Application of Network Visualization to Identify Gaps in Complex Information System Architectures

Carlos E. Martinez

Sheila A. Cane

Salwa Abdul-Rauf

Kevin Smith

Kristin Lee

POC: Carlos E. Martinez

The MITRE Corporation

7515 Colshire Dr.
McLean, VA 22102

703-983-7477

cmartinez@mitre.org

Approved for Public Release; Distribution Unlimited
MITRE Tracking Number: 08-0109

Application of Network Visualization to Identify Gaps in Complex Information Systems Architectures

Abstract

This paper follows earlier work aimed at ensuring a connection between the information sharing needs of a large number of participants comprising a distributed command, control, and coordination network and the underlying communications capabilities that they possess; and on the relational model and database built as part of a subsequent effort to analyze the complex relationships involved.

The relationships among participants and their varied means of communications is captured in a relational database, which provides the means to automatically catalog needs and capabilities as well as to enumerate gaps between participants. However, complexity due to the large numbers of entities and relationships makes it difficult to readily identify the most critical gaps, or identify patterns in gaps among various classes of participants by using tabular reports alone. This paper examines the applicability of network visualization techniques linked directly to the underlying relational data elements to analyze and portray information in a more intuitive way, including identification, comprehension, and presentation of participant interdependent relationships and capability gaps.

The paper will present a summary of the problem, a discussion of visualization techniques that will be used to conduct analyses, and an assessment of the utility of various visualization approaches. Since the paper is being written on the basis of current ongoing research and analysis in the classified arena, the paper will present notional examples rather than actual results.

Introduction

Organizations are increasingly called upon to work in concert to achieve a common purpose. Examples include multi-national military coalitions formed to establish and maintain peace in a distant land; local police, fire, and emergency rescue units jointly tasked to respond to a major disaster; and metropolitan transportation authorities directed to ensure smooth traffic flows in and out of large cities that straddle jurisdictions. Usually the individual entities in these partnerships were created to meet different objectives; governed by different entities; and characterized by very different organizational structures, processes, and communications infrastructures. These and similar entities are often federated into joint organizations to ensure proper command, control, and coordination of crisis management activities. The necessary business processes, information flows, and supporting information systems of these joint entities can be represented as a complex, information systems architecture. Each member brings missions, organizations, information elements, facilities, communications, and information technology components to bear on any particular scenario. A notional model of this complex architecture is shown in Figure 1 (Lee, Cane, Abdul-Rauf, & Martinez, 2007).

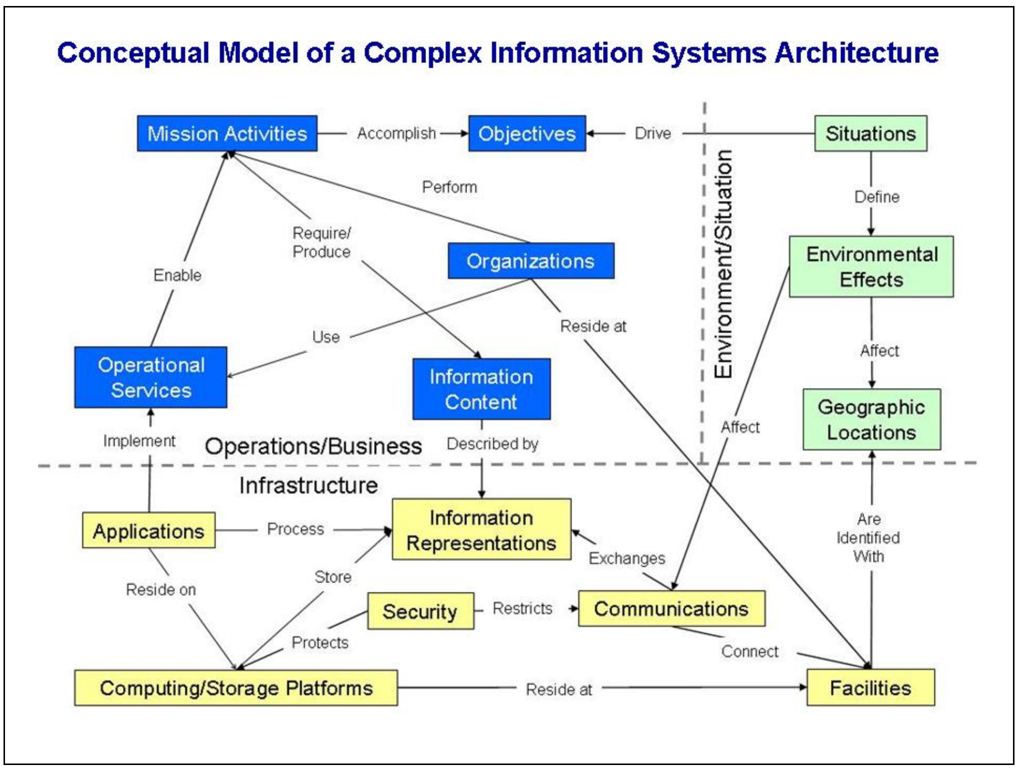


Figure 1: Conceptual Model of a Complex Information Systems Architecture (Lee, et. al, 2007)

To address the complexities represented by this architecture, the authors developed a relational database that would permit *pulling analytical threads* through the architecture to answer a rich variety of questions regarding the ability of a complex organization to meet its mission objectives given the infrastructure components its members possessed (Martinez, Mullins, & Sullivan, 2006) . Lee, et al. (2007), describes the analytical objectives of the architectural database, as shown in Figure 2. This paper addresses the analysis required to identify operational gaps, and to associate proposed changes to these gaps, as shown in Figure 2 by the arrow between the as-is operational/business architecture and the to-be operational/business architecture. It also begins to address at a high level the analysis of organizational infrastructure required to support operational needs.

