Planned Maintenance Tool (PMT): A Data-Driven Approach to Recommending the Best Time for Planned-Maintenance

A Technical Paper prepared for SCTE by

Pete Quesada
Sr. Principal Engineer
Comcast Cable
4100 E Dry Creek Rd, Centennial, CO 80122
720.692.8683
Pete_Quesada@Comcast.com

Julianne Heinzmann
Software Developer & Engineer
Comcast Cable
1800 Arch Street, Philadelphia PA, 19103
267.586.7928
Julianne_Heinzmann@Comcast.com

Nishesh Shukla
Software Developer & Engineer
Comcast Cable
4100 E Dry Creek Rd, Centennial, CO 80122
303.263.8121
Nishesh_Shukla@Comcast.com

Resmi Vijayan
Software Developer & Engineer
Comcast Cable
1800 Arch Street, Philadelphia PA, 19103
817.908.7353
Resmi_Vijayan@Comcast.com

Mike O'Dell
Director Network Maintenance
Comcast Cable
Virtual Location
412-417-0481
Michael_Odell@cable.comcast.com

May Merkle-Tan
Lead Machine Learning Researcher
Comcast Cable
4100 E Dry Creek Rd, Centennial, CO 80122
Heng-RuMay_Tan@comcast.com

SCTE CABLE-TEC EXPO 22
SEPTEMBER 19-22 • PHILADELPHIA, PA

2022 Fall Technical Forum
SCTE • CABLELABS • NCTA

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Examples

List of Tables

1. Introduction

1.1. Problem Statement

Planned maintenance is a daily activity for any number of complex systems, including cable plants. It is important to think of a cable plant as a living, breathing organism that requires care and feeding, involving replacing parts that are continuously exposed to the elements. Repairs often include identifying problems, making repairs, and replacing parts while temporarily interrupting a customer's service. A service interruption event (SIE) averages between five to ten minutes. In an ideal world, all SIEs would be performed during the evening maintenance window, but in practice, most short-duration SIEs must be performed outside of this maintenance window. Currently, SIEs are scheduled without the benefit of knowing which hours would have the least or the highest amount of subscriber impact. This information would be invaluable in optimizing the ideal time to perform an SIE.

We set out to find if a data-driven system could be developed to determine the best time to conduct SIEs. Performing SIEs during times when they will have the least impact on subscribers would not only provide a better subscriber experience but also could potentially cut the expenses incurred by responding to customer interactions (CI), such as calls to our care agents, unnecessary truck rolls, chat sessions, and other triaging events.

1.2. Proposed Solution

Our research involved identifying data that would have sufficient signal to indicate the least and the most impactful times to perform an SIE for the set of subscribers that each SIE would impact; we chose to calculate recommendations on each subscriber's hourly high speed data usage, which we will refer to, here, as customer usage information (CUI).

The main objective of our research was to develop a Planned Maintenance Tool (PMT) to assist with field operations. The algorithm that drives the PMT evaluates the hourly customer usage information (CUI) for each set of customer accounts that an SIE will impact. Then the PMT returns the hours when the SIE will be the least impactful toward those customers. We assessed the validity of our algorithm with historical SIEs and corresponding CI data from a geographic area that we will refer to as the 'test region.'

We have found that CIs typically increase when there is an unexpected SIE. We theorized that if we could create an algorithm to identify the best time(s) to perform an SIE, we would see a less severe CI increase around the hour the SIE is performed. Our assessment, though limited in breadth, appeared to follow this expected trend, and the findings pinpointed subsets of CIs we could monitor and assess periodically for financial and customer impact.

In the following sections, we discuss the constraints that motivated the initial version of the PMT user application and the refinements we think would be necessary further to improve the user experience and reliability of the PMT.

2. PMT Core Components

At its core, the PMT tool comprises two components that provide the data-science backbone of the system. They are the data processing and the recommendation algorithm, which are described in detail in this section.

2.1. Data Sources

It was posited that the best time to perform SIEs with the least impact on customers would be the hours when the least amount of data was consumed. We aggregated data from the following sources to collect a good source of data consumption for a group of homes.

2.1.1. Customer Usage Information (CUI)[1]

Customer usage information is aggregated, in bytes, for upstream and downstream traffic for each DOCSIS (Data Over Cable Service Interface Specification) -capable device at hourly intervals.

2.1.2. Device to Network Mapping (DNM)

We needed to map individual MAC addresses to the network elements, including nodes, regions, and CMTS (Cable Modem Termination System). DOCSIS-capable network infrastructure allows the implementation of a system that polls all devices six times a day to check for any impairments, noise, and other factors. This data also records all devices connected and active in the network, along with their mapping to customer account numbers, location, and elements in the network.

2.1.3. Plant Topology Information (PTI)

The plant topology information includes data from various source systems to provide a hierarchical view of multiple elements in the network, such as head-end devices, CMTSs, RF cables, power supplies, taps, nodes, drop cables, and customer devices in the network. Although not used in the initial assessment, plant topology will be used by the developed application.

2.1.4. Service Interruption Events (SIE)

Service interruption events are initiated when an interruption in the plant is needed to correct RF system impairments, if the plant needs to be disconnected to replace a network component, or if periodic maintenance needs to be performed. Data about service interruption events is recorded and includes the time they occurred, the length of time that service was interrupted and a list of accounts that were affected.

2.1.5. Customer Interactions (CI)

Customer interaction events are identified as indicators of the impact SIEs could have on customers. Customer interaction events consist of logs of activity in a customer's timeline. They include billing, communications, customer chats, customer request tickets, inbound calls, tech appointments, speed tests performed, and equipment orders. The events selected for this analysis were limited to inbound customer

[1] *We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.*

calls, chats, speed tests, device reboots, self-service device health tests, virtual assistant chats, and truck rolls.

2.2. Data Processing

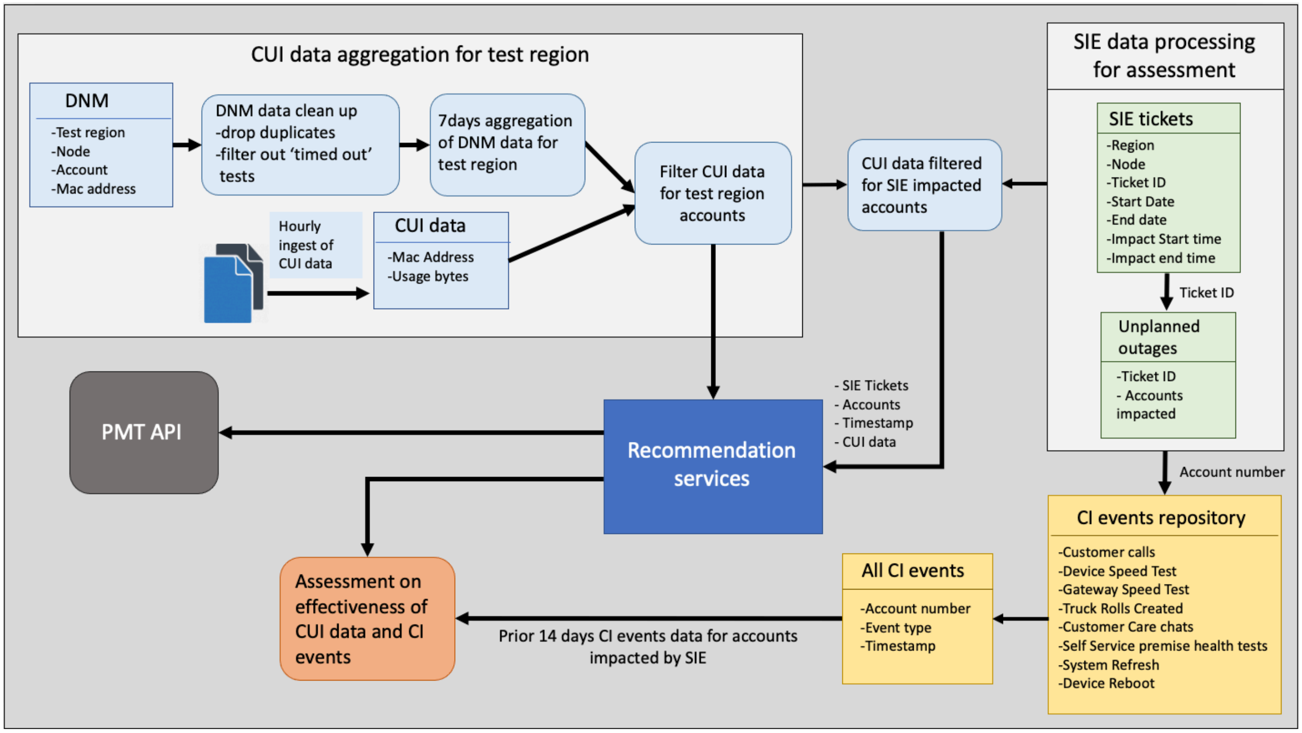


Figure 1 - Data Processing Workflow for the PMT Recommendations Process

2.2.1. CUI data filtering for test region

Daily polls of the DNM from the test region are collected and cleaned to filter out 'timed out' and inactive devices. A 7-day aggregate of DNM data is performed to accommodate account additions or deactivations changes. This seven-day aggregated DNM is then used to filter the stream of hourly CUI. We derive an hourly aggregate of CUI at account levels for the total 24 hours daily, which serves as the input into the PMT algorithms for any accounts (households) likely to be impacted by a network SIE.

