SAND 2011-4203 C

The Value of Utilizing Stochastic Mapping of Food Distribution Networks for Understanding Risks and Tracing Contaminant Pathways

Stephen H. Conrad, Walter E. Beyeler, and Theresa J. Brown

***Abstract*— Difficulties in adequately characterizing food supply chain topologies contribute perhaps the major uncertainty in conducting risk assessments in the food sector. The capability to trace contaminated foods forward (to consumers) and back (to providers) is needed for rapid recalls during food contamination events. The objective of this work is to develop an approach to risk mitigation that protects us from an attack on the food distribution system. We developed a general methodology for stochastic mapping of fresh produce supply chains and applied it to a single, relatively simple case - edible sprouts in one region. The sprout case study, selected using a system-scale risk assessment process, demonstrates how mapping the network topology and modeling the potential relationships allows us to determine the likely contaminant pathways and sources of contamination. The stochastic network representation improves our ability to explicitly incorporate uncertainties and identify vulnerabilities. The initial model runs show how: 1) Information on the relationships between entities in large food distribution supply chains aid in delineating network topologies and reducing uncertainty regarding potential contaminant pathways; 2) inventory management practices affect arrival times at the retail level and the overall residence times for contaminants in the supply chain; and, 3) the relative size of the contamination incident influences the dispersion of contamination across the retail end of the supply chain.**

I. Introduction

Quantifying the potential distribution of contaminants and providing faster, more accurate tracing of contamination sources requires a better understanding of the topology of food supply networks and the persistence of the relationships between network elements. Several factors contribute to the variability and uncertainty in food supply topology:

- Supply chain topology can vary markedly from one food marketing system and agricultural sector to another.
- Even within a single agricultural sector, some portions of the supply chain may be vertically-integrated and characterized by enduring supplier/customer relationships; while adjacent portions may be market-based and highly transitory.
- Customer/supplier relationships are sometimes considered to be proprietary information and can be closely held.
- Even among industry insiders, knowledge about supply chain topology can be relatively myopic. Many entities within the industry only know “one up and one down” – that is, they only know their direct supplier (one up) and their direct customer (one down).
- Spot-market relationships can be ephemeral, with suppliers changing from one day to the next.

Stochastic network representation provides the capability to incorporate and express the uncertainties using probability maps of the seed-to-consumer pathways. We call this process stochastic mapping. Stochastic mapping of food networks provides a repository of data and data sources that can be used to map topologies; providing the structural information that is an essential element for identifying vulnerabilities and designing risk mitigation strategies. In short, it provides capability that is needed for risk analysis and to design robust food defense strategies.

II. Methodology

We utilize a network-based, probabilistic modeling approach to develop the stochastic-mapping methodology for food supply chain risk analysis. The methodology was then applied to a case study for a food supply chain that is known to be vulnerable to contamination to identify data sources, illustrate how the methodology is implemented and evaluate the utility of stochastic mapping.

To develop a stochastic network model of a specific agricultural sector, we define the nodes, rules that describe the operational behavior of each node, rules that describe the interaction between nodes (links) and probabilities of interactions. Stochastic network maps are used to display and analyze the results of stochastic modeling.

The network characteristics were developed based on data from open, restricted access and trade sources, and through information elicited in interviews with industry experts [1]-[5]. Incomplete data represent a significant source of uncertainty in creating topologies of food agricultural sectors.

Manuscript received June 10, 2011. This work was supported in part by the U.S. Department of Homeland Security’s National Infrastructure Simulation and Analysis Project. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

S. H. Conrad, is with Sandia National Laboratories, Albuquerque, NM 87185 USA (phone: 505-844-5267; e-mail: shconra@sandia.gov).

W. E. Beyeler, is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: webeyel@sandia.gov).

T. J. Brown is with Sandia National Laboratories, Albuquerque, NM 87185 USA (e-mail: tjbrown@sandia.gov).

The uncertainty is explicitly incorporated in the simulation and analysis using probabilistic mathematical techniques. The probabilistic analysis provides an understanding of the plausibility and likelihood of specific food pathways.

A. *Agriculture-Food Supply Chain*

Agriculture sector supply chains that go from agricultural product production to the end food product consumer include multiple processes and markets. Vertically integrated supply chains, such as large grocery chains, have their own distribution system and contract directly with the farm, processor and/or a packer/shipper for specific products. Smaller retailers such as independent restaurants and grocery stores purchase products from a regional distributor and sometimes purchase un-processed (e.g., fresh produce) from local producers. Produce supply chains also include the seed growers, seed sellers on whom the produce growers depend. Similar entities exist in the non-produce agricultural sectors (e.g., cow and calf operations provide the “seed stock” for new herds raised by beef producers). The simplest supply chain is the one where food is grown or raised by the end consumer, eliminating many of the intermediate nodes between seed (or initial breeding stock seller) and food consumer.