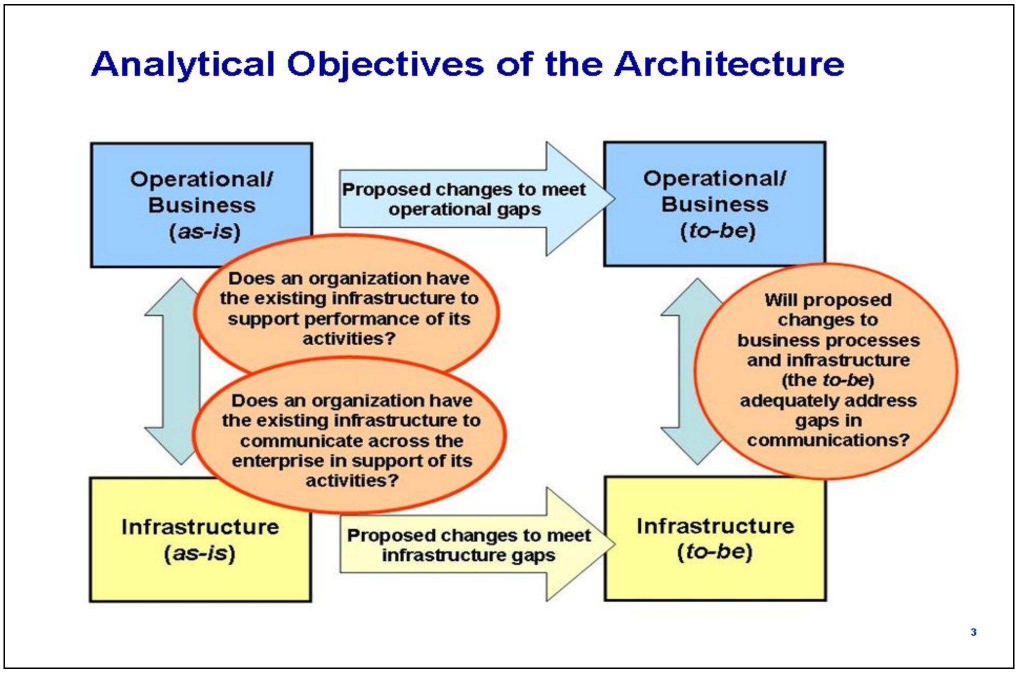


Figure 2: Analytical Objectives of the architecture (Lee et al., 2007).

The use of a model such as this presupposes that the operational/business model has been adequately defined. Where a central authoritative source exists, it may be fairly straightforward to define and document the operation/business model, although for a large-scale business enterprise, it may still be fairly laborious to do so.

However, for a large-scale complex system such as that represented by a distributed command, control, and coordination system, where a set of independent organizations must work in concert to accomplish a set of common objectives, it is not likely that there is a single authoritative source to provide the necessary information to define the operational/business model. Each organization may have to be polled independently and their multiple inputs integrated into a single model. To do so, their inputs need to be reconciled to ensure that the organizations possess a common understanding of their respective roles and responsibilities and a common perspective of their needs to communicate between organizations. To do so, questions such as the following must be answered:

1. What functions must each organization perform to fulfill their responsibilities within the overall context of the complex system?
2. Which organizations believe they have a need to communicate with which other organizations to accomplish their respective functions?
3. By what means do these organizations need to communicate with each other?
4. Do the various organizations agree with respect to their mutual need to communicate and the means by which they must do so?
5. Do the organizations with identified needs to communicate with each other represent reasonable groupings of organizations that should be expected to communicate with each other in a distributed command, control, and coordination system?

It is to answer questions such as these that the authors have proposed the use of visualizations of the architecture model—visualizations drawn directly from the data. Assuming one has collected the necessary information from each of the organizations and entered the results into an integrated relational database, the first three of these questions can be answered via fairly simple queries with the results presented either in tabular or graphic form. For the fourth question, more complex queries are needed, and while the results may still be presented in tabular form, the advantages of visualization techniques are apparent. Finally, to answer the last question, which requires the considerable application of human judgment, some form of visualization becomes almost a necessity. Beyond answering pre-defined questions from the database, another advantage of visualization is its ability to inspire the analyst. Visualizations help suggest additional queries: when an analyst sees a picture, it may stimulate another question.

This paper presents the results of the research team's work in prototyping a visualization approach to presenting and analyzing a complex information system architecture, with a particular emphasis on inter-organizational communication support for key missions. It begins with a discussion of alternative architecture tools and visualization methods considered, describes the team's methodology for conducting analyses to help establish the operational/business model for a complex, distributed command, control, and coordination system using a visualization approach, and presents comparative results of the use of tabular versus visual representation in support of establishing the operational/business model. The paper provides initial conclusions, and offers recommendations for further study.

Enterprise Architecture Frameworks and Tools

Enterprise architecture provides a framework or taxonomy of systems analysis models and is used to align information technology with organizational strategy (Rico, 2006). Enterprise architecture is a "strategic information asset base, which defines the business, the information necessary to support the business operations, and the transitional processes necessary for implementing new technologies in response to the changing needs of the business" and is used to explain how the information technology elements of the organization works together, including people, processes, systems and organizations (Morganwalp and Sage, 2004). Enterprise architectures, by their very name, are designed to address large-scale, complex systems and employ the use of visualizations to support comprehension of the information they present.

Many architecture frameworks have been defined in the recent past. Notable examples include Zachman (1987), Spewak (1992) and the US Department of Defense Architecture Framework (DoDAF). The Zachman framework includes architectural models developed by considering classic building architecture processes, and aims to provide the different perspectives of all of the different participants. Spewak's enterprise architecture planning method is less complex than Zachman's, and attempts to provide an overarching plan for an enterprise's technology, including the definition of architectures representing the information used to support the business and the plan to implement those architectures. Spewak's architecture is often represented as a "wedding cake." The top layer includes the initiation of planning; the second layer describes the as-is state of the organization and includes models of the business and a description of the current systems and technology; the third layer identifies the data, applications and the technology architecture required for the future state; the fourth layer describes the implementation and migration plans required to make the architecture a reality. The DoDAF principally is a set of architecture models and data that are driven from the perspective of supporting engineering design and analysis, which provides much more detailed views of an architecture, but consequently serves a narrower set of constituents.

