

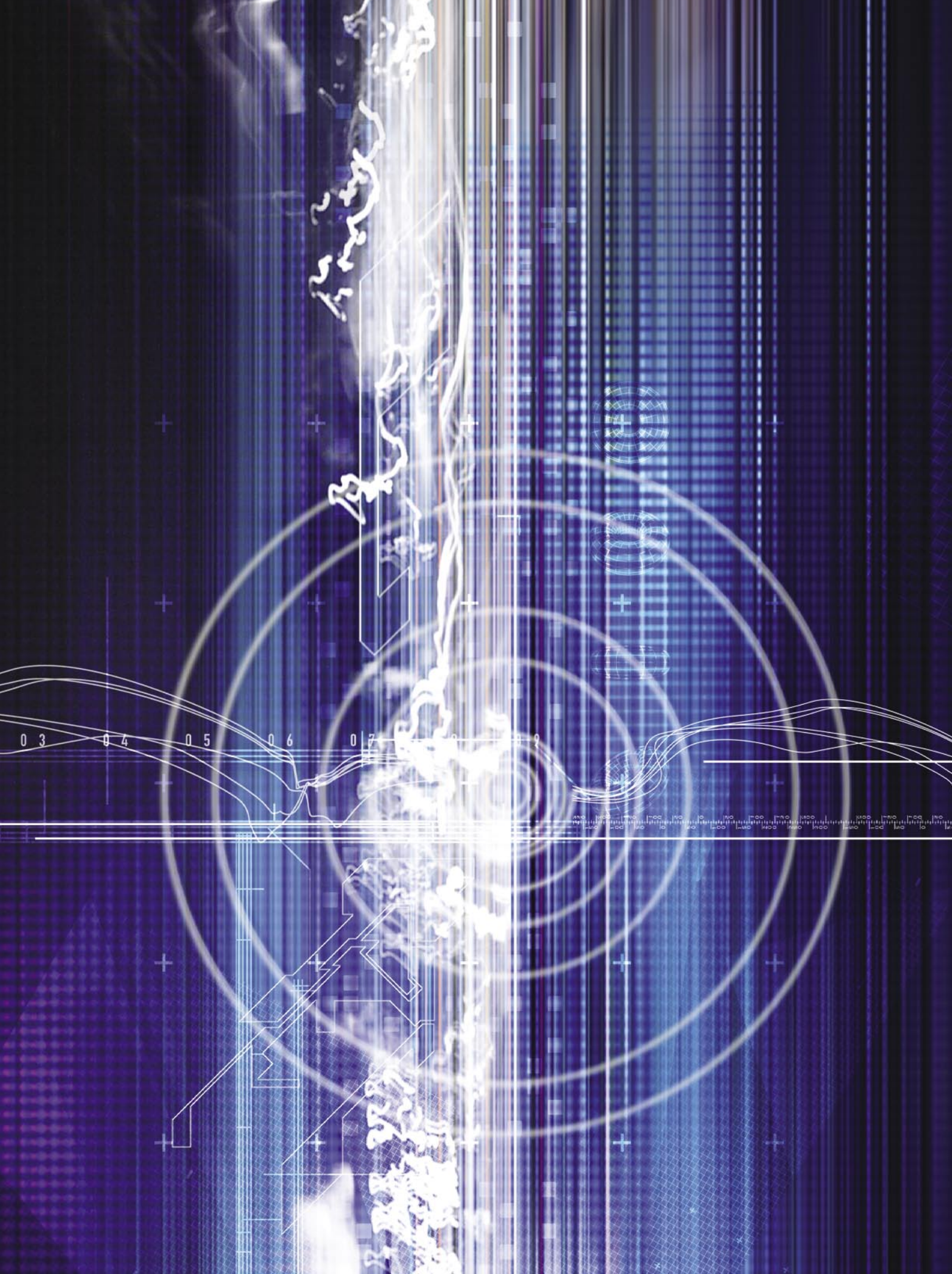
RFID (radio frequency identification) has received a great deal of attention in the commercial world over the past couple of years. The excitement stems from a confluence of events. First, through the efforts of the former Auto-ID Center and its sponsor companies, the prospects of low-cost RFID tags and a networked supply chain have come within reach of a number of companies. Second, several commercial companies and government bodies, such as Wal-Mart and Target in the United States, Tesco in Europe, and the U.S. Department of Defense, have announced RFID initiatives in response to technology improvements.

