Architecting Principles for Systems-of-Systems

Mark W. Maier, Ph.D.
EB-267
University of Alabama in Huntsville
Huntsville, AL 35899

Abstract. While there is growing recognition of the importance of "Systems-of-systems," there is little agreement on just what they are or on by what principles they should be constructed. This paper proposes a taxonomy of these super-systems and exhibits a basic set of architecting principles to assist in their design. While several heuristics are particularly applicable to systems-of-systems, the key insight is the central role played by communication standards. The enabling architecture of systems-of-systems is non-physical, it is set of standards that allow meaningful communication among the components. This is illustrated through existing and proposed systems.

SUMMARY

While the term "system-of-systems" has no clear and accepted definition, the phenomena is widespread and generally recognized. There is an emergent class of systems which are built from components which are large scale systems in their own right. Prominent examples include integrated air defense networks, the Internet, intelligent transport systems, and enterprise information networks. What factors set these systems-of-systems apart from large and complex, but monolithic, systems? Does the process of architecting and/or developing these systems differ in important ways from other types of systems?

Systems-of-systems should be distinguished from large but monolithic systems by the independence of their components, their evolutionary nature, emergent behaviors, and a geographic extent that limits the interaction of their components to information exchange. Within these properties are further subdivisions. For example a distinction between systems which are organized and managed to express particular functions, and those in which desired behaviors must emerge through voluntary and collaborative interaction.

The independence and extent of these super-systems results in an even greater emphasis on interface design than in traditional system architecting and engineering. Since the components are often developed independently of the supersystem, the supersystem emerges only through the interaction of the components. The system-of-systems architect must express an overall structure largely (or even wholly) through the specification of communication standards.

Systems-of-systems are defined by communication standards. Different problems require standards at different levels. Some applications, an intelligent transport example stands out, require a unique standard built from physical transmission up. As computer-to-computer communication becomes ubiquitous, however, the standards that enable each particular system-of-systems will be high level standards, operating above the transport layer, that define the semantic content of messages passed among the components.

DISCUSSION

This analysis of systems-of-systems architecting divides into three parts. The first part sets out characteristics which distinguish systems-of-systems from large but monolithic systems, a simple taxonomy. The second sets out architectural principles for the construction of systems-of-systems. Since systems-of-systems are built from communications, the nature and type of relevant communication standards forms the third part. Integrated air defense, the Internet, and intelligent transport systems form examples used throughput to illustrate principles and the role and types of communication standards.

What is a "System of Systems?" While the term "System of Systems" is now often used, no clear definition of such super-systems is accepted. What distinguishes a system-of-systems from other large systems? In a formal sense anything can be regarded as a system-of-systems. A personal computer is a system. Its disk drive, its video monitor, its processor, and so forth are also systems. So is a commercial personal computer a system-of-systems in the same sense as a continental air defense network? Intuitively, it is not.

Five principal characteristics are useful in distinguishing very large and complex but monolithic systems from true systems-of-systems.

1. Operational Independence of the Elements: If the system-of-systems is disassembled into its component systems the component systems must be able to usefully operate independently. The system-of-systems is composed of systems which are independent and useful in their own right.
2. Managerial Independence of the Elements: The component systems not only *can* operate independently, they *do* operate independently. The component systems are separately acquired and integrated but maintain a continuing operational existence independent of the system-of-systems.
3. Evolutionary Development: The system-of-systems does not appear fully formed. Its development and existence is evolutionary with functions and purposes added, removed, and modified with experience.
4. Emergent Behavior: The system performs functions and carries out purposes that do not reside in any component system. These behaviors are emergent properties of the entire system-of-systems and cannot be localized to any component system. The principal purposes of the systems-of-systems are fulfilled by these behaviors.
5. Geographic Distribution: The geographic extent of the component systems is large. Large is a nebulous and relative concept as communication capabilities increase, but at a minimum it means that the components can readily exchange only information and not substantial quantities of mass or energy.

