Introduction
In this paper, the author suggests a new class of hash functions and apply it for data storage and retrieval.

- Notation
- Properties of Universal Classes
- Some Universal$_2$ Classes
- Importance
- Future Research
- Acknowledgements and References
Notation
\[ |S| \quad [x] \quad x \oplus y \quad \mathbb{Z}_n \]
Hash function $f : A \rightarrow B$ with $|A| > |B|$.

Define

$$\delta_f(x, y) := \begin{cases} 
1 & \text{if } x \neq y \text{ and } f(x) = f(y) \\
0 & \text{otherwise}
\end{cases}$$
Hash function $f : A \to B$ with $|A| > |B|$.
Define
\[
\delta_f(x, y) := \begin{cases} 
1 & \text{if } x \neq y \text{ and } f(x) = f(y) \\
0 & \text{otherwise}
\end{cases}
\]
If $H$ is a collection of hash functions, $x \in A$ and $S \subset A$ then
\[
\delta_H(x, S) := \sum_{f \in H} \sum_{y \in S} \delta_f(x, y)
\]
The order of summation does not matter.

For $S, T \subset A$ with $S \cap T = \emptyset$,
\[
\delta_f(x, S \cup T) = \delta_f(x, S) + \delta_f(x, T).
\]
Properties of Universal Classes
**Definition**

Let $H$ be a class of functions from $A$ to $B$. We say that $H$ is universal $2$ if $\forall x, y \in A, \delta_H(x, y) \leq \frac{|H|}{|B|}$.

Subscript "2" is intended to emphasize that this definition constrains the behavior of $H$ only on pairs of elements of $A$. 
Let $H$ be a class of functions from $A$ to $B$. We say that $H$ is universal$_2$ if
\[ \forall x, y \in A, \delta_H(x, y) \leq \frac{|H|}{|B|}. \]

\[ \sum_{h \in H} \delta_h(x, y) \leq \frac{1}{|B|} \]

$H$ is universal$_2$ if no pair of distinct keys collide under more than $\frac{1}{|B|}$ fraction of the functions in $H$.

Subscript “2” is intended to emphasize that this definition constrains the behavior of $H$ only on pairs of elements of $A$. 
Proposition 1 shows the bound on $\delta_H(x, y)$ in the definition of universal$_2$ is tight when $|A|$ is much larger than $|B|$.

**Proposition 1**

*Given any collection $H$ of hash functions (not necessarily universal$_2$), there exists $x, y \in A$ such that*

$$\delta_H(x, y) > \frac{|H|}{|B|} - \frac{|H|}{|A|}.$$
Proposition 1 shows the bound on $\delta_H(x, y)$ in the definition of universal_2 is tight when $|A|$ is much larger than $|B|$. 

**Proposition 1**

*Given any collection $H$ of hash functions (not necessarily universal_2), there exists $x, y \in A$ such that*

$$\delta_H(x, y) > \frac{|H|}{|B|} - \frac{|H|}{|A|}.$$ 

**Proof.**

Let $a = |A|, b = |B|$ and $f \in H$. For each $i \in B$, let $A_i = \{x \mid x \in A \land f(x) = i\}$ and $a_i = |A_i|$. 

Lin Lv (SJTU CIS Lab)
Proposition 1 shows the bound on $\delta_H(x, y)$ in the definition of universal$_2$ is tight when $|A|$ is much larger than $|B|$.

**Proposition 1**

*Given any collection $H$ of hash functions (not necessarily universal$_2$), there exists $x, y \in A$ such that*

$$\delta_H(x, y) > \frac{|H|}{|B|} - \frac{|H|}{|A|}.$$  

**Proof.**

Let $a = |A|, b = |B|$ and $f \in H$.

For each $i \in B$, let $A_i = \{x \mid x \in A \land f(x) = i\}$ and $a_i = |A_i|$.

For $i \neq j$, $A_i \cap A_j = \emptyset$ and $\bigcup_{i \in B} A_i = A$.