2.2.2. Selected Accounts for Recommendations

Based on CMTS and node segment ID from PTI data, the geolocation latitude/longitude for customer service addresses are derived. The latitude/longitude information is passed to a Grouping Service (see section 4.3.1) to find all field topology information such as cables, taps, buildings, and addresses inside the node boundary. Accounts are then formed into multiple groups based upon the topology information by performing a fuzzy comparison of street addresses. Additionally, it was found useful to resolve some addresses by comparing their latitudes/longitudes using proximity. The account groupings[2] are further used in generating recommendations (See section 2.3.1).

---

[2] In Section 3 where we detail the assessment of the CUI and PMT algorithm, these "account groupings" are simply the set of SIE-impacted accounts, derived directly from SIE logs.

2.2.3. SIE Data Processing for Assessment

We collected SIE data for the test region for this assessment. The accounts impacted by each SIE are derived from an 'unplanned outages' dataset. For each account affected by an SIE, we collected customer interaction events from 14 days before the onset of a SIE, which we used to derive 3 different CI baselines. The SIE ticket information and CI events were then used to create the assessment detailed in Section 3.

CI event types include:

- Customer calls
- Virtual assistant chats
- Device speed tests
- Gateway speed tests
- Truck rolls created
- Customer care chats
- Self-service premise health tests performed through a web application
- System refreshes
- Device reboots

We summarize the various data processing pipelines in Figure 1.

2.3. Recommendation Algorithm

The PMT Tool uses existing network usage data to create a historical picture of how customers interact with our services within their homes. This section details the two main components of our recommendations: Data Analysis and the Ranking-Based Algorithm.

2.3.1. Data Analysis

CUI data is ingested and filtered down to Account ID, Timestamp, and Total usage on a given Hour (upstream bits + downstream bits = total). Then, Data Aggregations are performed on the filtered dataset. The important fields to note are those that are shown in Figure 2.

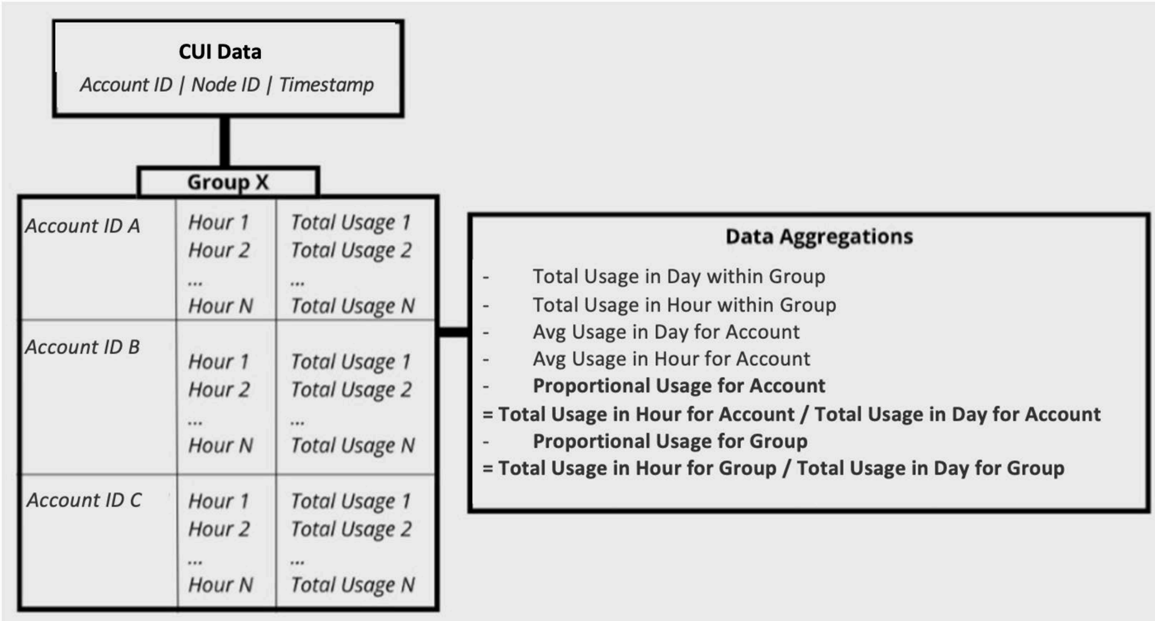


Figure 2 - How CUI data are aggregated to derive proportional usage for account and selected accounts

Proportional Usage for Account: We take the total usage seen in an hour for each Account ID and divide it by the sum of the total usage for the Account ID over the entire day (or however many hours are being considered in the comparison).

Proportional Usage for Group: We collect the total usage seen in an hour for the selected accounts (abbreviated as 'Acct' in the following Examples); this selected set of accounts is also referred to as a "Group". We then divide it by the sum of the total usage for the Group over the entire day (or the specific hours considered in the comparison).

Consider Example 1 for an applied example of calculating the Proportional Usage per Account:

Acct A (Proportional Usage)	**Acct B** (Proportional Usage)
Total Usage for Day = 99	*Total Usage for Day = 102*
Hour 1 = 10 / 99 = 0.101	Hour 1 = 9 / 102 = 0.088
Hour 2 = 40 / 99 = 0.404	Hour 2 = 6 / 102 = 0.059
Hour 3 = 3 / 99 = 0.03	Hour 3 = 5 / 102 = 0.049
Hour 4 = 15 / 99 = 0.151	Hour 4 = 12 / 102 = 0.118
Hour 5 = 20 / 99 = 0.202	Hour 5 = 10 / 102 = 0.098
Hour 6 = 5 / 99 = 0.051	Hour 6 = 20 / 102 = 0.196
Hour 7 = 6 / 99 = 0.061	Hour 7 = 40 / 102 = 0.392

Example 1 - Calculating Proportional Usage per Account

With Proportional Usage (PU) per Account derived, we can move on to describe how we ranked an account's PU relative to others within the selected set of accounts ("Group" and "Acct" are used interchangeably).

2.3.2. Ranking-Based Algorithm

Scoring ***N*** = # Hours in Group // 2	
Best Hours Hours ~ **N** LOWEST Total Usage / Acct % of Acct for each Hour Returned	**Worst Hours** Hours ~ **N** HIGHEST Total Usage / Acct % of Acct for each Hour Returned

Figure 3 - Scoring of best and worst hours

Figure 3 outlines the theory behind making recommendations based on CUI usage data. All recommendations are made according to the number of hours to be considered before creating the recommendation. The defined hour arrangements are as follows:

Morning Hours: *6 am-5 pm (inclusive)*

Extended Working Hours: *6 pm-11 pm (inclusive)*

All Day: *6am-11pm (inclusive)*

The number of hours being compared in each arrangement is divided by 2 (dropping any remainder) to get *N*.

Morning Hours: $N = 5$

Extended Working Hours: $N = 2$

All Day: $N = 8$

N is used to determine the best and the worst hours to perform maintenance for each Account ID. For example, looking at a 7-hour window of time, *N* would be equal to 7 // 2 = 3. We apply this in *Example 2*.

Acct A (Total Usage)	**Acct B** (Total Usage)
Hour 1 = 10	Hour 1 = 9
Hour 2 = 40	Hour 2 = 6
Hour 3 = 3	Hour 3 = 5
Hour 4 = 15	Hour 4 = 12
Hour 5 = 20	Hour 5 = 10
Hour 6 = 5	Hour 6 = 20
Hour 7 = 6	Hour 7 = 40
***N* Best Hours Acct A** = Hours 3, 6, 7	***N* Best Hours Acct B** = Hours 3, 2, 1
***N* Worst Hours Acct A** = Hours 2, 5, 4	***N* Worst Hours Acct B** = Hours 7, 6, 4

Example 2 - Determining Best and Worst Hours according to CUI

Next, the results are accumulated to get the percentage of each Account ID returned for each hour within a group. *Example 3* is expanded to demonstrate this:

Group AB (Acct A and Acct B)

Best Hours Group AB:	**Worst Hours Group AB:**
Hour 1 = 0.5	Hour 1 = 0.0
Hour 2 = 0.5	Hour 2 = 0.5
Hour 3 = 1.0	Hour 3 = 0.0
Hour 4 = 0.0	Hour 4 = 1.0
Hour 5 = 0.0	Hour 5 = 0.5
Hour 6 = 0.5	Hour 6 = 0.5
Hour 7 = 0.5	Hour 7 = 0.5

Example 3 - Proportional Network Usage According to Group

We now take our example's Proportional Users (PU), where PU = % Accts at Best Hour - % accounts at the worst hour. *Example 4* breaks this down:

Best Hours Group AB:		**Worst Hours Group AB:**	**(PU):**
Hour 1 = 0.5	-	Hour 1 = 0.0	= 0.5
Hour 2 = 0.5	-	Hour 2 = 0.5	= 0.0
Hour 3 = 1.0	-	Hour 3 = 0.0	= 1.0
Hour 4 = 0.0	-	Hour 4 = 1.0	= -1.0
Hour 5 = 0.0	-	Hour 5 = 0.5	= -0.5
Hour 6 = 0.5	-	Hour 6 = 0.5	= 0.0
Hour 7 = 0.5	-	Hour 7 = 0.5	= 0.0

Example 4 - Calculating Proportional Users

We define the Weight of our Prediction based upon the summation of the Proportional Usage for the accounts over each of the Hours in Best Hours and Worst Hours, as briefly explained in *Example 1.* We then combine that with the results from *Example 2* to get the Proportional Usage of Accts across the Best Hours and the Proportional Usage of Accts across the Worst Hours. This is written out in *Example 5*. The Best Hour results corresponding to each Account are marked with a "*", while the Worst Hours are denoted with a "-" minus sign.