Aaron Shenhar proposed a system classification matrix that includes many of these notions. An “array” system in Shenhar’s classification [Shenhar94] is defined as:

> A large widespread collection or network of systems functioning together to achieve a common purpose.

While this definition is very close to that given for a system-of-systems, there are distinctions it may be useful to draw. Some large array type systems may exist in which the components perform no purposes related to the overall purpose of the super-system. For example, in a highly distributed sensor surveillance network it may be impossible for any of the individual sensor nodes to perform a useful surveillance function. The computers and components that make up the network do something when disconnected from the supersystem, but that remaining function is distinct from that of the supersystem. In contrast, many components of an integrated air defense system can still perform an air defense function when disconnected, albeit at a reduced level.

A distinction not discussed by Shenhar is how the system elements of an “array” are developed and procured. In his discussion of management techniques across the spectrum of systems, reference is made to arrays being managed by central program offices that coordinate the elements through formal programmatic management, but do not exercise detailed technical control. In contrast, the notions of [Rechtin91] would suggest the reverse, central technical control of a framework or architecture with loose programmatic control. In any case, no distinctive management style for such distributed systems is used. The nature of such systems has resulted in management styles with considerable variation. Some systems-of-systems are procured and managed centrally like conventional systems. Others have no central authority, they seem to emerge from a web of voluntary actions carried out within a framework.

Virtual, Voluntary, and Directed Systems. Not all systems-of-systems of similar complexity and extent should be regarded as equivalent. An additional dimension, that of managerial control, is critical to identifying appropriate principles. There appear to be three basic categories of system-of-system, at least when classified by managerial control.

Directed: Directed systems are those in which the integrated system-of-systems is built and managed to fulfill specific purposes. It is centrally managed during long term operation to continue to fulfill those purposes, and any new ones the system owners may wish to address. The component systems maintain an ability to operate independently, but their normal operational mode is subordinated to the central managed purpose. For example, an integrated air defense network is usually centrally managed to defend a region against enemy systems, although its component systems may operate independently.

Collaborative: Collaborative systems are distinct from directed systems in that the central management organization does not have coercive power to run the system. The component systems must, more or less, voluntarily collaborate to fulfill the agreed upon central purposes. The Internet is a collaborative system. The Internet Engineering Task Force works out standards, but has no power to enforce them. Agreements among the central players on service provision and rejection provide what enforcement mechanism there is to maintain standards. The Internet began as a directed system, controlled by the US Advanced Research Projects Agency, to share computer resources. Over time it has evolved from central control through unplanned collaborative mechanisms.

Virtual: Virtual systems lack a central management authority. Indeed, they lack a centrally agreed upon purpose for the system-of-systems. Large scale behavior emerges, and may be desirable, but the supersystem must rely upon relatively invisible mechanisms to maintain it.

A virtual system may be deliberate or accidental. Familiar examples of what is called here a virtual system are the World Wide Web and national econcomies. Both "systems" are distributed physically and managerially. The World Wide Web is even more distributed than the Internet in that no agency ever exerted real central control. Control has been exerted only through the publication of standards for resource naming, navigation, and document structure. Web sites choose to obey the standards or not at their own discretion. The system is controlled by the forces that make cooperation and compliance to the core standards. The standards do not evolve in a controlled way, rather they emerge from the market success of various innovators.

National economies and the social "systems" that surround us might be thought of as virtual systems. Politicians regularly try to architect these systems, sometimes through forceful means, but the long-term nature is determined by highly distributed, partially invisible mechanisms.

Introduction to the Examples. Given the fundamentally nebulous nature of the subject, concrete examples are necessary to see principles in action for systems-of-systems. Three examples of existing and emergent systems should serve to span the range of size and management direction.

