

Project for a ®Evolution in Data Network Routing: the Kleinrock Universe and Beyond

Dima Krioukov

[<dima@krioukov.net>](mailto:dima@krioukov.net)

Midnight Sun Routing Workshop

June 18, 2002

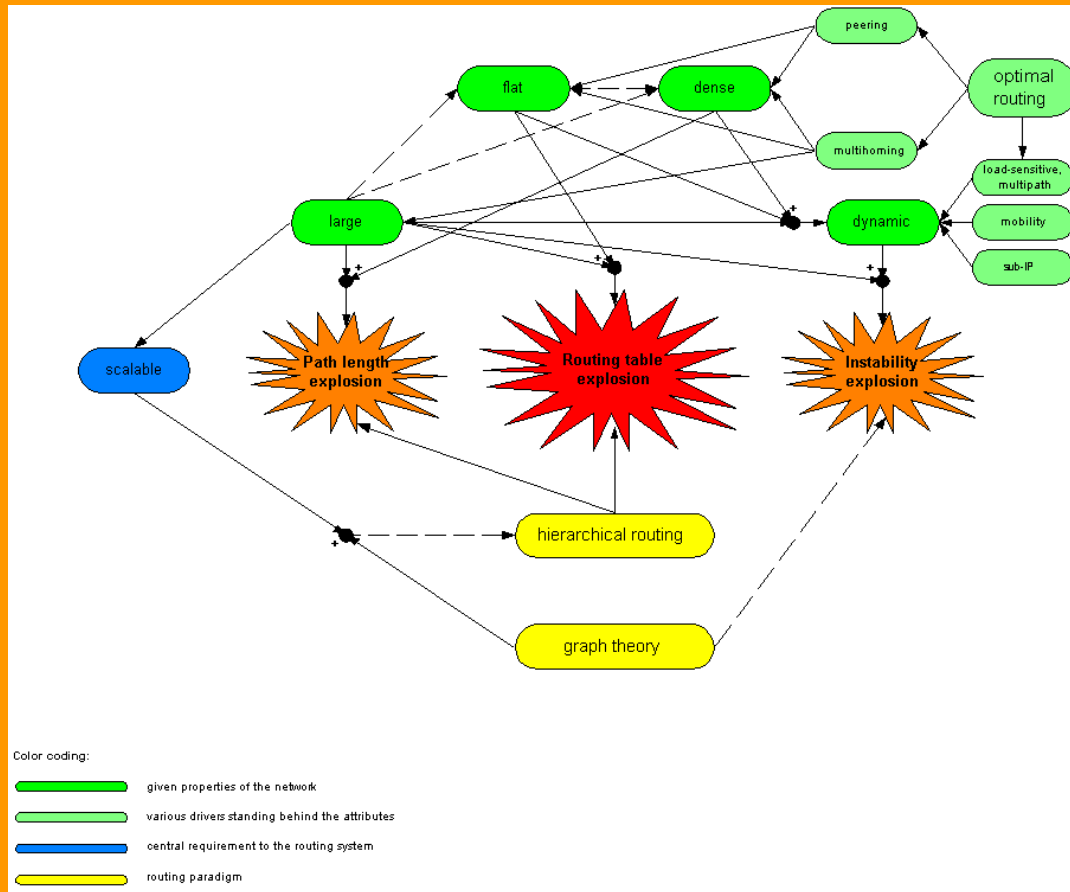
Outline

- **Present**
- **Past**
- **Future**

Outline

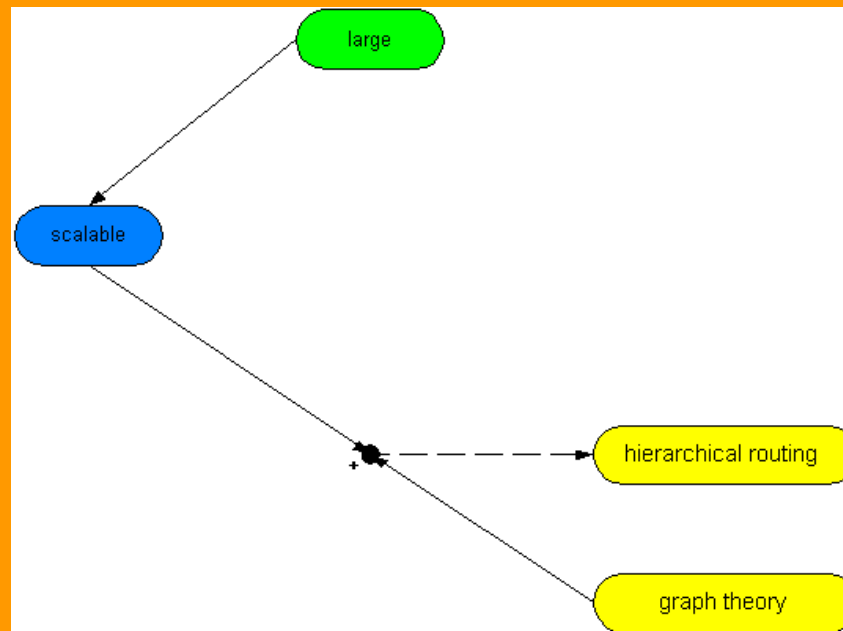
- **Present**
- **Past**
- **Future**

Present picture



Present routing paradigm

- **Network is modeled as a graph** ⇒
- **Topology information exchange and asynchronous distributed computation**
- **Scalability is the central requirement for large networks** ⇒
- **Information hiding is inevitable**
- **Hierarchical routing (areas and aggregation/abstraction) is the only *known* way of doing this**