The enterprise architecture frameworks as implemented in many tools consist of sets of useful abstract views of an enterprise, each capable of being represented by one or more visual aids. However, these abstractions have purposes which may be better suited for one end and not for another. The visual aspects of enterprise architecture tools are used to conceive of and portray technical or operational solutions. While these enterprise architecture tools provide techniques for design, when the scope is large they have limited capability to visually present analytical results. The focus of architecting tools (to date) has been geared towards system engineering and graphics generation or modeling, and not on the data or the analysis of the data captured in the models. That is, in developing a architecture a user is directed into developing a graphic model, either based on some standard or a template provided by the tool. For example, in developing a business process the user is taken directly into a drawing environment to model the process. Similarly, to develop an organization chart a drawing environment is used to model the organization.

There are numerous architecture tools that have been developed to help portray enterprise architectures (e.g., System Architect, Metis). These tools are principally designed for the user to *draw* or *model* an architecture (usually for a system or a business process). While modeling a system, it is as important to fully understand and develop the architecture's underlying data/information constructs, relations, and rules, as it is to draw a model of the system and organizational relationships. While the tools provide very robust capabilities to visually depict various views of the architecture, they presume that the architect or analyst already possesses a sufficient understanding of the underlying complexities of the system being modeled to accurately and adequately depict these relationships using the visualization capabilities provided by the tool. Furthermore, once the models are drawn, the tools do not provide off-the-shelf analysis of the data/information portrayed in these models to, for example, examine gaps in capabilities either internal to an organization, or across organizations.

Herein lies the problem with the use of most architectural tools available today. If the analyst does not already possess the necessary knowledge and insight into the relationships in the architecture model, such as is likely to be the case in a complex distributed command, control, and coordination system, he or she is at a major disadvantage in presenting an accurate visual representation of the architecture. Our experience shows that it is easier to for an analyst to grasp complex information in a graphic format rather than text. Consequently, a means must be provided to enable the analyst to enhance comprehension of the relationships inherent in an architecture model. Furthermore, while enterprise architecture tools may provide adequate capabilities to support engineering level decisions, they have limited utility in summarizing the information to support executive level decision making.

Database-Driven Analytical and Visualization Approach

Our approach takes the opposite approach of most enterprise architecture tools and creates visualizations from data, rather than data from visualizations. Specifically, our approach is to use visualization of the data to provide the capability for analysts to literally "see" complex relationships and immediately recognize disconnects in capabilities, both operational and technical. To do so, requires the development and population of an integrated relational database as described in Lee, et al (2007). The team's approach to development of the database consisted of the following steps:

1. Define the issues and questions to be answered by the architecture. For example, what are the relationships between organizations that allow them to communicate to execute missions, and what are the gaps in those relationships? Definition of the issues and questions helps direct and define the data modeling process, scope the breadth of the data collection effort, and define the level of detail that needs to be captured.

2. Build a relational database to provide explicit links between the elements of the architecture. The database should include explicit data elements defined to levels of detail required to address the issues and answer the questions defined in Step 1. While there is room here for a cost-benefits assessment, in general the team believed the best strategy was to build a database that models the actual environment with as much fidelity as possible, and provide custom views of the data in the form of tables, charts, and other visualization aids to meet a variety of specific analytic needs. The database employs key fields that provide the capability to implement complex queries. In developing the database, the team focused on the data, its constructs, relations, and rules. By "construct", we mean the definition of the data object and its attributes; the "relations"

define the relationships that exist between the defined objects; and the "rules" define specific rules associated with a particular attribute or collection of attributes.

3. Build and use tools/mechanisms to enter and maintain the data in the database. The team considered collecting and displaying the data in spreadsheets. While that may serve a particular need well, it may be of no or limited value for other needs. Attempts at simplifying data collection and abstracting away complexity may have short term utility, but often lead to analytic dead ends. To make the data available for later analysis, the team developed rule-driven data entry screens that employ "pick lists" of standardized terms to facilitate data entry and ensure a consistent data set that can support automated queries. The team employed a combination of MS Access and SQL Server to provide data entry and query capabilities. Given that we built an open database, we could easily connect or import data to any number of tools (e.g., develop an organization chart in System Architect).

4. Build and use tools/mechanisms to generate graphic visualizations directly from the database. The team selected netViz to provide visualizations of the database contents. The netViz software is an off-the-shelf network analysis tool that allows you to search and filter data, view connectivity between nodes, and portray data in tabular format. In addition, netViz provides the capability to show gaps in connectivity. While netViz is typically marketed as a telecommunications analysis tool, we used it to analyze operations and gaps. Having chosen netViz, we found we could model objects using any set of visualization icons and any number of types of lines to connect those objects.

Proposed Visualization Methodology

As described in the Introduction, the first problem set the team addressed was to develop a consistent operational/business model for a set of organizations comprising a complex, distributed command, control, and coordination system. A complex endeavor is "characterized by a large number of disparate entities that include not only various military units but also civil authorities, multinational and international organizations, non-governmental organizations, companies, and private volunteer organizations" (Alberts and Hayes, 2007, p.9-10). The complex endeavor we are concerned with is loosely organized, with no centralized authority, where the components need to work together to achieve a set of objectives. To enable development of a consistent operational/business model among the organizations within the complex endeavor requires identification of inconsistencies among participants, both with regard to their perceived needs to communicate with to each other, and the means by which they must communicate.

The relational database was developed to capture information on communication needs from each of the participating organizations for each of their functions. Communications needs include who the partner organizations are as well as whether or not they communicate using voice, video or data. The team defined a set of visualizations to permit recognition of inconsistencies and disconnects among the data. A data inconsistency occurs when two different data elements with the same meaning have different names. Inconsistencies are likely to be found in the data because each organization involved in the project will provide inputs to the database from their own perspective with little or no opportunity to coordinate responses with other organizations. The team provided pick-lists to minimize the occurrence of data inconsistencies for simple entities. However, for complex functions, it is not feasible to provide pick lists. Inconsistencies can be found by examining the data. We define a disconnect (or gap) as occurring when two organizations identify conflicting partnerships. For example, if organization A identifies organization B as a partner, but organization B does not identify organization A as a partner, there is a disconnect between organizations A and B. Inconsistencies can be found in organizational relationships as well. For example, if organization A and B each identify the other as a partner, but they identify differing communication needs, their relationship is considered "inconsistent".