There is no textbook describing the characteristics of food marketing networks. Luckily, abundant tacit knowledge about these systems exists within industry and academia. However, this knowledge tends to be more fragmentary and localized than comprehensive in nature. Each participant knows a lot about their company’s operations but typically they would be hard pressed if asked to describe the characteristics of the entire system. Understanding at the system level is required to understand the risks because any introduced pathogen or toxin will likely traverse many parts of the system before the consequences occur.

The first step in understanding the system is developing a comprehensive system description which requires collecting a broad set of information. Once the system has been described, it can be mapped in sufficient operational detail and a topological map can be constructed. Where data are changing too quickly, or are unavailable, there will be uncertainties. Stochastic modeling incorporates these uncertainties using standard probabilistic techniques. A probabilistic representation shows the possible pathways from producer through to consumer. Using the probability map produced by the model results, we should be able to see temporally and geospatially where: (1) products definitely went, (2) where they definitely did not go, (3) where they might have gone and the likelihood of each pathway, and (4) where additional data collection is warranted.

There are lots of data on produce production, processing and retail sales but little information about the flow of goods between businesses. Information about the movement of agricultural products will never be complete, even if historical information is perfect (which it never will be). Changes in the industry, such as changes in business practice or response to market dynamics, will change the pattern of flows in the future. The parts of the system involving markets are inherently stochastic. We use a model to study the possible movements of produce both in the current system and the possible flows under changed conditions. The model simulates the movement of contamination, introduced at any point in the system. This capability helps us identify locations or processes that might create unexpected vulnerabilities; simulate the effects of mitigation measures that might be used to prevent contaminant spread; and look for patterns in contaminant spread that could help pinpoint the source based on detection events downstream.

The model consists of “entities” that represent individual businesses of various kinds (farms, processors, distributors) and “markets” that organize the movement of goods from one entity to another. Markets can represent actual locations where products are exchanged but more often model a set of buyer/seller relationships that form over time (due to proximity, historical experience, and mutual business or social contacts). Products can move from any seller to any buyer in a market, depending on their rules for managing inventory and on their current circumstances.

B. *Sprouts Case Study*

The processor operations and general network characteristics for the sprouts supply chain are shown in Figure 1. This network is composed of multiple, connected networks. Alfalfa and Alfalfa seed are grown in many countries. Seed growers sell to large companies (seed brokers) that sell seeds nationally or internationally to the many sprout growers. Sprout growers vary in production capacity and the size of their markets. Sprouts are transported from the grower to the retailers by distribution companies. Large, vertically integrated chains use distribution companies for transportation, where small independent restaurants and grocers purchase from regional distributors.

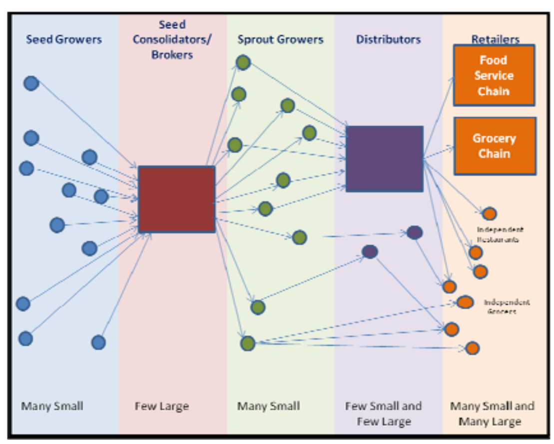

Fig. 1. Sprout Network Characteristics. Illustrating the supply chain entities and the order and potential types of relationships (e.g., one-to-one, one-to-many)

Information on business models for sprouts was obtained through interviews with food retailers, distributors and sprout growers; extrapolation of standard business models and information gleaned from past contamination events. It is assumed that all large grocery retail chains are vertically integrated due to the efficiencies gained in delivery time, inventory space, and transportation costs. Each class of business (e.g., producer, grocery chain, independent restaurant) has a different set of business models that may be applied. The network topology, that is the entities and connectivity of the entities, will be a function of the business models each entity utilizes and the logistical constraints created by geographic location and transportation infrastructure configuration.

Some of the key characteristics of the sprout network are that:

- ~ Seeds are grown domestically and internationally primarily by small producers
- ~ Most contamination is thought to occur in the seed growing step
- ~ Alfalfa seeds are used to grow animal fodder and sprouts for human consumption
- ~ There is no kill step for bacteria and the 3 main contaminants are E-coli, Salmonella and Listeria
- ~ Seed growers sell to seed companies who distribute sprout growers
- ~ Large sells to large - sprout growers sell to consolidators/distributors that sell to large food chains (restaurant chains and grocery chains)
- ~ Sprouts have a 7 day refrigerated shelf life
- ~ The market for sprouts is generally local to regional

The New Mexico sprout supply chain is shown in Figure 2 along with the results of a single run of the network model for a hypothetical contamination scenario. The network diagram provides useful information. The retail chains purchasing from "Sprouting Co. 1" are an isolated network starting from the seed company through the retail end of the chain. If a seed company or sprouting company has a contamination incident, the network diagram shows the possible pathways.