Present picture of the Internet interdomain topology

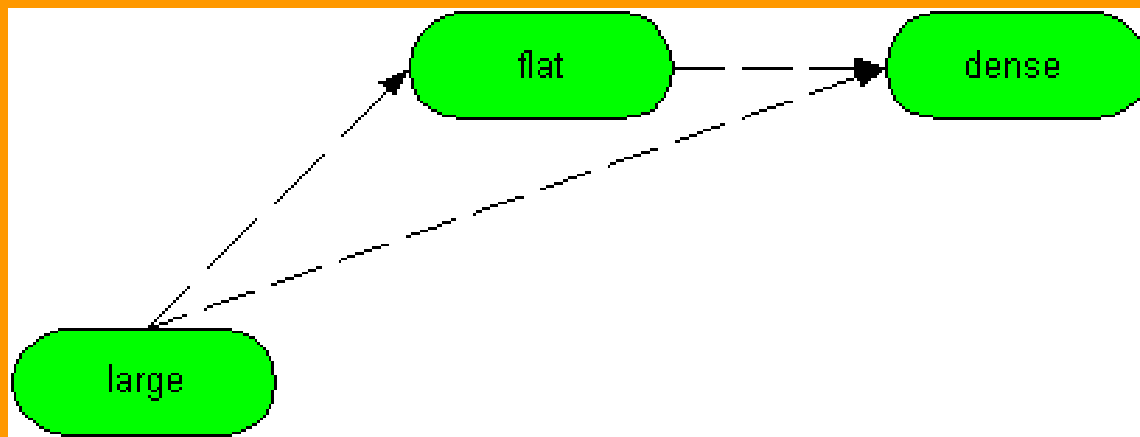
- **Why Internet?**

- **Because it's large**

- **Why interdomain?**

- **Split between what one can and cannot control will always be there; our task is to find scalable routing between islands of *independent* control**
- **No single point of full and strict external control
⇒ intrinsic properties of data network evolutionary dynamics (defined by data network design principles) exhibit themselves there first (“emerging behavior”)**

Large network with flat and densely meshed topology

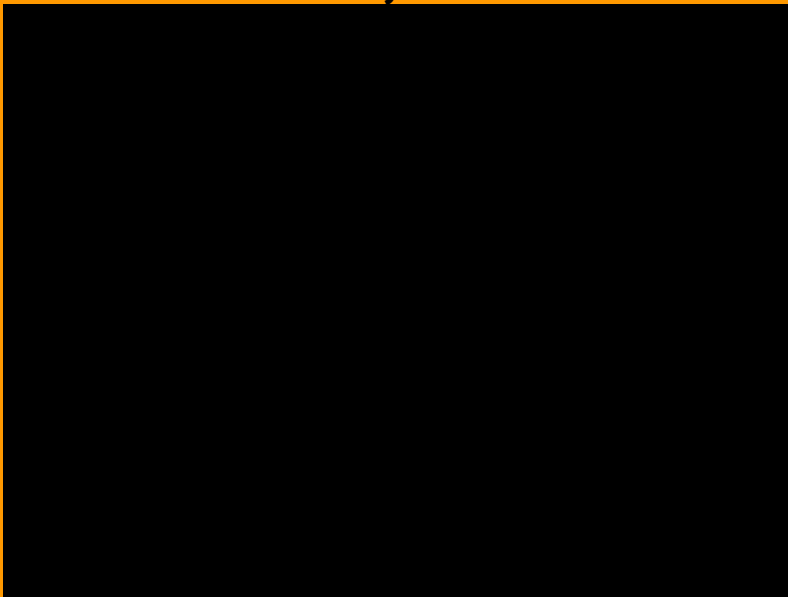


Completely flat topologies

In random/exponential networks, $P_d(k) \sim \langle d \rangle$ and exponentially drops around this value (Poisson distribution)

- **Sparse topology**

- $\langle d \rangle \ll N$
- $\langle h \rangle \sim N^\alpha$, where $\alpha \sim 1/2$



- **Dense topology**

- $\langle d \rangle \sim N$
- $\langle h \rangle \sim 1$ ($\alpha \sim 0$)



Not-so-flat topologies

In scale-free/power-law networks, $P_\gamma(k) \sim k^{-\gamma}$

- **Hub-and-spoke topology**
 - $\langle d \rangle \ll N$, but $P_\gamma(k)$ for large k is greater than in the exponential case
 - $\langle h \rangle \sim N^\alpha$, where $\alpha \ll 1$



Power law distribution as an emerging phenomenon

- **Examples of scale-free networks**
 - **Internet ($\gamma_{AS} = 2.2$, $\gamma_{router} = 2.5$)**
 - **WWW ($\gamma_{in} = 2.1$, $\gamma_{out} = 2.4$)**
 - **Airport networks**
 - **Bio-cell metabolic process diagrams**
- **Using the formalism of statistical mechanics, it was formally shown that the power law distribution emerges from these two assumptions about network evolutionary dynamics:**
 - **Addition of nodes**
 - **Preferential attachment**

Real Internet interdomain topology deviates slightly from the power law

- **Not only additions of nodes, but also deletions of nodes and additions&deletions of links**
- **Edges are directed by customer-provider relationships, which are very non-symmetric (90% of ASes are customer ASes)**

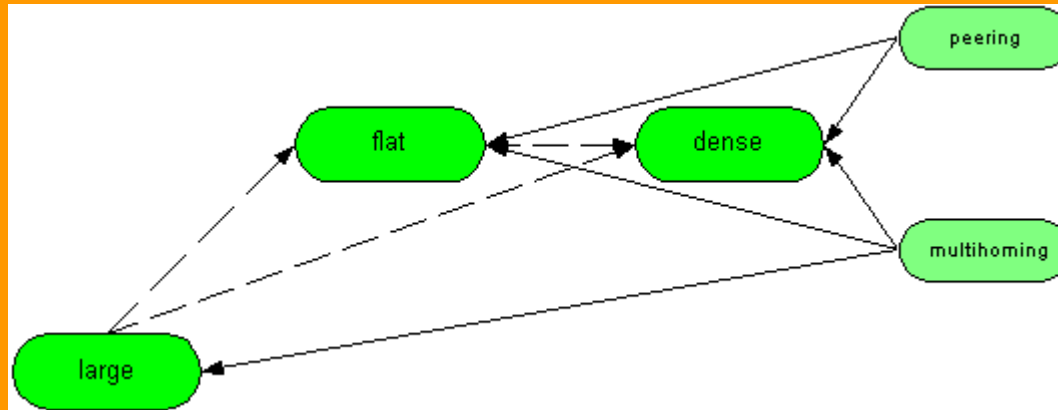
Thorough studies of the interdomain topology