Figures 3 and 4 present the team's proposed visualization approach to enable analysts as well as senior managers to quickly recognize inconsistencies and disconnects among the operational/business model of the organizations with respect to their inter-organizational communications needs.

Figure 3 describes the symbology used to describe organizational relationships as well as high-level communications needs.

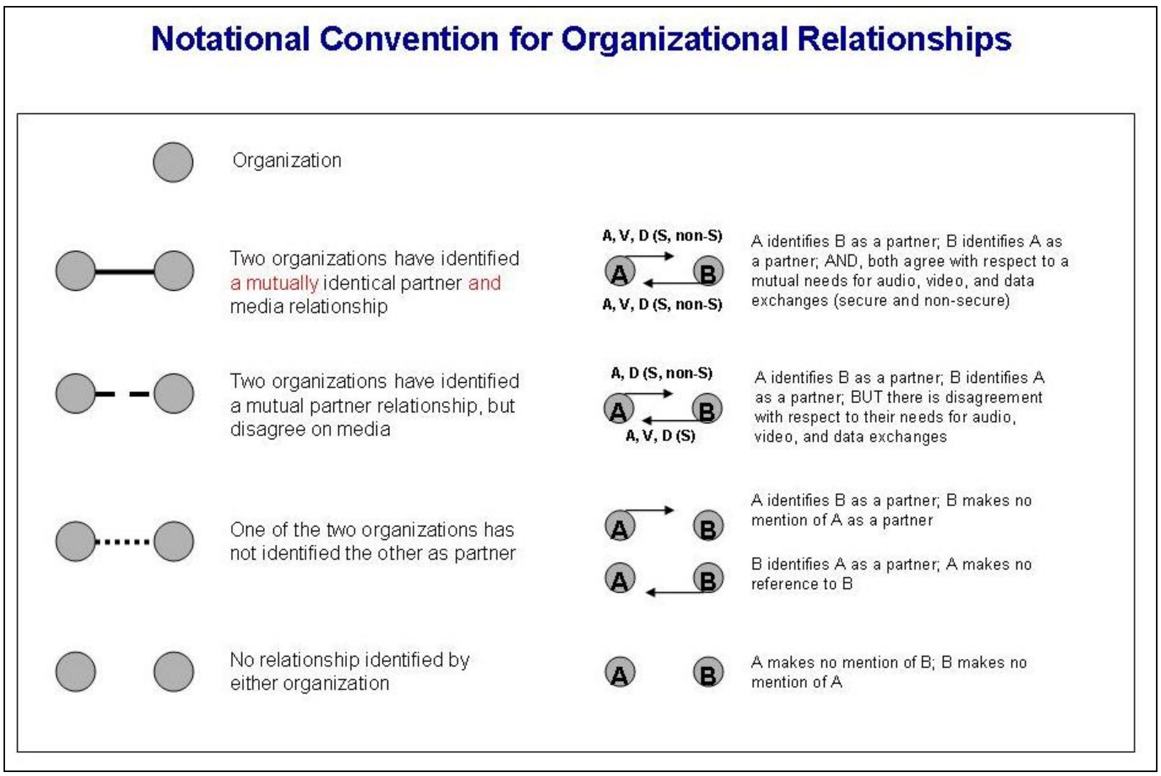


Figure 3: Visualization Symbology to Describe Organizational Relationships

Organizations are portrayed using circles (nodes). The lines between nodes represent the relationships between the organizations. For any two organizations, if each identifies the other as a partner and they agree on media (voice, video or data) the relationship is fully consistent and the line is solid. If they both agree they are partners but don't agree on the communications media, the relationship is partially consistent. If one has identified the other as partner, but the partner has not reciprocated, the relationship is inconsistent. The right side of figure 3 articulates the logic behind these relationships.

Figure 4 contains an example visualization of an executive level view of the status of the complex endeavor's communications. This view shows an overall picture of nodes and their partners. Organizations are green when all of their relationships are fully consistent, i.e., they know who their partners are, and they know at a high level how they will communicate. Organizations are yellow when only some of their partner relationships are fully consistent. Organizations are red when none of the relationships with their partners are fully consistent. While a complex organization has no central authority, there will usually be an architect or executive that has overall responsibility for identifying issues. This view provides the architect or executive a means to understand the organizational disconnects, as well as help them to prioritize risk mitigation. This view provided over time also gives the executive a means of tracking progress in correcting gaps.

