Master Course
Computer Networks
IN2097

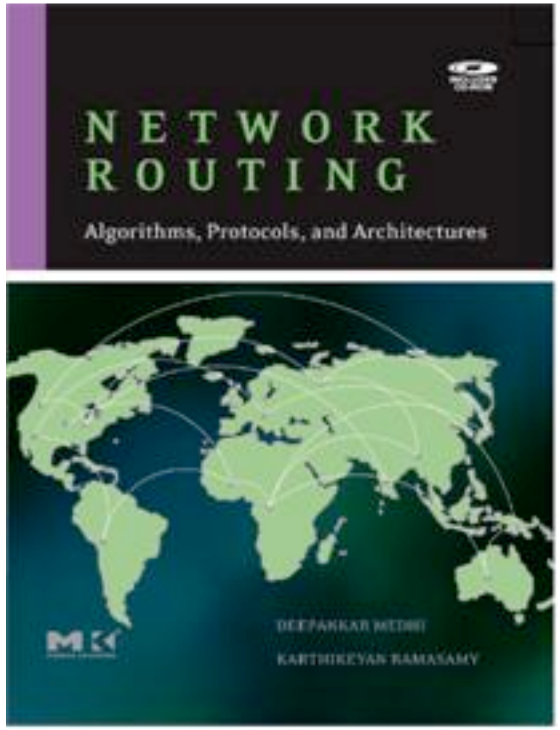
Prof. Dr.-Ing. Georg Carle

Chair for Network Architectures and Services
Department of Computer Science
Technische Universität München
http://www.net.in.tum.de

Time Plan

- This Week
  - Monday, 16. 12. 2013 – Lecture
  - Tuesday, 17. 12. 2013 – Exercise
- Next Week
  - Monday, 23. 12. 2013 – No lecture (teamwork on project)
- Christmas break
- 2014
  - Tuesday, 7. 1. 2014 – Lecture (first lecture in new year)

- Book
  - Network Routing
  - Authors: DEEP MEDHI
    and KARTHIK RAMASAMY
  - Morgan Kaufmann Publishers
  - http://www.networkrouting.net/
  - Chapter 8: BGP



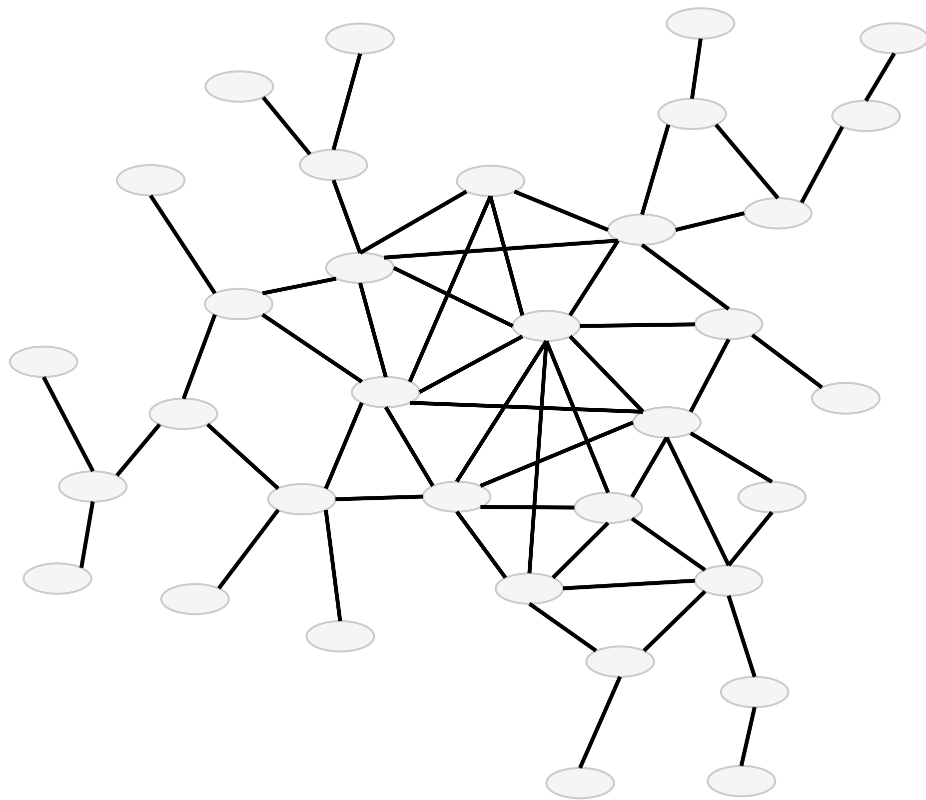
Outline

- Interdomain Routing (cont.)
  - BGP Analysis
    - BGP anomalies and hijacking detection
  - Business considerations
    - Traffic engineering
- Multicast Routing

BGP Analysis

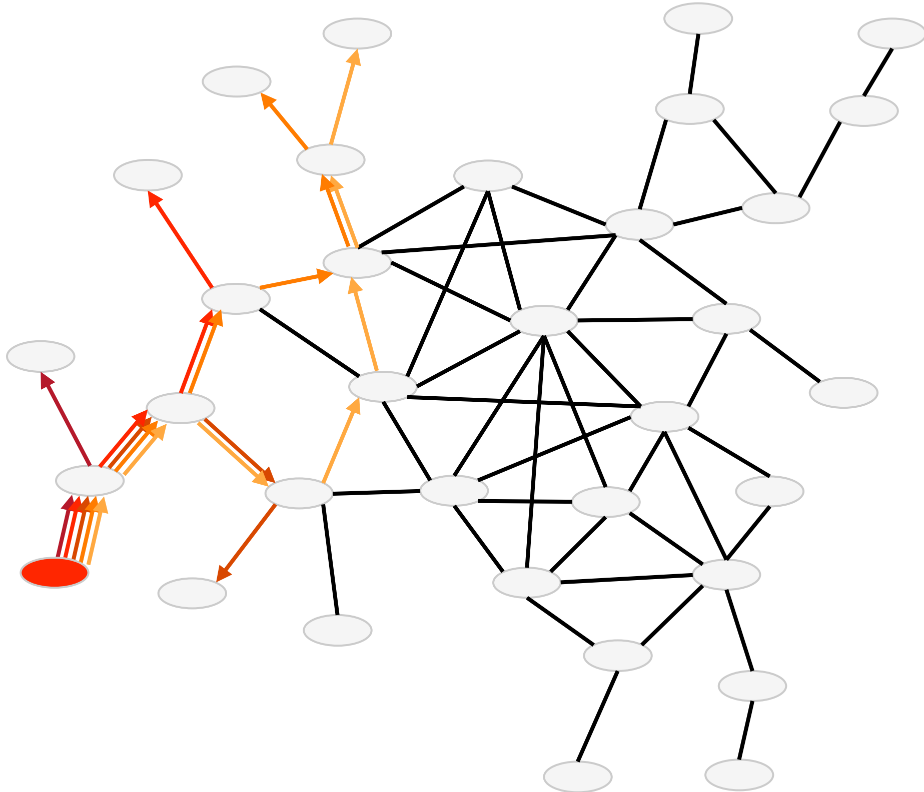
BGP Path Analysis

- Graph analysis
  - ASes as nodes
  - Links in AS path als edges
  - „Snapshot“ of Internet routes
  - Router-specific viewpoint



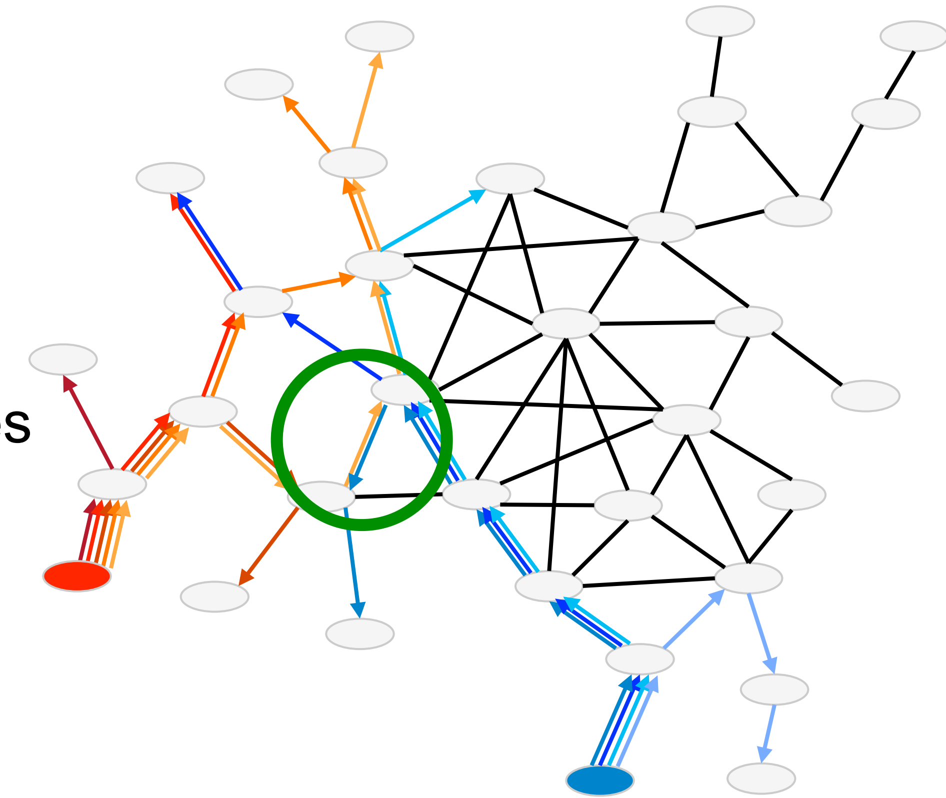
BGP Path Analysis

- Graph analysis
  - ASes as nodes
  - Links in AS path als edges
  - „Snapshot“ of Internet routes
  - Router-specific viewpoint



