# Root Cause Analysis for Fulfillment Decisions

Table of Contents

[Root Cause Analysis for Fulfillment Decisions 1](#_Toc135666273)

[Preliminaries 1](#_Toc135666274)

[Notation 1](#_Toc135666275)

[Assumptions 2](#_Toc135666276)

[Events 3](#_Toc135666277)

[Directed Follow Graphs 4](#_Toc135666278)

[Event Relationships and Causality 7](#_Toc135666279)

[Directed Causal Graphs 12](#_Toc135666280)

[Problem Statement for Root Cause Analysis of Fulfillment Decisions 12](#_Toc135666281)

[Algorithm For Root Cause Analysis 12](#_Toc135666282)

[Appendix: Probabilistic Temporal Logic 12](#_Toc135666283)

[Bibliography 12](#_Toc135666284)

[Downloadable Links for the Bibliography 13](#_Toc135666285)

## Preliminaries

Before we can formulate the problem statement and the algorithm providing a solution, we need to start with a set of notational conventions and definitions.

### Notation

- with capital Latin letters we will denote *scalar quantities* which are either essential algorithm parameters or constants which will not change during the algorithm execution; for example, *number of feasible nodes for the current bundle* (scalar constant) will be denoted with and *inventory for given SKU on given node* (algorithm parameter) will be denoted with . Graphs will also be denoted with capital Latin letters for historical reasons.

– with small Latin letters we will denote *variable/unknown (integral or not) quantities*, not necessarily scalar. For example, with we can denote the number of order-lines fulfilled at a given node.

– with small Greek letters we will denote *variable/unknown (integral or not) quantities*, not necessarily scalar.

– with capital Script letters we will denote a *set* (ordered or unordered) of quantities of the same type; for example, with we will denote the set of SKUs in some bundle of some order

– with capital Greek letters we will denote a *concept*, *logical statement* or a *logical expression* of *logical terms / statements* which is adorned with *semantic meaning*. In case of a logical statement, the latter can be either true or false depending on the context. The capital Epsilon letter will be reserved to denote an event type or event of interest. For instance, will denote the event of type “*an order has been received*”.

- with capital Fraktur letters we will denote a *map* over several arguments where at least one of those arguments is of type logical expression, a logical statement or a set of logical statements. For example, denotes graph representation of the concept by the set of events .

- with double struck Latin capital letters we will denote standard number sets. For example

- the set of complex numbers

- the set of natural numbers

- the set of the real numbers

- the set of integer numbers

Reserved letters for quantities, sets and concepts:

– number of bundles in the order .

– order received at moment .

– the -th bundle of order ; alternatively, denoted as .

or - the set of SKUs for the -th bundle will be denoted with .

- denotes some quantity related to the -th bundle of the -th order.

– denotes some quantity related to the SKU at node e.g., inventory for SKU at node .

- directed graph

- the vertex set of the directed graph

– the arc set of the directed graph

– denotes the *relative frequency of occurrence* of the event given event with the dataset

– denotes *Average Degree Of Causal Significance* (*ADCS*) of event for event given the background contexts

denotes the statement that event follows event

denotes the statement that event precedes event

– denotes graph representation of the concept

- denotes complete representation of the concept with the event set

– denotes static dependency map

– denotes static association map

Reserved symbols for relations and operations

- denotes logical conjunction

- denotes logical disjunction

- denotes logical negation

- denotes *follows in time* relation between two concepts

- denotes *is reachable* relation between two concepts

- denotes *static dependency* between two concepts

- denotes *dynamic dependency* between two concepts

- denotes *association (static, dynamic)* between two concepts

– denotes -*spurious cause* relation between two concepts

- denotes *causal association* between two concepts

- *prima facie* causal relation between two concepts

- denotes Eells causal relation between two concepts

- denotes *matching* between directed follow graph (DFG) and a concept

### Assumptions

All orders can be ordered in an increasing sequence of moments in time . That is, we assume that no two orders will arrive at the same moment in time. Thus, the time will take the form of a discrete variable on the natural numbers i.e., . Therefore, any order will be uniquely identified by a subscript .

