

Arquitetura da Internet

Dorgival Guedes, Professor Associado, Departamento de Ciência da Computação
Ítalo Cunha, Professor Adjunto, Departamento de Ciência da Computação

I. OBJETIVOS

Este curso é sobre a arquitetura da Internet, uma das invenções mais significativas dos últimos 30 anos e de claro impacto na sociedade. O sucesso da Internet é devido em parte à sua arquitetura, que inclui soluções para problemas complexos de escalabilidade, desempenho, gerência, robustez e custo. Este curso irá cobrir teoria, heurísticas, algoritmos e técnicas utilizadas em várias soluções empregadas na arquitetura da Internet. O conteúdo do curso tem utilidade direta para auxiliar no desenvolvimento de aplicações, operação, gerência e resolução de problemas de rede.

II. DINÂMICA E AVALIAÇÃO

As aulas serão como um *workshop* de discussão. Cada semana terá uma lista de dois a cinco artigos para leitura. O professor irá fazer uma apresentação dos trabalhos relacionados, os alunos irão apresentar (entre um e três) seminários sobre os artigos lidos e o resto da aula será utilizado para discutir o tema.

Aprovação será avaliada através da qualidade de resenhas escritas pelos alunos sobre os artigos discutidos, da qualidade dos seminários, e da participação nas discussões. A outra metade do curso será avaliada através da execução de um projeto orientado de medição ou avaliação de alguma técnica na Internet. O aluno deverá entregar um documento de aproximadamente seis páginas descrevendo metodologia e resultados bem como realizar uma apresentação de 20 minutos sobre seu projeto.

A. Prerequisitos

O aluno deve ter cursado Redes de Computadores (DCC023) ou disciplina equivalente; em particular o aluno deve conhecer os conceitos fundamentais dos protocolos IP, UDP, DNS e protocolos de roteamento.

III. PROGRAMA

Textbooks. We will use Peterson and Davie as textbook for background information on routing and transport protocols [1], as well as other more specific references as shown below.

Design Principles. Internet architecture principles [2], [3], the end-to-end argument [4], naming and binding [5], provider competition and its impact on the Internet's architecture [6].

Unicast Routing. Intradomain distance vector and link-state routing; BGP [7].¹ Mechanisms and practices: Route redistribution [8], intradomain route dissemination [9], routing convergence [10], [11], hot-potato routing [12], flat-label routing [13]. Properties and consequences: Path inflation [14], synchronization problems [15], routing stability and oscillation [16], routing outages and causes [17], and routing granularity [18].

Switching and Forwarding. Distributed spanning trees computation [19], packet switching [20], high-speed routing [21], and optical switching [22], tunneling [23].

Transport and Congestion Control. Reliable end-to-end transmission and flow control [24], congestion avoidance and control [25]–[28], TCP throughput [29], and transport on low-latency high-throughput datacenter networks [30].

Queue Management. Router buffer sizing [31], random early detection [32], fair queueing [33], [34], queueing delay [35], explicit congestion notification [36], and differentiated services [37].

Multicast and Applications. Multicast routing [38]–[40], application-layer multicast [41], erasure codes and file distribution [42], and on-demand streaming [43].

Network Measurement and Characterization. Network characterization with network support [44], [45]. Measurement methodology [46], reverse engineering [47]. Topology mapping [48]–[52], IP

¹BGP tutorials can be found at:
<http://www.academ.com/nanog/feb1997/BGPTutorial/> and
http://www.ittc.ku.edu/EECS/EECS_800.ira/bgp_tutorial/.

aliasing [53], performance measurements [54], [55], broadband measurements [56], [57], traffic characterization [58], [59], anomaly detection [60]–[62].

AS-level Topology. Inter-domain routing relationships [63], [64], AS-level topology properties [65], AS-level topology evolution [66], AS-level topology modeling [67].

Overlays and Content Distribution. Overlay networks [68], distributed hash tables [69], P2P content distribution [70], content distribution networks [71].