*** Best Hours for Mac**
- Worst Hours for Mac

Mac A	Mac B	B-Level	W-Level
*Hour 1 = 0.101	*Hour 1 = 0.088	(0.088 + 0.101) = 0.189	(0) = 0
-Hour 2 = 0.404	*Hour 2 = 0.059	(0.059) = 0.059	(0.404) = 0.404
*Hour 3 = 0.03	*Hour 3 = 0.049	(0.03 + 0.049) = 0.079	(0) = 0
-Hour 4 = 0.151	-Hour 4 = 0.118	(0) = 0	(0.151 + 0.118) = 0.269
-Hour 5 = 0.202	-Hour 5 = 0.098	(0) = 0	(0.202 + 0.098) = 0.3
*Hour 6 = 0.051	-Hour 6 = 0.196	(0.051) = 0.051	(0.196) = 0.196
*Hour 7 = 0.061	-Hour 7 = 0.392	(0.061) = 0.061	(0.392) = 0.392

Example 5 - Calculating Account Level Prediction Weights

Finally, we get the Weight of Our Prediction (WP) by taking W-Level – B-Level, shown in Example 6.

Hour	W-Level		B-Level	(WP)
1	0	-	0.189	= -0.189
2	0.404	-	0.059	= 0.345
3	0	-	0.079	= 0.079
4	0.269	-	0	= 0.269
5	0.3	-	0	= 0.3
6	0.196	-	0.051	= 0.145
7	0.392	-	0.061	= 0.331

Example 6 - Calculating Prediction Weights

Our final step is to take our PU from *Example 4* and subtract the WP to get our overall recommendation. This final step is calculated in *Example 7*.

Recommendation = (PU) – (WP)

Hour	(PU)	(WP)	Recommendation
1	0.5	-0.189	= 0.689
2	0.0	0.345	= -0.345
3	1.0	0.079	= 0.921
4	-1.0	0.269	= -1.269
5	-0.5	0.3	= -0.8
6	0.0	0.145	= -0.145
7	0.0	0.331	= -0.331

Example 7 - Final Recommendation Values

The recommendation value is then divided into three separate categories: Do-Not-Recommend (DNR), Caution (CAU), and Recommend (REC). To be categorized as DNR, the recommendation value will be <= -0.1. Conversely, a REC result will be >= 0.1. This leaves "CAU" to be between -0.1 and 0.1.

Figure 4 summarizes the above Examples as our PMT algorithm:

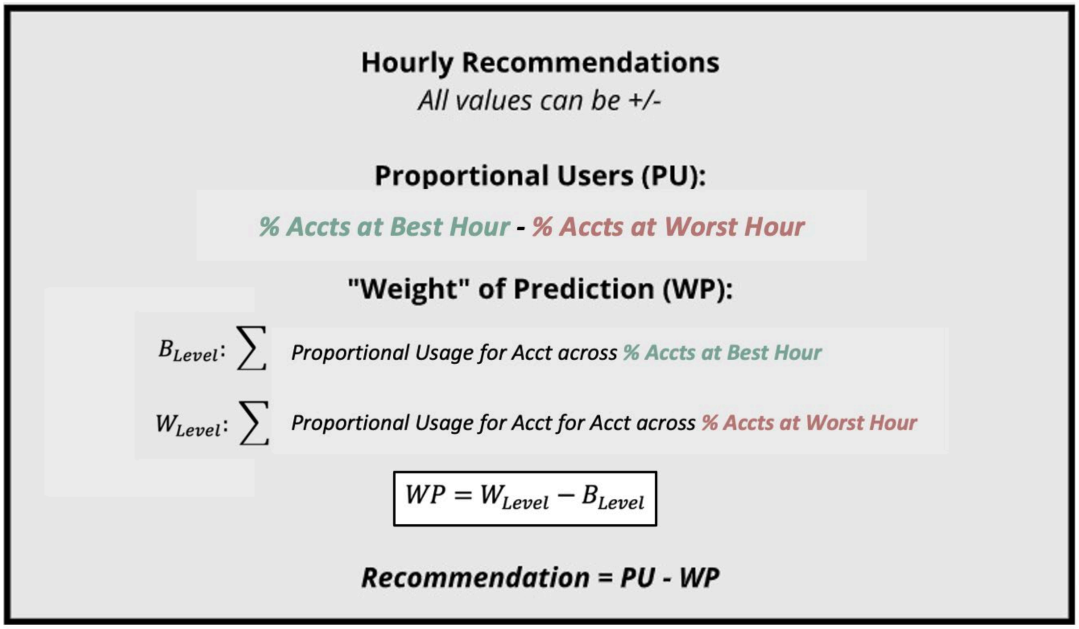


Figure 4 - PMT Algorithm Formula

3. Assessment

The PMT algorithm leverages CUI to find the optimal times when customers served by the same branch of a network have the lowest usage pattern levels relative to other hours of the day. This algorithm assumes that customers will likely experience less service impact during periods of proportionally lower data usage relative to their respective total use across the day. Identifying a collective lower usage pattern helps drive the recommendation towards the most optimal hours for a group of customers whose network branch requires planned maintenance.

3.1. Aim

Since the PMT algorithm is dependent on the CUI, we must determine if the identified CUI has sufficient signal to allow the PMT to make meaningful recommendations.

Specifically, we want to ensure that PMT suggestions of REC, CAU, or DNR for performing an SIE show variation in customer impact (approximated by the volume of CIs). We anticipate that if CUI has sufficient signal, SIEs occurring during PMT's "*recommended*" hours will show less relative CI increase than SIEs performed during PMT "*do-not-recommend*" hours. Otherwise, we would expect that the difference in these relative SIE increases across different PMT recommendations is negligible.

3.2. Method

We took advantage of the availability of historical CUI, CI, and SIE data to perform the assessment. This approach mimics an "idealized" scenario because we already know i) the SIE occurred, ii) the exact set of households impacted by the specific SIE, and iii) have access to the household's corresponding CUI for the same day as the SIE, which is used to derive the PMT recommendation for the hour each SIE occurred. Given that we were planning to trial an application based on the PMT algorithm in one of the regions of our service footprint (aka the 'test region'), we focused our assessment on the same region of interest from February 01 to March 11, and April 14, 2022. This time period avoids date ranges that could potentially be affected by daylight savings in 2022 (13th March (USA) and 27th March (UK)) as well as 2021 winter holiday seasonal effects[3].

A total of 3341 SIEs in the test region occurred during our collected data sample's assessment period of interest. We derived PMT recommendations for each SIE and its set of corresponding households' CUIs for each hour from 6 am to 11 pm ET on the day each SIE took place. The hourly recommendations for the same day as the SIE allowed us to determine if the start of each SIE occurred at an hour the PMT yielded i) REC, ii) CAU, or iii) DNR result.

Categorizing the time of SIEs onset by PMT recommendations allowed us to compare the average volume of CI associated with each set of affected households during the start hour of SIE relative to their corresponding mean CI baseline across these PMT categories of SIEs. We refer to this metric of relative change in the mean volume of customer interactions as our 'delta ratio' ($dRatio_{CI}$):

$$the\ \ dRatio_{CI} = \frac{CI_{SIE_{startHr}}}{CI_{baseline}}$$

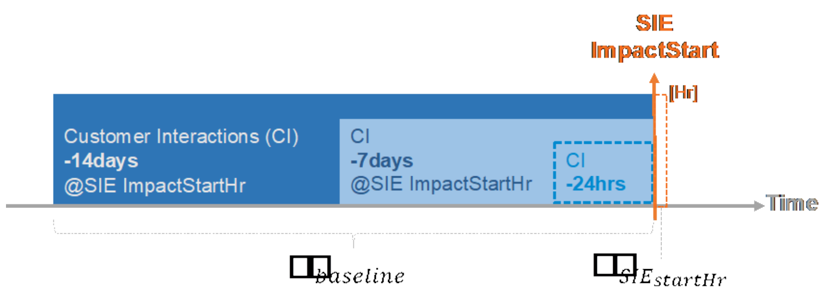


Figure 5 - Three different CI baselines were considered in our assessments.

For each SIE, the start of SIE impact (denoted as SIE ImpactStart) is used as a reference to derive the average CIs from the prior 24 hours, the average CIs from the same first hour since the start of SIE impact (denoted as a dotted orange bar) in the prior 7 or 14 days.

We derived the *delta ratio* metrics using three different CI baselines (see Fig 5): 24hrs prior, seven days prior, and 14 days before the same SIE$_s$. It is helpful to get a sense of the fluctuations in CI in the 24hrs before the SIE, as it can show the potential time-of-day effects of customer interactions (e.g., customers may tend to interact at certain hours of the day). This comparison also prompted us to consider the volume of CI in the last 7 or 14 days during the same period as the hour following the start of the service impact event (*@SIE*). At the same time, this does not fully account for potential day-of-the-week effects (for which we needed a sample of historical CI data going back for more weeks than we were able to

[3] There is a +5hrs (before 13th March) +4hrs (13-26th March) +5hrs (27th March onwards) UTC – EST/EDT difference. While there is potentially no differences in terms of Service Interruption Event (SIE) impactHr timestamps, the associated Customer Interactions (CI) events would include some days with +4/+5 ET depending on number of days or hours (7 or 14days | 24hrs baselines) one looks back relative to the CI data time zone. As such, the date-range we work with is primarily to avoid dealing with daylight savings conversions.

sample at the time of writing). We at least established a prior seven or 14-day baseline for the same period as the hour following each SIE.