Integrated Air Defense: The air defenses of modern military forces are clear examples of systems-of-systems. An integrated air defense system is composed of a geographically dispersed network of semi-autonomous elements. These include surveillance radars, passive surveillance systems, missile launch batteries, missile tracking and control sites, airborne surveillance and tracking radars, fighter aircraft, and anti-aircraft artillery. All units are tied together by a communications network with command and control applied at local, regional, and national centers.

When operating as an integrated system, the network exhibits network wide emergent behavior. For example, optimized missile firing and engagement strategies and selective radar use to make targeting of individual elements difficult. However, the uncertainties of warfare make it essential that the system be able to effectively fall back to less integrated configurations, and to make such transitions suddenly and in the heat of battle.

The Internet: The Internet, the global computer-to-computer communications network, is an example of a collaborative system-of-systems. Its elements are themselves computer networks and major computer sites. Some of these component networks may also be composed of further subnetworks. Internet component sites collaboratively exchange information using documented protocols. Protocol adherence is largely voluntary with no central authority with coercive power. Coercive power emerges through agreements among major sites to block traffic and sites observed to misbehave.

The Internet exhibits a rich set of emergent behaviors represented by the complex distributed applications that run on top of the communication substrate. The most complex of these is the World Wide Web, itself a virtual or collaborative system-of-systems that exists solely at protocol layers above the basic enabling protocols of the Internet which is its home.

Intelligent Transport Systems: Intelligent Transport Systems (ITS) covers a wide range of potential applications of information and computer technology to road and transport networks. These run the gamut from improved public service vehicle communication to automated highways with robotically driven cars. As an example here consider only the portions of ITS generally known as Advanced Traveler Information Services (ATIS) and Advanced Traffic Control Systems (ATCS) and their fusion [IVHS America92].

The goal of ATIS is to provide real-time information on traffic conditions and transportation options to travelers in any location. ATIS systems could allow a traveler to scan traffic conditions and choose the transportation mode with predicted least travel time. They could also allow a driver to get real-time traffic state and adapt his driving route accordingly.

The goal of ATCS is to allow a wide range of traffic control methods to be applied across metropolitan areas using strategies optimized from the information available. The information to be used would include real-time and predictive estimates of link times throughout the traffic network, and could include real-time statistics on driver start-destination points and planned route.

ATIS/ATCS fusion yields a very large, collaborative system-of-systems. Fusing the components principally requires communication standards to allow interpretable data exchange. Building a fused system that works will require understanding the incentives needed for collaboration.

