THE GHOSTS IN CAN TECHNOLOGY FIND

Every age gets the war its technology gives it. Mass and power defined industrial warfare, brought to bear literally to crush an enemy's people and machines. Its ultimate weapon was the atom bomb, matter transformed into annihilating energy.

Information Age conflict is different, focused on symbols, fears, disruption. War is fought by manipulating rivers of information moving at light speed, and by extracting crucial knowledge about the enemy from ever-spreading oceans of data. Mere matter is secondary. "It's not going to be a cruise missile or a bomber that will be the determining factor," Defense Secretary Donald Rumsfeld said over and over in the days following September 11. "It's going to be a scrap of information."

Make that multiple scraps. Hundreds of scraps. Thousands of scraps. Millions of scraps of information. Somewhere in which lurk the bits of data that, plotted together, will tease out the shapes of ghostly enemies, and bring them into sharp enough focus for a Tomahawk missile, a squad of Army Rangers or an FBI man with his gun drawn to finish the job.

Rumsfeld is right. Lethal weaponry still has its place—if a waning one. But when a commercial airliner is a missile, when the front lines are everywhere and pixels beamed by satellite to a billion TVs around the world are the measure of success, the landscape of battle irretrievably changes. High-speed networks, fast-access databases, biometric identification devices, mind-boggling processing power, neural network pattern recognition systems, analytical programs based on machine intelligence—those become the weapons of choice, and of necessity. Those and the most intangible resource of all, human creativity and imagination.

Which of the 1.5 million people who cross U.S. borders each day is the courier for the smuggling operation that sends dope money back to the Taliban? Which tiny shred

Russ Mitchell can be reached at vortumnuf@yahoo.com

THE MACHINE TERRORISTS?

BY RUSS MITCHELL

of the world's $1.5 trillion in daily foreign exchange transactions is the payment from an al-Quaida cell for a loose Russian nuke? Who happens to have a degree in microbiology, a purchase order for a high-end milling machine and a sudden interest in the writings of Ted Kaczynski? Which of the hundreds of thousands of tips flowing into the FBI are a matter of life or death?

Finding answers to those kinds of questions would tax the finest information technology professionals and the best of our computer hardware and software systems.

Too bad Uncle Sam is so far from being ready for the task.

BUNGLING THE BITS

Would you be surprised to learn that some of the FBI's key computer systems were outmoded a decade ago? Or that the CIA has trouble putting together all the intelligence gathered by its foreign agents? Or that the super-secret National Security Agency lacks the management discipline to effectively process the information spit out by its billion-dollar eavesdropping equipment and its breathtaking arrays of supercomputers? Or that the Immigration and Naturalization Service has completely mangled a border-control alien identification system that was supposed to be up and running long before Sept. 11?

Would it surprise you to know that only now are politicians discussing a highly secure federal government data network called Govnet that would connect the various agencies? Or that State Department personnel, using those old computer screens with the green letters, only within the past 24 months were given e-mail addresses? Or that the government's national law enforcement communication system, used by local police, is based on teletype?

Okay, you're not surprised. But you might be shocked to learn just how bad the situation really is.

Years of congressional monkey spanking, bureaucratic arteriosclerosis, federal-employee-job-for-life-guarantees, management

ineptitude and general systemic neglect—plus some old-fashioned stupidity—have conspired to leave the nation's law enforcement and intelligence agencies in sorry shape indeed. Saddled with ancient computer equipment and arthritic ideas about how to use it, the guardians of our peace are as ill-equipped for the 21st century's information battlefield as horse cavalry was for the trenches of World War I. But their success or failure will determine the freedom and prosperity of Americans, their friends and allies for generations to come. No hyperbole intended.

Armies can't hide in a world of satellites. Tanks have heat signatures. A squad of guerillas leaves tracks on jungle trails. The enemies of today's peace also leave electronic footprints. Trouble is, so does everybody else.

Today's raw data is tomorrow's valuable knowledge, which—combined with the plunging cost of computer networks and memory—is why the Information Age hangs on to everything. Anyone with a stake in the modern world –including those aiming to destroy it—inevitably spreads tracks across countless databases and hard drives. It's the cost of admission to the 21st century. That cost is what Sun Microsystems Chairman Scott McNealy had in mind when he made his famous statement about privacy in 1999: "You have zero privacy anyway. Get over it."

We'll leave the debate on privacy to other forums. The fact is that embracing technology makes everyone's lives more transparent: the good, the bad, and the ugly.

