TOWARDS ROBUST PART TRACKING IN AN AUTOMATED MANUFACTURING PROCESS USING RFID

James Brusey Duncan McFarlane

Institute for Manufacturing, Cambridge University Engineering
Department, Mill Lane, Cambridge CB2 1RX, UK

Abstract: Automated Identification and in particular, Radio Frequency Identification (RFID) promises to assist with the automation of mass customised production processes. RFID has long been used to gather a history or trace of part movements, but the use of it as an integral part of the control process is yet to be fully exploited. Such use places stringent demands on the quality of the sensor data and the method used to interpret that data. In particular, this paper focuses on the issue of correctly identifying, tracking and dealing with aggregated objects with the use of RFID. The presented approach is evaluated in the context of a laboratory manufacturing system that produces customised gift boxes.

Keywords: RFID, manufacturing plant control, mass customisation

1. INTRODUCTION

Customers are increasingly buying goods over the Internet, but still expect to be able to mix and match components as they did at a store. An example would be the purchase of a new computer where there are a number of parameters (memory, hard disk, video card) for the customer to select. This sort of customised manufacture is often referred to as *late-stage customisation* since all of the different options that the customisation provides can be handled during the last phases of the manufacturing process (Tseng and Piller, 2003).

Given this trend toward more flexible production processes where hundreds of different types of end-product are produced by combining component parts in different ways, any automation will need to be more sophisticated. Without having separate lines for each product type, and assuming that the end-product is produced to order rather than to stock, there is the need to rapidly switch between one sort of operation and another. Implicitly, such flexible machines must be able to quickly determine what operation to perform. At least some automation of such built-to-order production is certainly achievable. At Dell Computer's OptiPlex plant, for example, the process of transporting parts around the factory is automated in such a way that each workstation receives only the parts it needs when it needs them (Perlman, 2001). However the final product assembly is still a manual process.

Completely automating such late-stage customisation requires more intelligent automation and better sensory information than have traditionally been available. This is because the decision making in a customisable process does not depend on the mere presence of the part, but on which type of part, and sometimes on the specific identity of that part. For example, computer chassis A will be shipped to customer X, who requires 256Mb of memory, while computer chassis B will be shipped to Y, who requires 1Gb of memory. Thus when a computer chassis arrives at a workstation where memory chips are inserted, A must be treated differently to B. The general problem of establishing and keeping track of the identity and location of physical objects is referred to here as the *tracking problem*.

A system that tracks items must provide, on demand:

1. the *location* of an item (where is the memory chip for chassis B?)
2. the *state* of an item (does chassis A already have memory installed?)
3. the *identity* of an item at a particular location (which chassis is this?).

For such a system to be useful in the control of a manufacturing process, the location, state and identity information must be as complete and accurate as possible. Also, the information must be provided in a timely manner to avoid delaying the control process. It can be derived purely from sensor data or from a model of the process, or some combination.

A sensor driven approach has the disadvantage that many sophisticated sensors will be required. They need to be sophisticated in the sense that they will need to be able to tell the type of part, sometimes identify the item uniquely, and possibly estimate its state. There will need to be many of these sensors since every operation will need to be able to check which part or parts are involved.

On the other hand, a purely model-based approach may have difficulty dealing with even small deviations between the model and reality. For example, the model of a car plant may say that car A is followed by B and then C. But when A is removed from the line to fix a fault, if no sensors tell it otherwise, the model may continue to show the order as A,B,C whereas it has become B,A,C. The consequences are trivial when merely trying to trace production progress but catastrophic when they affect the production process.

This paper provides building blocks towards an approach that combines both sophisticated sensors that supply identity information with a model-based approach that allows parts to be tracked when they are out of range of the sensors. In particular, it focuses on the issue of tracking the location not merely of individual objects but aggregates. The following section provides some background on Radio Frequency Identification (RFID), which is used as the main sensor technology. Note that there are several other possible sensors that could be used to determine the identity of products, such as bar code scanners or vision systems. Section 3 presents an approach to deriving a meaningful model of the structure and contents of aggregated objects based on the use of RFID sensors. This approach is then evaluated in the context of a laboratory manufacturing system developed at Cambridge University.