**Definition**: *Atomic Proposition*

A basic proposition (or *atom*) which cannot be represented as a set of other atoms connected using conjunction , disjunction , negation , implication and equivalence .

### Events

**Definition**: *Event*

The word *Event* will be used to denote a *specific kind* *of* an *event* which is relevant for the causal analysis. *Event* can be viewed as *a template* from which a specific event can be *instantiated*. We will denote each event with capital Greek letter. Where it will be clear from the context, we will use interchangeably the word “*event*” to denote either *Event of specific kind* or an *Event instance*.

**Definition**: *Parameters of Event*

Each event has a *set of parameters* which will be denoted with . The set of parameters of an event together with the semantic description of the event uniquely identify the event. One can think of the semantic description as sort of *“semantic” template* (or *predicate* from some first order logic) identifying this event type. The template parameters will be given obviously with the parameter set which is an ordered set. Thus, each event is defined with the pair . An Event Instance additionally to and is given a specific value for each . We will denote the value space of an Event Instance with . Thus, an event instance is defined with the triplet .

**Note**: Every event additionally to its standard parameter set will have an implicit timestamp parameter which will always be present without regard of the nature of the event. We will not include explicitly the timestamp among the event parameters unless it is necessary in order to define uniquely the event instance.

We define the following *Events* which are relevant for the analysis of Fulfillment decisions causing splits:

*Set of events for analysis of the cause of splits in Fulfillment Decisions*

- order is received. Event parameters:

- the -th bundle of order is being processed. Event parameters: ,

- SKU in the -th bundle of order is being processed. Event parameters: , ,

- node has sufficient inventory for SKU in with order . Event parameters: , , ,

- node has sufficient capacity for SKU in with order . Event parameters: , , ,

- node is shipping eligible for SKU in with order . Event parameters: , , ,

- node is deprioritized; node has SKU in with order . Event parameters: , , ,

- node is turned on; node has SKU in with order . Event parameters: , , ,

- node is turned off; node has SKU in with order . Event parameters: , , ,

- node is soft capacity; node has SKU in with order . Event parameters: , , ,

- service level for node is overridden; node has SKU in with order . Event parameters: , , ,

- carrier for node j is overridden; node j has SKU in with order . Event parameters: , , ,

- backlog days for node is overridden; node has SKU in with order . Event parameters: , , ,

- node is with depleted inventory; node has SKU in with order . Event parameters: , , ,

- node is with depleted capacity; node has SKU in with order . Event parameters: , , ,

- node contains an orphan for SKU which is in with order . Event parameters: , , ,

– splitting decision is made; that is, at least for one bundle the SKUs in are fulfilled by more than one node with order . Event parameters: ,,

We will visualize an event with an ellipse and a Capital Greek letter denoting the event. For example:

We will visualize an Event instance with an ellipsis and will use a Capital Greek letter to denote the specific kind of event which it is an instance of. We will attach a set of labels where each label will represent an *atomic proposition* with pertinent semantic information for this instance. For example, in case of we can have:

Thus, the parameter set of is given with . Since the value space for the parameters of the event instances of is given with .

Let us consider the event : . Since the value space for the parameters of the event instances of is given with .

For we have accordingly and .

For we have accordingly and . Here

For

//TODO: finish this

**Definition**: Event *follows in time* Event

We say that event *follows in time* (denoted with ) if event has occurred in time *after* event and there does not exist a third event which occurs *after* and *before* .

*Follows in time* relation implies the timestamp associated with is newer than that associated with i.e. . See the **Note** on the event parameters.

**Definition**: Event *is reachable from* Event

We say that event *follows in time* (denoted with ) if event has occurred in time *after* event .

*Is reachable from* relation implies that the timestamp associated with is newer than that associated with i.e. . See the **Note** on the event parameters.

For example, let the event denotes the node being turned off, and event denotes order received at time . Then can be interpreted as “node was turned off prior to receiving order ”.

**Lemma**: *is reachable from* relation is the transitive closure of the *follows in time* relation.

### Directed Follow Graphs

We use *Directed Follow Graphs* (DFG) to depict order fulfillment scenarios which we are interested to capture.

