

## **Validating Access Network Maps for Plant Extensions and Capacity Upgrades**

A Technical Paper prepared for SCTE by

**John Schlack**

Distinguished Engineer  
Comcast Cable  
1800 Arch St  
Philadelphia, PA 19103  
+1 267-592-8063  
John\_Schlack2@comcast.com

**Rina Hayashi**

Senior Director, Product Management  
Comcast Cable  
1800 Arch St  
Philadelphia, PA 19103  
+1 267-562-9437  
Rina\_Hayashi@comcast.com

## Table of Contents

<b>Title</b>	<b>Page Number</b>
Table of Contents .....	2
1. Introduction.....	4
2. Pre-Vet Data Collection and Analysis .....	5
2.1. System Overview .....	5
2.2. RF Network Graph .....	8
2.3. Linking Addresses to the Graph.....	10
2.3.1. Addresses Inside Node Boundary .....	11
2.3.2. Addresses for RF Taps Outside Node Boundary .....	13
2.3.3. Addresses Associated with Node Segments .....	13
2.3.4. Linking Addresses to RF Network.....	15
2.4. Linking Devices to RF Network Graph.....	18
2.5. Data Aggregation .....	19
2.6. Bus Leg to Node Segment Identification.....	22
2.7. Opportunities for Map Improvement .....	23
2.7.1. Unconnected Subscriber.....	23
2.7.2. Address Not Serviceable .....	23
2.7.3. Unconnected RF Tap.....	23
2.7.4. Incorrect Node Segment.....	23
2.7.5. Node Segment to Bus Leg Mismatch .....	23
2.7.6. Outside Node Boundary.....	24
2.8. Pre-Vet Analysis Results.....	24
3. Maps and Data Output .....	24
4. Conclusion.....	28
Abbreviations .....	29
Bibliography & References.....	<b>Error! Bookmark not defined.</b>

## List of Figures

<b>Title</b>	<b>Page Number</b>
Figure 1 - Simplified Plant Map.....	4
Figure 2 - Adding Serviceable Addresses and Device Counts.....	5
Figure 3 - Using Device Locations to Identify Map Improvement Opportunities.....	6
Figure 4 - Pre-Vet System Diagram.....	7
Figure 5 - Example RF Network Graph.....	9
Figure 6 - Correlating RF Network Graph with Drafted Network .....	10
Figure 7 - Node Segments to Bus Leg Association .....	11
Figure 8 - Address Locations within Node Boundaries.....	12
Figure 9 - Drafted Supports and RF Taps Relative to Service Addresses .....	15
Figure 10 - Address to Support/Tap Assignment by Proximity and House Count.....	16
Figure 11 - Example Data Model to Associate Addresses to Supports/Taps.....	18
Figure 12 - Address Locations in RF Network Graph .....	18
Figure 13 - Aggregating Device Counts at the RF Output Port of Amps and Splitters.....	19
Figure 14 - Depth First Search to Calculate Device and Homes Passed Counts .....	20

Figure 15 - Map of Plant with Bus Legs Highlighted and Address Errors..... 25  
 Figure 16 – Device and Homes Passed Counts at Amp in Node Boundary View..... 26  
 Figure 17 – Device and Homes Passed Count at Amp (Focused View)..... 27  
 Figure 18 - Bus leg table..... 27

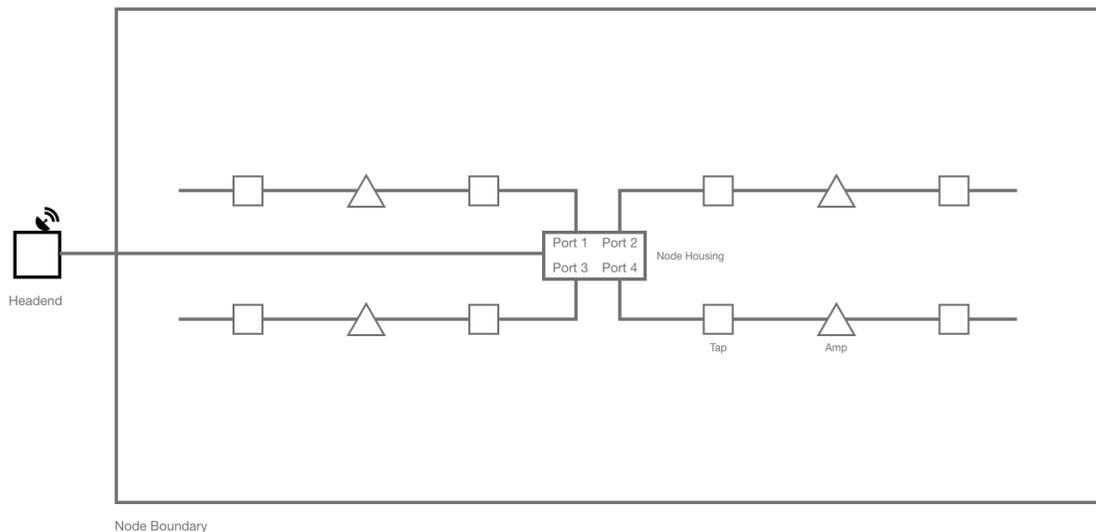
### List of Tables

<b>Title</b>	<b>Page Number</b>
Table 1 - Example Port to Port Connectivity from RF Flood Trace.....	8
Table 2 - Example Service Address Information <sup>1</sup> .....	13
Table 3 - Example Node Segments for HSD Devices <sup>2</sup> .....	14
Table 4 – Example Device and Homes Passed Count by Aggregator <sup>3</sup> .....	22

## 1. Introduction

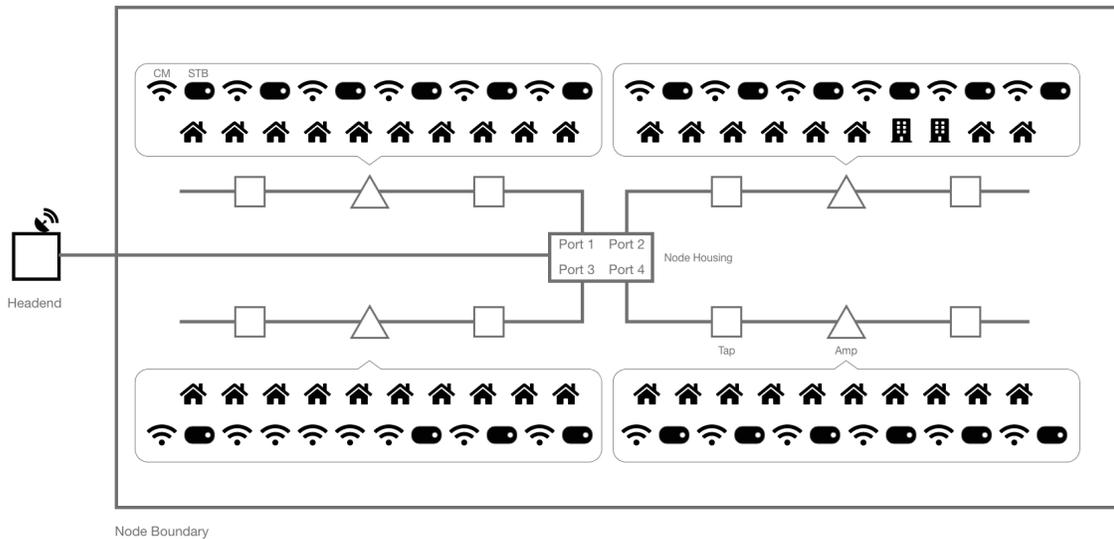
Operators of Hybrid Fiber Coaxial (HFC) have a long history of delivering continuous advancements in speed, capacity and performance. With the rise of DOCSIS 4.0 and 10G, operators are poised to take the next huge leap in delivering faster speeds, greater capacity, advanced security, and improved reliability. Anytime we update the HFC plant, we must conduct a thorough analysis of the drafted plant to decide whether the existing maps are accurate and if so, where and how plant upgrades should be made. This document will refer to this analysis as “pre-vet.” While HFC plant maps are a reliable source for beginning a pre-vet, changes to the area (new streets, buildings, plant repair, etc.) may not always make it into the plant map in time for the analysis. Layering additional data onto the plant maps can further enrich the data and increase the level of accuracy for the designer to make an informed decision on where capacity upgrades and plant extensions should take place.

The pre-vet process begins by capturing the node boundary of the existing plant map that has been documented. **Figure 1** shows a very simplified example of an HFC plant map. Information such as street addresses, parcel size, and where the node and other HFC equipment is located is represented on the map. While this helps show where the equipment is and the type of equipment used, it does not address capacity concerns on its own.



**Figure 1 - Simplified Plant Map**

Next, as shown in **Figure 2**, serviceable addresses and device counts connected to the CMTS are added to the plant map. The combination of the two data points indicates to the designer where the strain is on the network. By aggregating this information at the amplifier (amp) level, the designer can pinpoint where on the bus leg to place a new node to alleviate the strain on the old node. In addition, this level of data can identify gaps in the plant maps. For example, each leg should serve around 50-500 homes. If the aggregated data shows more than that it is a clear indication that either the plant map requires adjustment, or the serviceable addresses associated with the plant requires investigation regarding accuracy.