$\delta_f(A_i, A_j) = 0$ for $i \neq j$ and $\delta_f(A_i, A_i) = a_i(a_i - 1)$.
Proposition 1

Given any collection \( H \) of hash functions (not necessarily universal), there exists \( x, y \in A \) such that

\[
\delta_H(x, y) > \frac{|H|}{|B|} - \frac{|H|}{|A|}.
\]

Proof.

\[
\delta_f(A, A) = \delta_f\left( \bigcup_{i \in B} A_i, \bigcup_{j \in B} A_j \right) = \sum_{i \in B} \sum_{j \in B} \delta_f(A_i, A_j) = \sum_{i \in B} \delta_f(A_i, A_i)
\]

\[
= \sum_{i \in B} (a_i^2 - a_i) = \left( \sum_{i \in B} a_i^2 \right) - a \geq b \cdot \left( \frac{a}{b} \right)^2 - a = a^2 \left( \frac{1}{b} - \frac{1}{a} \right)
\]

So we have

\[
\delta_H(A, A) \geq a^2 |H| \left( \frac{1}{b} - \frac{1}{a} \right)
\]
Proposition 1

Given any collection $H$ of hash functions (not necessarily universal), there exists $x, y \in A$ such that

$$\delta_H(x, y) > \frac{|H|}{|B|} - \frac{|H|}{|A|}.$$ 

Proof.

Since $\delta_H(A, A) = \sum_{x \in A} \sum_{y \in A} \delta_H(x, y) = \sum_{x \in A} \sum_{y \in A \setminus \{y \neq x\}} \delta_H(x, y)$ is a summation of $a^2 - a$ items. The average value of these items is

$$\frac{\delta_H(A, A)}{a^2 - a} \geq \frac{a^2}{a^2 - a} |H| \left(\frac{1}{b} - \frac{1}{a}\right) > |H| \left(\frac{1}{b} - \frac{1}{a}\right).$$

Then there must exists one of them which is above average, i.e.,

$$\exists x, y \in A, x \neq y, \delta_H(x, y) > |H| \left(\frac{1}{b} - \frac{1}{a}\right).$$
Proposition 2

Let \( x \in A, S \subset A, H \) is a universal_2 class of functions and \( f \leftarrow H \). Then

\[
E[\delta_f(x, S)] \leq \frac{|S|}{|B|}.
\]
Proposition 2

Let \( x \in A, S \subset A \), \( H \) is a universal\(_2\) class of functions and \( f \leftarrow \$ H \).
Then

\[
E[\delta_f(x, S)] \leq \frac{|S|}{|B|}.
\]

Proof.

\[
E[\delta_f(x, S)]
= \frac{1}{|H|} \sum_{h \in H} \delta_h(x, S)
= \frac{1}{|H|} \sum_{y \in S} \delta_H(x, y)
\leq \frac{1}{|H|} \sum_{y \in S} |H| |B|
= \frac{|S|}{|B|}.
\]

(by def. of universal\(_2\))
A simple application of hash functions is to implement an associative memory.

Briefly an associative memory can perform three operations:

- **Store(Key, Data)** which stores “Data” under the identifier “Key” and overwrite the old data.
- **Retrieve(Key)** which returns the data associated with “Key” or “⊥” if there is no data.
- **Delete(Key)** which deletes the associated data.
A simple application of hash functions is to implement an associative memory.

Briefly an associative memory can perform three operations:

- **Store(Key, Data)** which stores “Data” under the identifier “Key” and overwrite the old data.
- **Retrieve(Key)** which returns the data associated with “Key” or “⊥” if there is no data.
- **Delete(Key)** which deletes the associated data.

Key space may be large. To reduce the space used, we apply a hash function $f$ to the key and store the data in the corresponding linked list (or balanced tree).

The linked list is searched linearly to determine if the key has been previously stored.
The time required to perform an operation involving the key $x$ is bounded by some linear function of the length of the linked list indexed by $f(x)$. 
The time required to perform an operation involving the key $x$ is bounded by some linear function of the length of the linked list indexed by $f(x)$.

If $S$ is the set of keys which have been used in operation “Store”, then the length of the linked list indexed by $f(x)$ will be $1 + \delta_f(x, S)$.

We ignore the constant factor and define the cost of an operation to the key $x$ to be $1 + \delta_f(x, S)$, where $S$ is the set of previously inserted keys.

