ARCHITECTURES FOR BUSINESS-CRITICAL MOBILE MAPPING

Paul M. Wilson
GE Energy, 5420 LBJ Freeway, Suite 1250, Dallas, Texas, USA
paulm.wilson@ge.com

KEY WORDS: AutoCarto, GIS, spatial infrastructures, data structures, mobile

ABSTRACT:

Although location is now a common part of mobile computing for consumers, there is a class of much more demanding applications we can call “high stress GIS.” These apps are time-critical as well as business-critical; they have to work whenever and wherever they are needed, and there are significant consequences when they don’t. This paper employs a specific type of user, the utility field crew, to illustrate the architectural issues that must be addressed in this kind of application. These users work at the edge of the enterprise, relying on hardware platforms and communications networks that are not optimal for the demands of a high-performance GIS. Complex data sets, covering a large geographic area, have to be kept updated and made available to crews that are spread around the utility’s service area. In the aftermath of storms, wireless communications networks may be unavailable, making it hard to update important data just when it is needed most. These systems play a vital role in managing our critical infrastructure so - despite these challenges – it is important that we can make them work. The paper outlines a number of these design challenges and suggests solutions that have been tested in current mobile applications.

1. INTRODUCTION

It’s hard to avoid maps on mobile devices these days. Location-based social networking with smart phones is now commonplace. Navigation, with GPS and other methods of determining position, is possible with a wide variety of devices. Given this pervasive usage of location in the field, can we assume that the key technical issues around mobile mapping have been resolved?

For many users, they have. But for business-critical mapping applications, where the importance of the application goes beyond convenience, the game is different. An example is the utility field crew, the “tech in the truck” responsible for getting the lights back on after a storm. These users are dealing with problems that are inherently spatial and have to be addressed on site. The application needs to work whenever and wherever it is needed. Failure to work can lead to economic and safety consequences.

This paper looks at those critical applications, with a focus on the geospatial information structure needed to support them. How do we design GIS systems that perform well in a challenging environment? Utilities are a good illustration, but certainly not the only example; emergency services and military users face many of the same challenges, so it’s important to design systems that can work in these “high stress GIS” conditions.

First, we’ll look at the context - maps and the electrical grid – and then describe the nature of utility field systems, with a brief look at how they will change in the near future. Then, we examine the design challenges for these applications and suggest architectural strategies for overcoming them – and identify areas where research is needed.

2. BACKGROUND - MAPS AND THE GRID

The electrical grid is inherently spatial in nature, spread over the geography of the service territory served by the utility. It’s a complex network of wires, supported by devices that control the flow of electrons through those wires. The complexity is multiplied by scale; for a large utility, the grid can cover thousands of square miles and millions of components.

To build and manage that network, the utility has to know the location and attributes of all those objects, and also how they are connected. A major part of this content is geospatial. A map is the natural mechanism for storing and disseminating this knowledge. For almost a century, the utility’s method for storing all of this spatial data was the paper map. The “data” was created and maintained by manual drafting, and distributed by making copies of large map books for each user.

Starting in the 1970s, the growth in computing technology offered utilities a better way to handle spatial data. Over the next two decades almost every large utility invested heavily in CAD and GIS. Early systems were complex; for many years, they were accessible only in the office and were used only by GIS professionals. Front-line users (the crews in their trucks) were, for the most part, still limited to using paper maps.

Digital maps finally started to reach the field in the mid-1990s. As mobile computers rugged enough to survive field conditions were deployed, paper disappeared, replaced by viewers that could search for objects (show me a pole based on its number or tag), display attributes (what are the voltage ratings on that transformer?), and even trace through the network to identify trouble spots. A rich set of visualization and analysis tools made GIS capabilities accessible to operations managers and field crews.

3. UTILITY FIELD AUTOMATION TODAY

In many ways, this drive to automation in the utility industry has mirrored technology trends in other sectors, where the platform for computing has moved steadily closer to the user's place of work.

Getting geospatial data to the field is key to a utility company because much of the work has to be done outside the office. The assets and the customers are all outside the building, arrayed across the service territory. Consequently, much of the utility workforce is also outside the office. Their job is inherently mobile, moving around the grid "in the geography" to be a job locations that change rapidly.

Some of this work, while business-critical, is not time-critical. Data collection, in most cases, can be rescheduled without major consequences. Similarly, many field inspections can be completed within a known time window (usually defined by the compliance regulations that drive the particular inspection type).