**Figure 2 - Adding Serviceable Addresses and Device Counts**

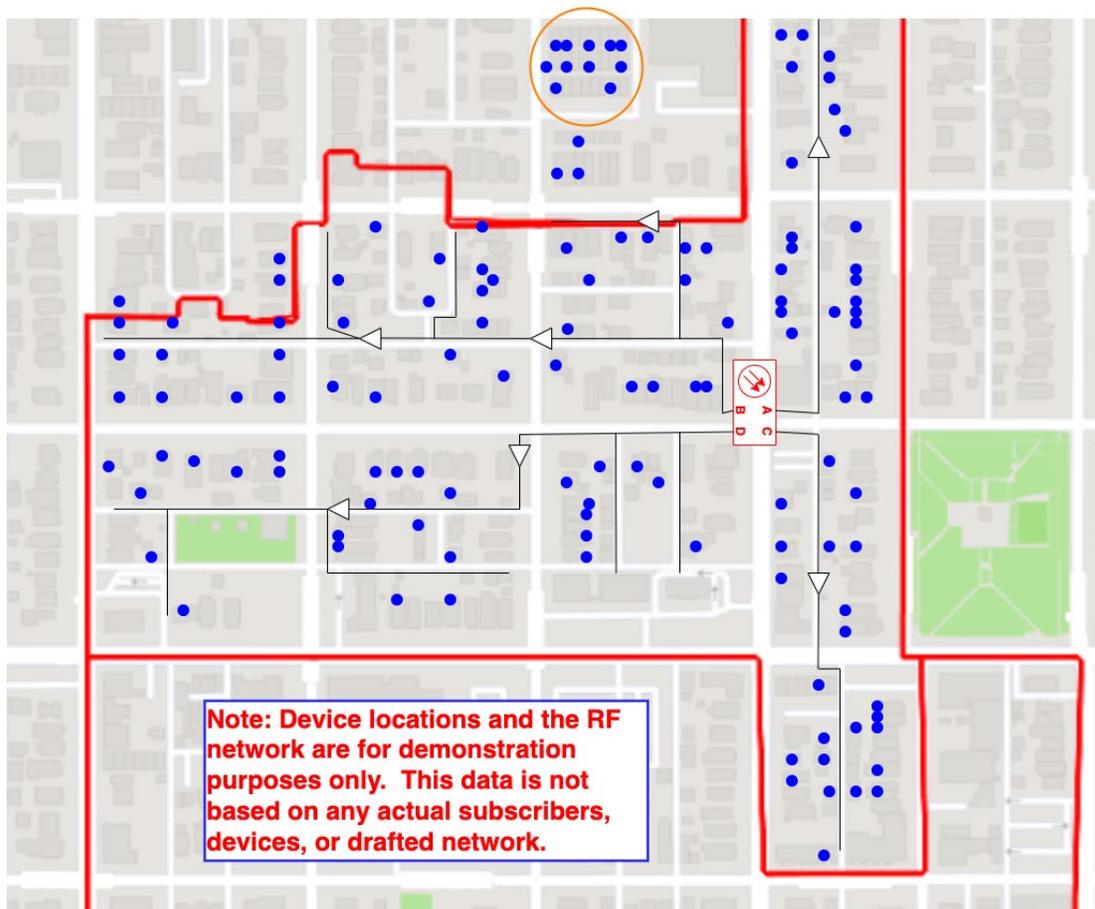
Historically, creating a pre-vet has been a manual, labor-intensive process, taking around 2-8 hours to create the package for a single node housing. As the need increases for upgrading the HFC plant, it will be pertinent to make this process more efficient by automating the aggregation of the serviceable addresses and device counts and visualizing the data on the plant map. This paper will discuss the data collection, analysis, and automation of creating a pre-vet package.

Note that the network maps and data shown in this paper do not include any names or personal data about subscribers.

## 2. Pre-Vet Data Collection and Analysis

### 2.0. System Overview

Subscriber service location (address) and device network communication location correlated against the drafted RF network map provides clues about the accuracy of the maps. Clusters of service locations that cannot be connected to the drafted network indicate opportunities for improving the drafted map. The example network shown below (which is NOT actual data) shows a cluster of service locations that cannot connect to the drafted network. An orange circle highlights these service locations. Notice that some service locations below the circled addresses reside outside the node boundary, but within acceptable distance to the drafted network.



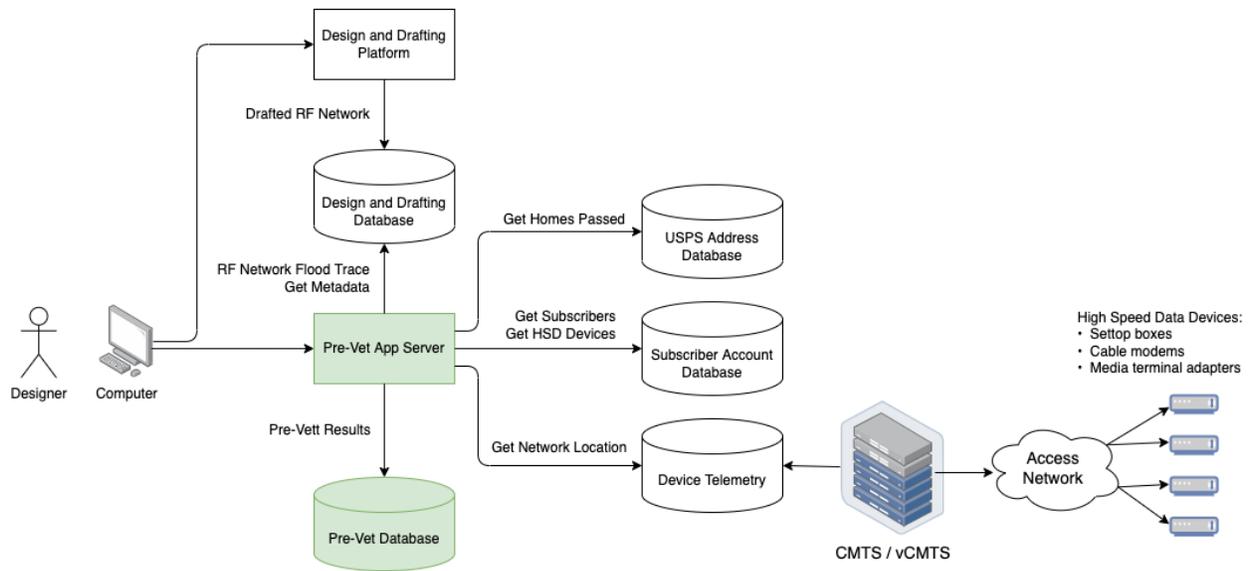
**Figure 3 - Using Device Locations to Identify Map Improvement Opportunities**

Once the drafted RF network map has been validated, the pre-vet process calculates homes passed and actual device counts at key locations throughout the RF network map to aid the designers in node segmentation or node split analysis.

To perform an automated pre-vet analysis of a node housing, data must be collected, correlated, and analyzed from several sources including:

- Design and drafting platform;
- USPS address database;
- Subscriber account database; and
- Device telemetry.

The figure below shows an example system for performing the pre-vet operation.



**Figure 4 - Pre-Vet System Diagram**

The design and drafting platform maintains the physical details of the support structure, cables, and equipment for the access network. This platform maintains connectivity information between elements, provides maps of the access network, and performs power and loss calculations. The drafted data may link the access network equipment to service address locations, but this linking is inconsistent and may be outdated.

The USPS address database contains the official USPS addresses and their geolocations (latitude and longitude). This database minimally must contain addresses for premises passed by the MSO access network.

The account database stores the subscriber account information, including service address and list of devices provisioned to the account. Each subscriber account links to a service address.

Device telemetry identifies the presence of communication with high-speed data (HSD) capable devices, such as cable modems (CM), set-top boxes (STB), and media terminal adapters (MTA). Telemetry data provides the set of ports on the cable modem termination system (CMTS) or remote PHY device (RPD) where messages from the HSD devices are received. The set of subscribers that communicate upstream to a specific CMTS or RPD port are described in this paper as a "node segment."

The steps to collect and correlate this data include:

- Build a graph representation of the RF network for a specific node housing.
- Retrieve the service location addresses based on the node boundary.
- Determine node segments if not already known.
- Retrieve service location addresses associated with node segments.
- Associate service locations to RF taps.
- Count homes passed and devices at amplifiers and splitters.
- Identify service address or network gaps.

Details for these steps appear in the following sections.

## 2.1. RF Network Graph

To analyze the RF network for a given node housing, one needs to understand the relationship between the equipment and cables downstream of the node housing. The RF network shall be represented as a graph data structure, which is a series of vertices and edges. A graph provides efficient mechanisms for determining connectivity between elements and also provides searching capabilities. Some example uses for the RF network graph include counting devices serviced by each amplifier and determining max amplifier cascade by bus leg.

Vertices in the graph represent the cables and equipment (node housing, RF cables, taps, amplifiers, splitters, etc.) while edges represent connectivity between the vertices. Notice that the edges do NOT represent cables. Edges are connectivity. A vertex associated with an RF cable will have two edges, representing other RF equipment or cables connected at each end of the cable in question. The vertex for a node housing will contain 4 edges, one per bus leg (RF output port). For the RF network analysis, the fiber network connecting to the node housing does not need to be represented in the graph. The RF network graph will have a root node that is the node housing.

Building the RF network graph is dependent on the design and drafting platform used by the MSO. In general terms, performing an RF flood trace on each bus leg of a node housing will specify a list of elements and connected ports. An example of port connectivity is shown below.

**Table 1 - Example Port to Port Connectivity from RF Flood Trace**

Source			Destination		
Type	ID	Port	Type	ID	Port
CoaxCable	DMTARCMC1853065972850	N/A	Splitter	DMTARPMC18536598	Input
Splitter	DMTARPMC18536598	Output-1	CoaxCable	DMTARCMC1853065991823	N/A
Splitter	DMTARPMC18536598	Output-2	CoaxCable	DMTARCMC1853065964851	N/A
CoaxCable	DMTARCMC1853065991823	N/A	Tap	DMTARTMC18536621	Input

Notice that port level connectivity means that most RF equipment or cables will have multiple entries in the RF flood trace details – one for each input and output port. The method to build the graph must account for this duplication so that only a single vertex is created for each RF element. Ports identified in the RF flood trace shall be annotated in the edges of the graph.

The graph building algorithm will also create a hash map that maintains the equipment ID to vertex relationship. This will be called the equipment map. The equipment map provides a method to quickly find the vertex of any RF element without requiring a graph search.