If $R$ represents a set of operations with orders and we use $C(f, R)$ to represent the summation of the costs of the individual requests in $R$. 
**Proposition 3**

Let $R$ be any sequence of $r$ operations which includes $k$ insertions. Suppose $H$ is a universal $2$ class of hash functions and $f \leftarrow H$, then

$$\mathbb{E}[C(f, R)] \leq r \left(1 + \frac{k}{|B|}\right).$$

*Proof.* Proposition 2 assures that the expected cost for an individual operation is no greater than $1 + \frac{k}{|B|}$. □

*Remark.* In the actual application, we estimate the value of $k$ and choose $B$ so that $k|B|$ is approximately 1. So the expected cost is linear in the number of operations.
Proposition 3

Let \( R \) be any sequence of \( r \) operations which includes \( k \) insertions. Suppose \( H \) is a universal \( 2 \) class of hash functions and \( f \leftarrow H \), then

\[
E\left[C(f, R)\right] \leq r \left(1 + \frac{k}{|B|}\right).
\]

Proof.
Proposition 2 assures that the expected cost for an individual operation is no greater than \( 1 + \frac{k}{|B|} \). \( \square \)
Proposition 3

$R$ be any sequence of $r$ operations which includes $k$ insertions.
Suppose $H$ is a universal_2 class of hash functions and $f \xleftarrow{} H$, then

$$
\mathbb{E}[C(f, R)] \leq r \left(1 + \frac{k}{|B|}\right).
$$

Proof.
Proposition 2 assures that the expected cost for an individual operation
is no greater than $1 + \frac{k}{|B|}$.

Remark.
In the actual application, we estimate the value of $k$ and choose $B$ so
that $\frac{k}{|B|}$ is approximately 1. So the expected cost is linear in the
number of operations for any sequence of operations.
Only average value is not enough for some applications of hash functions. Proposition 4 gives a loose bound for the probability that a cost is intolerable.

**Proposition 4**

Let \( x \in A \), \( S \subset A \), \( t > 1 \), \( H \) is a universal class of functions and \( f \leftarrow H \).

Let \( \mu = E[\delta_f(x, S)] \), then

\[
\Pr[\delta_f(x, S) > t\mu] < \frac{1}{t}.
\]

**Proof.** For each function with \( \delta_f(x, S) > t\mu \), there must be more than \( t - 1 \) functions with \( \delta_f(x, S) < \mu \) to make the mean down to \( \mu \). □

**Remark.** The result is the same for \( C(f, R) \).
Only average value is not enough for some applications of hash functions. Proposition 4 gives a loose bound for the probability that a cost is intolerable.

**Proposition 4**

Let $x \in A$, $S \subset A$, $t > 1$, $H$ is a universal class of functions and $f \leftarrow H$. Let $\mu = \mathbb{E}[\delta_f(x, S)]$, then

$$\Pr[\delta_f(x, S) > t\mu] < \frac{1}{t}.$$  

**Proof.**

For each function with $\delta_f(x, S) > t\mu$, there must be more than $t - 1$ functions with $\delta_f(x, S) < \mu$ to make the mean down to $\mu$.  

□
Only average value is not enough for some applications of hash functions. Proposition 4 gives a loose bound for the probability that a cost is intolerable.

**Proposition 4**

Let \( x \in A, \ S \subset A, \ t > 1, \ \text{H is a universal class of functions and} \ f \leftarrow H. \) Let \( \mu = \mathbb{E}[\delta_f(x, S)] \), then

\[
\Pr [\delta_f(x, S) > t\mu] < \frac{1}{t}.
\]

**Proof.**

For each function with \( \delta_f(x, S) > t\mu \), there must be more than \( t - 1 \) functions with \( \delta_f(x, S) < \mu \) to make the mean down to \( \mu \).

**Remark.**

The result is the same for \( C(f, R) \).
We conclude this section by showing that our result does not entail a poorer expected performance than the tradition approach.

**Proposition 5**

Given any single hash function, let $E_1$ be the expected cost with respect to that function of a randomly operation after $k$ random insertions have been made. Let $E_2$ be the expected cost (averaging over a universal class of hash functions) of any request after any $k$ insertions have been made. Then

$$E_1 \geq (1 - \varepsilon)E_2,$$

where $\varepsilon = \frac{|B|}{|A|}$. 


We conclude this section by showing that our result does not entail a poorer expected performance than the tradition approach.

**Proposition 5**

Given any single hash function, let $E_1$ be the expected cost with respect to that function of a randomly operation after $k$ random insertions have been made. Let $E_2$ be the expected cost (averaging over a universal class of hash functions) of any request after any $k$ insertions have been made. Then

$$E_1 \geq (1 - \varepsilon)E_2,$$

where $\varepsilon = \frac{|B|}{|A|}$.

