Workgroup: Network Working Group
Internet-Draft: draft-birrane-tvr-use-cases-00
Published: 24 October 2022
Intended Status: Informational
Expires: 27 April 2023
Authors: E. Birrane
JHU/APL

TVR (Time-Variant Routing) Use Cases

Abstract

Time-Variant Routing (TVR) refers to the calculation of a path or subpath through a network where the time of message transmission (or receipt) is part of the overall route computation. This means that, all things being equal, a TVR computation might produce different results depending on the time that the computation is performed without other detectable changes to the network topology or other cost functions associated with the route.

This document introduces use cases where TVR computations could improve message exchange in a network.

About This Document

This note is to be removed before publishing as an RFC.

Status information for this document may be found at https://datatracker.ietf.org/doc/draft-birrane-tvr-use-cases/.

Discussion of this document takes place on the Time Variant Routing Working Group mailing list (mailto:tvr@ietf.org), which is archived at https://mailarchive.ietf.org/arch/browse/tvr/. Subscribe at https://www.ietf.org/mailman/listinfo/tvr/.

Source for this draft and an issue tracker can be found at https://github.com/NasaDtn/tvr-bof-use-cases.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Existing Routing Protocols expect to maintain contemporaneous, end-to-end connected paths across a network. Changes to that connectivity, such as the loss of an adjacent peer, are considered to be exceptional circumstances that must be corrected prior to the resumption of data transmission. Corrections may include attempting to re-establish lost adjacencies and recalculating or rediscovering a functional topology.

However, there are a growing number of use cases where changes to the routing topology are an expected part of network operations. In these cases the pre-planned loss and restoration of an adjacency, or formation of an alternate adjacency, should be seen as a non-disruptive event.

The expected loss (and planned resumption) of links can occur for a variety of reasons. In networks with mobile nodes, such as unmanned aerial vehicles and some orbiting spacecraft constellations, links are lost and re-established as a function of the mobility of the platforms. In networks without reliable access to power, such as networks harvesting energy from wind and solar, link activity might be restricted to certain times of day. Similarly, in networks prioritizing green computing and energy efficiency over data rate, network traffic might be planned around energy costs or expected user data volumes.

This document defines three use cases where a route computation might beneficially consider time information. Each of these use cases includes the following information.

1. An overview of the use case describing how route computations might select different paths (or subpaths) as a function of time.

2. A set of assumptions made by the use case as to the nature of the network and data exchange.

3. Specific discussion on the routing impacts of the use case.

4. An exemplar of a network conformant to the use case.

The document may not represent the full set of cases where time-variant routes could beneficially impact network performance - new use cases are expected to be generated over time. Similarly, the concrete examples within each use case are meant to provide an existence proof of the use case, not to present any exhaustive enumeration of potential examples. It is likely that there exist multiple example networks that could be claimed as instances of any given use case.

2. Conventions and Definitions

The key words "**MUST**", "**MUST NOT**", "**REQUIRED**", "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**RECOMMENDED**", "**NOT RECOMMENDED**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Resource Preservation

Some nodes in a network might operate in resource-constrained environments or otherwise with limited internal resources. Constraints such as available power, thermal ranges, and on-board storage can all impact the instantaneous functionality of a node. In particular, resource management on such a node can require that certain functionality be powered on (or off) to extend the ability of the node to participate in the network.

When power on a node is running low, non-critical functions on the node might be turned off in favor of extending node life. Alternatively, certain functions on a node may be turned off to allow the node to use available power to respond to an event, such as data collection. When a node is in danger of violating a thermal constraint, normal processing might be paused in favor of a transition to a thermal safe mode until a regular operating condition is reestablished. When local storage resources run low, a node might choose to expend power resources to fuse, delete, or transmit data off the node to free space for future data collection.

In addition to power, thermal, and storage, other resource constraints may exist on a node such that the preservation of resources are necessary to preserve the existence (and proper function) of the node in the network. Nodes operating in these conditions might benefit from TVR computations as the connectivity of the node changes over time as part of node preservation.

3.1. Assumptions

To manage on-board functionality as a function of available resources, a node must understand certain elements of how resources are used and replenished. It is assumed that patterns of the environment, device construction, and operational configuration exist with enough regularity and stability to allow meaningful planning. The following assumptions are made with this use case.