Other kinds of work – and the real focus of this paper - are those functions that are both critical to the business and where timeliness is a key component. It's the daily work of crews as they maintain the grid. Most importantly, it also includes the times when things go wrong, when gas is leaking or there's a storm and the lights go out.

To do their job well, the crews need a wealth of information. A great deal of it is locational in nature: where am I, where do I need to be, how do I get there, what's there, what is it connected to? In many cases the "there" element is based on an object that's part of the grid. GIS systems are a natural fit for this type of information.

A natural fit, but not an easy one. Several characteristics of a utility combine to make field applications hard to implement. The data volume is high, usually with a lot of detail for a large area.

When the lights are out after a storm, data can start to change very quickly. For example, as the system is reconfigured to reroute power and resolve outages, the status of switches may change. Some of the information about the grid is less likely to be valid if there is damage. And public demand to get the lights back on adds to the time pressure; making this a true "high-stress GIS" situation.

4. SMART GRID – NOW LET'S MAKE IT HARDER

After many years of development, advances in hardware and software technology have finally been able to address most of the difficulties noted above. Current GIS practice is, for the most part, adequate for handling the demanding task of automating today's utility field force.

But that is about to change.

We are on the edge of another new era of applying geospatial tools to managing the grid. This time the changes are driven not by gains in geospatial technology but by the transformation of the grid itself: the emergence of the smart grid. Even though large-scale deployment of smart grids is in its infancy, it seems clear that the nature of the new grid will force major changes in how spatial data is managed and distributed.

The electric grid of today has remained fundamentally unchanged for a century: a one-way, static network where a flow of electrons is created at a small number of power generation plants and distributed to customers. A lot changes with smart grid. It becomes a resilient, self-healing network that combines electric and information flows. It adds increased control for both the utility and the consumer, along with the ability to handle dispersed, renewable energy sources.

There's a big impact of the utility's crews. It's a much more complex system that changes much faster. The GIS that models the grid will change dramatically, with new objects that behave differently and relate to each other in new ways. Everything scales up and moves faster.

The bottom line is that the utility's business-critical field tools, which already stretch the bounds of available geospatial technology, become even more difficult to build and maintain. These applications, then, can be called "high stress GIS" is two ways: they are used in situations where the stakes are high, and they push the envelope of GIS techniques.

5. DESIGNING FOR BUSINESS-CRITICAL APPLICATIONS

The forces described above pose significant architectural challenges to current geospatial design practice. This section looks at a number of these challenges and suggests design strategies for addressing them.

In essence, this type of critical application forces GIS designers to deal with the edge. The user is literally at the edge of the enterprise, and has to rely on a hardware platform (and, in some cases, a communications network) that are far from optimal for the demands of a complex GIS.

5.1 Maintaining availability

Availability is key. The application must be accessible, with at least a minimally useful set of data, under all conditions. It has to work whenever and wherever it is needed. Waiting until later is simply not an option; too much is at stake

This mandate limits the available choices for system architecture. The application has to work at some useful level even if the communications infrastructure is damaged or unavailable, so it must be designed with the assumption that communications is not there. A large utility's service territory may extend beyond the coverage of commercial networks, which presents problems even with normal access. And in recovering from storms or other events – the very cases when the applications are most critical to the utility – the network may not be available at all (or even when it is, bandwidth will likely be limited).

This means that a pure thin-client architecture will not work. The popular web 2.0 style approach, serving map data as needed, will not be able to guarantee availability. It is noteworthy that Google, which has been at the forefront of web-based mapping, has recognized this limitation and now provides an option (in Google Earth Enterprise Portable) of downloading a designated area to the mobile device.

The business-critical mobile system, then, has to be an application and a set of data residing on the mobile computer, capable of supporting the field crew in a standalone or "island" mode.

This constraint does not mean that apps cannot take advantage of wireless networks where they are available. Access to data updates, as discussed below, is very important. One design guideline is to allow for, where possible, the ability to access multiple networks. This should increase the chances of some communications availability.

The bottom line, however, is that the system must gracefully handle limited access, and must work in a pure standalone mode when necessary.

5.2 Keeping it up to date

Although most of the spatial data used by utility field crews is relatively static, some can change rapidly, and outdated data can have major safety and operational consequences.

