Communication and Network

Academic and research staff
Professor Vincent W. S. Chan

Visiting scientists and research affiliates
John Chapin, Professor Guevara Noubir (Northeastern University), Professor Ashwin Gumaste (IIT, Mumbai), Steven Finn, Alan Kirby (CISCO), Terrence McGarty, Eric Swanson, Lillian Dai (CISCO), Guy Weichenberg (The RAND Corporation)

Postdoctoral Fellow
Ingrid Lin

Graduate students
Matthew Carey, Anurupa Ganguly, Etty Lee, Katherine Lin, Andrew Puryear, Mia Yinuo Qian, Rui Li, Lei Zhang

Undergraduate students
Katherine Lin, Rui Jin

Technical and support staff
Donna Beaudry

Overview of the research in the Communication and Network Group
Research in the group includes topics on heterogeneous networks and communication systems. The work extends to applications in satellite, wireless and optical communication, and data networks. The objective is to develop the scientific base needed to design data communication networks that are efficient, robust and architecturally clean. Wide-area and local-area networks, high-speed networks, and point-to-point and broadcast communication channels are of concern. Topics of current interest include network architectures at all network layers; power control; multiple antenna techniques; media access control protocols; routing in optical, wireless and satellite networks; quality of service control; failure recovery; topological design; and the use of efficient resource allocation for network connectivity and QoS. An important new research frontier that calls for new modeling, analysis and architecture optimization is the emergence of heterogeneous networks combining fiber, wireless and satellite systems. Heterogeneous networks must provide agile and economical service delivery in the face of challenges that include rapidly changing communication channel quality, link connectivity, and traffic flows. An example of the research frontier is for one internetwork to simultaneously and efficiently support: mobile wireless communication in the presence of fast fading and changing connectivity, microwave satellite communication and free space optical communication affected by rapidly changing atmospheric phenomena, and unscheduled, bursty, large-granularity traffic flows. The problems in this direction are important, rich and very challenging. The theme of heterogeneous networking with performance guarantees is a challenging and important problem for defense and high-end commercial applications. A significant component of our research is the creation of new architectures guided by the understanding of technology and fundamental system limits and the realization and validation of these architectures with hardware and algorithms and system demonstrations. As such, we work closely with industry to extend our research reach and impact beyond the academic boundary.

1. Future Optical Network Architecture

Sponsors
DARPA – FONA: HR0011-08-1-0008
National Science Foundation (NSF) – FIND:CNS-0626800 and GOALI:CNS-0831612

Project Staff and other Participants
Vincent W. S. Chan, Anurupa Ganguly, Ori Gerstel (Cisco Systems), Kyle Guan, Ashwin Gumaste (Indian Institute of Technology), Eric Swanson, Guy Weichenberg, Lei Zhang, Ingrid Lin, Dan Kilper (Alcatel Lucent Bell Labs.)

Optical networking has enabled the Internet as we know it today, and is central to the realization of Network-Centric Warfare in the defense world. Sustained exponential growth in communications bandwidth demand, however, is requiring that innovation in optical networking continue, in order to ensure cost-effective communications in the future.

Cost optimized optical network architecture. In the current telecom environment, carriers have deployed huge capacity in their long-haul networks. In the meantime, end user access to higher data rates is still costly. Bridging the gap between the bandwidth glut at the backbone and the high access cost for the end-users requires architectural innovation in next-generation optical MANs (metropolitan area networks). The objective is to design networks that combine low installation cost with excellent scalability—a decreasing cost-per-node-per-unit-traffic as the number of users and the transaction size increase. Scalability is essential for any commercially deployed network to attract serious providers and investors to commit to the venture as part of a sensible business.

The central result of our research is to identify scalable network architectures that are consistent with the limitations and tradeoffs of optical networking technology: fiber connection topologies, switching technology selection and dimensioning, as well as routing and wavelength assignment. Considering all these factors, we emphasize optimization of network characteristics over possible fiber connection topologies and the interaction between cost and architecture tradeoffs. This optimization problem is intrinsically complex. To get traction, the first part of the research takes an analytical approach by concentrating on networks with regular topologies and (static and random) uniform all-to-all traffic. These assumptions are idealizations; nevertheless, the analyses provide insights into more complex problems and act as points of departure for the analysis of more realistic scenarios (such as irregular networks and non-uniform traffic) in the later part of this research.

The search for a scalable fiber MAN architecture hinges on analyzing the tradeoffs among expensive network resources. In our parametric, first-order, and homogeneous cost model, the constituent parts, which are closely related to fiber topology, are fiber cost and switching cost. To build a network, one can support lightpaths by deploying direct fiber connections among all source-destination nodes. This design obviously requires minimal switching resources but a maximal amount of fibers. Another way to establish lightpaths is by hopping through one or more nodes. Such a design requires less fiber connections but more switching resources. As such, the optimal connectivity of a fiber connection topology is determined by the fiber-to-switching cost ratio. Further, we show that the amount of switching resources used at nodes is proportional to the average minimum hop distance (for regular topologies and uniform traffic). To support the same set of (uniform) demands, we show that regular topologies with the smallest average minimum hop distances have lowest fraction of pass-through traffic and thus require less switching ports.

We have developed guidelines for rigorous studies of cost-effective network architectures. For a few representative classes of regular networks (e.g., Δ-nearest Neighbors and Generalized Moore Graphs), we first derive or approximate the closed-form expressions for important

parameters, such as average minimum hop distance, switch size, and network cost. We then set up the corresponding optimization formulations. We have found that for regular networks and uniform traffic, the joint design problems of fiber connection topology, dimensioning, and routing can be solved optimally and analytically for a special class of regular graphs — Generalized Moore Graphs.[1] That is, we prove that with minimum hop routing, Generalized Moore Graphs, whose average minimum hop distances scale favorably as $\log_{\Delta}^{N}$, achieve the lower bound on network cost and are good reference topologies. We also show that topologies with structures close to Generalized Moore Graphs can achieve close-to-minimum cost. The investigation of cost scalability further demonstrates the advantage of Generalized Moore Graphs and their close relatives as benchmark topologies: with linear switching cost, the minimal normalized cost per unit traffic decreases with increasing network size (Figure 1.1). In comparison, for less efficient fiber topologies (e.g., Δ-nearest Neighbors) and switching cost structures (e.g., quadratic cost), the minimal normalized cost per unit traffic plateaus or even increases with increasing network size. Our study has also revealed other attractive properties of Generalized Moore Graphs in conjunction with minimum hop routing. When minimum hop routing is employed for uniform traffic in Generalized Moore Graphs, the aggregate network load is evenly distributed over each fiber. Further, to support a given uniform traffic demand, Generalized Moore Graphs require the minimum (or close to the minimum) number of wavelengths, which directly affects the complexity and the cost of dispersive elements and filters in the network. This is the first time that Generalized Moore Graphs have been identified as optimum architectures in the context of network cost efficiency. These architectures are very different from the currently used ones in MANs, such as rings, interconnected rings, or non-optimized mesh networks.

Given our parameterized network cost analysis, closed-form solutions of the optimal degree and cost as functions of various network design parameters (such as network sizes and wavelengths of traffic between node pairs) are obtained. These results show that for a MAN of moderate size (a few tens to a few hundred nodes), under certain cost structures, neither rings nor fully connected mesh networks are optimal topologies. The optimal network connectivity is in the range of 0.03*N* to 0.1*N* (i.e. each node is connected to 3–10% of the total number of nodes). The advantage of analytical approaches is self-evident: they provide a valuable reference for the scaling of optimal network connectivity as design parameters change. More importantly, the results demonstrate that switching technology has a tremendous impact on the final choice of topology. The optimal topologies connecting the same set of nodes can differ significantly when different switching fabrics are used, even when these topologies are designed to serve the same traffic demand. Among all-optical technologies currently available, for smaller networks (a few to a dozen nodes) and light traffic, quadratic switching cost structures (e.g., 2-D switching fabrics) have a cost advantage. However, as the size of the network and the demand among node pairs increase, linear switching cost structures (e.g., 3-D switching fabrics) have the best scalability. Thus, the cost benefit of deploying 3-D switching technology for the future network is apparent. Moreover, a comparison of the cost benefit between OXC and OEO switches (optical cross connect vs. optical-electronic-optical designs) shows that at low data rate (e.g., below 1Gb/s for every source-destination pair for a 50-node network, as shown in Fig. 1.2), it is economical to use OEO switches; at high data rate (e.g., above 1Gb/s for every source-destination pair for a 50-node network, as shown in Fig. 1.2), it is more cost-advantageous to use OXC switches.

We also took steps to broaden the scope of this work to address more realistic design scenarios.

- We investigated irregular network topologies and (static) non-uniform traffic, which represent most existing networks. We show that if the switching cost is linear with port counts, minimum hop routing is still optimum. The results of Generalized Moore Graphs can be used to provide useful estimates for the cost of irregular networks. A Generalized

[1] The assumptions of uniform traffic and symmetric topology, as well as the special construction of Moore Graphs make this joint problem solvable. For non-uniform traffic and other classes of graphs, the joint problem remains difficult to solve.

Moore Graph with a node degree that equals to the minimal node degree of an irregular network provides lower bound on network cost. Also Generalized Moore Graphs can be exploited to suggest improvements for irregular fiber connection topologies. We show examples that connecting the same set of nodes via a Generalized Moore Graph results in savings on the number of switching ports.[2] For irregular topology under arbitrary traffic, we provide the lower bound on network cost using the concept of minimum flow tree, which is based on the unique construction of Generalized Moore Graphs (each of its nodes has a full or almost full Δ-ary routing spanning tree) and an (idealized, not always realizable) permutation of traffic matrix (the Δ destinations with the largest traffic are connected by one hop paths, the next Δ^2 destinations in descending order of traffic are connected by two-hop paths, and so on). More importantly, the construction of the minimum flow tree, though idealized, yields a general yet crucial design guideline: a cost-effective physical topology should minimize the propagation of large traffic flows.

- We also investigated designing networks that are robust to demand uncertainties, which are caused by diversification of services, changing of usage patterns, and data-dominated traffic in the metro environment, etc. We developed a framework to assist network designers to dimension optical MANs, incorporating uncertainty in demand. In this framework the interplay among topology design, resource provisioning, and routing is analyzed based on two stochastic optimization models that use probability distributions of demands as inputs. In one model, the weighted sum of network installation cost and expected penalty cost for unsatisfied traffic is minimized. In another model, the network installation cost is minimized subject to certain service level requirements. The optimization results enable us: (1) to identify the Generalized Moore Graphs as the physical architectures that are most robust (in cost) to demand uncertainties among rich classes of regular topologies, assuming the (random) uniform all-to-all traffic, and (2) to provide analytical references on how optimal dimensioning, network connectivity, and network costs change as functions of the designer's level of risk aversion, service level requirements, and probability distributions of the demand.

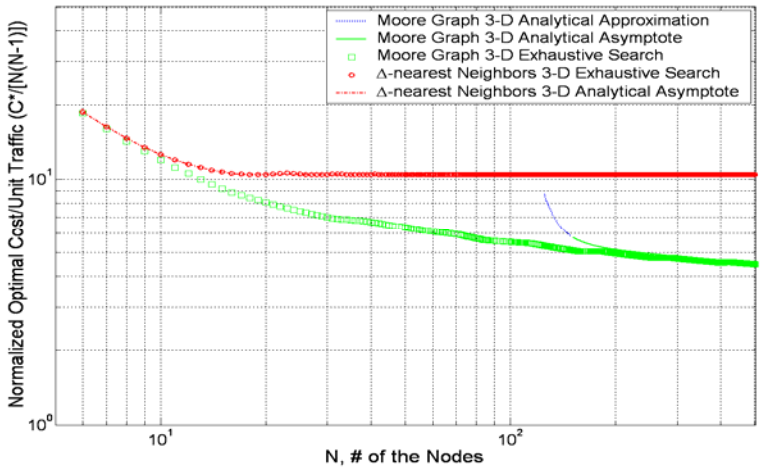

Figure 1.1: Optimal normalized network cost per unit traffic $C^*/[N(N-1)]$ as a function of network size *N* for the Δ -nearest Neighbors and the Moore Graphs. The points represent the results of exhaustive searches; while the lines represent the analytical asymptote. The switching fabric is 3-D with $F_1(K_o) = \beta_1 K_o$. The fiber-to-switching cost ratio α/β_1=40.