Security and Privacy. Denial-of-service attacks [72], botnets [73], onion routing [74], [75], ethics [76], differentially-private analysis [77].

Alternate Architectures and Solutions. Delay-tolerant networking [78], [79], software-defined networking [80], [81], content-based networking [82], network virtualization [83], [84], exchange points [85], and extensible architectures [86].

REFERENCES

- [1] L. Peterson and B. Davie, *Computer Networks: a Systems Approach*. Morgan Kaufmann, 2012.
- [2] D. Clark, “The Design Philosophy of the DARPA Internet Protocols,” in *Proc. ACM SIGCOMM*, 1988.
- [3] S. Shenker, “Fundamental Design Issues for the Future Internet,” *IEEE J. Selected Areas in Communications*, vol. 13, no. 7, pp. 1176–1188, 2006.
- [4] J. Saltzer, D. Reed, and D. Clark, “End-to-end Arguments in System Design,” *ACM Trans. Comput. Syst.*, vol. 2, no. 4, pp. 277–288, 1984.
- [5] J. Saltzer, “On the Naming and Binding of Network Destinations,” in *IFIP Symposium on Local Computer Networks*, 1982.
- [6] D. Clark, J. Wroclawski, K. Sollins, and R. Braden, “Tussle in Cyberspace: Defining Tomorrow’s Internet,” *IEEE/ACM Trans. Netw.*, vol. 13, no. 3, pp. 462–475, 2005.
- [7] J. W. Stewart, *BGP4. Inter-Domain Routing in the Internet*. Addison-Wesley, 1998.
- [8] F. Le, G. Xie, D. Pei, J. Wang, and H. Zhang, “Shedding Light on the Glue Logic of the Internet Routing Architecture,” in *Proc. ACM SIGCOMM*, 2008.
- [9] N. Gvozdiev, B. Karp, and M. Handley, “LOUP: The Principles and Practice of Intra-Domain Route Dissemination,” in *Proc. USENIX NSDI*, 2013.
- [10] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, “Delayed Internet Routing Convergence,” *IEEE/ACM Trans. Netw.*, vol. 9, no. 3, pp. 293–306, 2001.
- [11] T. G. Griffin and G. Wilfong, “An Analysis of BGP Convergence Properties,” in *Proc. ACM SIGCOMM*, 1999.
- [12] R. Teixeira, A. Shaikh, T. Griffin, and J. Rexford, “Dynamics of Hot-potato Routing in IP Networks,” in *Proc. ACM SIGMETRICS*, 2004.
- [13] M. Caesar, T. Condie, J. Kannan, K. Lakshminarayanan, I. Stoica, and S. Shenker, “ROFL: Routing on Flat Labels,” in *Proc. ACM SIGCOMM*, 2006.
- [14] N. Spring, R. Mahajan, and T. Anderson, “The Causes of Path Inflation,” in *Proc. ACM SIGCOMM*, 2003.
- [15] S. Floyd and V. Jacobson, “The Synchronization of Periodic Routing Messages,” *IEEE/ACM Trans. Netw.*, vol. 2, no. 2, pp. 122–136, 1994.
- [16] A. Shaikh, A. Varma, L. Kalampoukas, and R. Dube, “Routing Stability in Congested Networks: Experimentation and Analysis,” in *Proc. ACM SIGCOMM*, 2000.
- [17] F. Wang, Z. M. Mao, J. Wang, L. Gao, and R. Bush, “A Measurement Study on the Impact of Routing Events on End-to-end Internet Path Performance,” in *Proc. ACM SIGCOMM*, 2006.
- [18] W. Mühlbauer, S. Uhlig, B. Fu, M. Meulle, and O. Maennel, “In Search for an Appropriate Granularity to Model Routing Policies,” in *Proc. ACM SIGCOMM*, 2007.
- [19] R. Perlman, “An Algorithm for Distributed Computation of a Spanningtree in an Extended LAN,” in *Proc. ACM SIGCOMM*, 1985.
- [20] N. McKeown, “The iSLIP Scheduling Algorithm for Input-queued Switches,” *IEEE/ACM Trans. Netw.*, vol. 7, no. 2, pp. 188–201, 1999.
- [21] C. Partridge, P. P. Carvey, E. Burgess, I. Castineyra, T. Clarke, L. Graham, M. Hathaway, P. Herman, A. King, S. Kohalmi, T. Ma, J. Mcallen, T. Mendez, W. C. Milliken, R. Pettyjohn, J. Rokosz, J. Seeger, M. Sollins, S. Storch, B. Tober, and G. D. Troxel, “A 50-Gb/s IP Router,” *IEEE/ACM Trans. Netw.*, vol. 6, no. 3, pp. 237–248, 1998.
- [22] B. Mukherjee, “WDM Optical Communication Networks: Progress and Challenges,” *IEEE J. Selected Areas in Communications*, vol. 18, no. 10, pp. 1810–1824, 2000.
- [23] J. Sommers, P. Barford, and B. Eriksson, “On the Prevalence and Characteristics of MPLS Deployments in the Open Internet,” in *Proc. IMC*, 2011.
- [24] V. Cerf and R. Khan, “A Protocol for Packet Network Intercommunication,” *IEEE Trans. on Comm.*, vol. 22, no. 5, pp. 637–648, 1974.
- [25] V. Jacobson, “Congestion Avoidance and Control,” in *Proc. ACM SIGCOMM*, 1988.
- [26] K. K. Ramakrishnan and R. Jain, “A binary feedback scheme for congestion avoidance in computer networks,” *ACM Trans. Comput. Syst.*, vol. 8, no. 2, pp. 158–181, 1990.
- [27] L. S. Brakmo, S. W. O’Malley, and L. L. Peterson, “TCP Vegas: New Techniques for Congestion Detection and Avoidance,” in *Proc. ACM SIGCOMM*, 1994.
- [28] K. Winstein and H. Balakrishnan, “TCP Ex Machina: Computer-generated Congestion Control,” in *Proc. ACM SIGCOMM*, 2013.
- [29] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, “Modeling TCP Throughput: A Simple Model and Its Empirical Validation,” in *Proc. ACM SIGCOMM*, 1998.
- [30] M. Alizadeh, A. Greenberg, D. A. Maltz, J. Padhye, P. Patel, B. Prabhakar, S. Sengupta, and M. Sridharan, “Data Center TCP (DCTCP),” in *Proc. ACM SIGCOMM*, 2010.
- [31] N. Beheshti, Y. Ganjali, M. Ghobadi, N. McKeown, and G. Salmon, “Experimental Study of Router Buffer Sizing,” in *Proc. IMC*, 2008.
- [32] S. Floyd and V. Jacobson, “Random Early Detection Gateways for Congestion Avoidance,” *IEEE/ACM Trans. Netw.*, vol. 1, no. 4, pp. 397–413, 1993.
- [33] A. Demers, S. Keshav, and S. Shenker, “Analysis and Simulation of a Fair Queueing Algorithm,” in *Proc. ACM SIGCOMM*, 1989.
- [34] I. Stoica, S. Shenker, and H. Zhang, “Core-stateless Fair Queueing: A Scalable Architecture to Approximate Fair Bandwidth Allocations in High-speed Networks,” *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 33–46, 2003.