2. BACKGROUND

2.1 RFID Primer

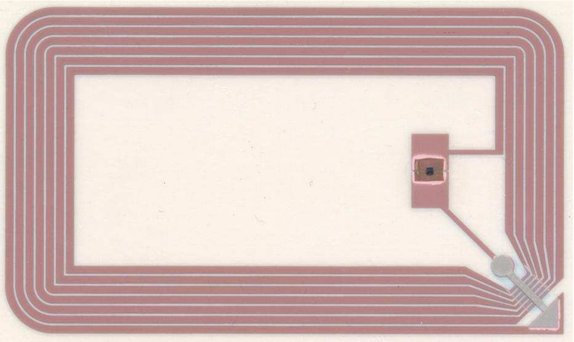

Fig. 1. Example of a passive RFID tag

RFID or Radio Frequency Identification (Finkenzeller, 2000) is a technology originally created for friend or foe transponders in aircraft during the second world war. It involves an asymmetric RF transmitter / receiver pair, where one is, on request, transmitting its identity to the other. The identity transmitter is usually referred to as a *tag*, whereas the identity receiver is known as a *tag reader* or sometimes simply *reader*. As long as the distance between tag and reader is small (within about 0.5 metres for HF or 10 metres for UHF), it is possible to use a *passive tag*, one that has no battery of its own. Passive tags (such as the one shown in figure 1) operate by absorbing some of the energy in the RF signal transmitted by the reader, and then transmitting back a short message. A key advantage of passive tags is that they are relatively small (around 50mm square and less than 1mm thick), and inexpensive. This paper deals exclusively with passive tags.

Since passive tags operate by absorbing energy in the RF signal, they tend to operate in bursts rather than continuously. Their response will depend on the local RF signal strength and their orientation relative to the local direction of the RF field. In turn, the local signal strength and direction will depend on what other objects are nearby. Specifically, conductive objects such as metal or liquid filled containers, will distort and deflect the field.

When two or more tags exist in the RF field, they may try to reply at the same time, distorting each other's response. This effect is referred to as a *collision*. Various algorithms exist for dealing with tag read collisions. The simplest of which involve causing the tags to wait for some amount of time before retransmitting (e.g. ALOHA (Finkenzeller, 2000)), while more sophisticated algorithms query specific ranges of tags until only a single one replies (e.g. binary search (Finkenzeller, 2000)). Even using such sophisticated algorithms, increasing the number of tags in a field will have the effect that any specific tag will be detected less often. In the worst case, and particularly for items moving quickly past a reader, it is possible for some items not to be detected at all.

2.2 *Related Work*

Hodges *et al.* (2002) developed one of the first laboratory manufacturing systems at Cambridge to make use of RFID in a customised manufacturing process. Their approach was to place RFID readers prior to every decision point. This meant that no model of the system was required and state information was limited to recent RFID tag reads. As each product arrived, it was processed according to its type.

When extending this system to allow products to flow into and out of each manufacturing cell—essentially dealing with routing of parts to appropriate destinations—it was discovered that some knowledge of the state of the system was required to avoid deadlocks (Brusey *et al.*, 2003*b*). Also, as processing times extended, more failures were traced to unreliability in the RFID sensory data. Two types of errors in interpreting RFID tag reads were identified: *false negatives* where an item is in range but not detected, and *false positives* where an item is outside the expected range but is still detected (Brusey *et al.*, 2003*a*). In that work, a simple filter was suggested. Further experimental results have been presented by Floerkemeier *et al.* (2003), who have extended the approach to use Bayesian techniques. Also, Hähnel *et al.* (2004) have implemented a variant of Monte Carlo Markov Localisation to make use of RFID tags as landmarks for a mobile robot.

The use of Petri nets in modelling manufacturing processes is well established (DiCesare, 1993). They are implicitly used to track the movement of parts for control purposes by supervisory Petri net control approaches (Cassandras and Lafortune, 1999).

3. RFID FOR AUTOMATED PART TRACKING

The work described in this paper represents a specific aspect of an activity examining the impact of RFID on tracking processes. In this section, an approach is presented to integrate RFID sensor data with a representation of the state of the manufacturing system and a model of how that state is changed. The aim of this approach is to enhance the accuracy of the identity information and thus improve the robustness of the manufacturing system. It relies on the fact that parts are not always seen in isolation, but often travel together. A common example is that of pallets and cases. Two cases on the same pallet will tend to both be detected by RFID sensors at around the same time. Similarly, the pallet will be detected along with the two cases. All together they form an *aggregate*. Aggregated objects provide an opportunity to improve the reliability of RFID information.