Driver's license records, credit card receipts, telephone logs, Internet usage, airline trips, car rentals—all are digitally recorded and stored on electronic databases. Surveillance cameras snap photos of license plates. EZ-Pass records travel patterns across urban bridges and toll roads. TV set-top boxes ship viewing habits back to marketers. Cell phones offer up triangulation information to pinpoint a caller's whereabouts. Moms, dads, junkies and terrorists put their mugs on camera when they use the ATM machine.

As anyone in the data-storage business will be happy to tell you, the amount of information being spewed is dizzying. Last year, more than 610 billion e-mail messages were delivered, a University of California at Berkeley survey shows. In 2000, 2.1 billion static pages graced the Web; it'll double this year. All the information created around the world—e-mail, snail mail, the Web, books, movies, TV, photographs, databases – last year totaled two exabytes, according to the Berkeley study. How much is an exabyte? It's 10^{18} bytes. Not concrete enough for you? If every word ever *spoken* by every human being on the planet throughout recorded history were added together, it would total five exabytes of information—or so says computer storage maker EMC. This year we'll do that easily.

And we're storing most of it. One reason is that the price is right. The cost of computing power is cut in half every 12 or 18 months, a phenomenon recognized as Moore's law. The cost of data storage is plummeting at a similar rate. In 1992, the cost of storage per gigabyte was $1,000. This year, it's about ten bucks. By 2010, you'll pay a quarter and get change back. The costs are shrinking because storage densities are shrinking. And when the physical limits of magnetic storage media are reached, holographic storage will take over. Within a few years, holographic storage will emerge from research labs and put a terabyte of 3D information on a DVD-sized disc – 200 full-length movies. Through an emerging fiber optic technology called wavelength division multiplexing that increases bandwidth by sending information down different colors of light, bandwidth down a single fiber will soon reach speeds of 5 terabits—just short of one of those movie-packed DVDs—*per second*. Make that a stream of, say, video conference calls between the United States and the Middle East, and you get an idea of what would-be watchers are up against.

So data warriors have their work cut out—too bad about those tools. Here's the situation today at the FBI: standard issue for many agents is a 486-generation desktop PC. The 486 Intel processor hit the market in 1989; it was already obsolete in 1993, when the Pentium hit the market. A G-men's 486 pokes along at 66 megahertz; low-end Dell computers sold for use by children at home clock in at 1.6 gigahertz. For our more politically inclined readers, that's 25 times faster.

How about networking? Managers at big companies use computers attached to T1 or T3 lines for high speed connections to the outside world. Even many small businesses are equipped with cable modems or DSL lines. Much of the FBI limps along on the same poky 56K dial-up modem you have at home.

The FBI manages huge databases. Many are based on old mainframe technology, some still on token ring networks, a slow-motion 1980s network architecture that was superseded long ago by Ethernet. The FBI's 17 or so major databases are "stove-piped"—islands unto themselves, with no robust connection to any other data banks. Whether they store criminal records or fingerprints or descriptions of stolen jewelry, they can't share their data easily with other FBI databases, much less with databases at the INS or the CIA. Not what you want when the name of the game is coaxing out patterns from across the datasphere.

And accessing that data? Agents use terminal emulators. That's an old-timey way to get a PC to act like the 1960s-style dumb terminals that used to be commonly attached to mainframes. To search for data, agents must deal with one database at a time: first you log in, find what you want, then you log out. Then you log in somewhere else, look for whatever you want there, then you log out. Not ideal, even for locating information you're pretty confident will be there. Prowling through data on a hunch? Most of the time, it's not worth the effort.

Agents aren't the only ones who are frustrated. At his confirmation hearing last summer, FBI Director Robert Mueller told senators: "I would like to be able to review…critical classes of cases, by turning on (my) computer and using the mouse to click on a series of cases to see what has been done the last three days, what you expect to be done in the next 30 days." Great idea! We'll get back to you.

"Yesterday's Technology Tomorrow"—the phrase has been thrown around the FBI for decades. It might be funny if it weren't so true. "The agents have encrypted radio; they've got night vision; they've got the surveillance stuff," says a former agent, one of several interviewed for this piece. "They have top of the line everything—everything except information technology." And now they have a thousand suspects and potential

witnesses for the September 11 disaster in detention, and 600,000 "leads" from worried citizens (that's as of late October) to try to sort through and correlate. Oops.

Since its formation in 1908 as a special investigative squad for the Department of Justice, the FBI's job has been solving crimes that already happened. Even its fight against organized crime depended on building evidence of committed deeds to nail Mafia bosses. Now the FBI is expected to anticipate crimes and stop them before they happen. The home front is also the front lines, which makes police work suddenly a lot more like what spies and soldiers traditionally do. At one time, modern computer systems would have helped the FBI do its job better. Without them now, the FBI may not be able to do its job very well at all.