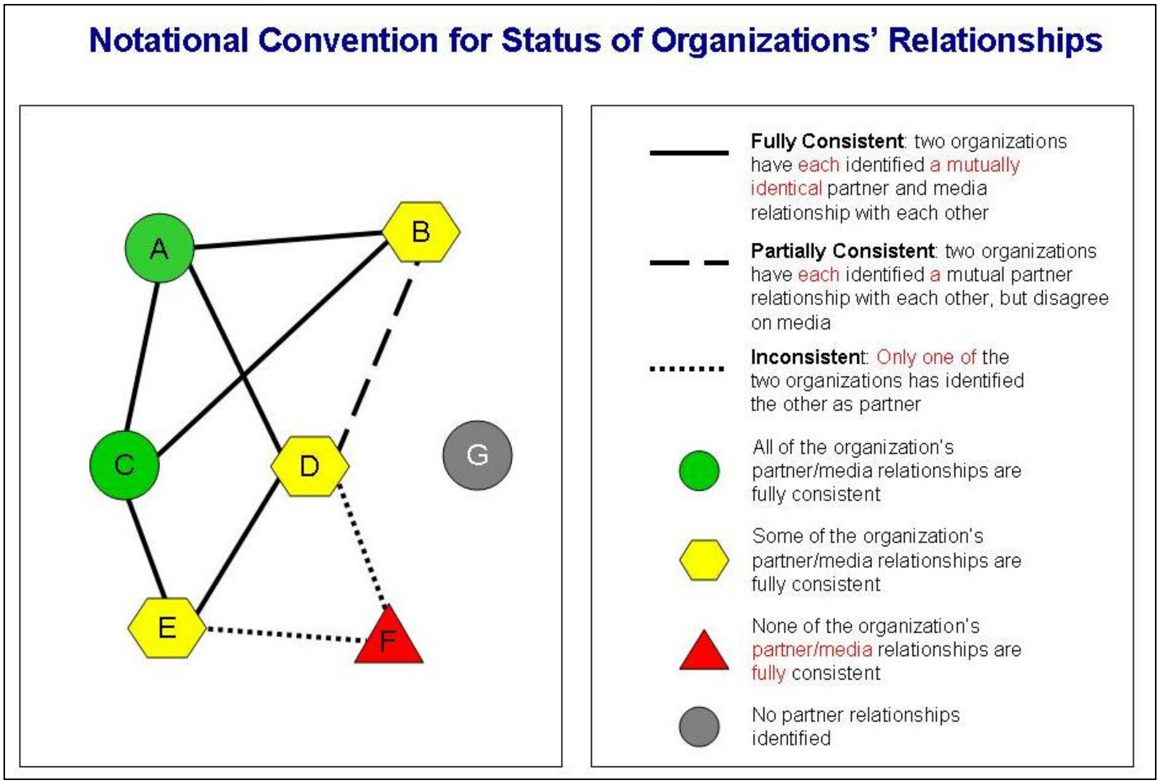


Figure 4: Sample Visualization of Organizational Relationships

Given the manually-developed symbology defined in Figures 3 and 4, the team evaluated netViz to see if it could emulate the symbology using the information from the database. Because the actual problem set against which the team's methodology is being employed is classified, the team defined a completely unclassified, yet comparably complex problem against which to test the methodology. The test set consists of the primary functions of a number of local government organizations that may have to work cooperatively within a given region to respond to a major emergency. Figure 5 presents the selected symbology as implemented in netViz. Figure 5 demonstrates another feature of data visualization—the layout of the various organizational elements to show relationships among different types of organizations such as mayor's offices, police departments, fire departments, etc.

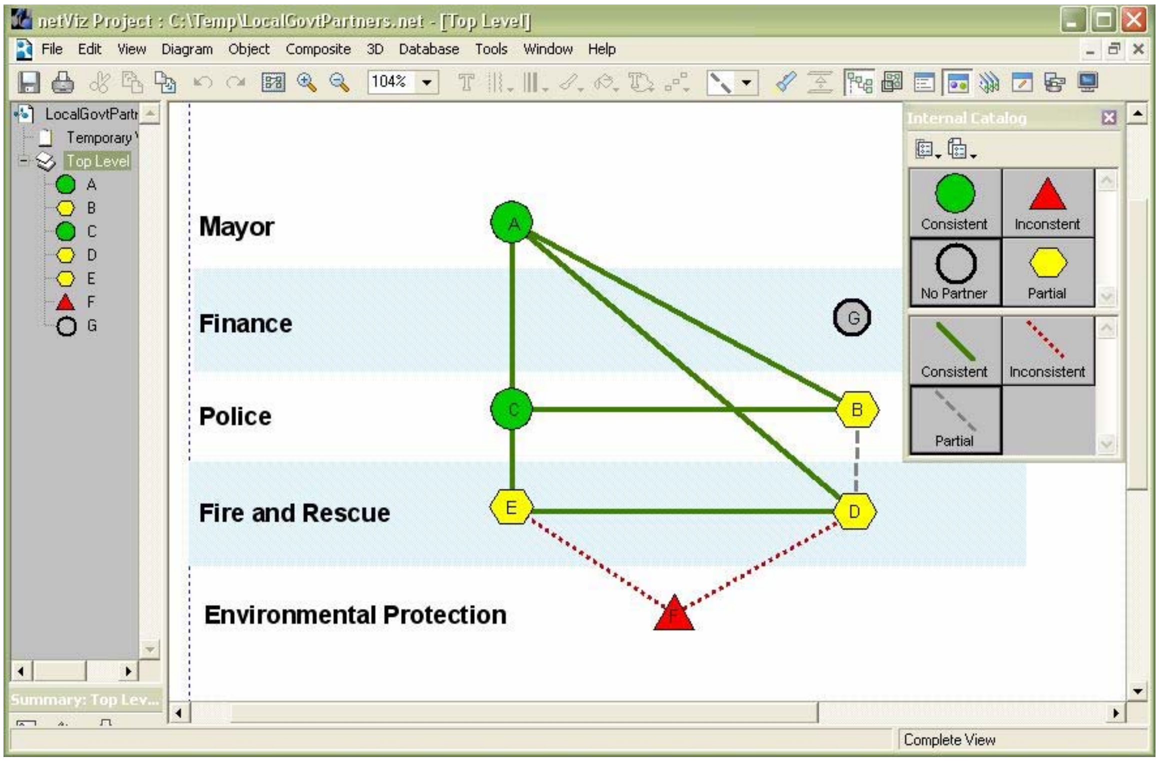


Figure 5: Sample netViz Representation of Inter-Organizational Relationships

Comparative Analysis of Techniques

The comparative analysis is based on a test data set, as examined by queries of the database. The test data set uses Washington, D.C. area government agencies' emergency response functions (i.e., fire, rescue, police, etc.) for context.

The test data set describes a set of notional emergency response functions performed by a variety of local government organizations. To build the data set, the team researched web pages to determine the actual organizational structures of several local government agencies, including the State of Virginia, Fairfax County, VA, and Arlington, VA[i]. Several organizations that would likely have emergency response responsibilities were selected, and a database populated with notional information regarding the emergency functions that these organizations would likely perform, other organizations with whom they would have to work to accomplish the identified function, and the means of communications they would likely need with each of these organizations. Three possible means of communication were identified for the test case:

- **Audio**. Simple voice communications such as landline telephones, ordinary cell phones, or radio.
- **Video**. Video teleconferencing communications to enable face-to-face meetings and expand upon audio communications to enable sharing of simple graphics.
- **Data**. Electronic exchange of digital media to include email with attachments and file transfers.