**Proof.**

Let $a = |A|$, $b = |B|$ and $S$ be the set of elements of $A$ which were inserted beforehand. Proposition 2 implies $E_2 \leq 1 + \frac{|S|}{b}$. 
Proposition 5

Given any single hash function, let $E_1$ be the expected cost with respect to that function of a randomly operation after $k$ random insertions have been made. Let $E_2$ be the expected cost (averaging over a universal class of hash functions) of any request after any $k$ insertions have been made. Then

$$E_1 \geq (1 - \varepsilon) E_2,$$

where $\varepsilon = \frac{|B|}{|A|}$.

Proof.

$$
\left(1 - \frac{b}{a}\right) E_2 \leq \left(1 - \frac{b}{a}\right) \left(1 + \frac{|S|}{b}\right) = 1 + |S| \left(\frac{1}{b} - \frac{1}{a}\right) - \frac{b}{a}
$$

$$
< 1 + |S| \left(\frac{1}{b} - \frac{1}{a}\right) \leq 1 + k \left(\frac{1}{b} - \frac{1}{a}\right).
$$

Then we show $1 + k \left(\frac{1}{b} - \frac{1}{a}\right) \leq E_1$. 
**Proposition 5**

*Given any single hash function, let $E_1$ be the expected cost with respect to that function of a randomly operation after $k$ random insertions have been made. Let $E_2$ be the expected cost (averaging over a universal class of hash functions) of any request after any $k$ insertions have been made. Then*

$$E_1 \geq (1 - \varepsilon)E_2,$$

*where $\varepsilon = \frac{|B|}{|A|}$.*

**Proof.**

In the proof of Proposition 1, it was shown that $\delta_f(A, A) \geq a^2 \left( \frac{1}{b} - \frac{1}{a} \right)$ for any hash function. Thus if we fix $f$ and let $x \xleftarrow{} A$, $y \xleftarrow{} A$, then

$$E[\delta_f(x, y)] = \frac{1}{a^2} \delta_f(A, A) \geq \frac{1}{b} - \frac{1}{a}.$$

Let $R$ be the set of insertions.

$$E_1 = E\left[1 + \delta_f(x, R)\right] = 1 + E\left[\delta_f(x, R)\right] = 1 + \sum E[\delta_f(x, y)] \geq 1 + k \left( \frac{1}{b} - \frac{1}{a} \right).$$
Some Universal$_2$ Classes
We present the first universal hash functions, $H_1$.

$A = \{0, 1, \cdots, a - 1\}$, $B = \{0, 1, \cdots, b - 1\}$ and $p$ be a prime with $p \geq a$. $g$ be any function from $\mathbb{Z}_p$ to $B$ which

$$|\{y \in \mathbb{Z}_p | g(y) = i\}| \leq \left\lceil \frac{p}{b} \right\rceil$$

for all $i \in B$.

A natural choice for $g$ is the modulo $b$ function. ($g(x) = x \mod b$)
We present the first universal hash functions, $H_1$.

$A = \{0, 1, \cdots, a - 1\}$, $B = \{0, 1, \cdots, b - 1\}$ and $p$ be a prime with $p \geq a$. $g$ be any function from $\mathbb{Z}_p$ to $B$ which

$$\left|\{y \in \mathbb{Z}_p \mid g(y) = i\}\right| \leq \left\lceil \frac{p}{b} \right\rceil \text{ for all } i \in B.$$ 

A natural choice for $g$ is the modulo $b$ function. ($g(x) = x \mod b$)

Let $m, n \in \mathbb{Z}_p$ with $m \neq 0$. Define $h_{m,n}(x) := (mx + n) \mod p$ and define $f_{m,n}(x) := g(h_{m,n}(x))$

Then define $H_1$

$$H_1 := \{ f_{m,n} \mid m, n \in \mathbb{Z}_p \text{ and } m \neq 0 \}$$

$$|H_1| = p(p - 1).$$
Let $m, n \in \mathbb{Z}_p$ with $m \neq 0$.
Define $h_{m,n}(x) := (mx + n) \mod p$ and define $f_{m,n}(x) := g(h_{m,n}(x))$
Then define $H_1$

$$H_1 := \{ f_{m,n} \mid m, n \in \mathbb{Z}_p \text{ and } m \neq 0 \}$$

Lemma 6

$\forall x, y \in A \text{ with } x \neq y, \delta_{H_1}(x, y) = \delta_g(\mathbb{Z}_p, \mathbb{Z}_p)$. 
Let \( m, n \in \mathbb{Z}_p \) with \( m \neq 0 \).
Define \( h_{m,n}(x) := (mx + n) \mod p \) and define \( f_{m,n}(x) := g(h_{m,n}(x)) \)
Then define \( H_1 \)

\[
H_1 := \{ f_{m,n} \mid m, n \in \mathbb{Z}_p \text{ and } m \neq 0 \}
\]