BGP Path Analysis

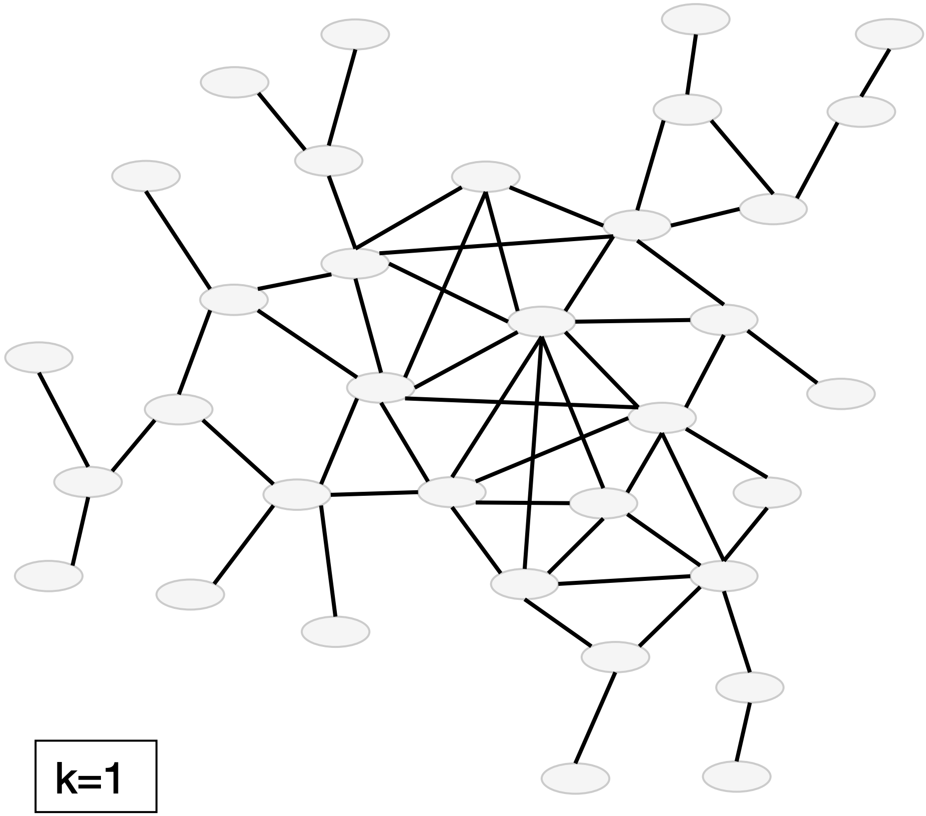
- Graph analysis
  - ASes as nodes
  - Links in AS path als edges
  - „Snapshot“ of Internet routes
  - Router-specific viewpoint
- Interesting nodes
  - large in- and out-degree
  - Internet fixpoints
- Route changes
  - observable in BGP updates
  - convergence process



Internet Fixed Points

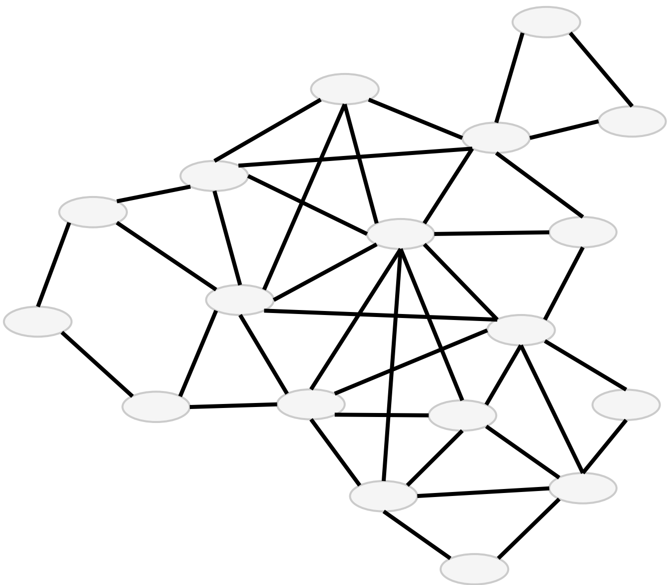
- Necessary properties of fixed point
  - Stable over long period of time
  - constant properties
  - Fixed point from different perspectives
  - Core as center of gravity: route length to fixed point is similar
- Candidates
  - Individual routers
  - Individual Autonomous System
  - Set of routers / Autonomous Systems
  - Structural components of Internet graph
- Core of the Internet
  - Set of Autonomous Systems
  - Stable (no significant fluctuation)
  - Fixed point from all perspectives
  - ⇨ k-core algorithm

k-core algorithm



1. removal of nodes with degree=1

k-core algorithm



1. removal of nodes with degree=1
2. removal of nodes with degree<= 2

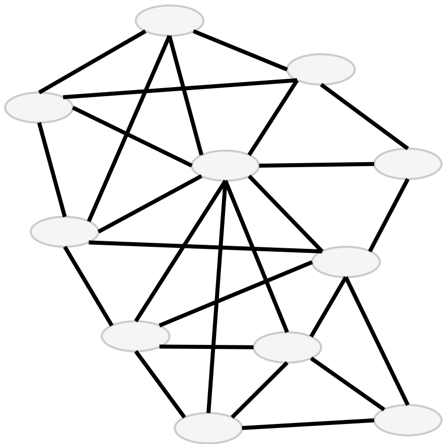
...

X. all nodes removed → (X-1)-core found

k=2

Core of the Internet

k-core algorithm

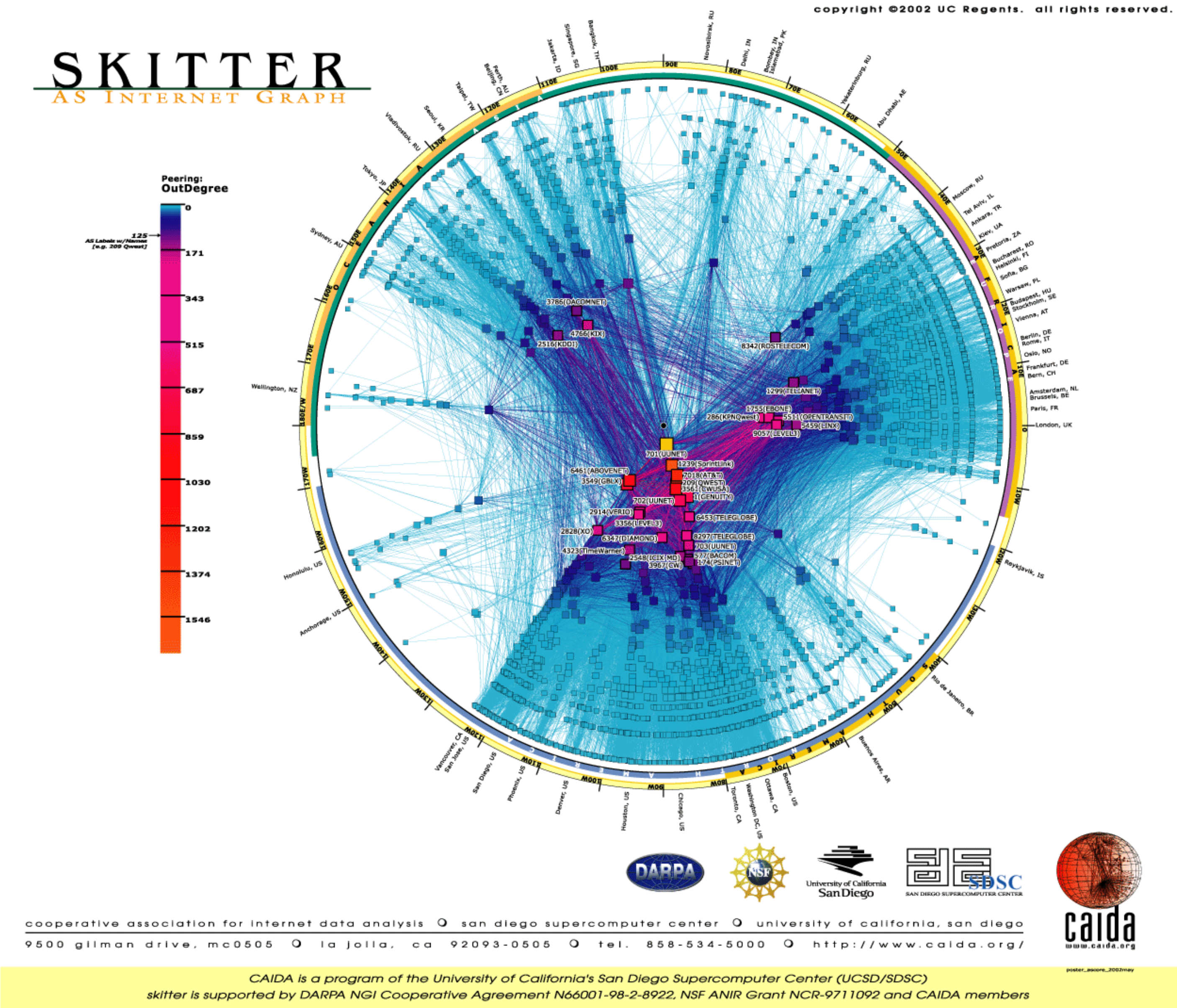


Internet AS core

- maximum k=23
- 49 AS (of 38.693 AS)