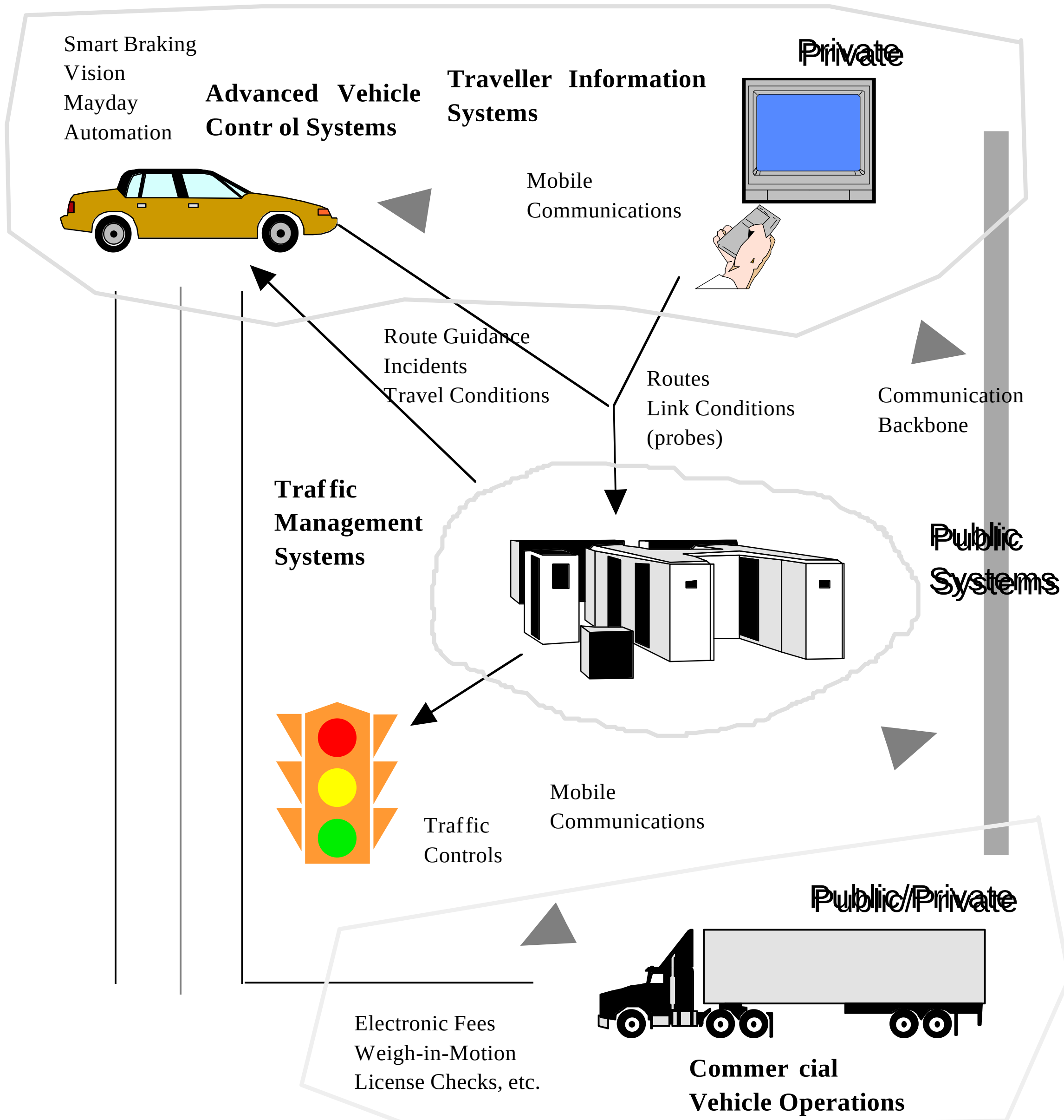

Figure 1: General Structure of ITS showing technical and possible administrative boundaries.

The structure or architecture of ITS is sketched in figure 1. Loose boundaries have been drawn on the figure to emphasize that portions of the overall system are broken across administrative and political as well as technical boundaries. The purchase of vehicles with advanced intelligence will probably continue to be a primarily private transaction between individuals and corporations. The provision of data on destinations, positions, routes and traffic state by private vehicles will probably be voluntary. On the other hand, highway network control will probably continue to be a public responsibility managed by politically chosen organizations. For the overall system to work well, not only must the technical components interface successfully, but the broader interaction of private choice and public policy must do as well, and it must do so compatibly with the technological architecture.

Architectural Principles. No system-of-systems is possible without communication. Since the elements of a system-of-systems are substantially independent, and cannot exchange significant matter and energy, they collaborate only through information exchange.

Models for integrated systems can be used to describe systems-of-systems [Maier96], but do not provide guidance for their structuring. So while the structure is driven by communications, which will be taken up in detail later in the paper, it is first necessary to explore the organizing principles for systems-of-systems. Some literature suggests the use of heuristics as structuring guides in these ill-formed situations. All the heuristics here are taken from [Rechtin91] and refined [Maier94] as appropriate to the circumstance.

Stable Intermediate Forms

The heuristic on stable intermediate forms is given in [Rechtin91] as:

> Complex systems will develop and evolve within an overall architecture much more rapidly if there are stable intermediate forms than if there are not.