Each node of the DFG will represent an *Event instance* of interest. We use an arc (or a directed edge) to denote a *follow in time* relation between two Event instances. Each arc has a label with a counter which counts how many times the current arc connecting a pair of events has been seen in the data log given specific dataset.

**Definition**: *Labeled Directed Follow Graph (LDFG)*

We extend the concept of Directed Follow Graph (DFG) by introducing a set of labels to each node and to each arc. Each label represents an atomic proposition which is relevant to the specific node or to specific arc.

We use LDFGs to represent the follow relationships between event types and event instances for a given dataset of orders.

Discussion on how DFG is constructed-

Let us consider a given order data set and let us assume we have Fulfillment Optimization engine processing the set of orders sequentially thereby generating order metrics events. Let us assume we have a parsing engine which combs through the order metrics created after the Fulfillment Optimization engine run. This parsing engine parses the events which it is configured to recognize and assembles the DFG instance based on the parsed events data. Let us denote with the events which the parsing engine is configured to recognize. Per our definition of *Parameters of Event* given earlier each event type is represented by the pair where is the template of the event which together with the parameter set uniquely identifies this type of event. Let us denote with the ordered set of values which correspond to each parameter for all instances of generated using . Since each event instance has a timestamp, we can construct DFG from the parsed events. Each arc between two event types and will be labeled with the final count showing how many times this pair of events have been seen in a follow relation .

For example:

**Example 1**:

An order for has 3 bundles. The first bundle has three SKUs – , , , the second bundle has two SKUs – and , the third bundle has 4 SKUs – , , ,.

Let us denote with the Event that an order with 3 bundles have been received. Also, we will denote with an instance of event type each of the bundles of the order . We denote with an instance of event type each of the SKUs in each bundle of the order

We depict this scenario with the following DFG:

**Definition**: *Directed Follow Graph Instance* *(DFGI)*

DFGI is a directed graph in which each node is a specific *event instance (or a token)* of an event type and each arc denotes a follow relation between the event instances. Each node (event instance/token) is labeled with the value set of parameters for this event type. For example:

**Definition**: *Aggregated Directed Follow Graph* (*ADFG*)

The ADFG corresponding to DFGI can be obtained by replacing each event instance by its corresponding event type and replacing a multi-set of arcs leaving an event instance of type and entering event instance of type with a single arc labeled with the corresponding instance count.

**Definition**: *Frequency Count* of pair of events and – this is the number of DFG instances in which directly follows i.e., .

**Definition**: *DFG Representation* of a concept over an event set

We say that DFG is a representation of if the graph constructed with the events in models *semantically* the internal structure of .

For example, the DFG shown in *Example 1* is DFG representation of the order . The DFG representing will be denoted with or shortly .

**Definition**: *Complete* *Representation* of a concept over an event set

We say that the DFG is a *complete* *representation* of the concept (e.g., fulfillment order, fulfillment decision) from the event set , denoted with , *iff* there does not exist DFG such that with and .

**Definition**: *Order Fulfillment Decision*

The process of fulfilling the order which can be viewed as a set of events relevant to the decision which was made. The events are pairwise related by the *follow in time* () relation. The events are represented by DFG over some set of events .

For example:

**Example 2**

An order with a single bundle and single SKU with unit quantity has been received. Let us define the following event set : The event “order has been received” will be denoted with . The event “The order bundle is being processed” will be denoted with . The event “SKU is being processed” will be denoted with . The event “node has inventory for SKU ” will be denoted with . The event “node has capacity for SKU ” will be denoted with . The event “Reward for node has been calculated” will be denoted with . The event “Fulfilling node has been chosen” will be denoted with .

This is visualized as:

**Definition**: *DFG* *Matching* of an order fulfillment decision

Let denotes the fulfillment decision of order . Let denotes some DFG. We say that the DFG *matches* (denoted with ) if is a representation of over some set of events .

//TODO: Finish this

### Event Relationships and Causality

**Definition**: Static (*semantic) dependency between events* and

We say that event is static dependent on (denoted with ) if each instance of can exists only *in the context* of some instance of for *any* order data set . That is, removing an instance of in will remove all instances of underneath from the event tree for any chosen .

if there is a static dependency then we can define a map (called *static dependency map*) such that when .