[2] In these examples, every improved network uses the same number of fiber connections as the original one does.

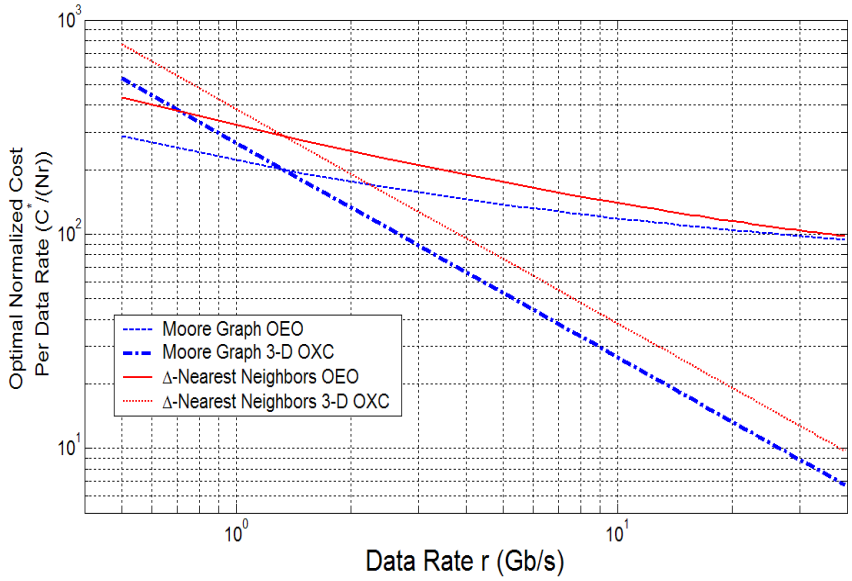

Figure 1.2: Minimal normalized network cost per data rate $C^*/(Nr)$ as a function of data rate per wavelength r for combinations of two classes of network topologies (Δ-Nearest Neighbors and Moore Graphs) and two types of switching fabrics (OEO switch and 3-D OXC). *N*=50, $\alpha = 20$, $\beta_1 = 1$, and $\beta_e = 7.5$. α and β denote the cost per fiber connection and the cost per port for linear switching cost structure, respectively.

In the area of optical network diagnostics, we have established a mathematical mapping between the fault diagnosis problem and the source-coding problem. This mapping has both theoretical and practical implications in efficient fault diagnosis algorithm design. From a theoretical perspective, it suggests that that the minimum average number of probes required is lower bounded by the entropy of the network states. From a practical perspective, it suggests an efficient fault diagnosis algorithm design via translating existing source-coding algorithms under physical probe constraints.

We have also developed a family of novel near-optimal polynomial time diagnostic algorithms for all-optical networks with probabilistic link failures. The algorithms are based on run-length codes and their performance asymptotically approaches the theoretical entropy limit for large networks.

Finally, we extended the application of the run-length probing scheme to practical all-optical networks with both node and link failures. To diagnose probabilistic node/link failures, we introduced a network transformation that converts the original undirected graph into a directed graph: each link in the original graph is replaced by two parallel directed arcs in opposite directions, and each node of degree *d* is replaced by a $d \times d$ directed complete bi-partite graph, where any node in the left column is connected to any node in the right column via a directed arc. Under this transformation, both link and node failures in the undirected graph are mapped into arc failures in the directed graph. Moreover, the directed topology can always be made Eulerian, rendering the run-length probing scheme applicable. Depending on the relative dominance between link failure probability and node failure probability, different probing strategies are obtained through analytical and numerical investigations.

Analytical and numerical investigation have also revealed a guideline for efficient probing schemes: each probe should be designed to provide approximately 1-bit of information on the network state and the number of probes required is approximately equal to the information entropy of the network states.

Optical Flow Switched Core Network Testbed

Optical flow switching has recently emerged as a promising new approach to address the exponentially growing traffic demands in optical networks. In order to realize this new paradigm for optical networks, the core network must be capable of wavelength re-configuration and service provisioning on time scales approaching one second or faster. The present methods used in wavelength division multiplexed (WDM) optical transmission are designed for static wavelength channels, which can increase in number over time and may be configurable to form different paths at the time of provisioning. Provisioning new wavelengths is currently a methodical, one at a time, step-wise process of 'turning up' the optical signal along the path, which requires minutes to hours in order to achieve the final stable state, appropriate for Generalized Multiprotocol Label Switching (GMPLS) networks. Furthermore, the extensive literature on optical transmission assumes this quasi-static network environment. The intent of this project is to determine both the practical and fundamental limitations on optical flow switched networks imposed by physical layer constraints, including optical transmission in a dynamic, wavelength switching environment. This goal is particularly challenging because of the multi-layer nature of the problem. The network topology and signaling protocols will dictate physical layer transmission requirements, such as how fast the optical signal must achieve an error free condition and over what transmission distance. Likewise the physical limitations both in terms of performance and cost will motivate design choices for the architectures and protocols.

Targeted experiments are conducted in a testbed at Bell Laboratories in conjunction with the optical flow switched network optimization studies at MIT. This new testbed has recently been used to study both the power dynamics and the transmission performance of signals in reconfigurable optical networks. These two elements are essential for understanding the physical layer in a transparent switched network and this facility is unique in that it combines these in a broadband, long haul transmission configuration, typical of core networks.

Increasingly, focus is being placed on the anticipated processing burden on IP routers in the face of steady network traffic growth. Many recent efforts to address this problem have centered on augmenting the speed and capacity of routers by using optical technologies [1]. However, the bottleneck in router processing is primarily due to power consumption and heat dissipation requirements in the electronics that perform advanced packet header operations, and it is still not clear how optics will adequately address this problem. Furthermore, an optical solution for complex header computations that is more power efficient than CMOS has not been identified [2,3]. An alternative to simple scaling of IP router technology from electronics to optics is to re-design the network itself in order to minimize the processing load on individual routers.

Optical flow switching in transparent networks is one such all-optical, end-to-end approach. It relieves routers from serving large transactions that are likely to comprise a substantial portion of future Internet traffic. The use of flow switching is particularly attractive in core networks because the traffic loading is large enough that wavelength granularity resource allocation efficiently handles traffic demands. For user bandwidths above ~200 Mb/s, optical flow switching becomes the more cost effective solution, surpassing GMPLS, which is preferred at low traffic loading [4]. In practice, one would expect that a combination of GMPLS and OFS would be used, in which groomed and sub-wavelength traffic would be handled through GMPLS whereas the large flows (so-called "elephants") generated by high-end applications would be addressed with OFS.

Both lowest cost network solutions, GMPLS and OFS, utilize the efficiency of transparent networking (data delivery through the network entirely in the optical domain without optical-electronic conversion). Today there are still many challenges associated with extracting the full benefit of transparent networks for GMPLS, particularly with regard to the interactions between the physical layer and higher network layers [5-7]. OFS requires even stronger coordination between layers. Traditionally, the complexity of optical transmission technologies has been

hidden from the higher layers because it was largely isolated and separately managed. In OFS, as opposed to IP flow switching [8], the switching is realized in the physical layer and therefore the details of the optical transmission must be visible to the network control system. Efficient utilization of the physical layer requires detailed consideration of the technical challenges in optical transmission. At present, however, this analysis is restricted to physical layer constraints associated with transmission in static networks, because of a lack data on dynamic networks. In many cases, such as short distance access networks, static constraints dominate so that dynamic transmission effects can be safely ignored. Dynamic networks are more challenging in the core, due to longer transmission distances and higher performance requirements.

Transparent and Dynamic Physical Layer Transport

Transparent optical networks have been a subject of study since the advent of WDM optical transmission. Early studies, however, tended to be restricted by the available optical technologies and understanding of optical transmission. The MONET project is a well known example [9]. While the benefits of MONET were arguably far reaching in terms of progressing the state of the art in WDM, the project did not include many of the key characteristics that dictate performance unique to mesh network transport today. The channel count for example was a maximum of 16 channels at 10 Gb/s, which does not reflect the wavelength dependent effects in a 40-80 channel system with mixed bit rates and modulation formats [10]. The widespread use of forward error correction coding has significantly moved the performance targets, leading to new transmission engineering rules [11]. Wavelength selective switches [12] were not available at the time. Transient effects and power control were just being studied for the first time in MONET [13]. Transparent systems today benefit from substantial technological progress as well as advances in simulation methods [14]. The simulator used in the MONET project [15] did not include Raman interactions in the transmission fiber, which have a significant impact on power evolution.

Recent experimental work related to dynamic mesh network transmission generally divides into two categories: (1) short reach or un-amplified network studies of burst or switched systems [16,17] and (2) long reach static transmission experiments with selective addition of components or features representative of dynamic networks [18-21]. The experiments in the first category do not consider the physical phenomena such as non-linear channel interactions mediated by the fiber plant or optical amplifier gain saturation, which are critical to transmission in core networks and are the primary subject of the experimental work. The second category of experiments include transmission studies with ROADM (reconfigurable optical add-drop multiplexer) or photonic cross-connect elements. In these experiments, the core network transmission effects are studied, however, they generally do not address the dynamic aspects. Power dynamics are studied as a separate topic and have primarily focused on large transient effects associated with fiber breaks.

Our experiments in the Bell Laboratories testbed specifically consider the transmission performance of signals in a transparent optical network. Long haul transmission experiments typically involve the use of a circulating loop apparatus to repetitively propagate signals through a particular amplified fiber plant in order to achieve long distances without exorbitant equipment requirements. Unfortunately, the repetitive nature of this method is not well suited for the study of transmission in reconfigurable networks. In our experiment, fast switches (sub-microsecond) are incorporated into the circulating loop in order to change fiber plant characteristics during the signal propagation. Multiple unique mesh paths are realized, through automated system control using this new circulating-loop method, to support more realistic experimentation.

Fundamental Limiting Factors and Testbed Modification

In order to understand limitations associated with dynamic functionality, we need to understand the different physical factors that may impact the optical adjustments required at a given location. In this project, we have identified and provided initial models for many key factors. We start by categorizing them into global and local physical power adjustment constraints. Global factors

involve events and quantities that may be distributed in the network and therefore require global knowledge of the network in order to be monitored or characterized. The global factors include:

1. Re-entrant wavelength paths
2. Diversity of wavelength paths
3. Path diversity of switched wavelengths
4. Overlap of old and new wavelength paths
5. Coupling (common wavelengths) between old and new paths
6. Wavelength transmission histories
7. Overall path length of switched wavelengths

On the other hand, local constraints on power adjustments originating from power-dynamics are due to factors that occur at one location and/or can be characterized by a local measurement. These include:

1. Number of wavelengths present before switch event
2. Number of switched wavelengths
3. Length (distance) of the link
4. Optical amplifier spectral tilt and output power ripple
5. Availability of controls and their characteristics
6. Wavelength placement arrangement (banded, spread, and etc)
7. Channel spacing (50GHz vs 100GHz, etc.)

The above mentioned factors are related to physical power constraints that influence the tuning, but may not directly impact the transmission performance.

Testbed Modifications

A dynamic distributed-distance circulation loop provides rapid wavelength redirection capability for mesh network emulation. Several key modifications to the distributed-distance loop setup were implemented in order to better study transmission in dynamic mesh networks. These include adding a second reconfigurable optical add-drop multiplexer (ROADM) node to the loop to attain the physical implementation as shown in Figure 1.3. Two erbium doped optical amplifiers (OA) are inserted between each ROADM for proper power and spectral tilt adjustment. Two spans of standard single mode fiber (SSMF) ~80 km/span are placed between the ROADMs for long haul operation [22].