- **Five classes of ASes that can be split in the two groups:**
 - **Core**
 - **Very dense part - almost a full mesh ($d_{\min} = N/2 \Rightarrow h_{\max} = 2$)**
 - **Transit part**
 - **Outer part**
 - **Shell**
 - **Customers**
 - **Regional ISPs**
- **The core is flattening and getting denser (in 2001: 25% growth of the total number of ASs, but the average AS path length was steady) \Rightarrow**
- **Tendency towards a very densely meshed core of provider ASes and a shell of customer ASes**
- **Less and less strict hierarchy in connectivity across the AS classes**

The blue points are analyzed in more detail

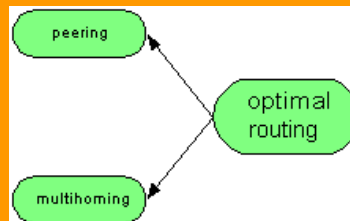
Drivers for flat and dense mesh

- **Peering and multihoming**



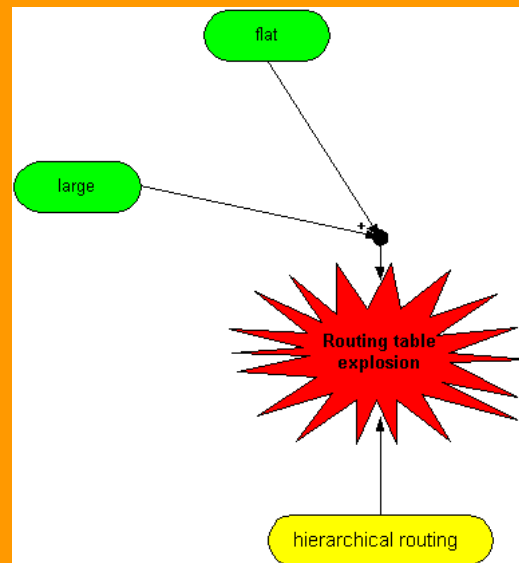
Drivers for peering and multihoming

- **Why more peering and multihoming recently?**
- **Because it became cheaper**
- **Still, why would one want to peer and multihome?**
- **Peering**
 - **Routing cost reduction (e.g. avoid transit costs)**
 - **More optimal routing**
 - **Higher resilience and routing flexibility**
- **Multihoming**
 - **Higher resilience**
 - **More optimal routing**
- **Optimal routing is min-cost routing, where cost model is a variable (by default: shortest delay \Rightarrow by default: shortest path); everything above fits this generic definition**
- **In summary: *optimal routing***
 - **Why does not this fundamental cause break strict hierarchies of PSTN connectivity topologies?**
 - **Because they are circuit-switched—in circuit-switched networks, delay does not depend that strongly on the number of switching nodes in a data path (no queuing!)**



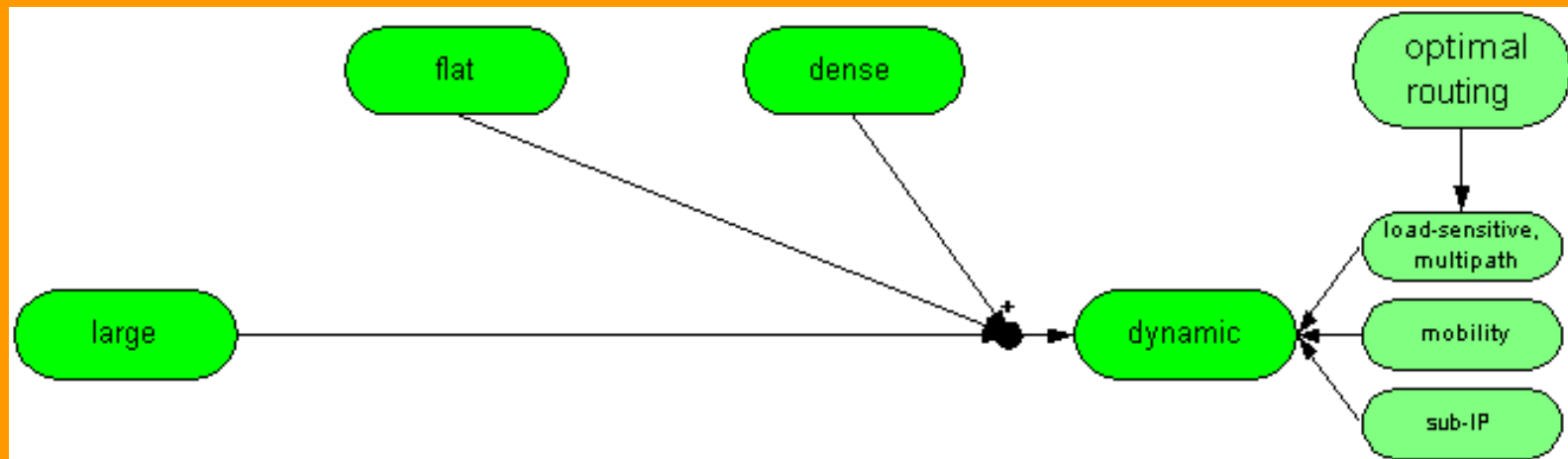
Explosive #1

- **Routing table size**
- **Might not the problem be fixed by a good routing architecture?**
- **The answer is in explosive #3**



