A General Approach to model traceability systems in food manufacturing chains

Lucia Lo Bello, Orazio Mirabella, Nunzio Torrisi

Dipartimento di Ingegneria Informatica e delle Telecomunicazioni – Facoltà di Ingegneria –
Università degli studi di Catania, ITALY
{llobello,omirabel,ntorrisi}@diit.unict.it

Abstract

Traceability in food chains is now mandatory for companies interested in food processing. Every food chain is made up of a variable number of companies which cooperate through the exchange of materials and services. Since the guarantee about a product, doesn't depend on a single company, but it depends on all the factories of the food chain, a traceability system can provide information useful to reconstruct fully the path the food has followed during its production.

To carry out a full traceability system we propose a general approach based on a distributed collaborative information system where every company exchanges traceability data with the others over a network. XML was used as the format to represent data, for its ability to cope with data structures of different size. WebServices based technology has been adopted to interface different suppliers which communicate through HTTP protocol. A prototype has been implemented and some experiments have been performed with the aim to test the usability of the system.

Keywords: Traceability, WebService, Distributed systems, XML, Digital Signature, OPC, Food Processing.

1. Introduction

Implementing a system for traceability is very important in industrial production systems as it allows the path taken by products during all stages of production to be monitored[1][3]. This feature is very important in many critical application (e.g. the construction of aircraft, nuclear power plants, etc.) in which, in order to provide the certification of a component the possibility of monitoring the whole production chain is required.

The problem becomes even more significant with reference to the production of foodstuffis [2]. In the recent years many problems with food produced a notable loss of confidence into industrial production. Let’s consider for example the Milk, which is the basic component in the preparation of several foods, known sourcing was essential to avoid international trade bans during BSE crisis. Similarly, if we consider the Belgian Dioxin Crisis in 1999, we can realize that there was no way to prevent it, so that within 3 working days it had become a global problem. All E.U. Member States were involved in this and it became clear that, in order to keep in business, it was necessary to perform and supply to customers, quickly, very detailed traceability reports.

Different issues come out from consumers and manufacturers. Consumers ask for technology to be able to trace the paths the food they have chosen has taken during production. In this way they can obtain some additional guarantees concerning the quality and safety of the food. Manufacturers require another kind of technology to link their products to a path which provides them and others operating in the food chain with proof of origin. This feature is important mainly for the final manufacturers because often their product is the result of previous food processing performed by other companies. For this reason, operators will need to have in place systems and procedures able to identify any business / person from which they have been supplied with a food, feed, food-producing animal, or any substance to be incorporated into a food or to which their products have been supplied. Moreover, all such Information have be made available to authorities on demand.

Traceability is referred to in ISO regulations[5], but it is based on company logic, its effects only extending as far as the suppliers and clients of a specific company. A broader definition is provided in national regulations [4], where traceability is taken to mean the process a product undergoes throughout the whole food chain; re-traceability is the reverse process.

Although there are currently no regulations requiring the use of a computer-based system for traceability, (most companies adopts paper based traceability systems) it is clearly convenient to adopt distributed information technology to draw up traceability reports. For the realization of a full
traceability system, two monitoring subsystems are necessary. A first subsystem, that operates at production cell level, will record the production phases through the reading the process data from every cell. A second subsystem, that operates at factory level, will receive the data from the first system and from the system for logistics. The subsystem at factory level will prepare the traceability reports of every lot, by inserting the documentation concerning production and logistics.

The general system will prepare, for every outgoing lot, a complete documentation of both the manufacturing process and the future destinations.

Through the use of a distributed information system we can obtain two main advantages:
- A short time for delivering and processing of all documents related to the production between two companies which are in a Producer/Consumer relationship.
- The possibility to automatize all shipment and delivering of goods along all the production chain. Since the number of steps from the raw material to the final customer can be considerable, as shown in Fig.1, an automated system for the management and monitoring of the transactions can be very helpful.

