Bottleneck Patterns in Provenance

Sara Boutamina[1], James D. A. Millington[2], and Simon Miles[1]

[1] Department of Informatics
[2] Department of Geography
King's College London, London WC2B 4BG, UK
{sara.boutamina, james.millington, simon.miles}@kcl.ac.uk

Abstract. A bottleneck, in general, is a point of congestion in a system which impacts its efficiency, productivity and may lead to delays. Identifying and then fixing bottlenecks is an important step in maintaining and improving a system. To detect bottlenecks, we must understand the flow of processes, and dependencies between resources. Thus provenance information is an appropriate form of input to address this matter. In this paper, bottleneck patterns based on provenance graphs are proposed. These patterns are used to define the structures bottlenecks may take based on their classification, and offer a way to detect possible bottlenecks. An example from soybeans distribution is used to illustrate this preliminary work.

Keywords: Bottlenecks · Bottleneck Patterns · Provenance Graphs

1 Introduction

Bottlenecks can have a great impact on systems leading, for instance, to an increase in the production and distribution times and costs, and hence, a reduction in satisfaction of customers. In global food distribution, for example, detecting bottlenecks in processes can be an important step in solving problems such as preventing food wastage.

There is an ambiguity in the literature regarding the definition of bottlenecks despite the existence of much work, especially in manufacturing, dealing with this topic. In particular, there is a conflation between the definition of bottlenecks and the methods used to detect them (e.g. [1], [2]). For instance, in *the inventory definition* [1], bottlenecks are considered to be where there is a queue of the most of work waiting to be processed. This definition focuses on how to identify a bottleneck based on the waiting queue before a station rather than on what independently characterises it. Moreover, the use of different characteristics to describe bottlenecks may lead to different bottlenecks being identified [2]. To have an acceptable definition of bottlenecks, the authors in [3] suggest that it is important to reduce the level of detail integrated in such definition. They propose the following definition: "The bottleneck of a system is the element (node or edge) that limits the system in attaining higher throughput beyond a certain threshold. This threshold is determined by the bottleneck's physical throughput capacity, organizational rules, or operational practices." [3]. It is notable from this that they consider that a bottleneck may occur not only because of physical capacity of a resource, for instance, but also because of the organisation of work at this resource.

In this poster, we present initial work towards specifying bottlenecks in terms of patterns within provenance graphs documenting past performance in a system. First, a classification of bottlenecks is considered, then bottleneck patterns, corresponding to provenance subgraphs, are proposed to match and suit this classification.

The patterns are illustrated using the movement of soybeans between a farmer in Brazil and a consumer in China. A trading company, acting as an intermediary, bought soybeans from the farmer, then transported it to their stores not far from the production area. After that, the soybeans were transported from those stores to port, to be exported to China. This company delegated to two transport companies to transfer soybeans firstly from farm to their storage units, then from these units to port terminal managed by the trading company.

2 Classification of Bottlenecks

In order to determine the forms of pattern to look for in the provenance, we need to first consider the types of bottleneck that can exist. We derive our taxonomy from existing classifications [3]. *Tangible bottlenecks* impede higher system throughput due to physical limitations. System elements that are tangible bottlenecks can be either *active* or *passive*. Active tangible bottlenecks represent elements which can influence system throughput by their own actions and behaviour, such as workers in food distribution. On the other hand, passive tangible bottlenecks are not able to change the throughput themselves, since they do not have the power to do so, but represent other physical limitations, such as streets slowing down transportation due to high traffic. *Intangible bottlenecks* represent processes that prevent higher system throughput, for instance, because of their poor design.

Other ways of classifying bottlenecks exist, not currently translated into provenance patterns in our work. For example, comparable but not quite the same as the above classification, one can distinguish *organisational*, *physical* and *operational* bottlenecks, corresponding to limitations in the way activities are planned, physical capabilities, and the way that work is conducted, respectively. A bottleneck can also be either with an *external* or *internal* locus of control, based on whether the measures used to manage bottlenecks are inside or outside the organisations. We can also classify bottlenecks on whether they are *unavoidable* at points when demand for a resource exceeds its capacity, or *avoidable* when its emergence is due to, for instance, careless preparation.

3 Bottlenecks Patterns

In this paper, we consider that the bottlenecks to be detected are classified as tangible (active and passive) and intangible. To structure and specify these bottlenecks, patterns are proposed which are defined as:

Definition 1. *A* bottleneck pattern *is a set of rules that apply to a provenance sub-graph indicating whether it presents a bottleneck in the process documented.*

These patterns are specified using PROV which is defined as "a specification to express provenance records, which contains descriptions of the entities and activities involved in producing and delivering or otherwise influencing a given object" [4]. It is based on three main core concepts which are *entities*, *activities* and *agents* [5]. Entities represent digital, physical or other things; activities are actions dealing with entities to create new ones or to use existing ones; and agents correspond to something that was responsible for an already happened activity, an entity or another agent.

We assume that the provenance graph within which we are trying to detect bottlenecks is representative of the process' past behaviour, or is a representative aggregation of information from multiple provenance documents effectively documenting the process' multiple executions over time. Each entity, activity and agent in a PROV graph can have attributes, which we use to record, for instance, past localised capacity and demand. The way in provenance graphs may be aggregated and the ontology for expressing capacity and demand is beyond the scope of this poster.