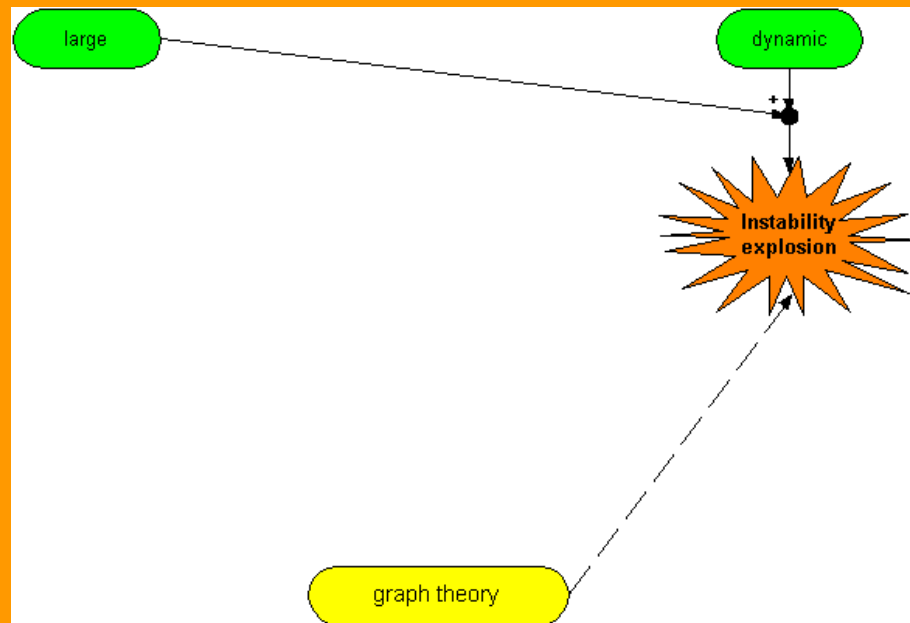
Explosive #2 next

Dynamic “topology”



Explosive #2

- **Instabilities**



Understanding of explosive #3 lies in the past

Outline

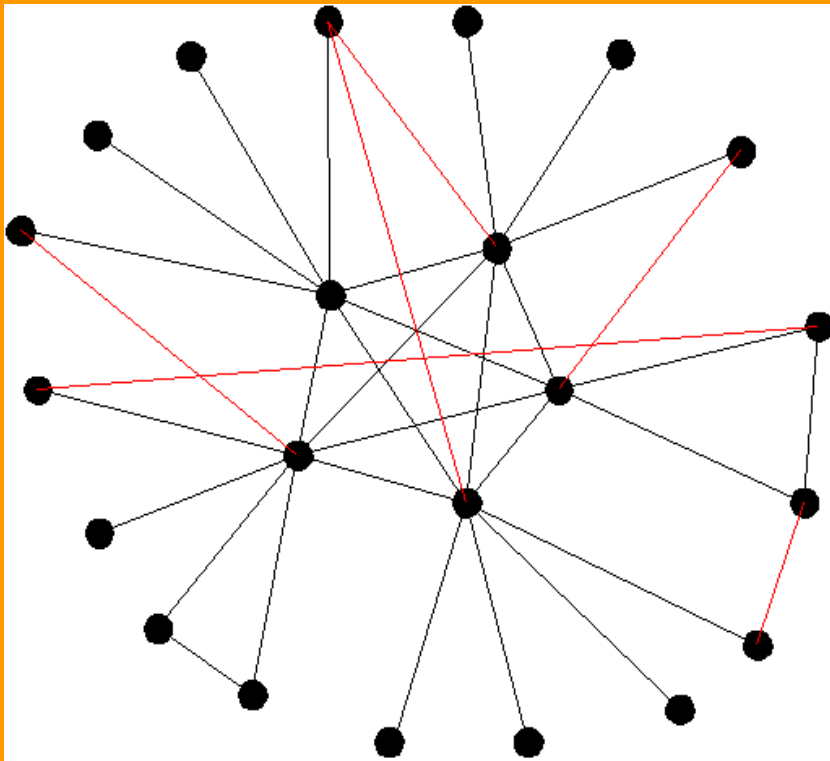
- **Present**
- **Past**
- **Future**

Is it future in the past or past in the future?

- **But what about present?**
- **Present discussions/ideas/proposals \Leftarrow Nimrod \Leftarrow L. Kleinrock and F. Kamoun (K&K)**
- **Small routing table for arbitrary topologies**
- **But...**
 - **Hierarchical topologies (cf. slide 12)**
 - **Path length increase**

Hierarchical topology

- **No strict hierarchy**

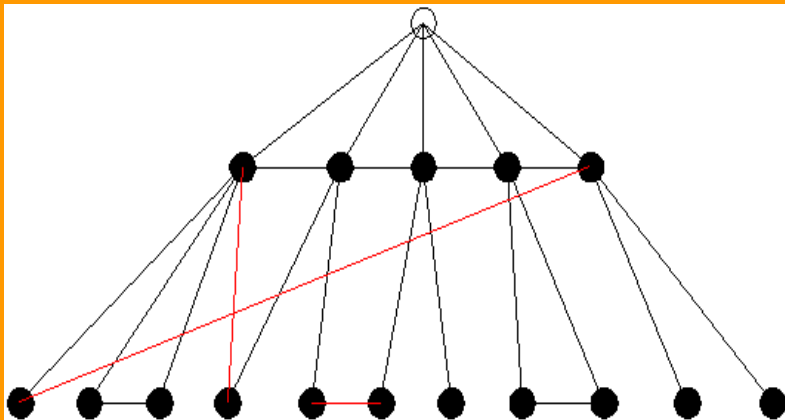


- **Strict hierarchy**



Hierarchical addressing

- **No strict hierarchy**



- **Strict hierarchy**



Hierarchical topologies and K&K

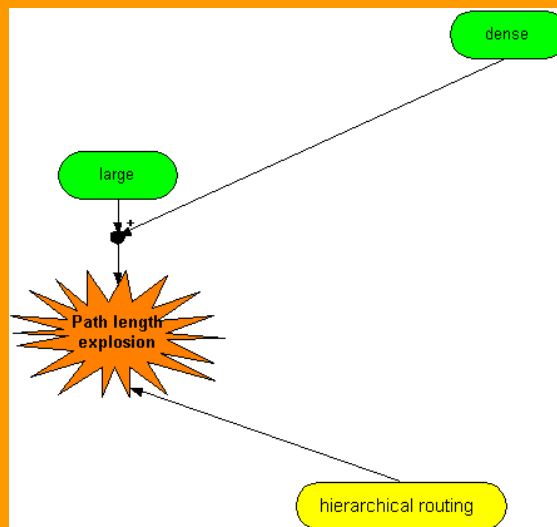
- **To be able to analyze any realistic characteristics of paths produced by their scheme, K&K need to assume (among other things) that:**
 - **Any pair of nodes in an area at any level of hierarchy are connected by a path lying completely in that area**
 - **The shortest path between any pair of nodes also lies within the area**
- **Hence, a hierarchical topology induces a specific structure of hierarchical addressing (as expected)**
- **The second assumption is not really necessary for one's being able to analyze the K&K path characteristics but the resulting path characteristics are much worse without it than with it**