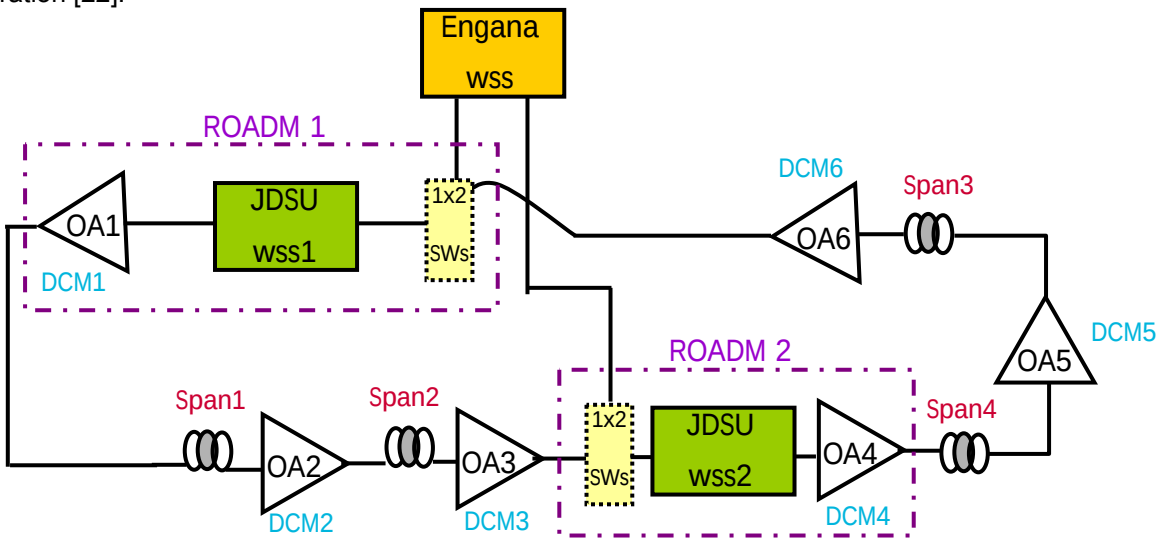

Figure 1.3. Experimental schematics of distributed distance circulation loop with a second ROADM inserted. (Span length ~80Km)

The other key testbed modification is configuring a new dispersion map to reduce distortion coming from the fiber nonlinear response, including self-phase modulation, cross-phase modulation, and four-wave mixing. Such effects give rise to signal intensity distortion during propagation and thus limit the maximum transmission distance in the fiber link. The choice of dispersion compensation module (DCM) in line at the mid stage of each OA facilitates the dispersion map arrangement. Different dispersion maps are used depending on the system design goals. For a preliminary setup, we chose a return-to-zero (RTZ) map for which dispersion pre-compensation at each ROADM and a residual dispersion per span (RDPS) are used, but the total accumulated dispersion is brought back to zero at each ROADM. This map is commonly used for its simplicity, but is known to be more sensitive to non-linear fiber effects than non-RTZ maps. Careful calibration of the launch power into each span was also carried out in this experiment to obtain an accurate estimation of the channel power evolution. Furthermore, we apply different lengths of SSMF and DCF for post-compensation to determine the performance of our dispersion map at various circulation distances.

Our modification to the testbed allows us to add channels within the circulation loop in addition to adding them at the beginning of the loop, include stronger filtering effects as a function of distance, enable more combinations of channel groups, and also provide twice the power equalization capability throughout one round trip. Briefly speaking, the modified configuration offers greater flexibility enabling better emulation of mesh networks.

Measurements and Device Evaluations

Distributed Distance Circulation Loop 80x10Gb/s OSNR

System performance with 80 channels of 10Gb/s wavelength-division multiplexed (WDM) signals was measured using the modified testbed elaborated previously. Both required optical signal-to-noise ratio (ROSNR) at bit error rate 10^{-3} and delivered optical signal-to-noise ratio (DOSNR) at the output of the ROADM are of interest. The difference between ROSNR and DOSNR decides the performance margin of each wavelength at a particular distance.

Measurements of ROSNR and DOSNR are taken for eight different transmission distances, i.e. from 1 round trip to 8 round trips as shown in Figure 1.4. We intentionally divide the eighty channels into two groups and load them at two different locations, one at ROADM 1 and the other at ROADM 2. Therefore two groups of channels are displaced by half a round trip. This can be seen from the OSNR measurements in Figure 1.4, where the wavelengths shorter than 1547nm are loaded at JDSU WSS #2 and wavelengths longer than 1547nm are loaded at JDSU WSS #1. The ROSNR increases from 8.5dB to 13dB at the longest distance. By subtracting ROSNR from DOSNR, we can also obtain the OSNR margin. We observed a margin decreases from 13dB to almost 0dB as a function of distance. These measurements form a performance baseline that can now be used as we introduce dynamic capabilities into the system.

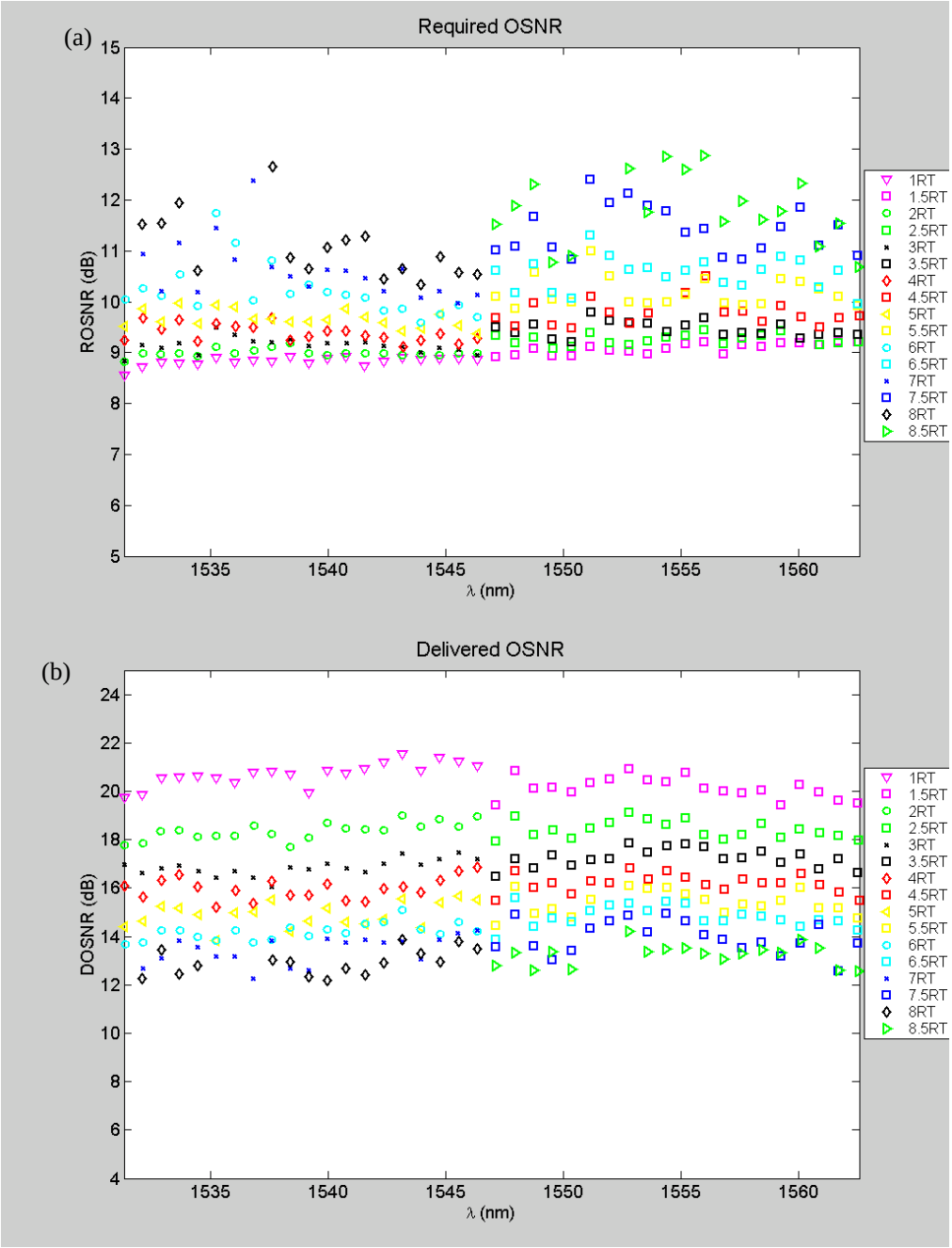

Figure 1.4. Transmission performance measurements at different numbers of loop round trips (RT) of 80 x 10Gb/s signals in the 2 ROADM distributed-distance loop (a) transmission penalties characterized by the ROSNR and (b) the amplifier noise penalty characterized by the DOSNR.

Single 40Gb/s OSNR Measurements in Circulation Loop

Introducing higher bit rate signals is another key testbed modification in which we engaged. In order to understand the penalty for higher data rate (40Gb/s) WDM channels, we placed a single 40Gb/s signal at 1538.19 nm among the other 10Gb/s channels at the entrance of the loop. Again we measured both ROSNR and DOSNR at different distances as performance metrics, varying

the 40Gb/s channel launch power as shown in Figure 1.5. The two adjacent 10Gb/s neighbors are turned on for 50GHz spacing measurements; and turned off for 100GHz spacing measurements. We observed a 0.6dB ROSNR penalty at 50GHz spacing setting. At 643.6 km (two round trips), the 40Gb/s signal starts to show a rapid ROSNR growth indicating our current dispersion map is limiting the maximum 40Gb/s transmission distance, as expected for the RTZ dispersion map. Using these measurements as a baseline, we can now study the impact of dynamic switching and mesh transmission on the 40 Gb/s signals.

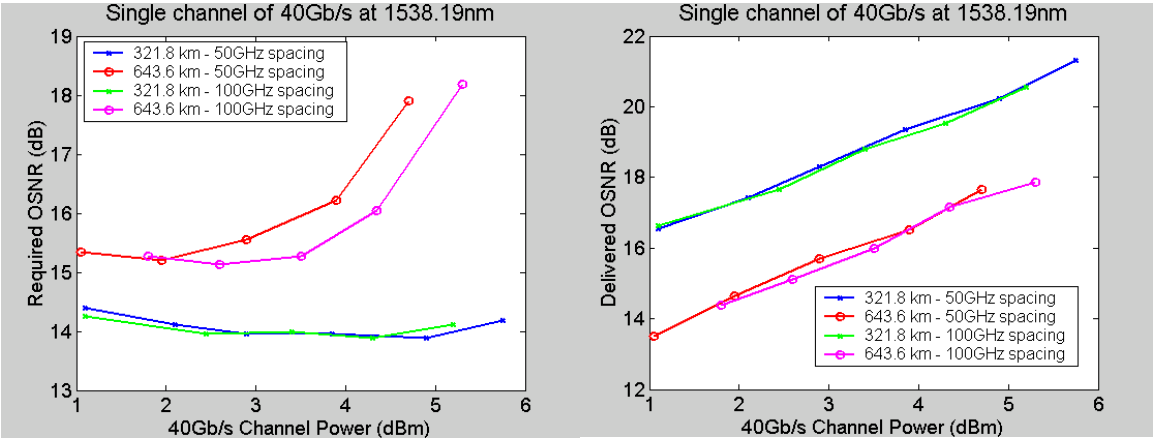

Figure 1.5. Measurements of a single 40Gb/s channel along with 79 x 10Gb/s channels at various launch powers (a) transmission penalties characterized by the ROSNR and (b) amplifier noise penalty characterized by the DOSNR

Reference:

[1] S. Yao, B. Mukherjer, and S. Dixit, "Advances in Photonic Packet Switching: An Overview," *IEEE Commun. Mag.*, pp. 84–94 (2000).

[2] C. Minkenberg, R. P. Luijten, F. Abel,W. Denzel, and M. Gusat, “Current issues in packet switch design,” *ACM SigComm Comput. Commun. Rev.* 33, pp. 119–124, (2003).

[3] R. S. Tucker, “Optical Packet-Switched WDM Networks—A Cost and Energy Perspective,” in *Proceedings of OFC/NFOEC*, paper OMG1 (2008).

[4] G. Weichenberg, V. W. S. Chan, M. Médard, “On the Throughput-Cost Tradeoff of Multi-Tiered Optical Network Architectures” in *Proceedings of IEEE Global Communications Conference* (2006).

[5] H. Zang, J.P. Jue, and B. Mukherjee, “A Review of Routing and Wavelength Assignment Approaches for Wavelength-Routed Optical WDM Networks” *Optical Network Magazine* 1, (2000).

[6] R. Parthiban, R. S. Tucker, C. Leckie, “Waveband grooming and IP aggregation in optical networks” *J. Lightwave Technol.* 21, pp. 2476-2488 (2003).

[7] J. Strand, and A. Chiu, “Realizing the advantages of optical reconfigurability and restoration with integrated optical cross-connects” *J. Lightwave Technol.* 21, pp. 2871-2882 (2003).