- [35] K. Nichols and V. Jacobson, "Controlling Queue Delay," *Queue*, vol. 10, no. 5, pp. 20–34, 2012.
- [36] D. Katabi, M. Handley, and C. Rohrs, "Congestion Control for High Bandwidth-delay Product Networks," in *Proc. ACM SIGCOMM*, 2002.
- [37] B. Davie, "Deployment Experience with Differentiated Services," in *SIGCOMM Workshop on Revisiting IP QoS*, 2003.
- [38] S. E. Deering, "Multicast Routing in Internetworks and Extended LANs," in *Proc. ACM SIGCOMM*, 1988.
- [39] S. Floyd, V. Jacobson, C.-G. Liu, S. McCanne, and L. Zhang, "A Reliable Multicast Framework for Light-weight Sessions and Application Level Framing," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 784–803, 1997.
- [40] S. Ratnasamy, A. Ermolinskiy, and S. Shenker, "Revisiting IP Multicast," in *Proc. ACM SIGCOMM*, 2006.
- [41] Y.-H. Chu, S. G. Rao, S. Seshan, and H. Z., "A Case for End System Multicast," *IEEE J. Selected Areas in Communications*, vol. 20, no. 8, pp. 1456–1471, 2002.
- [42] J. W. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A Digital Fountain Approach to Reliable Distribution of Bulk Data," in *Proc. ACM SIGCOMM*, 1998.
- [43] D. L. Eager, M. K. Vernon, and J. Zahorjan, "Bandwidth Skimming: a Technique for Cost-effective Video on Demand," 1999.
- [44] D. Turner, K. Levchenko, A. Snoeren, and S. Savage, "California Fault Lines: Understanding the Causes and Impact of Network Failures," in *Proc. ACM SIGCOMM*, 2010.
- [45] A. Markopoulou, G. Iannaccone, S. Bhattacharyya, C. N. Chuah, Y. Ganjali, and C. Diot, "Characterization of Failures in an Operational IP Backbone Network," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 749–762, 2008.
- [46] V. Paxson, "Strategies for Sound Internet Measurement," in *Proc. IMC*, 2004.
- [47] N. Spring, D. Wetherall, and T. Anderson, "Reverse Engineering the Internet," *SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 3–8, 2004.
- [48] B. Augustin, B. Krishnamurthy, and W. Willinger, "IXPs: Mapped?" in *Proc. IMC*, 2009.
- [49] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP Topologies with Rocketfuel," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, 2004.
- [50] I. Cunha, R. Teixeira, D. Veitch, and C. Diot, "Predicting and Tracking Internet Path Changes," in *Proc. ACM SIGCOMM*, 2011.
- [51] E. Katz-Bassett, H. Madhyastha, V. Adhikari, C. Scott, J. Sherry, P. van Wesp, A. Krishnamurthy, and T. Anderson, "Reverse Traceroute," in *Proc. USENIX NSDI*, 2010.
- [52] H. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iPlane: an Information Plane for Distributed Services," in *Proc. USENIX OSDI*, 2006.
- [53] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, "Internet-scale IPv4 Alias Resolution with MIDAR," *IEEE/ACM Trans. Netw.*, vol. 21, no. 2, pp. 383–399, 2013.
- [54] R. Kapoor, L.-J. Chen, L. Lao, M. Gerla, and M. Y. Sanadidi, "CapProbe: A Simple and Accurate Capacity Estimation Technique," in *Proc. ACM SIGCOMM*, 2004.
- [55] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Accurate and Efficient SLA Compliance Monitoring," in *Proc. ACM SIGCOMM*, 2007.
- [56] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzer: Illuminating The Edge Network," in *Proc. IMC*, 2010.
