Set Reconciliation and Error-Correcting Code, via IBLTs

**Set Reconciliation**

Alice

Bob

\[ S \subseteq \mathbb{N} \quad T \subseteq \mathbb{N} \]

**Goal:** Identify \( S \setminus T \) using communication \( O(\|S\| \log \|S\| + \|T\| \log \|T\|) \).

**Solution:** Alice inserts all elements of \( S \) into IBLT of size \( O(\|S\|) \). Sends IBLT to Bob.

Bob deletes each element of \( T \) from IBLT, then calls the IBLT lookup procedure to recover \( S \setminus T \).

**Error-Correcting Codes**

- **Goal:** Convey the message \( m \) to Bob. Assume we know an upper bound \( B \) on the number of symbols of the message that the channel will corrupt. Need to add redundancy to \( m \) so that Bob can correct errors introduced by the channel.

- **Solution:** For \( i = 1, \ldots, l \), Alice inserts \( (i, m_i) \) into an IBLT of size \( O(B) \).

- Alice sends IBLT to Bob, as well as the IBLT. Assume IBLT is uncorrupted by the channel.

- Bob recovers the corrupted message \( m' \) and deletes \( (i, m_i) \) for \( i = 1, \ldots, l \) from IBLT.

Bob then calls IBLT lookup procedure to recover the difference between \( m \) and \( m' \).

This ignores the issue that the channel might corrupt the IBLT as well. To address this, one can apply a different error-correcting code to the IBLT.
Goal: Given a turnstile stream, output a uniform random item i with \( \frac{1}{N} \), i.e., if N items have \( f_{c}(i) \), output each with probability \( \frac{1}{N} \).

Note: In insert-only streams, this problem is trivial (e.g., store item w/ smallest hit value seen).

Solution: For \( j = 1 \ldots \log N \), let \( T_j \) be an IBLT of size \( O(1) \) and let \( h_j : [N] \rightarrow [2^j] \) be a random hash function.

While processing update \( (a_i, f_i) \):
  
  For \( j = 1 \ldots \log N \):
  
  if \( h_j(a_i) = 1 \) then call insert \( (T_j, a_i, f_i) \).

At end of stream, find a \( T_j \) for which \( LST(T_j) \) succeeds and returns at least 1 item. Output a random item returned by \( T_j \).

Total space: \( O(\log^3 N) \) bits.

Claim: Let \( j = \left\lceil \log \frac{F_0}{10} \right\rceil \). Then with probability at least \( 1 - \frac{1}{10} \cdot \frac{1}{2} > 0.53 \), between 1 and 10 items have non-zero frequency satisfies \( h_j(i) = 0 \).

\[
\mathbb{E} \left[ \min_{i \in [N]} f_{c}(i) \text{ and } h_j(i) = 1 \right] = \frac{F_0}{2^j} \mathbb{E} [X_{hid}] \quad \text{so Markov } \Rightarrow \quad \Pr \left[ \min_{i \in [N]} f_{c}(i) = 1 \right] \leq \frac{1}{10}.
\]

\[
\Pr \left[ \min_{i \in [N]} f_{c}(i) = 1 \right] \leq \left(1 - \frac{1}{2^j}\right)^{F_0} \leq \frac{1}{e}.
\]