**Lemma 6**

\( \forall x, y \in A \text{ with } x \neq y, \ \delta_{H_1}(x, y) = \delta_g(\mathbb{Z}_p, \mathbb{Z}_p) \).

**Proof.**

For \( x \neq y \),

\[
f_{m,n}(x) = f_{m,n}(y) \iff g(h_{m,n}(x)) = g(h_{m,n}(y)).
\]

So \( \delta_{f_{m,n}}(x, y) = \delta_g(h_{m,n}(x), h_{m,n}(y)) \).
Let $m, n \in \mathbb{Z}_p$ with $m \neq 0$.

Define $h_{m,n}(x) := (mx + n) \mod p$ and define $f_{m,n}(x) := g(h_{m,n}(x))$

Then define $H_1$

$$H_1 := \{ f_{m,n} \mid m, n \in \mathbb{Z}_p \text{ and } m \neq 0 \}$$

**Proof.**

For fixed $x \neq y$, if we define the mapping $F(m, n) = (h_{m,n}(x), h_{m,n}(y))$ for $m, n \in \mathbb{Z}_p$ with $m \neq 0$. And let

$$S := \{ (r, s) \mid (r, s) \in (\mathbb{Z}_p \times \mathbb{Z}_p) \land r \neq s \}.$$

Then $F$ is a one to one mapping from $\mathbb{Z}_p^* \times \mathbb{Z}_p$ to $S$. ($|S| = p(p - 1)$)
Let $m, n \in \mathbb{Z}_p$ with $m \neq 0$. Define $h_{m,n}(x) := (mx + n) \mod p$ and define $f_{m,n}(x) := g(h_{m,n}(x))$

Then define $H_1$

\[ H_1 := \{ f_{m,n} \mid m, n \in \mathbb{Z}_p \text{ and } m \neq 0 \} \]

**Proof.**

For fixed $x \neq y$, if we define the mapping $F(m, n) = (h_{m,n}(x), h_{m,n}(y))$ for $m, n \in \mathbb{Z}_p$ with $m \neq 0$. And let

\[ S := \{(r, s) \mid (r, s) \in (\mathbb{Z}_p \times \mathbb{Z}_p) \land r \neq s \}. \]

Then $F$ is a one to one mapping from $\mathbb{Z}_p^* \times \mathbb{Z}_p$ to $S$. ($|S| = p(p - 1)$)

So

\[
\delta_{H_1}(x, y) = \sum_{(m, n) \in \mathbb{Z}_p^* \times \mathbb{Z}_p} \delta_{f_{m,n}}(x, y) = \sum_{(m, n) \in \mathbb{Z}_p^* \times \mathbb{Z}_p} \delta_g(h_{m,n}(x), h_{m,n}(y)) = \sum_{(r, s) \in S} \delta_g(r, s)
\]

\[
= \sum_{(r, s) \in \mathbb{Z}_p \times \mathbb{Z}_p} \delta_g(r, s) = \delta_g(\mathbb{Z}_p, \mathbb{Z}_p). \]

□
**Proposition 7**

\( H_1 \) is universal \( _2 \).

**Definition**

Let \( H \) be a class of functions from \( A \) to \( B \). We say that \( H \) is universal \( _2 \) if \( \forall x, y \in A, \delta_H(x, y) \leq \frac{|H|}{|B|} \).
**Proposition 7**

$H_1$ is universal$_2$.

**Definition**

Let $H$ be a class of functions from $A$ to $B$. We say that $H$ is universal$_2$ if $\forall x, y \in A, \delta_H(x, y) \leq \frac{|H|}{|B|}$.

**Proof.**

Let $n_i = |\{t \in \mathbb{Z}_p \mid g(t) = i\}|$ and $g$ was chosen because $n_i \leq \left\lceil \frac{p}{b} \right\rceil$ for all $i \in B$.

