Concept of Traffic Engineering (TE)

- Some possible definitions:
  - Traffic Engineering = Performance Optimization (of network, applications or both)
  - Traffic Engineering is a field of communications engineering that tries to make network operations more effective and reliable
  - Application of technology and scientific principles to the measurement, characterization, modeling, and control of Internet traffic

- RFC 3272: Overview and Principles of Internet Traffic Engineering

Some problems that TE tries to solve

- Effective bandwidth utilization on the path that packets are currently using
- Effective bandwidth utilization within an Autonomous System
- Optimal traffic relaying between Autonomous Systems (BGP TE)
- Fast connectivity restoration after a component breakdown (IGP Fast Convergence, MPLS Fast Re-Route, etc.)

Result: Happier users = more money
Some TE goals illustrated (1/4): Better throughput

Without TE

With TE

Some TE goals illustrated (2/4): More equal link utilization

Without TE

With TE

Some TE goals illustrated (3/4): Smaller delay (variation)

Without TE

With TE

Some TE goals illustrated (4/4): Faster route convergence

Without TE

With TE
Need for TE

- Optimization does not seem so important at the moment
  - Many applications (like P2P) behave unpredictably
    - Optimization assumptions, like traffic matrix, are hard to obtain
  - Backbone capacity is relatively inexpensive (100 GBit/s ethernet is coming soon)
    - Most ISP networks are heavily over-provisioned (peak link utilizations around 10%)
  - Optimization introduces additional complexity into the network, which might lead to connectivity problems
- Availability maximization (route convergence) is very important – over 0.2 second connectivity loss is noticed by VoIP customers

Requirements for TE: Policy System

- Policy: A definite goal, course or method of action to guide and determine present and future decisions
- TE systems incorporate a set of rules (Policy) which are propagated to enforcement points
  - Policy must be enforced to ensure that the users are behaving properly
- Examples of simple policies:
  - OSPF link weights, MPLS Label-Switched Paths, Maximal Reservable Bandwidth
- Hierarchical policy systems possible
  - Action Policies, Goal Policies, Utility function Policies

Hierarchical policy system

- Three policy levels may be used
  1) *Utility function* policies (automatically determine the most valuable goal in any situation) – highest level
    - Elements employing utility function policies MUST have sophisticated modeling and *optimization* capabilities to translate utility functions into actions
  2) *Goal* policies ("Response time must not exceed 2 sec.")
    - Elements employing goal policies MUST possess sufficient modeling or *planning* capabilities to translate goals into actions
  3) *Action* policies (IF condition THEN action)
    - Elements employing action policies MUST measure and/or synthesize the quantities stated in the condition

TE in practice

- High-level TE concepts (ISP revenue models, conceptual frameworks, policies, flowcharts, heuristic functions, etc.) must be mapped (coded) into the network's components
- Issue: can we manage complex systems by adding more complexity?
TE method classification

- TE methods can be classified in many ways:
  - Short term vs. Long term TE
  - Intra-domain vs. Inter-domain TE
  - Centralized vs. Distributed TE
  - On-line vs. Off-line TE
  - Performance optimizing vs. Availability maximizing TE
  - Host-based vs. Network-based TE
  - ...

Classification 1/6: Short-term vs. Long-term TE

- This classification actually deals with TE decisions more than methods themselves
- A single TE method may include decisions at many time scales

1. Short-term TE decisions
   - Decisions made at this timescale affect milliseconds
   - Example: What should a router do with a single data packet: drop it, forward it or put it to queue?

2. Medium-term TE decisions
   - Decisions affect seconds-minutes
   - Example: Should a router allow a connection to be established?

3. Long-term TE decision
   - Decisions affect hours-days (or more)
   - Example: Which BGP-policy should be configured to the AS?