![Fig. 1: The many transfers from Supplier to Customer.](image)

In this paper we will describe a solution devised to comply with the forthcoming ISO 22519 standard[6], which lays down guidelines for the design of a system of re-traceability and establishes that, for complete product traceability it is necessary to record not only all incoming and outgoing movements of the production lots, but also all the procedures and processing operations applied to them.

The traceability system proposed in this paper is designed to manage and deliver XML based data about food processing. Any software engines can access traceability data via Microsoft XML API, Java XML API, and many other cross-platform XML tools in the market. Each installation of the traceability system has a database repository to store the XML traceability data. Most enterprise database solution can import and export data in XML format without the need of external tools.

The paper is organized as follows: Section 2 describes shortly the structure of a food chain, section 3 presents the architecture of the system, section 4 discusses some issues concerning security, section 5 proposes some solution to retrieve data at the field level during processing. Finally in section 6 some conclusion are presented and some open issues are discussed.

2. The food chain structure

In a food production system, for each lot processed, n lots from an equal number of production cycles can be used. Each food chain, which may extend from nodes producing the raw materials to distribution, may also be interwoven with other chains, sharing one or more intermediate processing nodes.

![Fig. 2: Food chain example](image)
1. Communication in the production system. This requires the production system in the factory is supported by a suitable SCADA system, able to perform a continuous monitoring of all process data.

2. Multiple Business To Business communications between companies connected by a producer/consumer link in several food chains. As each company in the chain has to be able to retrace the production paths, querying not only direct suppliers but also those not directly connected, the proportions of the application scenario and the volumes of information exchanged can be enormous. It is therefore necessary to evaluate carefully the choice of technology and infrastructure that will ensure reliability, security and performance.

3. Main Architecture

The food chain traceability system is designed as a distributed computer system in which all the nodes in the chain take part. We assume that there is a single identification code for each lot in any node. We will call this the traceability code and we will discuss in section 5 some solutions to assign an unique ID to each job. Whenever the system has to trace a lot, it carries out a search for a given traceability code. The result of the search will be a list of traceability codes corresponding to the lots used, through which the search can be taken further to find other lots used in connected nodes in the chain.

There are two main features in the traceability system:

1. The lack of a reference node for each chain which is responsible of coordinating and organising the nodes it is assigned.

2. The possibility for each node to perform a search in the whole system to monitor, even in part, incoming and outgoing lots.

The lack of a reference node means that each node will only answer a query regarding the processing documents it contains, and will insert traceability codes into each reply as references for the nodes relating to steps immediately following the processing chain. Hence, if a query comes from the end of a food chain, referring to a production lot coming from a number of intermediate processing operations, to obtain a complete traceability document it will be necessary to make as many queries as there are nodes involved. What may occur is therefore a chain reaction of requests propagating back up the food chain.

Information will be acquired in the following way:
- each node will ask the nodes of directly connected chains for references to the traceability documents for the lots used (Fig 4a);
- in reply to a request it will return a document containing references to the other nodes corresponding to the lots used by the directly connected nodes.

All requests generated will finally have to be executed by the agent making the initial request (as shown in Fig.4a), which will receive a reply from the single nodes as showed in fig 4b.
It is therefore necessary to include in the replies each node returns, exact references to any other requests referring to other nodes. This is made possible by choosing the XML format to represent both requests and replies using the XML Link technique [10]. XML Link permits to contain in a XML file multiple references to remote sources through. In Fig.4b the main request is decomposed into multiples different queries for each node referenced in the main XML result document. If some nodes cannot process the sub-request, the main XML result document will be partially filled with the data incoming from the others nodes, or completely if the missing data are cached by others nodes.

The reply represented in XML with multiple references, has a number of advantages:

1. Each node is independent from the others and has no response time constraints. There is no dedicated protocol for the exchange of traceability data between two nodes.
2. As XML is an open standard supported by different Operating Systems and multi-platform development environments, each node can implement its reply to traceability requests with maximum flexibility.

Each node is therefore responsible not only for its own data but also for the quality of the traceability service it provides in collaboration with other nodes to obtain all the documentation related to a given lot.