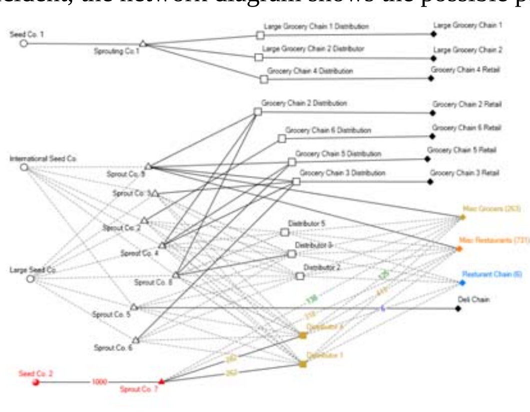

Fig. 2. New Mexico sprout supply chain with results for a single run with hypothetical contamination of 1000 units from a single Seed Company and the resulting distribution of contaminated units in the system (numbers of units sent to each location in the model) run). Solid lines indicate known relationships. Dashed lines indicate possible relationships.

The uncertainties are represented in the model by varying the rules and the parameters to evaluate their effect on the results. Uncertainties in the contamination incident, seed stock management policies and model parameters were evaluated. The size and origin of the contamination event influence the potential spatial distribution of contamination in the network. Seed stock management policies influence the rate of contaminant movement through the system. Figure 3 shows the results of 50 stochastic modeling runs for the random seed management process (seeds are drawn from stock randomly) and compares the distribution to that for a first-in-first-out (FIFO) management process. The same amount of contamination moves through the system during a much narrower time frame in the FIFO case verses the Random case.

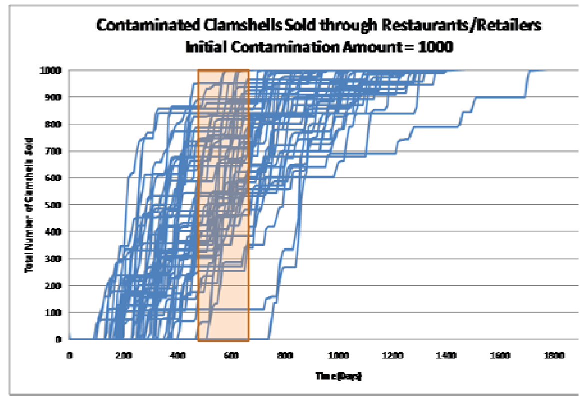

Fig. 3. Total Retail Sales of Contaminated Sprouts over Time for Random Seed Inventory Management Policy for 50 Runs (First-in-first-out management results fell within the range marked in orange).

Model simulations can also be used to assist backward tracing when a contamination event occurs. The problem can be framed as a kind of probabilistic inference, in which observation of contamination at one or more locations (or even an assigned probability that contamination occurred at a location) is used to infer probabilities that the contamination will also be found at other locations in the network. We illustrate this idea using the current set of results. Table 1 shows the set of retail outlets along the left column, and a sub-set of the sprout growers along the top row. Entries in the table show the fraction of simulations, with 1000 units initially contaminated and with random inventory management, in which contamination passed through the indicated grower given that it was found at (one or more outlets of) the indicated retailer. The "unconditional probability" shown in the table is the fraction of simulations in which the grower produced contaminated spouts, regardless of which retail outlets received them. It is the probability that would be assigned to the individual growers if we did not have any information on which retailers sold the contaminated sprouts. The red and green shading show examples where the stochastic mapping provides a significant change in the probability (more than double or less than half) and therefore refining the understanding of the potential contamination pathways. In the cases with the greatest dispersion of contamination (large contamination incident, random seed inventory management) the information providing the greatest value is the knowledge gained in delineating the network structure. The red and green areas are primarily related to delineation of the Arizona Sprouts portion of the supply chain.