Early struggles with RFID have all involved hardware. Readers, tags, and even wiring and infrastructure are likely to be the first challenges early adopters will face. In fact, these constraints have already caused some of the early adopters to relax their timelines. Compared with these struggles, software seems secondary. In the haste of adopting RFID, the question that will often be asked is whether RFID readers are simply new-fangled replacements for bar-code scanners. In this article I present the view that RFID systems are fundamentally different from bar-code systems and that careful software and architecture design is necessary to achieve not only near-term performance, but also long-term return on investment.

DATA MANAGEMENT AND INVENTORY CONTROL  
ARE ABOUT TO GET A WHOLE LOT MORE INTERESTING.

# *Integrating* RFID

SANJAY SARMA, OATSYSTEMS AND MIT





# Integrating RFID



## A SHORT HISTORY

The Auto-ID Center was a research lab that I cofounded at MIT in the late 1990s. Over a five-year period, it grew to encompass five other labs around the world and drew sponsorship from more than 100 companies. Its basic mission was to make RFID tags cheap and ubiquitous. When we started our research, RFID tags cost upward of \$1 each. As a comparison, when bar codes were introduced in the early 1970s, they cost about 3 cents (in 1970s dollars).

We quickly determined that if RFID tags were ever going to have a shot at being widely used, a 5-cent price target was important for both psychological and commercial reasons. In return, though, the volumes would have to be very high—for example, more than 5 billion bar codes are scanned daily today. The problem with RFID tags at the time was that the industry was “stuck” in a higher-margin, lower-volume mind-set. At the Auto-ID Center, we set about flipping it to a high-volume, low-margin approach.

We proposed a two-pronged strategy. First, we recognized that the primary and most rigid component of tag cost is chip cost, which in turn is largely proportional to chip area. We therefore spent a great deal of effort minimizing the complexity of the state machine and the memory required on the chip. We did this, ironically, by developing simpler, lower-weight protocols that on the one hand reduced cost, while on the other hand were also applicable to a larger range of applications. For example, earlier versions of RFID tags used complex anti-collision techniques, supported large, complex memory structures, and included encryption.

We insisted on reducing the memory on the tag to a simple “license plate,” and we simplified the anti-collision technology to simpler tree-walking or Aloha-like variants. We eliminated encryption from the simplest tags because there was no memory to protect. In doing so, we had to address a number of issues ranging from digital signal processing to semiconductor manufacturing issues.

This minimalist approach permitted us to reduce the size of the chip, and since cost is roughly proportional to size, it permitted us to reduce cost. Of course, very small chips opened up new challenges in chip packaging, so we

also had to invent new ways to handle small chips.

The second part of our two-pronged strategy was to put much of the data and intelligence associated with tagged items, which had hitherto resided on the RFID tags themselves, on the network instead. We achieved this by proposing a new, unique numbering scheme called the EPC (Electronic Product Code). The EPC would act as a pointer to data on the network in much the same way as a license plate on a car can be used to refer to the traffic tickets associated with that car. We then developed an infrastructure for associating these EPC tags with databases across the network using a variant of the DNS (Domain Name System), which we called the ONS (Object Name System). The ONS can be used to find the authoritative owner of the original data associated with an EPC tag. Other infrastructure components include the EPCIS (EPC Information Service), which is being standardized using a Web Services architecture. It can be used to extract information about an EPC from either a trading partner or another EPC-related application or repository within the enterprise.