It's not just the FBI. The INS, the State Department, Treasury, even the Central Intelligence Agency and the National Security Agency are disturbingly ill-equipped to process, analyze, and communicate the data they already possess, let alone the data storm that's already hitting.

Granted, when everything is clicking just right, the NSA can intercept a satellite communication that a CIA analyst can interpret as a bomb threat that a team of FBI agents can defuse. But at 7:30 a.m. at Logan Airport on a fine Tuesday morning in September, well.....

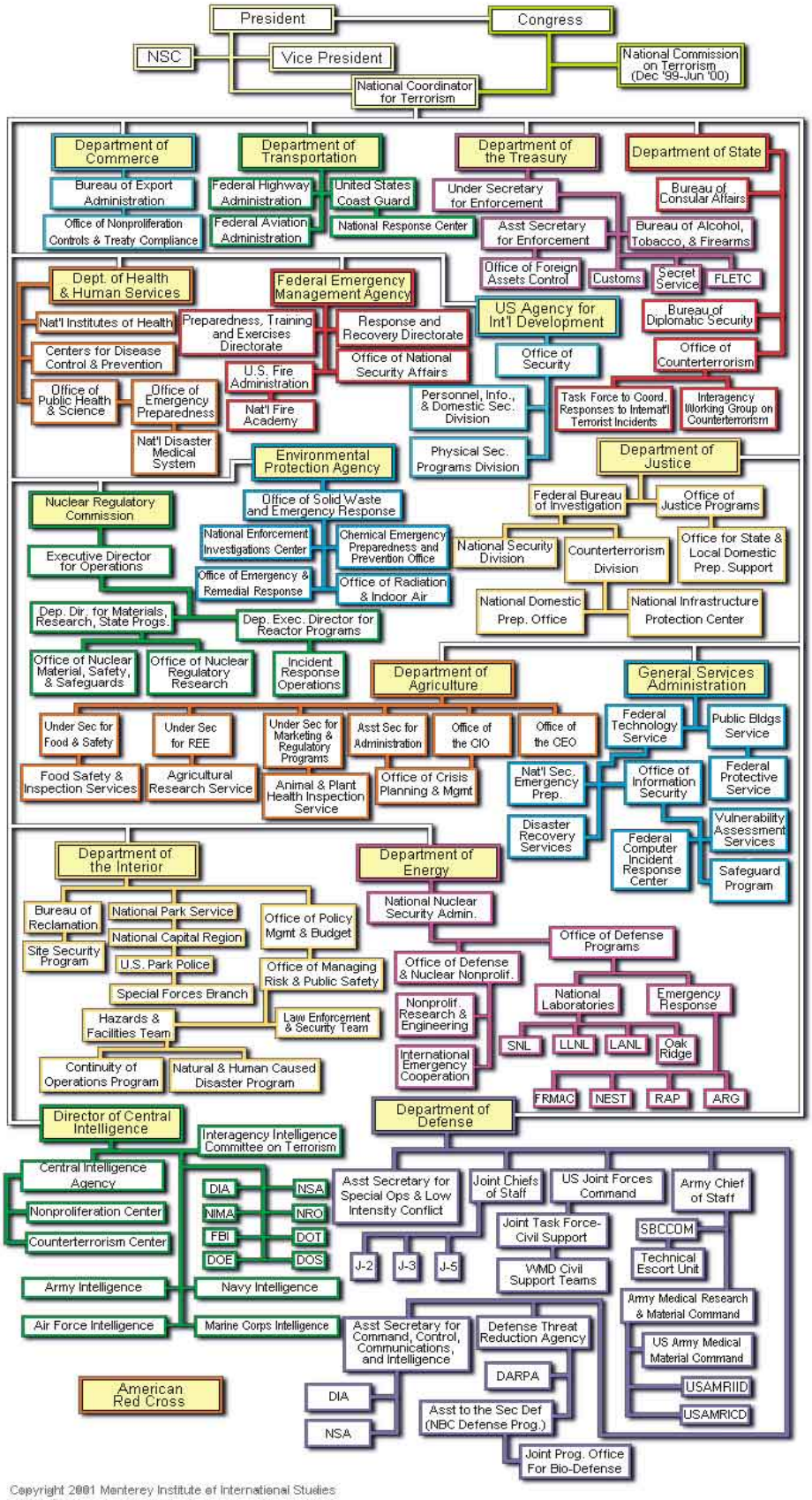
I-WAR, G-MEN

Not to dwell too much on Sun's Scott McNealy, but he did popularize the phrase "The network is the computer." This idea, which the Fortune 500 started grasping five years ago, is as anti-stovepipe as you can get. It sees individual computers as communication devices, whose ability to link up with the rest of the world's computers is more important than the processing power they possess on their own. It sees the network as a platform on which knowledge can be shared, amplified and re-created in new innovative forms, the networked pieces adding up to far more than the sum of their parts.