Example of static (semantic) dependency-

Let the event denote the statement that an order with two bundles was received. Let the event denotes the statement that the current bundle being processed is the second bundle for order . Then for the order .

Example of absence of static (semantic) dependency

Let the event denote the statement that an order with two bundles was received.

Let the event denotes the statement that a splitting decision for both bundles in order is made; that is, the SKUs in both bundles are fulfilled by more than one node with order .

Let the event denotes the statement that node contains an orphan for SKU which is in with order .

Then we can write ; that is, the splitting decision for order can be reached only after order has been received. However, there is an absence of static dependency between and . The relevant for instance of event could have been received in an earlier time than that of order ; that is, . Removing the instance of corresponding to order has no impact on the existence of .

**Lemma**: Static (semantic) dependency implies *follow in time* relation

That is, . Note that the *follow in time* relation between and will hold for all pairs of instances of those events.

**Definition**: Static (semantic) descendant

We say that the event is a static descendant of another event if there exist a chain of event such that:

for some or if .

**Lemma**: *Static (semantic) dependency* defines a directed follow graph of statically associated events

Refer to the DFG shown for **Example 1** as an illustration.

**Definition**: *Dynamic dependency* betweenevents and

We say that event is dynamic dependent on (denoted with ) if all of the following is true:

* removing an instance of *may* trigger the removal of an event which is static dependent of or removal of itself.

//TODO: Finish this

Example of dynamic dependency-

In the previous Example of absence of static (semantic) dependency we considered three events - (order has been received), (node contains an orphan for SKU in the order), (split fulfillment decision has been made). Clearly, there is some kind of dependency between event and as triggering of event at a time earlier than the time the split decision is made can potentially impact the instance. Clearly, this dependency is not static as event at a time later than the time the split decision is made hence this is not a static dependency. Thus, the relation between and matches the definition of dynamic dependency.

**Definition**: Event is *associated (statically, dynamically*) with Event

We say that event is *associated* (*statically*)with (denoted with ) if each instance of is (static, dynamic) descendant of some instance of or vice versa. That is, if the

//TODO: Finish this

In order to find how a set of events are associated statically we can define a map (called *static association*) such that when .

Discussion on static association map:

How can we define the *static association* map ? The answer can be found in the definition of *Parameters of Event* given earlier. We have the parameter spaces of the two events - and and the value sets and of the corresponding event instances. Note that where . Here denotes the cartesian product of the value sets for each parameter of event ; thus, we have:

.

In general, the map should be defined over the cartesian product of the event tuples .

Let us consider this question from the context of our Fulfillment Decision **Example 1**.

Clearly, we expect that the instance and all instances under the parent instance are statically associated. We expect that each instance and all instances which are children of the current instance are statically associated as well. Let us denote the instance in this example with . Let us denote the three instances of with , , and . Then we obviously we have:

(1)

The relation (1) represents the fact that each child is statically associated to its parent.

Additionally, we can write:

(2)

The relation (2) represents the fact that the children of the same parent are statically associated.

Similarly, we continue with writing the static association relations involving the instances of

(3)

Additionally, all instances of the same parent are statically associated with each other - we write this as:

and (4)

Here denotes the index set of .

Also, all instances are associated with their grandparent which is expressed with:

and (5)

Let us construct the map for the sets (1)-(5). We start with (1):

//TODO: finish this

Discussion on causal association between events

What does it mean that certain event types can be associated causally with each other? Let us consider two event types - and . Per our definition of *Parameters of Event* given earlier the event is characterized with the pair where is the template of the event which together with the parameter set uniquely identifies this type of event. Similarly, we will consider another event type represented by . Now let us consider a given order data set and let us assume we have Fulfillment Optimization engine processing the set of orders sequentially thereby generating order metrics events. Let us denote with the ordered set of values which correspond to each parameter for all instances of generated using . Similarly, with we denote the ordered set of values which correspond to each parameter and generated for all instances of using order data set . For an instance of we will denote the values of the instance parameters with . Thus, for each instance of in (denoted as ) we have . Similarly, for we have .