Recent history shows that standards are clearly an important determinant in the success of any networking activity, so standards were very much on our minds from the get-go at the Auto-ID Center. The center proposed standards for the RF air interface between readers and tags, for communicating with RFID readers, for the ONS, and for a predecessor to the EPCIS called Savant software. In 2003, a new not-for-profit entity called EPCglobal was created and is now carrying the RFID standardization effort forward as commercial deployments proliferate. The keep-it-simple approach of the Auto-ID Center and EPCglobal is designed to enable a small number of standards to address a wide range of applications. The hope is that minimalist, shared standards will enable economies of scale in infrastructure.

## IMPEDANCE MATCHING IN THE SMALL

RFID systems are often thought of merely as glorified bar-code systems. This is a dangerously limiting approach that could lock a user into a much smaller subset of the potential benefits of RFID. Comparing RFID systems with bar-code systems is useful for understanding how these

systems differ and what challenges they face.

The difficulties of effective RFID interfacing stem from the very features of RFID that, ironically, are also its advantages. The best analogy to describe the problem of connecting RFID systems to today's enterprise systems is a concept from electrical engineering called *impedance matching*, which refers to the balancing of the dynamic properties of connected components—for example, a speaker and a hi-fi amplifier. An impedance mismatch results in a distortion of the signal and poor sound. With RFID, and in general with sensing systems, I speculate that the connections to existing software infrastructure will result in a mismatch of capabilities and requirements.

#### CHALLENGE 1: NON-LINE-OF-SIGHT READING


The ability to read without line-of-sight is a principal advantage of RFID systems over bar-code systems. The fact that every bar-coded item needs to be handled to enable a successful read makes bar codes fundamentally manual. In the few cases where bar codes are scanned automatically, the system is very structured: all the boxes need to be rectangular, they all need to be aligned fairly accurately, the scanner needs a motion sensor to assist in locating the box, and so on. The supply chain rarely affords this much structure. The result is that scanning an item is cumbersome and expensive, and bar codes are therefore read infrequently in the supply chain. For example, the bar codes on an individual item, such as a pack of chewing gum, might be scanned only once in the lifecycle of the object: at checkout.

The one area in which bar-code systems are used extensively is in courier and specialized parcel delivery applications such as FedEx and UPS. This is more affordable because, unlike the traditional supply chain, parcel delivery involves a lot of manual handling, anyway, and the incremental cost of a bar-code scan with a hand-held reader is minimal. Furthermore, packages in these industries tend to be of standard shapes and sizes, with the bar codes at predictable locations, so scanning can even be automated. The standard supply chain, however, offers neither the homogeneity to permit automation, nor the incidental opportunity to perform manual scanning of bar codes.

RFID readers, on the other hand, can sense items even when their tags are hidden, or sometimes, within the bounds of physics, when the tagged item is hidden behind other tagged items. This enables automation. Unfortunately, the very “locational tolerance” that makes RFID tags easier to read also makes it difficult to understand whether a tag is in fact in the reader's prescribed

zone, or whether the read tag is simply passing by. Missed reads are also an unfortunate reality with RFID systems.

While reader performance is improving, cost pressures will dictate that RFID systems will always be used at the limits of performance. This means that very often, a tag that should have been read will go unread. Furthermore, problems of reader interference, multipath fading, or sometimes more exotic or transient effects will cause many reads to be missed. For all these reasons, RFID readers must be dealt with very differently from bar-code scanners. Industrial deployments in distribution centers and stores will eventually have hundreds of readers. Managing the readers, handling interference, scheduling their operation, filtering the data that the readers produce, and interpreting reader data are all functions that traditional bar-code interfaces are fundamentally not designed to handle.



**An impedance mismatch**  
results in a distortion of the  
signal and poor sound.

#### CHALLENGE 2: HANDLING SERIAL NUMBERS

Placing a serial number on a bar code requires either a very long symbol or a two-dimensional variant that is difficult to scan and fit into the available space. Furthermore, today's scanners can't read 2D bar codes. Given the other drawbacks of bar codes, the CPG (consumer packaged goods) industry has shown substantial interest in RFID for a few years as the next-generation replacement for bar codes.