1. Known resource expenditures. It is assumed that there exists some determinable relationship between the resources available on a node and the resources needed to participate in a network. A node would need to understand when it has met some condition for participating in, or dropping out of, a network. This is somewhat similar to predicting the amount of battery life left on a laptop as a function of likely future usage.

2. Predictable resource accumulation. It is assumed that the accumulation of resources on a node are predictable such that a node might expect (and be able to communicate) when it is

likely to next rejoin a network. This is similar to predicting the time at which a battery on a laptop will be fully charged.

3. Consistent cost functions. It is assumed that resource management on a node is deterministic such that the management of a node as a function of resource expenditure and accumulation is consistent enough for link planning.

3.2. Routing Impacts

Resource management in these scenarios might involve turning off elements of the node as part of on-board resource management. These activities can affect data routing in a variety of ways.

1. Power Savings. On-board radios may be turned off to allow other node processing. This may happen on power-constrained devices to extend the battery life of the node or to allow a node to perform some other power-intensive task.

2. Thermal Savings. On-board radios may be turned off if there are thermal considerations on the node, such as an increase in a node's operating temperature.

3. Storage Savings. On-board radios may be turned on with the purpose of transmitting data off the node to free local storage space to collect new data.

Whenever a communications device on a node changes its powered state there is the possibility (if the node is within range of other nodes in a network) that the topology of the network is changed, which impacts route calculations through the network. Additionally, whenever a node joins a network there may be a delay between the joining of the node to the network and any discovery that may take place relating to the status of the node's functional neighborhood. During these times, forwarding to and from the node might be delayed pending some synchronization.

3.3. Exemplar

One example of a network where nodes must perform resource preservation is an energy-harvesting, wireless sensor network. In such a network, nodes may be powered solely by the environment (such as through solar panels). On-board power may fluctuate as a function of the sensors on each node, the processing performed on each node, and position and orientation of the node relative to its energy source.

Consider a contrived three node network where each node accumulates power through solar panels. Power available for Radio Frequency (RF) transmission is shown below in Figure 1. In this figure, each of the three nodes (Node 1, Node 2, and Node 3) have a different plot of available power over time. This example assumes that a node will not power its radio until available power is over some threshold, which is shown by the horizontal line on each plot.

```
           Node 1                   Node 2                   Node 3
P |                      P |   -------            P |          --
o |  ----       --       o |  /       \           o |         /  \
w |~/~~~~\~~~~~/~~\~~    w |~/~~~~~~~~~\~~~~~~    w |~~~~~~~~/~~~~\~~~~~
e |/      \   /    \     e |/           \         e |       /      \
r |        ---      -    r |             -----    r |-------        ---
  +---++----++----++-      +---++----++----++-      +---++----++----++-
      t1    t2    t3           t1    t2    t3           t1    t2    t3
           Time                     Time                     Time
```

Figure 1: Node Power Over Time

The connectivity of this three node network changes over time in ways that may be predictable and are likely able to be communicated to other nodes in this small sensor network. Examples of connectivity are shown in Figure 2. This figure shows a sample of network connectivity at three times: t1, t2, and t3.

- At time t1 Node 1 and Node 2 have their radios powered and are expected to communicate.

- At time t2 it is expected that Node 1 has its radio off, but that Node 2 and Node 3 can communicate.

- Finally, at time t3 it is expected that Node 1 may be turning its radio off and that Node 2 and Node 3 are not powering their radios and there is no expectation of connectivity.

```
            +----------+        +----------+        +----------+
       t1   |  Node 1  |--------|  Node 2  |        |  Node 3  |
            +----------+        +----------+        +----------+

            +----------+        +----------+        +----------+
       t2   |  Node 1  |        |  Node 2  |--------|  Node 3  |
            +----------+        +----------+        +----------+

            +----------+        +----------+        +----------+
       t3   |  Node 1  |        |  Node 2  |        |  Node 3  |
            +----------+        +----------+        +----------+
```

Figure 2: Topology over Time

4. Operating Efficiency

Some nodes in a network might alter their networking behavior to optimize metrics associated with the cost of a node's operation. While the resource preservation use case described in Section 3 addresses node survival, this use case discusses non-survival efficiencies such as the financial cost to operate the node and the environmental impact (cost) of using that node.

When a node operates using some pre-existing infrastructure there is (typically) some cost associated with the use of that infrastructure. Sample costs include items such as the following.

1. Nodes that use existing wireless communications such as a cellular infrastructure must pay to communicate to and through that infrastructure.