Then

$$n_i \leq \left\lceil \frac{p}{b} \right\rceil \leq \frac{p - 1}{b} + 1$$

Then

$$\delta_{H_1}(x, y) = \delta_g(\mathbb{Z}_p, \mathbb{Z}_p) = \sum_{r \in \mathbb{Z}_p} \delta_g(r, \mathbb{Z}_p) = \sum_{r \in \mathbb{Z}_p} (n_{g(r)} - 1) \leq \frac{p(p - 1)}{b} = \frac{|H_1|}{|B|}.$$
Proposition 7

$H_1$ is universal$_2$.

Remark.
Choose $b$ as a power of 2 and $g$ as the modular function so the mod $b$ operation is fast.
Proposition 7

$H_1$ is universal.

Remark.
Choose $b$ as a power of 2 and $g$ as the modular function so the mod $b$ operation is fast.

Choose $p$ to make the mod $p$ operation fast. For example, choose $p = 2^j - 1$ for some $j$. Then the intermediate value in the calculation of $f$ is at most $2j$ bits long. Assume $mx + n = 2^j t_1 + t_2$ for some $t_1, t_2 < 2^j$. Then $mx + n \equiv t_1 + t_2 \mod p$. 
**Proposition 7**

$H_1$ is universal$_2$. 

**Remark.**
Choose $b$ as a power of 2 and $g$ as the modular function so the mod $b$ operation is fast.

Choose $p$ to make the mod $p$ operation fast. For example, choose $p = 2^j - 1$ for some $j$. Then the intermediate value in the calculation of $f$ is at most $2^j$ bits long. Assume $mx + n = 2^jt_1 + t_2$ for some $t_1, t_2 < 2^j$. Then $mx + n \equiv t_1 + t_2 \mod p$.

Suppose we drop $n$ from the definition, i.e., define $h_m(x) := mx \mod p$, $f_m(x) := g(h_m(x))$ and $H := \{ f_m \mid m \in \mathbb{Z}_p \land m \neq 0 \}$, then

$$\delta_H(x, y) \leq 2 \cdot \frac{|H|}{|B|}.$$
Proposition 7

\( H_1 \) is universal\(_2\).

Remark.
Choose \( b \) as a power of 2 and \( g \) as the modular function so the mod \( b \) operation is fast.

Choose \( p \) to make the mod \( p \) operation fast. For example, choose \( p = 2^j - 1 \) for some \( j \). Then the intermediate value in the calculation of \( f \) is at most \( 2^j \) bits long. Assume \( mx + n = 2^jt_1 + t_2 \) for some \( t_1, t_2 < 2^j \). Then \( mx + n \equiv t_1 + t_2 \mod p \).

Suppose we drop \( n \) from the definition, i.e., define \( h_m(x) := mx \mod p \), \( f_m(x) := g(h_m(x)) \) and \( H := \{ f_m \mid m \in \mathbb{Z}_p \land m \neq 0 \} \), then

\[
\delta_H(x, y) \leq 2 \cdot \frac{|H|}{|B|}.
\]

\( H_1 \) may not be convenient when the keys are too long to be multiplied using a single machine instruction.
Proposition 8 gives a method for extending a class of functions for long keys.

**Proposition 8**

Suppose \( B = \{0, 1, \cdots, b - 1\} \) where \( b \) is a power of 2 and \( H \) is a class of functions from \( A \) to \( B \) with the property that for some real number \( r \), for each \( x, y \in A \) with \( x \neq y \), and for each \( i \in B \), \(|\{ f \in H \mid f(x) \oplus f(y) = i\}| \leq r|H|\).

Define the class \( J \) of hash functions from \( A \times A \) to \( B \) as follows:

For \( f, g \in H \), define \( h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2) \), let \( J := \{ h_{f,g} \mid f, g \in H\} \).

Then for all \( x, y \in A \times A \) with \( x \neq y \), and for all \( i \in B \),

\(|\{ h \in J \mid h(x) \oplus h(y) = i\}| \leq r|J|\).
Proposition 8 gives a method for extending a class of functions for long keys.

**Proposition 8**

Suppose $B = \{0, 1, \cdots, b - 1\}$ where $b$ is a power of 2 and $H$ is a class of functions from $A$ to $B$ with the property that for some real number $r$, for each $x, y \in A$ with $x \neq y$, and for each $i \in B$, $|\{ f \in H \mid f(x) \oplus f(y) = i \}| \leq r|H|$. Define the class $J$ of hash functions from $A \times A$ to $B$ as follows:

For $f, g \in H$, define $h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2)$, let $J := \{ h_{f,g} \mid f, g \in H \}$. Then for all $x, y \in A \times A$ with $x \neq y$, and for all $i \in B$, $|\{ h \in J \mid h(x) \oplus h(y) = i \}| \leq r|J|$. 