The EPC, which is a serialized numbering scheme, is one reason RFID shows so much promise. Every pallet, case, and, eventually, item with an EPC tag will have a unique serial number. Serial number information is extremely powerful in understanding, diagnosing, and controlling the supply chain. Serial numbers can be used to track individual entities and provide much more detailed behavior of the supply chain than can nonserialized bar codes such as UPC (Universal Product Codes) and EAN (European Article Numbering), which are used today

# Integrating RFID



around the world. These bar codes cannot really be used to count unambiguously. A scanner that sees the same bar code twice will conclude that there are two items because there is no serial number to identify it as the same item. In contrast, an RFID reader that sees the same tag twice can read the serial number and conclude that there is only one item.

Serial numbers can also be used to diagnose problems such as food freshness/expiration. Today, all we can know about a supply chain is the inventory levels at various stages in the chain. We cannot tell how long an item has been in the supply chain without manually looking for lot numbers (which are not captured in the bar code.) The common assumption is that the supply chain is entirely a FIFO (first-in first-out) queue. It is not. Unfortunately, items may get shuffled, or may spend considerably more time in the supply chain than simple inventory numbers might indicate. Serial numbers would address this problem, permitting one to measure the sojourn time of an item in the supply chain accurately.

Finally, because it may be possible to fake an item's EPC, but not its history (because the history is maintained on a server), the tracking information associated with an EPC is a powerful tool against brand problems such as product loss, counterfeiting, and diversion.

**RFID systems have more of  
a sensor-network or  
monitoring-system flavor  
to them than do bar-code systems.**

Unfortunately, many software systems used in enterprise systems today are not designed to handle serial numbers at the resolution that RFID enables. For example, most ERP (enterprise resource planning) systems can deal with individual pallet numbers, but not with

individual numbers at the case level. Very rarely can ERP systems handle serial numbers at the item level, except in special circumstances such as defense applications or high-value applications. Even in these applications, the serial number is used merely for after-the-fact traceability for auditing, warranty, or quality assurance purposes. The ubiquity of the serial number, along with the places where RFID tags will be read, makes RFID fundamentally different from bar codes.

## CHALLENGE 3: REALTIME DATA VOLUMES

It is important to understand that one of the benefits of RFID over bar codes is, as described earlier, the ability to read automatically. This has pronounced implications for data volumes and arrival rates. For example, today, if a pallet arrives at a dock door, it will not be registered until an operator walks over and scans it. The number of operators available to perform the scanning is automatically a limit on the data rate that a bar-code system can generate. In RFID, however, the acceptance of the pallet might happen automatically because a reader at the dock constantly monitors its space for incoming shipments.

So, first, the data is asynchronous. Second, the reader at the dock door might continue to read its space and generate read data because the organization wants to monitor theft or other aberrant patterns. The consequence is that the data flow is likely to be continuous. Third, a central system might request certain readers to perform reads on demand to verify certain facts. For example, it might want a certain dock reader to look for items that might have fallen off the pallet while it was being loaded, because the end-customer has detected some shrinkage.

In all these ways, RFID systems have more of a sensor-network or monitoring-system flavor to them than do bar-code systems. The data rates, volumes, and variability of data are all different from systems designed for bar codes.

## THE SOLUTION: RFID INFRASTRUCTURE

Impedance-matching RFID systems to ERP systems is a challenging task. A graceful way to address this challenge is to introduce a layer between the readers and the application software. This has come to be known, for lack

of a better term, as *RFID middleware*. It needs two levels of functionality to be effective: a lower-level device and data management level and a higher-level interpretation level, as shown in figure 1.

#### DATA MANAGEMENT

I have already described the challenges of reading RFID tags. Of these, two—namely, intermittent and unreliable reads, and high-volume data—can be addressed with appropriate data management. The data management layer must provide a buffer in case of surges in read-rates—using a queue, for example—and it must provide some basic filtering functionality to remove repeated, useless reads and, often, fake reads. The objective is to report events that are useful for higher-level reasoning such as “tag read for the first time” or “tag disappeared from view.”