Victory in the information war depends on the ability to use that information—to understand it, to react to it quickly. "September 11 was all based on controlling information," says Carver Mead, the CalTech physicist, one of the fathers of the microprocessor and a pioneer in neural networking, an approach to artificial intelligence based on replicating the connective miracles

WHERE'S WALDO?

U.S. Government Anti-Terrorism Agencies, Sept. 11, 2001



of the human brain. (*Spectator* Interview, September/October 2001). "The fact is the hijackers controlled all the information. They had *all* the information on the first three planes—they were the only ones who knew what was going to happen. On the fourth plane, there was a tiny amount of information available to the passengers; they were able to use it to thwart the plan.

"We actually *knew* a lot about some of these people," says Mead. "We actually did know a lot, but we didn't put it together."

Putting it together requires acceptance of the idea that the network is the computer–and that the bigger and broader it is, the better. In many ways law enforcement agencies are an ideal starting place for intelligent networking. Take the FBI: attached to 56 field offices in the U.S. and 44 international outposts, agents and analysts serve as information-rich network nodes. They possess the facts and data that constitute the collective knowledge of the FBI. Add the spook agencies, and local law enforcement, and a potentially huge collective intelligence—if anyone can figure out how to collect and actually tap it.

In any robust network, of course, effectively channeling data from node to node is key. This is where the FBI's information systems start breaking down: slow transfer rates, unreliable connections, incompatible databases, ad hoc network architectures, inadequate search tools—all those conspire against would-be cybersleuths. If the nodes can't communicate effectively, the result is hardly a network at all, but an archipelago. The agents, the analysts, the databases – all are isolated digital islands.

How did things get this way? Back in the 1980s, the FBI began moving beyond index card files by combining its paper records unit with its main computer unit, the Information Resources Division, or IRD. An IT hotshot named William A. Bayse was brought over from the Pentagon to run it–a brilliant innovator, by all accounts, but perhaps the wrong man for the job. "Artificial intelligence was a big thing for him," a former agent says. "He was like a French chef. He could make all these incredible sauces, but couldn't put meat and potatoes on the table. The real world stuff was not getting done. The other divisions weren't happy. They wanted IT that worked."

Perhaps Bayse should have been put in charge of research. Artificial-intelligence programming was in its infancy, and the clunky mainframe computers of the time weren't powerful enough anyway to do much with it. The experience gave AI a bad name around the FBI, and today the agency is way behind the curve on the subject—just when the technology has matured enough to really do law enforcement some good.

So the IRD limped along. The equipment was bad, but management and technical support were worse. The FBI promotes from within. Most of its IT managers and support technicians are home grown, moving to technical jobs from positions as agents and analysts, or anywhere else they could find someone looking for a better career path. "A lot of these people don't have the aptitude for this," says a former agent. "They'd take a steno who could do WordPerfect, who they were paying as a Grade 5, ask if they'd like to become a support technician and promote them to Grade 7. They'd say 'Oh, *yeah*!'" Of course, if it didn't work out, there wasn't much anybody could do about it; this is the public sector. By steering clear of fireable offenses, even a clueless techie has the job for life. It's a problem the agency has only recently begun to address. Says Clinton Van Zandt, another retired agent: "The FBI has realized kicking and screaming that you need people from the outside."

The situation got so bad at IRD that two of the agency's "big dog" divisions—Criminal Investigative and National Security, which conducts espionage—cut IRD out of the loop and began running their own IT. By the mid-90s, when client-server architecture and high-speed networking were standard operating procedure for big businesses, and when it had become clear the Web and standard protocols were radically changing the way information is structured and communicated, it finally dawned on the FBI that it needed an IT overhaul.

