













For our  
multiple  
to form

Chunks that have one or more features (maximal values) in common are likely to be very similar, but small changes to the data are unlikely to perturb the maximal values

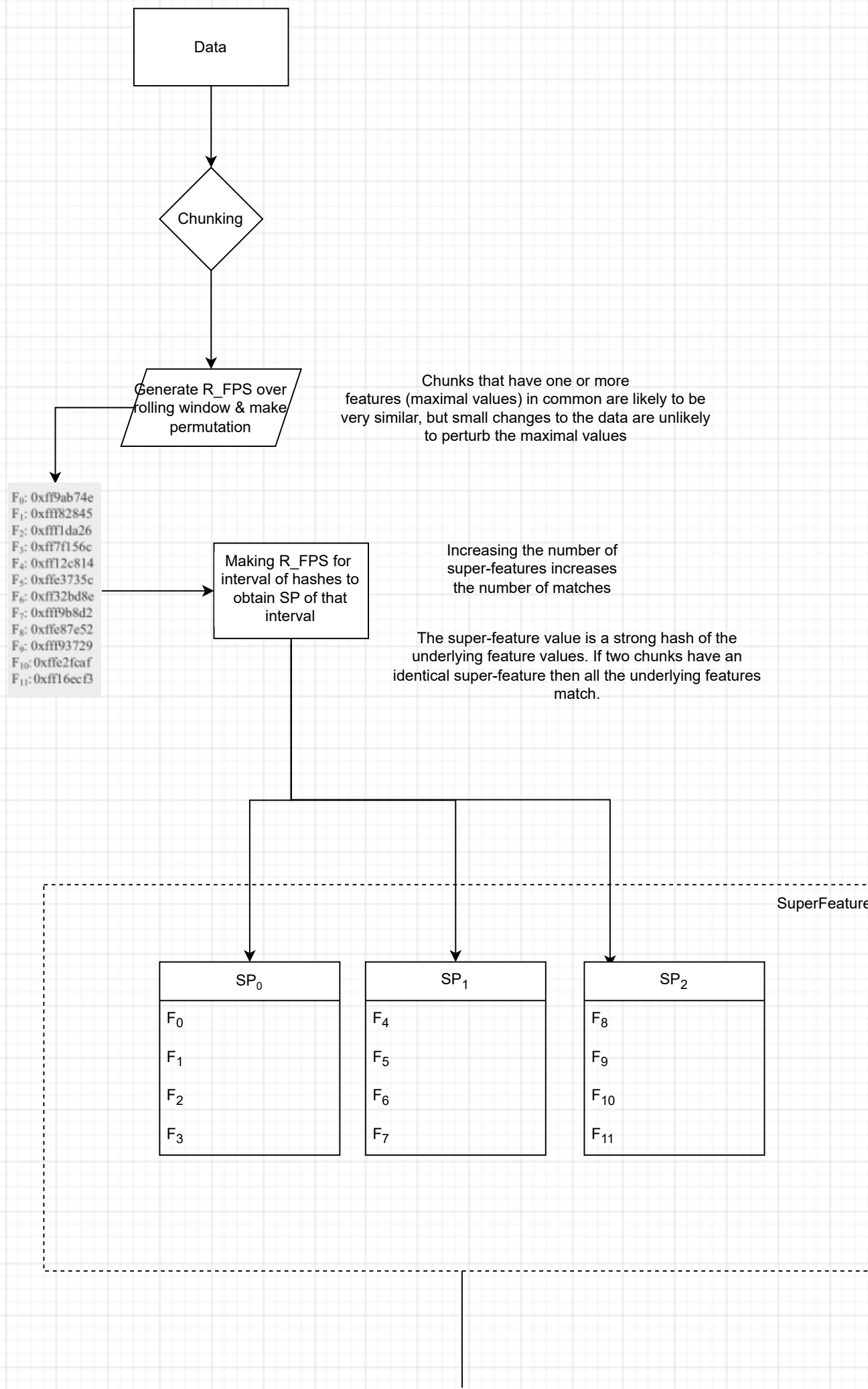
Data

Chunking

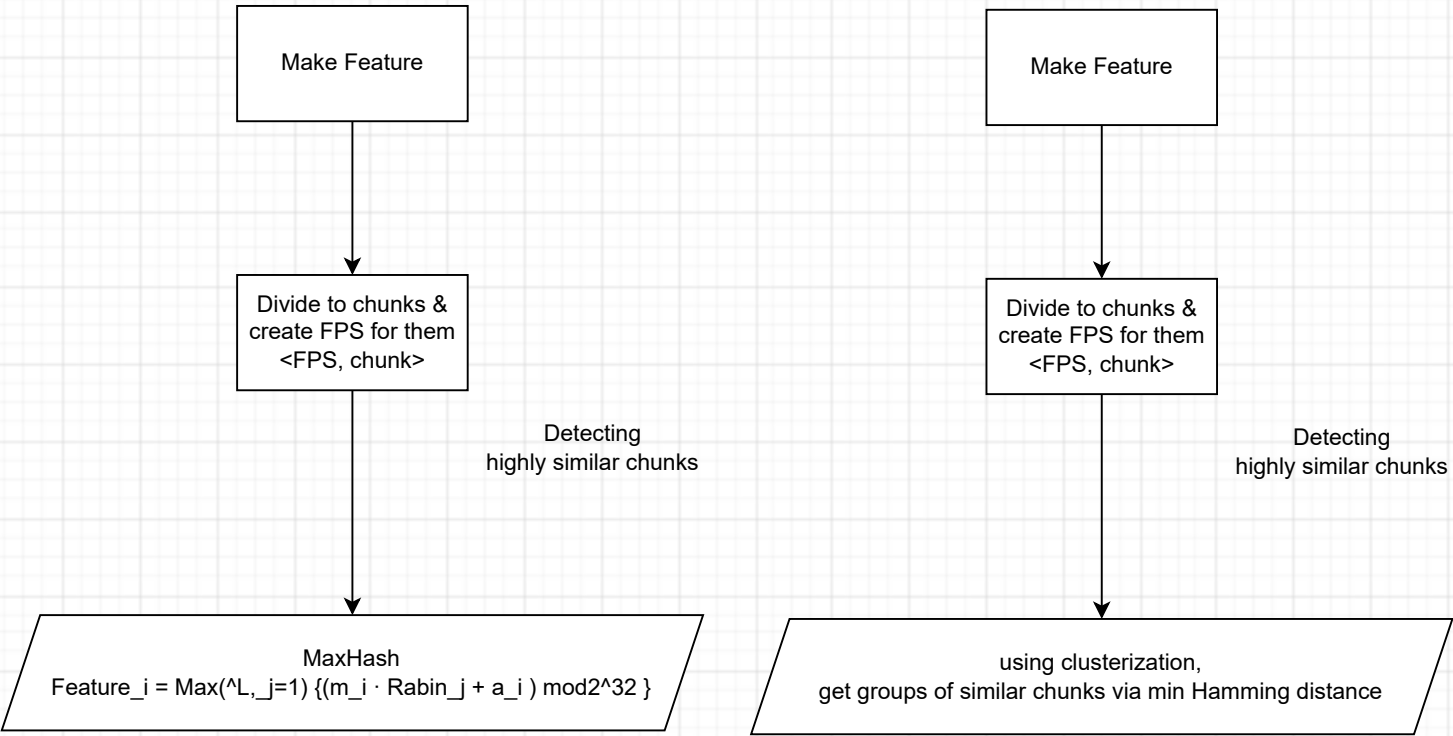
sketches we group  
e features together  
n "super-features"