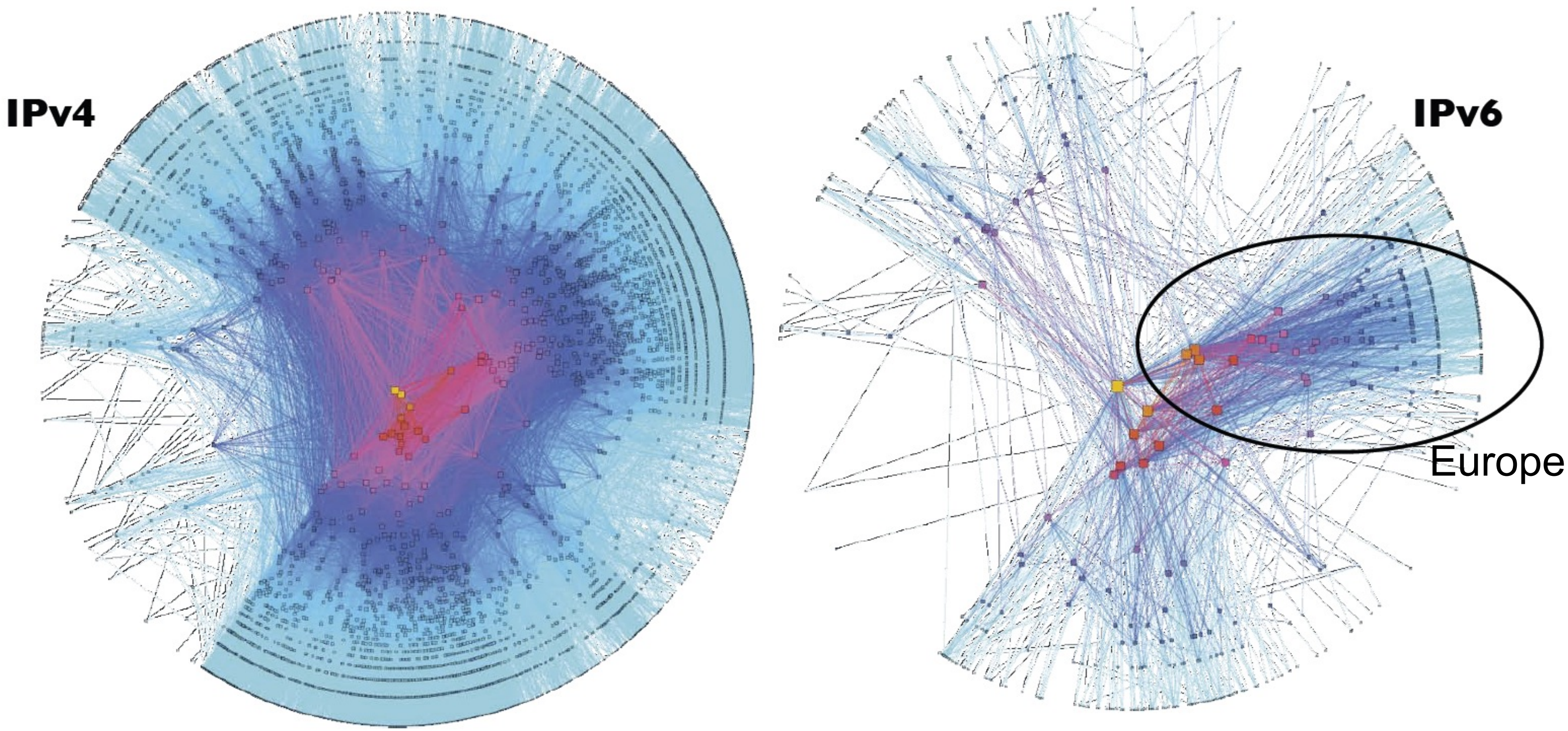
AS174 COGENT-174	AS4436 AS-NLAYER
AS209 ASN-QWEST	AS4637 REACH
AS286 KPN	AS5400 BT
AS293 ESNET	AS5413 UNKNOWN
AS701 UUNET	AS6453 UNKNOWN
AS812 ROGERS-CABLE	AS6461 ABOVENET
AS852 UNKNOWN	AS6539 GT-BELL
AS1239 SPRINTLINK	AS6762 SEABONE-NET
AS1273 CW	AS6939 HURRICANE
AS1299 TELIANET	AS7018 ATT-INTERNET4
AS1668 AOL-ATDN	AS7473 SINGTEL-AS-AP
AS2497 Asia Pacific NIC	AS8001 NET-ACCESS-CORP
AS2516 KDDI	AS8075 MICROSOFT-CORP
AS2828 XO-AS15	AS8928 INTEROUTE
AS2914 NTT-COMM	AS9002 RETN-AS
AS3257 TINET-BACKBONE	AS10026 PACNET
AS3292 TDC	AS10310 YAHOO-1
AS3303 SWISSCOM	AS11164 TRANSITRAIL
AS3320 DTAG	AS13030 INIT7
AS3356 LEVEL3	AS15169 GOOGLE
AS3491 BTN-ASN	AS15412 FLAG-AS
AS3549 GBLX	AS19151 WVFIBER-1
AS3561 SAVVIS	AS20940 AKAMAI-ASN1
AS4134 APNIC	AS22822 LLNW
AS4323 TWTC	

ISP Peering Relations



source: caida.org

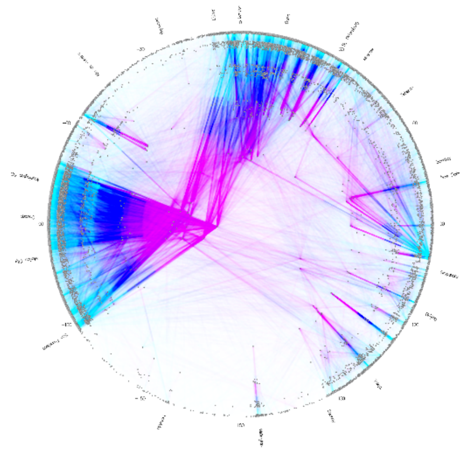
IPv4 vs. IPv6 Graphs



Ases are ranked by their customer cone size:
the number of their direct and indirect customers

source: caida.org

AS Topology Exploration



- **Custom Tool “A3View”**
  - Explorative AS-graph inspection
  - Visualizes arbitrary BGP dumps
  - Extends CAIDA visualization “otter”

- **Goal**
  - Provide a flexible tool for further research
  - Reduce AS-Graph complexity through layout and clustering
  - Enable efficient access to BGP dumps

Mathias Helminger. **Interactive visualization of global routing dynamics**. *Bachelor thesis supervised by Johann Schlamp*, TUM, June 2011.

BGP Update Process

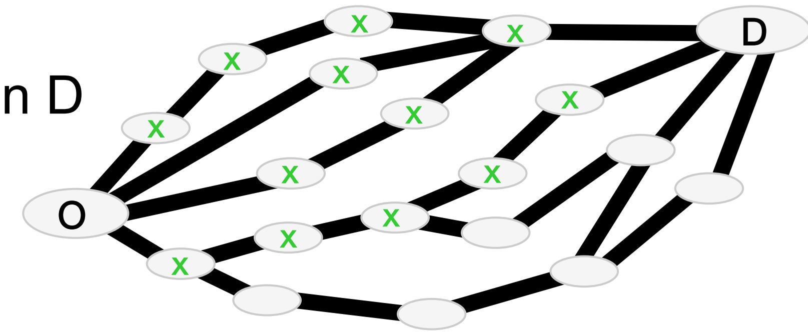
- Neighboring node „announced“ route to destination prefix
  - Propagation of best route only
  - However: several routes to destination prefix known
  - Selection of best route as part of BGP Path Selection Process; influences include AS path length

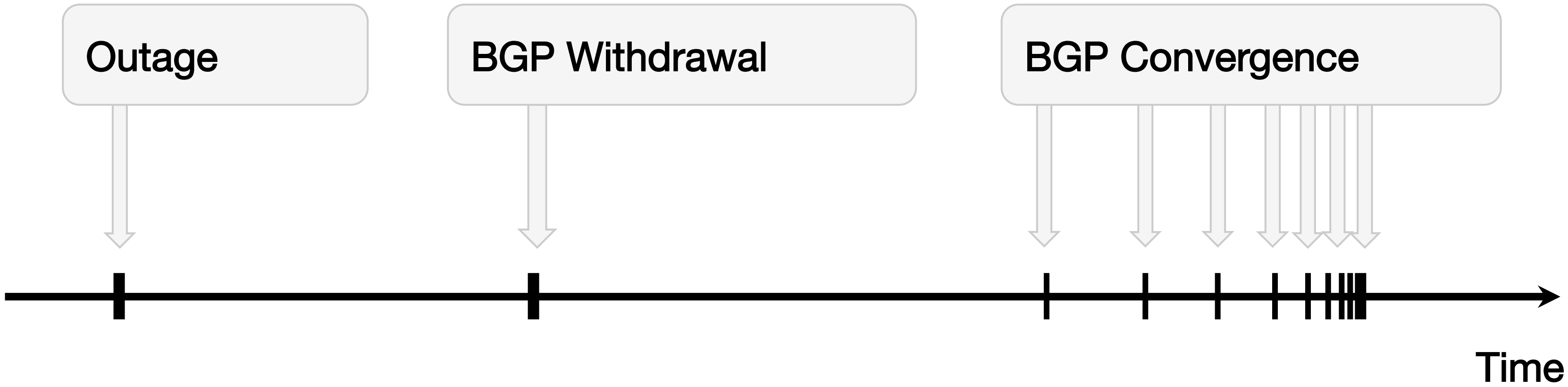
- Evaluation
  - Statistical analysis (e.g. „number of route updates per prefix and time)
  - Quantitative analysis (e.g. number of topological changes of BGP graph)