XML format was devised to represent data in text format, so it is suitable for both web contents and information resulting from database queries. Thanks to this characteristic the data format used by XML is compatible with any hardware or software platform.

The software architecture for the food chain traceability system has to meet requests for a given traceability code by generating a traceability report. It was therefore structured as follows:

1. The QueryDispatcher module, dedicated to evaluating and sorting requests;
2. The QueryManager module, which maps the traceability code into specific requests and commissions details from the ReportBuilder;
3. The ReportBuilder module, for retrieving data from the business layer and plant layer subsystems and drawing up the traceability document to be returned to the QueryManager;

The QueryDispatcher module, developed in C#, receives a traceability code via a webform or ad hoc program.
The choice of the web service format allows our project to enable the exchange of traceability codes between nodes as long as a URL has been defined for each of them and the web service, irrespective of the platforms and languages used for the implementation. This guarantees the interoperability among different systems which exchange data in XML format.

Our web service module can thus be queried about a given traceability code directly by a remote node. This occurs when a remote request made by a user triggers off sub-requests to other nodes for lots connected with the initial request.

For each request made by an user, the system will activate a series of requests according to the amount of lots used to making up the lot whose code has been requested.

We assuming the presence of one or more UDDI(Universal Description, Discovery and Integration) server to enforce and increase the flexibility when the URI link embedded in every request has to matched to the right web service. There are many standard and improvements discussed on UDDI on[13].

3.1. Robustness

The system described provides a user making a request with a partial result even when some of the nodes involved in the search are unavailable. However, it should be pointed out that each node has a different influence on the final traceability document returned to the user, according to its position in the chain.

If we consider the example in Fig. 4a, unavailability on the part of node G would mean that the traceability report would be almost empty as compared with the complete one generated by all the nodes involved. Unavailability by node C, on the other hand, would reduce the contents of the final report by about 25%. It is therefore necessary to find a solution to improve reliability in processing nodes on which several lots from more than one production chain converge.

One solution would be to oblige each node to keep at least a partial archive of all the traceability information for the immediately adjacent nodes. This is unadvisable, however, because the number of adjacent nodes is unpredictable and the archives might be much bigger than that of the single node involved. In addition, the costs of maintaining the services on line would increase due to the greater amount of traffic converging on the node. This could multiply the cost of providing a traceability service on single nodes.

An alternative would be to add a dedicated external server on which to keep a replica of the archives, to be used in the event of faults. A redundant server for each chain could duplicate several traceability archives and would cost less than duplicating the archives of adjacent nodes.

Moreover, using a dedicated server to overcome faults occurring in the traceability servers in the single nodes may be advantageous in heavy traffic situations to reduce the workload on the duplicated nodes.

If several replications of different nodes in the chain merge in the redundant server, it is necessary to balance the load of the queries among the different replicas.

Therefore, we can define a cost function for every node replication on the redundant server of the chain, in order to obtain a suitable reference index in the allocation of its resources. Such resources are above all the ability of satisfying the requests for the traceability service. This means that the maximum number of requests the server can satisfy, must be divided according to a suitable policy, among the nodes replications on that servers.

In the case of food chains, the production lots are mainly worked foods or to be worked. In this area, most of the produced lots has a life cycle no longer than four years. In addition, most of the short-life food production has at most a life cycle of one year. These considerations bring us to suppose that most of traceability request will refer to lots whose production belongs to the current year, or to the previous one. According to such considerations we can define the most important weights of a cost function to minimize the memory used to serve the requests.

If we name $p_i$ the actual cost, in terms of memory space, necessary to store the replicated (data)file belonging to the node $i$, we have:

$$p_i = \alpha \cdot p_i \text{ last 4 years} + b \cdot p_i \text{ last 3 years} + c \cdot p_i \text{ last 2 years} + d \cdot p_{\text{last 1 year}}$$

with $a < b < c < d$ in order to give a greater weight to the most recent production. The relation $a < b < c < d$ is justified by the usual consumption of foods. This one, especially for fresh foods, is temporally closer to the production date.