Divide to chunks &  
create FPS for them  
<FPS, chunk>



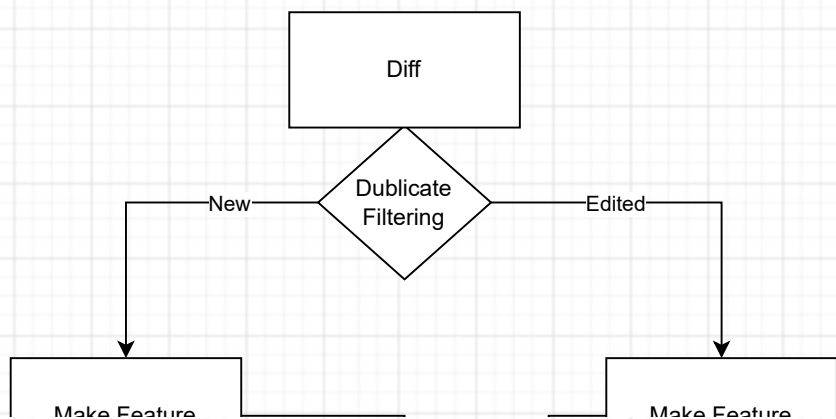
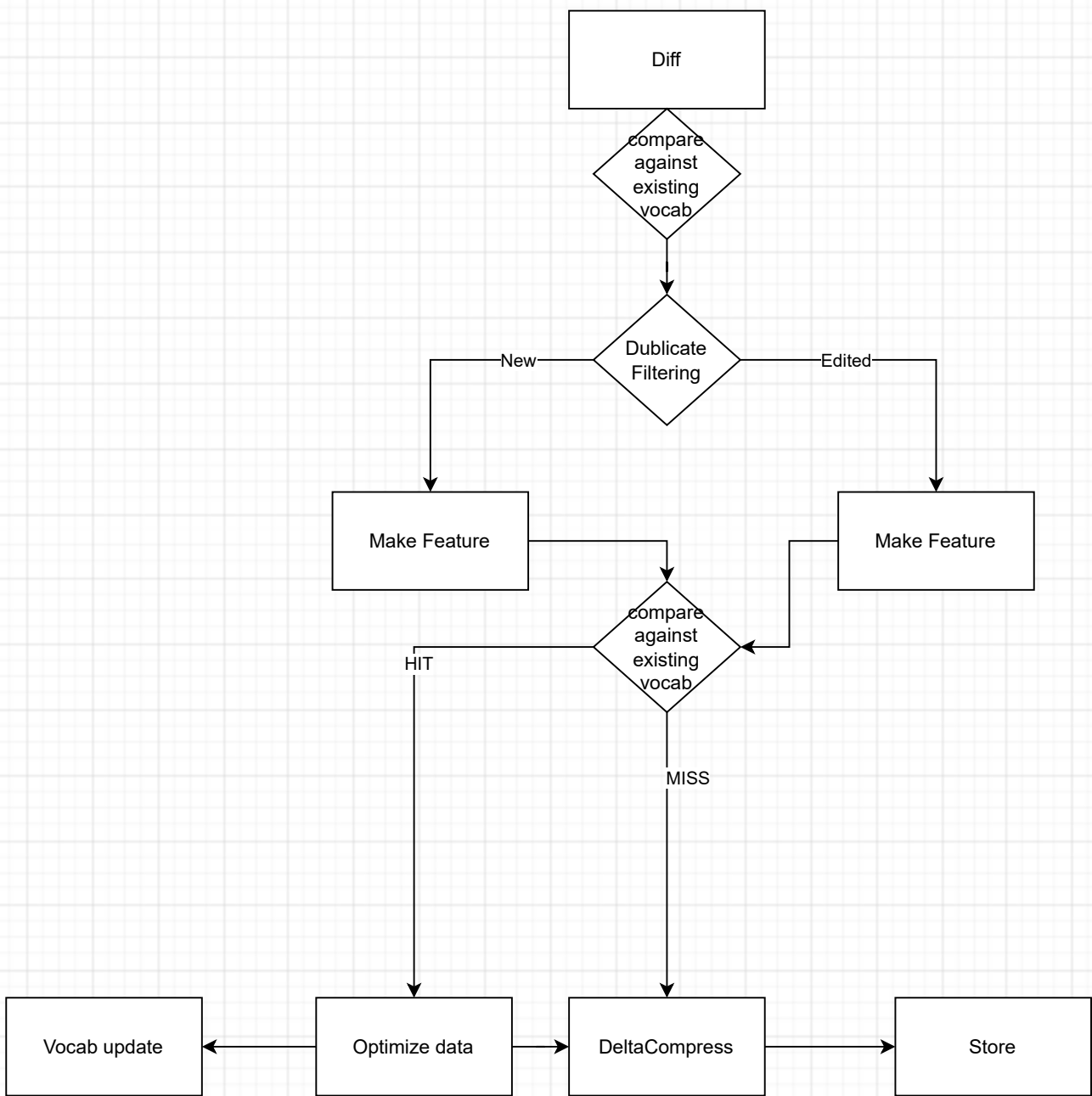