The need for stable intermediate forms during a system-of-systems evolution, and during its operation, should be obvious. It is not a monolithic system. It is assumed to be evolutionary and is often collaborative. In a collaborative system-of-systems it cannot be assured that all participants will continuously collaborate, and evolution based on new self-assessments of their objectives for collaboration should be assumed. Even directed systems-of-systems, like an integrated air defense system, may be exposed to sudden (and violent) “reconfiguration.”

Stability is a somewhat complicated concept since several types of large scale stability and instability exist. Stability of intermediate forms means the intermediate forms can “stand on their own.” In the original usage in civil architecture this means buildings should be physically self supporting at intermediate stages of their construction. Taken more generally on systems, this means intermediate systems should be capable of operating and fulfilling useful purposes before full deployment or construction is achieved. An even more general interpretation is that intermediate forms in an evolutionary system should be technically, economically, and politically self-supporting. It should be possible to build and operate the intermediate forms within the economic and political framework of the planned full system.

All three examples show this heuristic at work. Integrated air defense systems are designed with numerous fall back modes, down to the anti-aircraft gunner working on his own with a pair of binoculars. The Internet allows components nodes to attach and detach at will. A still existing subset of the net (the UNIX-to-UNIX Copy Protocol (UUCP) system) is based on intermittent telephone modem connections among its members. The ITS will be deployed piecemeal and unevenly based on the preferences of local and state governments and the willingness of the public to invest in in-car systems. At least in the United States, a monolithic ITS with a distinct startup date is impractical.

Policy Triage

This heuristic gives guidance in selecting and supporting components for a system-of-systems. It is given in [Rechtin91] as:

> The triage: Let the dying die. Ignore those who will recover on their own. And treat only those who would die without help.

The lesson for the system-of-systems architect is to assess the component elements and possible modes of interaction and collaboration. When the system is directed, components must be selected based on the directing authorities power to control the development of component systems. Collaborative systems imply very little power in a directing authority, perhaps only the authority to higher an architect and publish a “vision” document or standard. Here the architect must carefully select components and interaction standards that will voluntarily be taken up by the participants. The architect can also attempt to design interaction mechanisms that will reinforce an incentive to collaborate rather than act independently.

The Internet technical oversight group, the IETF, has had to carefully choose its standards. It has had to avoid putting large efforts into developing standards or extensions that could be implemented only if a central authority financed or dictated their use. Their approach has been to try to validate and standardize those approaches which have developed a consensus through use, and proactively establish standards that would then be the least cost option in emerging function areas.

Leverage at the Interfaces

Two heuristics, here combined, discuss the power of the interfaces.

> The greatest leverage in system architecting is at the interfaces. The greatest dangers are also at the interfaces.

When the components of a system-of-systems are highly independent, operationally and managerially, the architecture of the system-of-systems *is* the interfaces. There is nothing else to architect. The Internet *is* the interfaces, in this case the Internet Protocol (IP). An integrated air defense system, once you extract the sensors and weapons as independent elements, *is* the command, control, and communications network.

Ensuring Cooperation

In a collaborative or virtual system-of-systems how are the components induced to collaborate? The simplest economist argument is that the costs and benefits of collaboration should be superior to the costs and benefits of independence. The Internet maintains this condition because the cost of collaboration is relatively low (using compliant equipment and following addressing rules) and the benefits are high (access to the backbone networks). Similar principles should be designed into other systems where collaboration is required.

Primacy of Communications

Communications is the principle enabling technology for systems-of-systems. The inability to exchange mass and energy implies that only information can be exchanged. Emergent properties can appear only if information exchange is sufficient. If elements are procured semi-independently the standards of communication are more important than any particular systems. Since communications are the principal substrate of systems-of-systems, we take up the nature of communications and communication standards as an independent topic.

Communications Substrates

What does a communication standard standardize? Understanding communication standards requires an understanding of the layered communication model. The layered model divides the communication process into a process stack, each component of the stack is referred to as a layer. The reference model for communication system layering is the seven layer OSI model [Tannenbaum89]. Following more recent practice, a better model may be to consider five layers.