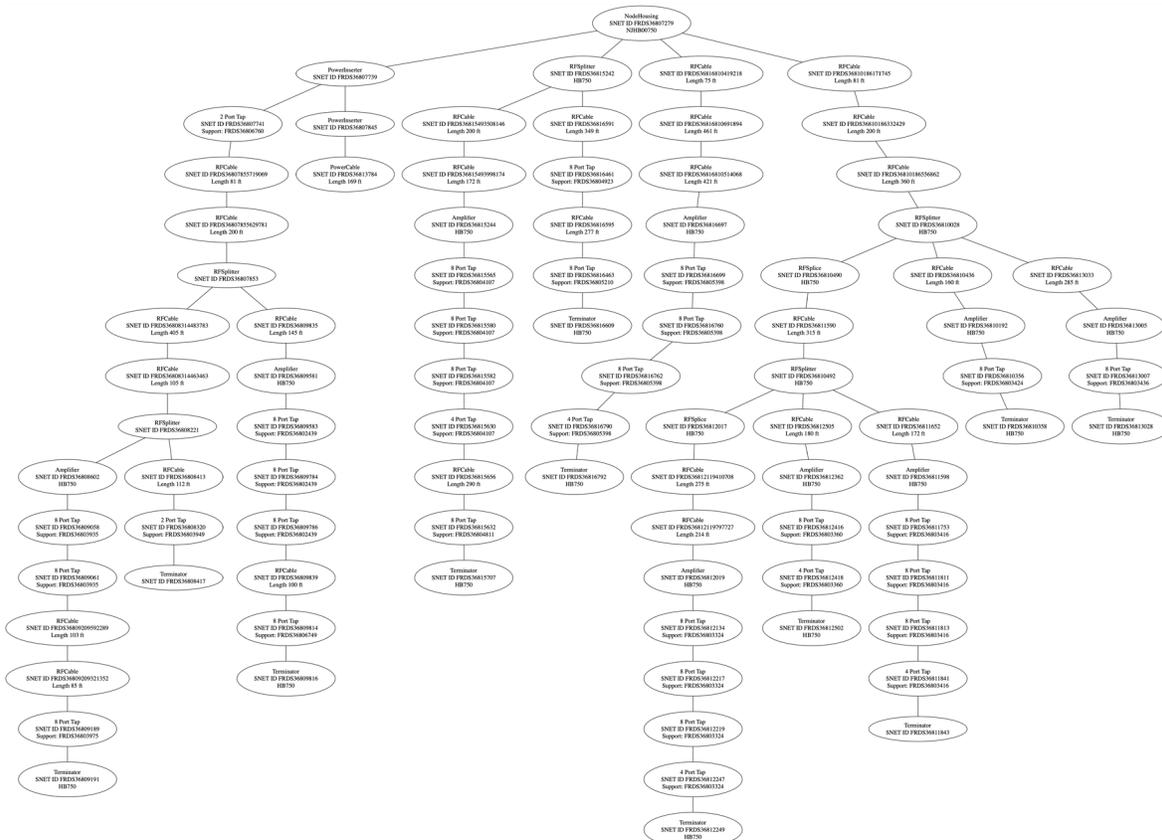
The algorithm to create the RF network graph and equipment map is specified below:

- Create vertex for the node housing and add the vertex as the root of the graph.
- Load the node housing metadata from the design and drafting system.
- Add node housing metadata to vertex.
- Add the ID and vertex of the node housing to the equipment map.
- For each RF output port on node housing:
  - Perform RF flood trace on RF output port; and
  - For each connection pair in RF flood trace:
    - Search equipment map using the equipment source ID and destination ID from the connection pair to find the equipment vertices.

- For each vertex not found:
  - Create vertex for the equipment and add to graph;
  - Load equipment metadata from the design and drafting system;
  - Add equipment metadata to vertex; and
  - Add the equipment ID and vertex to the equipment map.
- Create an edge between equipment vertices.
- Add the port IDs and port types of the equipment to the edge.

Port types can include coax ports, power ports, and drops. Tracking the port type is important since the pre-vet analysis will not analyze power and thus power ports can be ignored. Although shown in the figure below, the power port edges can be pruned from the graph.

An example RF network graph is shown below for node housing “NJHB00750”. Each of the four bus legs can be seen connected to the root node of the graph, which is the node housing.



**Figure 5 - Example RF Network Graph**

To better visualize the correlation of the RF network graph with the drafted plant data, the figure below focuses on bus leg A from the figure above. The below figure shows a side-by-side comparison of the RF network graph and the drafted network. Each RF cable or RF equipment element is represented by a

vertex in the graph. Notice that the power supply is missing from the graph. This may be a drafting issue indicating that the power cable running to the power supply is not actually connected.

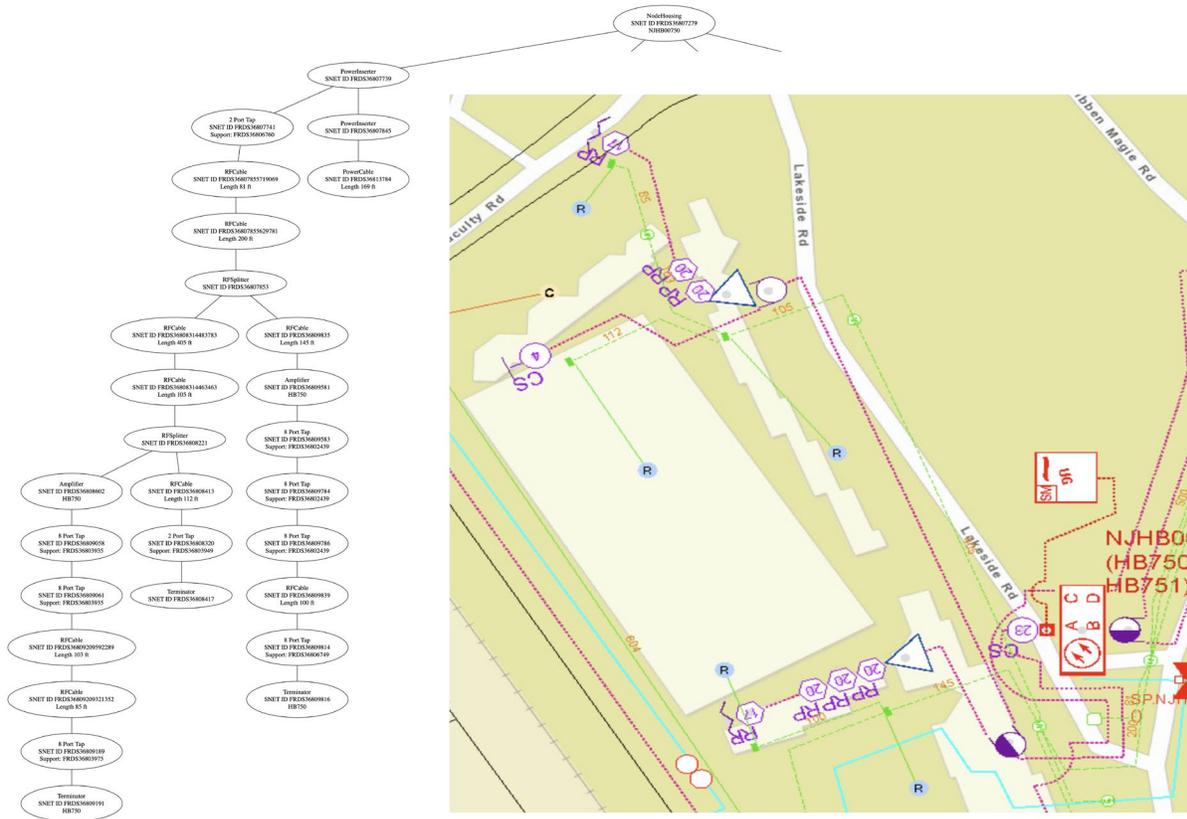
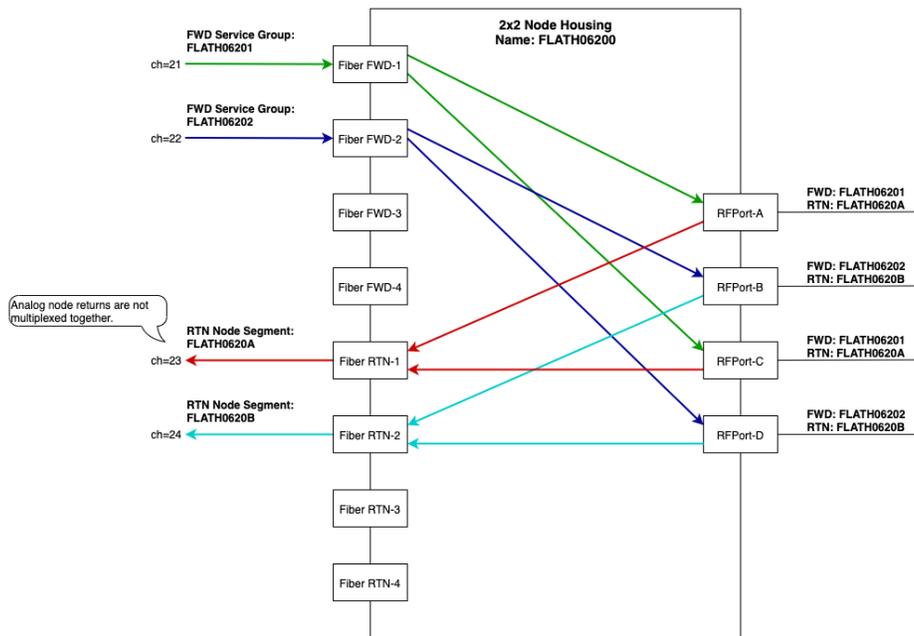


Figure 6 - Correlating RF Network Graph with Drafted Network

## 2.2. Linking Addresses to the Graph

An important method to evaluate the accuracy of the drafted RF network is to examine the subscriber locations and the high-speed data (HSD) capable devices on the network. HSD devices communicate with the cable modem termination system (CMTS). The upstream communication for a set of subscribers – the “node segment” – is typically on a specific CMTS port for a physical CMTS or remote PHY device (RPD) port in the case of a virtual CMTS (vCMTS). A node segment may be comprised of one or more bus legs from the node housing. The figure below shows an analog 2x2 node housing, with two upstream node segments.