**Proof.**

Given $x, y \in A \times A$ with $x \neq y$, assume $x = (x_1, x_2), y = (y_1, y_2)$ and wlog. assume $x_1 \neq y_1$. Given $i \in B$:

$$|\{ h \in J \mid h(x) \oplus h(y) = i \}| = |\{ f, g \in H \mid f(x_1) \oplus g(x_2) \oplus f(y_1) \oplus g(y_2) = i \}|$$

$$= \sum_{g \in H} |\{ f \in H \mid f(x_1) \oplus f(y_1) = i \oplus g(x_2) \oplus g(y_2) \}|$$

$$\leq |H| \cdot r|H| = r|J|. \qed$$
**Proposition 8**

Suppose $B = \{0, 1, \cdots, b - 1\}$ where $b$ is a power of 2 and $H$ is a class of functions from $A$ to $B$ with the property that for some real number $r$, for each $x, y \in A$ with $x \neq y$, and for each $i \in B$, $|\{f \in H \mid f(x) \oplus f(y) = i\}| \leq r|H|$. Define the class $J$ of hash functions from $A \times A$ to $B$ as follows:

For $f, g \in H$, define $h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2)$, let $J := \{h_{f,g} \mid f, g \in H\}$. Then for all $x, y \in A \times A$ with $x \neq y$, and for all $i \in B$, $|\{h \in J \mid h(x) \oplus h(y) = i\}| \leq r|J|$.

**Remark.**

Proposition 8 can be used to get universal$_2$ classes which work on long keys.
Proposition 8

Suppose $B = \{0, 1, \cdots, b - 1\}$ where $b$ is a power of 2 and $H$ is a class of functions from $A$ to $B$ with the property that for some real number $r$, for each $x, y \in A$ with $x \neq y$, and for each $i \in B$, $|\{ f \in H \mid f(x) \oplus f(y) = i \}| \leq r|H|$. Define the class $J$ of hash functions from $A \times A$ to $B$ as follows:

For $f, g \in H$, define $h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2)$, let $J := \{ h_{f,g} \mid f, g \in H \}$. Then for all $x, y \in A \times A$ with $x \neq y$, and for all $i \in B$, $|\{ h \in J \mid h(x) \oplus h(y) = i \}| \leq r|J|$.

Remark.

Proposition 8 can be used to get universal$_2$ classes which work on long keys.

Suppose $H$ is a class of functions which can be applied to keys of length $\alpha$ and satisfying the above condition with $r = 1/|B|$. ($H$ is universal$_2$)

Then the resulting $J$ is a class of functions with key length $2\alpha$. 

Lin Lv (SJTU CIS Lab)
Proposition 8

Suppose $B = \{0, 1, \ldots, b - 1\}$ where $b$ is a power of 2 and $H$ is a class of functions from $A$ to $B$ with the property that for some real number $r$, for each $x, y \in A$ with $x \neq y$, and for each $i \in B$, $|\{ f \in H \mid f(x) \oplus f(y) = i\}| \leq r|H|$. Define the class $J$ of hash functions from $A \times A$ to $B$ as follows:

For $f, g \in H$, define $h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2)$, let $J := \{ h_{f,g} \mid f, g \in H \}$.

Then for all $x, y \in A \times A$ with $x \neq y$, and for all $i \in B$,

$|\{ h \in J \mid h(x) \oplus h(y) = i\}| \leq r|J|$.

Remark.

Proposition 8 can be used to get universal$_2$ classes which work on long keys.

Suppose $H$ is a class of functions which can be applied to keys of length $\alpha$ and satisfying the above condition with $r = 1/|B|$. ($H$ is universal$_2$)

Then the resulting $J$ is a class of functions with key length $2^\alpha$.