ps



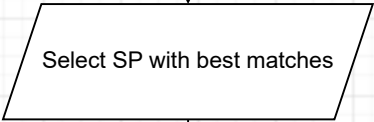
F<sub>0</sub>: 0xff9ab74e  
F<sub>1</sub>: 0xfff82845  
F<sub>2</sub>: 0xfff1da26  
F<sub>3</sub>: 0xff7f156c  
F<sub>4</sub>: 0xff12c814  
F<sub>5</sub>: 0xffe3735c  
F<sub>6</sub>: 0xff32bd8e  
F<sub>7</sub>: 0xfff9b8d2  
F<sub>8</sub>: 0xffe87e52  
F<sub>9</sub>: 0xfff93729  
F<sub>10</sub>: 0xffe2fcaf  
F<sub>11</sub>: 0xff16ecf3

F<sub>0</sub>: 0xff9ab74e  
F<sub>1</sub>: 0xfff82845  
F<sub>2</sub>: 0xfff1da26  
F<sub>3</sub>: 0xff7f156c  
F<sub>4</sub>: 0xff12c814  
F<sub>5</sub>: 0xffe3735c  
F<sub>6</sub>: 0xff32bd8e  
F<sub>7</sub>: 0xfff9b8d2  
F<sub>8</sub>: 0xffe87e52  
F<sub>9</sub>: 0xfff93729  
F<sub>10</sub>: 0xffe2fcaf  
F<sub>11</sub>: 0xff16ecf3

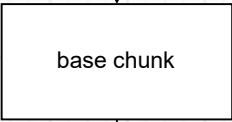




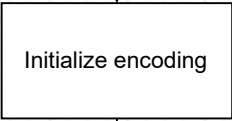




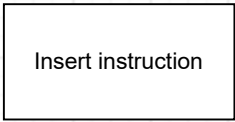
Chunks that match on more SP's are considered better matches than those that match on fewer SP => base chunk



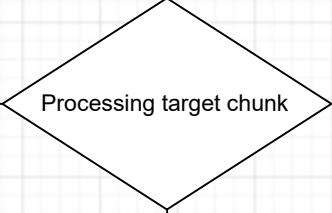
base used for delta compression



Путем перебора base chunk мы вычисляем хеш-функции в выбранных позициях, сохраняем хеш-функции и смещения в index.



Если байты не совпадают, мы выполняем инструкцию вставки, чтобы вставить байты target фрагмента в буфер вывода. Также добавляем эту область в индекс хешей

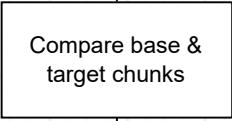


Вычисляем хеш-функции в позициях скользящего окна. Ищем соответствие хеш-функции в индексе, чтобы найти совпадение с базовым фрагментом.

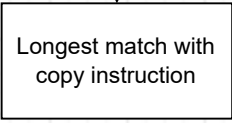
Miss

Intersection

Match



Если есть совпадение, мы сравниваем байты в base и target фрагментах вперед и назад от начальной позиции, чтобы создать максимальное возможное совпадение.