This can be achieved by setting simple time thresholds to ignore intermittent appearances and disappearances that are deemed to be aberrations from physics (like interference) rather than from real physical removal of objects. For example, you could tell the software to ignore tags that appear for only a second before disappearing (assuming that they are stray tags that are just passing by), or you could tell the software to record tags as missing only after they haven’t been seen for three seconds. These simple rules would reduce the rate of false-positive and false-negative reads. These thresholds must be set with care, however, because they are obviously preemptive in the sense that they cause data to be rejected from the system.

Some have argued that part of this functionality will eventually end up in readers themselves. This is not unlikely, although it is worth pointing out that some of this functionality is *intra-reader* reasoning, and some of it is *inter-reader* functionality. It is obviously easier to

reduce intra-reader functionality into readers; inter-reader functionality is more difficult to capture and standardize in the near term.

#### DEVICE MANAGEMENT

In most practical RFID implementations, readers must interact with other devices such as motion sensors, PLCs (programmable logic controllers), and human interfaces. Readers must also be scheduled to avoid interference with each other and with other communication devices. RFID devices operate in free ISM (industrial, scientific, medical) bands—13.56 megahertz, 915 megahertz, and 2.45 gigahertz, the last of which is used by wi-fi. Bandwidth is always at a premium.

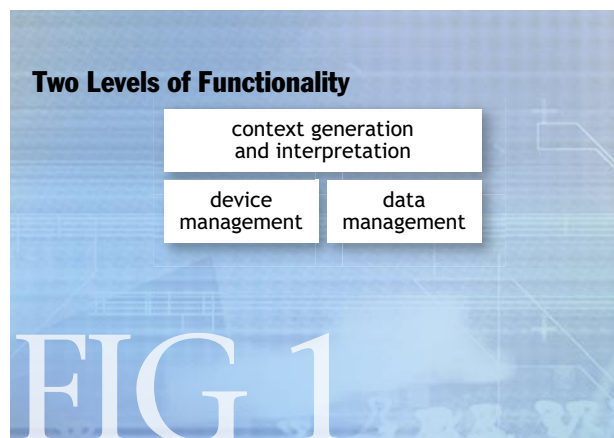
The situation is further complicated by differing standards around the world. For example, bandwidth and power allocations in Europe are very different from bandwidth and power allocations in the United States, and the way in which readers are scheduled will be very different between the United States and Europe.

Efficient device management is necessary to squeeze the maximum read rate out of larger RFID deployment. Device management is also necessary to monitor and maintain readers and other devices in an RFID deployment, to upgrade firmware, and to detect security intrusions. This is because, unlike bar-code scanners, which are operated by humans, RFID readers will operate autonomously. RFID will likely be one of the most extensive examples of ubiquitous computing in the coming years, and device management will be one of the major challenges. Device management is an example of inter-reader functionality, which will be difficult to reduce to within a reader in the short term.

#### DATA INTERPRETATION

Lower-level device management and data management yield coherent, clean RFID data. The next task, at the higher level, is to extract inferences that can be used by the applications the RFID system feeds. Today, these applications depend on a human operator to provide much of the context necessary to fulfill a task. For example, a WMS (warehouse management system) might instruct an operator to make a pallet with a certain manifest of cases and to hit Enter when done. This is done on faith—the operator is trusted to make a number of judgments, ranging from whether the right cases have been placed on the pallet to whether the pallet has been placed at the correct dock door.

RFID attempts to automate much of this functionality and must interface with the warehouse management





# Integrating RFID



system at the same level of sophistication that the human provides. Raw RFID data is fundamentally too low level to supplant the human. For example, to conclude from 97 reads generated by three different readers around a dock door that (a) they correspond to a single tagged pallet, (b) the pallet carries 96 additional tagged cases, and (c) this pallet has successfully exited that dock door and entered the right truck—this takes a lot more interpretation than simply noting 97 tags. A forklift that passes by the dock door carrying a different pallet might confuse a less sophisticated system, but a more able system can ignore the reads that are spurious to the business event at hand. Higher-level reasoning of this type can involve a number of inferences and associations. Tags can be associated with each other (when they are assembled); or they can be associated with a location or a business event—for example, a sales order.