The advent of smart grid technology, with more data changing more rapidly, adds to the need for frequent updates. An example is the spread of renewable energy generation. Wind farms and rooftop solar add new sources of power that are not controlled by the utility. When these are added to the grid, flows on the electric network can change and crews cannot rely on traces.

An obvious solution for this issue would be a browser-based system that serves updated map views from an enterprise server. (Keeping a utility's spatial data updated on the server poses a whole set of other problems, but that's outside of our scope here.) That design, unfortunately, is in direct conflict with the need for availability described above.

How do we resolve this contradictory set of requirements: a need to constantly refresh data in the field, but working under the constraint of not relying on a communications channel? There is, of course, no perfect answer to this dilemma. There are, however, some guidelines that can minimize the chance of problems.

One way to help is to establish a regular protocol that regularly updates data on field computers when there is connectivity. This minimizes the amount of data that needs to be updated at any given time and reduces the likelihood of outdated data if there is no communications. For example, many utilities now employ a process that updates each mobile computer every day.

Another strategy revolves around the geographic coverage of the data sets stored on each mobile computer. Even though a given crew usually works within a relatively small area, many utilities prefer for the mobile app to have data stored locally for the entire service territory so crews can be moved around as needed in an emergency. This approach can lower the likelihood of having to rely on data communications at the time when it is most likely to be compromised. Crews can be moved around without going outside of the geographic coverage on their machines.

When bandwidth is available, but limited, it has to be utilized effectively. A third strategy, then, is to have clear priorities for updates, compensating for lower bandwidth by selectively updating based on factors like layer, location, and user type. Key grid devices like switches, for example, should always be given priority since their open /closed status has safety implications for the crews

Finally, the application should be designed to make age of data visible to the user. The field user will be able to make more informed decisions with knowledge of what data may be outdated.

These methods have helped many utilities continue effective field operations when the communications infrastructure was unavailable. A large utility in the Midwest, for example, was able to move crews to help out in damage areas after an ice storm because all of its mobile computers stored data for the full service area.

5.3 Coping with scale

Data volumes for utility apps are already high, with detailed landbase and complex facility maps that often cover a large geographic area. It is not uncommon to see utility GIS databases in the hundreds of gigabytes, not counting related data sets like customer information and work schedules. (And, as noted above, there are operational reasons for storing data for the entire service territory on each mobile machine.)

Everything gets bigger with the smart grid. The combination of new objects and increased detail, replicated over a large area, means data volumes scale rapidly. Data structures that work reasonably well in the current environment may not meet performance expectations. It's important, too, to keep the hardware platform in mind; database products that work well in the office may be hampered by the processor and storage constraints of the typical mobile computer.

Although the increased data volume can be mitigated somewhat with good data models and compact data representation, the critical nature of these applications makes it hard to leave out any data that might be essential to some field operation.

It would seem that some form of hybrid data structures, recognizing the static nature of some data types but also capable of handling large volumes of transactions, will be necessary. Today's commercial GIS software will require fundamental redesign to cope with the high volume, near-real-time data demands of the future grid.

5.4 Supporting geocollaboration

One of the paradoxes of the smart grid is that, even though the system is designed to be self-healing, its deployment will put even more burden on the people who maintain it. The growing complexity of the network will increase the need to collaborate: multiple crews, along with supervisors in an office, will work together to handle a situation like a large outage. These users would benefit from a higher degree of interactivity with the back-office GIS (data input or sketches) to communicate changes; this can be seen as a need for geocollaboration on a large scale.

Almost all of a utility's critical field applications are highly spatial in nature. They revolve around problems that are "of" the geography being worked on by crews that are "in" the geography. We can take advantage of this characteristic by designing mapcentric systems that utilize spatial data as the basis for communicating.

The mobile application will also have to support a higher degree of interactivity, with tools to record field conditions or sketch on the map. And, once this information is recorded, there will have to be a quick way to move it to other members of the group (crew-to-crew communications).

How do we support interaction among distributed users? This poses another challenge to communications infrastructure, which in the aftermath of a storm is most likely to be diminished just when it is needed the most.

In some cases it may be necessary to use backup methods for moving data. One large utility, for example, equips crews with flash drives that are configured to handle manual transfer of data. This facilitates the task of collecting and communicating damage data after storms. Their protocol is to utilize wireless when it is available, but to manually transfer data using the flash drives when the crews have to operate in "island" mode.