When considering aggregates, structure plays some role. For example, it is easier to remove a case from a pallet than to remove the pallet from underneath several cases. Typically this structure is hierarchical. A pallet may contain several cases, each of which may contain some bottles. When a case is removed from the pallet, those bottles that were in the case will move with it.

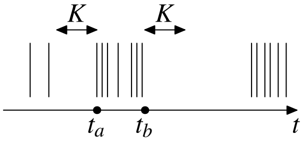

Fig. 2. Identifying an aggregate based on the time associated with a series of tag reads.

In general, to understand how parts and containers of parts move from one location to another, some form of model is required.

3.1 *Discovering Aggregates*

In this paper, a time-based aggregation approach is proposed. This approach relies on constraining the flow of each aggregate as it moves past the RFID tag reader. Specifically, there must be a delay both before and after each aggregate is detected by the reader where no tags are detected. In addition, while the aggregate is "seen", the associated tag read events should not be separated by too much of a delay.

Define a string of tag read events occurring at a particular tag reader r as

$$s(r) = (e_1,t_1),(e_2,t_2),\ldots,(e_n,t_n),$$

where e_k is a tag read event that occurred at time t_k. This string is ordered by time, and so $a < b \Rightarrow t_a \leq t_b$. If the aggregate moves past the reader over a particular interval of time, and if there are no other tagged objects within the read range at the same time as the aggregate, then the set of tag read events associated with a particular aggregate must be separated in time from other read events. Formally, the read events $\{e_a,\ldots,e_b\}$ belong to a single aggregate iff $t_a - t_{a-1} > K$, $t_{b+1} - t_b > K$ and also that $t_{i+1} - t_i \leq K$ for all $a \leq i < b$, where K is a suitable time interval. In other words, the time interval without any tag read events before and after the aggregate being seen is at least K. Also, the largest interval between any two successive tag reads within the aggregate between e_a and e_b is less than K. An example of how this works is shown in figure 2.

The choice of the parameter K must be sufficiently large to ensure that a single aggregate is not considered to be two separate objects but at the same time, sufficiently small so that two aggregates arriving one after the other are not considered as though they were a single object. For example, K should be large enough so that if a tag at the leading edge of the aggregate is seen as soon as the part moves into the field followed by a tag at the trailing edge being seen when the aggregate leaves the field, then the aggregate is still

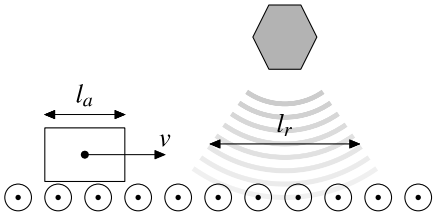

Fig. 3. Example of an aggregate moving along a conveyor, past an RFID antenna. The size of the aggregate l_a, its velocity v, and the size of the antenna's field l_r determine the minimum K parameter.

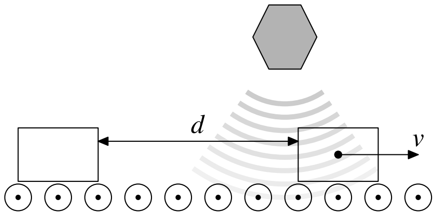

Fig. 4. The distance between items d constrains the maximum value of K.

seen as a single object. Specifically, for an aggregate of length l_a travelling at an even positive velocity v past a read field of length l_r (as shown in figure 3), then we require that

$$K > (l_a + l_r)/v.$$

In some cases, it may be necessary to constrain the flow of aggregates to ensure that each arrives at the reader a small time after the prior aggregate has moved out of the way. Specifically, let the distance between two aggregates (from trailing edge of the first to leading edge of the next) be d, as shown in figure 4. Then an additional constraint is $K < (d - l_r)/v$ which can be rearranged to give a spacing requirement of

$$d > Kv + l_r.$$

Another issue is that of whether it is allowable for the aggregate to stop near the reader. The main difficulty with this is due to the existence of regions near the reader where a tag can be placed indefinitely without generating a tag read (Mallinson, 2003). For the above approach to work, it would be necessary to set K to be at least as large as the maximum time spent stopped.