**Figure 7 - Node Segments to Bus Leg Association**

The node segment for an HSD device can be determined when the device communicates with the CMTS or RPD. Through the provisioning process, HSD devices are linked to subscriber accounts. These accounts have service address locations. Under ideal conditions, the drafted network includes drops that link the RF taps to drafted address locations. These drafted address locations exactly match the service locations assigned to subscribers and their associated provisioned devices. Thus, a direct link can be made from the drafted RF tap to the subscriber HSD devices and the associated node segments. Unfortunately, there are several factors that could prevent that direct association:

- Constructed network has changed since the design was drafted (e.g. plant extension or capacity upgrade).
- No drop was drafted to link the RF tap to the address location.
- Address has changed since the network was originally drafted (e.g. street renamed).
- Originally drafted address was not the official USPS address, but a place holder for new construction (e.g. drafted network has lot numbers instead of USPS addresses).

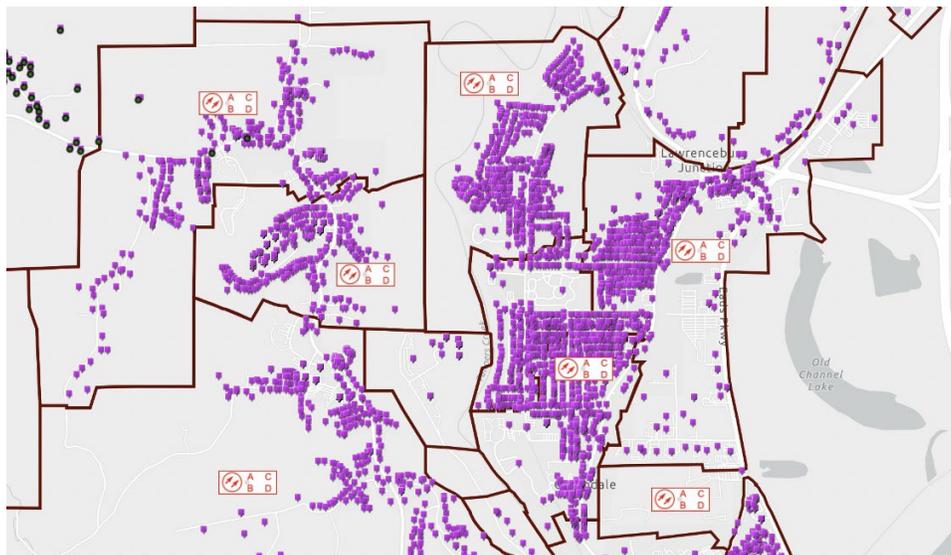
The pre-vet analysis will perform several steps to overcome these conditions. These steps include querying address locations based on the node boundary, querying addresses associated with node segments of devices, and linking address locations to RF taps by proximity.

### **2.2.1. Addresses Inside Node Boundary**

Designers draft boundaries to indicate the coverage area for each node housing. Most address locations serviced by the RF network of a node housing reside within the drafted boundary. Therefore, querying the drafted boundary provides most of the service addresses needed for the network analysis. Note that depending on the address database that the MSO queries, the address locations may be homes passed (e.g.

USPS address database) by the RF network or subscribers' service addresses (e.g. subscriber account database).

Designers may draft distinct types of node boundaries including node housing, node segment, and bus leg boundaries. One or more bus leg boundaries should comprise the node segment boundary. Node segment boundaries comprise the node housing boundary. Only one type of node boundary needs to be queried to retrieve the service addresses associated with the node housing. Querying the service address locations within the node's drafted boundaries and associating those addresses with active devices will allow discovery of the node segments associated with the node housing. The figure below shows address locations of homes passed (not subscribers) for a set of node housing boundaries.



**Figure 8 - Address Locations within Node Boundaries**

An example Oracle query to retrieve the service address locations of homes passed within a node boundary is shown below:

```

SELECT A.ID, A.STREET_NUM, A.STREETNAME, A.UNIT_TYPE, A.UNIT_VALUE, A.CITY,
       A.STATE, A.POSTALCODE, A.LATITUDE, A.LONGITUDE
FROM ADDRESSES A
WHERE
SDO_RELATE (
  A.LOCATION,
  SDO_GEOMETRY(
    2003,
    4326, -- WGS84 coordinate system
    NULL,
    SDO_ELEM_INFO_ARRAY(1,1003,1), -- one polygon (exterior polygon ring)
    SDO_ORDINATE_ARRAY(
      <LIST OF COORDINATES>
    )
  ),
  'mask=anyinteract' ) = 'TRUE'

```

where <LIST OF COORDINATES> are the node boundary polygon coordinates. Coordinates are specified as longitude and latitude pairs.

Example:

```
SDO_ORDINATE_ARRAY (
  -87.28482973277373, 37.25299509730186,
  -87.27747974176278, 37.25299509730186,
  -87.27747974176278, 37.24625992935859,
  -87.28482973277373, 37.24625992935859,
  -87.28482973277373, 37.25299509730186 )
```

Note that some addresses in this query may not actually be serviced by the node housing in question. Drafted node boundaries are not precise and RF networks from different node housings may overlap such that it is extremely difficult to draft boundaries that are completely distinct. Additional processing of the data, discussed below, will isolate and remove address locations associated with other node housings.

Fields returned in the query include a unique ID assigned to the address, street number, street name, unit type and value (for multi dwelling units), city, state, postal code, latitude, and longitude. The latitude and longitude are based on the geocoder for the address database, not the drafted location.

**Table 2 - Example Service Address Information<sup>1</sup>**

ID	Number	Street	Unit Type	Unit Value	City	State	Zip	Latitude	Longitude
2198101	101	Main St.	Apt.	A	Philadelphia	PA	19103	39.9547	75.1685
36198354	222	First Ave.			Philadelphia	PA	19103	39.9621	75.0981

**2.2.2. Addresses for RF Taps Outside Node Boundary**

The design and drafting platform will have the latitude and longitude of the RF taps (and the location of the support structure where the RF tap connects). Query all RF taps within the node boundary used to query the service addresses. In an ideal case, all RF taps from the RF network graph are in the query results. However, it is possible that some RF taps in the graph are outside the drafted node boundary. It is also possible that some RF taps from the RF network of other node housings are in the query results. Save these “extra” RF taps for use in the proximity-based address to RF tap correlation step below.

On the graph, RF taps that are significantly outside the node boundary will require an additional address query to capture the associated service addresses. Drops that connect the RF tap to an address location are typically limited to about 300 feet. To account for white space in the drafting process, if the location of the support structure where the RF taps are attached is more than 500 feet outside the drafted node boundary, then an additional address query needs to be executed to capture the locations.

For each RF tap outside the node boundary, query all addresses within a 500-foot circle of the support structure on which the RF tap is connected. Add these addresses to the addresses from the node boundary query.

**2.2.3. Addresses Associated with Node Segments**

To identify missing plant extensions from the drafted map, it is not sufficient only to examine the addresses in the drafted node boundary or addresses associated with drafted RF taps since the data

---

<sup>1</sup>We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.

returned can only be associated with the drafted access network. The address search needs to expand to include any address that has devices in the node segments associated with this node housing.

Using the assumption that the drafted network is at least partially accurate, one can use the service addresses queried from the map to obtain the set of node segments associated with the node housing. Some of the queried addresses passed by the drafted network will be actual subscribers. These subscriber accounts link to provisioned devices, whose telemetry includes the node segment on which the device reports.

Join the addresses retrieved above with the subscriber account database to identify subscriber accounts among these addresses. Find the device IDs of the provisioned HSD devices for the accounts and use that to retrieve the node segment from telemetry data. These queries are MSO specific. Below is an example table with the joined account to device to node segment data. Note that the data below is not actual customer data and was created only for demonstration purposes.

**Table 3 - Example Node Segments for HSD Devices<sup>2</sup>**

Account ID	Device ID	Device Type	Node Segment
335056384	A3:98:31:C9:ED:71	CM	FLD0010A
242012225	22:3E:C1:B5:CC:8E	CM	FLD0010C
263370322	B7:C0:39:D2:AC:4F	STB	FLD0010D
619042765	FA:45:22:13:D2:1C	CM	FLD0010A
390653365	6F:4A:70:D6:E8:44	MTA	FLD0010B
133703447	E5:7B:B2:84:0A:FC	CM	FLD0009D
423544086	9A:16:8C:CB:0E:B5	STB	FLD0010C
195661518	37:7E:08:47:DC:F4	STB	FLD0010B
444325235	0E:03:81:4C:60:D5	CM	FLD0010B
386919575	AA:6C:D7:AE:61:FA	MTA	FLD0010D
339519483	FD:00:9A:C0:04:9E	CM	FLD0010A
110301259	9F:F7:FD:E6:0B:5C	STB	FLD0010A

In an ideal case, the result should be between 1 and 4 node segments. There cannot be more than 4 node segments since the node housing has a maximum of 4 bus legs. Note that some of the address locations may be associated with other node segments and may have been retrieved since node boundaries of other node housings are adjacent or overlap. If more than 4 node segments are found, use the 4 instances that occur most frequently. In the example table above, the node segments FLD0010A, FLD0010B, FLD0010C, and FLD0010D appear most frequently.

It is still possible to identify more node segments than are associated to the node housing. For example, the node housing may actually contain only 2 node segments, but 4 node segments were identified due to node boundary overlap with adjacent nodes. These extra node segments will be removed when individual bus leg to node segment association is calculated below.

---

<sup>2</sup>We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.

