Detecting Unpreventable Collisions

Pradyumna Kannan

July 20, 2008

Suppose two keys k_1 and k_2 such that $k_2 > k_1$ collide unwantedly using the 64-bit magic hashing scheme to the same index given the guessed bits in the magic g, the unguessed bits in the magic u, and the number of bits in the index x, where $3 \leq x \leq 12$. The goal is to find out if the collision can be prevented by guessing additional bits in u. If the collision cannot be prevented, then the search can cutoff at this point. Let all arithmetic (numbers and operations) here on out be in 64-bits.

Definition For the sake of easy notation, the binary & operator has precedence over + and - and * but not over juxtaposed multiplication (eg. $k_1 g \& IB = (k_1 * g) \& IB$) for these notes. This is different from precedence rules in C.

Definition A full bitboard is defined as $U64FULL = 0xFFFFFFFFFFFFFFFF$.

Definition The index bits is defined as $IB = \sim (U64FULL >> x)$.

Definition The number corresponding to the smallest non-zero index is defined as $SI = C64(1) << (64-x)$.

Definition The difference between the input keys is defined as $d = k_2 - k_1$.

Proposition 0.1 $k_2 = k_1 + d$

Proposition 0.2 $a - a = 0$

Proposition 0.3 *If* $a = b$ *and* $c = d$, *then* $a + c = b + d$.

Proposition 0.4 *If* $a = b$ *and* $c = d$, *then* $a - c = b - d$.

Proposition 0.5 *If* $a = b$ *and* $c = d$, *then* $a * c = b * d$.

Proposition 0.6 $a * (b + c) = a * b + a * c$.

Proposition 0.7 $a = (a/b) * b + a\%b$.

Proposition 0.8 $(a + b\&IB)\&IB = a\&IB + b\&IB$

Proposition 0.9 $(a - b\&IB)\&IB = a\&IB - b\&IB$

Proposition 0.10 $(a + b\&IB)\& \sim IB = a\& \sim IB$

Proposition 0.11 $(a - b\&IB)\& \sim IB = a\& \sim IB$

Proposition 0.12 $[a * (b\&IB)]\ \& \sim IB = 0$

Proposition 0.13 $(a\&IB) * (b\&IB) = 0$

Proposition 0.14 $(a\& \sim IB + b\& \sim IB)\& \sim IB = (a + b)\& \sim IB$

Proposition 0.15 $(a\& \sim IB - b\& \sim IB)\& \sim IB = (a-b)\& \sim IB$

Proposition 0.16 $(a\& \sim IB * b\& \sim IB)\& \sim IB = (a*b)\& \sim IB$

Proposition 0.17 $a = a\&IB + a\& \sim IB$

Proposition 0.18 $0 = a\&IB$ *is true if* $a < SI$.

Proposition 0.19 $a\& \sim IB + b\& \sim IB \leq \sim IB + SI$

Proposition 0.20 *If* $[a\& \sim IB + b]\,\&IB = 0$ *then either* $b\&IB = 0$ *or* $b\&IB = IB$.

It is given that

$$[k_1 g]\,\&IB = [k_2 g]\,\&IB.$$

Try to assert that for all possible u,

$$[k_1(g+u)]\,\&IB = [k_2(g+u)]\,\&IB.$$

$$\begin{aligned}
[k_1(g+u)]\,\&IB &= [k_2(g+u)]\,\&IB \\
[k_1 g + k_1 u]\,\&IB &= [k_2 g + k_2 u]\,\&IB \\
[k_1 g + k_1 u]\,\&IB &= [k_2 g + (k_1+d)u]\,\&IB \\
[k_1 g + k_1 u]\,\&IB &= [k_2 g + k_1 u + du]\,\&IB
\end{aligned}$$

$$\begin{aligned}
0 &= [k_2 g + k_1 u + du]\,\&IB - [k_1 g + k_1 u]\,\&IB \\
0 &= [k_2 g\& \sim IB + k_1 u\& \sim IB + du\& \sim IB + k_2 g\&IB + k_1 u\&IB + du\&IB]\,\&IB \\
&\quad - [k_1 g\& \sim IB + k_1 u\& \sim IB + k_1 g\&IB + k_1 u\&IB]\,\&IB \\
0 &= [k_2 g\& \sim IB + k_1 u\& \sim IB + du\& \sim IB]\,\&IB + k_2 g\&IB + k_1 u\&IB + du\&IB \\
&\quad - [k_1 g\& \sim IB + k_1 u\& \sim IB]\,\&IB - k_1 g\&IB - k_1 u\&IB \\
0 &= [k_2 g\& \sim IB + k_1 u\& \sim IB + du\& \sim IB]\,\&IB + du\&IB \\
&\quad - [k_1 g\& \sim IB + k_1 u\& \sim IB]\,\&IB \\
0 &= [k_2 g\& \sim IB - k_1 g\& \sim IB + k_1 g\& \sim IB + k_1 u\& \sim IB + du]\,\&IB \\
&\quad - [k_1 g\& \sim IB + k_1 u\& \sim IB]\,\&IB \\
0 &= [k_2 g\& \sim IB - k_1 g\& \sim IB + (k_1 g\& \sim IB + k_1 u\& \sim IB)\& \sim IB + (k_1 g\& \sim IB + k_1 u\& \sim IB)\&IB + du\& \sim IB]\,\&IB \\
&\quad - [k_1 g\& \sim IB + k_1 u\& \sim IB]\,\&IB + du\&IB \\
0 &= [k_2 g\& \sim IB - k_1 g\& \sim IB + (k_1 g\& \sim IB + k_1 u\& \sim IB)\& \sim IB + du\& \sim IB]\,\&IB + du\&IB \\
0 &= [k_2 g\& \sim IB - k_1 g\& \sim IB + (k_1 g\& \sim IB + k_1 u\& \sim IB)\& \sim IB + du]\,\&IB \\
0 &= [k_2 g\& \sim IB - k_1 g\& \sim IB + (k_1(g+u))\& \sim IB + du]\,\&IB
\end{aligned}$$

If $du\&IB$ is non-zero, the collision may be preventable; otherwise,

$$[(k_1 g + k_1 u)\& \sim IB + dg\& \sim IB + du\& \sim IB]\,\&IB = 0$$

The above is true if the following is true:

$$\begin{aligned}
(k_1 g + k_1 u)\& \sim IB + dg\& \sim IB + du\& \sim IB &< SI \\
(k_1 g + k_1 u)\& \sim IB + du\& \sim IB &< SI - dg\& \sim IB
\end{aligned}$$