K&K path characteristics

- **Increased average length (cf. slide 19); that is, *less optimal* routing**
- **Analytically**
 - **If $\langle h \rangle \sim N^\alpha$, $E = \langle h_{\text{K&K}} \rangle / \langle h \rangle - 1$, then $E = E(N, \alpha)$**
 - **The exact form of $E(N, \alpha)$ is somewhat complex; the two of its limits are (α is a measure of density of connectivity):**
 - **$E(N = \infty, \alpha \rightarrow 0) \rightarrow 1/\alpha$**
 - **$E(N \rightarrow \infty, \alpha = 0) \rightarrow \ln(N)$**
 - **The average path length *increase* is *unbounded***
- **Practically**

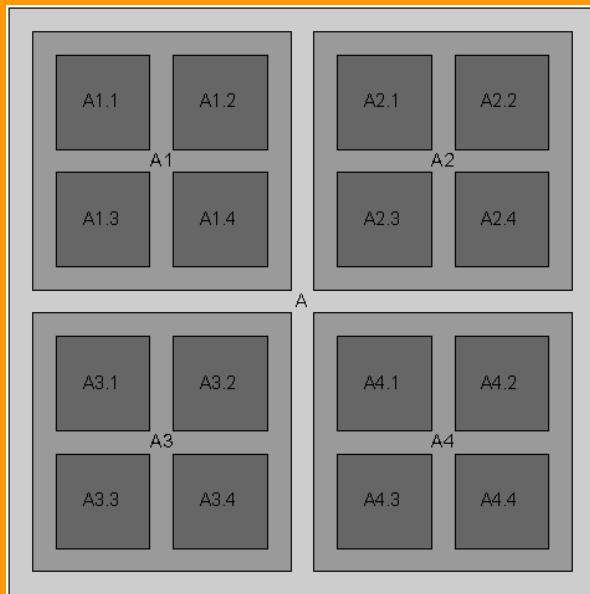
Explosive #3: K&K path length increase for the Internet

- $3 \leq \langle h \rangle \leq 4 \Rightarrow \langle h \rangle \sim e, N \sim 10^4 \Rightarrow$
- $\alpha \sim 10^{-1} \Rightarrow$
- $E \sim 10 \Rightarrow$
- $\langle h_{K\&K} \rangle$ is ~ 10 (6 in the most optimistic calculations) times longer than $\langle h \rangle \Rightarrow$
- $30 \leq \langle h_{K\&K} \rangle \leq 40$ (in AS hops \Rightarrow hundreds of IP hops!)

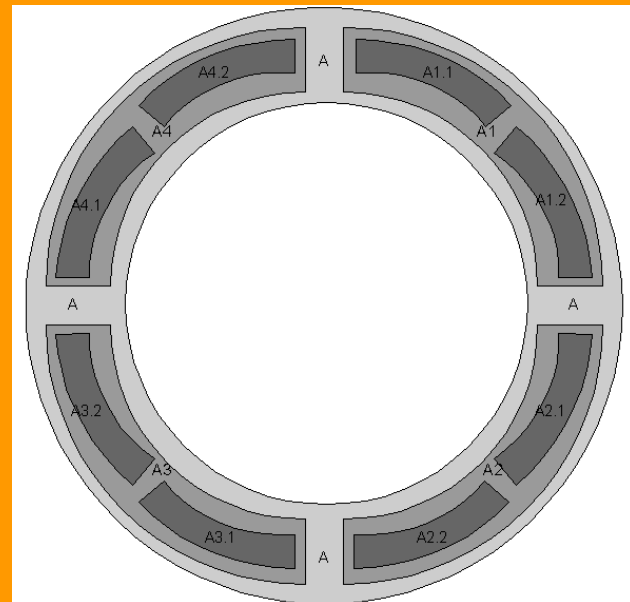


K&K path length increase for dense topologies is intuitively expected

- Area organization on a sparse topology
- $\langle h \rangle \rightarrow \infty$, $\langle h_{K\&K} \rangle \rightarrow \infty$ so that $\langle h_{K\&K} \rangle / \langle h \rangle \rightarrow 1$
- There are remote points



