Set Reconciliation and Error-Correcting Codes via IBLTs

Set Reconciliation

Alice $S \subseteq [n]$

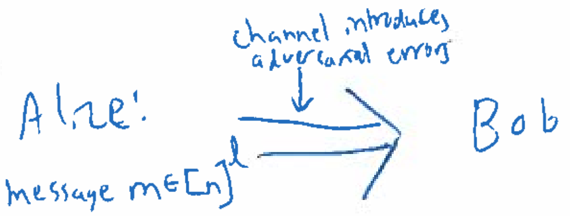
Bob $T \subseteq [n]$

Assume $|S \Delta T|$ is known

Goal: Identify $S \Delta T$ using communication $\tilde{O}(|S \Delta T|)$.

Solution:
- Alice inserts all elements of S into IBLT of size $O(|S \Delta T|)$, sends IBLT to Bob
- Bob deletes each element of T from IBLT, then calls the IBLT listing procedure to recover $S \Delta T$.

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 $i = 1, \ldots, \ell$, Alice inserts (i, m_i) into an IBLT of size $O(B)$
- Alice sends m to Bob, as well as the IBLT. Assume IBLT is uncorrupted by the channel.
- Bob receives a corrupted message m' and deletes (i, m'_i) for $i = 1, \ldots, \ell$ from IBLT. Bob then calls listing 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.

L_0-Sampling:

Goal: Given a turnstile stream, output a uniform random item i with $f_i \neq 0$. I.e. if N items have $f_i > 0$, output each with probability $\frac{1}{N}$.

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

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

```
While processing update $(a_i, \delta_i)$:
    For $j = 1 \dots \log n$:
        if $h_j(a_i) = 1$ then call insert$(T_j, a_i, \delta_i)$.
```

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

Total space: $O(\log^2 n)$ bits.

Claim: Let $j = \lfloor \log F_0 \rceil$. Then with probability at least $1 - \frac{1}{10} - \frac{1}{e} > .53$,
between 1 and 10 items with non-zero frequency satisfy $h_j(i) = 0$.

$\mathbb{E}[|i : f_i > 0 \text{ and } h_j(i) = 1|] = \frac{F_0}{2^j} \in [1, 2)$ so Markov $\Rightarrow$ $\Pr[|i : f_i > 0 \text{ and } h_j(i) = 1| >$
$\leq \frac{1}{10}$

$\Pr[|i : f_i > 0 \text{ and } h_j(i) = 1| = 0] = \left(1 - \frac{1}{2^j}\right)^{F_0} \leq \frac{1}{e}$ $\blacksquare$