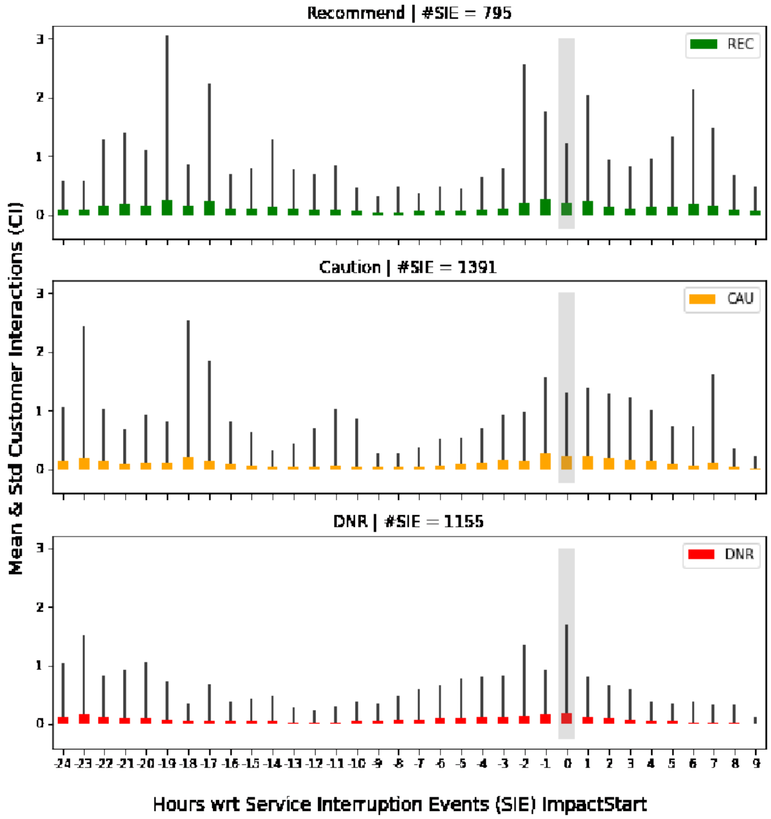
3.3. Findings

3.3.1. Comparison of REC and DNR delta_ratios

We are particularly interested in the difference between DNR and REC delta_ratios. Specifically, if a DNR delta_ratio is more extensive compared to REC delta_ratio, we can infer that the SIE at a PMT REC hour is a less customer "impacting" time.

Figure 6 illustrates how delta_ratio(s) are derived for the three PMT categories of SIEs for the assessment using 24hrs before baseline with all customer interaction types considered. The analytical approach is similarly applied for the evaluations performed using the same SIE for each day in the previous 7 or 14 days as baselines (Figs 9—11).

a)



b)

CI	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
CI All Types	REC	0.119	0.199	1.673
	CAU	0.105	0.238	2.273
	DNR	0.083	0.182	2.200

@SIE_Hr / prior24Hrs

Higher DNR delta_ratio cf. REC delta_ratio indicates successful PMT recommendation of less customer impacting time

Avg # Customer Interactions per hour / SIE over 24 hours prior to SIE impact onset

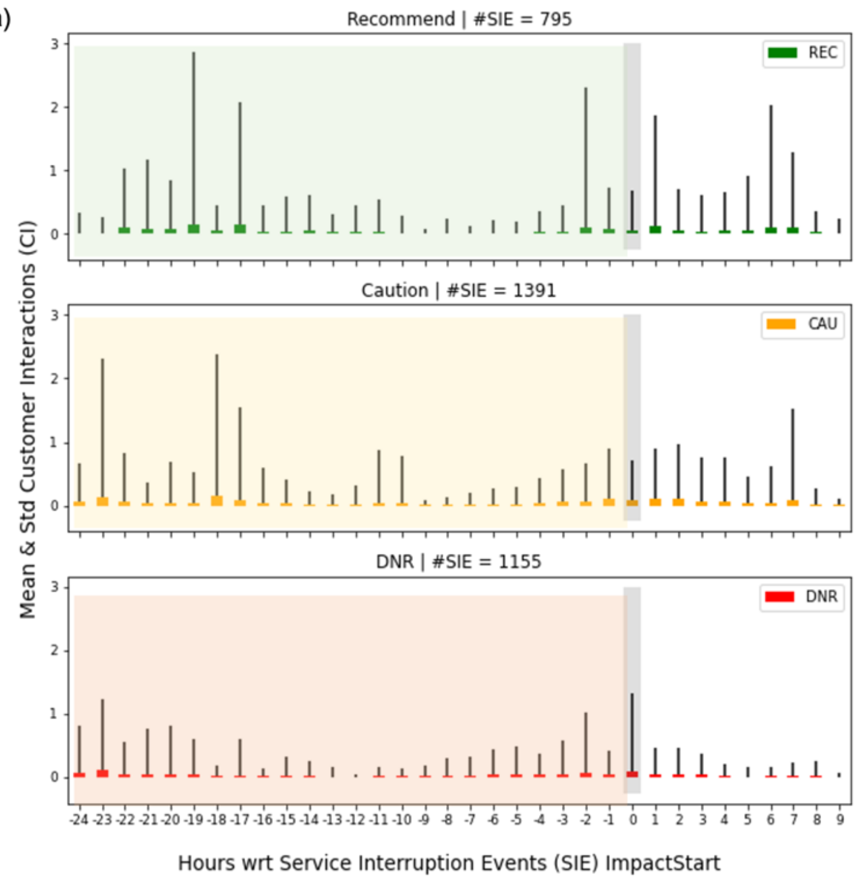
Avg # Customer Interactions per SIE during the hour after the start of SIE impact

Figure 6 - CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with prior 24hrs as a baseline

a) Hourly mean and standard deviation of CI across all SIEs concerning impact start hour (i.e., 1—24 hours prior and 1—9 hours after). All types of CIs are considered. The prior 24 hrs as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 24 hours (prior24Hrs), ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category.

Figures 7 and 8 provide the summary of customer interactions 24 hours before the SIE impact start hour for assessments with sub-types of customer interactions.

a)



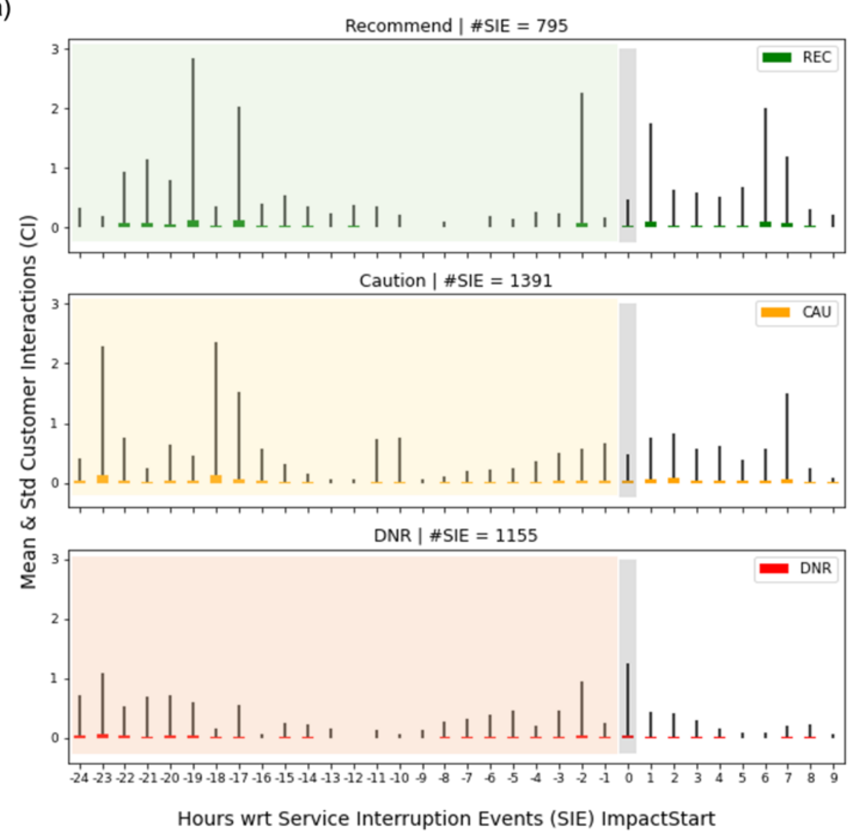
b)

	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
CI Omitting Self-service Types	REC	0.051	0.060	1.174
	CAU	0.045	0.081	1.783
	DNR	0.033	0.082	2.482

Figure 7 - Non-self-service CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with prior 24hrs as a baseline.

a) Hourly mean and standard deviation of CIs across all SIEs concerning SIE impact start hour (i.e., 1—24 hours prior and 1—9 hours after). CIs without self-service event types are considered. The prior 24hrs as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 24 hours (prior24Hrs), ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category.

a)



b)