In the same way, calling $r_i$ the cost in terms of queries received by the node $i$, we have :

$$r_i = a \cdot r_i \text{ last 4 years} + b \cdot r_i \text{ last 3 years} + c \cdot r_i \text{ last 2 years} + d \cdot r_{\text{last 1 year}}$$

with $a < b < c < d$ in order to give a greater weight to the queries concerning the most recent food production.

A generalized total cost of resources for a node $i$, served by a dedicated server, will be therefore the sum of $\alpha \cdot r_i + d_i$ with $\alpha > 1$ because the most critical resource is the ability of satisfying multiple queries.
4. Security

Each node in the chain is responsible for its own production. Each operator in the production system is in turn responsible for his production activity. When the lots on a node conclude the processing cycle, the documents relating to production on that node should not be modifiable, in order to guarantee their authenticity. A possible solution is to attach a digital signature at each step in the production process using digital certificates each node should possess. However, the following considerations should be made:

1. If a traceability code indirectly refers to several nodes in the chain, the document produced will be structured into as many parts as there are referenced lots.
2. A node will not guarantee the authenticity of traceability documents for other nodes representing different companies.

It follows that each node will have to authenticate its own information and send this information on request. The outcome of a generic request for a traceability code will therefore be a document made up of several separately authenticated parts. If it is assumed that all the documents involved have the structure described above, it will be possible to adopt the standard XML Digital Signature of the W3C consortium \[7\][8] to achieve an XML document authenticated on several nodes by different subjects.

The result of a digital signature applied to the external document held on a remote network resource would allow a control authority outside the production chain to verify any alterations made after generation of the document.

Another problem is security and authentication in communications between the nodes in the chain. There are two possible strategies to adopt. The first, which is the most widely used nowadays especially for bank transactions, is to make the channel between the two endpoints secure. Solutions such as SSL, VPN and HTTPS are used for this purpose. The second strategy, which is currently spreading along with web service, is to realise a system of exchanging secure messages at the application level using methodologies known and tested in the Secure Socket Layer. A solution could be implemented by integrating XML Encryption [9] in the web services exchanging reserved information.

5. Retrieving food processing data from the shop floor.

To achieve the traceability inside a single node of the food chain we need the identification of food items. This is based essentially upon two categories of identifier\[11\]:

- Primary identification (based on the use of biological markers and feature extraction based upon anatomical, physiological, biochemical or molecular, including DNA, methods of identification).
- Secondary or data carrier-based identification techniques in which a number or alphanumeric string is used for identification purposes and may be accompanied by other data or information for traceability or process support purposes.

A secondary identifier may also be linked to a primary identifier, particularly where the primary identifier is held as a data template in a data carrier or database. A list is presented in Fig.8.

Fig. 7. - Part of an XML document containing the Digital Signature of an external document.

Fig. 7 shows part of an XML document containing the result of a digital signature applied to the external document indicated in the Reference field. The possibility of referencing the digital signature of a document held on a remote network resource would allow a control authority outside the production chain to verify any alterations made after generation of the document.
levels. The technologies considered particularly relevant to the needs for food traceability include:

- Linear bar codes
- Two-dimensional codes (multi-row bar, matrix and composite codes), including direct marking
- Contact memory devices
- Radiofrequency identification (RFID)
- Smart labels (passive and active devices)

For open systems usage, which is invariably the case for food supply chains, it is essential that the identification codes and any additional information relating to the item, such as batch number, weight, volume, other identification numbers, use or sell-by-date, adhere to a particular identification standard.

Through standardization the hierarchy of packaging and item traceability can be better achieved. The EAN.UCC standard[12] provides the necessary coding structures for identification of items, and other entities such as location. In our approach we add another identification layer before the packaging to bind the data processing incoming from the shop floor to specials ID generated for each step of production. We adopted a simply codify strategy to create external links to the database with the processing info and other repository with data required for traceability.