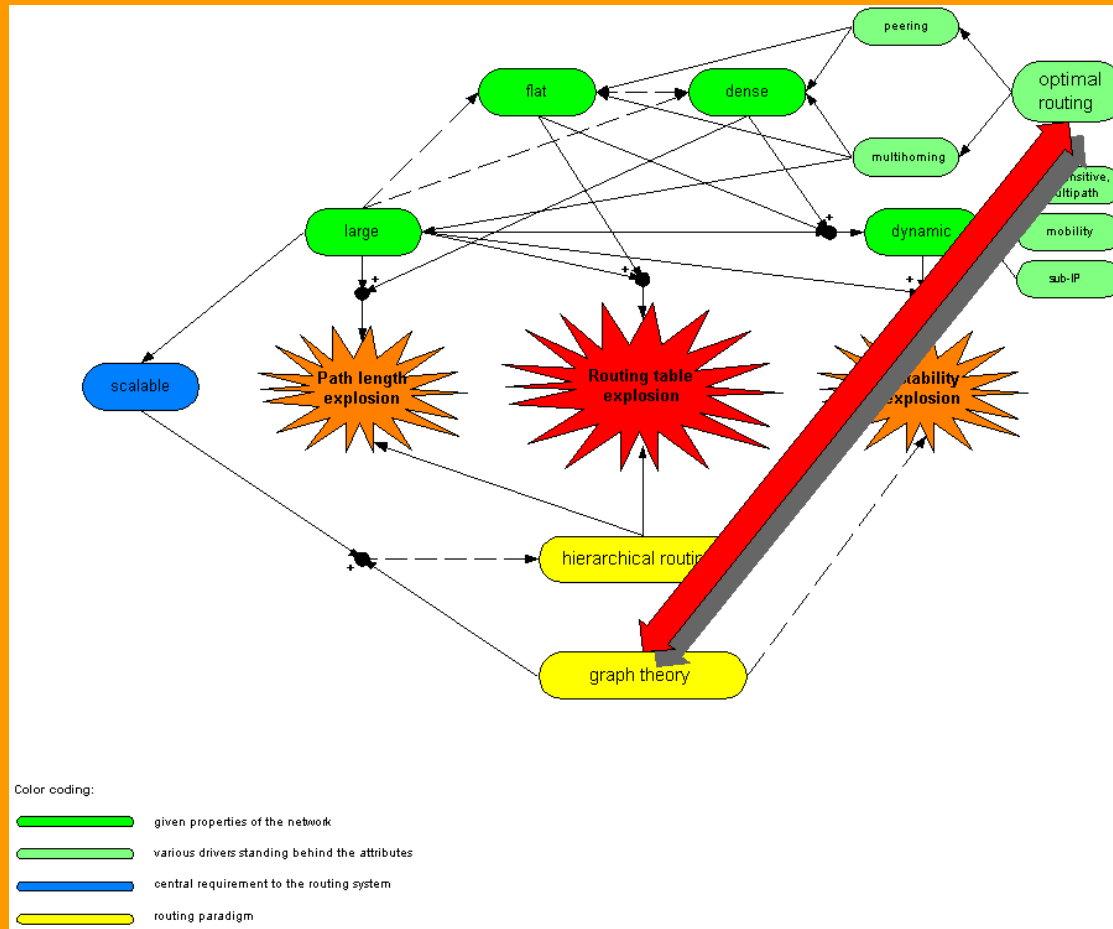
- Area organization on a dense topology
- $\langle h \rangle$ is steady ($\langle d \rangle \rightarrow \infty$ instead) but $\langle h_{K\&K} \rangle \rightarrow \infty$ so that $\langle h_{K\&K} \rangle / \langle h \rangle \rightarrow \infty$
- There are no remote points, so that one cannot usefully aggregate, abstract, etc., anything remote—everything is close



No path length increase is allowed in reality

- If two ASes peer, then they do so to exchange traffic over the link (subject to their policies); one has to consider this as an integer constraint to the routing system, as a requirement
- If this link violates an imposed hierarchical structure (a red link), then it's a "hole" in the hierarchy leading to an extra routing table entry—an extra portion of topological information being propagated at the higher (than intended) levels of hierarchy
- When the total size (strict portion + "red" portion) of the hierarchical routing table becomes comparable with the size of the non-hierarchical routing table, the value of hierarchical routing drops to zero
- In the extreme example of a fully meshed network (cf. previous slide), the non-hierarchical routing table size is $\sim N$, the hierarchical one is $\sim \ln(N)$, but if optimal routing is a constraint, then the total hierarchical routing table size is $\sim \ln(N) + N$; that is, hierarchical routing, not bringing any benefit, just *increases* the routing table size

Collecting all pieces together: a satellite photo



A little dip in philosophy

● Left keywords

- **Hierarchy**
- **Order**
- **Circle**
- **Top-down**
- **Planned**
- **Controlled**
- **Reductionism**
- **The Mind**
- **Mathematics**

● Right keywords

- **Anarchy**
- **Chaos**
- **Fractal**
- **Bottom-up**
- **Self-organizing**
- **Self-governing**
- **Emergence**
- **The Nature**
- **Physics**

Outline

- **Present**
- **Past**
- **Future**

Routing research program

- **Practical/engineering research subprogram**
- **Theoretical/fundamental research subprogram**

Engineering (re)search subprogram: no paradigm shift

- **The task at hand is a new *Internet* routing architecture**
- **However, the problem is fundamental and cannot be solved within the present routing paradigm; therefore, all potential IxTF solutions seem to be temporary (e.g. PTOMAINE—shorter term, RRG—longer term (hopefully)), although a formal proof is still needed**
- **For example, routing on AS numbers (as the first step, AS numbers (and their K&K-like aggregates) become addresses, IP addresses become just src/dst tags)**

Routing on AS numbers

● Pros

- **A very simple and straightforward thing to do; in fact, this whole talk discusses a situation where it's already done!**
- **Routing table size reduction is ~10 times (10^5 IP prefixes but 10^4 ASes), and all associated consequences (higher stability, etc.)**

● Cons

- **This whole talk discusses a situation where it's already done! Given the interdomain topology structure and its evolutionary trends, it is impossible to usefully aggregate anything at and above the current AS level of hierarchy**
- **The proposal does not solve anything, it just shifts the problem to another level (winning some time, though)—tomorrow's AS numbers might pretty quickly obtain the semantics of today's IP addresses (ASes from the customer shell requiring ~1 public IP address but connecting to a number of ASes from the provider core with distinct routing policies—AS number-IP address 1-to-~1 correspondence)**

List of engineering problems

- **Given the split between the customer AS shell and the provider AS core, can a hierarchical scheme utilizing it be devised?**
- **Search for other hierarchical schemes that would solve the problem and that would not conflict with the tendencies rooted in optimal routing**
- **The same for *non*-hierarchical schemes**
- **Can the “flat/dense” tendencies be fought against (e.g. “multihomers should pay”)?**