- Convergence of BGP

BGP Update Process

- Example: process after route outage
  - Outage of link/system at destination D
  - Propagation of BGP messages
  - Convergence at observer O
- Process influenced by
  - BGP timeout (90s)
  - Number of different routes to destination
  - Withdrawal of all affected routes required for convergence





BGP Update Process

- Analysis: BGP Beacons
  - Periodic announcements and withdrawals of a specific subnets (e.g. every 2 hours)
  - Observation of propagation and convergence by route collectors *(e.g. RIPE, RouteViews)*
  - Additional measurements at IP level from specific collectors *(LookingGlasses)*

Business Considerations:

Traffic Engineering

Routing: Optimization purposes

- Inter-AS routing
  - Optimality = select route with highest revenue/least loss
  - Mainly policy driven – we've seen that now
- Intra-AS routing
  - Optimality = configure routing such that network can host as much traffic as possible
  - Traffic engineering methods

Traffic Engineering (TE)

1. Collect traffic statistics: Traffic Matrix
   - How much traffic is flowing from A to B?
   - Often difficult to measure!
     - Drains router performance
     - Therefore often estimated – active research area
     - Alternative: Build lots of MPLS tunnels, measure each tunnel
2. Optimize routing
   - E.g., calculate good choice of OSPF weights
   - Typical goal: minimize maximum link load in entire network; keep average link load below 50% or 70%
3. Deploy new routing
   - Performance may deteriorate during update
   - E.g., routing loops during OSPF convergence

Dynamic Traffic Engineering

Why static? Why don't we do it dynamically?

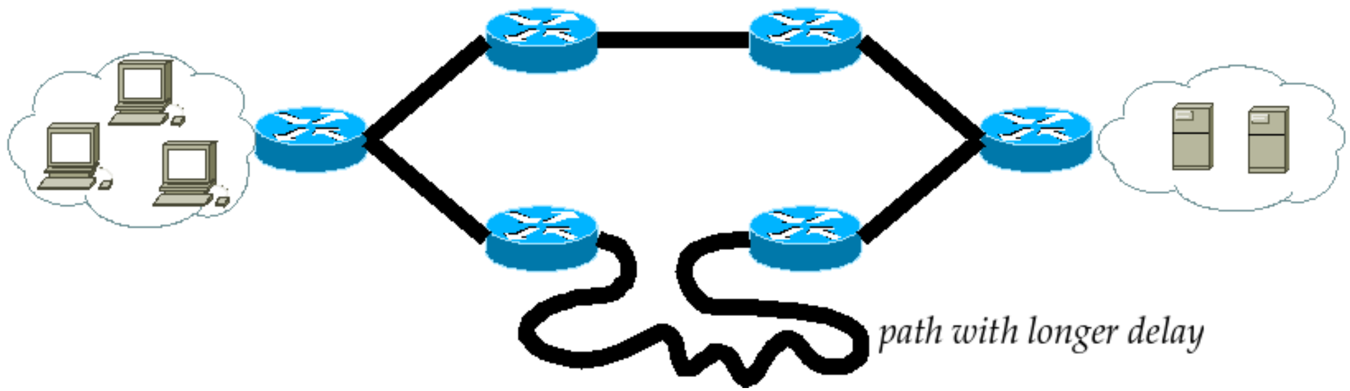
- Prone to oscillations and chaotic behaviour
  - Bad experiences in the ARPANET
  - Ex.: Route A congested, route B free
    → Everyone switches from A to B
    → Route A free, route B congested → …
- Routing loops during convergence → packet losses
- Packet reordering:
  - Packet P1 arrives later than Packet P2
  - TCP will think that P1 got lost! ⇒ congestion control!
- Actually, a difficult problem
  - Stale information
  - Interaction with TCP congestion control
  - Interaction with dynamic TE mechanisms in other ASes
- Thus: Congestion control in end hosts (TCP), usually not in network

Multipath Routing

- Routing = finding best-cost route
- But: What if more than one best route exists?
- Some routing protocols allow Equal-Cost Multi-Path (ECMP) routing, e.g., OSPF
  - ≥ 2 routes of same cost exist to destination prefix?
    → Evenly distribute traffic across these routes

Multipath routing: TCP problem

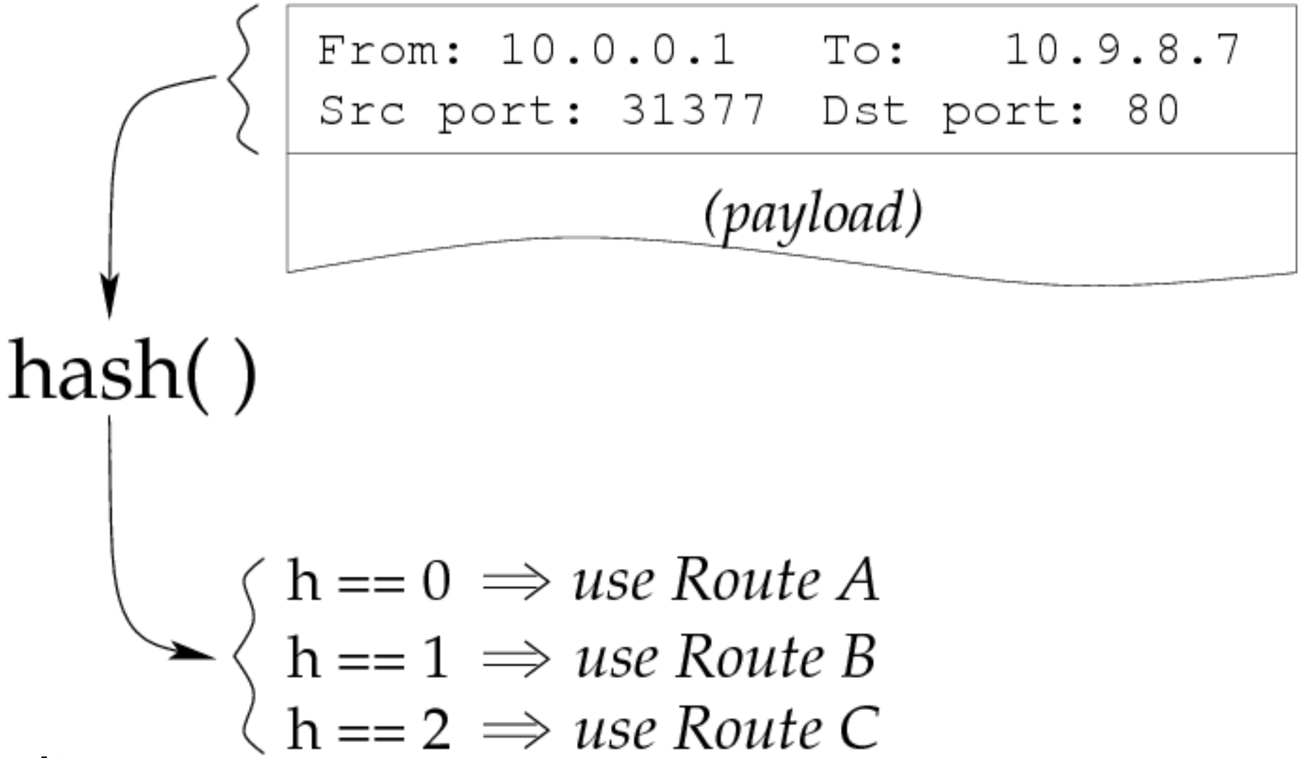
- How to distribute traffic? Naïve approaches:
  - Round-robin
  - Distribute randomly
- Equal cost does not mean equal latency:



- Problem with TCP = Packet reordering!
  - Packets sent: P1, P2
  - Packets received: P2, P1
  - Receiver receives P2 → believes P1 to be lost → triggers congestion control mechanisms → performance degrades

Multipath routing: Solution

- Hash “consistently”…
- …and use packet headers as “random” values:



- Result:
  - Packets from same TCP connection yield same hash value
  - No reordering within one TCP connection possible