As an example, we consider here a generic shop floor organized in various production lines and model each line as a tidy succession of working steps from the beginning to the end of the line. The two main issues in the realization of a traceability reports are:

- To associate to each job processed, the data flows of the production cells that have been employed in the working of the jobs.
- For each data flow, to differentiate the values of production related to the single job.

![Fig. 9 – Retrieving data from the shop floor.](image)

By modeling the production line into processing steps we have the possibility to record on a database the quantity of incoming and outgoing products from one step to another, for all the jobs that belong to a specific processing line. We will have therefore three types of codes or IDs for every processing step:

- Input codes have the following format:
  IN.LINEXX.STEPPYY[JOBIDZZZ] where XX represents the number of the production line and YY indicates the step of the processing sequence. The sub-code JOBIDZZZQ3 represents the code of the job which has been received from the previous step and is linked through the letter Q to the quantity of material which has been received;

- Processing code have the following format:
  PROC.LINEXX.STEPPYY[JOBIDZZZ] where XX represents the number of the production line, and YY is the processing step. JOBIDZZZQ3 represents the code of the jobs received from the previous step, linked with the letter Q followed by the amount received; the code of the processed job can be linked with another code by means of the character & in order to indicate the joint processing.

- Output codes from the processing step are in the format:
  OUT.LINEXX.STEPPYY[JOBIDZZZ] where XX represents the number of the production line, and YY is the step. JOBIDZZZQ3 is the code of the job received from the previous step again linked with the letter Q which is followed by the quantity used;

![Fig. 10 – Traceability IDCodes Generation flow for each processing step](image)

Moreover, the use of the OPC XML-DA standard [13] in order to access processed data allows the codes generated by the process, to record all processing data through the same data structure for various machinery when they are performed in the same processing step.

The enormous advantage in our case derives from the possibility to have one single data table for every
processing step of working with all operational data coming from various machinery.

The data structure obtained this way is homogenous for all the production line and simplifies the realization of the Report Builder that will behave like a standard interface in order to query the traceability database for data inherent the manufacturing activity.

![Image](84x645 to 98x663)

![Image](81x590 to 97x608)

![Image](81x619 to 100x636)

![Image](194x611 to 269x622)

![Image](198x602 to 203x608)

![Image](202x580 to 228x592)

![Image](220x580 to 222x582)

Fig. 11 – Storage of data food processing referred to a single job.

6. Some conclusions and final remarks

In this paper, we presented a general architecture of a distributed traceability system to be used in food manufacturing chains. The system allows a user to trace all steps during the production, distribution, and sale of food, this way gaining precious information about the quality of the food. The architecture has been discussed from several points of view and it has shown to be robust, flexible and featured by high security.

Other important aspects refer to the usability of the system with reference to the availability of a simple, user-friendly interface, and to the access time to the traceability information.

A User-friendly interface requires an effective organization of information to be displayed and a simple Web navigation strategy. These features can be provided into several ways and thanks to the wide availability of languages and tools, it is not a problem to design such an interface.

The access time of the traceability information represents another key point of the system performances. If we consider any real scenario, with one or more intersecting production chains, the system processing and storing traceability documents will grow continuously as regards both the volume of documentation and the extent of the chain, as new forms of production are inserted. So, according to the complexity of the traceability chain, a considerable time can be required to collect all the information spread into several nodes. For this reason it is important in the system design to provide suitable resources in order to reduce the access time to the information.

In the architecture described above, the nodes involved in retrieving traceability information can be queried for information they contain and for information produced and stored by other nodes. Therefore, the system’s response will also depend on the response times of the caches in the nodes involved in the query. These times will obviously, change according to whether the information is retrieved directly from the Web Service Cache or from remote company databases. The cache management policy appears to be a key point for the improvement of the performances. Each cache should know which requests are satisfied by adjacent and/or remote nodes and which are the main request flows.

For this reason, our future work will cover the study of dynamic strategies to minimise the access delay in the traceability systems.

7. References


[6] ISO 22519, TC 34/WG 9, Traceability system in the agriculture food chain - General principles for design and development.