For example, the team postulated that the Fairfax County Emergency Fire and Rescue organization would need to perform the function *Contain and extinguish Fairfax County fires*. For this function a number of *partner organizations* were identified with whom the Fairfax County Emergency Fire and Rescue organization would likely need to work, such as the Fairfax County Office of Emergency Management, the Virginia Department of Emergency Management, and the Fairfax County Police Department. The team also postulated that among these partners, only audio communications would be needed.

The team repeated the process for several government agencies within Fairfax and Arlington Counties in Virginia and for selected state-level government agencies. The data was entered into a relational database against which queries could be conducted to determine which organizations had identified which others as partners and what means of communications each of these organizations believed they needed with each of their partners. To demonstrate the utility of the analytical technique in finding anomalies, the team intentionally excluded several organizations and relationships that would ordinarily exist.

Figure 6 presents a typical tabular presentation of the contents of the database for the hypothetical set of organizations for which test data was entered. This particular tabular view is an NxN table that shows the same set of organizations down the side as across the top. Each cell in the table presents the required means of communications (a = audio, v = video, d = data) between the pairs of organizations that form the cell intersection. The data was compiled and presented from the perspective of the organizations listed in the rows. For example, as shown by the symbols “a,d” at the intersection of the 4th row and 5th column of the table, hypothetically, the Fairfax County Office of Emergency Management has expressed a need for audio and data communications with the Fairfax County Fire and Rescue Department. However, as shown by the symbol “a” at the intersection of the 5th row and 4th column of the table, the Fairfax County Fire and Rescue Department only believes that it needs audio for its communications with the Office of Emergency Management of Fairfax. Similarly, while the Arlington County Fire Department has indicated no need for any communications with the Arlington County Sheriff (blank cell at the intersection of row 9 and column 11), the Arlington County Sheriff has indicated a need to have audio communications with the Arlington County Fire Department (“a” at the intersection of row 11 and column 9).

Notional Sample of Required Organizational Partnerships and Communications Needs

Organizations and Their Partners	Governor of Virginia	Virginia Department of Emergency Management	County Executive (Fairfax)	Office of Emergency Management (Fairfax)	Fire and Rescue Department (Fairfax)	Police Department (Fairfax)	County Manager (Arlington)	Office of Emergency Management (Arlington)	Fire Department (Arlington)	Police Department (Arlington)	Sheriff (Arlington)
Governor of Virginia		a,v,d	a,v,d				a,v				
Virginia Department of Emergency Management	a,v,d			a,v,d				a,v,d			
County Executive (Fairfax)	a,v,d			a,v							
Office of Emergency Management (Fairfax)		a,v,d	a,d		a,d	a,d		a			
Fire and Rescue Department (Fairfax)				a				a		a	
Police Department (Fairfax)				a						a	
County Manager (Arlington)	a							a,d	a,d	a,d	
Office of Emergency Management (Arlington)		a,v,d		a,d			a,v,d		a,v,d	a,v,d	
Fire Department (Arlington)					a			a		a	
Police Department (Arlington)						a	a	a,d	a		a
Sheriff (Arlington)									a	a,d	

Key:

- Vertical axis indicates polled organizations.
- Horizontal axis indicates partner organizations identified by polled organizations.
- Letters in cells indicate means of communications identified as needed by polled organization (a=audio, v=video, d=data).

8

Figure 6: Notional Partners and Communications Needs among Organizations

As can be seen from this simple hypothetical example, however, it is difficult to recognize which organizations have a common perspective regarding the need to communicate with each other, which organizations have a common perspective regarding their required mutual needs of communication, or even if there are some organizational relationships that should have been identified, but were not.

Using the notational symbology presented in Figure 4, the same information can be presented visually as depicted in Figure 7. This figure was hand drawn from the data presented in Figure 6 via a laborious process that required careful examination of the contents of the table and was designed to establish a benchmark against which to test the software queries and resulting netViz graphics that facilitate analysis of the data contents via visualizations. The actual netViz-generated graphics are presented in Appendix A. As can be seen from this figure, it is relatively easy to see which organizations have a common perspective with respect to their mutual communications needs (solid lines), which disagree with respect to the means of communication (dashed lines), and which do not share a common perspective with regard to the need to communicate with each other at all (dotted lines).