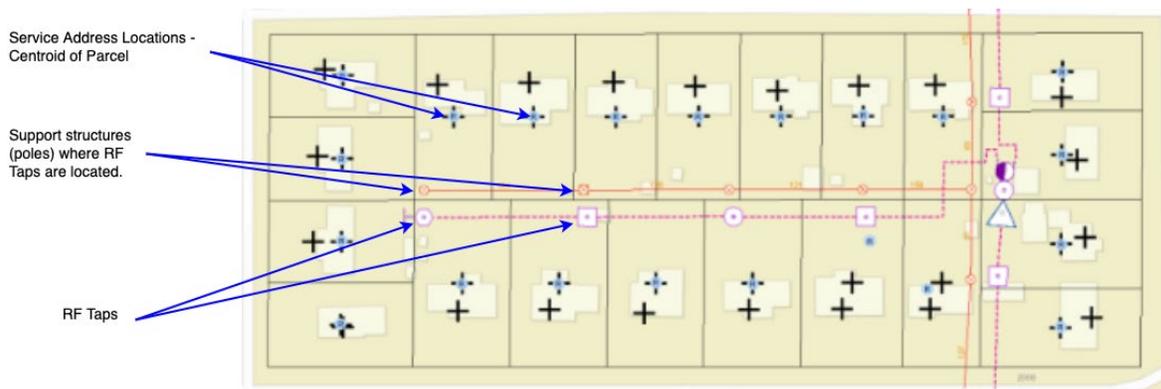
Given the set of node segments, query the device telemetry data for the full set of HSD devices that communicate on these node segments. Then use the device provisioning data to determine the service addresses for these accounts.

Add these addresses to the previously collected addresses.

### 2.2.4. Linking Addresses to RF Network

Addresses must be associated to RF Taps in the Graph to find map inconsistencies, calculate homes passed and device counts, and find the node segment per bus leg. In an ideal case, the service addresses from the USPS database or subscriber account database can be connected to the addresses specified in the design and drafting platform, which are then linked to the RF taps by drops on the drafted RF network. However, in many cases, drop details are missing or the address specified in the drafted network does not match USPS addresses or subscriber account addresses. Another method must be used to link addresses to the RF taps in the RF network.

The algorithm outlined below uses proximity of the address geolocation to the support structure on which the RF tap is attached. Support structure locations are typically more accurate than the RF equipment locations, which may be offset for white space management purposes. The figure below shows a drafted set of support structures and RF equipment relative to the parcels of service addresses.



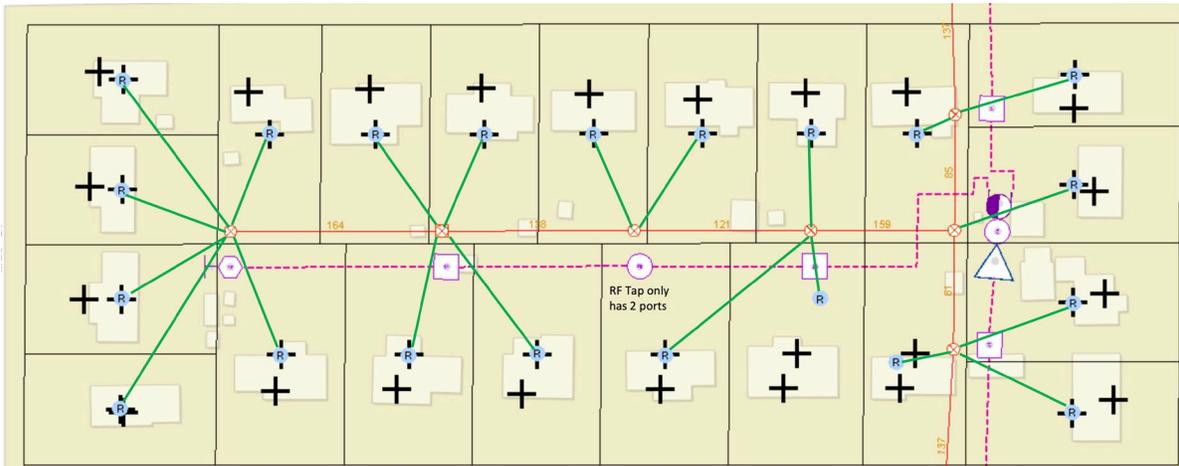
**Figure 9 - Drafted Supports and RF Taps Relative to Service Addresses**

Each address location shall be linked to the closest support structure location that has capacity for the address if the distance does not exceed 500 feet. Note that long drop lengths are typically limited to 300 feet. However, allowances must be made for the geocoder assigned location of the address not matching the actual drop location of the building or the drafted position of support structures not being exact.

The figure below shows proximity-based assignment of addresses to support structures associated with RF taps. Each parcel contains a cross with an “R,” which is the geocoder assigned location of the address. For comparison, the estimated drop location of the building is the plain cross. The pink circle, square, and hexagon connected by a pink dashed line are RF Taps, representing taps with 2, 4, and 8 ports, respectively. The small red circles with "X" connected by solid red lines are support structures (poles and strand).

The RF taps are physically located on the adjacent support structure. Notice that the two-port tap in the center of the image has two connections. There is an address below it which should connect to the two-

port tap based solely on shortest distance. However, since the RF Tap already has both ports occupied, the address must connect to the more distant RF tap to its "northeast."



**Figure 10 - Address to Support/Tap Assignment by Proximity and House Count**

The proximity algorithm requires a sorted list of distances for each combination of address and support structure with an RF tap. The algorithm uses this list to assign the next address based on shortest distance. The pseudo-code to track distances is shown below:

- Create an empty distance list that will store objects with the triplet of distance, address ID, and support ID.
- For each combination of address and support:
  - Calculate the distance from the address lat/lon to the support lat/lon; and
  - If the distance is less than max drop distance (e.g. 500'), add the triplet of distance, address ID, and support ID to the distance list.
- Sort the distance list by distance value in ascending order.

Support structures may contain multiple RF taps. It does not matter how the addresses are distributed across these RF taps if the RF taps are on the same bus leg. This is because the goals are determining address assignment per bus leg and determining the device and homes passed counts aggregated at amplifiers and splitters.

Each support will have a maximum number of addresses that can be connected. This will be the smaller of the total output ports of the RF taps on the support and the house count associated with the support. For example, a support may have 2 RF taps - one with 2 ports and one with 4 ports. The support may also have a house count of 5. The maximum number of addresses that can be connected is 5, since the total RF tap port count is 6, but the house count is 5. The converse may also occur. If the house count was 7, then the maximum number of addresses would be 6 since there are only 6 ports available between the 2 RF taps on that support.

The algorithm to assign addresses to supports by proximity will iterate over the sorted list of distance calculations. If the address has not yet been assigned to a support, and the referenced support has space for another address, the address shall be assigned to the support. The available capacity for the support is decreased by one.

There is no actual pruning of addresses from the distance list when assigning the address to an RF Tap. Each address appears multiple times in the list. Removing each address would require sequential traversal of the list to find all address instances. It is more efficient to traverse the distance list once and keep track of assigned addresses in the hash map of assigned IDs.

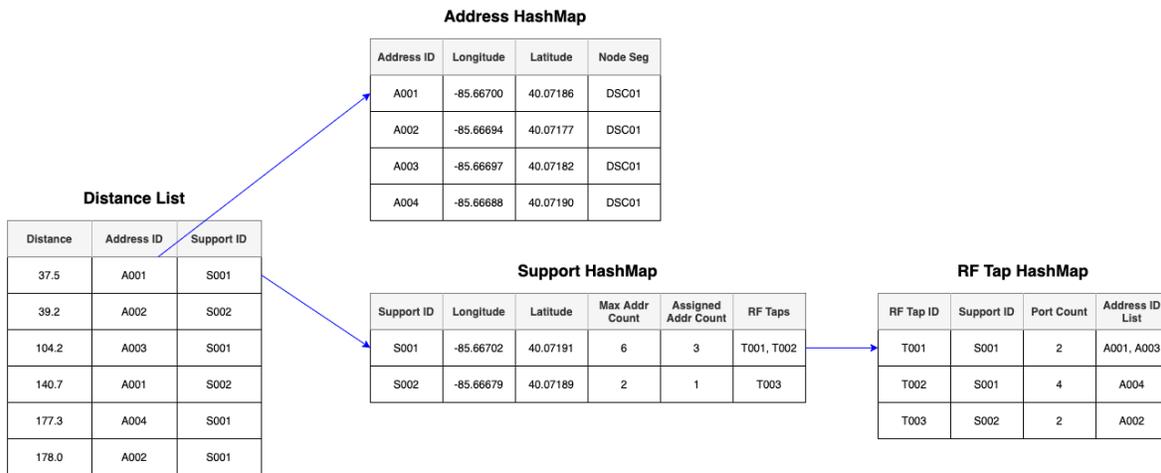
The following pseudo-code assigns addresses to RF taps based on support distance and capacity:

- Create an assigned hash map to store the list of assigned addresses. It is simply a lookup table whose key is the address ID but contains no data.
- For each element in the distance list (distance-address-support triplet) traversed in order of ascending distance:
  - Search assigned hash map. If a match is found, go to the next element in the distance list.
  - If support does not have capacity (assigned address count  $\geq$  max address count), go to next element in the distance list.
  - Find an RF tap for this support with available ports (connected address count less than port count):
    - Assign the address ID to this RF tap;
    - Decrease the support's available capacity by one; and
    - Add the address ID to the assigned hash map.

It is possible that some addresses cannot be assigned when there is insufficient RF tap port or house count capacity. It is equally possible that some RF tap ports do not have assigned addresses.

