_2}(...)} h_2$$

A new bound for an $\mathcal{E}$-AU (K,M,H)

- Assume that $M = H * t + H'$. If we repeat the above process t (or $t + 1$) times depending the value of H', then we derive the following:

A new bound for an ε-AU (K,M,H)

- Assume that $M = H * t + H'$. If we repeat the above process t (or $t + 1$) times depending the value of H', then we derive the following:

- When $H' = 0$ (or M is a multiple of H), our new AU-bound is same as the one derived from the Singleton bound:

$$K \geq \log_2 \left[\varepsilon^{-1} \left(\frac{M}{H} - 1 \right) \right]$$

A new bound for an ε-AU (K,M,H)

- Assume that $M = H * t + H'$. If we repeat the above process t (or $t + 1$) times depending the value of H', then we derive the following:

- When $H' = 0$ (or M is a multiple of H), our new AU-bound is same as the one derived from the Singleton bound:

$$K \geq \log_2 \left\lceil \varepsilon^{-1} \left(\frac{M}{H} - 1 \right) \right\rceil$$

- But when $H' \neq 0$ (or M is *not* a multiple of H), our new AU-bound is tighter than the one derived from the Singleton bound since we now have

$$K \geq \log_2 \left(\varepsilon^{-1} \left\lfloor \frac{M}{H} \right\rfloor \right)$$

Comparison between two *AU*-bounds

- When *M* is *not* a multiple of *H*, any $\mathcal{E}$-*AU* (*K*,*M*,*H*) can still be converted into an equivalent error-correcting code.

Comparison between two *AU*-bounds

- When M is *not* a multiple of H, any $\mathcal{E}$-*AU* (K,M,H) can still be converted into an equivalent error-correcting code.

- But the resulting code never satisfies the Singleton bound with equality.

Comparison between two *AU*-bounds

- When *M* is *not* a multiple of *H*, any $\mathcal{E}$-*AU* (*K*,*M*,*H*) can still be converted into an equivalent error-correcting code.

- But the resulting code never satisfies the Singleton bound with equality.

- Therefore the *AU*-bound derived from the Singleton bound is not tight in this case (*M* is *not* a multiple of *H*) since equality cannot be met.

Conclusions

- Our results demonstrates that the equivalence between *AU* and error-correcting codes does not always give a tight bound for an *AU*.

Conclusions

- Our results demonstrates that the equivalence between *AU* and error-correcting codes does not always give a tight bound for an *AU*.

- This therefore opens the way for re-examining the accuracy of many other bounds for almost (strongly) universal hash functions which are *indirectly* derived from various combinatorial objects such as error-correcting codes, orthogonal arrays, difference matrices, and balanced incomplete block designs.

- Because universal hash bounds derived from other combinatorial objects might only fit certain of universal hash parameters tightly.

$$\left|\left\{k \in \left\{1..2^K\right\} : h_k(m) = h_k(m')\right\}\right| \leq \varepsilon 2^K$$

$$\Pr_{k \in \left\{1...2^K\right\}}\left[h_k(m) = h_k(m')\right] \leq \varepsilon$$

$$K \geq \log_2\left[\varepsilon^{-1}\left(\frac{M}{H} - 1\right)\right]$$

$$K \geq \log_2\left(\varepsilon^{-1}\left\lfloor\frac{M}{H}\right\rfloor\right)$$

$$\left\lfloor 2^{M-H} \right\rfloor \qquad \left\lceil\frac{2^M}{2^H}\right\rceil$$

$$\left\lfloor 2^{M-2H} \right\rfloor \qquad \left\lceil\frac{n}{m}\right\rceil$$

$$\xrightarrow{h_{k_1}(...)}$$