TABLE I

STOCHASTIC MAPPING FOR NM SPROUT RETAILERS TO SPROUT GROWERS FOR RANDOM INVENTORY MANAGEMENT CASE WITH 1000 UNITS OF CONTAMINATED SEED

	Sprout Co 1	Sprout Co 7	Sprout Co 2	Sprout Co 4	Sprout Co 10	Sprout Co 8
Large Grocery Chain 1	1.00	0.00	0.00	0.00	0.00	0.00
Deli Chain	0.00	0.46	0.95	0.76	1.00	0.95
Sprout Co. 7 Customers	0.00	1.00	0.86	0.59	0.95	0.86
Grocery Chain 2 Retail	0.00	0.50	0.95	0.73	1.00	0.93
Misc. Grocers	0.00	0.51	0.93	0.70	0.98	0.91
Misc. Restaurants	0.00	0.51	0.93	0.70	0.98	0.91
Grocery Chain 3 Retail	0.00	0.47	0.95	0.68	1.00	0.92
Large Grocery Chain 2	1.00	0.00	0.00	0.00	0.00	0.00
Restaurant Chain	0.00	0.51	0.93	0.70	0.98	0.91
Grocery Chain 4 Retail	1.00	0.00	0.00	0.00	0.00	0.00
Distributor 5 Customers	0.00	0.50	0.95	0.71	1.00	0.93
Grocery Chain 6 Retail	0.00	0.48	1.00	0.73	1.00	0.93
Grocery Chain 5 Retail	0.00	0.50	0.95	0.71	1.00	0.93
Unconditional Probability	0.14	0.44	0.80	0.60	0.84	0.78

A subset of the full results showing the probability of each sprout grower having contamination given that contamination is found at the specified retailer. Shading indicates a large increase (red) or decrease (green) in probability, based on knowledge of the retailer

TABLE II

STOCHASTIC MAPPING FOR NM SPROUT RETAILERS TO A SUB-SET OF SPROUT GROWERS FOR FIFO INVENTORY MANAGEMENT CASE WITH 1000 UNITS OF CONTAMINATED SEED

	Sprout Co 1	Sprout Co 7	Sprout Co 2	Sprout Co 4	Sprout Co 10	Sprout Co 8
Large Grocery Chain 1	1.00	0.00	0.00	0.00	0.00	0.00
Deli Chain	0.00	0.18	0.27	0.18	0.09	0.09
Sprout Co. 7 Customers	0.00	1.00	0.00	0.00	0.20	0.00
Grocery Chain 2 Retail	0.00	0.07	0.13	0.40	0.53	0.20
Misc. Grocers	0.00	0.24	0.39	0.15	0.22	0.10
Misc. Restaurants	0.00	0.24	0.38	0.14	0.24	0.10
Grocery Chain 3 Retail	0.00	0.00	0.25	0.38	0.50	0.06
Large Grocery Chain 2	1.00	0.00	0.00	0.00	0.00	0.00
Restaurant Chain	0.00	0.24	0.38	0.14	0.24	0.10
Grocery Chain 4 Retail	1.00	0.00	0.00	0.00	0.00	0.00
Distributor 5 Customers	0.00	0.11	0.11	0.11	1.00	0.00
Grocery Chain 6 Retail	0.00	0.00	1.00	0.00	0.06	0.13
Grocery Chain 5 Retail	0.00	0.00	0.13	0.40	0.53	0.20
Unconditional Probability	0.16	0.20	0.32	0.12	0.20	0.08

A subset of the full results showing the probability of each sprout grower having contamination given that contamination is found at the specified retailer. Shading indicates a large increase (red) or decrease (green) in probability, based on knowledge of the retailer.

As the contamination event gets smaller and dispersion decreases, the value of information provided by the stochastic modeling increases. Table 2 shows the results for a large event with FIFO management.

In the cases with small contamination (10 units), the tables are almost entirely green or red.

III. SUMMARY

This project developed a general methodology for stochastic mapping of fresh produce supply chains and applied it to New Mexico edible sprouts. The sprout case study demonstrates how mapping the network topology and potential relationships allows us to determine the likely contaminant pathways (forward tracing) and sources of contamination (backward tracing). The initial model runs show the value of having some specific information on the network topologies in reducing uncertainty in the possible pathways. The way seed stock is managed (random selection or first-in-first out) has significant impact on the arrival times at the retail level, overall residence times for contaminants in the supply chain, and the tendency of the network to disperse contaminants. Some knowledge about the retail site from which a contaminated product was purchased can reduce the uncertainty about other possible locations with contaminated products, particularly if the network dispersion is low.

REFERENCES

[1] RedBook Marketing Database (updated continuously and available by subscription at: http://www.rbcs.com/rbm)

[2] US Census Bureau, Foreign Trade Statistics 2009 (available at: http://www.census.gov/foreign-trade/data)

[3] US Department of Agriculture (USDA), Agriculture Census, 2007 (available at: http://www.agcensus.usda.gov)

[4] USDA, Vegetables 2010 Summary National Agricultural Statistics Service (NASS) 2011 (available at: http://www.ers.usda.gov/Publications/VGS/#yearbook)

[5] U.S. Census Bureau, Foreign Trade data 2010 (available at: http://www.census.gov/foreign-trade/data).