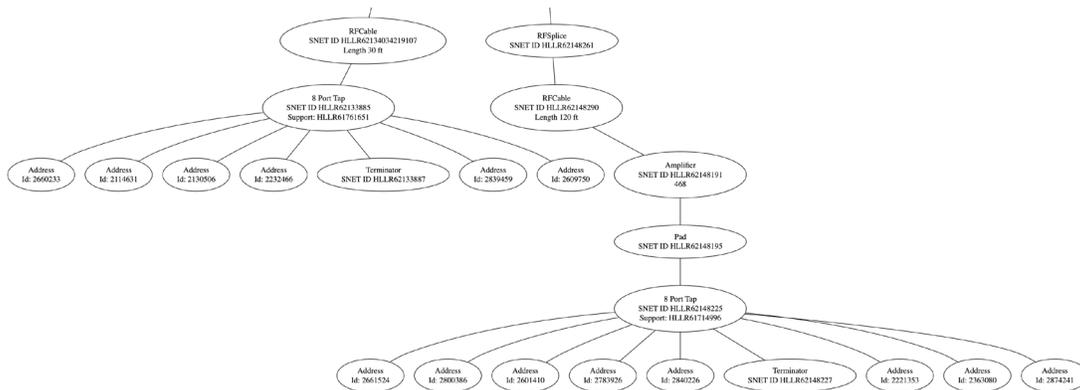
Below is an example of the data model used by the algorithm:

- **Distance List** - List of distances between an address and a support. List is sorted by distance in ascending order. Every combination of address and support distance is included in this list if the distance is less than the maximum drop distance.
- **Address Hash Map** – Stores the service addresses retrieved in the boundary and node segment queries above. The hash map uses the Address ID as a key. It contains the geolocation of each service address and may also store the node segment (if determined by device association).
- **Support Hash Map** – Stores the support structures on which the RF taps reside. The hash map uses the support ID as a key. Each support structure entry contains the list of RF taps on this support, maximum address count, and assigned address count.
- **RF Tap Hash Map** - RF tap lookup table using RF tap ID as a key. Contains all RF taps downstream of a node housing. Each RF tap entry stores the port count and the addresses associated with its support based on distance. The assigned address count must not exceed the port count.



**Figure 11 - Example Data Model to Associate Addresses to Supports/Taps**

Use the results in the RF tap hash map to update the RF network graph with the RF tap to service Address association. Create a new vertex for each assigned address and create an edge between the address Vertex and the RF tap vertex.



**Figure 12 - Address Locations in RF Network Graph**

### 2.3. Linking Devices to RF Network Graph

With the service address locations linked to the appropriate RF taps in the RF network graph, it should be possible to link the service locations to subscriber accounts and obtain the provisioned HSD devices for those accounts. A similar operation occurred above when identifying the possible node segments for the node housing. Remember that not all homes passed are subscribers and therefore only a subset of the service locations will have associated HSD devices.

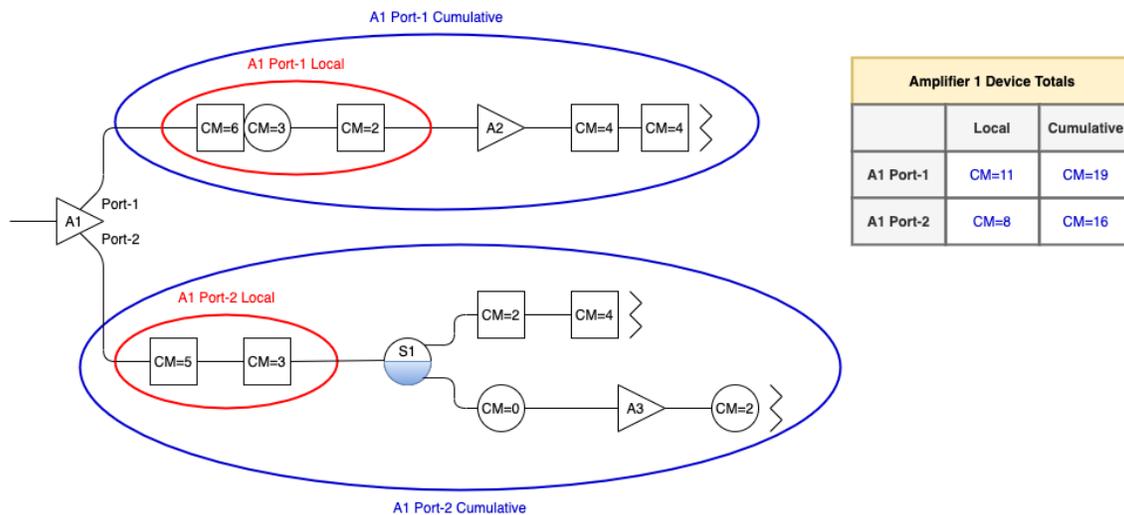
The provisioned HSD devices may be added to the RF network graph as vertices connected to the address vertices. However, for the Comcast implementation, the devices were simply added to the address vertex itself as additional data.

Note that the graph contents are never provided to the designer since it identifies individual subscriber service locations. The device count data will be aggregated and anonymized before being presented to the user.

## 2.4. Data Aggregation

HSD device counts and homes passed counts are aggregated at each of the node housing RF ports as well as output RF ports for amplifiers and splitters. The equipment where the data is aggregated shall be referred to as “aggregators” in this document. The device and homes passed counts provide the designer a view into the distribution of devices and service locations across the network, which will assist in the node segmentation or node split operation. Each of these elements will track both "local" counts and the “cumulative” counts. The local counts are the device and homes passed count from the specified aggregator to the next downstream aggregator for each of the output ports. The cumulative counts are the device and homes passed count from the specified aggregator to all downstream components and branches.

The figure below shows the port level data aggregation. Focusing on amplifier A1, there are two output ports: port-1 and port-2. The local counts are based on the device and homes passed counts on all RF taps from output port to the next aggregator (amplifier A2 and Splitter S1 for A1 port-1 and A1 port-2, respectively). The cumulative counts for a given output port is the summation of device and homes passed counts of all RF taps downstream of the port.



**Figure 13 - Aggregating Device Counts at the RF Output Port of Amps and Splitters**

MSOs can divide HSD device counts into multiple categories. Comcast uses 3 categories for device counts: cable modem (CM), set-top box (STB), and media terminal adapter (MTA). Homes passed can also be categorized as well, such as residential or commercial properties.

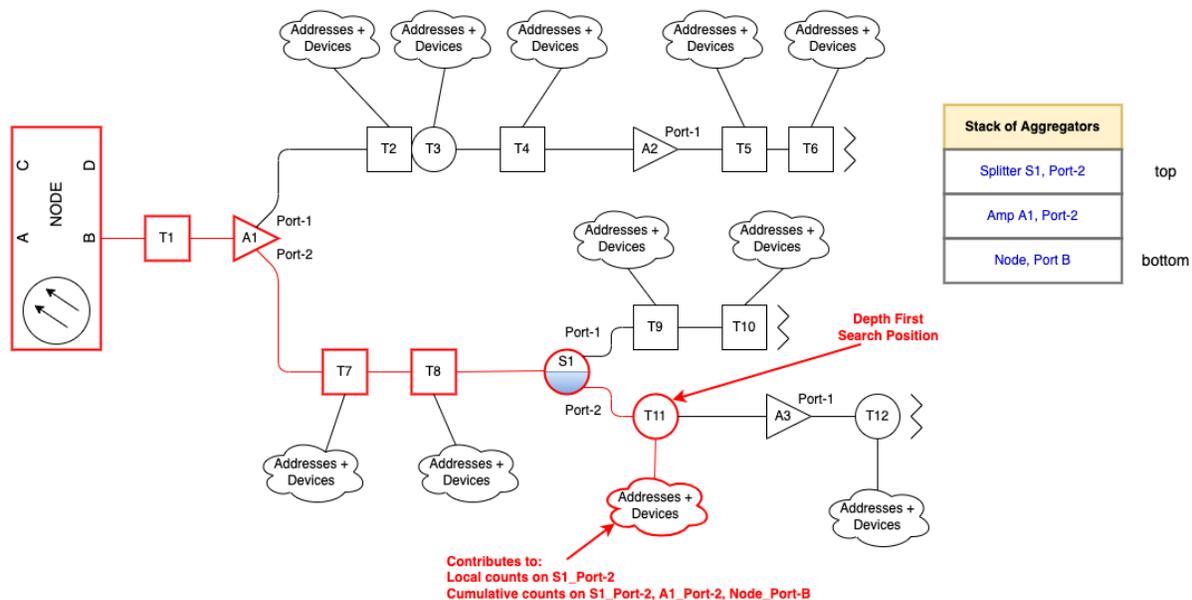
The algorithm to calculate the device and homes passed counts of each aggregator in the RF network uses a depth first search (DFS) to visit each of the vertices of the graph. A stack data structure keeps track of

the visited aggregators during the search. The last aggregator placed on the stack will be used for updating local counts. All aggregators on the stack will be used for calculating cumulative counts.

The DFS executes for each bus leg on the node housing. It visits each vertex in the graph, traversing as deeply as possible along one path before backtracking to visit alternate branches in the graph. When a DFS passes in the downstream direction through the RF output port of an aggregator, the RF output port of that aggregator gets pushed onto the stack. Conversely, when the DFS passes in the upstream direction through an RF output port on an aggregator, that RF output port pops off the stack.

When the DFS reaches an RF tap, it will traverse into the connected addresses. Each address contributes to the aggregated homes passed. If the address has associated provisioned HSD devices, these devices contribute to the aggregated device counts. The aggregator at the top of the stack will have its local counts updated. All aggregators in the stack will have the cumulative counts updated.

The figure below shows an example bus leg for a node housing. It includes a set of amplifiers, splitters, RF taps, and addresses. The RF equipment highlighted in red shows the elements traversed in a depth first search and the resulting stack at a point in the DFS when visiting vertex T11.



**Figure 14 - Depth First Search to Calculate Device and Homes Passed Counts**

The DFS began at Bus Leg B on the node housing. This RF output port is the first entry on the stack (bottom of the stack). The DFS proceeds to visit vertices T1 and A1, following the edge connected to Amp A1 Port-2. Amp A1 Port-2 shall be added to the stack since amplifiers are aggregators. The DFS proceeded to vertices T7, T8, and through Splitter S1 Port-2. Splitter S1 Port-2 shall be added to the stack since splitters are aggregators.