One reason that it may be necessary to slow down or stop the aggregate as it passes through the read range is to allow all of the tags to be read. If an aggregate involves many sub-components, and at least some components are tagged, then multiple tags will be in range of the tag reader simultaneously. Obviously if all tags attempt to respond simultaneously then their signals will interfere. For this reason, tag readers and tags typically employ some form of anti-collision protocol, such as ALOHA or binary search (Finkenzeller, 2000). ALOHA is the simplest mechanism and relies on each RFID tag only responding intermittently thus reducing the probability of a collision. However as the number of tags increase, the length of time needed to be reasonably confident that all tags have been detected also increases. For a 99.9% confidence level, Finkenzeller suggests that for HF tags, 0.5 seconds is required to see 2 tags, whereas for 8 tags, 2.7 seconds is required. Different anti-collision protocols have different characteristics but all require longer periods to recognize larger numbers of tags.

Given that collisions and other environmental factors may result in some tags in the aggregate being missed, tracking the movement of the aggregate, rather than the individual part, allows such missed tag reads to be inferred. This is a key benefit of this approach.

Once aggregates have been discovered, prior knowledge about the characteristics of the tagged objects can help to infer the aggregate's likely structure.

3.2 *Inferring Containment Relationships*

When a set of objects form an aggregate, it is usually the case that at least one of the objects acts as a container. For example, a pallet that supports cases can be considered to "contain" those cases, in the sense that if the pallet moves, then so do all of the associated cases. The converse is not necessarily true. Sometimes a case will be removed from a pallet. The notion of containment is naturally hierarchical, and so cases may contain, say, bottles of jam. When the case is removed from the pallet, the bottles contained within that case will move too.

In any given application, there are typically only a few levels of the containment hierarchy, and also only a few ways that containment can occur. To infer the likely containment structure, it is usually sufficient to know the likely containment *level* of each type of object. For example, a pallet might be of level 1, a case of level 2, and a bottle of level 3. In an application where a bottle should never appear on a pallet on its own, the appearance of a single pallet, a single case and a single bottle allows us to infer that the bottle is contained by the case and that the case is contained by a pallet.

When several items exist at the same level, for example, two cases are detected but only a single bottle, then it is not possible to infer the location of the bottle. However it is possible to say that, in the absence of any other information, that it is equally likely for the bottle to be in either case. This probabilistic representation of the position of the bottle may not be useful immediately, but if subsequently one of the cases is removed, and the pallet subsequently passes by a reader, the absence of the bottle at this stage implies that the bottle is more probably in the case that was removed.

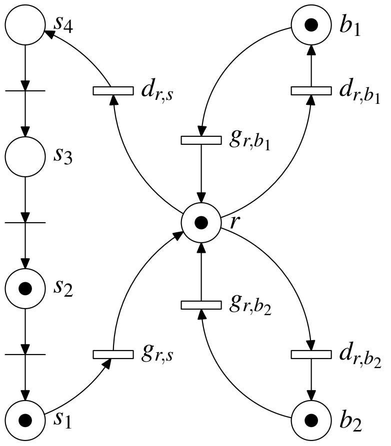

Fig. 5. Part of the Petri net transition model.

3.3 *Transition Model*

To estimate the change in state caused by an action, some form of model is required. A form of high-level Petri net (Murata, 1989) was used to describe the part tracking problem examined in this paper. This net represented the possible locations for parts as *places* while actions are represented by *transitions*. A *token* in a place represents a part being at a location. Since parts are identified uniquely, the corresponding token has a corresponding identity. A portion of the model used is shown in figure 5. Note that, for conciseness, the token identity is not shown in the diagram. Controlled transitions, shown as boxes in the diagram, are labelled with the corresponding action. Uncontrolled transitions, unlabelled and represented as lines rather than boxes, can occur at any time as long as there is a part at the source place and nothing at the target place. Requiring that the target place be empty is not usual for Petri nets, however it is helpful here since tokens correspond to uniquely identified items, and their order, for example in the work-in-process buffer, must be preserved in the model. When a transition fires, a part is moved from the transition's input place to its output place. For example, a typical action is for a robot gripper r to "grab" an item from the stack s, and this is denoted $g_{r,s}$. The places correspond to locations, such as r for the robot gripper, $s_1, \ldots, s_4$ for the four positions in stack s, and b_1 and b_2 for two locations in a box. There are two types of action shown in the diagram: $g_{r,x}$ being a "grab" from x to r, $d_{r,x}$ being a "drop" from r to x.

A key issue with the development of the transition model was the correct handling of asynchronous updates to the world state from the transition model and RFID sensors. Network and processing delays can mean that the last few tag reads for a part that has just been moved away from a reader arrive after the transition model has updated the location of the part. In early versions, the part apparently "jumped" back to its previous location. To resolve this, RFID reads events are timestamped close to the source and any events older than updates from the transition model are ignored.