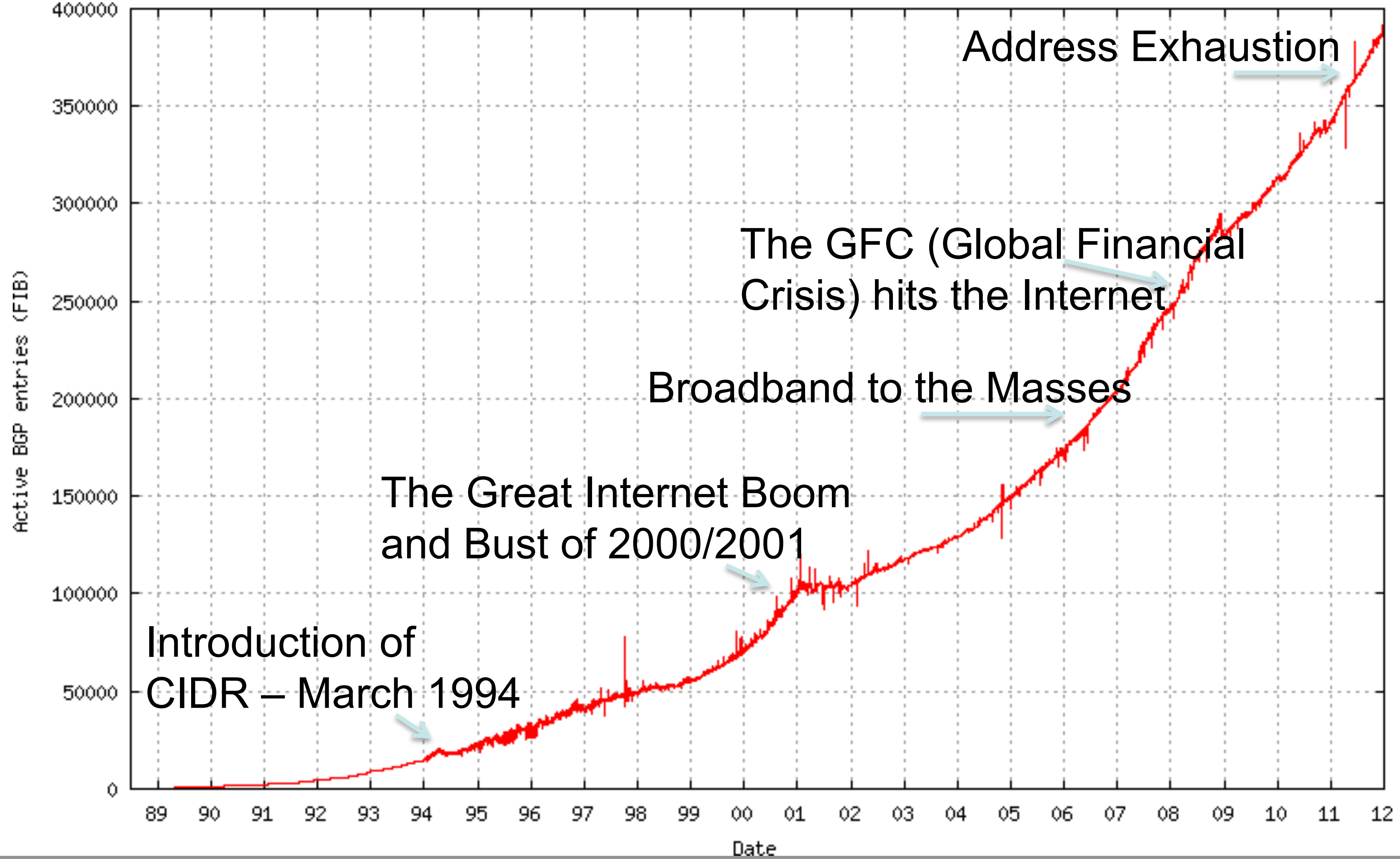
BGP Table Growth

- Geoff Huston

  - Chief Scientist at APNIC
  - Books
    - Quality of Service: Delivering QoS on the Internet and in Corporate Networks, by Paul Ferguson and Geoff Huston, Wiley, 1998
    - ISP Survival Guide, Wiley, 1998
    - Internet Performance Survival Guide, Wiley, 2000
  - Blog „The ISP Column“at http://www.potaroo.net/
  - Article: „Analyzing the Internet's BGP Routing Table“, http://www.potaroo.net/papers/2001-3-bgptable/4-1-bgp.pdf published in The Internet Protocol Journal, Volume 4, Number 1, http://www.cisco.com/ipj

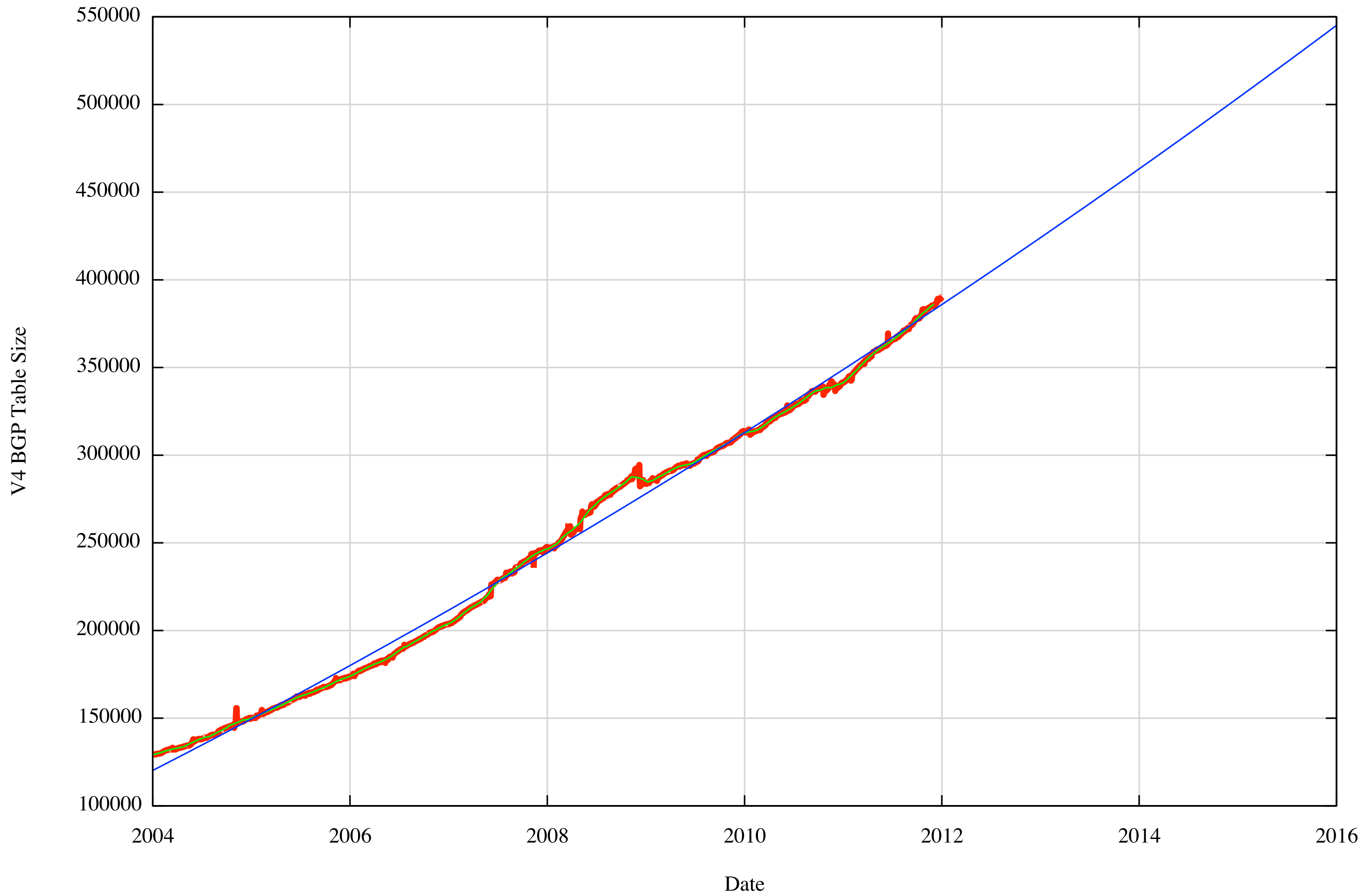
The Big Picture of the v4 Routing Table

- Geoff Huston, An Introduction to Routing in the Internet, IGF Workshop, IGF, Baku, 9 November 2012, http://www.potaroo.net/presentations/index.html



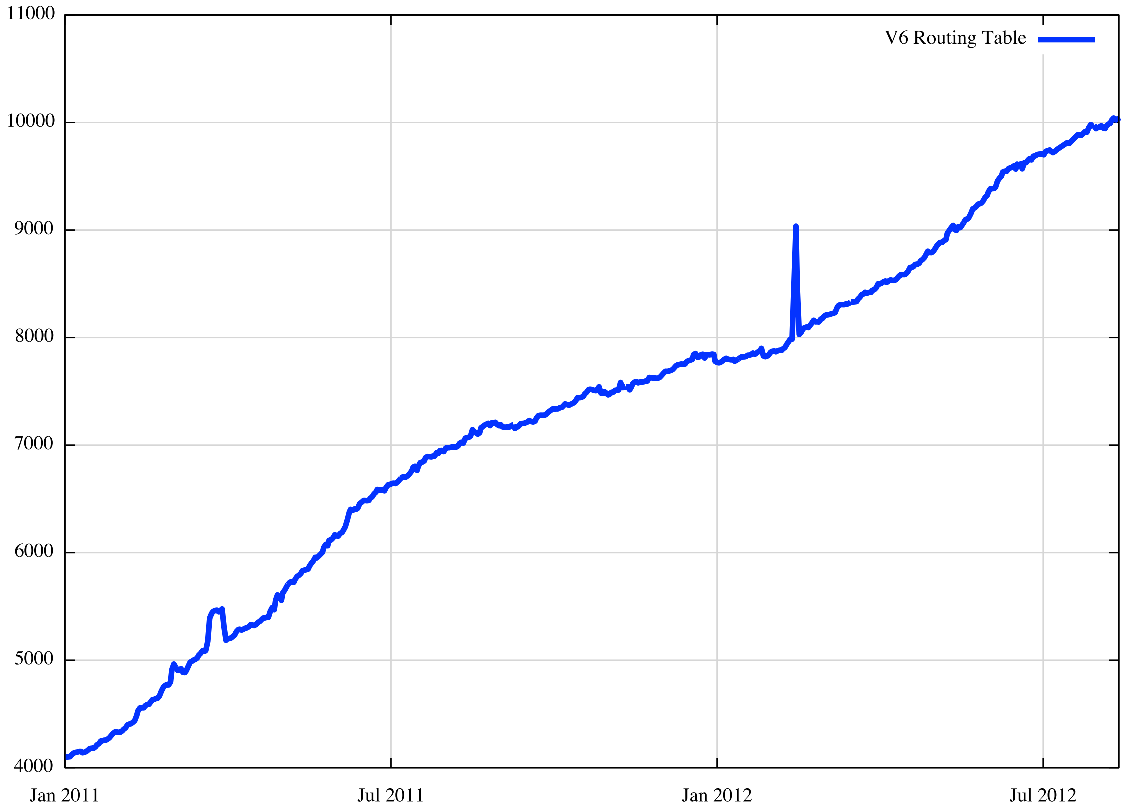
Hot Topic: Scaling of the v4 Routing Table

