

Exploratory Data Analysis on CDETS

Rajesh Ekkaladevi, Satya Pradhan, Baldeep Jain, Suneel Yeluru

Feb 19, 2018

Introduction

CDETS is a centralized application to track all types of defects across all CISCO products. The application captures many fields some of them are mandatory and important while some of them are ancilliary info and not mandatory. Looking at this data can provide many useful insights on the defects life cycle and how they are handled.

The primary goal of this exercise is to explore the mean time to assign, resolve and verify (mtta, mttr, mttv) metric across multiple dimensions namely PIN, Category, Severity, Found, Qtr, etc and also across defect status changes to see if anything obvious or interesting observations can be made. The analysis help familiarize with the data, identify data quality issues and facilitate in considering right metrics and dimensions on data science analysis and prediction projects. For the case the analysis was performed to a max of 2 dimensions. Where needed further deep dive can be performed at multiple granularity and metrics.

Synopsis

For the analysis the CDETS defects data is extracted for the period 2015-01-01 till 2018-12-31 (recent 4 years) but the analysis focus is for the period 2016-01-01 till 2018-12-31 (recent 3 years). This is to include defects that are created before 2016 but have some update in 2016, 2017 and 2018. The defects are extracted using the PIN to Defect mapping queries provided by the Quality Team that is in use to monitor all PINs under Ravi C's org. The query is based on Submit_On date and considers defect Severity of type 1,2,3 (Catastrophic, Sever, Moderate). In total there were 188k unique defects either created, assigned, resolved or verified or ending in other states like Duplicated, Junked, etc. There were 10k duplicate defects falling into multiple PINs. The analysis considers the current state of defect whether it is closed or open. Also, for the average time statistics only 2018 defects (had some modifications) are used and they are in total 78k defects. The following fields are considered for analysis. For reference the CDETS field definitions are available at <http://www.in-metrics.cisco.com/doc/fields.html>.

- Dimensions:
 - **PIN** : DNAC, IOT, Prime-Infra, Routing, SPA, Switching, Viptela, Wireless, XI-PI-Common, XI-PI-Polaris
 - **Category** : Categories under PIN (AP, ASR1K, BB, Cat3k(NG), CAT9K, CDB, CSR, ISR4K, ISRV, SDWAN, VG, WLC, etc).
 - **Class** : Class namely (CSC.labtrunk, CSC.sys, CSC.embu, etc).
 - **Found** : The stage where the defects are found (customer-use(CFD), devtest, func-test, regression, sys-test, etc).
 - **Severity** : Severity of the defect (Catastrophic, Severe, Moderate).
 - **Qtr** : Calendar quarter is considered. If needed this can be modified to represent CISCO financial quarter.
 - **DE_Manager** : Top 15 DE Managers in terms of numbers of bugs resolved.
 - **Created_By** : Top 15 Engineers in terms of numbers of bugs resolved.
 - **Defect_State** : State of the defect (A - Assigned, C - Closed, D - Duplicate, F - Forwarded, H - Held, I - Information required, J - Junked, M - More, N - New, O - Open, P - Postponed, R - Resolved, S - Submitted, U - Unreproducible, V - Verified, W - Waiting)
- Metrics:
 - **Defect Count** : Defects by selected dimension.

- **Defect State Change Cycle** : States through which the defect went through (e.g. NARV, NAIRV, etc)
- **Defect State Change** : Between 2 states or a hop (e.g. NA, AR, RV, etc).
- **Freq** : Frequency of change of defect from one state to another. Can be more than the no. of defects.
- **Avg TA** : Average Time Taken to Assign in days.
- **Avg TR** : Average Time Taken to Resolve in days.
- **Avg TV** : Average Time Taken to Verify in days.

Processing Performed

- The quarterly numbers consist of number of defects found in each state Created (C), Assigned (A), Resolved (R), Verified (V) in that quarter. It is likely that same defect can go through multiple states in same quarter.
- The max values for TTA, TTR, TTV are restricted to the 95th percentile given there are some outliers.
- Size of the points for Avg TTA, Avg TTR, Avg TTV are based on volatility (standard deviation) of the observation.