числяем  
циях,  
е в temp

циях

ий в  
ие с

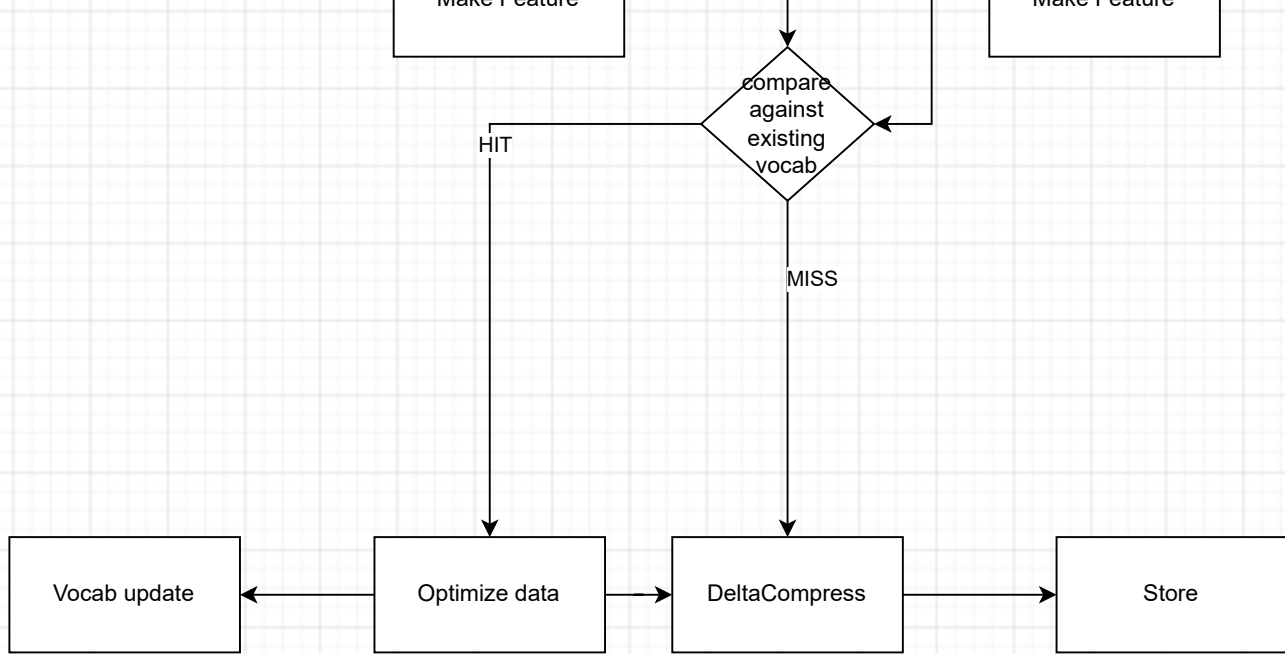
В процессе обратных  
проходов мы можем  
столкнуться с регионом,  
который был ранее  
закодирован.

Мы решаем, следует ли  
сохранить предыдущую  
инструкцию или обновить ее  
для достижения лучшей  
степени сжатия.

Save/Update  
previously encoded







$F_0: 0\text{xf}9\text{ab}74\text{e}$   
 $F_1: 0\text{xf}\text{f}82845$   
 $F_2: 0\text{xf}\text{f}1\text{da}26$   
 $F_3: 0\text{xf}\text{f}7\text{f}156\text{c}$   
 $F_4: 0\text{xf}\text{f}12\text{e}814$   
 $F_5: 0\text{xf}\text{f}3735\text{c}$   
 $F_6: 0\text{xf}\text{f}32\text{bd}8\text{e}$   
 $F_7: 0\text{xf}\text{f}9\text{b}8\text{d}2$   
 $F_8: 0\text{xf}\text{f}32\text{bd}8\text{e}$   
 $F_9: 0\text{xf}\text{f}93729$   
 $F_{10}: 0\text{xf}\text{f}2\text{fca}\text{f}$   
 $F_{11}: 0\text{xf}\text{f}16\text{ec}\text{f}3$

$F_0 < F_2 < F_1$   
 $F_4 < F_3 < F_5$   
 $F_6 < F_8 < F_7$   
 $F_{11} < F_{10} < F_9$

$SF_0: \text{hashing}\{F_1, F_5, F_7, F_9\} = 0\text{xf}794\text{de}5\text{e}$   
 $SF_1: \text{hashing}\{F_2, F_3, F_8, F_{10}\} = 0\text{x}4\text{b}6\text{f}535\text{e}$   
 $SF_2: \text{hashing}\{F_0, F_4, F_6, F_{11}\} = 0\text{x}\text{d}07267\text{d}6$

