The Fortune Teller's Attack

In today's lecture for *Shared Public Ledgers*, Professor Micali described a novel Byzantine agreement protocol in an idealized model achieving soundness $\sigma = 1 - 2^{-R}$ in R rounds. The idealized model supposed that in each round, a uniformly random "coin from the sky" is revealed to all parties at some specific step, and no earlier.

The question of how the protocol breaks if the coins are known in advance was raised and left as an exercise. We describe one such attack. In the modified protocol achieving perfect soundness, this attack can be adapted to prevent the protocol from terminating.

The Protocol We consider the Byzantine agreement problem with $n \geq 3t + 1$ parties of which at most t may be malicious in the setting of authenticated communication channels between each pair of parties. We additional assume that, at the appropriate moment in each round r, a uniformly random coin $c[r]$ is revealed to all parties. Each party i receives as input a bit $\mathsf{in}_i \in \{0, 1\}$ and outputs a bit $\mathsf{out}_i \in \{0, 1\}$. The protocol specifies that each party i executes the following program:

1. Let $b_i[0] := \mathsf{in}_i$.
2. For each round $r \in \{1, \ldots, R\}$, do:
   (a) Send $b_i[r-1]$ to each party j (including $j = i$).
   (b) Correspondingly, receive $b_j[r-1]$ from each party j.
   (c) Let $\sharp_0 = |\{j : b_j[r-1] = 0\}|$ be the number of received zeros, and $\sharp_1 = |\{j : b_j[r-1] = 1\}|$ be the number of received ones.
   (d) The heavens reveal $c[r]$.
   (e) If $\sharp_0 \geq 2t + 1$, let $b_i[r] = 0$.
   (f) Else, if $\sharp_1 \geq 2t + 1$, let $b_i[r] = 1$.
   (g) Else, let $b_i[r] = c[r]$.
3. Return $\mathsf{out}_i := b_i[R-1]$.

In lecture it was proved that with probability $1 - 2^{-R}$ all honest parties agree on a single value out and moreover, if there exists in such that for all honest parties i, $\mathsf{in}_i = \mathsf{in}$, then $\mathsf{out} = \mathsf{in}$.

The Attack The power that the adversary has is in the protocol to choose which bits to send in Step 2a, or to send none at all. In particular, the adversary may send different values from the same malicious party to different honest parties.

For example, let $n = 3t+1$ with exactly t malicious players. For each round r, define $n_0[r] = \{\text{honest } i : b_i[r-1] = 0\}$. If $t+1 \leq |n_0[r]| \leq 2t$, then the adversary has the power to choose whether these players (or any subsets thereof) will execute Step 2e or Step 2g. In this case, we say that $n_0[r]$ is **vulnerable**. We define $n_1[r]$ and vulnerability of $n_1[r]$ analogously.[1]

What can an adversary do if she knows the heavenly coins $c[1], \ldots, c[R]$ in advance?

In round 1, suppose that $n_0[1]$ is vulnerable, and that $c[1] = 1$. Then the adversary choose whether to make $n_0[2]$ or $n_1[2]$ vulnerable in the next round by influencing the behavior of the parties in $n_0[1]$:

- For the former, force all parties in $n_0[1]$ to execute Step 2e, keeping their value.
- For the latter, force $t + 1 - |n_1[1]|$ parties in $n_0[1]$ execute Step 2g and take the value of the coin, and the remainder of $n_0[1]$ to keep their value as above.

The above generalizes to any round r and bit b where $n_b[r]$ is vulnerable and $c[r] \neq b$. Therefore, if this dangerous situation arises, the fortune-telling adversary can perpetuate it by choosing the next round's vulnerable parties based on $c[r+1]$.

[1] The proof of the protocol above uses the fact that for each r, at most one of $n_0[r]$ and $n_1[r]$ can be vulnerable.