- Geoff Huston, An Introduction to Routing in the Internet, IGF Workshop, IGF, Baku, 9 November 2012, http://www.potaroo.net/presentations/index.html



Growth of the v6 Routing Table

- Geoff Huston, BGP Progress Report, APNIC 34, Phnom Penh, 30 August 2012, http://www.potaroo.net/presentations/index.html



- Overall IPv6 Internet growth in terms of BGP is 50 % p.a.
- If relative growth rates persist then the IPv6 network would span the same network domain as IPv4 in 2018

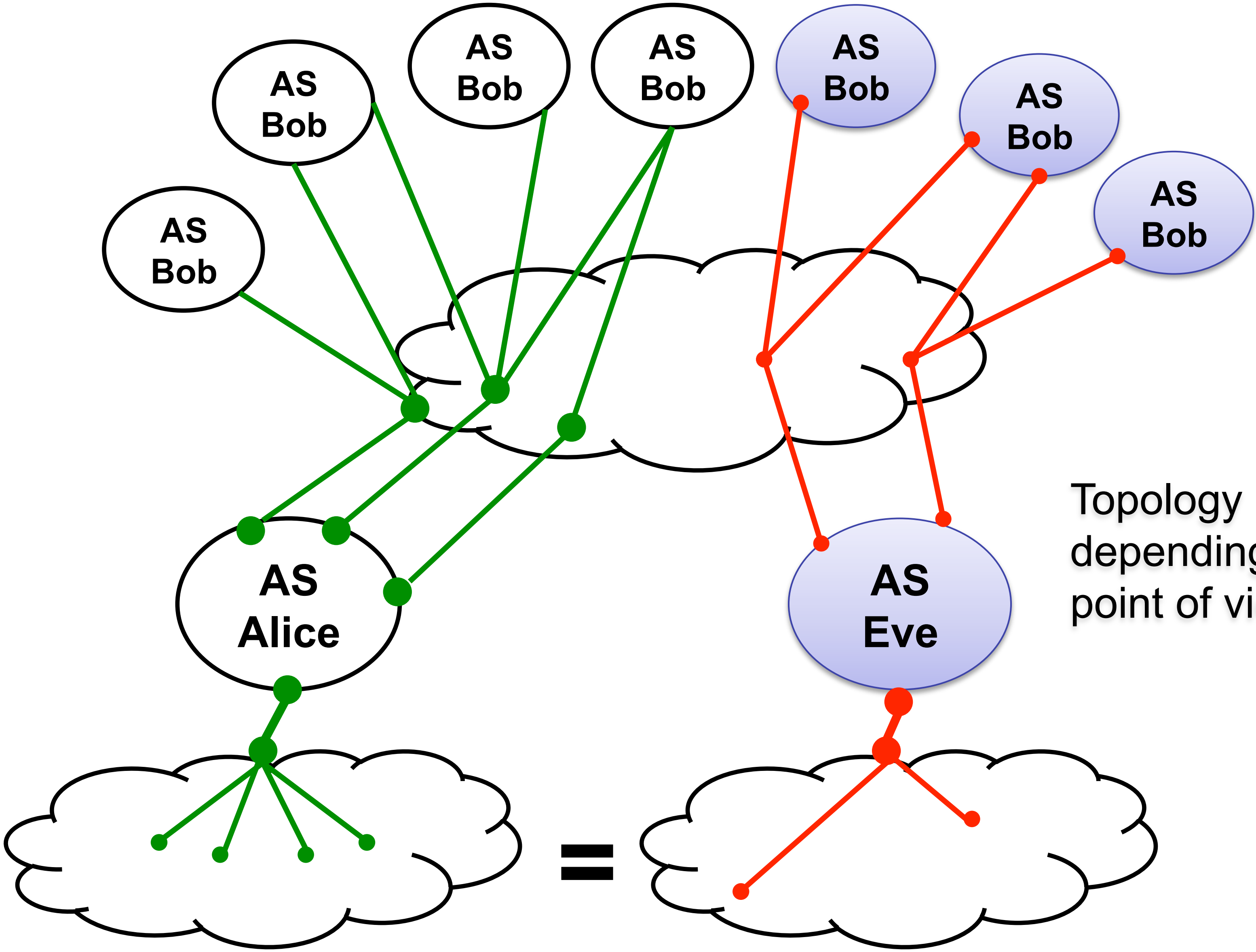
Growth of NATTed Internet

- Geoff Huston, BGP Progress Report, APNIC 34, Phnom Penh, 30 August 2012, http://www.potaroo.net/presentations/index.html
- Growth of mobile devices
  - In 2012, approximately 400 mio smartphones were sold
  - This does not include tablet s (Kindles, iPads, etc.)

  ⇨Estimation: NATTed Internet grew by ~600M devices in 2012

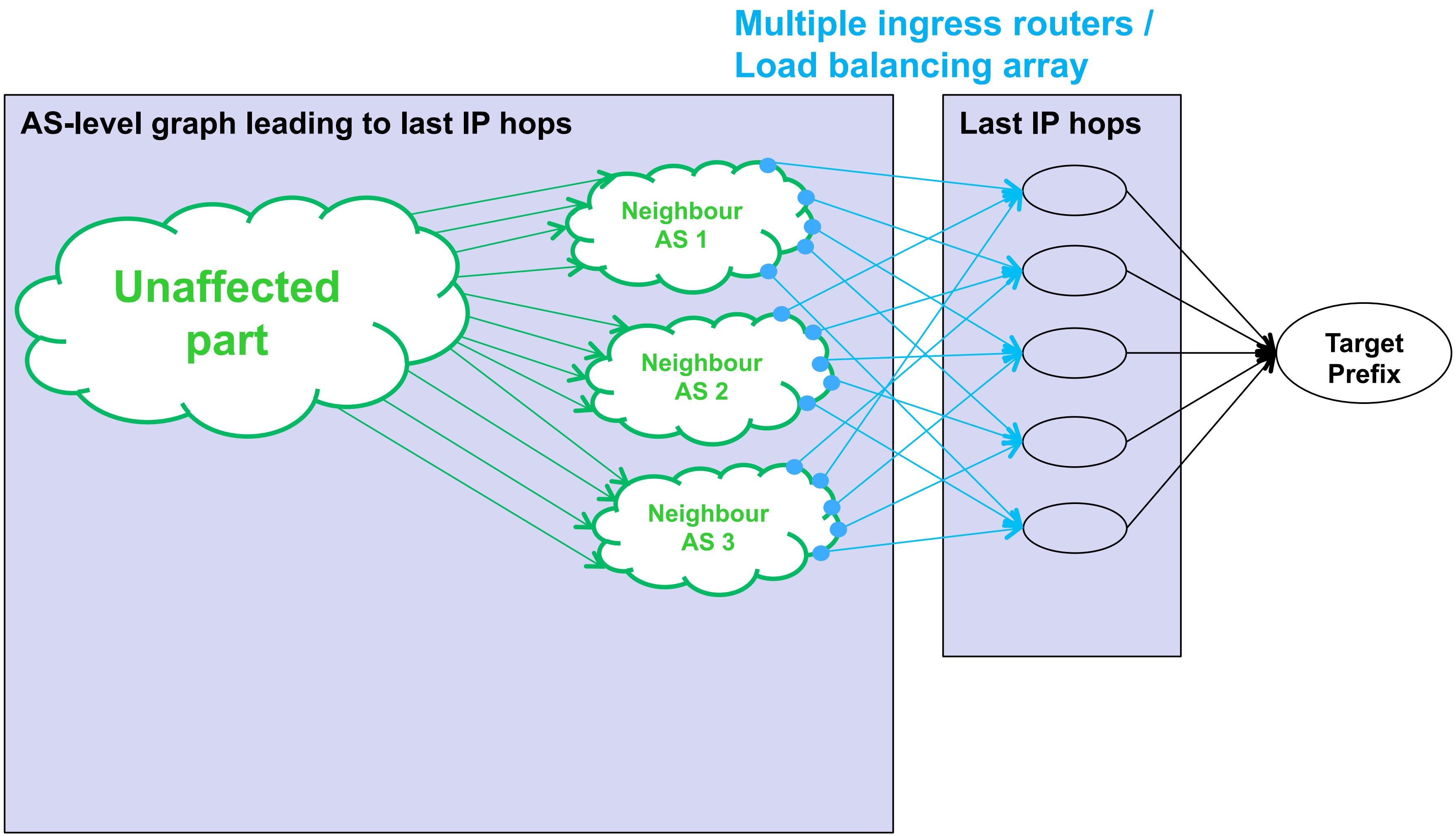
Hijack Detection (cont.)