1. The application layer. The user level application processing.
2. Upper layer(s): Object standards, global naming, standards for semantic content in user to user message passing.
3. Transport: End-point to end-point arbitrary message transfer.
4. Network: End-point to End-point unreliable single packet transfer with an upper bound on packet size. Convergence on IPv6 is likely in the future over a broad range of applications.
5. Physical, Media and Data Link Layers: Point to point data transfer including low level reliability, contention and access control, and modulation issues.

The structure of a five layer model is shown schematically in figure 2. The traditional seven layer model divides the bottom layer into two, and calls out two specific upper layers, the session and presentation layers. In practice the choice of physical and/or data link layer is becoming less significant. A wide variety of bit level transport media are available and it is rarely desirable to design a new one for a specific application. Hence the architecture of communication for a system-of-systems is more likely to concentrate on the layers above bit transfer to focus on the areas unique to the system. Bit transfer will usually be provided by the emerging backbone of wide area communication services.

The upper layer situation is less well defined. Neither of the OSI models two upper intermediate layers (session and presentation) have seen wide implementation. Moreover, the communication abstractions suggested by those layers do not seem to match well the actual structures of computer and convergent communication. Instead a variety of upper intermediate toolkits have appeared. Many of the newer ones are based on object oriented abstractions of inter-program communication or other models of computer-to-computer communication [Next96, OSF96].

A communication standard may encompass any set of layers. Standards typically cover only a single layer, but an integrated set of multi-layer standards may be needed for a particular application. The example systems contain a diversity of cases. Some of the most widely known standards are those for the network and transport layers, such as TCP/IP and IPv6 [Comer95], SPX/IPX and AppleTalk [Sidhu90]. Asynchronous Transfer Mode (ATM) largely fits into this category.

Network Layer: The Internet

The Internet is an existing system-of-systems, and it is built on standards. In particular, it is built on a critical standard, the Internet Protocol (IP). This standard—several standards, actually—defines the structure of a data packet, globally routable addressing, routing methods, and an inter-node control protocol. In addition, standards exist for mapping IP onto various lower layers for the diversity of physical interconnections, and for various higher layers that operate on top of IP. While IP itself is all that is necessary for Internet data exchange to exist, useful applications, the emergent behaviors observed by the users, require higher layer standards from which the emergent applications are built.

Physical Layer Up: ITS Beacon System

In a system as diverse as the ITS it is no surprise to find a very wide diversity of communication systems and standards. In one particular case there is a potential requirement for an ITS unique communication system that would include standards from the physical layer up. This case is the short range vehicle beacon system. The beacon system is envisioned as an infrared (or possibly microwave) based system that will communicate between vehicles and roadside transceivers over distances of a few meters to tens of meters. This system has several important attributes:

1. Because it is short range and directed it provides an enormous aggregate bandwidth for communications over the entire vehicle population of a metropolitan area. It could support near real-time independent interaction with every vehicle if so desired.
2. Since roadside beacon locations are known it combines communication and position determination and reporting. It does so at low cost since it combines the functions and the unit itself is potentially very low cost.
3. It could be used for inter-vehicle communication and cooperative sensing as well, paving the way for automated highway operation.

All of these attributes require a standard for the system from the physical layer up. Commonalty over a nations road network is required in all of modulation, wavelength, data linking, and message content.