[8] P. Newman, G. Minshall, and T. L. Lyon, "IP Switching - ATM Under IP," *IEEE/ACM Transactions on Networking* 6, pp. 117-129, (1998).

[9] R. E. Wagner, R. C. Alferness, A. A. M. Saleh, and M. S. Goodman, “MONET: Multiwavlength Optical Networking” J. Lightwave Technol. 14, pp. 1349-1355 (1996).

[10] M. D. Feuer, D. C. Kilper, and S. Woodward, “ROADMs and Their System Applications,” in *Optical Fiber Communications V B*, eds. I. P. Kaminow, T. Li, and A. E. Willner (Academic Press, San Diego, CA), 2008.

[11] P. J. Winzer and R.-J. Essiambre, “Advanced modulation formats for high capacity optical transport networks” *J. Lightwave Technol.* 24, pp. 4711-4728 (2006).

[12] D. T. Neilson, C. R. Doerr, D. M. Marom, R. Ryf, M. P. Earnshaw, “Wavelength selective switching for optical bandwidth management” *Bell Labs Tech. J.* 11, pp. 105-128 (2006).

[13] W. T. Anderson, et. al., “The MONET project—a final report”, *J. Lightwave Technol.* 18, pp. 1988-2009 (2000).

[14] C. Checkuri, P. Claisse, R.-J. Essiambre, S. Fortune, D. C. Kilper, W. Lee, N. K. Nithi, I. Saniee, B. Shepard, C. A. White, G. Wilfong, and L. Zhang, “Design tools for transparent optical networks” *Bell Labs Tech. J.* 11, pp. 129-143 (2006).

[15] I. Roudas, N. Antoniades, D. H. Richards, R. E. Wagner, J. L. Jackel, S. F. Habiby, T. E. Stern, and Al F. Elrefaie, “Wavelength-domain simulation of multiwavelength optical networks” *IEEE J. Sel. Top. Quantum Electron.* 6, pp. 348-362 (2000).

[16] C. Qiao and M. Yoo, “Optical burst switching OBS – a new paradigm for an optical internet,” *J. High Speed Networks* 8, pp. 69–84, (1999).

[17] T. Battestilli and H. Perros, “An introduction to optical burst switching,” *IEEE Optical Communications* 41, pp. S10–S15 (2003).

[18] N. Madamopoulos, D. C. Friedman, I. Tomkos, and A. Boskovic, “Study of the Performance of a Transparent and Reconfigurable Metropolitan Area Network” *J. Lightwave Technol.* 20, pp. 937-945 (2002).

[19] G. Raybon, S. Chandrasekhar, A. H. Gnauck, B. Zhu, and L. L. Buhl, “Experimental investigation of long-haul transport at 42.7-Gb/s through concatenated optical add/drop nodes” in *Proceedings of OFC/NFOEC,* paper ThE4 (2002).

[20] A. R. Pratt, B. Charbonnier, P. Harper, D. Nesset, B. K. Nayar, and N. J. Doran, “40x10.7 Gbit/s DWDM transmission over a meshed ULH network with dynamically re-configurable optical cross connects” in *Proceedings of OFC/NFOEC*, paper PD9 (2003).

[21] P. Peloso, D. Penninckx, M. Prunaire, and L. Noirie, “Optical transparency of a heterogeneous pan-European network” *J. Lightwave Technol.* 22, pp. 242-248 (2004).

[22] A. Morea, D. C. Kilper, I. S. Lin, F. Leplingard, S. Chandrasekhar, T. Zami, J.-C. Antona, “Testbed methods to study physical layer path establishment in long haul optical wavelength switched networks”, *ICTON Conference*, Portugal, 29 June – 2 Jul. 2009

[23] Applied Constant Gain Amplification in Circulating Loop Experiments, Frank Smyth, Daniel C. Kilper, S. Chandrasekhar, and L.P. Barry, *Lightwave Technology, Journal of* : Accepted for future publication, 2009

[24] Scaling of Most-Likely Traffic Patterns of Hose- and Cost-Constrained Ring and Mesh Networks, Steven K. Korotky, Kostas N. Oikonomou, *Optical Fiber Communication Conference*, 2006

[25] Dispersion map optimisation and dispersion slope mismatch effects for 40 channel · 10 Gbit/s transmission over 3000 km using standard SMF and EDFA amplification, Bo-Hun Choi *, Manik Attygalle, Yang Jing Wen, Sarah D. Dods, *Optics Communications* Volume 242, Issues 4-6, 8 December 2004, Pages 525-532

2. Coherent Free Space Optical Communication and Network

Sponsors
DARPA –TACOTA: W91NF-06-1-0062
ONR- Free space optical heterodyne communication and network: N00014-09-1-0131

Project Staff
Vincent W. S. Chan, Etty Lee, Andrew Puryear, Rui Jin

Our research addressed the feasibility of an optical wireless network over the atmosphere in the presence of turbulence and interference, Figure 2.1. These networks will have significant impact for over-land and over-ocean applications as well as links to aircraft and satellites. Data rates of over 100 Gbps can be supported but only with the right mitigation techniques against various impediments. The most significant challenge is to handle atmospheric turbulence/weather and intentional interference so network protocols can run over these links.

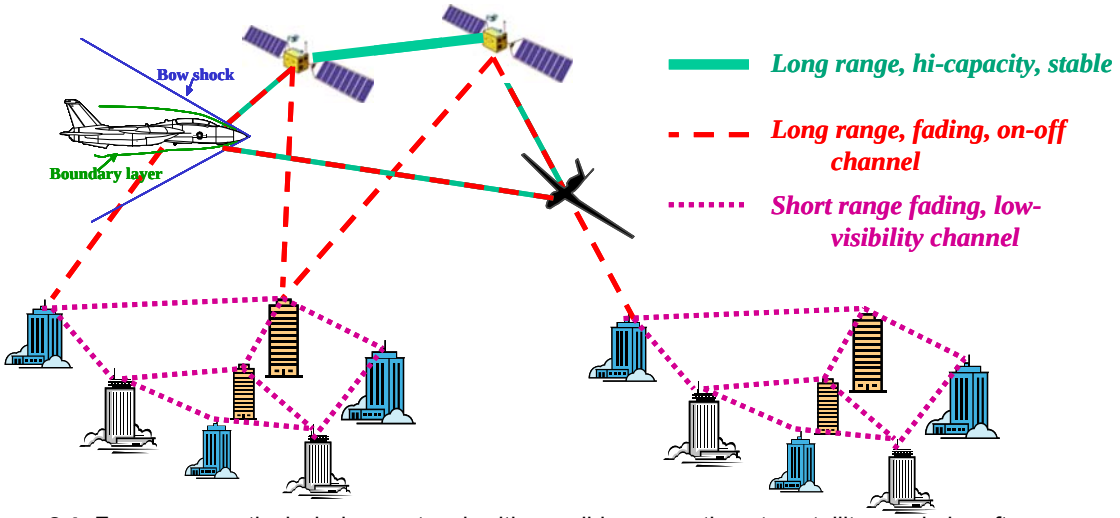

Figure 2.1. Free space optical wireless network with possible connections to satellites and aircraft.

The program was conducted with the following two goals:

> *1. Demonstrate the superiority of diversity coherent systems over direct detection systems in the mitigation of fading due to atmospheric turbulence and background noise rejection.*
>
> *2. Demonstrate the superior performance of coherent systems against jamming including using near-field transmitter and receiver phase and amplitude control with feedback*

Diversity Systems

Diversity is a sensible technique to mitigate fading experienced in free space optical communication because of the improvement in average fading statistics that it provides. Moreover, in the presence of background noise and an intelligent interferer, diversity *coherent* detection is preferable over diversity *direct* detection. This is due to coherent detection's ability to limit the amount of unwanted background noise and interference detected.

The available diversity equals the product of the number of independent (in the sense of being in different phase coherence cells) transmitter elements and receiver elements. If a low rate (~10kbps) feedback link is available between the receiver and the transmitter, there is the possibility of transmitter phase pre-distortion to help focus the optical energy on the receiver

array, enhancing energy delivery efficiency of the free space channel. Receiver phase tracking element by element either via local-oscillator (LO) tuning or signal path phase modulation is demonstrated in this research program. The data is fed back to the transmitter for phase and amplitude modulation of the individual array elements.

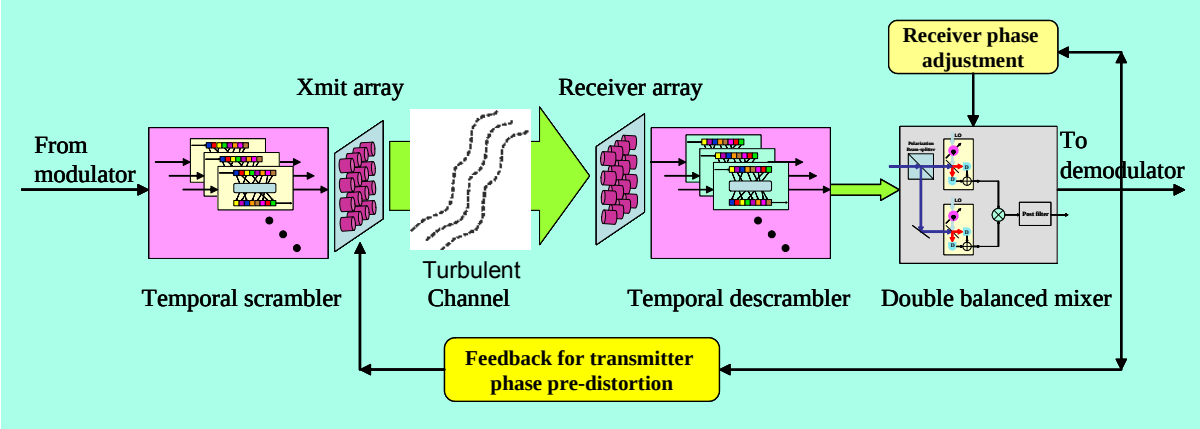

Figure 2.2. Transmission system with dynamically and adaptively reconfigurable receiving plane field patterning, crypto-key driven frequency spreading and time interleaving.

TCP Shortcomings in Free Space Optical (FSO) Networks and Modified TCP

TCP's congestion control which has been used in the Internet for roughly two decades has been successful in preventing congestion collapse. However, when data rates and geographic spans increase, and communication links without wires or cables (such as satellite and free-space optical links) are added into the network, TCP has performance issues leading to low throughput. This is partly due to the TCP sender's limited rate of window increase. It is also due to the TCP sender unnecessarily reducing its window upon link losses (it assumes every packet loss is due to congestion and that it should thus reduce its window to relieve the congestion).

FSO optical links are different from fiber optic and radio frequency (RF) wireless links because they have long fades. It is typical for an outage to last tens of milliseconds even when reasonable amounts of link margin and diversity are used. These long link outages cause a large number of consecutive packets to be dropped and can cause the TCP sender's retransmission timer (RTO) to expire thereby resulting in reduction of the sender's window to one packet in flight without acknowledgement. In fiber optic and RF wireless links, the typical losses are of single packets which usually cause three duplicate ACKs and window halving rather than timeouts. The fact that FSO outages can cause timeouts is an observation that has not been considered before. When high bandwidth-delay product paths are combined with FSO links with atmospheric turbulence, the outages cause severe throughput degradation. The throughput degradation is severe because after a long outage occurs and the sender's window timeout and is reduced to one packet in flight, it takes the sender a long time to increase the window to a size large enough to make good use of the available rate in the network.

We studied a modified version of TCP that does not have the problem of reduced throughput due to outages causing window reduction to one packet. Specifically, the Modified TCP does not respond to a large number of outage losses by timing out. Rather, it is able to distinguish whether the packet loss is due to congestion or link outage. See Figure 2.3 for the expected outage lengths for direct and coherent detection. One way this can be accomplished is by having the router feed back a 'Congestion Loss' message to the sender for each packet the router drops (and optionally try to guarantee that this feedback gets to the sender by having routers give priority to these feedback messages or by using out-of-band feedback). If a 'Congestion Loss' feedback packet is received, the sender responds to the congestion by reducing its window. If no

'Congestion Loss' feedback packet is received and the RTO expires, the sender assumes the timeout is due to a link outage and the sender does not cut its window. The Modified TCP sender's ability to distinguish outage losses and congestion losses (and respond appropriately) significantly improves sender throughput.