Furthermore, $J$ is universal$_2$. $(r|J| \geq |\{ h \in J \mid h(x) \oplus h(y) = 0\}| = \delta_J(x, y))$
Proposition 8

Suppose $B = \{0, 1, \cdots, b - 1\}$ where $b$ is a power of 2 and $H$ is a class of functions from $A$ to $B$ with the property that for some real number $r$, for each $x, y \in A$ with $x \neq y$, and for each $i \in B$, $|\{ f \in H \mid f(x) \oplus f(y) = i\}| \leq r|H|$. Define the class $J$ of hash functions from $A \times A$ to $B$ as follows:

For $f, g \in H$, define $h_{f,g}(x_1, x_2) := f(x_1) \oplus g(x_2)$, let $J := \{ h_{f,g} \mid f, g \in H\}$.

Then for all $x, y \in A \times A$ with $x \neq y$, and for all $i \in B$,

$|\{ h \in J \mid h(x) \oplus h(y) = i\}| \leq r|J|.$

Remark.
Proposition 8 can be used to get universal_2 classes which work on long keys.

Suppose $H$ is a class of functions which can be applied to keys of length $\alpha$ and satisfying the above condition with $r = 1/|B|$. ($H$ is universal_2)

Then the resulting $J$ is a class of functions with key length $2\alpha$.

Furthermore, $J$ is universal_2. $(r|J| \geq |\{ h \in J \mid h(x) \oplus h(y) = 0\}| = \delta_J(x, y))$

Repeat using this proposition to get functions for arbitrarily long keys.
The following universal class of functions, denoted $H_3$ for historical reasons, does not require multiplication and may be better for many applications.

Let

- $A = \{0, 1\}^i$ and $B = \{0, 1\}^j$.
- $M$ be the set of $i$ by $j$ Boolean matrix.
- For $m \in M$, let $m(k)$ be $k$th row of $m$.
- $x = x_1 x_2 \cdots x_i$ for $x \in A$.
- Define $f_m(x) := x_m = x_1 m(1) \oplus x_2 m(2) \oplus \cdots \oplus x_i m(i)$.
- $H_3 := \{ f_m \mid m \in M \}$. 
Proposition 9

$H_3$ is universal$_2$. 

Proof.
Induction on $i$ using Proposition 8. When $i = 1$, we have $A = \{0, 1\}$, $M = B$, and for $m \in B, f_m(0) = 0$ and $f_m(1) = m$. The condition of Proposition 8 is satisfied with $r = 1/|H|$ since the only possible choice for $x \neq y$ are $x = 0, y = 1$ (or $x = 1, y = 0$), and for each $m$, $f_m$ is the only function for which $f_m(0) \oplus f_m(1) = m$.

Proposition 8 supplies the induction step and then $H_3$ is universal$_2$. □
Proposition 9

\( H_3 \) is universal_2.

Proof.

Induction on \( i \) using Proposition 8. When \( i = 1 \), we have
\[ A = \{0, 1\}, \quad M = B, \quad \text{and for } m \in B, \quad f_m(0) = 0 \quad \text{and} \quad f_m(1) = m. \]
The condition of Proposition 8 is satisfied with \( r = 1/|H| \) since the only possible choice for \( x \neq y \) are \( x = 0, y = 1 \) (or \( x = 1, y = 0 \)), and for each \( m \), \( f_m \) is the only function for which \( f_m(0) \oplus f_m(1) = m. \)
Proposition 8 supplies the induction step and then \( H_3 \) is universal_2.

\[ \square \]
This paper also give an example $H_2$ which is similar to $H_3$ but requires less time and more space. The analysis for $H_2$ can be found in [2].

We omit this part.
Importance
We summarize the results proved in this paper in the next two theorems.

**Theorem 10**

*Using a standard model of computation, where multiplication, choosing of random numbers, and memory references take unit time, any sequence of $r$ operations to an associative memory can be processed in expected time $O(r)$.*

**Proof.**

Apply Proposition 3 by choosing $|B|$ approximately equal to $r$ and using $H_1$. 

If the keys are too long, the assumption that multiplication takes unit time is unrealistic. And we show:

**Theorem 11**

*Using a standard model of computation, where Boolean operations on machine addresses, choosing of random numbers, and memory references take unit time, any sequence of operations to an associative memory can be processed in expected time linear in the number of bits in the input.*

**Proof.**

Use class $H_2$ or $H_3$. □
The practical and theoretical value of universal class of hash functions is immense. Example of its application is shown in this paper. (The analysis of [4], [5], [8] can benefit from it)
Future Research
See textbook [Shoup08] for the discussion about extensions of universal $\text{univ}_2$ hash functions.
Acknowledgements and References
Thanks