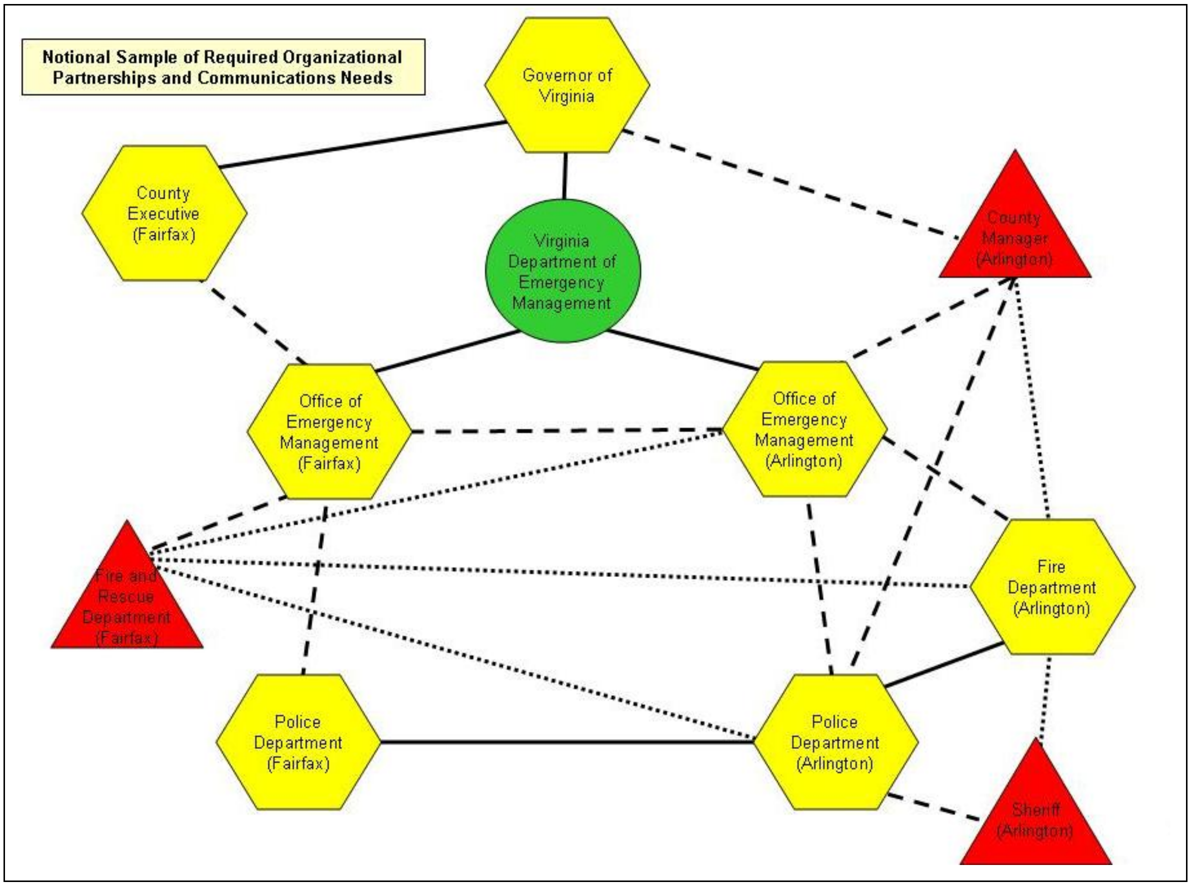


Figure 7: Visual Presentation of Communications Needs among Organizations

While database queries can be written to list these inconsistencies in tabular or narrative reports, the authors believe that visual presentations portray the information in a more readily comprehendible way, in particular for its use by analysts and managers.

In addition, the visual presentations can highlight inconsistencies in the data that the human mind can immediately detect but would be difficult to detect using database queries. In Figure 7, the organizations have been laid out with one county's organizations on the left and the other county's functional counterparts on the right, such that differences between the two become readily apparent. For example, the blue circles in Figure 8 highlight information that is presented on one side of the graphic but not the other. There appear to be major differences between the information captured for one county's (notional) organizations than for the other. This could be due to incomplete data gathering, inconsistencies in organizational perspectives, or differences in organizational structures and relationships. The visualizations make these differences readily apparent and prompt the analyst to probe further into the data.