Data Analysis

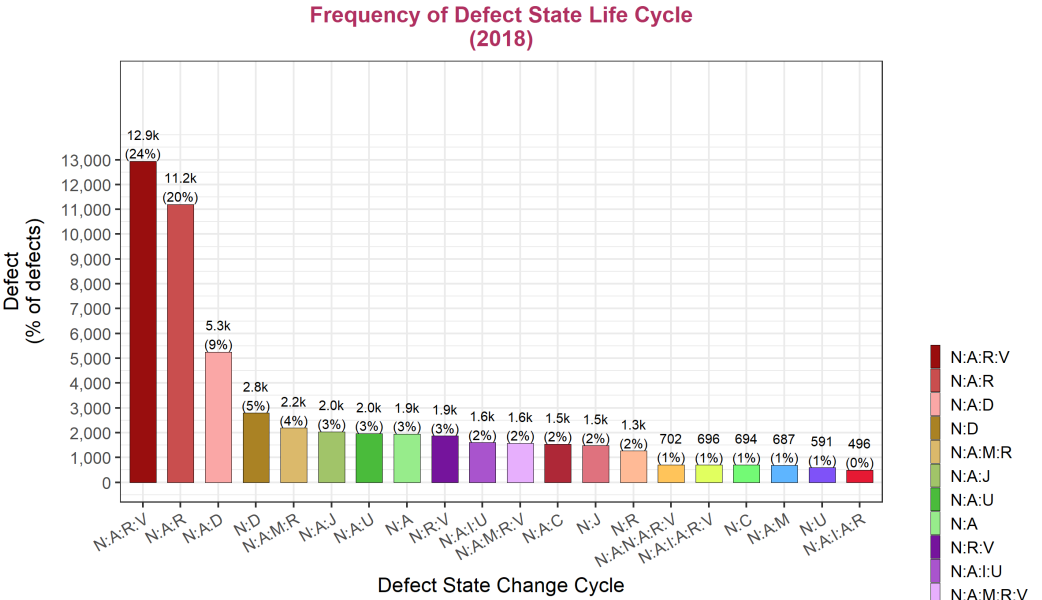
Defect Status Life Cycle (2018)

Data: There are more than 1600 different bug life cycles found. Only bugs cycles having 1% of the total defects are considered This brings the data to 85% coverage for the chart.

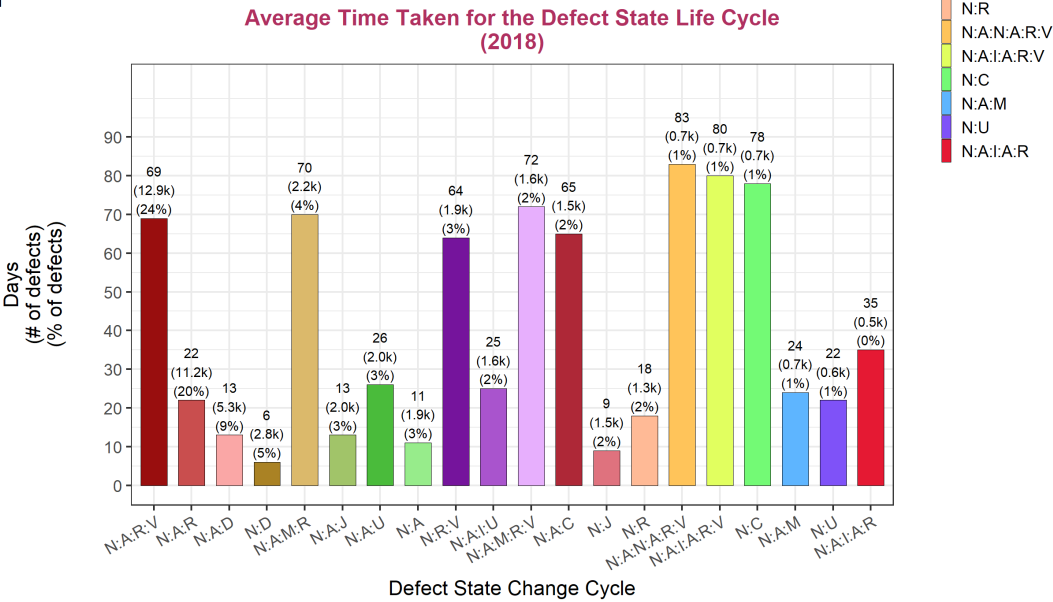
Observations:

1. N:A A:R R:V (New to Assigned, Assigned to Resolved, Resolved to Verify) is the top pattern as expected.
 2. About 42% ($29/(41+29)$) defects are not validated.
 3. About 5% of defects go to MoreOn stage where the defect move to different group (state change from A:M M:R).
 4. 2% of defects go directly from New to Resolved state by passing Assignment.
1. The standard defect cycle (N,A,R,V) takes 53 days on average to close.
 2. The defects that require more info (state I) take on average 17 days more from the normal A,R,V stage (70-53).

SC



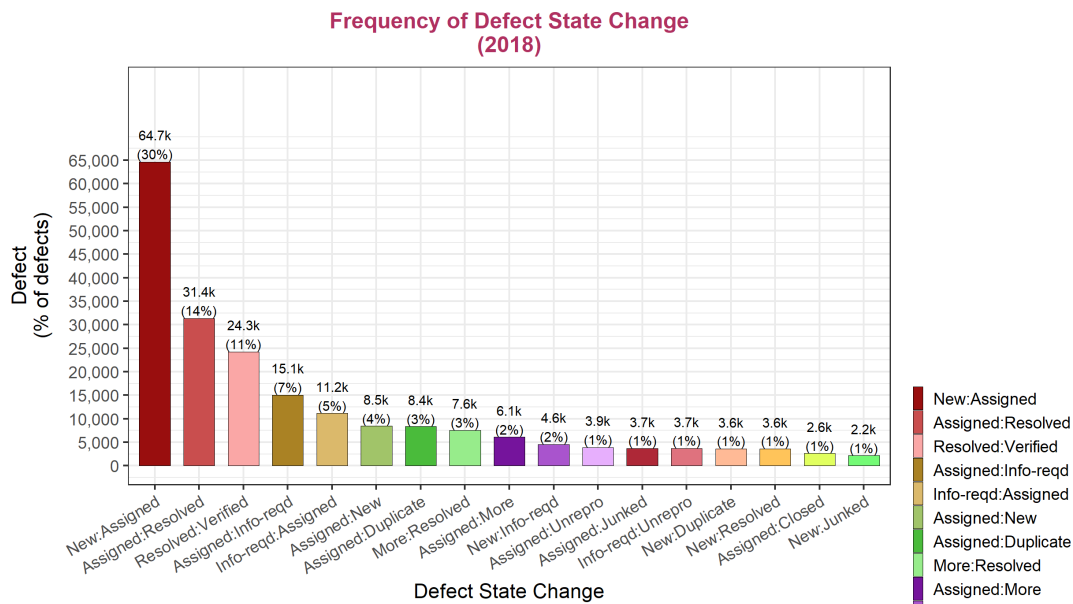
sd



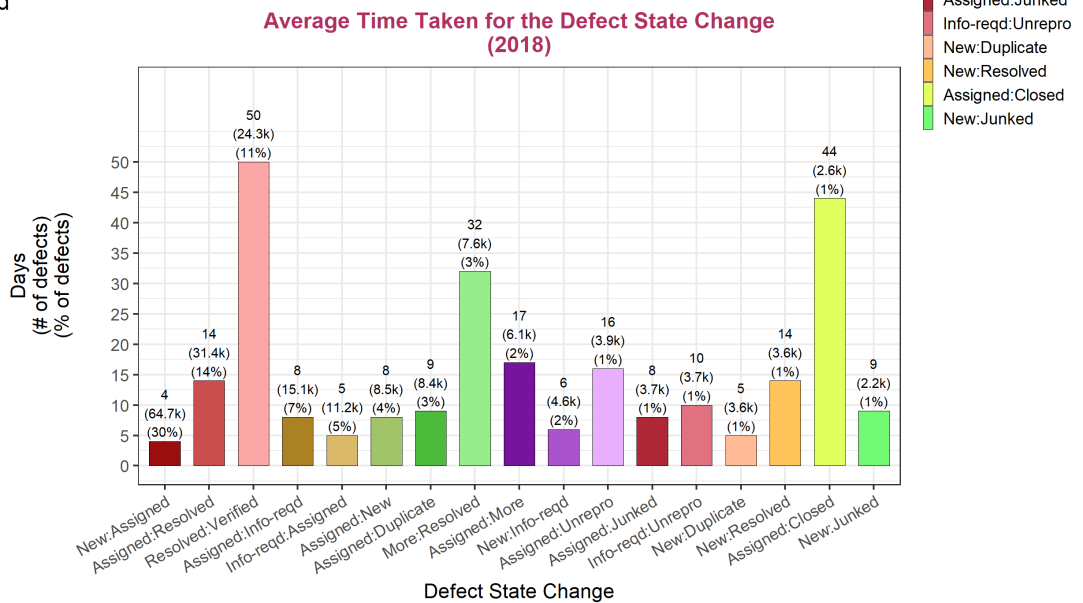
Defect Status Life Cycle by End Status (2018)

Observations:

sc



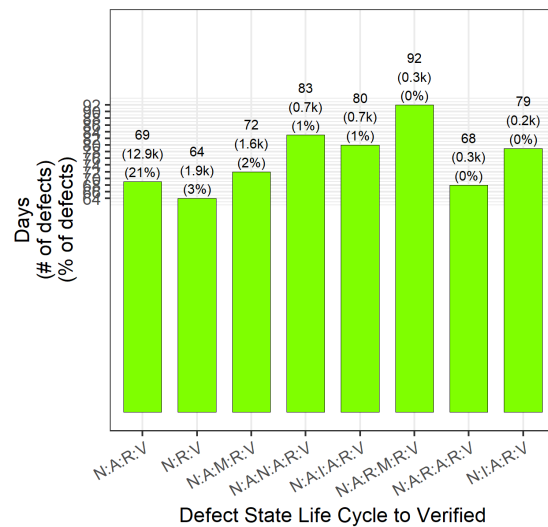
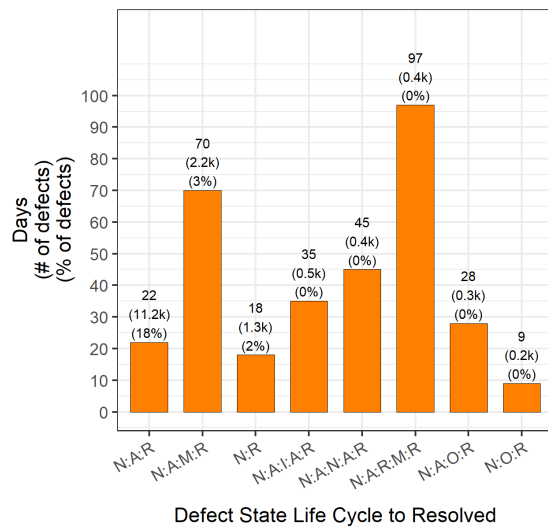
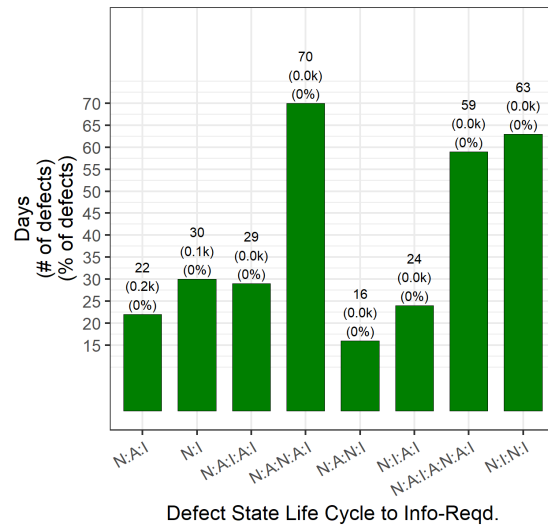
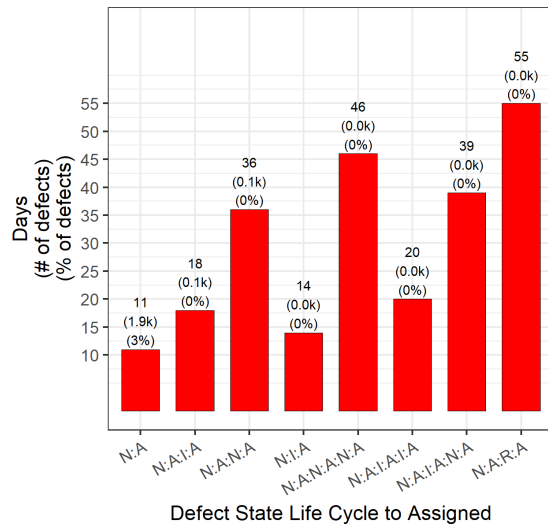
sd