Prefix Hijack Detection

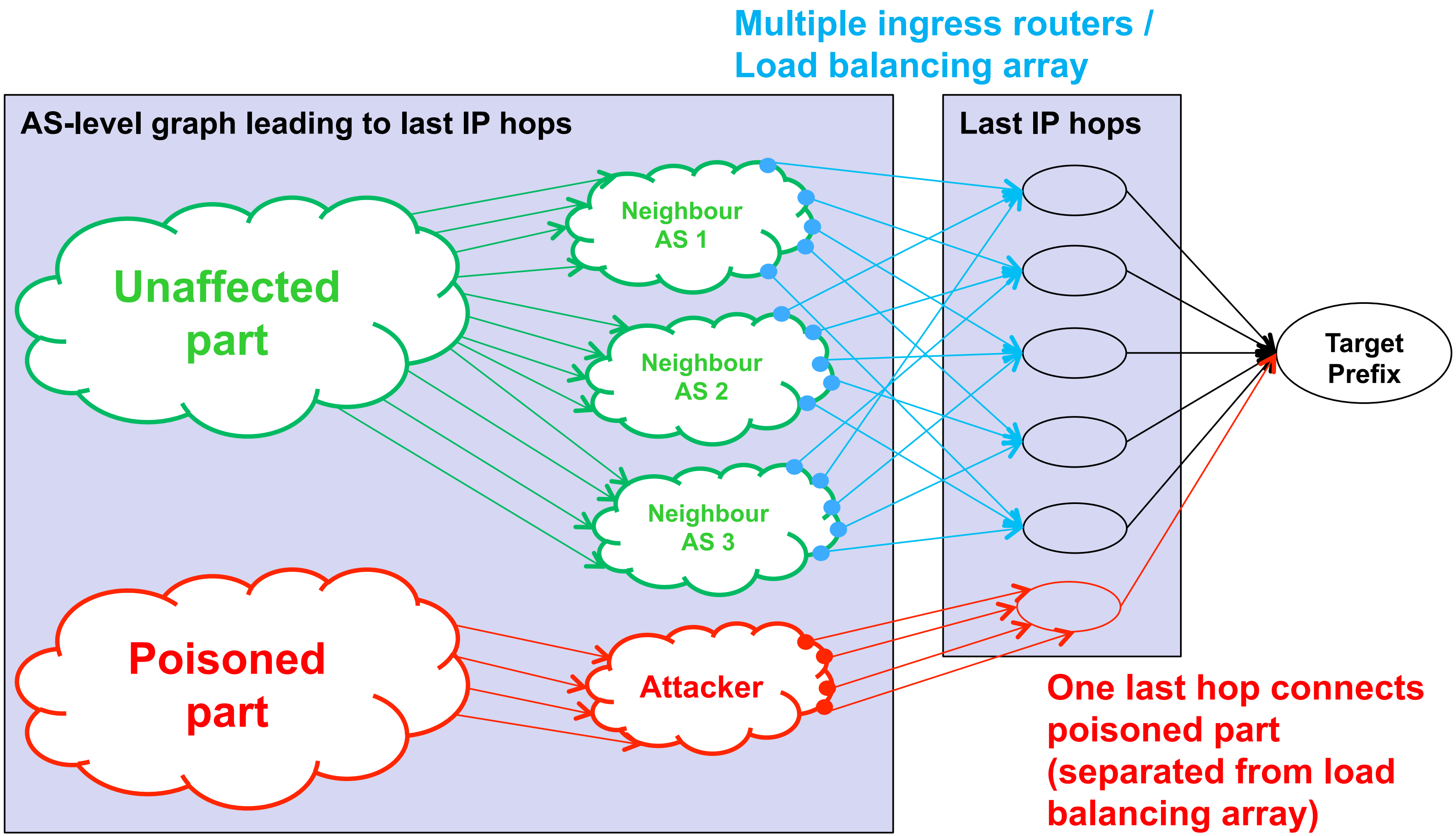


Topology depending on point of view

Prefix Hijack Detection – Background



Concurrent Hijack Detection – Background



Concurrent Hijack Detection

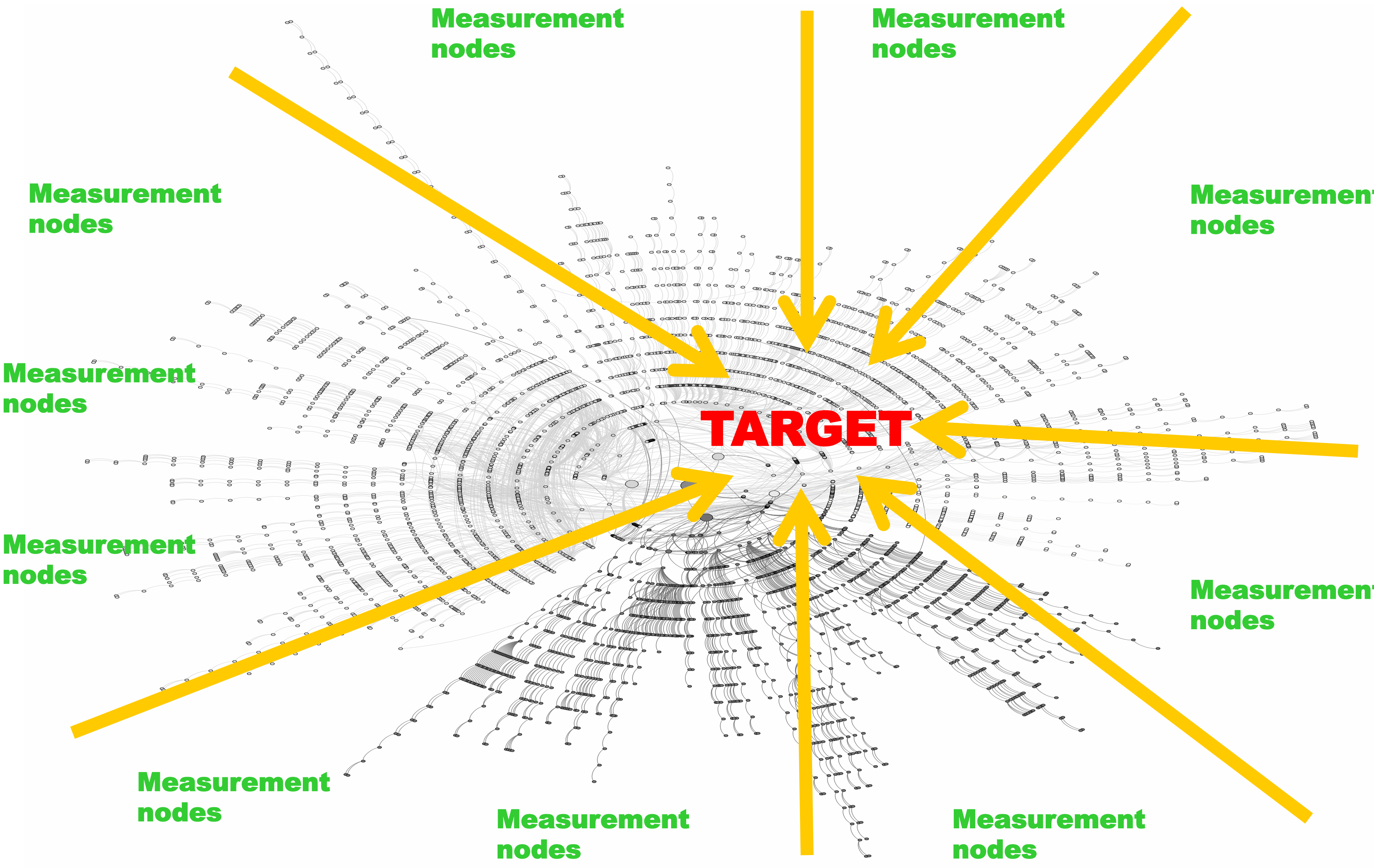
- **Traceroute measurements from multiple vantage points**
- **Try to identify unaffected and poisoned part**
  - Compare last hops to target prefix
  - Compare „downstream graph“ of last hops
  - Quantify differences
- **Promising metrics for last hops**
  - **Examples for Hijacking detection metrics**
    - Odd distribution of first-rank countries
      For example, five neighbours located in Germany, one in New Zealand
    - Odd distribution of first-rank ASes
    - Odd correlation of downstream AS distributions
    - Odd Round-Trip Time
  - **“ExAS” metric**
    - Detects how many Autonomous Systems in the whole graph are **connected exclusively** through one neighbour of the target
    - Segmentation of the Graph into a possibly hijacked and valid part
    - Can specify the *impact* of a possible hijack

Quirin Scheitle. **Active Detection of BGP Prefix Hijacking**. *Diploma thesis supervised by Johann Schlamp*, TUM, September 2012.