2. Nodes supplied with electricity from an energy provider pay for the power they use.

3. Nodes that cluster computation and activities might increase the temperature of the node and incur additional costs associated with cooling the node (or collection of nodes).

4. Beyond financial costs, assessing the environmental impact of operating a node may also be modeled as a cost associated with node operation, to include achieving carbon credits or other incentives for green computing.

When the cost of using a node's resources changes over time, a node can benefit from predicting when data transmissions might optimize costs, environmental impacts, or other metrics associated with operation.

4.1. Assumptions

The ability to predict the impact of a node's resource utilization over time presumes that the node exists within a defined environment (or infrastructure). Some necessary characteristics of these environments are listed as follows.

1. Cost Measureability. The impacts of operating a node within its environment can be measured in a deterministic way. For example, that the cost-per-bit of data over a cellular network or the cost-per-kilowatt of energy used are known.

2. Cost Predictability. Changes to the impacts of resource utilization are known in advance. For example, if the cost of energy is less expensive in the evening than during the day, there exists some way of communicating this change to a node.

3. Cost Persistent. Changes to the cost of operating in the environment persist for a sufficient amount of time such that behavior can be adjusted in response to changing costs. If costs change rapidly or near continuously it is likely not possible to meaningfully react to their change.

4. Cost Magnitude. The magnitude of cost changes are such that a node sees a minimum threshold cost reduction as a result of optimization.

4.2. Routing Impacts

Optimizing resource utilization can affect route computation in ways similar to those experienced with resource preservation. The significant difference being that when optimizing costs the overall network topology is not changing. Even without a changing topology, cost optimization can impact route calculation in a variety of ways, some of which are described as follows.

1. Link Filtering. Data might be accumulated on a node waiting for a cost-effective time for data transmission. Individual link costs might be annotated with cost information such that adjacencies with a too-high cost might not be used for forwarding. This effectively filters which adjacencies are used (possibly as a function of the type of data being routed).

2. Burst Planning. In cases where there is a cost savings associated with fewer longer transmissions (versus many smaller transmissions), nodes might refuse to forward data until a sufficient data volume exists to justify a transmission.

3. Environmental Measurement. Nodes that measure the quality of individual links can compute the overall cost of using a link as a function of the signal strength of the link. If link quality is insufficient due to environmental conditions (such as clouds on an optical link or long distance RF transmission in a storm) the cost required to communicate over the link may be too much, even if access to infrastructure is otherwise in a less expensive time of day.

In each of these cases, some consideration of the efficiency of transmission is prioritized over achieving a particular data rate. Waiting until data rate costs are lower takes advantage of platforms using time-of-use rate plans - both for pay-as-you-go data and associated energy costs. Accumulating data volumes and choosing more opportune times to transmit can also result in less energy consumption by radios and, thus, less operating cost for platforms.

4.3. Exemplar

One example of a network where nodes might seek to optimize operating cost is a set of nodes operating over cellular connections that charge both On-Peak and Off-Peak data rates. In this case, individual nodes may be allocated a fixed set of "On-Peak" minutes such that exceeding that amount of time results in expensive overage charges. Generally, the concept of On-Peak and Off-Peak minutes exists to deter the use of a given network at times when the cellular network is likely to encounter heavy call volumes (such as during the workday).

Just as pricing information can act as a deterrent (or incentive) for a human cellular user, this pricing information can be codified in ways that also allow machine-to-machine (M2M) connections to prioritize Off-Peak communications for certain types of data exchange. Many M2M traffic exchanges involve schedulable activities, such as nightly bulk file transfers, pushing software updates, synchronizing datastores, and sending non-critical events and logs. These activities are usually already scheduled to minimize impact on businesses and customers, but can also be scheduled to minimize overall cost.

Consider a contrived three node network, similar to the one pictured in Figure 1, except that in this case the resource that varies over time is the cost of the data exchange. This case is illustrated below in Figure 3. In this figure, a series of three plots are given, one for each of nodes Node 1, Node 2, and Node 3. Each of these nodes exists in a different cellular service area which has different On-Peak and Off-Peak data rate times. This is shown in each figure by times when the cost is low (Off-Peak) and when the cost is high (On-Peak).

```
 Node 1                     Node 2                   Node 3

C |       +---------     C |--+                    C |-------------+
o |       |              o |  |                    o |             |
s |       |              s |  |                    s |             |
t |-------+              t |  +----------------    t |             +----
  |                        |                         |
  +---++----++----++--     +----++----++----++--     +----++----++------+
      t1    t2    t3            t1    t2    t3            t1    t2     t
           Time                      Time                      Time
```

Figure 3: Data Cost Over Time

Given the presumption that peak times are known in advance, the cost of data exchange from Node 1 through Node 2 to Node 3 can be calculated. Examples of these data exchanges are shown in Figure 4.