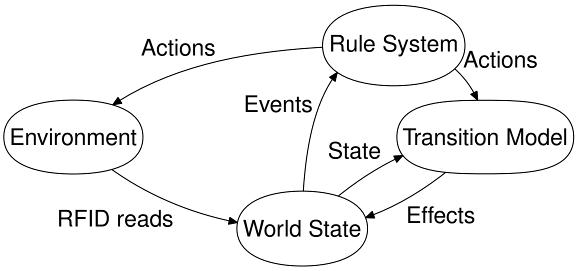

Fig. 6. Using an internal world state representation and a transition model to track the effects of actions. The world state tracks the location of parts first by interpreting RFID tag reads, but also by interpreting the effect of actions on the current state using a transition model.

Note that actions are not derived from the transition model, but rather come from a reactive rule system. The interaction of RFID sensor data, the transition model, the rule system and the world state representation is shown in figure 6.

4. EVALUATION

To evaluate this approach, it was applied to the Cambridge laboratory manufacturing system mentioned previously. This system packs Gillette™gift boxes. As with previous developments, it packs to order rather than to stock. It extends earlier work by both routing parts and boxes to the appropriate cell and flexibly handling the packing operation of a single box across several cells. It also removes the finished product from the shuttle and puts it into a warehouse. The order can be changed at any stage during production, causing the gift box to be repacked in an efficient manner. A schematic diagram of part of the manufacturing system is shown in figure 7.

To allow the location of parts to be identified, RFID tags are attached to the individual items, the boxes, the trays carrying the boxes and the shuttles. RFID readers are positioned at the base of the work-in-process (WIP) stacks (see figure 8) and along the monorail track just prior to the gates and docking stations. Although the original design called for readers prior to every decision point, some readers were able to be disabled, although some slight changes were required to the transition model to cope with this.

4.1 *Results*

A statistical summary of logs produced by demonstration runs of the laboratory manufacturing system is given in table 1. A demonstration of the system usually takes about 30 minutes and consists of placing several orders to demonstrate the ability of the system to cope

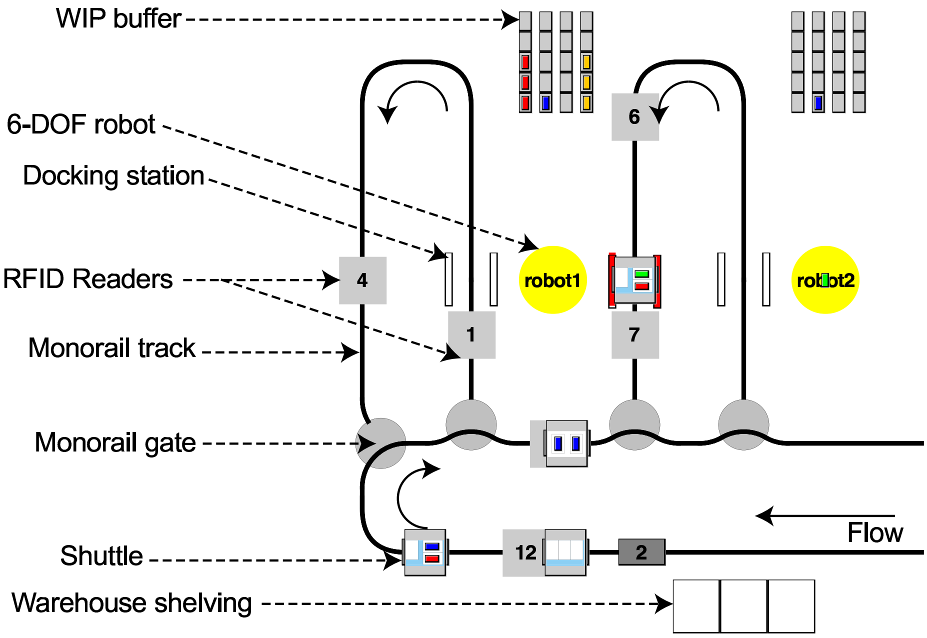

Fig. 7. Plan (schematic) view of part of the gift-box packing system.