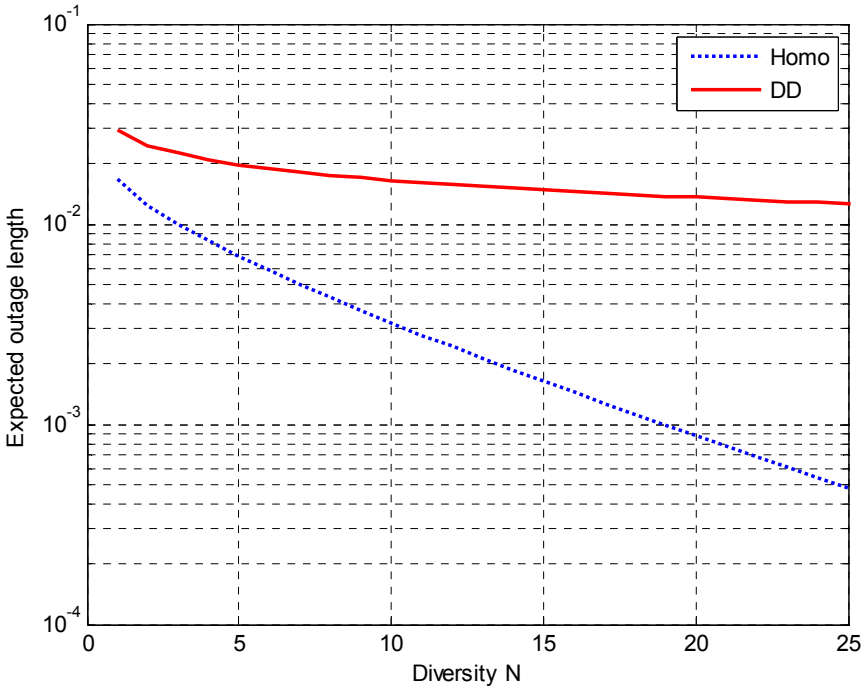

Figure 2.3: Expected outage length versus diversity for fixed link margin of 5 dB when σ_χ^2=0.5, P_e^{thresh}=0.1, N_n=1, transverse wind speed is 10 km/hr, path distance=20km

Steady State Analysis (Average Throughput)

The average throughput of TCP in steady state is given by

$$\text{steady state TCP throughput} = \frac{E[\text{window size}]}{RTT}$$

Using a linear window increase Markov chain, we obtain a lower bound on TCP throughput at steady state

$$\text{steady state TCP throughput} \geq \frac{\sum_{n=1}^{n_{\max}} n\pi_n^{linear}}{RTT} \text{ packets/sec}$$

where the superscript "linear" denotes the linear window increase Markov chain, and π_n is the steady state probability of being in state n. Using an exponential window increase Markov chain, we obtain an upper bound on TCP throughput at steady state

$$\text{steady state TCP throughput} \leq \frac{\sum_{n=1}^{n_{\max}} 2^{n-1}\pi_n^{\exp}}{RTT} \text{ packets/sec}$$

where the superscript "exp" denotes the exponential window increase Markov chain.

Transient Analysis

When a user sends a small or moderate sized file (for example <1 MB), TCP is often not in steady state for most of the file transfer, particularly if the round-trip distance is long. Thus, steady state throughput does not give an accurate estimate of the time it takes to send the file. Transient throughput of TCP and Modified TCP after session initiation is the critical performance metric.

The expected number of packets sent in the m^{th} round-trip time is given by

$$E\left[\text{number of packets sent in m}^{\text{th}}\ \text{RTT}\right] = \begin{cases} \sum_{i=1}^{n_{\max}} i p_i(m) & \text{for linear increase(lower bound) of both Modified TCP and TCP} \\ \sum_{i=1}^{n_{\max}} 2^{i-1} p_i(m) & \text{for exponential increase (upper bound) of both Modified TCP and TCP} \end{cases}$$

where $p_i(m)$ is the probability of being at a window size of i in the m^{th} RTT. The $p_i(m)$ can be obtained from the following evolution of probability distribution across the states:

$$\begin{aligned} \vec{p}(m) &= \vec{p}(m-1)P \\ &= \vec{p}(1)P^{(m-1)} \end{aligned}$$

where $\vec{p}(m)$ is a row vector of probabilities of being in each of the n_{max} TCP window size states in the m^{th} RTT, P is the probability transition matrix for the Markov chain, and $P^{(m-1)}$ is the matrix product of P with itself (m-1) times and represents the transition matrix from any given RTT to (m-1) RTTs later. TCP starts with an initial window size of 1. Thus, $\vec{p}(1) = [1 \quad 0 \quad 0 \quad \ldots \quad 0]$.

The expected number of packets sent in K round-trip times is given by:

$$E\left[\text{number of packets sent in K RTTs}\right] = \sum_{m=1}^{K} E\left[\text{number of packets sent in m}^{\text{th}}\ \text{RTT}\right]$$

Thus, we can calculate the expected number of packets sent in the m^{th} RTT and in K RTTs if we know the transition matrix of the Markov chain.

Figure 2.4 and 2.5 show the steady state throughput of TCP and Modified TCP. The links have a maximum transmission rate of 10 Gb/s, packet size of 1500 bytes, 8 dB of link margin, error probability threshold of 0.1, transverse wind speed of 10 km/hr and background noise photons per symbol per mode of one.

The steady state TCP throughput goes down when atmospheric turbulence increases. This is because packet losses due to link outages cause the sender to timeout and reduce its window size to one packet, and higher turbulence means more frequent window reduction. Furthermore, for long round-trip time sessions (such as a path from the Earth to a geosynchronous satellite with round-trip time of a quarter of a second), the throughput is very low when turbulence and congestion losses exist. This is because after each window reduction, the window size is increased gradually (every quarter of a second), and it takes the TCP sender a large number of round-trip times to increase its average rate. Even if the window increase is exponential, the rate of increase is still slow due to the long round-trip time. Moreover, before the window builds up to full value, another outage will cut the window again preventing the TCP sender from reaching the maximum possible number of packets in flight. For large round-trip time links through the atmosphere (and large sender transmission rate), TCP congestion control is far from ideal as a congestion control protocol.

Modified TCP gains back the performance loss due to outages. This is because Modified TCP does not cut its window in response to outage losses and thus also does not have to suffer the slow increase in its window after the outage. Over large round-trip time free-space optical links, Modified TCP performs significantly better than TCP when the turbulence is strong such that the outages dominate the congestion losses. However, TCP and Modified TCP have an issue of slow window increase over high bandwidth-delay products, which is yet to be addressed in future research. Moreover, this slow window increase over high bandwidth-delay product paths results in unfairness to sessions with large round-trip times because the sessions with small round-trip times increase their window more quickly (due to ACKs returning more quickly) and are able to use up more of the available network capacity.

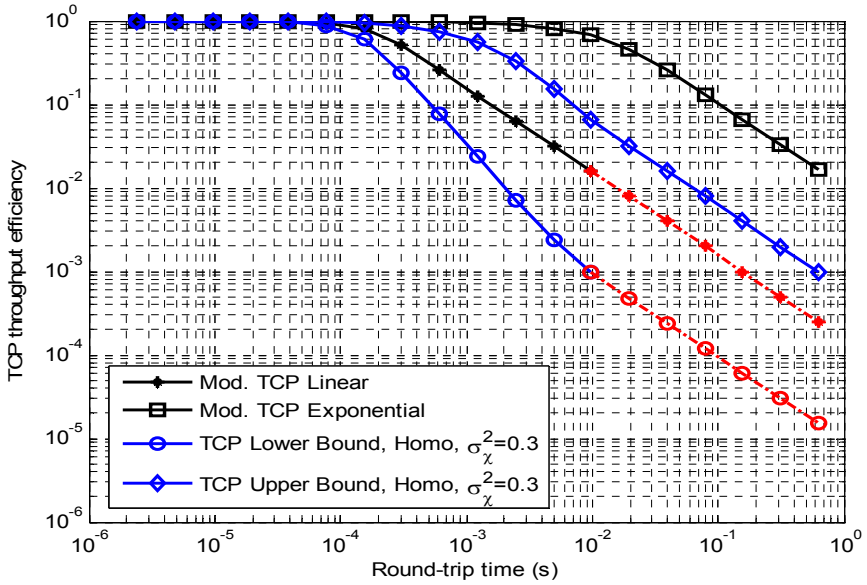

Figure 2.4: Throughput efficiency for congestion loss per packet of 10^{-4} and σ_χ^2=0.3

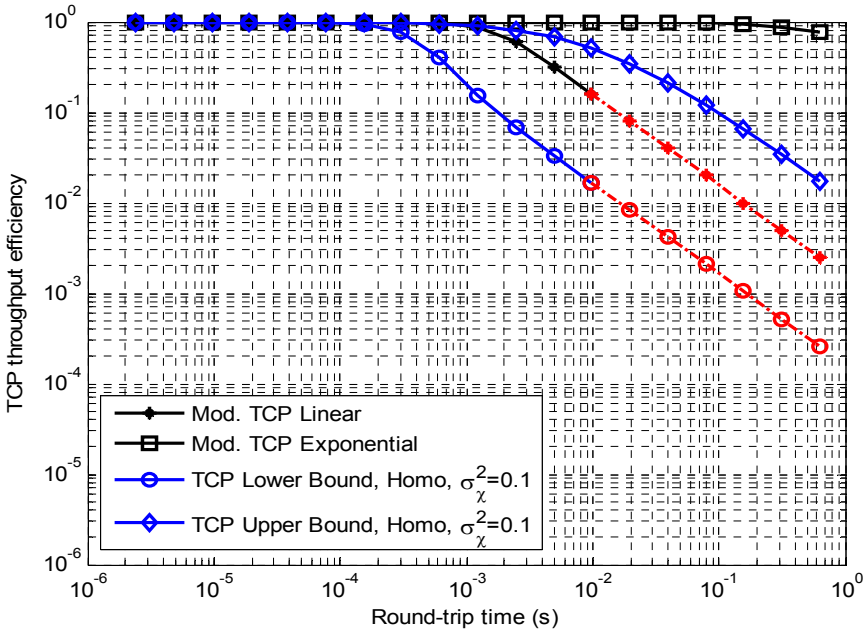

Figure 2.5: Throughput efficiency for congestion loss per packet of 10^{-6} and σ_χ^2=0.1

Transient Throughput of TCP and Modified TCP

Figures 2.6 and 2.7 plot the amount of data transmitted in the K^{th} RTT and up to the K^{th} RTT by TCP and Modified TCP for various congestion loss probabilities and turbulence levels. These plots use the same parameters as used for the steady state throughput plots but where the round-trip time is fixed to 0.3 seconds.