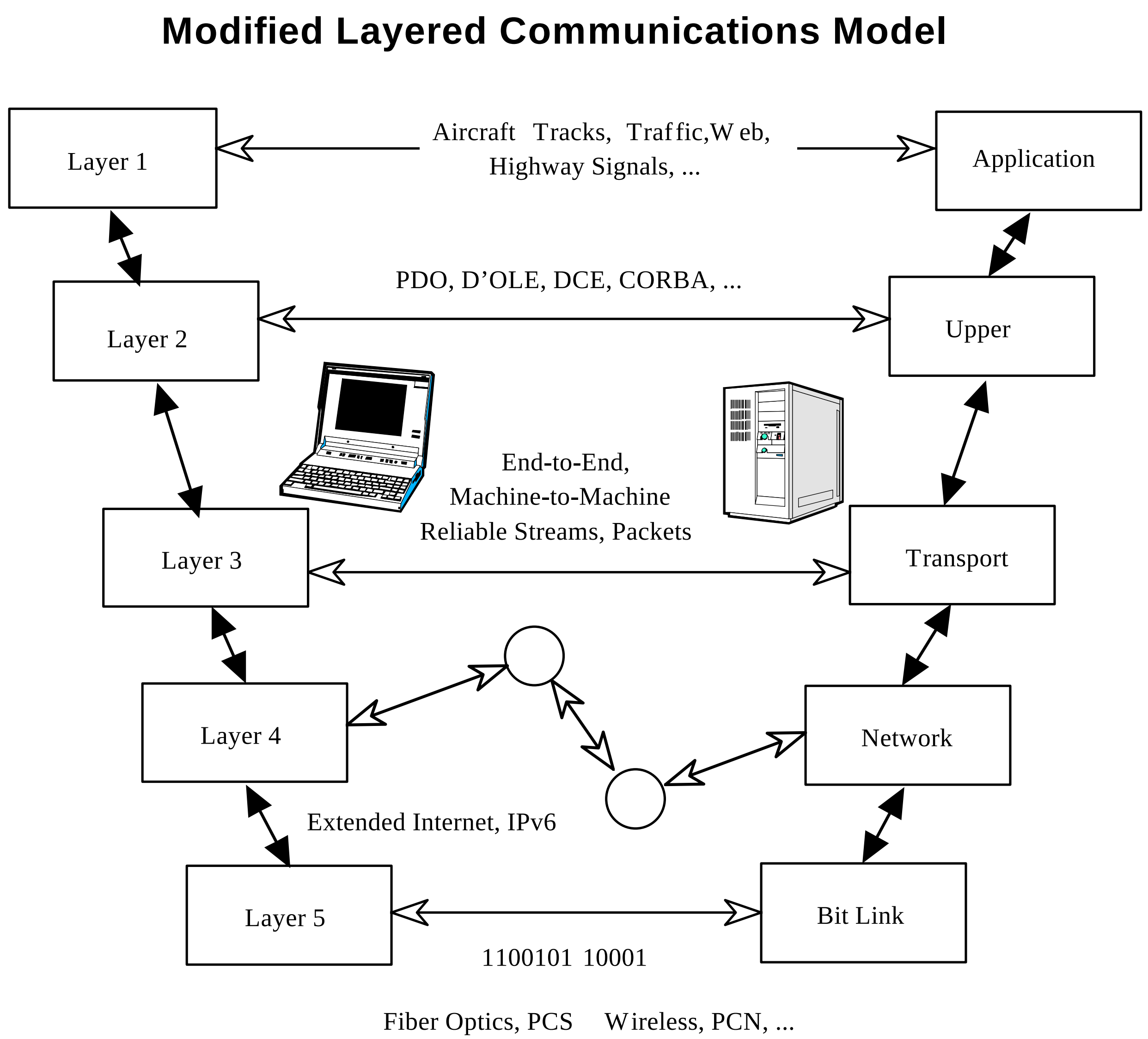

Figure 2: A Five Layer Conceptual Model of Communications. The five layers emphasize the ubiquity of physical links, and the emerging importance of intermediate structures between reliable data transport and applications.

Upper Layers: ITS Information Exchange

ATIS and ATCS fusion in ITS does not require a beacon system. It does require a set of standards at higher layers. If one assumes that Internet like communications will be ubiquitous for computer-to-computer nodes in the ITS time frame (the next twenty years), several not-now-existing standards are relevant to ITS.