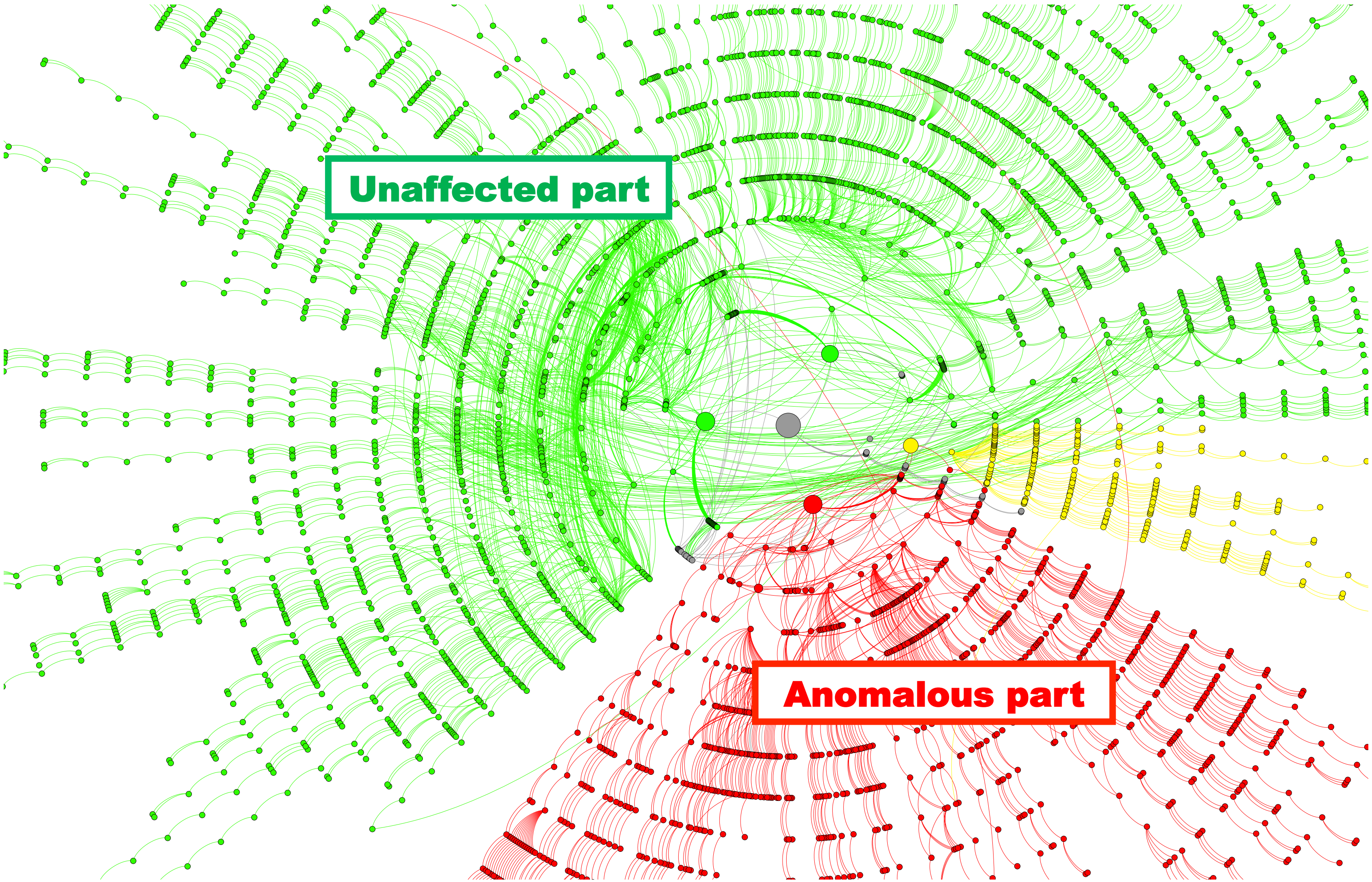
Looking Glass Framework

- Volatile measurement framework
  - Based on traceroute interfaces on publicly accessible BGP routers
  - >500 nodes connected, 5% fluctuation / week
  - Measurement of ping, traceroute and AS paths
- Available measurement nodes
  - 5 continents, 67 countries
    Argentina | Australia | Austria | Bahrain | Bangladesh | Belgium | Bolivia | Brazil | Bulgaria | Canada | Chile | China | Cyprus | CzechRepublic | Denmark | Djibouti | Egypt | Estonia | Finland | France | Georgia | Germany | Ghana | Greece | Hungary | Iceland | India | Indonesia | Ireland | Italy | Japan | Kazakhstan | Kenia | Korea | Latvia | Lithuania | Luxembourg | Malaysia | Mexico | Mozambique | Nepal | Netherlands | Norway | Peru | Philippines | Poland | Portugal | Qatar | Romania | Russia | SaudiArabia | Saudi Arabia | Serbia | Singapore | Slovakia | South Africa | Spain | Sweden | Switzerland | Taiwan | Thailand | Turkye | UAE | Ucraine | Ukraine | UnitedKingdom | USA
  - 101 providers
    AARnet Australian Research Network | AboveNet USA | Aconcaguared Telecom Chile | Adnet Telecom Romania | Algar Telecom Brazil | ALOG Datacenters do Brasil | americana digital brazil | Anders Telecom Russia | ATMAN Poland | ATT | Bangladesh Research Network | Bell Canada | BIT Netherlands ISP | BroadBandTower Japan | Broadnet Finland ISP | BTCL Bangladesh Telecom | BT Ireland, former ESAT | Colt Europe | Comstar Direct MTU Russia | Cooperative Telefonica Pinamar | Corbina Telecom | Cyfra Ukraine | cz.nic czech ISP | DFN - German Research Network | DTAG | Eastlink Atlantic, CA | Eastlink CA Eastern | Eastlink CA Pacific | Faroe Telecom | fiord russian ISP | FUNET CSC Finnish Research Network | GTD Internet Chile | GTS Poland | HurricaneElectric | Hutchinson Global Communications HongKong | Init7 Swiss | Inteliquent | IP Exchange German ISP | IP TriplePlay Netherlands | ITgate Italian ISP | JSC Kazakhtelecom | Korea Telecom | lancernet brazil | level3 | Linxtelecom | MegaLink Bolivia | MTN Ghana | NeoTelecom Russia | Neotelecom Russian Tier3 ISP | Netia Poland | Netnod Sweden IX | NTT | OJSC MegaFon Russian ISP | optus Australia | Orange Business Services Russia | Orange/TPnet poland | OSJC Vimpelcom gldn.net Russia | pipe networks, australia | Primus Telecommunications, Australia | Rack66 | Rede Rio de Computadores | Registro.br | Reliance Globalcom UK/India | RETN Russia | RHNet Iceland | RNC Brazil | Rogers Telecom Rest of Canada | Rogers Telecom Toronto | Runnet Russian ISP | Rusnet | savvis | Skylink Russia | South African Internet Exchange SAIX | Spacenet Russia | SP Tel | Sunrise Communications, Suisse | Swisscom | Switch Switzerland Research Network | tata | Telecom Italia | Telecom Srbjia | TeliaSonera | Telmex Chile | TELUS West Canada | Thunderworx Cyprus | TimeWarner Telecom | Titan Networks German ISP | TKP 3s polish isp | TK Telekom Poland | TrueIntergateway Thailand | UARnet ucrainian research network | UkrCom Ukraine | Unigate Taiwan | Universidade Estadual Paulista | VersaTel DE | Vodafone Iceland | Volia/Telesweet Ucraine | Woodynet Nairobi Kenia IXP | WV Fiber / Broadband One US | XO Communications | ZapSib Transtelecom
  - 130 cities
    Accra | Adelaide | Aktau | Alexandria | Almaty | Astana | Athens | Atlanta | Bangkok | Barcelona | Belgorod | Belgrade | Belo Horizonte | Berlin | Boston | Brasilia | Bratislava | Brisbane | Brussels | Bucharest | Budapest | Buenos Aires | Cairo | Calgary | CapeTown | Catania | Chelyabinsk | Chennai | Chicago | Cleveland | Cologne | Copenhagen | Dallas | Dammam | Denver | Dhaka | Djibouti | Doha | Donetsk | Dubai | Dublin | Faroe Islands | Frankfurt | Gdansk | Geneva | Halifax | Helsinki | HongKong | Istanbul | Jakarta | Jeddah | Johannesburg | Kathmandu | Kazan | Kharkiv | Kiev | Kochi | Krakow | Krasnodar | KualaLumpur | La Paz | Lima | Lisbon | Lodz | London | Los Andes | LosAngeles | Luxembourg | Lviv | Madrid | Manama | Manila | Maputo | Melbourne | Mexico City | Miami | Milan | Montreal | Moscow | Mumbai | Munich | Nairobi | NewYorkCity | Nicosia | Nizhny Novgorod | Novosibirsk | Odessa | Osaka | Oslo | Palermo | Palo Alto | Paris | Perth | Poznan | Prague | Reykjavik | Riga | Rio de Janeiro | Riyadh | Rome | Rostov-on-Don | SaltLakeCity | SanDiego | SanFrancisco | SanJose | Santiago | Sao Paulo | Saratov | Seattle | Seoul | Singapore | Sofia | Stockholm | St. Petersburg | Sydney | Taipei | Tallinn | Tbilisi | Tokyo | Toronto | Tula | Uberlandia | Valencia | Vancouver | Vienna | Vilnius | Voronezh | Warsaw | Wroclaw | Zurich
- Quirin Scheitle. **Active Detection of BGP Prefix Hijacking**.
  Diploma thesis supervised by Johann Schlamp, TUM, September 2012

Concurrent Hijack Detection – Example



Concurrent Hijack Detection – Example



Excursus: KLIK Team

- **KLIK Team**
  annual party

- Gifts
  - Macbooks
  - Briefcase with money
  - Car