Stand-alone databases are the number-one nightmare. The biggest is the National Crime Information Center—it's the database traffic cops use to find out if you're wanted in another jurisdiction. FBI agents and analysts also make heavy use of the Automated Case System, which stores internal reports known as 302s, memos, and electronic messages. The database is searchable by case number and by text string—"U-Haul truck" or "anthrax at NBC." The limits of this kind of search without context—it's the same problem faced by Internet search engines—were made clear by the detective played by Al Pacino in the movie *Heat*: "Run 'Slick' as an alias to the FBI. You're going to get the phone book. Do it anyway."

The problem sparked some forward thinking: how about at least giving FBI agents and analysts access to multiple databases through a common browser. And while they were at it, the systems planners figured, why not add up-to-date desktops and laptops, Internet-access in the field offices and multimedia databases that allow storage of audio and images, with scanners and digital cameras to go with that.

Another idea was high-speed connections, not just to speed up communications but also to make official networks more secure: more bandwidth allows stronger encryption of sensitive data. Former FBI director Louis Freeh led the mid-1990s battle against civil libertarians over software that allows private citizens and businesses – crooks and terrorists included – to wrap electronic messages in secret code. Legalities aside, that technological horse was already long out of the barn; Freeh lost. But meanwhile the FBI's own encryption-laden internal communications were crawling through tiny digital pipes, slowing agents down and making them less productive.

The result of all this was the Information Sharing Initiative, or ISI. (Sorry for the acronyms—this is Washington.) Serious planning began in 1997, and Congress appropriated a large chunk of funding to get the $430 million project rolling. The contractors were ready to go.

But suddenly Appropriations subcommittee staffers began asking for more details. Months went by, then years. Conflicting versions explain the delay. Suffice to say that the prime contractors competing on the project—Raytheon, Lockheed Martin, secretive San Diego-based SAIC—got tired of being on hold, and the FBI killed the project in 1999. "The system would have been in place well before Sept. 11," says a former agent. "Would ISI have prevented the attack? Of

course, it's impossible to say. But it's certainly clear they would have had a whole lot stronger set of analytical tools."

The saga doesn't end there. ISI was soon born again, as eFBI—an unfortunate name that was changed to Trilogy when the dot-com stock boom busted. IBM veteran Bob Dies was brought in last year to run it. Congress not only appropriated funds, but began releasing them—$100 million this year—and the FBI says Trilogy is on schedule. If it stays that way, a fit-for-the-1990s computer system will be in place at the agency by the end of 2003.

That's just the FBI. Its sister agencies aren't much better. In the case of INS—guardian of border-crossings, and arguably the real front line force in the battle against terrorists—it's even worse. Despite endless pledges to Congress, the White House and the public, federal agencies don't cooperate. It's not just a matter of one agency telling another, "Screw you—this is our case." There are different cultures, different responsibilities, different motivations, different incentives. It's the nature of organizations.

But wouldn't it make sense, for example, for Uncle Sam to have a single, common fingerprint database to catch criminals, terrorists, and deport-worthy immigrants? The FBI keeps fingerprint records of criminals—ten prints each, one from each finger. The INS keeps fingerprint records of deported aliens—two prints each, from the index finger of each hand. The systems are not integrated, and they are not compatible. Under pressure, the FBI and the INS began talking about linking their systems when the immigration service's Ident database was introduced in 1994. They're still talking. On the current schedule, the two systems won't become integrated until 2006—unless somebody prods them along a bit quicker. Tom Ridge, white courtesy phone!

Of course, the FBI may have good reason for dragging its feet. A lot of INS agents don't understand how even their own system works. Whether it's a system design problem, a training problem, a personnel problem, or a combination of the three, Ident's track record has been poor. The most egregious example involves a man with two names: Angel Maturino Resendez and Rafael Resendez-Ramirez. Now in custody, he's a suspected serial killer who was listed on the FBI's 10 Most Wanted List in 1998 and 1999, when he crossed the border at least seven times in the course of allegedly killing nine people. The INS maintains a "lookout list" as part of Ident. The suspect was a wanted man when the INS deported him at least once; no one had put him on the lookout list. Four of the victims were killed after the deportation. The General Accounting Office noted drily, "The INS has failed to effectively train its employees on Ident." For good measure, the system archives the records of anyone who has not encountered the INS for 15 months or more, criminals and criminal suspects included. A normal Ident search routine won't find them.