The ability to make these associations also has an impact on the ROI of an RFID implementation. Consider the utility of the case-pallet association. As pallets travel at high velocity through dock doors, the reader at a dock

door might not pick up the pallet tag. Ordinarily, the business would have to invest capital in more readers or antennas or reduce the speed of the forklift (thus suffering a loss of throughput) to read the pallet tag more reliably. With a more sophisticated interpretation system that has access to the case-pallet association, however, the reader can read any of the 97 tags and infer that the pallet has passed through. This gives the reader 96 more opportunities to succeed. In other words, inference and contextualization permit the system to operate more robustly in the face of unreliable or lower-investment installations.

## IMPEDANCE MATCHING IN THE LARGE

Thus far, I have listed the challenges of integrating RFID into an existing enterprise. Now comes the challenge of actually extracting value for the enterprise at a systemic level. I will make the case that impedance-matching RFID with the enterprise in the large is actually a difficult task involving systemic rethinking.

In the management of any system, there is a spectrum of purposeful actions, which ranges from planning to

control. In almost every task that humans undertake, we create a long-term roadmap, which we refer to as a plan, and then we execute *to* the plan. This involves detecting and compensating for realtime disturbances through a process called control. So, for example, a pilot might create a flight plan, but the pilot, or the autopilot system, then controls the aircraft to follow that plan. Planning is usually done less frequently, using more long-term data. Control requires realtime data.

Lower animals, such as cats, have excellent control, but poor planning

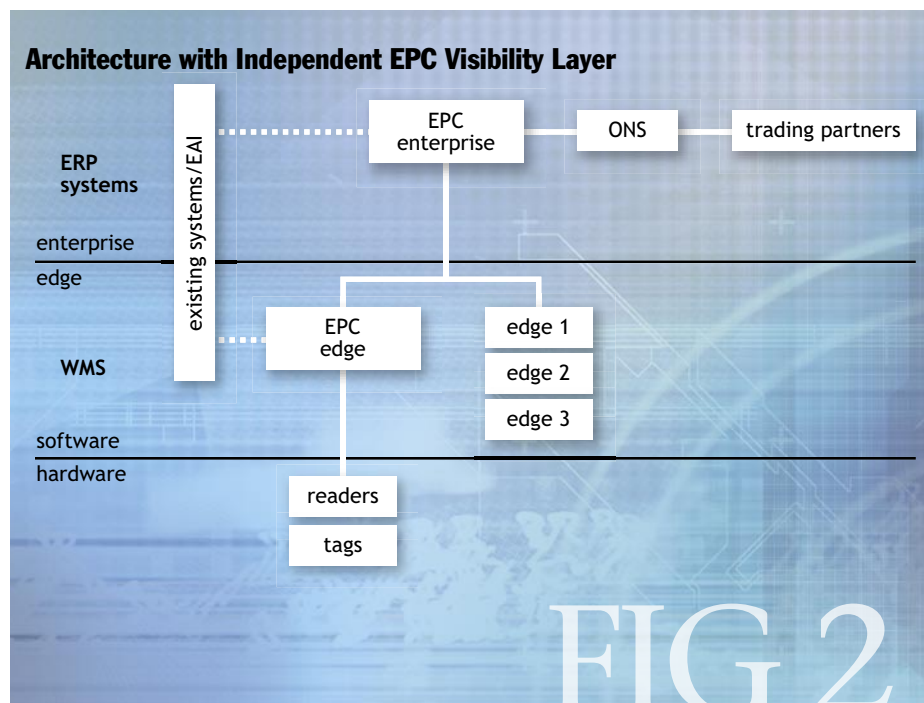


FIG 2

skills. The supply chain, oddly, is the opposite. In the supply chain, we perform good planning, but because of the lack of feedback, we have evolved fewer “reflexes” to actually absorb that feedback and perform control. This is likely to be the more difficult challenge in incorporating and using RFID data in the most efficient way.