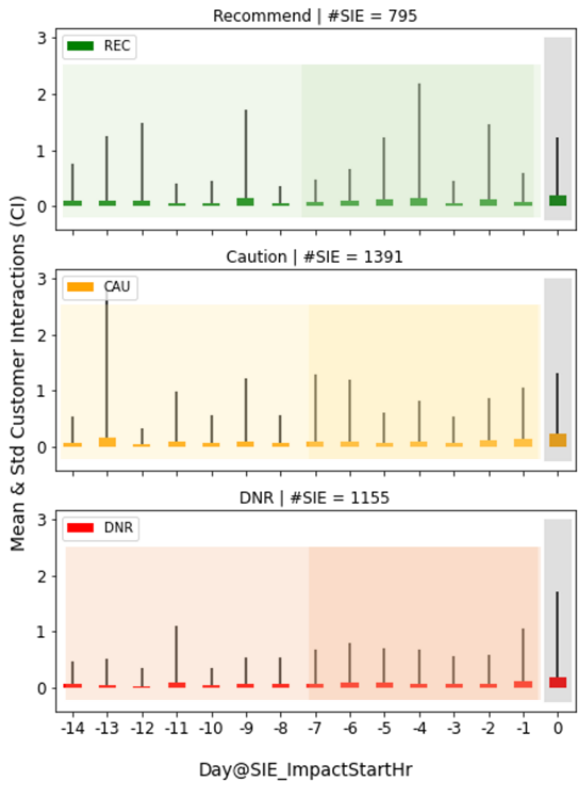
CI Cost Incurring Types	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
	REC	0.036	0.034	0.949
	CAU	0.031	0.035	1.123
	DNR	0.022	0.054	2.399

Figure 8 - Cost-incurring CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with prior 24hrs as a baseline

a) Hourly mean and standard deviation of CI across all SIEs concerning SIE impact start hour (i.e., 1—24 hours prior and 1—9 hours after). Cost-incurring CI types (e.g., technician visit scheduling; repair call/chats) considered. The prior 24hrs as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 24 hours (prior24Hrs), ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category.

Similarly, the baselines of assessments performed using the same SIE in the prior seven days and 14 days are summarized in Figures 9—11.

a)



b)

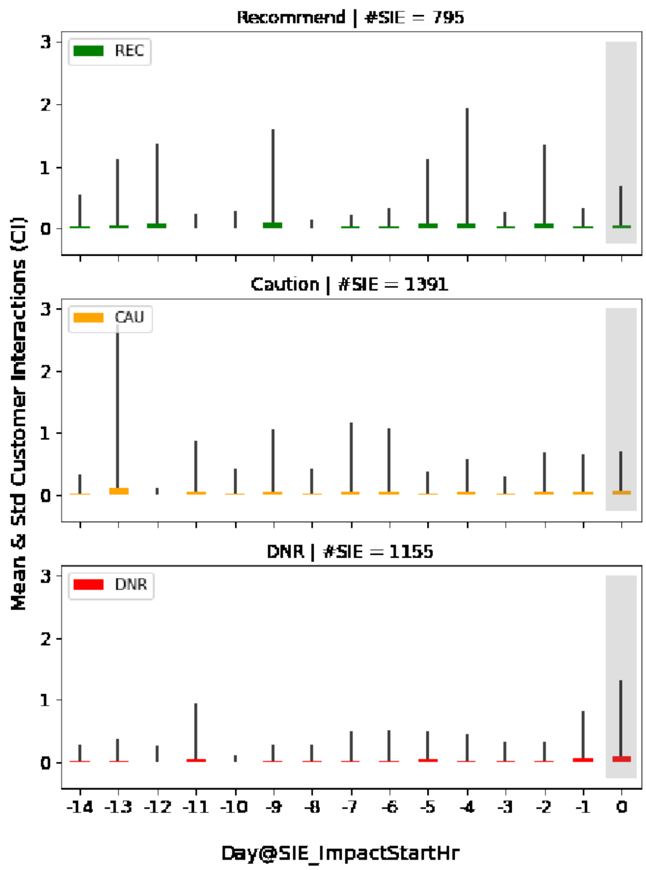
	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
CI All Types	REC	0.099	0.199	2.004
	CAU	0.097	0.238	2.452
	DNR	0.089	0.182	2.047

	SIE category	prior14Days@Hr	@SIE_Hr	delta_ratio
CI All Types	REC	0.093	0.199	2.139
	CAU	0.092	0.238	2.576
	DNR	0.075	0.182	2.428

Figure 9 - CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with SIE_ImpactStartHr in prior 7 or 14 days as a baseline

a) Daily mean and standard deviation of CI (at SIE impact start hour across all SIEs. All types of CIs are considered. The prior 7 or 14 days CI at SIE impact start hour as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 7 or 14 days, ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category.

a)



b)

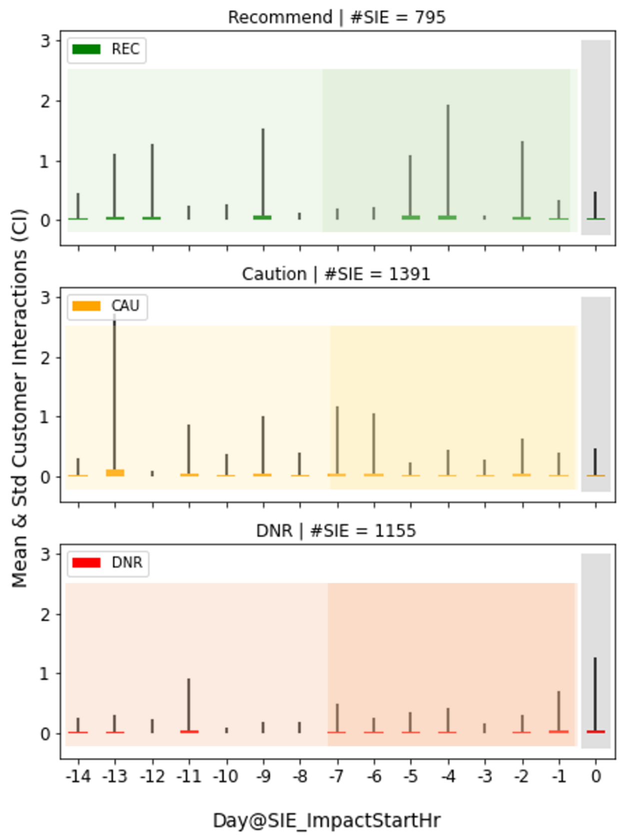
	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
CI Omitting Self-service Types	REC	0.045	0.060	1.333
	CAU	0.044	0.081	1.828
	DNR	0.036	0.082	2.293

	SIE category	prior14Days@Hr	@SIE_Hr	delta_ratio
CI Omitting Self-service Types	REC	0.043	0.060	1.415
	CAU	0.045	0.081	1.798
	DNR	0.030	0.082	2.777

Figure 10 - Non-self-service CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with SIE_ImpactStartHr in prior 7 or 14 days as a baseline.

a) Daily mean and standard deviation of CI at SIE impact start hour across all SIEs. CIs without self-service event types are considered. The prior 7 or 14 days CI at SIE impact start hour as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 7 or 14 days, ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category.

a)



b)

CI Cost Incurring Types	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
	REC	0.036	0.034	0.955
	CAU	0.031	0.035	1.140
	DNR	0.024	0.054	2.249

CI Cost Incurring Types	SIE category	prior14Days@Hr	@SIE_Hr	delta_ratio
	REC	0.034	0.034	1.005
	CAU	0.035	0.035	1.012
	DNR	0.020	0.054	2.704

Figure 11 - Cost-incurring CIs relative to SIE_ImpactStartHr and derived delta_ratios for the PMT categories of SIEs with SIE_ImpactStartHr in prior 7 or 14 days as a baseline.

a) Daily mean and standard deviation of CI at SIE impact start hour across all SIEs. Cost-incurring CI types (e.g., technician visit scheduling; repair call/chats) considered. The prior 7 or 14 days CI at SIE impact start hour as a baseline is denoted with SIE categorical shading, relative to SIE start hour, as indicated by the grey vertical bar; b) Summary of average CI i) across all hourly means of prior 7 or 14 days, ii) for an hour after SIE impact start (@SIE_Hr), as well as the respective delta_ratio for each SIE category

We observed that our assessments across the different baselines (24hrs prior; 7 days prior; 14 days prior) and the various combinations of customer interaction types (e.g., all types; excluding self-service troubleshooting; inclusion of only cost-incurring types) all show a consistent trend: Service Interruption Events (SIEs) performed during PMT "recommended" hours show less relative customer interaction (CI) delta_ratio increase compared to SIEs performed during PMT "do-not-recommend" hours (see Tables 1—3).