A grander INS project that could make the borders a lot more secure is crawling along at the usual Washington pace. The biometric Border Crossing Control Card (innovatively, bBCC) would apply fingerprints, face-recognition, iris scans, retinal scans and other body-based input-output data to border control. It would also issue immigrants an ID card containing biometric data. The project was mandated in 1996, along with another large-scale system to track foreign students in the U.S. Both were scheduled to be up and running by now; they're still undergoing tests—behind-the-curve technology, guaranteed. This is not a joke: Washington's Industrial Age requisition-budgeting-bidding-approval process, which can take years, may be great for generating congressional committee work. And yes, oversight of public funds is important, for a whole lot of reasons. But the system we've inherited is clearly inadequate to the real-time demands of running something as complex as a nation (let alone, the United States) in an information age. Unless this gets fixed, our guardians will always be stuck with yesterday's technology, while Osama bin Laden remains free to order himself the latest new satellite phone.

At least the super-secret spook outfits are stocked with spectacular technology and the expertise to use it, right? The secrecy that surrounds the CIA and the NSA shroud them in a kind of mythologic mystery—right down to those famous "black" budgets—that makes it easy to imagine a lot more power than they actually possess. Secrecy works two ways: it can cover up tools and weapons you don't want anyone to know about. It can also hide ineptitude. Think about the Wizard of Oz.

Former FBI agents are willing to talk to the press. It's significantly more difficult to find talkative sources at the NSA. Secrecy laws get in the way, not to mention the spook culture. But Congress has begun exposing the NSA and CIA to greater scrutiny. The failure to anticipate Sept. 11—or even react to it quickly enough to protect the Pentagon, hit more than 45 minutes after the first WTC attack—will only turn up the heat.

The House Permanent Select Committee on Intelligence last year ripped into the insular culture of the Cold War-nurtured NSA for its failure to keep pace with commercial technology. The agency boasts the world's most sophisticated listening devices and fastest supercomputers. How fast? Details—or anything else—about NSA equipment are classified, but it's widely known in the supercomputer industry that when NSA buys, it buys the best. The fastest machine commercially available now is the IBM ASCI White, which whips along at about 7 teraflops per second. In English? Let's say it takes you one second to do a single operation on your hand calculator. If you spent 24 hours a day, 365 days a year, doing nothing but punching out single calculations at a steady rate of speed, it would take you 221,963 years to do what ASCI White can do in less than a second.

Lovely stuff, but when it comes to analyzing, prioritizing, and communicating actual information—all the streams of stuff from all those satellites and listening devices and (presumably) the world's ever-thickening net of data pipes—the agency falls short. "As the global network has become more integrated," the intelligence committee concluded, "NSA's culture has evolved so that it is seemingly incapable of responding in an integrated fashion."

Zoom in a bit (which is all you're allowed) and the picture is not edifying: a not-invented-here culture on steroids. NSA, the House committee reported, "must take a hard look at the extent to which a relatively small number of government engineers, however talented, can be expected to keep pace with the commercial industry." Referring to system crashes—including one that took down key computer operations three days straight—the committee noted that this was "not the result of terrorist attacks or hacker gamesmanship," but mismanagement of outdated IT systems. Maybe the CIA is better? Here's the intelligence committee—these are the agency's *friends*, remember—assessing CIA information capabilities, in a post mortem to the Sept. 11 debacle. "Thousands of pieces of data are

never analyzed after the fact…" Potentially crucial information—that's a key thing about information: you can't know what's important until you actually look at it—may "sit for months, sometimes years."

THINKING MACHINES

Not surprisingly, some well-informed critics suggest just starting over. Oracle chairman Larry Ellison sparked controversy in September with a call for a single, centralized national database that would "ensure that all the information in myriad government databases was integrated." He said Oracle—whose first customer was the CIA—would provide the software free. Civil libertarians went apoplectic. Ellison watchers accused him of grandstanding. For the record, the marginal cost to a company like Oracle of providing its software approximates zero; and the installation, support and maintenance revenue opportunity would be immense.

Political land mines notwithstanding, Ellison's idea for an all-government uber-database is probably unrealistic—it's just too big and too ambitious. But ratchet the vision down a notch or two and the calculus starts to change. New clean-sheet inter-agency databases, built from the ground up, are certainly doable, given enough political will.

One step that does seem short-term possible would be a nationwide—even international—fingerprint database that ties I.D. cardholders to their fingerprints. Some states already require a thumbprint for a driver's license. A print could be required for airline passengers. Biometric fingerprint sensors are fast becoming commodity items—in large volume cost per unit would be trivial. The cards needn't contain any information other than name and physical characteristics. Would air passengers be willing to place their finger on a biometric sensor and have the database compare it to a photo? Probably. What if the whole thing were done from the ground up with the specific idea of using encryption and other technologies to protect citizen's privacy at every stage and level—while still letting law enforcement zero in on the "bad actors"? It's possible new technology rather than new laws may be an easier way to square the privacy-versus-security circle.