- [57] S. Sundaresan, W. de Donato, N. Feamster, R. Teixeira, S. Crawford, and A. Pescapé, "Broadband Internet Performance: a View from the Gateway," in *Proc. ACM SIGCOMM*, 2011.
- [58] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," in *Proc. ACM SIGCOMM*, 2010.
- [59] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, and C. Diot, "Packet-level Traffic Measurements from the Sprint IP Backbone," *IEEE Network Magazine*, vol. 17, no. 6, pp. 6–16, 2003.
- [60] M. Dischinger, M. Marcon, S. Guha, K. P. Gummadi, R. Mahajan, and S. Saroiu, "Glasnost: Enabling End Users to Detect Traffic Differentiation," in *Proc. USENIX NSDI*, 2010.
- [61] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing Network-wide Traffic Anomalies," in *Proc. ACM SIGCOMM*, 2004.
- [62] F. Silveira, C. Diot, N. Taft, and R. Govindan, "ASTUTE: Detecting a Different Class of Traffic Anomalies," in *Proc. ACM SIGCOMM*, 2010.
- [63] M. Luckie, B. Huffaker, K. Claffy, A. Dhamdhere, and V. Giot-sas, "AS Relationships, Customer Cones, and Validation," in *Proc. IMC*, 2013.
- [64] L. Gao, "On Inferring Autonomous System Relationships in the Internet," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 733–745, 2001.
- [65] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang, "The (in)Completeness of the Observed Internet AS-level Structure," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 109–122, 2010.
- [66] A. Dhamdhere and C. Dovrolis, "Twelve Years in the Evolution of the Internet Ecosystem," *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1420–1433, 2011.
- [67] W. Mühlbauer, A. Feldmann, O. Maennel, M. Roughan, and S. Uhlig, "Building an AS-topology Model That Captures Route Diversity," in *Proc. ACM SIGCOMM*, 2006.
- [68] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris, "Resilient Overlay Networks," in *Proc. ACM SOSP*, 2001.
- [69] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: a Scalable Peer-to-peer Lookup Protocol for Internet Applications," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 17–32, 2003.
- [70] M. Piatek, T. Isdal, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "Do Incentives Build Robustness in Bit Torrent," in *Proc. USENIX NSDI*, 2007.
- [71] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy, "An Analysis of Internet Content Delivery Systems," in *Proc. USENIX OSDI*, 2002.
- [72] A. Hussain, J. Heidemann, and C. Papadopoulos, "A Framework for Classifying Denial of Service Attacks," in *Proc. ACM SIGCOMM*, 2003.
- [73] S. Staniford, V. Paxson, and N. Weaver, "How to Own the Internet in Your Spare Time," in *USENIX Security Symposium*, 2002.
- [74] R. Dingleline, N. Mathewson, and P. Syverson, "Tor: The Second-generation Onion Router," in *USENIX Security Symposium*, 2004.
- [75] I. Clarke, S. G. Miller, T. W. Hong, O. Sandberg, and B. Wiley, "Protecting Free Expression Online with Freenet," *IEEE Internet Computing*, vol. 6, no. 1, pp. 40–49, 2002.
- [76] M. Bailey, D. Dittrich, E. Kenneally, and D. Maughan, "The Menlo Report," *IEEE Security and Privacy*, vol. 10, no. 2, pp. 71–75, 2012.
- [77] F. McSherry and R. Mahajan, "Differentially-private Network Trace Analysis," in *Proc. ACM SIGCOMM*, 2010.
- [78] K. Fall and S. Farrell, "DTN: An Architectural Retrospective," *IEEE J. Selected Areas in Communications*, vol. 26, no. 5, pp. 828–836, 2008.