Consider, for example, the problem of an incorrectly assembled pallet. Today, a warehouse management system might ask an operator to create that pallet. If a mistake has been made, the WMS essentially leaves it to the operator to detect and correct that error. With RFID, a verification tunnel could detect the error. The WMS has limited courses of action, however, to deal with this error—simply because the WMS is not used to receiving this form of feedback. In the short term, process workarounds might be possible. In the long term, all warehouse management systems will develop ways to deal with these exceptions automatically. In the middle term, however, as scale increases, there will likely be a gap between what the WMS can do and what the RFID data enables.

An unfortunate reaction to this impedance mismatch would be to throw the RFID data away. For example, you might be tempted to discard the serial-number information from RFID data simply because an ERP system is not designed to accept the serial numbers of the specific cases in a pallet. For the very short term, this might seem acceptable, but when you try to leverage RFID to perform recalls, this will seem like a shortsighted decision. The alternative to this approach is to build an independent EPC visibility layer that keeps RFID data in many levels of detail. This will permit many systems, both existing and new, to draw upon this data as new applications and functions come online. The architecture for such a system might look as shown in figure 2.

The Enterprise EPC repository in this figure would then be the single source of all EPC data. Instead of percolating, and therefore thinning, EPC data through existing systems, the enterprise system would keep a true and multiresolution record of EPC data across the enterprise, permitting different applications—old and new—to access EPC data at the appropriate resolution. This approach avoids the impedance-matching problem in the large, where the temptation would be to commit to an attenuated approach simply because it might seem expedient in the short term. Over time, new enterprise functions that can make full use of EPC data will emerge. Examples include track-and-trace for recalls, automatic shipping and receiving, counterfeit detection, and so on.

At the Auto-ID Center, we developed a software suite

called the Savant, which served as the edge and enterprise software. We also built a prototype of the ONS. Today, EPCglobal operates the ONS. EPCglobal also sells EPC codes to users who want to place EPC tags on their products. Furthermore, EPCglobal runs a series of standards activities for both the hardware and software modules of the EPC system. The EPCglobal ecosystem includes a number of emerging standards for communicating with readers, for middleware at the edge, and for edge and enterprise EPCIS systems. A number of vendors today sell software that performs these functions, and there have been hundreds of implementations along the lines of this architecture around the world.

#### LONG-TERM OPPORTUNITY

The EPC will create a new wave of supply-chain thinking in which RFID data will drive the supply chain. The emergence of this new sixth sense will challenge the way the supply chain is operated today. It will be natural and understandable to try and view RFID in a minimal, incremental way, either as if it were a new bar code or as if the extra information carried by the EPC were unnecessary. This approach might serve the short-term need, and it may even provide short-term value, but it will preclude a much more exciting long-term opportunity with RFID. The main challenge facing software professionals will be impedance matching in this gap. The approach I have described uses the common solution to all impedance-matching problems—namely, the use of a buffer. RFID infrastructure will permit RFID users to have a system that provides incremental value today, but can provide revolutionary value in the future. Q

#### LOVE IT, HATE IT? LET US KNOW

[feedback@acmqueue.com](mailto:feedback@acmqueue.com) or [www.acmqueue.com/forums](http://www.acmqueue.com/forums)

**SANJAY SARMA** is an associate professor of mechanical engineering at MIT. He was a cofounder of the Auto-ID Center, and until recently, chairman of research. Sarma received his B.A. from the Indian Institute of Technology, his M.A. from Carnegie Mellon University and his Ph.D. from the University of California at Berkeley. In between degrees, Sarma worked at Schlumberger Oilfield Services in Aberdeen, U.K., and at the Lawrence Berkeley Laboratories in Berkeley, California. Sarma is an executive board member of OAT Systems, a leading company in RFID middleware. He has authored over 50 academic papers in computational geometry, RFID, automation, and CAD, and is the recipient of numerous awards for teaching and research.

© 2004 ACM 1542-7730/04/1000 \$5.00