But clean-sheet solutions require a vision, plus the clout to knock heads until the vision is achieved. Tom Ridge made more intelligent analysis a top priority of his new Homeland Security organization, and has—we are told—the President's and Congress' full backing. Let's see.

Meanwhile, local law enforcement needs help, too—at the end of the day it's intelligence at (and from) the edge that makes or breaks networked data systems. "Until maybe two years ago most police officers didn't even have e-mail," says Jim "Gator" Hudson, a veteran former cop who ran a computer crime division in Portland, Oregon. "A lot of detectives still don't have PCs on their desk. Five or six of them share one PC. It's like the early 1900s when the telephone first came out: 'What do you need a telephone on your desk for? We've got one in the hallway.'"

Unlike Washington, turf wars and rivalries aren't as much of a problem at the local level—it's sheer numbers, more than 13,000 local police departments across the United States, each with its own internal database that contains, in effect, proprietary information. The FBI's NCIC is accessible from a squad car, but as Hudson points out, "that information is almost all on convicted criminals. It has very little on suspects." There's also the National Law Enforcement Telecommunications System, run by the states to help police departments cooperate on investigations. The system runs on—get this—teletype, requiring a human operator to punch in the text. In most businesses, telex *went out* in the 1970s.

Hudson actually quit the police to start a private business called Amcrin.net—the American Criminal Investigators Network—which rolled out commercially last May with support from major banks like Wells Fargo. Aimed at fighting check fraud, Amcrin links the information resources of commercial investigators with police departments that join the network. "A crook can pass bad checks at a department store, a bank, a utility. If they're in different jurisdictions, nobody will put it all together," says Hudson. Amcrin lets investigators at banks, retailers and other check fraud victims to load their open cases into an online database. "Hotlinks" connects private investigators and detectives in far-flung jurisdictions. The system is up and running in Oregon and Washington State, with San Francisco and Los Angeles next.

The fact is that it isn't hard for even the dimmest public agencies to figure out which way the technologic wind is blowing: just watch the private sector. For a lot of businesses, the kind of systems the FBI is struggling with are already old hat. Companies and industries are overwhelmed by data, too, but sense opportunity, not nightmare. They're moving beyond standard computing and adding the technologies of machine intelligence to find profit-making patterns in oceans of data too wide and too deep for human minds to explore on their own.

Decades of research into machine intelligence are beginning to pay off. Technologies like rule-based expert systems and neural networks, once the province of academic researchers, are hitting mainstream business. Data mining is the rubric, and businesses are applying it to marketing, drug manufacturing, semiconductor yield improvement, network intrusion detection, money laundering, fraud detection and more. Law enforcement isn't entirely asleep; the FBI's controversial Carnivore system, which can read e-mail and track suspects' Web activity, reportedly has a data mining component called Coolmine to help it analyze the information. But for most of Uncle Sam's security-related data, mining is something done by keyboard- and hunch-wielding humans, by hand.

Here's an example. Google, the leading Internet search company, jumped to prominence by applying data-mining methods to the billions—literally—of pages of Web content. The way Google ranks search results is the key innovation: first it finds pages that match the user's search term. Then it counts how many other sites have links to each page; the most popular—those linked to the most other sites—go to the top of the list. In other words, Google taps the collective intelligence of the Web to determine which pages are likely to be most relevant. Other tricks in the works or already operating include things like remembering the results you've zeroed in on during past searches; from that Google learns more about your interests and tastes, and can adjust its sights accordingly.

Contrast this with the FBI's text-string document searching: typing in "white powder" turns up a flurry of documents that include that phrase. It could refer to cocaine, it could be talcum, it could be anthrax—the system has no smart way to sort out the chaff. Surfing random data can sometimes spark intuition and insight. Poring by mouse clicks through thousands of irrelevant documents does no one any good.

Another big problem is unstructured

data. Traditional databases can only handle information in rigid formats—name, address, zip code, income, etc. They're great for performing specific, programmed tasks at high speeds. But the Web and other computer networks are flooding the world with *unstructured* data. Every day, it gets cheaper and easier to create digital text, image, photo, audio and video files – all stuff that the FBI, for example, might produce in the course of an investigation. Unstructured data is piling up at accelerating rates, not just on the Web but on internal mainframes, servers, and individual PCs (think about yours).. Some of it will eventually find its way into structured databases. Most of it will never be organized at all.