**Definition**: *Causal association* *between events* and – Given the dataset we say that and are *associated (causally)* if one of the following is true:

* both and are *causes* of another event in //do we need this?
* both and are *caused* by another event in //do we need this?
* either *causes* or *causes*
* there is a semantic association between and

Note: we denote causal association between the events and with the symbol i.e. .

For example, a *prima facie causal association* implies that all causal relationships in its definition are *prima facie causes* (defined in the paragraph below).

**Definition**: *Conditional probability of an event*

Let us consider the event type . Per our definition of *Parameters of Event* given earlier the event is characterized with the pair where is the template of the event which together with the parameter set uniquely identifies this type of event. Similarly, we will consider another event type represented by . Now let us consider a given order data set and let us assume we have Fulfillment Optimization engine processing the set of orders sequentially thereby generating order metrics events.

Let us run the Fulfillment engine with the given order set and we find that in out of the instances in which event has occurred there has been an instance of *associated with* each instance of .

Then given the data set the relative frequency of occurrences of given is obtained as:

(6)

We say that the relative frequency given is an estimate for the conditional probability i.e.

(7)

**Definition**: Event is *prima facie cause* of Event

Given the data set let us denote with the set of instances of which follow the set of instances of , denoted with. That is, .

We say that event is a *prima facie cause* ofevent (denoted with ) *iff*:

the sets and are non-empty

**and**

(8)

**Lemma**: *Prima facie* cause between event and event implies dynamic dependency between the two events

That is, .

**Definition**: Event is -*spurious cause* of an Event

Let us consider the event type given with its template and parameter space .

Let us consider another event type given with its template and parameter space .

Given the data set we denote with the set of all events with which is associated such that .

Let is an event such that:

* it is not necessarily in :
* is reachable from i.e.

Then we say that is-spuriouscause of an Event *iff*

* over
* over

We denote -spuriouscause with

**Definition**: Event is *Suppe’s* cause of an Event (a.k.a. *Suppe’s causality*)

We define and as in the definition of -*spurious cause*.

Given the data set we denote with the set of all events with which is associated such that .

Let is an event such that:

* it is not necessarily in :
* follows i.e.,

Then we say that Suppe’scause of an Event *iff*

* over (*Suppe*’s causal relation hypothesis)

We denote *Suppe*’scausality relation with

**Definition**: Event is *Eells* cause of an Event (a.k.a. *Eells’ causality*)

Let us consider the event type given with its template and parameter space .

Let us consider another event type given with its template and parameter space .

Given the data set we denote with the set of all events with which is causally associated such that .

Additionally, we define the following causal *background contexts* . Those are formed by holding fixed the set of all factors causally associated with . For instance, given a set of three associated with events one possible background context will be

Let is an event such that:

* it is not necessarily in :
* is reachable from i.e.,

Then we say that is *Eells-*caused by Event *iff*

over ( )

We denote *Eells*-causality with

**Definition**: *Average Degree Of Causal Significance* (*ADCS*) of event for event in given context

The *Average Degree Of Causal Significance* (*ADCS*) of event for event given the background contexts is defined as:

We use the Latin capital letter to denote *ADCS* from the Lat. *significatio* (significance).

**Lemma**:

Static dependency implies *Eells* causality. However, *Eells* causality does not imply static dependency.

That is, if then it is true . However, if it does not necessarily follow that .

Example of *Eells*-causality-

Let the event denote the statement that an order with one bundle was received. Let the event denotes the statement that the capacity feasible nodes for order are node and node .

//TODO: finish this

***Note***: the *caused by* relations do **not** impose a total order; that is, for **every** pair of events and it does **not** follow that either or is true. Therefore, a set of events cannot be visualized as an ordered sequence; instead, we will use *Directed Causal Graph* for the purpose.

### Directed Causal Graphs

**Definition**: *Directed Causal Graph (DSG)*

A directed graph in which each node represents an Event Type, and each arc represents causal relation. Each arc is labeled with causal significant factor, a real number between and , describing how significant is the causal relationship between the two event types.