5.5 Insuring Usability

In the past most complex GIS apps were used in the office by professionals. Mobile settings were handled with "light" versions or specialized viewers, typically using a subset of the enterprise data and a restricted set of functionality. For business-critical processes, this approach may no longer be viable. Complex problems require more complete solutions. A crew needs analytical functionality that goes beyond simple map display and attribute query (e.g., a network trace to find the switch controlling power flow in a specified area).

We are, in effect, extending the full power of GIS to the mobile workplace. But as the complexity of the app grows, it is vitally important that we emphasize usability.

This is especially true in field settings. It's a challenging work environment; the display screen is typically smaller than is common in the office and viewing conditions are rarely ideal. In events like storms, there is a great deal of pressure on the user (and consequently the system) to work quickly. True situational awareness is a key and the application must support it, not get in the way.

There is an understandable tendency to start with desktop systems and minimally modify them for mobile users. After all, a mobile computer does share most of its hardware components (and often the same operating system) with its desktop counterpart. While this approach does extend the full power of GIS to the field, it almost always falls short in usability. Assumptions about screen size and input devices don't translate to mobile machines.

Recognizing the nature of the user is equally important. The GIS professional in the office, backed with adequate training and experience, can navigate through nested menus and toolbars with ease. The field user is expert in something other than GIS, and typically has very limited training in spatial applications. And in business-critical settings, the field user is often under intense time pressure, with very little patience for a complex interface.

There are some fairly obvious design guidelines, like an uncluttered screen layout that maximizes the screen area devoted to the map. Methods like gesture-based interaction are easy to use and can help reduce the training burden.

We also have to keep field conditions in mind when we design map symbology. Certain colors are often used to designate specified categories of electric lines. Some colors, like yellow, show up well in the controlled conditions of office lighting but can be very hard to see with sunlight glaring on the screen of a mobile computer.

One area that deserves research is how to help users pick out key information in a system with more potential for clutter. Is it possible to define key data based on context (location, time of day, current user activity)?

In general, it seems like we have much to learn from consumer markets. User interfaces designed for the casual user do not necessarily translate directly to business-critical apps, but many usability principles (multitouch interaction, simplicity as a philosophy) do apply.

5.6 More Open, More Closed

Like many critical systems, the utility GIS has always been a very closed environment. Security is a real concern; the grid, as it becomes more automated, is subject to electronic as well as physical attacks. Utilities protect critical data by treating it as proprietary and storing it on servers inside the enterprise firewall. The cloud is not an accepted tool for storing utility GIS data sets.

Another reason for the closed nature of utility systems has been a concern for data integrity. Companies collect and update facility data through structured (and often time-consuming) processes. Since data accuracy is so important the utility demands control over all aspects of the GIS database.

These concerns are particularly acute for field applications if critical data sets are stored on computers that are more subject to damage or theft than office machines. Security should be an integral part of the architecture for these apps.

Paradoxically, the changing grid will also demand that systems be more open to sources of data outside the normal utility GIS.

There will be RFID tags on many devices and smart meters at the point of delivery. The grid will be, in some ways, a perfect example of the "internet of things." Although most of this data will only affect back-office systems, some of it will undoubtedly make its way into field applications.

In addition to these inputs that are controlled by the utility, there will be a need to integrate external data sources. Current and forecast weather data, for example, will be an important tool when trying to predict power flows from geographically-dispersed solar and wind sites. Electric vehicles, which represent both a load and a potential source of backup power for the grid, will pose an additional complication for utility systems since their location can vary (in both XY terms and in how they connect to the grid topology).

How can systems be designed to handle these conflicting directions – more closed to enhance security, more open to embrace new data sources? Hybrid data structures, employing added security for sensitive data sets, might be combined with more conventional methods for landbase or public data.

6. CONCLUSIONS

Geospatial technology – even when the technology was ink on paper – has always been an essential part of managing the electric grid. Extending that capability to field crews is critical to the utility business now, and will become even more critical as the grid becomes smarter.

Further, as the grid undergoes these major changes, it seems clear that the GIS technology being used today will not be able to keep up. We will need to develop systems that can scale to handle massive, rapidly changing data sets. Systems will need to take advantage of wireless methods for updating under constraints, but gracefully handle situations with limited or no bandwidth.

Design of business-critical mobile systems must also recognize the realities of the field environment, with particular attention to usability.

Research in several aspects of geospatial technology will be needed to overcome the design challenges that are inherent in extending the power of GIS to field users.