Defect State Change

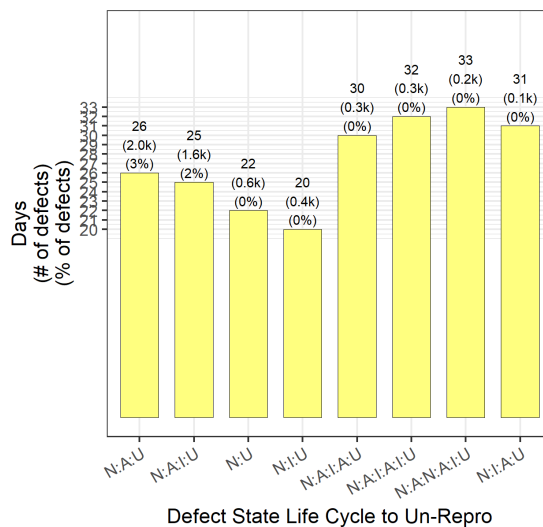
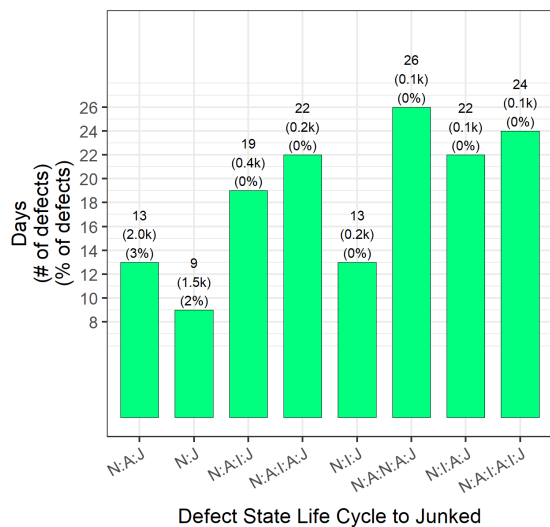
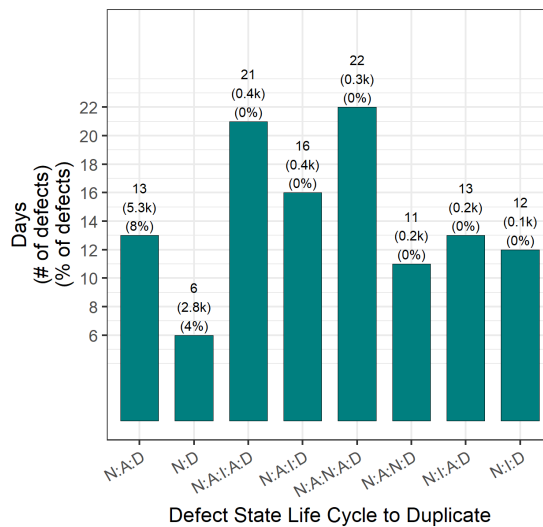
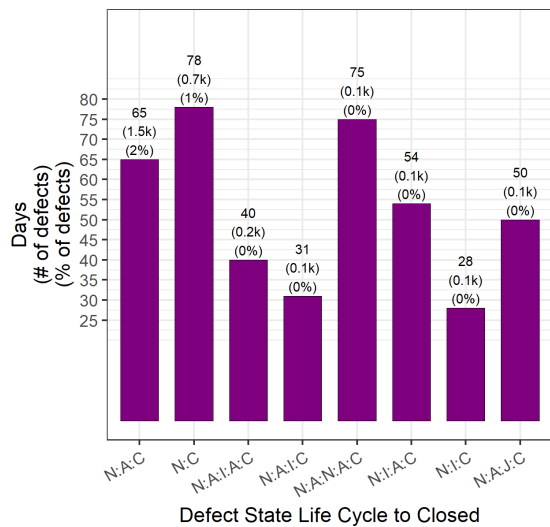
Observations:

Average Time Taken to Move to (A,I,R,V) States



Observations:

Average Time Taken to Move to (C,D,J,U) States

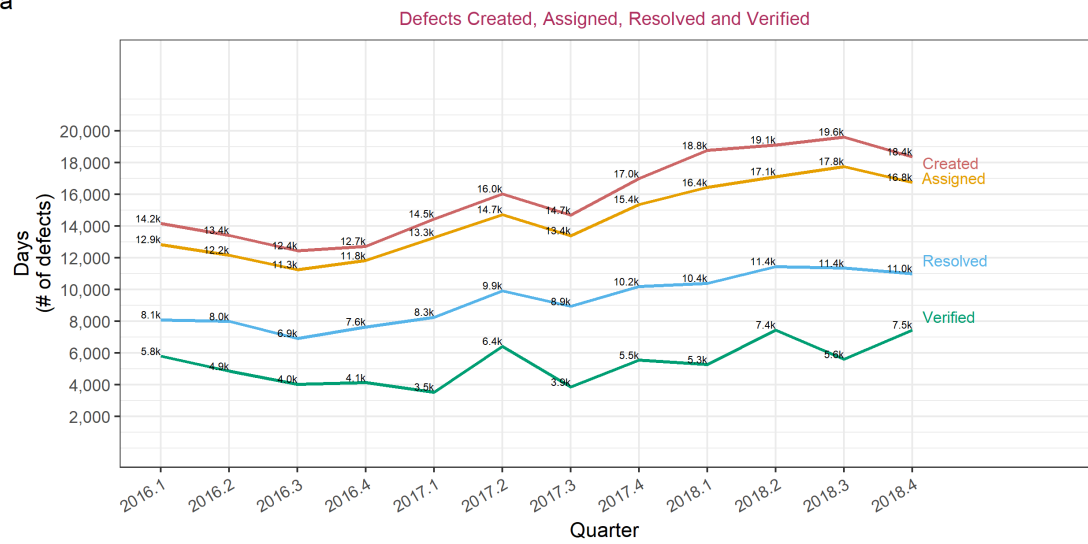


Defects and MTT Trend across EN