## Problem Statement for Root Cause Analysis of Fulfillment Decisions

The goal of the RCA algorithm applied to Fulfillment Optimization events is to understand and analyze causal relationship between predefined set of events based on the order metrics payloads. Each detected causal relationship will be assigned a significance factor which will indicate based on the supplied dataset how significant was this causal relationship inferred from the dataset and the configured set of events . Thus, the result of a single RCA algorithm run with a given dataset will be a Directed Causal Graph instance , where the vertex set will be a subset of the events set i.e., . Each arc will represent a causal relationship between the connected events, and it will be labeled with a causal significance factor (abbrev. ).

For instance, for the set of events shown earlier (see *Set of events for analysis of the cause of splits in Fulfillment Decisions*) we can have the following output of the RCA algorithm:

//TODO: finish this

## Algorithm For Root Cause Analysis

*Brief description of the RCA algorithm*

1. Choose a set of events of interest. ,
2. Compile order sequence from the given events dataset
3. Using the given dataset create Directed Follow Graph instances for each order in the dataset.
4. From the created , construct Aggregated Directed Follow Graph with the set of events of interest ,
5. Using Eell’s definition of causality calculate the Average Degree of Causal Significance (ADCS) for each pair of nodes in using the already calculated in 3. frequency counts for each pair of events in .
6. Given a minimum significance level construct Directed Causal Graph (DCG) using and for each pair of events in such that every arc in will have significance factor larger or equal to .

//TODO: finish this

## Appendix: Probabilistic Temporal Logic

Probabilistic Temporal Logic is a tool for state machine model checking which is a more complete alternative of the Labeled DFG defined earlier. A somewhat reduced subset of Probabilistic Temporal Logic is defined with the help of *Kripke* structures. With randomness introduced *Kripke* structure is roughly equivalent to a Discrete Time Markov Chain, and it is just another tool to validate specific first order logic statements relevant for RCA against our process model.

**Definition**: *Kripke structure*

Let be a set of atomic propositions. A *Kripke* structure over is defined as the tuple where

* is a finite set of states
* is the set of initial states
* is a total transition relation, such that
* : is a function which labels each state with a set of atomic propositions that are true within it.

//TODO: Finish this or remove it in case we do not need to deal with Kripke structures. Most likely we will not need this abstraction as the concept of LDFG I have defined earlier will be sufficient. (Hansson & Jonsson, 1994)

# Bibliography

Clarke, E. M., & Schlingloff, B. H. (2001). Model Checking. In A. Robinson, & A. Voronkov, *Handbook of Automated Reasoning* (pp. 1369-1520). Elsevier Science Publishers B.V.

Eells, E. (1991). *Probabilistic Causality.* Cambridge UK: Cambridge University Press.

Hansson, H., & Jonsson, B. (1994). *A Logic for Reasoning about Time and Reliability.* Uppsala, Sweden: SICS Research Report SICS/R90013.

Houdth, G. V., Depaire, B., & Martin, N. (2022). Root Cause Analysis in Process Mining with Probabilistic Temporal Logic. *ICPM 2021 Workshops* (pp. pp. 73–84). Eindhoven, The Netherlands: LNBIP Volume 433.

Kleinberg, S., & Mishra, B. (2009). The Temporal Logic Of Causal Structures. *The Conference on Uncertainty in Artificial Intelligence*, (pp. 303-312). Montreal, Canada.

Spirtes, P., Glymour, C., & Sheines, R. (1993). *Causation, Prediction and Search.* New York: Springer Verlag.

## Downloadable Links for the Bibliography

(Clarke & Schlingloff, 2001): [box location](https://nike.box.com/s/tbsi5sl9eaewvzch6jhm0xqpezkbs2vy)

(Eells, 1991): [box location](https://nike.box.com/s/pdotw4v4hfqrp366g00eli46x5zt76cr)

(Hansson & Jonsson, 1994): [box location](https://nike.box.com/s/v0ffvsisd4vax6n241nkqk6xji3yxdfo)

(Houdth, Depaire, & Martin, 2022): [box location](https://nike.box.com/s/5mzr05f2pwm033npe0eonlrkuohn95ce)

(Kleinberg & Mishra, 2009): [box location](https://nike.box.com/s/x6vq7z281hl6h1z5pf63ynrlcx9h9vfa)

(Spirtes, Glymour, & Sheines, 1993): [box location](https://nike.box.com/s/3y23dyhch3zxykalqn6a09lsrnsswr9w)