From this figure, both times t1 and t3 result in a smaller cost of data exchange than choosing to communicate data at time t2.

```
      +-----------+          +-----------+          +-----------+
 t1   |  Node N1  |---LOW----|  Node N2  |---HIGH---|  Node N3  |
      +-----------+          +-----------+          +-----------+

      +-----------+          +-----------+          +-----------+
 t2   |  Node N1  |---HIGH---|  Node N2  |---HIGH---|  Node N3  |
      +-----------+          +-----------+          +-----------+

      +-----------+          +-----------+          +-----------+
 t3   |  Node N1  |---HIGH---|  Node N2  |----LOW---|  Node N3  |
      +-----------+          +-----------+          +-----------+
```

Figure 4: Data Exchange Cost over Time

While not possible in every circumstance, a highly optimized plan could be to communicate from Node 1 to Node 2 at time t1 and then queue data at Node 2 until time t3 for delivery to Node 3. This case is shown in Figure 5.

```
      +-----------+          +-----------+
 t1   |  Node N1  |---LOW----|  Node N2  |
      +-----------+          +-----------+
                             +-----------+          +-----------+
 t3                          |  Node N2  |----LOW---|  Node N3  |
                             +-----------+          +-----------+
```

Figure 5: Data Cost using Storage

5. Mobile Devices

When a node is placed on a mobile platform, the mobility of the platform (and thus the mobility of the node) may cause changes to the topology of the network over time. To the extent that the relative mobility between and amongst nodes in the network can be understood in advance, the associated loss and establishment of adjacencies can also be planned for.

Mobility can cause the loss of an adjacent link in several ways, such as the following.

1. Node mobility can cause the distance between two nodes to become large enough that distance-related attenuation causes the mobile node to lose connectivity with one or more other nodes in the network.

2. Node mobility can also be used to maintain a required distance from other mobile nodes in the network. While moving, external characteristics may cause the loss of links through occultation or other hazards of traversing a shared environment.

3. Nodes that can change the orientation of their communication terminals will also establish and lose connectivity with other nodes as a function of that motion.

The impacts of node mobility are separate concerns from either resource preservation or cost efficiency. Unlike with resource preservation, there is no expectation that mobile nodes are resource constrained. Unlike cost efficiency, there is no expectation that concepts such as peak-hours or other computation-based metrics need to be optimized. This use case is solely concerned with understanding the routing implications of motion-related changes to a network topology.

5.1. Assumptions

Predicting the impact of node mobility on route computation requires some information relating to the nature of the mobility and the nature of the environment being moved through. Some information presumed to exist for planning is listed as follows.

1. Path Predictability. The path of a mobile node through its environment is known (or can be predicted) as a function of (at least) time. it is presumed that mobile nodes using time-variant algorithms would not exhibit purely random motion.

2. Environmental Knowledge. When otherwise well-connected mobile nodes pass through certain elements of their environment (such as a storm, a tunnel, or the horizon) they may lose connectivity. The duration of this connectivity loss is assumed to be calculable as a function of node mobility and the environment itself.

5.2. Routing Impacts

Changing a network topology has a straightforward impact on the computation of paths (or subpaths) through that topology. In particular, the following features can be implemented in a network with mobile nodes such that different paths might be computed over time.

1. Adjacent Link Expiration. A node might be able to predict that an adjacency will expire as a function of that node's mobility, the other node's mobility, or some characteristic of the environment. Determining that an adjacency has expired allows a

route computation to plan for that loss, rather than default to an error recovery mechanism.

2. Adjacent Link Resumption. Just as the loss of an adjacency can be predicted, it may be possible to predict when an adjacency will resume.

3. Data Rate Adjustments. The achievable data rate over a given link is not constant over time, and may vary significantly as a function of both relative mobility between a transmitter and receiver as well as the environment being transmitted through. Knowledge of both mobility and environmental state may allow for prediction of data rates which may impact path computation.

