

How harmful is CTUF?

Aug 2022

Author: [Vivek Pandey](#)

Collaborators: [RobinSomya Ranjan SahuSheekha VermaAshish Jain](#)

Abstract:

CTUF (Connected to Transacting User Filter) is a prominent anti fraud filter. There is a general concern that CTUF blocks too many users. In this article we do data analysis and find that if we were to remove CTUF completely, we expect at most 0.4pp increase in approval rates and an increase of at most 600 new transacting users everyday. We then provide some notes and provide an opinion on the way forward.

I encourage readers to look into the data and discuss if they disagree with the methodology and request for additional investigation if necessary. I have provided some of the queries for anyone to be able to replicate these results.

Impact of CTUF on approval rates

I take one representative day: Aug 21, 2022 (Appendix A contains same analysis for Aug 21-24)

Number of users sending us approval calls in the day: 5,366,050

Number of users getting approval failures in the day: 3,674,565

Number of users getting approval failure because they are transactionally blocked: 85,586

Query 1A: Number of users sending us approval calls in a day:

```
SELECT
  COUNT(
    DISTINCT(
      CAST(json_extract(payload, '$.user.phone_number') AS VARCHAR)
    )
  )
FROM
  vivek_all_events
WHERE
  bucket_id = 'date_2022_08_21' AND
```

```
event_name IN ('UserApprovedEvent', 'UserApprovalFailedEvent')
```

Query 1B: Number of users getting approval failures in the day:

```
SELECT
  COUNT(
    DISTINCT(
      CAST(json_extract(payload, '$.user.phone_number') AS VARCHAR)
    )
  )
FROM
  vivek_all_events
WHERE
  bucket_id = 'date_2022_08_21' AND
  event_name IN ('UserApprovalFailedEvent')
```

Query 1C: Number of users getting approval failure because they are transactionally blocked

```
SELECT
  COUNT(
    DISTINCT(
      CAST(json_extract(payload, '$.user.phone_number') AS VARCHAR)
    )
  )
FROM
  vivek_all_events
WHERE
  bucket_id = 'date_2022_08_21' AND
  event_name IN ('UserApprovalFailedEvent') AND
  CAST(json_extract(payload, '$.message') AS VARCHAR) = 'User
transactionally unauthorized'
```

Out of 86K disapprovals because of transaction unauthorized reason, 20K are because of CTUF (remaining 66K disapprovals are because the user is blocked due to some other reason). Justification of this estimate will be provided in the next section.

Thus, around 20K disapprovals are because of CTUF blocks. So, if we take away CTUF complete, increase in approval rate will be around $20K/5.4M = 0.4pp$

Impact of CTUF on daily new transacting users

New users acquired on a daily basis is another metric that we track. We currently get around 10K new users everyday. We ask the question: What will be the bump in daily new users if we were to stop CTUF entirely and also unblock all the users who have been blocked by CTUF so far. We find that we expect a bump of 400 daily new users.

On Aug 21, 86K users were not approved because they were transactionally blocked.(Query 1C)

Next, let's try to estimate: of the 86K users who were not approved because they were transactionally blocked, what fraction were transactionally blocked due to CTUF.

To estimate this, we adopt the following methodology.

We find the list L of all the users who have been blocked for some reason (including but not limited to CTUF) since Jan 1, 2022. Here are sample rows of L.

1	Realtime Fraud Blocker: Connected to transacting users.	7077994891
2	Realtime Fraud Blocker: Connected to transacting users.	8050244096
3	Automated_Cluster_Block	7017925178
4	Automated_Cluster_Block	7849900776
5	Automated_Cluster_Block	7983082689
6	Automated_Cluster_Block	9609925206
7	Realtime Fraud Blocker: Connected to transacting users.	7510332298
8	Realtime Fraud Blocker: Connected to transacting users.	8017327150
...	...	...

Next, amongst the 86K users whose approval failed because of being blocked, we found the following, by joining with L:

- 15K users were blocked by CTUF
- 51K users were blocked due to other reasons
- 20K users were not found in L. Since L only contains users blocked since Jan 1, 2022, we presume they were blocked prior to 2022.

Using above data, we estimate that amongst the 86K users blocked on Sep 21, 2022, users blocked due to CTUF were:

$$86K * (15K / (15K + 51K))$$

$$= 20K.$$

Two notable points about above estimate:

- We assume that the 20K users blocked before 2022 will have same proportion of CTUF vs non CTUF blocked users.
- We are entirely discounting the fact that 2/3 of all CTUF blocked users are later unblocked.

Due to the second point, actual impact of CTUF blocks on new user per day is going to be smaller than what we estimate.

Next we assume a CTR of 3% for new users.

Thus, if there were not CTUF, additional new users that we would get is: $20K * 3\% = 600$.

Few more notes

1. CTUF blocks some 4K-5K users daily. However, there are two factors that dissipate its impact.

- Firstly, a good fraction of these users have high PD, and thus they don't have a credit line any way. Here is an analysis by [Somya Ranjan Sahu](#)

		pd score		
		≤ 0.25	> 0.25	All
Network Score	≤ 0.27	1,37,549	44,465	1,82,014
	> 0.27	42,154	16,833	58,987
	All	1,79,703	61,298	2,41,001

Above analysis is on data from June 2nd cycle to Aug 1st cycle (total 4 cycles):

- 241K users were blocked by CTUF
- 182K of them were later unblocked
- Of the 59K users that continued to be blocked, 16K had high PD score.

Thus, $16K/59K = 27\%$ of the users won't have the credit line anyway.

- Secondly, a large fraction of $182/241 = 75\%$ of the blocked users are unblocked in an hour.

2. Note also that the sequence model is also blocking around 1000 users per day. CTUF will have some overlap with that. So, stopping CTUF will not unblock all the CTUF blocked users. [We have not yet checked how much overlap is there]

3. CTUF serves as a bulwark against a large class of organized fraud that we may not have seen earlier.

Path forward

Here is our recommendation

1. No dropping CTUF till we have an alternative

Above calculations show that we expect a 0.4pp improvement in approval rates and additional 600 new users daily if CTUF were completely dropped. We believe these numbers are not high enough to take the risk of stopping CTUF altogether.

2. Work on network model

We are currently working on Network Analytics as a Service which will replace CTUF. NAaaS will have network rules and an improved network model. We now have much better links from TU and even our own data. This will take a few months.

3. Check if unblock can be realtime

We don't believe making unblock realtime is going to impact metrics significantly (since getting rid of CTUF completely itself provides a gain of 600 new users max per day, so this project will yield a small fraction of that).

However, we will check if this is something easily doable.

4. Can CTUF block fewer users

We will check if CTUF can be relaxed in some ways easily. Again we don't expect a big impact if we find something.

[I am interested in items (3) and (4) less because of potential impact, and more because doing such exercise will help me understand problem space better]

Appendix A

	21-8-2022	22-8-2022	23-8-2022	24-8-2022
Num users sending approval calls (A)	5366050	4611902	4720751	4723014
Users not approved because they were transactionally blocked (B)	85586	80755	82367	83037
Of the users in (B) how many were blocked by CTUF in 2022 (C)	15473	14301	13838	13664
Of the users in (B) how many were blocked by system other than CTUF in 2022 (D)	51311	49638	50224	50244
Of the users in (B), how many overall users we estimate to be blocked by CTUF = $B * C / (C + D)$ (E)	19823	18056	17791	17977
Bump in approval rates if there were no CTUF = E/A	0.37%	0.39%	0.38%	0.38%
New users bump if there were no CTUF, assuming CTR of 3% for new users = $E * 0.03$	594	541	533	539