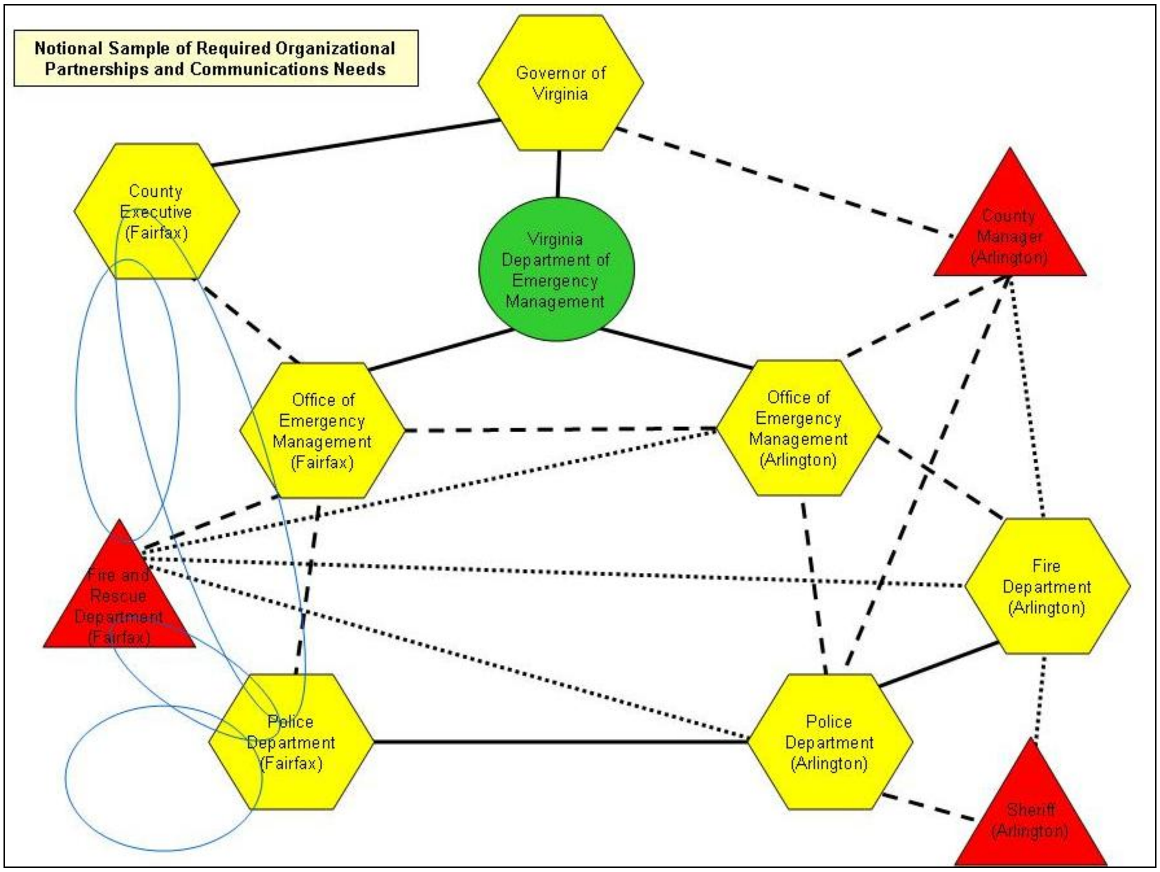


Figure 8: Demonstration of Advantages of Visualization Layout

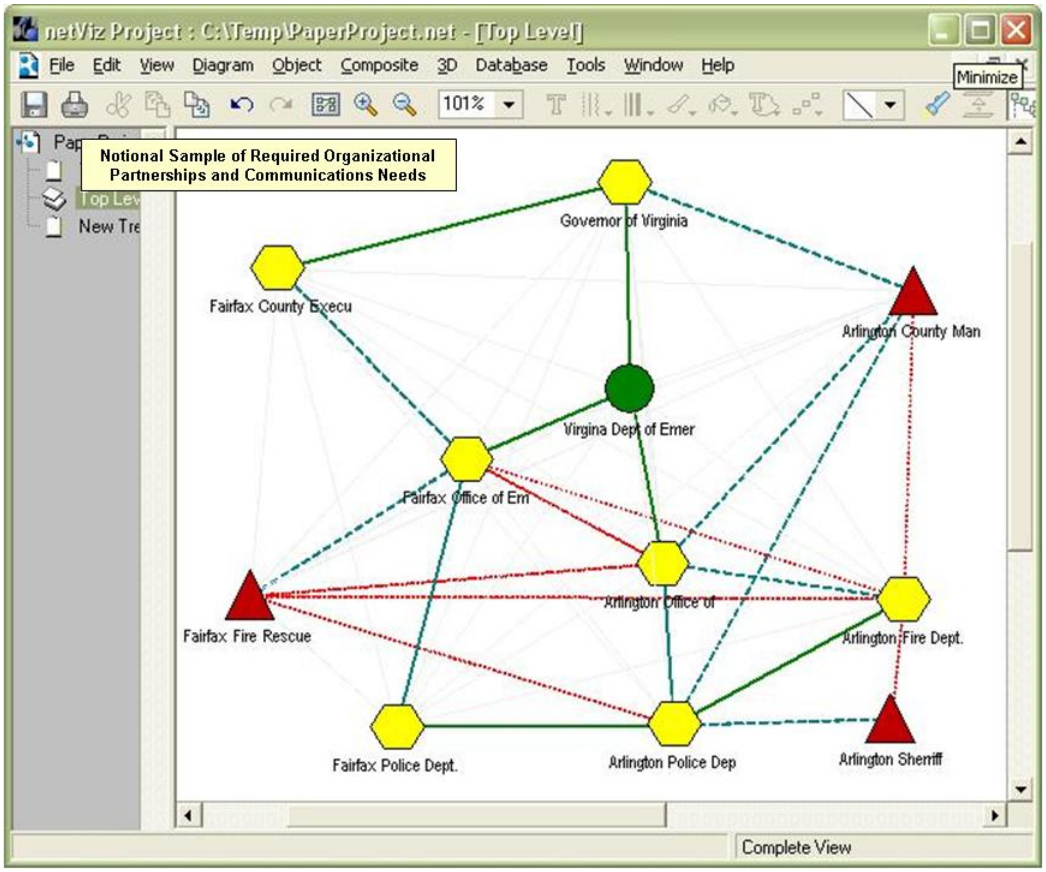
Results and Conclusions

From the fictitious example constructed for this paper, the authors have concluded that graphical visualizations of the data can significantly reduce the time needed to comprehend the results of analysis of the complicated relationships that are characteristic of complex endeavors. Even for this simple example that contains only a handful of organizations, there were relationships (or absences thereof) that would have been next to impossible to detect without presenting graphics that were directly derived from the data. For a real example, say all of the counties and emergency response organizations within a large metropolitan area, the analytical problem would become almost intractable. However, the ability to model the elements of the architecture in a relational database coupled with an automated query capability and visualization tool to generate the graphics directly from the data enables the complexities of such information system architectures to be more readily studied.

Recommendations for Follow-on Research

Future work in the use of database-driven analysis coupled with automated visualizations is being pursued with respect to identifying gaps between "communications needs" as described in the sample presented herein, and “capabilities” to communicate, as can be captured with respect to the existing means of communications that exist among participating organizations. The results of these analyses and how they can support operational planning and investment decisions will be the subject of additional future study.

Appendix A: Sample netViz Results



References

Alberts, D.S. and R.E. Hayes, 2007. *Planning: Complex Endeavors*. Washington, DC:CCRP

DoD Architecture Framework, Version 1.0, Volume 1: Definitions and Guidelines. 2004. Downloaded from http://www.defenselink.mil/nii/doc/DodAF_v1_Volume_1.pdf.

Lee, K., S. Cane, S.Abdul-Rauf, and C. Martinez. 2007. An Architectural Approach for Command, Coordination, and Communications Capabilities Assessment. Paper presented at Command and Control Research and Technology Symposium, June 19-21; in Newport, RI.

Martinez, C.E., K. Mullins, and K.S. Sullivan. 2006. A Framework for Architecture Based Planning and Assessment to Support Modeling and Simulation of Network Centric Command and Control, Paper presented at Command and Control Research and Technology Symposium, June 20-22; in San Diego, CA.

Morganwalp, J.M., and A.P. Sage. 2004. Enterprise Architecture Measures of Effectiveness. *International Journal of Technology, Policy and Management,* 4(1): 81-94.

Rico, D.F. 2006. A Framework for Measuring ROI of Enterprise Architecture. *Journal of Organizational and End User Computing*, 18(2): i-xii.

Spewak, S.H., and S.C. Hill. 1992. *Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technol*ogy. New York, N.Y.:John Wiley and Sons, Inc.

Zachman, J.A. 1987. A Framework for Information Systems Architecture. *IBM Systems Journal*, 26(3): 276-292.

---

[i] Organizational names were selected from the following web pages: the State of Virginia at http://www.virginia.gov/cmsportal2/government_4096/branches_of_state_government_4097/executive_branch.html, Fairfax County, VA at http://www.fairfaxcounty.gov/government/departments/, and Arlington, VA at http://www.arlingtonva.us/Portals/Departments/DepartmentsPortal.aspx