Memex, based in Scotland, sells search technology specifically to law enforcement agencies, to retrieve information from both structured and unstructured data. Its "Crime Workbench" lets users create links in unstructured data for future use, and leaves a "thought chain" that keeps track of how the investigator went about his search, to help him or others do better next time. The London Metropolitan Police (a.k.a. Scotland Yard) is the company's largest customer; the FBI is trying it out on its national NCIS system.

Even more intriguing are artificial neural networks. Neural computing grew out of research into the human brain, about whose real workings we still know surprisingly little. But even our crude understanding of how the brain's neurons communicate has been translated into rough but useful models for computation. Neural nets represent a great breakthrough in computing. Their programs can actually learn—a key capability for sorting through ever-growing mountains of data.

Neural nets start by finding rough patterns in the data. The program learns from its mistakes; as it makes adjustments to the model, the results get more reliable, to the point where accurate predictions are possible. A cell phone company might use a simple neural net to identify bad subscriber risks. Telemarketers use them to identify what kinds of people are most susceptible to which kinds of cross-selling. (Oh joy.) They can make a good guess that your credit card has been stolen by seeing that you've rarely used it anywhere but the strip in Santa Monica, when suddenly it's being charged for pricey electronics equipment in Amsterdam.

The Defense Department is funding more advanced research to retrieve multimedia files in many languages from vast unstructured collections of data. "Show me the files that describe a suitcase nuke like this one" is the type of inquiry an investigator might make. The contractor, HNC Software in San Diego, already works with credit card companies, financial institutions, retailers and telecom providers, mainly to fight fraud and assess credit risks. They also winnow bad customers from good ones by predicting future behavior. That's the kind of help investigators would love to have.

Neural net predictions are hardly foolproof. Just because the data show a connection between dealing in street drugs and large purchases of small plastic bags at the 7-11 doesn't mean police ought to question a shopper just for buying a bunch of Baggies. The best neural net models today depend on huge sets of data; for detecting credit card fraud, the model might involve millions of sample transactions. Such huge data sets naturally require that most of the records analyzed be honest credit card users. And therefore the FBI is unlikely to turn to neural nets anytime soon to, say, track the travel patterns of suspected terrorists. Not only would privacy advocates squawk; the data set of known terrorists is simply too small.

Where the FBI could put neural nets immediately to work in a big way is managing the tips and threats flooding in since Sept 11—an average of 80 threats and 10,000 tips (or more accurately, would-be tips) a day. By analyzing those that have proven most useful (or credible, in the case of threats), and assessing the factors that good tips and tipsters share, a neural net could sort out the best. Worthy tips and real threats would be less likely to go missed, and immense blocks of investigators' time would be saved.

Other technologies are more powerful, but farther away. Face-recognition technologies, much talked about in the aftermath of the Sept. 11 attacks, are actually quite crude. They "recognize" faces by plotting measurements between the eyes and other parts. Scientists at the University of Illinois and elsewhere, however, are working on systems that would enable a computer to recognize a face much the way a baby does—by taking a glance at all its features and instantly recognizing its unique appearance.

A key to making this work will be advances in data storage and computer processing power still in the lab. Within a decade, maybe sooner, many engineers believe that holographic storage will be commercially available. Rather than electronic bits, holographic storage uses light waves to etch three-dimensional patterns; terabytes of data can be stored in a box the size of a sugar cube, and searched virtually instantaneously (see "Through a Glass Smartly." page 70.) Combined ubiquitous high-bandwidth networking, the possibilities are staggering.

Here's a scenario, for roughly 2006: New airport check-in systems require a fresh photo and biometric fingerprint for a boarding pass to be issued. The data is instantly matched against a locally stored holographic database, updated constantly in real time through high-speed connections to the integrated databases of the FBI, the INS, the DEA, State and Customs. When a criminal or terrorist suspect is matched, the system quickly pings airport security, the local FBI office and local police, while simultaneously creating a "hot" database file for easy access on a smoothly integrated cross-platform system.

Here's another, for 2016: Intelligence agents learn "something will happen" at the Super Bowl LI. Police officers scan stadium crowds with handheld cameras that capture moving images of each person. Seeking a match, the camera's optical processor runs live footage through thousands of pre-loaded 3D images of suspected terrorists, lights up when it picks one out of the crowd, and makes a quick connection to a database identifying who else to look for and what sorts of mayhem they are likeliest to be planning.

Technology's powers truly are awesome. A culture of freedom and democracy that allows entrepreneurship to flourish is the fastest—the only—way to make it happen. We've got the tools we need, and we're continuing to develop more. What we need is the vision, brains and will to use them.

Meantime, decent desktops and DSL lines for the FBI will help.