- [79] K. Fall, "A Delay-tolerant Network Architecture for Challenged Internets," in *Proc. ACM SIGCOMM*, 2003.
- [80] M. Casado, M. J. Freedman, J. Pettit, J. L., N. Gude, N. McKeown, and S. Shenker, "Rethinking Enterprise Network Control," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1270–1283, 2009.
- [81] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, and A. Vahdat, "B4: Experience with a Globally-deployed Software Defined Wan," in *Proc. ACM SIGCOMM*, 2013.
- [82] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking Named Content," in *Proc. ACM CoNEXT*, 2009.
- [83] T. Anderson, L. Peterson, S. Shenker, and J. Turner, "Overcoming the Internet Impasse Through Virtualization," *IEEE Computer*, vol. 38, no. 4, pp. 34–41, 2005.
- [84] Z. Ben-Houidi and M. Meulle, "A New VPN Routing Approach for Large Scale Networks," in *Proc. ICNP*, 2010.
- [85] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger, "Anatomy of a Large European IXP," in *Proc. ACM SIGCOMM*, 2012.
- [86] D. Han, A. Anand, F. Dogar, B. Li, H. Lim, M. Machado, A. Mukundan, W. Wu, A. Akella, D. G. Andersen, J. W. Byers, S. Seshan, and P. Steenkiste, "XIA: Efficient Support for Evolvable Internetworking," in *Proc. USENIX NSDI*, 2012.