Theoretical research subprogram: problems within the present paradigm

- **“Barabasi++” studies: evolutionary dynamics of data networks with more significant insight on data networks specifics ⇒ a formal demonstration of the “flat/dense” tendencies (dotted lines between the “large” and “flat/dense” boxes on the diagram)**
- **Having a theoretical answer above, can the “flat/dense” tendencies be undermined at the fundamental level (e.g. by modifications to the cost models for optimal routing in data networks); one of interesting sub-problems is a theoretical comparison with circuit-switched networks, where delay does not depend on the number of switching nodes and, hence, strict hierarchies of connectivity are possible**
- **A formal proof that a hierarchical scheme from the previous slide does or does not exist (problem: conflict with topology)**
- **The same for a non-hierarchical scheme (problem: information hiding—dotted line between the “scalable” and “hierarchical routing” boxes on the diagram)**

Theoretical research subprogram: paradigm shift

- **The proposed first step is to review potentially relevant areas of the current academic research—a set of chapters, each chapter including:**
 - **Introduction to and description of the research area in a reasonably accessible form**
 - **The most important recent results and current problems (internal to the research area)**
 - **The history of the research—how it was originated, what initially perceived problems it was to solve**
 - **Interdisciplinary aspects (if any)**
 - **Data network (in general) and Internet (in particular) routing applicability considerations:**
 - **Why the chapter is included in the review**
 - **No chapter is expected to describe a ready solution—what problem(s) must be solved within the research area for it to be applicable to what degree**
 - **Check against the requirements with a special emphasis on scalability**
 - **Attempt to estimate complexity levels of these problems (the chapter should not be included if there are any strong reasons to believe that the problems cannot be solved *in principle*)**
 - **If the problems get solved, attempt to estimate complexity levels of associated engineering and operational efforts**

Proposed chapters (cf. the references)

- **Control theory and related areas:**
 - **Q-routing, reinforcement learning (RL), collective intelligences (COINs), neuro-dynamic programming (NDP)**
 - **Game theoretical approaches**
- **Bio-networks, adaptive routing, application routing, active networks, etc.**
- **Packet routing and queuing theories**
- **Routing in mobile ad-hoc networks (?)**
- **...**
- **Physical routing**

Physical routing: the ball-and-string model as an initial example



- **Given: a graph with links of the shown costs**
- **Find: the shortest path tree with root R**



- **Given: a set of heavy balls connected by inelastic strings of the shown lengths**
- **Find: the equilibrium state when the system is left to hang suspended at ball R**

The ball-and-string system is a computer



- **Computation complexity is $O(|E|+|V|\log(|V|))$ (with Fibonacci heaps as priority queues)**

- **Computation complexity is $O(L_{\max})$**

The two problems are equivalent

- **The both problems are minimization problems:**
 - **The shortest path problem is equivalent to the min-cost flow problem: find the minimum cost flow subject to the constraints imposed by the graph**
 - **The ball-and-string system: find the minimum potential energy of the system in the uniform scalar field (the gravitational field) subject to the constraints imposed by the strings**
- **The standard mathematical formalism used to solve minimization problems (in mathematics, theoretical physics, as well as many network optimization problems) is the Lagrangian formalism**
- **The reason why the two problems are equivalent is that their *Lagrangians are equivalent***
- **There are other similar examples (e.g. the Maxwell electromagnetic energy minimization problem for a linear resistive circuit satisfying Kirchhoff's and Ohm's law is an example of the equilibrium theorem for the network optimization problem for networks with generic convex cost functions)**

The physical routing problem

- **Find a physical system with the Lagrangian equivalent to the Lagrangian of the data network routing problem \Rightarrow inherent scalability as opposed to almost all other paradigm-shifting proposals**
- **Motivation: the Lagrangian of the data network routing problem is similar to many Lagrangians in theoretical physics (the scalar field theory, in particular)**
- **Minor differences:**
 - **Continuous (physics) vs. discrete (networks)—the continuous shortest path problem is known**
 - **“Material” (field, liquid, etc.) flow (physics) vs. information flow (data networks)—information flow can be represented by propagation of field strength alterations**
- **Major difference(s):**
 - **Single commodity (physics) vs. multicommodity (data networks)—commodities are defined by source-destination pairs—no direct analogy in physics**

A proposed research program on physical routing

- **Find a continuous form of the data network Lagrangian function**
 - **If impossible, work with discrete forms of Lagrangians of physical systems**
- **Perform an analytical comparison of the Lagrangian functions for data networks and for various physical systems including systems naturally appearing in:**
 - **theoretical mechanics**
 - **scalar field theory**
 - **tensor field theory**
 - **quantum versions of the above**
 - **...**
- **Given the results of the analysis, try to find any correlations indicating how some known physical system might be modified so that its Lagrangian becomes “closer” or equivalent to the data network Lagrangian**
- **The research methodology would probably borrow from the methodology that led to discoveries of quantum computing, biological computing, etc.**

Summary

- **Certain fundamental problems/conflicts in data network routing seem to start exhibiting themselves in the Internet**
- **Formal proofs are needed of how profound those problems really are**
- **The proofs and associated research would provide deeper insight on what (temporary) engineering solutions might be and how much time is really left before a paradigm shift**
- **It is better to start preparing for a paradigm shift now**

References

- **BGP statistics and Internet interdomain topology**
 - “BGP Table Data,” <http://bgp.potaroo.net/>
 - “The Skitter Project,” <http://www.caida.org/tools/measurement/skitter/>
 - S. Agarwal, L. Subramanian, J. Rexford, and R. H. Katz, “Characterizing the Internet hierarchy from multiple vantage points,” *IEEE Infocom*, 2002, <http://www.cs.berkeley.edu/~sagarwal/research/BGP-hierarchy/>
- **Network evolutionary dynamics**
 - R. Albert and A.-L. Barabasi, “Statistical mechanics of complex networks,” *Reviews of Modern Physics* 74, 47 (2002), <http://www.nd.edu/~networks/PDF/rmp.pdf>
 - “Study of Self-Organized Networks at Notre Dame,” <http://www.nd.edu/~networks/>

References (contd.)