For each of the three bottleneck types in the classification, we have a provenance pattern. These are defined informally at this point, but we illustrate their effect with the case study. A bottleneck pattern comprises a *potential bottleneck pattern*, identifying a PROV graph structure within which a bottleneck of the given type may be found, and then rules over the attributes attached to the nodes of the structure, such as demand and capacity, which can be used to confirm the bottleneck exists. Examples of possible bottlenecks, based on the proposed patterns, from the soybeans distribution example are presented in Figures 1–5.

- A potential intangible bottleneck may occur when there is an agent or a number of agents acting on behalf of another agent (e.g. Fig. 1). Moreover, it may occur when there is a collection representing for instance a regulation, and there is at least a member who had business with a non-member (e.g. Fig. 5). *Being a member* of this collection means that the agent had accepted and signed the regulation, and *having a business* means that there is an activity which is associated with at least two agents where one of them is a buyer and the other is a seller.
- A potential passive tangible bottleneck may occur when there is an entity used by one or more activities (e.g. Fig. 2). Moreover, it may occur when there is a collection representing for instance a line of production where the members (such as machines) play a role to define the bottleneck.
- A potential active tangible bottleneck may occur when one or more activities are associated with an agent (e.g. Fig. 3) or when one or more entities are attributed to an agent (e.g. Fig. 4).

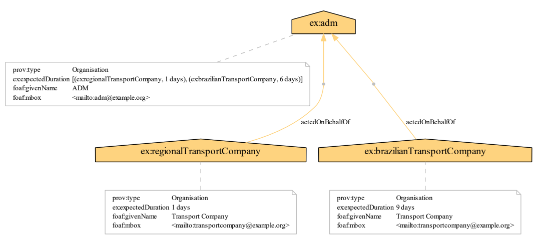

Fig. 1: Agent-Agent relationship example

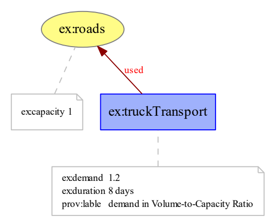

Fig. 2: Entity-Activity relationship example

Within the case study, we see bottlenecks matching the rules applied to attributes in the potential bottleneck patterns, illustrated as follows. In Fig. 1, the trading company, ADM, delegated to two transport companies with the goal that soybeans were collected from a farm in one day, and after that reach port terminal in six days. The first company acts according to the set goal, but the second had another goal (9 days), and this discrepancy is an intangible bottleneck. In Fig. 2, roads are a resource shared between activities, with a volume-to-capacity ratio representing the number of vehicles passing through divided by the number of vehicles specified in the designed capacity (if it is greater than 1, then it is over capacity). The farmer was responsible for different entities with damage percentages not to exceed (e.g. the expected percentage of damage in storing bags was 10%, however it was 25%) (Fig. 3), and different activities which would be completed in specific durations (e.g. the expected duration for harvest was 31 days, however it took 40 days) (Fig. 4). As an example of an intangible bottleneck, the Soybean Moratorium (SoyM) [6] is a voluntary agreement to boycott trading soybeans grown on lands deforested after July 2006 in the Brazilian Amazon. SoyM is represented as a collection of companies, traders, etc. who signed it (Fig.5). Members of this collection had to avoid trading with other parties which did not belong to this collection.

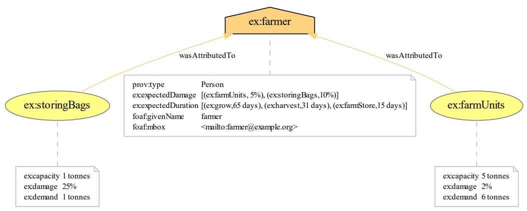

Fig. 3: Agent-Entity relationship example

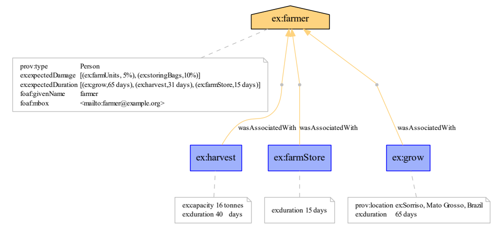

Fig. 4: Agent-Activity relationship example

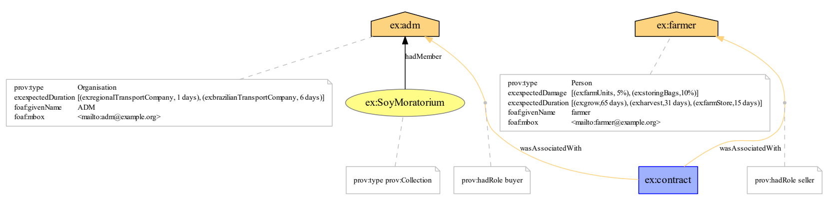

Fig. 5: collection example

References

1. Lawrence, S.R., Buss, A.H.: Economic analysis of production bottlenecks. Mathematical Problems in Engineering **1**(4) (1995) 341–363
2. Li, L., Chang, Q., Ni, J.: Data driven bottleneck detection of manufacturing systems. International Journal of Production Research **47**(18) (2009) 5019–5036
3. Beer, J.E.: Analysis and management of bottlenecks in supply networks: Towards a structured approach to stabilization of inbound material flow. (2015)
4. W3C Working Group: PROV Model Primer (2013)
5. Moreau, L., Groth, P., Cheney, J., Lebo, T., Miles, S.: The rationale of PROV. Web Semantics: Science, Services and Agents on the World Wide Web **35** (2015) 235–257
6. Gibbs, H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., Soares-Filho, B., Barreto, P., Micol, L., Walker, N.F.: Brazil's Soy Moratorium. Science **347**(6220) (2015) 377–378