Table 1 - Derivation of delta_ratios for comparison across SIE categories using 24hrs prior CIs as a baseline.

a)

CI	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
All Types	REC	0.119	0.199	1.673
	CAU	0.105	0.238	2.273
	DNR	0.083	0.182	2.200

b)

CI	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
Omitting Self-service Types	REC	0.051	0.060	1.174
	CAU	0.045	0.081	1.783
	DNR	0.033	0.082	2.482

c)

CI	SIE category	prior24Hrs	@SIE_Hr	delta_ratio
Cost Incurring Types	REC	0.036	0.034	0.949
	CAU	0.031	0.035	1.123
	DNR	0.022	0.054	2.399

Table 2 - Derivation of delta_ratios for comparison across SIE categories using CIs in the prior 7 days at the same impact start hour as a baseline

a)

CI	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
All Types	REC	0.099	0.199	2.004
	CAU	0.097	0.238	2.452
	DNR	0.089	0.182	2.047

b)

CI	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
Omitting Self-service Types	REC	0.045	0.060	1.333
	CAU	0.044	0.081	1.828
	DNR	0.036	0.082	2.293

c)

CI	SIE category	prior7Days@Hr	@SIE_Hr	delta_ratio
Cost Incurring Types	REC	0.036	0.034	0.955
	CAU	0.031	0.035	1.140
	DNR	0.024	0.054	2.249

Table 3 - Derivation of delta_ratios for comparison across SIE categories using CIs in the prior 14 days at the same impact start hour as a baseline

		SIE category	prior14Days@Hr	@SIE_Hr	delta_ratio
a)	CI All Types	REC	0.093	0.199	2.139
		CAU	0.092	0.238	2.576
		DNR	0.075	0.182	2.428
b)	CI Omitting Self-service Types	REC	0.043	0.060	1.415
		CAU	0.045	0.081	1.798
		DNR	0.030	0.082	2.777
c)	CI Cost Incurring Types	REC	0.034	0.034	1.005
		CAU	0.035	0.035	1.012
		DNR	0.020	0.054	2.704

What is particularly interesting to note is that when omitting self-service CI types or considering only cost-incurring CIs – such as scheduling a technician visit and calls and chats with a customer agent – we observed relatively larger differences in DNR—REC delta_ratios. Although the differences in delta-ratios observed were not statistically significant ($p>0.05$[4]; likely due to data sample sizes), these findings are encouraging because they point to specific CIs that we could potentially monitor and assess for the financial and customer impact on an ongoing periodic basis.

3.3.2. Time-of-day Effects

In addition to deriving and assessing the DNR—REC delta_ratios, categorizing the time of SIEs onset by PMT recommendations also allowed us a view into when different categories of PMT recommended SIEs tended to occur over the period of 06:00 hrs—23:00 hrs in the test region. Figure 12(a) shows the distributions of SIEs for each PMT recommendation category over time.

Additionally, our assessments highlighted a time-of-day effect: hours that the algorithm would recommend tended to occur earlier in the day (06:00-14:00hrs), while the caution hours shifted to later (08:00-15:00hrs), and for hours that PMT yielded a do-not-recommend, we saw the latest (09:00-19:00hrs).

Combining the separate categorical plots in Fig 12(b) as a relative proportion of all SIEs performed during the onset hour of service interruption across the period of 06:00 hrs—23:00 hrs, we can begin to appreciate the time-of-day effects together with the previously described trend in delta-ratios. Specifically, we observe that performing an SIE after 15:00 hr almost always is associated with high customer impact compared to performing an SIE before 08:00 hr when it shows low impact for customers.

[4] Our t-tests were performed with probability (p) significance test against the threshold $a = .05$; we assume and allow for a 5% chance level of how extreme our observed results must be to reject the null hypothesis of no delta_ratio difference. At the set threshold of $p< .05$, an observed test probability p below 0.05, would indicate the alternative hypothesis of a delta_ratio difference is 'statistically significant' and that we could reject the null-hypothesis. In our case we observe that the t-test performed on the delta_ratios derived from our limited sample exceeds the set threshold of acceptable chance level, and we conclude that the observed trend is not statistically significant.

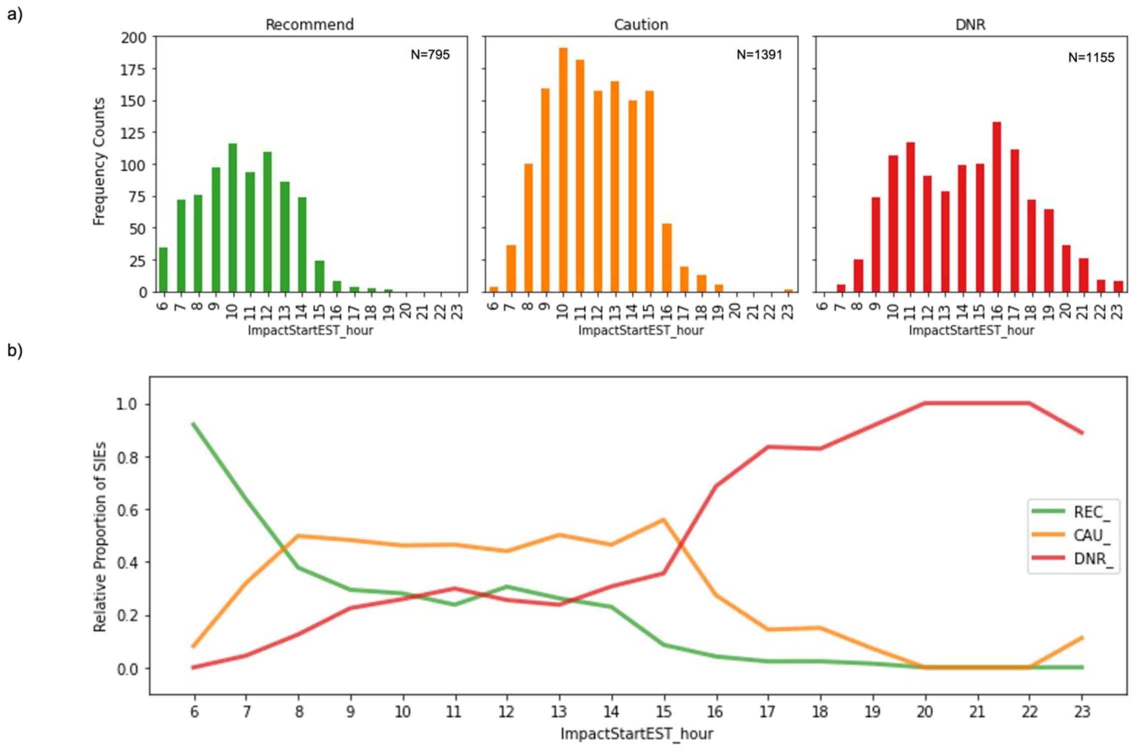


Figure 12 - Distribution and relative proportion of SIEs for each PMT category.

a) Hourly distribution of SIEs for each PMT recommendation category – REC; CAU; DNR – over time; b) Relative proportion of all SIEs in data sample during onset hour of service interruption across the period of EST 06:00 hrs—23:00 hrs.

3.4. Assessment Summary

Overall, the assessments for the test region over the date range highlighted the following:

1) CUI has sufficient signal to drive the PMT algorithm;
2) Different categories of PMT recommendation are associated with a consistent trend: SIEs performed during PMT "recommended" hours show less relative customer interaction (CI) increase (as measured by delta_ratio) compared to SIEs served during PMT "do-not-recommend" hours;
3) Operationally, a simple and effective way to improve customer experience is by retroactively measuring CI for SIEs in defined geographic regions and using the results to provide guidance. PMT recommendation is most effective if it is run:
   i. Only when a planned maintenance job needs to be scheduled (i.e., beforehand, before being in location)
   ii. Derived with CUI of the exact SIE-affected accounts

We reiterate that our data-driven assessment approach mimics an idealized scenario, which may not be easily achieved in the field without a robust data ingest and computational platform. To approximate the idealized system, future customer usage information must be forecasted. This relies on the stability of historical data, which may be affected by seasonal events – an area of research we hope to pursue.

Notwithstanding, the findings discussed provide valuable insights and recommendations for how the PMT could be applied in the field. Importantly, given that the data sampled was over a short period in the current year and for only the selected test region, periodic assessments would be needed to monitor if our observed delta_ratios and time-of-day trends hold across seasons, time, and indeed, if there may be regional differences in such movements.

Next, we discuss practical considerations of applications in the field and how our findings could be best incorporated into deploying an early version of the PMT.

4. Application

An application was developed, for use by technicians, to determine the best time to perform SIEs, specifically for 'planned maintenance' events to be performed during regular and extended working hours. For better or worse, the application was developed with scalability and performance, which means that its function does not match that of the Assessment described above in Section 3. Although much further testing would be required to measure its effectiveness, it helped expose the challenges faced by developing a field tool that would rely on data generated by various corporate systems. A description of the application follows.

4.1 Architecture

4.1.1. Overview

The application is divided into three major systems, including the same Recommendation Service (Section 2.3) developed for the Assessment (Section 3). The additional components include a Grouping Service (Section 4.1.3), designed to help with performance, and an Application Programming Interface (API), where aggregated data and field tech requests were processed. Figure 13 provides an overview of the PMT Application and its different components.

As described earlier, the Recommendation Service uses CUI metrics and groups of selected accounts to calculate recommendations. However, in the PMT application, we created a 'Grouping Service' that pre-determined groups of up to 40 accounts for which recommendations would be calculated. This approach allows the recommendations to be calculated before the technicians need them, thus minimizing query latency, unlike systems that must perform data pulls and calculations on-demand. We provide a summary of this in Figure 14 and Section 4.1.2.

The Grouping Service (Section 4.1.3) consumes data about the structure of the cable plant (see description of PTI in Section 2.1.3) from each CMTS down to each account. It then creates groups of up to 40 accounts on the same network branch. This is then stored in the MySQL database via the API, which is used to fetch this data by the Recommendation Service. These pre-created groups are approximations of areas that could be affected by a representative SIE event based on their connectivity to the cable plant.