Fig. 8. Work in process buffer for packing robot.

with customised demand, and then changing the orders to demonstrate its ability to react to a changed demand. It is reasonably common for the system to receive a false positive RFID read in the work-in-process stack, since the WIP tag readers sometimes read the item second from the base of the stack as well as the item at the base. This leads to two items being considered to be at the base of the stack. Roughly half the time this is resolved when the probability estimate for one of the items reduces below a threshold (a value of 0.2 was used for this threshold). Since only a single item can fit at the base of the stack, the probability of an item being at the base decreases when another item is detected there. In the rest of the cases, the uncertainty was removed after an action was taken to move the item at the base, and subsequently one of the items was detected elsewhere.

The process of forming an aggregate has proven useful in reducing problems caused by false negatives for a shuttle tag. Although the shuttle tag is in close proximity with each reader as it passes by, it is sometimes the case that the shuttle tag is not detected at all. Since seeing the shuttle tag is used to identify the movement of the shuttle and therefore to take actions such as switching a gate, it is critical that the shuttle can be identified. Based on previously gathered aggregate information, it was possible in most cases to correct for the missed tag and thus to keep operating without intervention. The two cases where this was not possible occurred when the shuttle tag was missed on the first occasion that the aggregate was seen.

Table 1. Accumulated results from 55 demonstration runs

Total running time (minutes)	1 649
Item movements detected or inferred	109 366
Actions taken	2 683
False positives in work-in-process stack	106
False positives pruned by probability threshold	57
False positives pruned after item movement	49
False negatives for shuttle corrected	41
False negatives for shuttle not corrected	2

5. CONCLUSION

RFID is a mature technology that is currently seeing a rise in prominence, largely due to its increased use in the retail sector. It has been applied to manufacturing, however it is mostly used as a means of establishing the geneology or history of the end product, rather than as a mechanism to support the automation of customisable production. However increased consumer demand for customisation may drive manufacturers to adopt RFID as a central part of the manufacturing control loop.

Tracking RFID tags in a stateless manner has been demonstrated to be sufficient for many applications, however more sophisticated use of RFID will require the integration of a model-based approach. In particular, when RFID is used for control, its reliability can be enhanced by modelling the movement of parts and thus detecting some sensor errors. As a side effect, this can allow a reduction in the number of RFID readers required and also address the problem of any temporary failure to read tags.

REFERENCES

Brusey, James, Christian Floerkemeier, Mark Harrison and Martyn Fletcher (2003*a*). Reasoning about uncertainty in location identification with RFID. In: *Proc. IJCAI-03 Workshop on Reasoning with Uncertainty in Robotics*. Acapulco, Mexico.

Brusey, James, Martyn Fletcher, Mark Harrison, Alan Thorne, Steve Hodges and Duncan McFarlane (2003*b*). Auto-ID based control demonstration - phase 2: Pick and place packing with holonic control. Technical report. Auto-ID Centre, Cambridge University.

Cassandras, Christos G. and Stéphane Lafortune (1999). *Introduction to Discrete Event Systems*. Kluwer Academic Publishers.

DiCesare, Frank, Ed.) (1993). *Practice of Petri nets in manufacturing*. Chapman & Hall.

Finkenzeller, Klaus (2000). *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*. John Wiley & Sons.

Floerkemeier, Christian, Matthias Lampe and Thomas Schoch (2003). The smart box concept for ubiquitous computing environments. In: *Proceedings of SOC'2003 (Smart Objects Conference)*. Grenoble.

Hähnel, Dirk, Wolfram Burgard, Dieter Fox, Ken Fishkin and Matthai Philipose (2004). Mapping and localization with rfid technology. In: *Proc. of the IEEE International Conference on Robotics and Automation (ICRA), 2004*.

Hodges, Steve, Alan Thorne, Andy Garcia, Jin-Lung Chirn, Mark Harrison and Duncan McFarlane (2002). Auto-id based control demonstration phase 1: Pick and place packing with conventional control. Technical Report CAM-AUTOID-WH-006. Auto-ID Center. `http://www.autoidlabs.org/whitepapers/CAM-AUTOID-WH-006.pdf`.

Mallinson, Hugo (2003). Enhancing identity with location. Master's thesis. University of Cambridge, UK. `http://www.ifm.eng.cam.ac.uk/automation/publications/`.

Murata, Tadao (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE* **77**(4), 541–580.

Perlman, Stacy (2001). Automate or die. *Business 2.0*. `http://www.business2.com/b2/web/articles/0,17863,513636,00.html`.

Tseng, Mitchell M. and Piller, Frank T., Eds.) (2003). *The Customer Centric Enterprise: Advances in Mass Customization and Personalization*. Springer. Also see `http://www.mass-customization.de`.