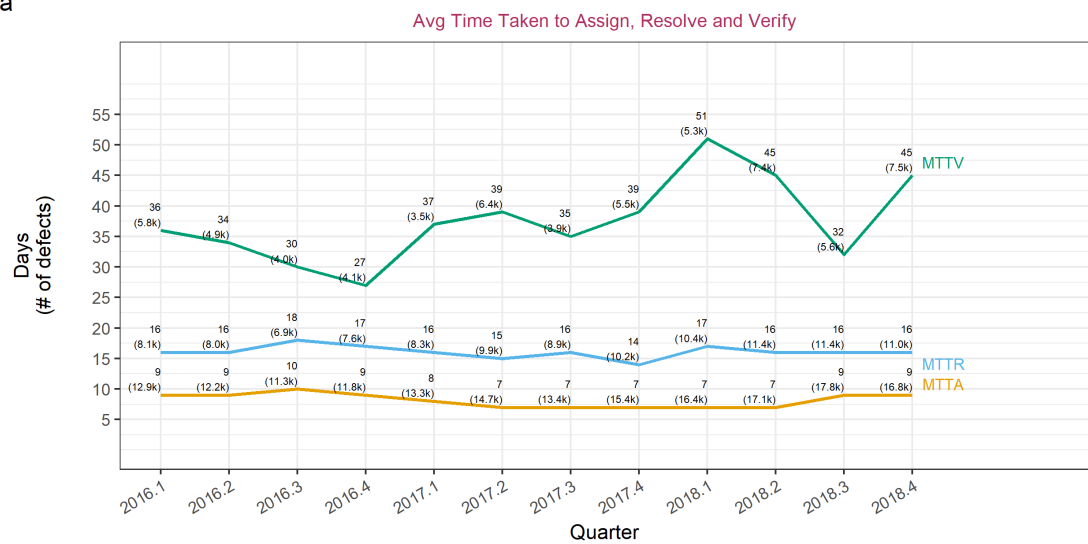
Observations:

Defects Trend and MTT across EN

ea



ea

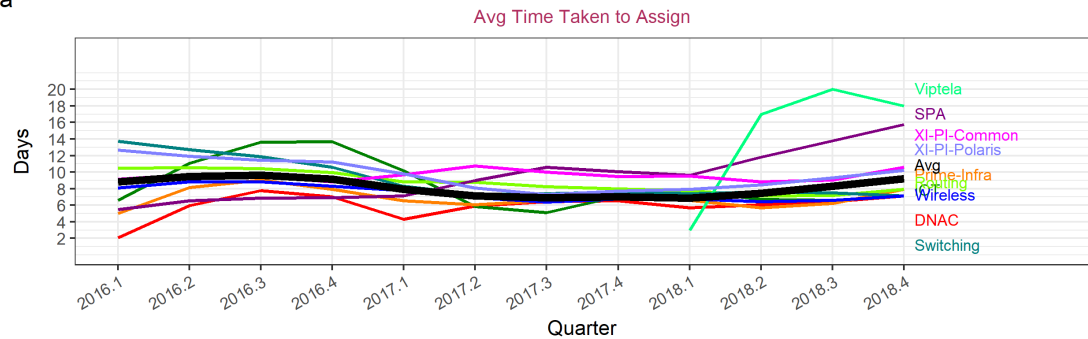


MTT Trend by PIN

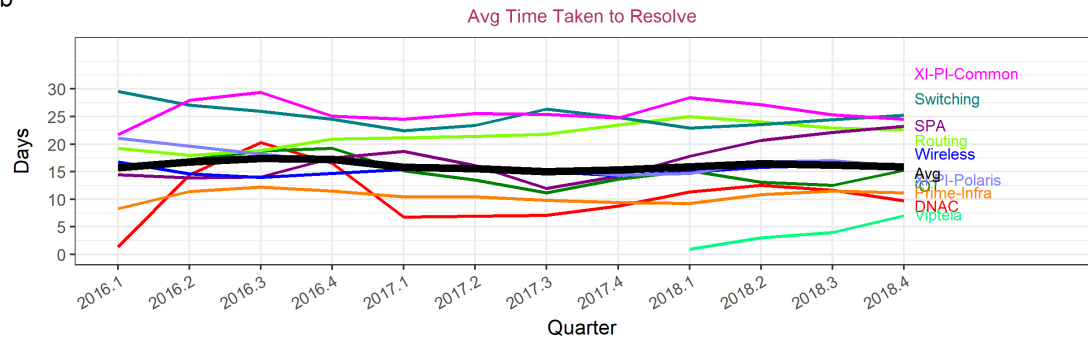
Observations:

Average Time Taken to Assign, Resolve, Verify by PIN

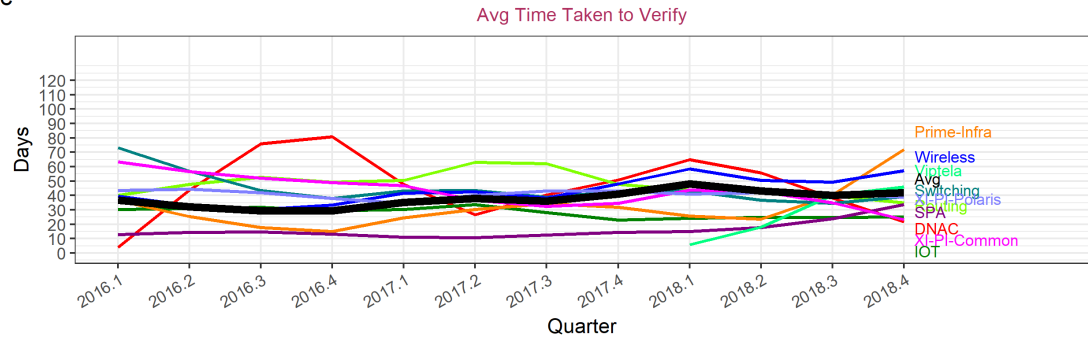
pa



pb



pc

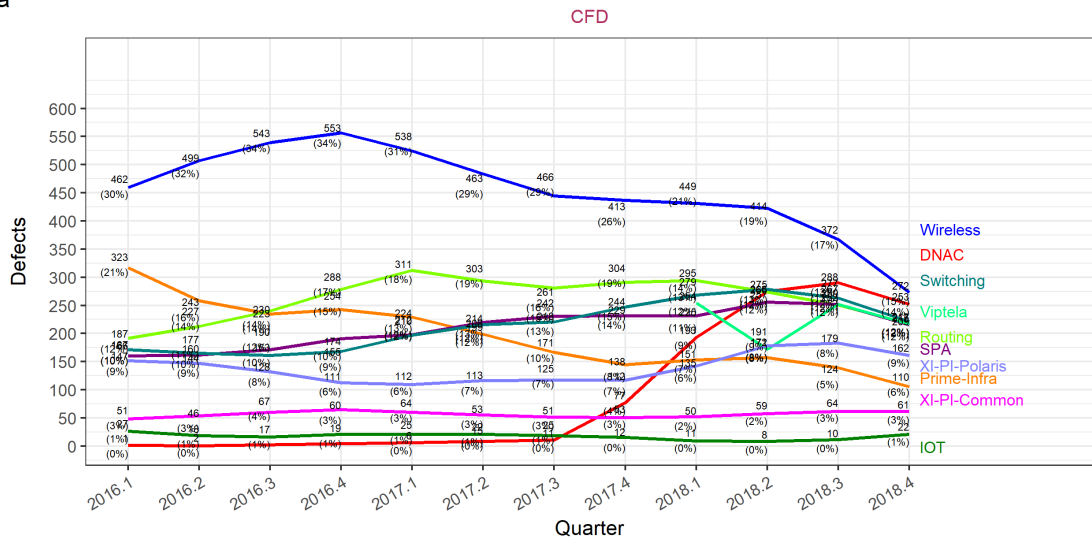


Found (CFD/IFD) Trend by PIN

Observations:

Found (CFD/IFD) by PIN

fa



fb