The application is built with a RESTful web service that takes requests from the Field Tech's Laptop through a user interface, which helps them generate a list of accounts that would be involved in an SIE. This list of accounts is then sent to the API, which determines the group it best represents, looks up the pre-calculated recommendations for that group, and returns the result to the user interface on the Field Tech's laptop, as illustrated in Figure 13.

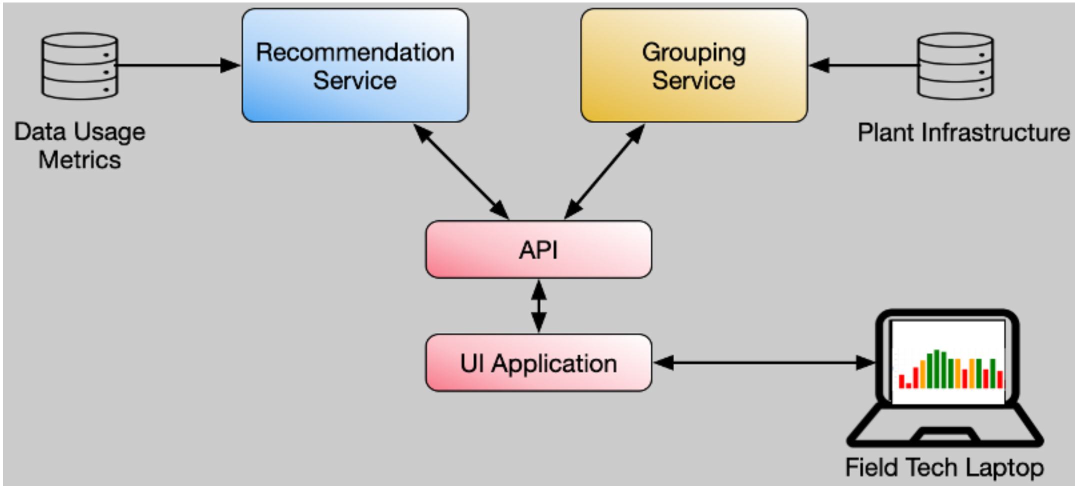


Figure 13 - Overview of PMT Application and its different components.

Detailed descriptions of each of these components are now discussed, except for the Recommendation Algorithm (Section 2.3), which has already been described.

4.1.2. Process Flow

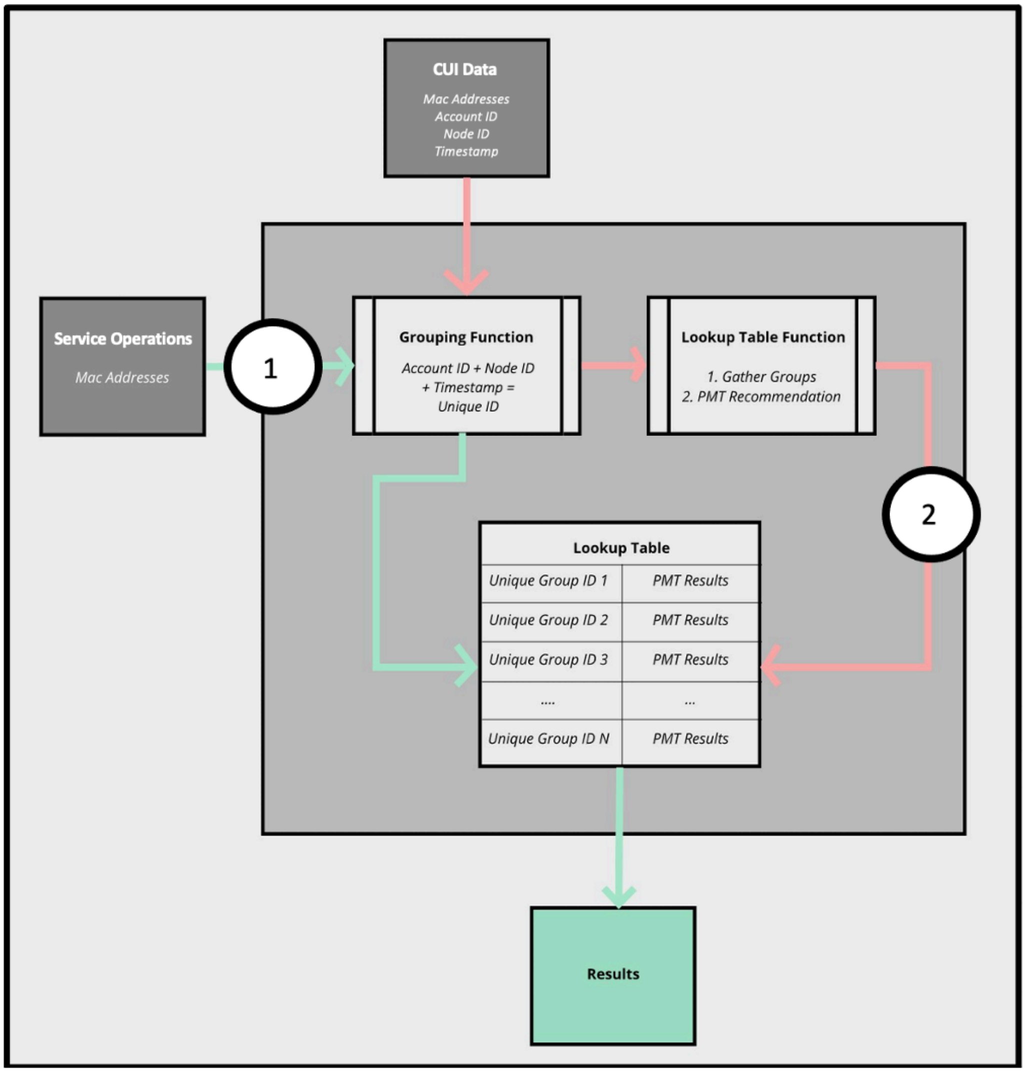


Figure 14 - High-Level Overview of PMT Recommendation Service

Flow 1, described by the **green** arrows, represents a technician's process when scheduling a job. It takes Mac Addressees provided by our Service Operations Tool and then matches them to Account and RF Node IDs. This data then gets passed through the Grouping function to create a Unique Identifier for the group. That Unique ID goes to the Lookup Table, and the PMT Result tied to the Unique ID is returned.

Flow 2, described by the **red** arrows, represents the backend flow where data is processed and prepared for a technician to query. It takes in CUI Data, comprised of Upstream and Downstream rates, and creates a unique ID using the same Grouping function as Flow 1. Then, this ID and CUI Data are passed to the Lookup Table Function, making a PMT Recommendation. The results are then populated into the Lookup Table to await a technician to access the results.

4.1.3. Grouping Service

An assumption was made that given a neighborhood where our services are provided, customers within closer proximity and/or customers who share a pedestal or an amplifier may experience similar plant activity. For example, fiber connects the CMTS to the RF Node in a neighborhood; there would be actives and amplifiers downstream. From there on, there would be taps with up to 8 ports that supply our services for up to 8 households in a residential neighborhood. If the amplifier is faulty, every household connected downstream from that amplifier would be affected. This led to creating groupings of 5-40 accounts.

An internal geographical topology tool, which is built on top of the Geographic Information System Framework (GIS), provides geographical and physical connection information regarding the plant's infrastructure: CMTS, Nodes, Actives, Passives, Amplifiers, Taps, and houses that have been set up to receive our services. Once information regarding a particular CMTS is retrieved, PMT's Grouping Service creates the groups based on common ancestors (such as taps or amplifiers) and geographical proximity within the infrastructure tree. It saves it into a graph database using a graph framework (See Figure 15 below). The last step in the Grouping Service process is to send lists of accounts for all the identified groups and underlying information, including account numbers, MAC addresses, and physical addresses, over to the MySQL database via the API. This information is then pulled into the Recommendation Service via an automated pull.

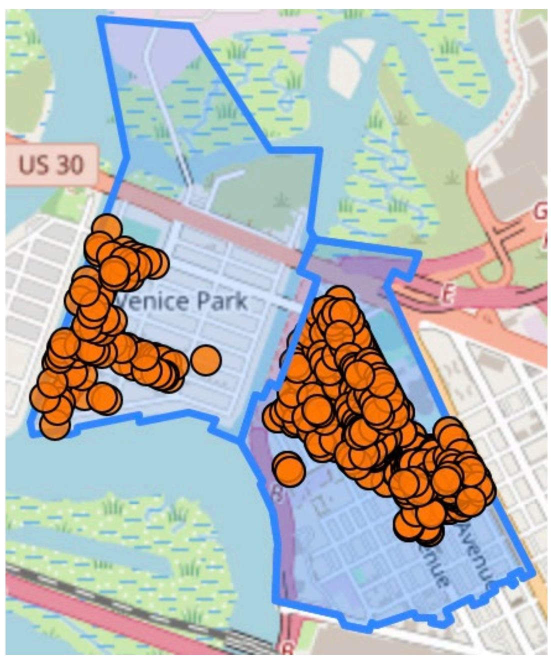


Figure 15 - Example groupings determined by Grouping Service.

Two groups being calculated based on their geographical proximity by the Grouping Service

There are tradeoffs in assuming that the data usage patterns are similar between customers sharing a common ancestor or geographical proximity. It would be more accurate for the PMT application to evaluate each account's statistically considered data consumption when calculating a recommendation for a SIE.