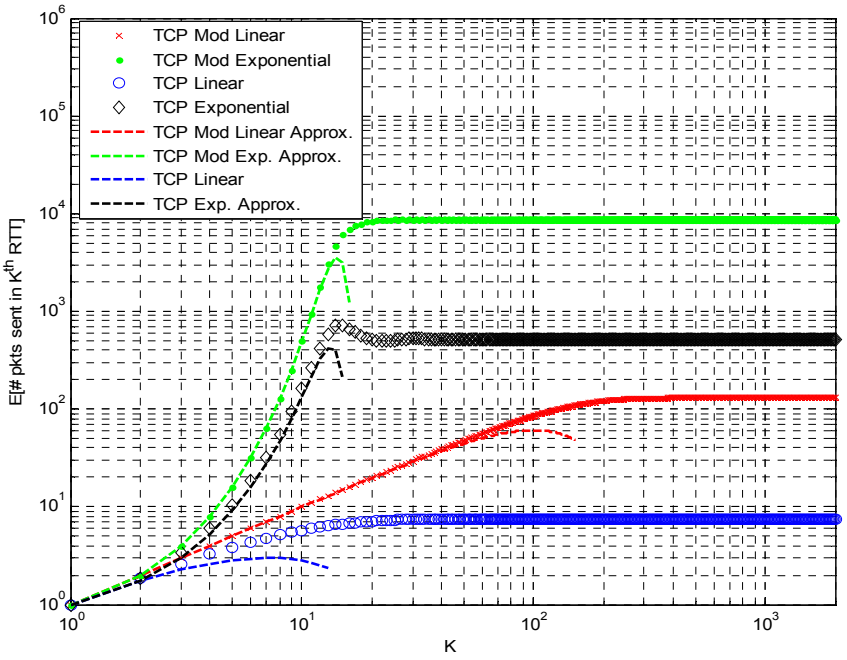

Figure 2.6. Expected number of packets sent in the K^{th} round-trip interval for congestion loss per packet of 10^{-4} and σ_X^2=0.3 and M=2^{18} (RTT=0.3 sec)

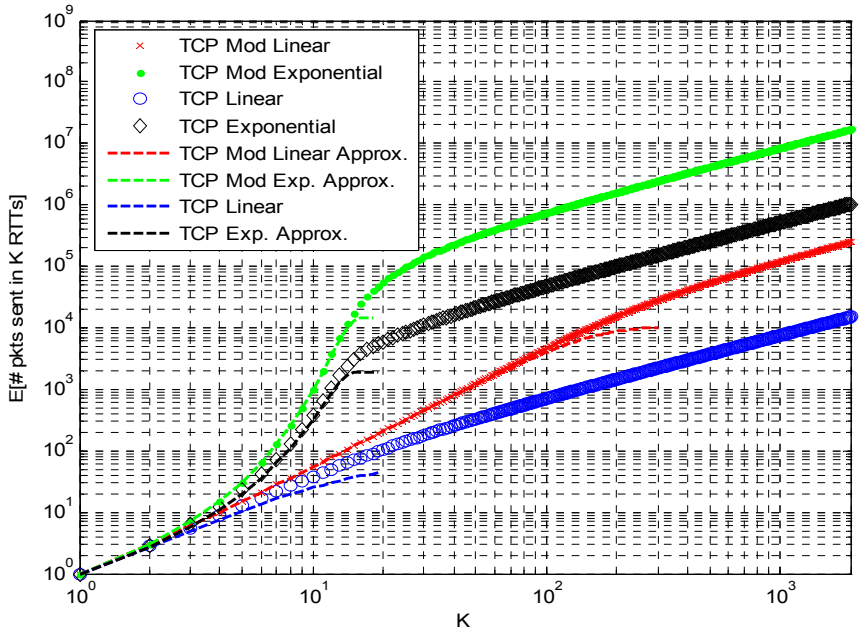

Figure 2.7. Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and σ_X^2=0.3 and M=2^{18} (RTT=0.3 sec)

For long round-trip distances, when packet losses due to congestion occur with independent probability $p_{congperpkt}$, the value at which the expected number of packets sent in the K[th] RTT levels off is approximately

$$E\left[\text{number of packets sent in K}^{\text{th}}\text{ RTT}\right] \to \min\left(M, \frac{0.75}{p_{congperpkt}}\right)$$

for Modified TCP exponential increase, and approximately

$$E\left[\text{number of packets sent in K}^{\text{th}}\text{ RTT}\right] \to \min\left(M, \sqrt{\frac{2}{p_{congperpkt}}}\right)$$

for Modified TCP linear window increase. M is the maximum possible number of packets in flight for the given RTT. For quite some time TCP has been suspected to have poor performance over long atmospheric links and these analytic results have upper and lower bounded its performance.

If the receiver is in the near-field of the transmitter, it is possible to modulate the received spatial mode of the signal and via feedback to the transmitter select a spatial mode with as small an overlap (small inner product) with the interfering field as possible and provide maximum power transfer to the receiver, Figure 2.8.

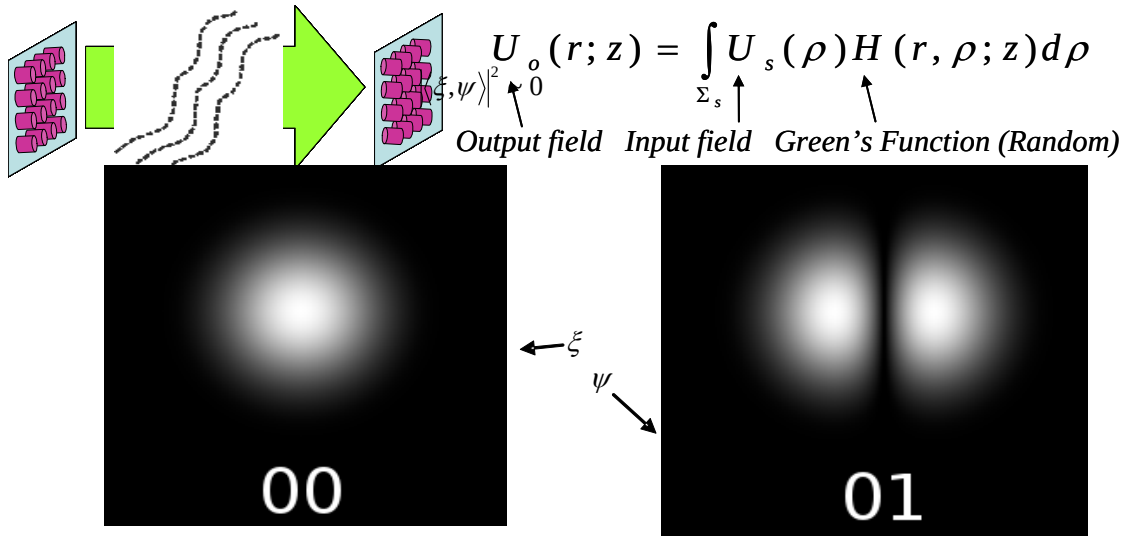

Figure 2.8. Spatial mode modulation for near-field AJ and multiple-access communication.

The objective of the research is to develop a cost-effective low-power free space optical communication system architecture suitable for rapidly deployable, dynamic, high data rate communication applications. Any mobile platform requiring high data rate communication, such as a drone, could potentially be a direct beneficiary of this research. Additionally there are examples when a fixed platform would benefit from this project: a company may require extra communication capacity to support a short-term high speed connection. Atmospheric turbulence, can however cause severe fading leading to the loss of 10^9 consecutive bits of data. As a result, this program focuses on developing an architecture to overcome turbulence induced fading. In general, there are several techniques used to overcome turbulence induced fading. Some current systems use a single transmitter and single detector and simply use extra power to overcome the 20-30dB fades. This approach is prohibitively expensive due to the high cost of multi-watt 30dB amplifiers. Another approach, borrowed from astronomy, is to employ a large deformable mirror to compensate for the

turbulence. This approach effectively uses spatial diversity to overcome the turbulence. Theoretically it performs very well, but is there a way to exploit spatial diversity without the need for large, expensive deformable mirrors? The answer is a sparse aperture system, which uses many small transmitters and receivers spaced far enough apart to achieve the same spatial diversity as the deformable mirror provides. A sparse aperture system is shown in Figure 2.9.

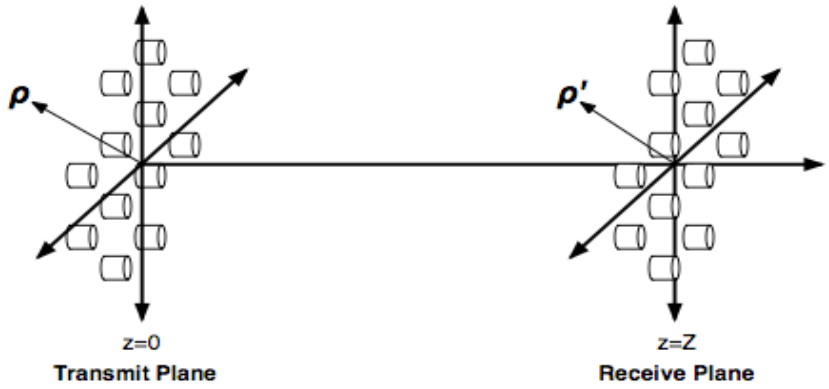

Figure 2.9: Sparse aperture free space communication system.

We investigate practical engineering questions related to the sparse aperture system, such as what is the optimal size and spacing of individual transmitters/receivers and how to fully exploit the spatial diversity provided by this setup. To fully exploit the diversity, the sparse aperture system must use wavefront predistortion based on an estimate of the turbulence state. Because the coherence time of the atmosphere is relatively long compared to the propagation time of light, we propose using a low-rate feedback link from the receiver to inform the transmitter of the channel state. This system can be realized, especially if the turbulence or the receiver is in the near-field of the transmitter. Other relevant engineering questions we address include:

1. Given a system geometry, what feedback rate R is needed to take full advantage of the diversity?

2. What information should be fed back?

3. Given a system geometry and performance requirements, what is the tradeoff between increasing transmit power and increasing feedback rate?

4. How many transmit and receive apertures does the system need to ensure a required outage probability? To increase average BER, should the system designer increase the number of transmitters, the number of receivers, or both?

5. How does increasing transmit power or increasing the number of transmitters impact the probability of outage?

Sparse Aperture Communication Performance

We found the spatial modulation that optimally mitigates the effects of the turbulent atmosphere under the assumption of independent control of the phase and magnitude of each transmitter. Through coherent detection, the receiver measures the phase and magnitude of the received wave. The architecture optimally allocates transmit power to the spatial modes with the smallest propagation losses to decrease bit error rate and mitigate turbulence-induced outages. Figure 2.10 compares the predicted theoretic average bit-error-rate (lines) of the optimal spatial modulation with a simulation (points) of communication through turbulence.

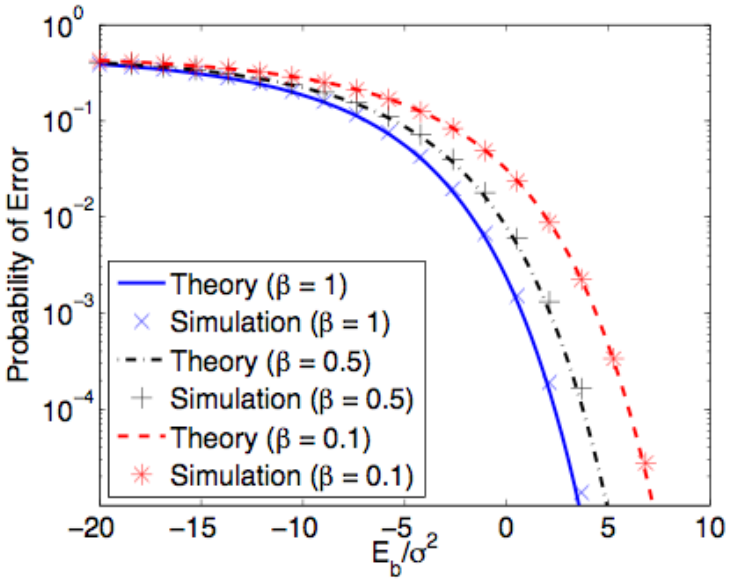

Figure 2.10: Comparison of simulation and theory for optimal spatial modulation verses signal-to-noise ratio. Turbulence strength $\sigma\chi^2$ = 0.3.

Average BER, while still a relevant metric, does not fully characterize free space optical systems which are limited by long deep outages: turbulence-induced constructive and destructive interference will cause significant variation in the metric around its mean performance. As a result we quantify the variation by developing closed-form expressions for outage probability and power efficiency. There are many ways to measure the variability in system performance due to fading. Outage probability defined in terms of BER is particularly useful because it guarantees at least some minimum performance some fraction of the time. Figure 2.11 shows the theoretic outage probability for the optimal spatial modulation. It is apparent from the figure that increasing the number of apertures dramatically reduces the outage probability.

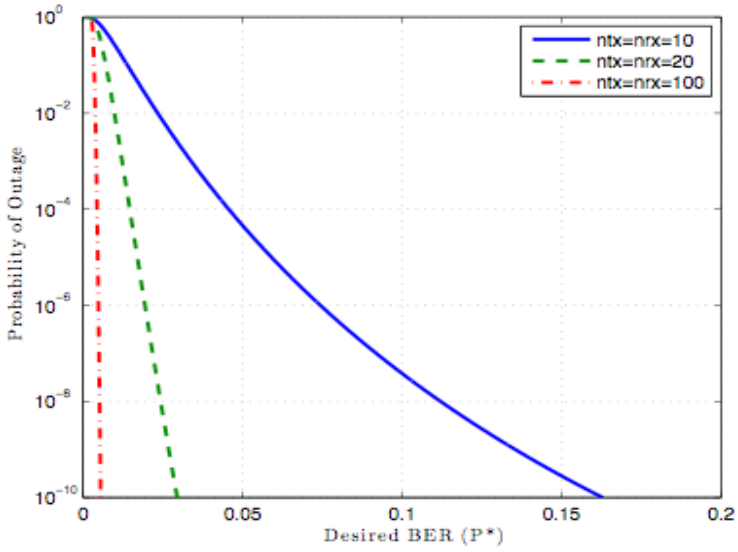

Figure 2.11. Probability of outage verses desired BER for various numbers of apertures and SNR=1. Ntx is the number of transmitter apertures, nrx is the number of receiver apertures.

