DFKChain Degradation Incident March 15, 2023 Post Mortem

Hello all. We had a degradation on DFKChain today and wanted to report on the cause, steps taken, and what we have learned.

Degradation:

Duration: 1 hour 24 minutes

From: Mar 15, 2023, 14:49:00 PM EST (block 14838525) To: Mar 15, 2023, 16:13:09 PM EST (block 14838661)

Only 2-3 blocks per minute were produced during this time of degradation. Additionally, most

RPCs were not responsive during this period of time.

Cause:

During a scheduled upgrade, one validator's config was incorrect, causing it to not properly track the subnet, but still reported as healthy (a false positive). At this point, only 87% of the validators were online, so things were still running smoothly. The next node then began the upgrade, causing the consensus to drop to 75%, which is right at the threshold for consensus to continue.

Due to being right at the threshold, the blockchain began to struggle to produce new blocks, slowing down to a rate of about 1 block per 30 seconds, with a few periods of 4-5 minutes between blocks. During this time, the chain was still intermittently usable in a degraded state.

The issue was compounded by the nodes that were being upgraded needing to regenerate their state, which can take over an hour to complete. Since there were now two nodes in that state instead of just 1, the degradation occurred.

Action:

Within seconds of the issue starting, the devs started troubleshooting and started a call to diagnose the issue. The issue was found and remedied, and the troubled node was restarted and began to sync. During this time, the other node that was being upgraded had been syncing and was about halfway through.

Contingencies were explored including temporarily removing the troubled validators from the subnet set to get back above a healthy consensus level, but in the end the decision was made to wait for the node to finish syncing rather than risk unknown issues that could happen by removing a validator.

Once the first validator was done regenerating, the network began to produce blocks again every 2-3 seconds, while the other validator continued to regenerate.

Learned and Next Steps:

Improvements to our health status check have been suggested to make it harder for a false positive to be reported, adding additional checks to the process before proceeding. Additionally, new steps to remove validators from the validation set prior to upgrading have been suggested and will be researched further to determine feasibility and ramifications.

A debrief call with AvaLabs has been slated for next week to discuss the incident and brainstorm ideas to make the process more robust.

We have been and continue to explore the possibility of additional validators, sharing the stake and making each validator less impactful, meaning more would have to go offline to fall below consensus threshold. This can be a good option, and we do have plans to open validation beyond ourselves. However, it is also a position that requires trust. If a validator is not updated or maintained properly, and we do not have control over it, we also cannot resolve the issue. It places the health of the network in the operators of those validators, as well as ourselves. While this is a medium-term plan already, it is not something we want to do hastily. We are continuing these discussions with AvaLabs to discover the best balance in decentralization and robustness, and plan a phased approach, beginning with validation by trusted third-party partners.