When the DFS traverses the address vertices connected to T11, each address contributes to the local homes passed count of S1 Port-2. Each address contributes to the cumulative homes passed counts for all elements on the stack, which include S1 Port-2, A1 Port-2, and Node Port B. Any provisioned HSD devices associated with the addresses will contribute similarly to the elements on the stack.

The DFS can also be used to calculate total amplifier counts and maximum amplifier cascade per bus leg. Each time an amplifier is encountered in the DFS, the amplifier total for that bus leg is incremented and the number of amplifiers on the stack are counted to update the maximum cascade count.

The pseudo-code for the local and cumulative device and homes passed count calculations are shown below. There will be an initialization function to kick off the processing of each bus leg. Aggregation begins at the node housing's RF output ports, which define each bus leg.

- For each node housing RF output port (bus leg), perform the following steps:
  - Create a "visited" hash map that tracks each of the vertices that have been visited.
  - Add node housing vertex to the "visited" hash map.
  - Create a stack that tracks RF port of aggregators visited as part of the DFS.
  - Push the node housing's RF output port onto the stack.
  - Set max cascade to 0.
  - Set amplifier count to 0.
  - Call ProcessVertex function for the vertex connected to this RF output port of the node housing.
  - Save max cascade and amplifier count data for this bus leg

Depth first search is performed by a recursive function that (1) skips visited nodes, (2) aggregates data from addresses connected to RF taps, and (3) manages the stack contents based on RF ports of aggregators that are visited.

- ProcessVertex function (SourceVertex)
  - Add SourceVertex to "visited" hash map.
  - If SourceVertex is a power supply, return without processing its connections.
  - If SourceVertex is an address:
    - For tracking local counts, add device and homes passed count from this address to the top element on the stack; and
    - For tracking cumulative counts, add device and homes passed count from this address to all elements in the stack.
  - If SourceVertex is an amplifier:
    - Increment amplifier count.
  - For each vertex connected to the SourceVertex:
    - If connected vertex is in the "visited" hash map, go to next connected vertex.
    - If SourceVertex is an aggregator:
      - Get RF output port from edge between this vertex and SourceVertex;
      - Add RF output port to aggregator stack;
      - Count the number of amplifier RF output ports on the stack;
      - If amplifier count on stack > max cascade,
        - max cascade = amplifier count on stack.
    - Call ProcessVertex function for this vertex.
    - If SourceVertex is an aggregator,
      - Pop stack.

The “visited” hash map is required so that searches do not loop back on themselves. Although this should not happen, it is good to have logic to prevent that case.

After executing the algorithm, the bus legs of the node housing and the RF output ports of amplifiers and splitters will have local and cumulative counts of devices and homes passed. The maximum number of amplifiers and maximum amplifier cascade will be known by bus leg.

The table below shows example device and homes passed counts for the aggregators in on Bus Leg D of a node housing:

**Table 4 – Example Device and Homes Passed Count by Aggregator<sup>3</sup>**

Type	Device ID	Port	Local				Cumulative			
			Homes	CM	STB	MTA	Homes	CM	STB	MTA
Node	UTAARMC18121634	D	0	0	0	0	96	39	43	69
Amp	UTAARAMC18121825	1	4	2	1	3	55	23	21	39
Amp	UTAARAMC18121825	2	3	1	2	2	41	16	22	30
Split	UTAARPMC18121923	1	2	1	1	1	21	8	8	17
Split	UTAARPMC18121923	2	0	0	0	0	30	11	12	19

Local device counts and homes passed counts may be zero if there are no RF taps between aggregators.

## 2.5. Bus Leg to Node Segment Identification

The set of node segments for the node housing has been discovered based on the service address locations linked to the RF network graph and the association of provisioned devices to accounts. The next step is to determine the actual node segment assignment by bus leg. All devices on the same bus leg must communicate with the same CMTS or RPD port. Therefore, they will share the same node segment.

Proximity based assignment of address locations to RF Taps is not deterministic, providing the chance for errors. Therefore, the node segment assignment to each bus leg shall be determined based on the most common node segment of the devices on the bus leg. This can be accomplished using a depth first search of each bus leg, counting each occurrence of a node segment for the devices associated with accounts on that bus leg. Details of the depth first search algorithm are provided above. In fact, this calculation can occur at the same time as the device aggregation.

Device telemetry indicates the node segment on which each device communicates. With devices associated with accounts, and the service address of the accounts linked to the RF taps in the RF network graph, one can calculate the node segment for the bus leg.

The pseudo-code to calculate the node segment assigned to a bus leg is shown below. Perform these steps for each bus leg of the node housing:

- Create hash map for node segments; key is the node segment, value is the count.
- Perform a depth first search on the RF output port of the node housing:
  - If vertex is an address, then for each device:
    - Use the node segment of the device to search the hash map;
    - If a match is found, increment the count for that node segment;
    - Else (no match found) add the node segment to the hash map with a count of 1.

<sup>3</sup>We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.

- Set the node segment of the bus leg to the node segment with the highest count in the hash map

The node segment assignment to the bus leg shall be used to identify areas of inconsistencies on the map.

## **2.6. Opportunities for Map Improvement**

After assigning node segments to each bus leg and linking homes passed or service addresses to the RF taps in the access network, one can examine each of the addresses to determine whether there is any mismatched data. Common exception conditions are outlined below.

It is possible that multiple exception conditions may occur per service address location. All exceptions that occur for each service address shall be recorded for display purposes to allow the designer to interpret the quality and accuracy of the map.

### **2.6.1. Unconnected Subscriber**

The address location is associated with a subscriber account and/or active device. However, there is no RF tap with which the address location is associated. This condition may be caused by a missed plant extension.

### **2.6.2. Address Not Serviceable**

The address location is not connected to an RF tap. Additionally, the address location is not associated with an active account and does not have a provisioned device. This may not be an error, since sometimes homes passed by the RF network are simply too far to connect. However, large numbers of non-serviceable addresses may indicate a map deficiency.

### **2.6.3. Unconnected RF Tap**

An address location connects to an RF tap that does not connect to the RF network of any node housing. The RF tap does exist and can be plotted on the map. However, tracing the RF cable that connected to its import port does not lead back to a node housing. This typically represents a drafting error where the RF tap input port was not connected to an RF cable.

### **2.6.4. Incorrect Node Segment**

The address location contains one or more provisioned devices whose node segment as reported by the telemetry data indicate it is on a different node segment than any of the node segments assigned to this node housing. If there are other RF taps from adjacent node housings within a typical drop distance, then the issue is either a drafting error or the proximity assignment algorithm.

### **2.6.5. Node Segment to Bus Leg Mismatch**

The address location contains one or more provisioned devices whose node segment as reported by the telemetry data indicate it is on a node segment for a different bus leg of the node housing. This may be a drafting error or error in the proximity-based address to tap assignment.

### 2.6.6. Outside Node Boundary

Address locations connected to RF taps in the graph reside outside the node boundary. This condition is minor and would just require an update to the drafted node boundary.

## 2.7. Pre-Vet Analysis Results

Pre-vet data collection, correlation, and analysis, generates a set of data for use by the designers to determine where drafted maps have opportunities for improvement and how to segment or split the RF network for a capacity upgrade.

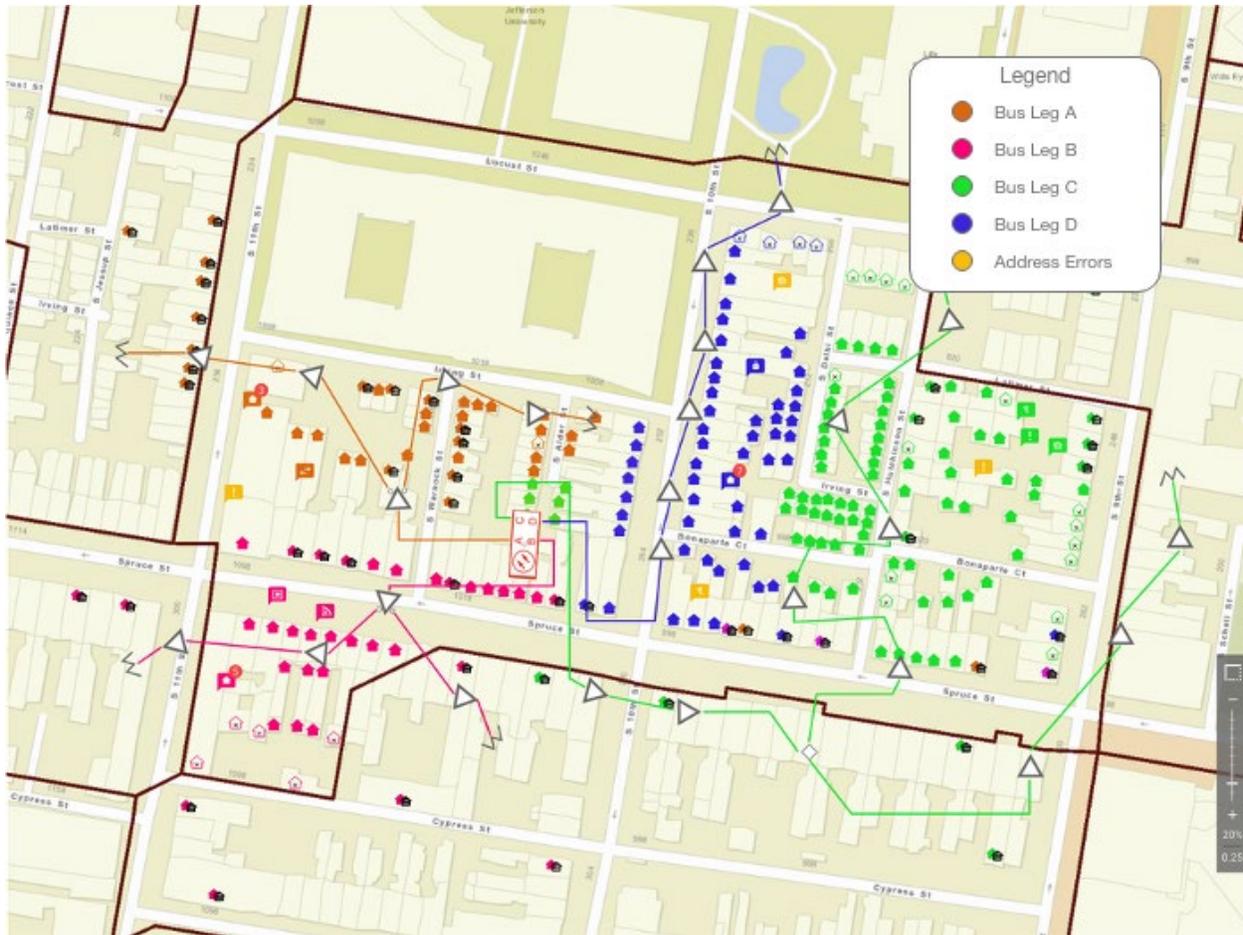
- Graph representation of all elements in the RF network.
- Service locations of homes passed and the connection points into the RF network
- Local and cumulative homes passed counts and device counts for each bus leg, amplifier, and splitter.
- Number of amplifiers and max amplifier cascade per bus leg.
- Service location exceptions including incorrect node segment, no connection to the RF network, and address not serviceable.