4. Adjacent Link Filtering. Separate from the instantaneous presence or absence of an adjacency, a route computation might choose to not use an adjacency if that adjacency is likely to expire in the near future or if it is likely to experience a significant drop in predicted data rate.

5.3. Exemplar

There are a significant number of mobile node use cases, to include vehicle-to-vehicle communications, swarms of unmanned aerial and underwater vehicles, ships in shipping lanes, airplanes following flight plans, and trains and subways. A (relatively) new type of mobile network that has emerged over the past several years is the Low Earth Orbit (LEO) networked constellation (LEO-NC). There are a number of such constellations being built by both private industry and governments.

Many LEO-NCs have a similar operational concept of hundreds-to-thousands of inexpensive spacecraft that can communicate both with their orbital neighbors as well as down to any ground station that they happen to be passing over. The relationship between an individual spacecraft and an individual ground station becomes somewhat complex as each spacecraft may only be over a single ground station for a few minutes at a time.

A LEO-NC represents a good example of planned mobility based on the predictability of spacecraft in orbit. Unlike other mobile vehicles that might experience traffic congestion or significant changes to speed, spacecraft operate in a less impactful environment. This determinism makes them an excellent candidate for time-variant route computations.

Consider three spacecraft (N1, N2, and N3) following each other sequentially in the same orbit (this is sometimes called a string of pearls configuration). Spacecraft N2 always maintains connectivity to its two neighbor spacecraft, N1 (which is behind in the orbit) and N3 (which is ahead in the orbit). This configuration is illustrated in Figure 6. While these spacecraft are all mobile, their relative mobility ensures that they are always in contact with each other (absent any true error condition).

```
      .--.                     .--.                     .--.
####-| N1 |-####  <--->  ####-| N2 |-####  <--->  ####-| N3 |-####
      \__/                     \__/                     \__/
```

Figure 6: Three Sequential Spacecraft

Flying over a ground station imposes a non-relative motion between the ground and the spacecraft - namely that any given ground station will only be in view of the spacecraft for a short period of time. The times at which each spacecraft can see the ground station is shown in the plots in Figure 7. In this figure, ground contact is shown when the plot is high, and a lack of ground contact is shown when the graph is low. From this, we see that spacecraft N3 can see ground at time t1, N2 sees ground at time t2, and spacecraft N1 sees ground at time t3.

```
       Spacecraft N1            Spacecraft N2             Spacecraft N3
G |                      G |                       G |
r |              +--+    r |         +--+          r |   +--+
o |              |  |    o |         |  |          o |   |  |
u |              |  |    u |         |  |          u |   |  |
n |--------------+  +-   n |---------+  +-------   n |---+  +-----------
d |                      d |                       d |
  +---++----++----++--     +----++----++----++--     +----++----++----++
      t1    t2    t3            t1    t2    t3            t1    t2    t3
           Time                      Time                      Time
```

Figure 7: Spacecraft Ground Contacts Over Time

Since the ground station in this example is stationary, each spacecraft will pass over it, resulting in a change to the network topology. This topology change is shown in Figure 8. At time t1, any message residing on N3 and destined for the ground could be forwarded directly to the ground station. At time t2, that same message would need to, instead, be forwarded to N2 and then forwarded to ground. By time t3, the same message would need to be forwarded from N2 to N1 and then down to ground.

```
       +------+          +------+
   t1  |  N2  |----------|  N3  |
       +------+          +---+--+
                             |
                            /|\
                           \___/
                            / \
                          Ground
                          Station
   ------------------------------------------------------------------
       +------+          +------+          +------+
   t2  |  N1  |----------|  N2  |----------|  N3  |
       +------+          +---+--+          +------+
                             |
                            /|\
                           \___/
                            / \
                          Ground
                          Station
   ------------------------------------------------------------------
                         +------+          +------+          +------+
   t3                    |  N1  |----------|  N2  |----------|  N3  |
                         +---+--+          +------+          +------+
                             |
                            /|\
                           \___/
                            / \
                          Ground
                          Station
   ------------------------------------------------------------------
```

Figure 8: Constellation Topology Over Time

6. Security Considerations

TBD

7. IANA Considerations

This document has no IANA actions.

8. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/

RFC2119, March 1997, <https://www.rfc-editor.org/rfc/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/rfc/rfc8174>.

Acknowledgments

TBD

Author's Address

Ed Birrane
JHU/APL

Email: Edward.Birrane@jhuapl.edu