- **Hierarchical routing**

- **L. Kleinrock and F. Kamoun, “Hierarchical routing for large networks: Performance evaluation and optimization,”** *Computer Networks*, vol. 1, pp. 155-174, 1977, <http://www.cs.ucla.edu/~lk/LK/Bib/PS/paper071.pdf>
- **P. Tsuchiya, “The landmark hierarchy: A new hierarchy for routing in very large networks,”** *Computer Commun. Rev.*, vol 18, no. 4, pp. 43-54, 1988
- **J. J. Garcia-Luna-Aceves, “Routing management in very large-scale networks,”** *Future Generation Computer Systems*, North-Holland, vol. 4, no. 2, pp. 81-93, 1988
- **I. Castineyra, N. Chiappa, and M. Steenstrup, “The Nimrod routing architecture,”** *RFC 1992*, August 1996, <http://ana-3.lcs.mit.edu/~jnc/nimrod/docs.html>
- **P. Tsuchiya, “Pip,”** <http://www.watersprings.org/pub/id/draft-tsuchiya-pip-00.ps>, <http://www.watersprings.org/pub/id/draft-tsuchiya-pip-overview-01.ps>
- **F. Kastenholz, “ISLAY,”** <http://partner.unispherenetworks.com/rrg/draft-irtf-routing-islay-00.txt>

References (contd.)

- **Control theory and derivatives**
 - **D. Bertsekas, *Dynamic Programming and Optimal Control*, Athena Scientific, 2000-2001, <http://www.athenasc.com/dpbook.html>**
 - **D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999, <http://www.athenasc.com/nonlinbook.html>**
 - **D. Bertsekas and J. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996, <http://www.athenasc.com/ndpbook.html>**
 - **J. Boyan and M. Littman. “Packet routing in dynamically changing networks: A reinforcement learning approach,” *Advances in Neural Information Processing Systems*, vol. 6, pp. 671-678, 1993, <http://www.cs.duke.edu/~mlittman/topics/routing-page.html>**
 - **D. Wolpert, K. Tumer, and J. Frank, “Using collective intelligence to route Internet traffic,” *Advances in Neural Information Processing Systems-11*, pp. 952-958, 1998, <http://ic.arc.nasa.gov/ic/projects/COIN/>**

References (contd.)

- **Game theory**
 - **R. La and V. Anantharam, “Optimal routing control: Game theoretic approach,” *IEEE Conference on Decision and Control*, 1997, <http://citeseer.nj.nec.com/la97optimal.html>**
 - **Y. Korilis, A. Lazar, and A. Orda, “Achieving network optima using Stackelberg routing games,” *IEEE Transactions on Networking*, vol. 5, no. 1, pp. 161-173, 1997, http://comet.columbia.edu/~aurel/papers/networking_games/stackelberg.pdf**
- **“Mobile ad-hoc networks (MANET),” <http://www.ietf.org/html.charters/manet-charter.html>**
 - **E. Royer and C.-K. Toh, “A review of current routing protocols for ad-hoc mobile wireless networks,” *IEEE Personal Communications Magazine*, pp. 46-55, April 1999, <http://alpha.ece.ucsb.edu/~eroyer/txt/review.ps>**

References (contd.)

- **Bio-nets, adaptive routing, application routing, active networks, etc.**
 - **G. Di Caro and M. Dorigo, “An adaptive multi-agent routing algorithm inspired by ants behavior,” *Proc. PART98 - Fifth Annual Australasian Conference on Parallel and Real-Time Systems*, 1998, <http://dsp.jpl.nasa.gov/members/payman/swarm/>**
 - **“Bio-Networking Architecture,” <http://netresearch.ics.uci.edu/bionet/>, and related works, <http://netresearch.ics.uci.edu/bionet/relatedwork/index.html>;**
application/content/peer-to-peer routing, in particular:
 - **S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, “A scalable content-addressable network,” *Proc. of SIGCOMM*, ACM, 2001, <http://citeseer.nj.nec.com/ratnasamy01scalable.html>**
 - **S. Joseph, “NeuroGrid,” <http://www.neurogrid.net/>**
 - **“Active Networks,” <http://nms.lcs.mit.edu/darpa-activenet/>**

References (contd.)

- **Packet routing and queuing theories**
 - **A. Borodin, J. Kleinberg, P. Raghavan, M. Sudan, and D. Williamson, “Adversarial queuing theory,” *Proc. ACM Symp. on Theory of Computing*, pp. 376-385, 1996, <http://citeseer.nj.nec.com/472505.html>**
 - **C. Scheideler and B. Vocking, “From static to dynamic routing: Efficient transformations of store-and-forward protocols,” *Proc. of the 31st ACM Symp. on Theory of Computing*, pp. 215–224, 1999, <http://citeseer.nj.nec.com/scheideler99from.html>**
 - **B. Awerbuch, P. Berenbrink, and A. Brinkmann, Christian Scheideler, “Simple routing strategies for adversarial systems,” *Proc. IEEE Symp. on Foundations of Computer Science*, 2001, <http://citeseer.nj.nec.com/awerbuch01simple.html>**

References (contd.)

- **Physical routing (starting points)**
 - **D. Bertsekas, *Network Optimization: Continuous and Discrete Models*, Athena Scientific, 1998, <http://www.athenasc.com/netbook.html>**
 - **Ball-and-string model**
 - **G. J. Minty, “A comment on the shortest route problem,” *Operations Research*, vol. 5, p.724, 1957**
 - **Multicommodity flow problem**
 - **“Multicommodity Problems,” <http://www.di.unipi.it/di/groups/optimize/Data/MMCF.html>**
 - **B. Awerbuch and T. Leighton, “Improved approximation algorithms for the multicommodity flow problem and local competitive routing in dynamic networks,” *Proc. ACM Symp. on Theory of Computing*, 1994, <http://citeseer.nj.nec.com/awerbuch94improved.html>**
 - **R. D. McBride, “Advances in solving the multicommodity flow problem,” *SIAM J. on Opt.* 8(4), pp. 947-955, 1998**
 - **T. Larsson and D. Yuan, “An augmented Lagrangian algorithm for large scale multicommodity routing,” *LiTH-MAT-R-2000-12*, Linkopings Universitet, 2000**

Thank you!