1. Geographic referencing. When a message says "I'm going from here to there," how do we define here and there so all receivers understand? This type of messaging requires a standard means of reporting location and correlation to maps.
2. Traffic message content. The ITS needs messages that report traffic state. Such messages must include location (already discussed above), but must also include state. In general the ITS needs a standard for message types and contents that map to underlying transport mechanisms.

The main emerging standards for upper intermediate layers are in computer-to-computer communication. Among well known are examples are PDO, new versions of Microsoft's OLE, and DCE.

Eclectic: Integrated Air Defense

Modern military systems often exhibit an eclectic diversity of communication systems and standards. Because of the long life of most military systems, old communication interfaces may be maintained long after their technology has become obsolete. This diversity of communications is often a burden to the military architect. The existence of this burden is a testimony to the importance of elegant and insightful communication standards in systems-of-systems.

The response of communication engineers is to emphasize layering. An insightful standard on one layer can fruitfully live on long past the obsolescence of the physical layers on which it originally ran. Equally important is the need to design in evolution. For example, the IP standard uses a version numbering system to allow packets from different versions of IP to coexist on the Internet. This ability has been critically important in allowing the Internet to evolve, and will be used again in the next few years during the evolution to IPv6 from the current IPv4.

CONCLUSIONS

Systems-of-systems are the children of modern communications and computing. They will exist more widely in the future as individual systems become "smarter" and communication interfaces become routine. The nature of the communication standards that enable and define individual systems-of-systems should shift from physical-up standard to higher layer standards that assume the existence of an IP-like data transport substrate.

Collaborative and virtual systems-of-systems will also become more common with the ubiquity of smart systems independently operated and managed. This will place a premium on the discovery and clever use of design principles that produce emergent behavior through voluntary collaboration. A fruitful area for such work may be in the use of pseudo-economic mechanisms.

REFERENCES

Comer, D. E., Internetworking with TCP/IP, Vol. 1, 3rd edition, Prentice Hall, 1995.

IVHS America, Strategic Plan for Intelligent Vehicle-Highway Systems in the United States, IVHS America, Report No. IVHS-AMER-92-3, Intelligent Vehicle-Highway Society of America, Washington, D.C., 1992.

Maier, M. W., Heuristic Extrapolation in System Architecture, in *Proceedings of the 4th International Symposium of the National Council on System Engineering*, NCOSE, Vol. 1, pp. 525-532, 1994.

Maier, M. W., Integrated Modeling: A Unified Approach to System Engineering, *Journal of Systems and Software*, 32:2, February. 96.

NeXT Computer, White Paper: Architecting for Change with Enterprise Objects Framework and Portable Distributed Objects, NeXT Computer White Paper.

Open Software Foundation, DCE Overview, OSF-DCE-PD-1090-4, 1996.

Rechtin, E, *System Architecting: Creating and Building Complex Systems*, Prentice Hall, 1991.

Sage, A., Methodology for Large Scale Systems (New York, NY: McGraw-Hill, 1977).

Shenhar, A., A New Systems Engineering Taxonomy, in *Proceedings of the 4th International Symposium of the National Council on System Engineering*, National Council on System Engineering, Vol. 2, pp. 261-276, 1994.

Sidhu, G. S, Inside AppleTalk, 2nd Edition, Addison Wesley, 1990.

Tannenbaum, A. S., Computer Networks, 2nd Edition, Prentice Hall, Englewood Cliffs, NJ, 1989.

AUTHOR BIOGRAPHY

Mark W. Maier received the B.S. and M.S. degrees from the California Institute of Technology and the Engineer and Ph.D. degrees in Electrical Engineering from the University of Southern California. While at USC he held a Hughes Aircraft Company Doctoral Fellowship, where he was also employed as a section head. Currently he is Assistant Professor of Electrical and Computer Engineering at the University of Alabama in Huntsville. Dr. Maier's research interests are in randomized radar waveforms, data compression, computer based system design, and system architecting for large scale systems.