4.1.4. User Interface

Due to expediency and the fact that technicians use a smartphone or their in-truck laptop, we decided to use a browser-based solution. In the PMT application, the UI was created with ReactJS, and the backend was developed via NodeJS, Express API, and MySQL as a database. This was done to provide a visualization of the recommendations to a technician that would be concise and relatively quick to read.

As mentioned in section 4.1.3, the input for the area affected by the SIE is acquired via an internal GIS-based tool in which the accounts within the impacted area are identified and provided to the PMT application. Within the UI, the technician selects future dates (as far as two days in the future) in which they desire to view the recommendations for that area. These inputs go to the API, which determines which accounts belong to which groups via a “majority rules” filter to select which pre-determined group has the most accounts for an assigned area. Using a graph, PMT then fetches the recommendations for the identified group from the MySQL database and displays them on the ReactJS frontend.

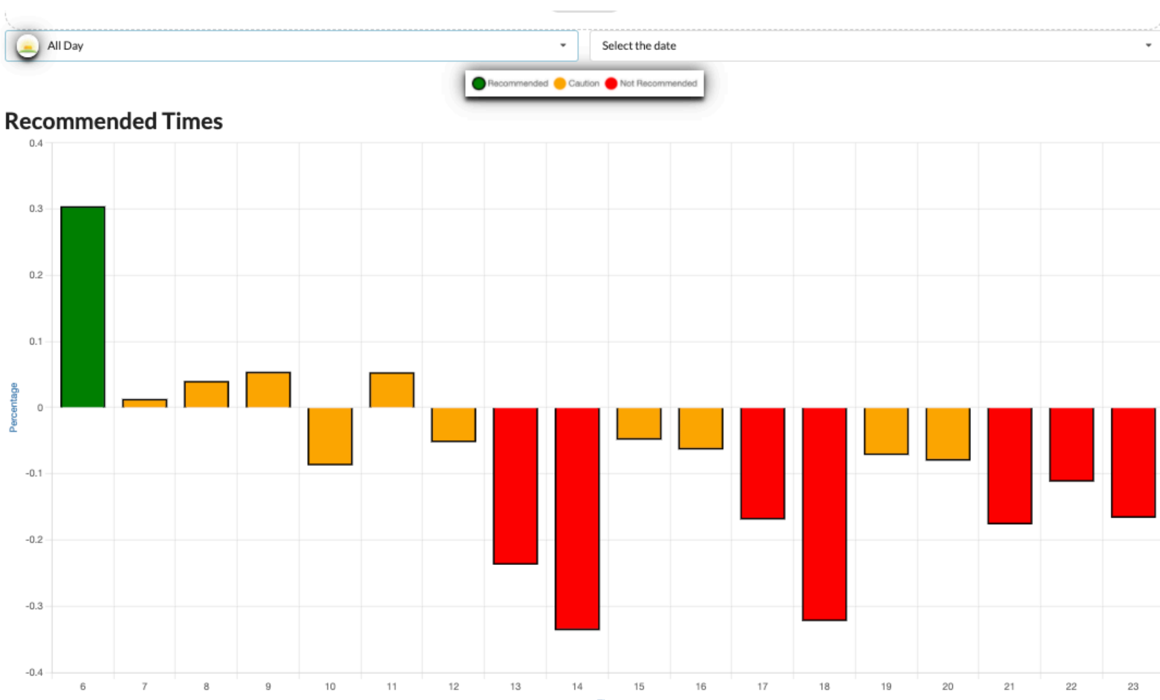


Figure 16 - Example PMT application recommendations

After the accounts are input by the user, PMT would run its microservices and reveal the hourly recommendation based on the hours of the day and the date selected.

Figure 16 shows the displayed recommendations in three color categories: green for REC, yellow for CAU, and red for DNR. In addition, the bar graph’s magnitude is determined by the values calculated by the Recommendation Service. Additionally, a user can look at today’s, tomorrow’s, and the day after tomorrow’s data to find an ideal time to perform an SIE.

5. Discussion

5.1 Learnings from the field and insights from the assessment

We discuss the learnings from our small-scale field trial and the insights derived from assessing our algorithm with historical SIE and CI data.

5.1.1. Trial

The PMT application was trialed in a limited area (0.15% of total nodes in the region), or approximately 30 nodes in a test region with tens of thousands of nodes. While the trial is ongoing, we have already discovered that deployment of the PMT on a fraction of the regional network plant does not generate sufficient data to concretely determine if the use of the developed PMT application improves customer experience.

We note that the assessment described in Section 3 was performed on a sample of a few thousand SIEs across the whole test region. As just mentioned, the application trial covered a small fraction of the test region by comparison. Therefore, to match the integrity of the assessment, the trial would need to be expanded to include the entire region. However, such an expansion would be equivalent to creating a large-scale production-level application deployment, which would require committing resources to an unverified design.

As of this writing, we determined it would be best to extend our assessment approach within the same region, as well as to the other areas, to verify that findings in the test region: scale up in volume of historical SIEs and CIs, and check if they are consistent across the geography of different regions and over time.

5.1.2. Learnings and Observations

We acknowledge that our assessment of derived PMT recommendations, based on historical data, simulated an *idealized* scenario wherein customer usage information was extracted for the set of SIE-affected accounts and all hours on the same day (that is still unfolding) when the SIE occurred. In a real-world scenario, this application would have to reliably and efficiently estimate future customer data used to calculate recommendations for any set of accounts and any day. Building such an application will require:

- Reliable forecasting of future customer data usage from historical CUI;
- Adequate computing and data ingest resources to continuously store, fetch and compute the stream of available CUI data;
- Adequate computing resources to calculate and return recommendations, at scale, in a performant fashion when a field technician makes a request;
- A system to continuously monitor system effectiveness to provide feedback for improvements.

A noteworthy finding is the time-of-day pattern of SIEs and their associated PMT recommendation category observed in our assessment: SIEs with PMT REC hours tended to occur earlier in the day. In contrast, an SIE during hours that PMT would yield CAU or DNR occurred at higher frequencies later in the afternoon. Field operations groups can immediately use this time-of-day insight to guide them to make sensible choices and possibly even schedule directives on when to perform investigations and preventative maintenance that results in a SIE.

5.1.3. Considerations

In the future, it may be prudent to store data about SIE's impact on customers for an entire year to understand annual patterns. Correlations could be based on locality, especially given the different schedules that universities, K-12 schools, and local holidays, among other events, will affect data usage. For example, university towns have transient populations that drop significantly between terms. K-12 schools have schedules imposed by local governments and are typically published before each school year. Knowing when children are not going to be at school means knowing that they're probably going to be at home sharing bandwidth with parents who may be working from home.

Moreover, events like the COVID pandemic have significantly impacted the use of residential gateways, making knowledge of data usage patterns even more critical as livelihoods have been, and continue to be, made from home offices. Although epidemics and pandemics do not have a known schedule, local health directives and infection rates could be studied and used for predicting increases in residential data usage.

6. Conclusion

Insights from our assessments can be used to recommend how a PMT user application could be architected, deployed, and used in real-time to guide technicians to the optimal time to schedule an SIE. While it should be cautioned that the data used in our assessments were limited in sample size, within one northeastern region in our national footprint, and over a limited time range (February 01 – March 11 and April 16th onwards), the findings we observed suggest that customer usage information could be a good indicator of potential customer impact during SIEs.

Observing trends using the assessment method described in section 3 can help simplify the deployment of the PMT application across our service footprint. It offers a way to standardize its use and potentially roll out randomized field trials across multiple regions for a more robust assessment of the value of using the PMT application in the field.

We also know that the recommendation algorithm may not always yield the expected trend during the year. This could be due to many factors, including:

- seasonal customer data usage patterns
- customer experience campaigns rolled out in parallel
- changes in data usage patterns due to pandemics or natural disasters

Overall, we learned that it helps to employ "hindsight," using retrospective analyses as we did in assessing our algorithm with historical SIE and CI data. We suggest this as good practice before building a field application for trial. The insights derived from a data-driven analytical approach can help guide how such an application would be best used and deployed. Since there is no way to predict which nodes in our network will require the most SIEs in the future, we determined retroactive data was the most efficient way, albeit idealized, to validate our assumption that customer usage data at particular times during the day would be indicative of customer impact.

In conclusion, we entered this trial with the premise that the best time to interrupt the network with a SIE is when nobody is using it. This premise was assessed using the methodologies described in this paper. The measurable reduction of CIs during SIEs that occurred during times recommended by the PMT indicates that a data-driven model to predict the best times to conduct invasive maintenance is possible and deserves further development.

Abbreviations

API	Application Programming Interface
CAU	Caution
CI	Customer interactions
CMTS	Cable Modem Termination System
CUI	Customer Usage information
DNM	Device to Network Mapping
DNR	Do Not Recommend
DOCSIS	Data Over Cable Service Interface Specification
ERSI	Environmental Systems Research Institute
GIS	Geographic Information System Framework
PMT	Planned Maintenance Tool
PTI	Plant Topology Information
PU	Proportional Users
REC	Recommend
SIE	Service interruption events
WP	Weight of Our Prediction

Acknowledgments

Many thanks for the support and contributions made to this project by our esteemed Comcast colleagues:

James Cortese, who joined billing data to plant topology using a combination of character matching and geolocation methods used in the Grouping Service.

Mai Lam, who provided project management support by driving project schedules, authoring status reports, and coordinating meetings for a project and team that changed course many times during its journey.

Susan Owens, for editing services on her first project with our team.

Peter Lester, who helped to refine the PMT assessments and asked critical questions.