To compare systems, we define spatial modulation gain to be the average power gain due to spatial modulation relative to the average power with no spatial modulation. The metric provides a rough value in terms of SNR of a particular modulation scheme. By definition, the open loop system attains 0dB of spatial modulation gain. For a balanced system, comparing the asymptotic BER given for wavefront predistortion with the asymptotic BER for open loop, we see that an open loop system, on average, performs as well as a wavefront predistortion system with 6dB less transmit power. Figure 2.12 shows the spatial modulation gain for all three systems verses number of transmitters for a fixed 100 receivers. As the number of transmitters becomes large compared to the number of receivers, the modulation gain grows unbounded. Intuitively, the more transmitters the higher the probability the system will experience a favorable state.

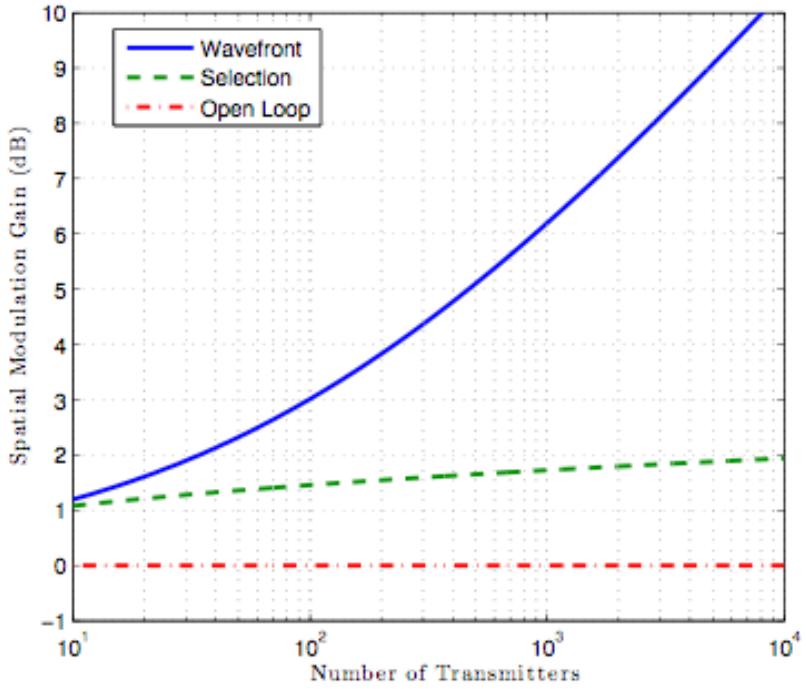

Figure 2.12. Spatial modulation gain for all three systems verses number of transmitters for a fixed 100 receivers

This average result, however, can be misleading for any real world system because instantaneous BER fluctuation around the mean can cause catastrophic outages. Figure 2.13 shows a comparison of outage probability for open loop, selection, and wavefront predistortion systems with $n_{tx}=n_{rx}=100$. Because the outage probability for the open loop system falls all relatively slowly, it will require more than the average power gain, the spatial diversity gain, to perform as well as the wavefront predistortion system with a reasonable outage probability. We define the following metric to quantify this effect, showing how much additional power an open loop system requires to operate with the same performance as a wavefront predistortion system: feedback power gain is the additional power that a system with no channel state information needs to perform as well as a system with perfect CSI at least P_{out} fraction of the time. The term feedback power gain is used because it represents the equivalent power gain enabled by feedback. From studying the feedback power gain, we find that the feedback power gain for the optimal system converges to four as the number of apertures approaches infinity. An interesting interpretation is, as the number of apertures grows, the value of the feedback diminishes.

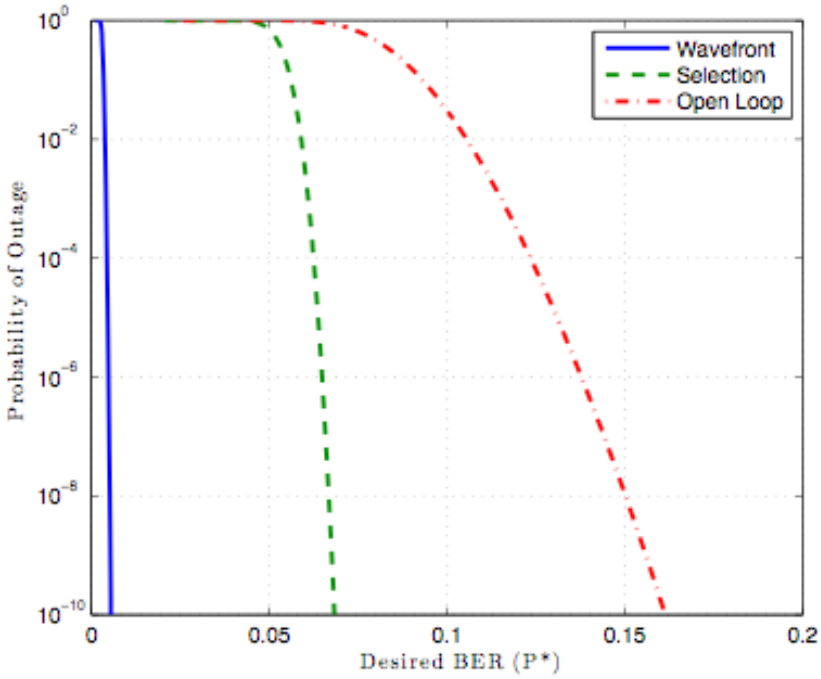

Figure 2.13. Outage probability verses desired BER for wavefront, selection, and open loop spatial modulation.

As the outage probability is reduced, the feedback power gain is greatly increased: if a system is required to operate at a low outage probability, the value of the feedback information is very high. As the number of apertures increases, the number value of feedback is diminished. Asymptotically, as the number of apertures gets very large the feedback power gain approaches the average value.

Conclusion

Optical communication over the turbulent atmosphere has the potential to provide reliable rapidly reconfigurable multi-gigabit class physical links over tens of kilometers. Such systems, however, are prone to long (up to 100ms) and deep (10-20dB) fades and are susceptible to interference. For this project, we have shown that a sparse aperture system with spatial mode control provides cost-effective protection against fades compared with a single aperture system.

Optical communication generally provides such high data rates that the added complexity involved in implementing a system that communicates over multiple spatial modes simultaneously is not typically justified by the added rate. As such, this work has focused on metrics related to communicating on one spatial mode at any given instant, such as BER and outage probability.

3. Heterogeneous Defense Network Study

Sponsors
DARPA: HR0011-09-C-0131
OSD-DDR&E – Future Defense Network Study:700087748

Project Staff and other Participants
Vincent W. S. Chan, John Chapin, Anurupa Ganguly, Lillian Dai, Greg Ackers (Cisco Systems), Alan Kirby, Eric Swanson, Guy Weichenberg, Ada Poon (Stanford University)

This project investigates the fundamental architectural issues of networks that combine heterogeneity with guaranteed performance. A heterogeneous network is one that incorporates multiple communications modalities (e.g. wired, wireless, satellite, optical). Heterogeneous networks must provide agile and economical service delivery in the face of challenges that include rapidly changing communication channel quality, link connectivity, and traffic flows. The combination of multiple modalities in a single data path requires significant modification to existing resource management and control solutions that largely assume homogenous end-to-end communication characteristics. This problem is of high interest both in military and in commercial applications. The system we study to develop generalizable solutions is the global US defense network.

Since this is a fundamental architectural study, the methodology aims to identify classes of system, protocol, topology, and control designs that may be used to address various classes of end user requirements. Constraints of current and future technologies and operational environments are carefully studied and used to limit the design space explored. Detailed designs, for example specific topologies, are considered when analyzing the optimal approaches in different parts of the space. However the project is not intended to result in a recommending a specific point solution. Instead the output is design rules and guidance that will feed into the various specific designs created to meet specific mission requirements and technology constraints.

To start the research, we worked to develop a set of high level architecture principles for the integrated global defense network.

1. The Defense Network should be heterogeneous and accommodate multiple types of networks of different generations with the integration into a single interoperable network via gateways between disparate subnets and a Master Control Plane for subnet integration.

2. The Defense Network should have the following two-tiered network services, Figure 3.1:
    a. *A robust MINIMUM "hard-core" network service necessary for successful defense operations. This service will have modest rates but solid connectivity at all times with low delays and designed and tested to be bullet proof.*
    b. *A much higher rate "soft-shell" network, which is less robust but higher rates and support high rate services such as web-based Service Oriented Architectures.*

3. The integrated network should support voice, video, data, hybrid IP and circuit based services.

4. There are critical outstanding architecture and technology issues of infrastructure-less wireless networks and some SATCOM networks. The integrated network must be designed with awareness of these issues.

5. It is important to examine the new concept of a Master Control Plane to facilitate internetworking and Information Assurance: critical for the integration across subnets/domains, interoperability, priority/policy enforcement, access control and security, Figure 3.2.

6. The proliferation of information and data services to the tactical edge will generate and consume vastly more bandwidth in the future and will saturate the backhaul network. A Network-aware Distributed Information Service (NDIS) should be deployed near the network edges to adjust user behavior as the network changes to maintain critical services on the network, Figure 3.2.

This project was started in the Fall of 2009 and will continue to evolve in the next few years in the architecture development of a robust and scalable heterogeneous network. We expect both high level strawman designs and also creation of new architectures for protocols at all layers including relayering.

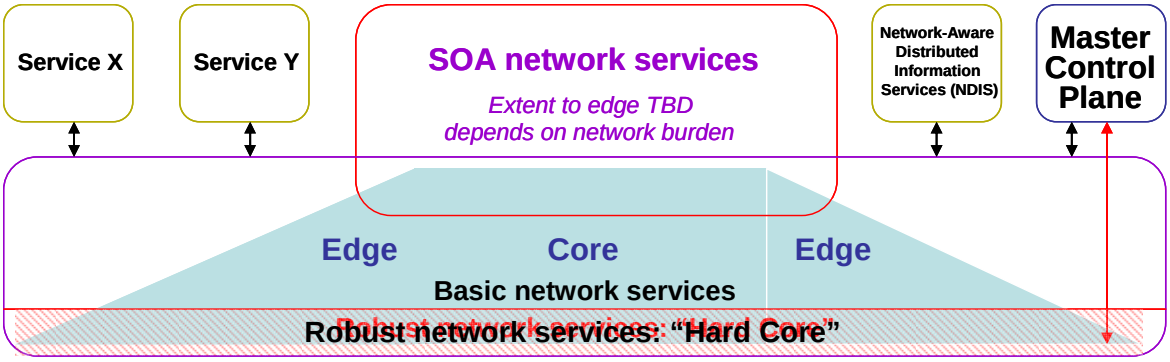

Figure 3.1. Defense network architecture with two-tiered robust "hard-core" network services (for time critical and network management control messages) and basic network services

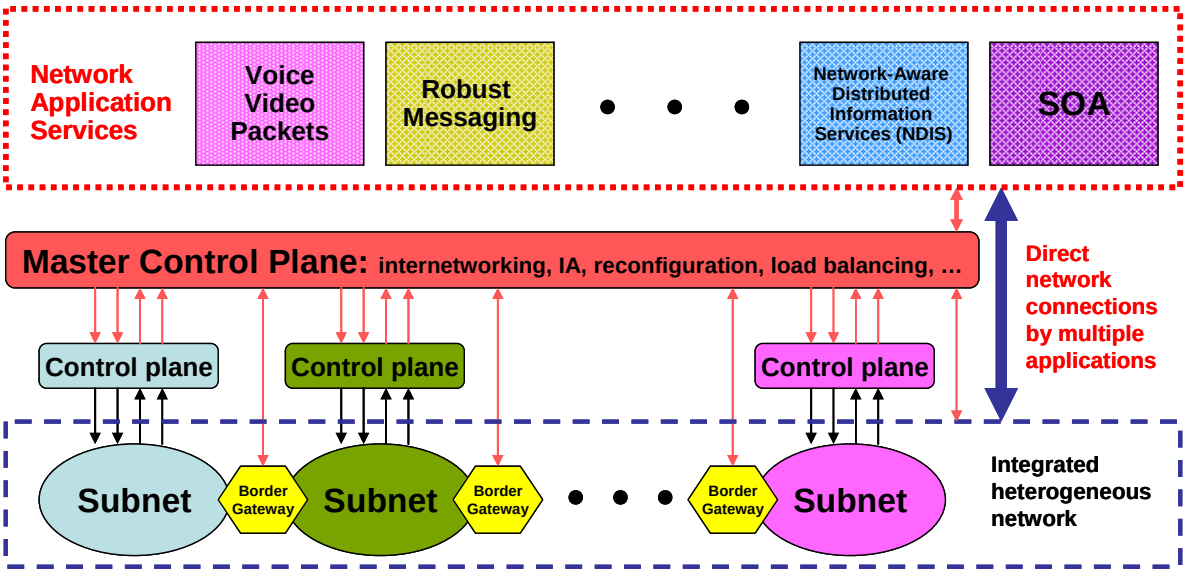

Figure 3.2 Integrated heterogeneous defense network with voice, video, data, hybrid IP and circuit based services.