This data will be overlaid on the drafted maps to visualize the analysis results and any inconsistencies that have been found.

## 3. Maps and Data Output

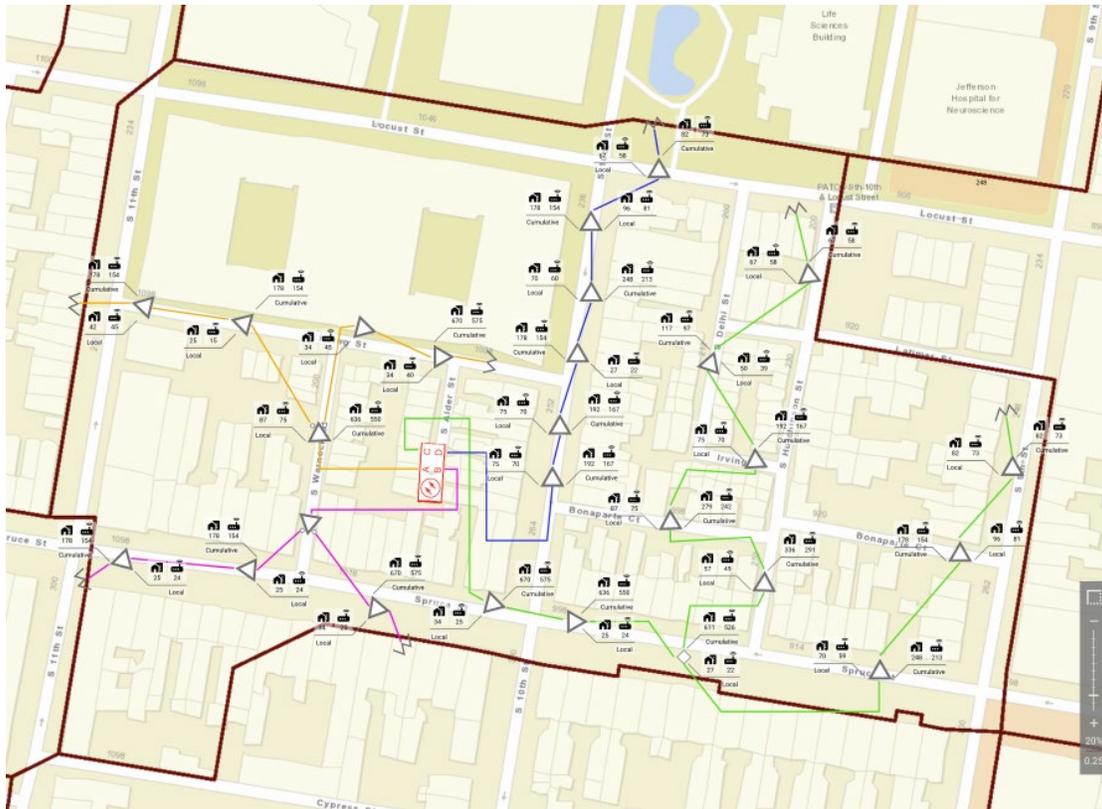
Visualization of the pre-vet analysis results is overlaid onto the drafted access network map. The pre-vet analysis shall be presented to the designer using multiple layers that can be individually toggled on and off. This capability permits the designer to view the overall set of results or focus on a specific aspect of the analysis.

In Figure 15, service addresses and address errors are color coded based on the associated bus leg. This helps designers identify which addresses are tied to which bus leg. It also provides a visual indication of how many addresses are on a single bus leg and identifies the opportunities for map improvements.



**Figure 15 - Map of Plant with Bus Legs Highlighted and Address Errors**

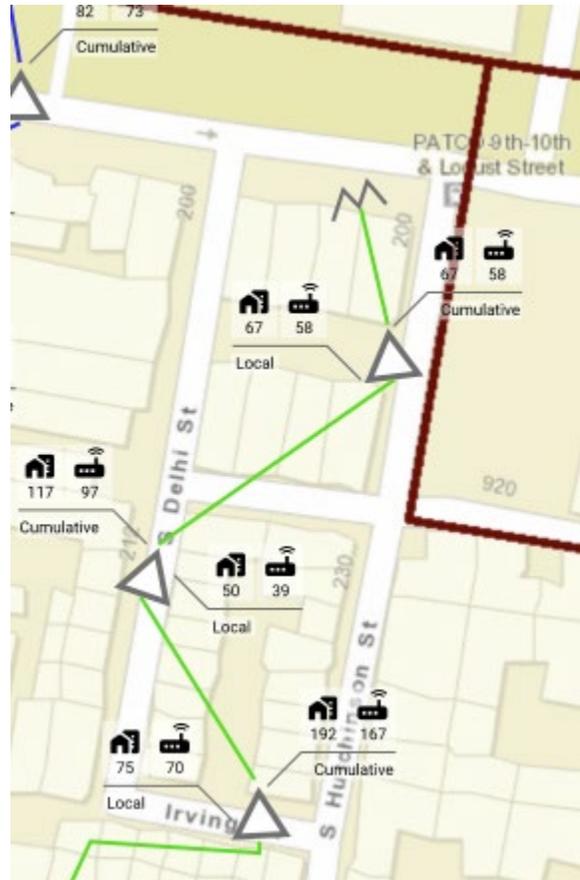
In the next two figures, users can see the device counts (aggregation of STB, CM, and MTAs) and homes passed at the amp level. The designer typically chooses an amp location to create a node split; Figure 16 shows how this view helps the user identify where there is pressure on the HFC plant.



**Figure 16 – Device and Homes Passed Counts at Amp in Node Boundary View<sup>4</sup>**

Figure 17 is a focused version of the previous figure. The image is centered on the end of a bus leg to view the cumulative and local homes passed and device counts at the amp level. Starting at the RF terminator, the cumulative counts add up the devices towards the node housing. The total number of devices and homes passed on a bus leg is shown at the node housing.

<sup>4</sup> We collect, store, and use all data in accordance with our privacy disclosures to users and applicable laws.



**Figure 17 – Device and Homes Passed Count at Amp (Focused View)**

In addition to the map view, the user interface provides a summary table of the bus leg. Figure 18 does not reflect real data, but a representation of how the data is displayed for users to easily digest.

Bus Leg Info				
Legs	A	B	C	D
# of Amps	12	0	18	0
Max Cascade	5	0	5	0
Serviceable Passings	78	1	103	3
Not Serviceable Passings	7	1	8	0
Device Count	66	0	110	2
Active Subscribers	46	0	72	1

**Figure 18 - Bus leg table**

The simplification and visualization of the data helps users understand the current state of the HFC plant in an efficient manner.

## 4. Conclusion

The pre-vet process aggregates and correlates data from multiple sources; the results are used to analyze the drafted RF network and determine the quality and accuracy of drafted maps. If these maps are deemed accurate, the pre-vet process provides information to the designer regarding where and how capacity upgrades should be applied to the RF network.

The process to automate these steps involves building a graph model of the RF network for a given node housing based on data in the design and drafting platform. The set of homes passed by the network are determined using both USPS addresses in the node boundary of the drafted network and HSD device telemetry data. These service locations are correlated against the drafted network to identify deficiencies in the drafted map.

A depth first search algorithm is employed on the graph to aggregate homes passed and device counts at the bus legs, amplifiers, and splitters. A view of the distribution of homes passed and devices across the network aids the designer when performing the capacity upgrade design. Additional data such as amplifier count and maximum amplifier cascade is also collected. The automation system presents the results to the designer through both a map-based interface and as tabular data.

The manual pre-vet process requires 2-8 hours to complete for a single node housing. Automating many of the pre-vet steps significantly reduces the amount of time a designer dedicates to this process. Time savings of 50% or more are possible, freeing the designer to focus more on the data analysis and design, rather than data collection and correlation.

Looking to the future, additional opportunities exist to enhance the pre-vet process. One example would be to collect upstream and downstream historical utilization data by node segment; this could be presented to the designer to augment the existing pre-vet's homes passed and device count data when designing the capacity upgrades. Beyond pre-vet automation can be applied to steps in the node capacity upgrade design process. Mid split designs, node segmentation, and even node splits share many common design steps, such as node housing swaps, amplifier upgrades, and power analysis. Automating these steps accelerates the design process, permitting faster delivery of network capacity upgrade designs.

## Abbreviations

Amp	amplifier
CM	cable modem
CMTS	cable modem termination system
DFS	depth first search
HFC	hybrid fiber coaxial
HSD	high-speed data
MTA	media terminal adapter
RF	radio frequency
RPD	remote PHY device
STB	set-top box
USPS	United States Postal Service
vCMTS	virtual cable modem termination system