Within the overall scope of the project, the focus during the last year has been in the following areas.

A. Interactions between physical layer effects in heterogeneous networks and transport layer performance.

Optical, wireless and satellite networks suffer from link quality variations that cause packet loss. The characteristic statistics of each modality strongly affect transport layer design and performance. The group studied the impact of free space optical link characteristics on the performance of Internet-standard Transport Control Protocol (TCP) and used these insights to develop and characterize a modified TCP with better performance (see earlier discussion in section 2). The group also investigated the use of packet replication across multiple communication paths to improve the packet loss statistics visible to the transport layer, and completed initial analysis of TCP performance and efficiency when replication is enabled in simple networks.

In a simplified network model where all parallel paths are fading channels with the same statistics (expected up time, expected down time, and distribution), we analyzed the overall efficiency of a network layer where packets sent by TCP are replicated across a number of parallel paths. The first replica to reach the receiver stimulates a TCP acknowledgment. Only if all replicas are lost does the sender detect a packet loss and potentially reduce its sending rate. Overall efficiency is upper bounded by the number of replicas used. If two copies of each packet are sent, then the maximum amount of user data transmitted is 50% of network capacity.

Figure 3.3 shows the efficiency of different numbers of replicas (n) in a system where link statistics (particularly round trip time) are chosen to model a geostationary satellite configuration. From many graphs like this, the replica count that leads to maximum efficiency can be read out and graphed as in Figure 3.4. We found that increasing roundtrip time increases the optimal number of replicas for maximum efficiency, which makes sense because TCP window closing leads to a greater loss of user throughput in high roundtrip systems where it takes a long time for TCP to ramp back up to full link rate. As shown in the figure, we also characterized the dependency of this effect on the statistics of the fading channels.

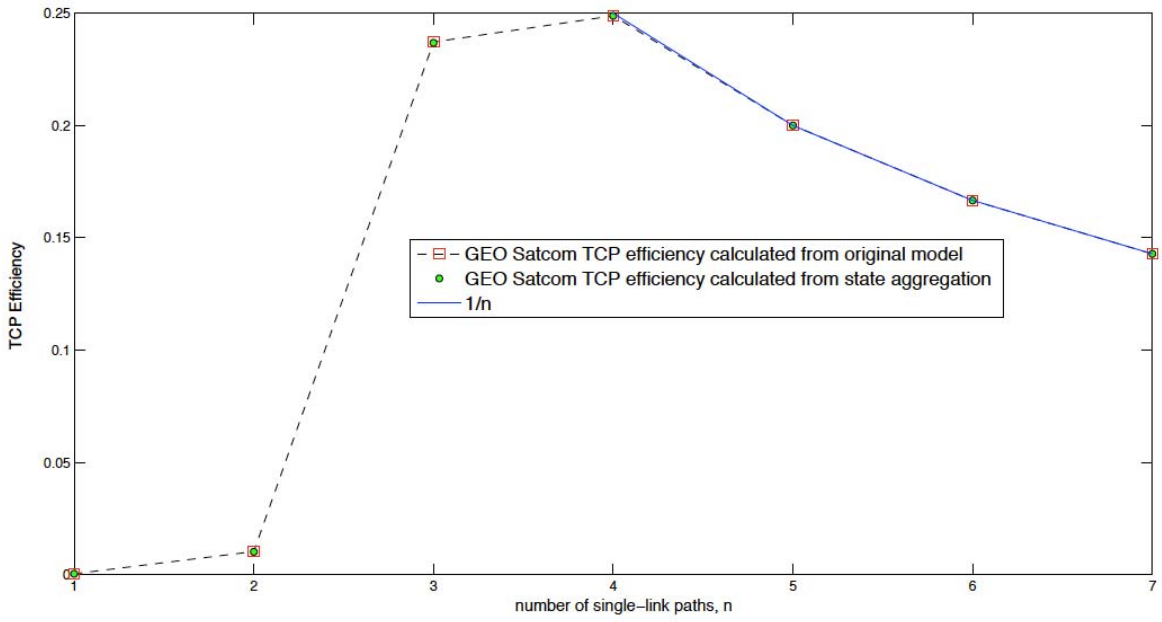

Figure 3.3. TCP efficiency in a GEO satellite network versus the number of replicas of each TCP packet (n). Link rate is 1 Gb/sec; packet size=10 Kb; expected uptime of the link is 10 seconds; expected disconnection time is 10 milliseconds; round trip time = 0.5 seconds; TCP maximum window size bounded only by delay-bandwidth product of the link.

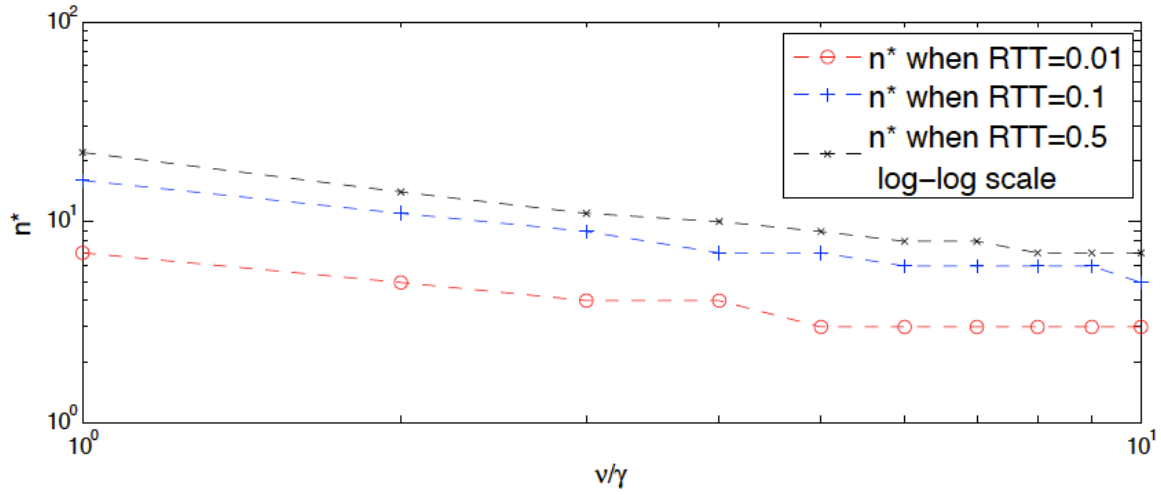

Figure 3.4. Optimal number of replicas of TCP packets (n*) versus different ratios of expected uptime to expected downtime, in a fading network with three different round trip times ranging from 10 milliseconds to 500 milliseconds. Expected uptime for each link is 1 second, while expected downtime varies. The rate of each link is 1 Gb/sec and packet size is 10 Kb.

B. Defense network architecture focused on highly assured data delivery over heterogeneous physical modalities.

The group initiated a cleansheet architecture project considering the interrelated problems of network topology, management, and protocols that can improve network performance, reliability, delivery guarantees and evolvability while still maintaining compatibility with off-the-shelf and legacy equipment and meeting the stringent security requirements of defense systems. The particular unmet end user requirement driving this part of the study is the need for highly assured data delivery. This work stands in direct contrast to the large recent research area of Delay Tolerant Networks. We focused on delay *intolerant* networks and developed multiple mutually supportive novel design ideas that work together to support this requirement. The following chart summarizes the key changes from a traditional high-capacity design to a high-connectivity service with high assurance of timely data delivery.

Feature	**High-capacity service**	**High-connectivity service**
1. Route discovery	Best route only	Best and alternate routes
2. Source routing	Not supported	Sender controls which of the alternates to use
3. Node location	Single IP addr in directories	Native support for multihoming
4. Route info shared by subnets	Vector of networks (e.g. BGP route)	Vector of networks with latency, headroom, other statistics
5. Response to packet loss	Reduce sender rate	Increase allocated network resources
6. Deadlines in packet headers	Ignored if present	Mandatory, used to control resource allocation
7. Behavior if link down or congested	Packet dropped	Option for store and forward by surrogate
Required lower layer behaviors for wireless networks		
8. Transmission control by network layer	Minimal	Give extra resources to packets near deadline
9. Channel access time	Can be >1 sec (eg. SATCOM)	<< 1 second

In conjunction with the above, we initiated study of network designs resistant to hostile intrusions of the control plane. Current systems are subject to catastrophic failure, which is not acceptable for a global-scale mission-critical network in which not all network managers and authorized users can be assumed to be trustworthy.

Deploying a new network stack such as suggested in the above chart is not feasible if off-the-shelf end devices all have to be replaced. We developed a method to interface standard unencrypted ("red") endpoints (using standard TCP, DNS, routing) through standard security devices used in US defense networks (HAIPE) to an innovative heterogeneous encrypted ("black") core network running a stack such as the high-connectivity stack outlined above. The key attributes of the method outlined in Figure 3.5 below are:

- TCP is an access protocol only, not end-to-end
  - A TCP proxy on the unencrypted or "red" side (which is mature off-the-shelf technology) spoofs the TCP link, making the endpoint think it is connected end-to-end via TCP to the other endpoint.
- The new Core Transport Protocol stack runs on the encrypted ("black") side of the security device (HAIPE)
  - In Core Transport Protocol Endpoint (CTPE)
  - The stack can do cross-layer optimization, interact with gateways, and adapt to the statistics of heterogeneous network modalities since it is running on the black side.
- Short-Haul Protocol (SHP) connects proxy and CTPE
  - A minimal, simple protocol with one end on red side, one end on black side
  - Uses innovative design techniques to perform flow control, demultiplexing and other necessary functions without violating security requirements

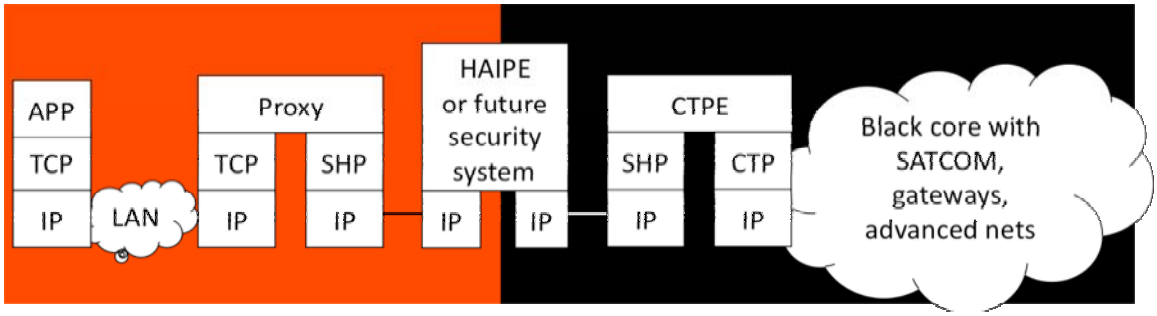

Figure 3.5 TCP-access protocol architecture.

As this project is in early stages, significant work remains in all of the above areas before an initial set of quantitative tradeoffs and design guidelines can be developed.