Classification 2/6: Inter-domain vs. Intra-domain TE

- Inter-domain TE
  - Inter-domain TE methods try to optimize network resource usage on a global (inter-domain) scale
    - Example: BGP TE

- Intra-domain TE
  - Try to optimize resource usage within the ISPs own network
    - Example: MPLS TE

Sub-categories of Inter-domain TE

1. Outbound TE
   - To which provider ISP should I send my data packets to (in case of multi-homing)
     - The best choice may depend on packet destinations

2. Inbound TE
   - From which provider ISP do I want to receive my data packets from
     - From the cheapest one, or
     - From all of them in a certain proportion

3. Transit TE (for transit ISPs)
   - To which peer/provider/customer should I relay my peer’s/provider’s/customer’s packets to
Classification 3/6: Centralized vs. Distributed TE

- Centralized TE has one Policy Decision Point (PDP) (= TE Server) and routers are Policy Enforcement Points (PEP)
- Benefit: Simple implementation
- Drawback: Single point of failure

- Distributed TE has no centralized PDP
- In distributed TE, routers may exchange measurement data and decide on the policy themselves
- Benefit: Robust to failures
- Drawback: Often complicated implementation

Classification 4/6: On-line vs. Off-line TE

- On-line TE basically means that policy is decided **dynamically** (and usually relatively **quickly**)
- On-line TE must usually be "autonomic" i.e. computer-controlled
  - Controller computer(s) must have certain utility function(s) to base their decisions on
- Off-line TE usually has slower response times and human operator may decide the functional policies that are used in the network
- Performance may be equally good

Classification 5/6: Performance optimizing vs. Availability maximizing TE

- Some TE methods try to maximize the network’s throughput, minimize the average delay that users experience, or minimize the jitter in delay
  - Note: In practice these problems are often cheaper to solve by simply adding hardware to the network
- Others try to maximize the resiliency (robustness, availability) of the network
  - This is where software actually helps
  - Note: Some people consider these methods to be outside the definition of TE

Performance optimizing TE

- Performance optimization is often synonymous for **Load Balancing**
  - However, TCP, which is not load balancing, could also be considered as optimization
- Performance optimization sometimes requires **Traffic Matrix** (how much traffic users try to send others)
  - Often hard to obtain in practice
  - However, in large networks traffic matrix may be quite similar across days
- Optimization problems are often NP-hard (It takes too long to find a globally optimal solution) -> **heuristic** algorithms needed
- A 'heuristic' is a method to help solve a problem.
  - It is used to produce a solution that is usually reasonably close to the best possible answer. Heuristics are “rules of thumb”, educated guesses, intuitive judgments or simply **common sense**.
- Important note: As network traffic grows all the time, optimization may only shift hardware acquisition some time (~ 1 year) to the future
Classification 6/6: Host-based vs. Network-based TE

- In host-based TE methods, each sender-receiver pair controls traffic based on some policy (e.g. TCP)
- In network-based TE, the networks controls traffic (e.g. queuing methods, MPLS, IGP Metric Based TE)
- Both methods used in most networks
- Non-trivial to quantify the interactions of Host-based and Network-based TE

Examples - TE method 1/9: TCP

- TCP tries to optimize the bandwidth utilization on the path that packets are currently using
- However, it cannot solve the problem of effective resource sharing within the network
- Inter-domain/Intra-domain, Distributed, On-line, Host-based, Performance optimizing

TE method 2/9: Queuing methods

- Try to optimize the bandwidth utilization on the path that packets are currently using
- Data packets can be placed in different queues within a router (packets must be classified before this)
- These queues may be served according to different policies so that the packets in those queues will get service that is most suitable (according to traffic characteristics and operator policies)
- Inter-domain/Intra-domain, Distributed, Network-based

TE method 3/9: BGP TE

- The de-facto business tool of ISPs – configurations are usually based on business goals
  - Hot potato – policy common (use shortest internal route)
  - Attempts to minimize internal network congestion
  - Cold potato is sometimes used in case of server-client asymmetries between peering ISPs to distribute traffic more evenly
- BGP TE could also be used to optimize Internet on a global (inter-domain) scale
  - Hasn’t been popular among ISPs for a couple of obvious reasons (lack of business incentives, lack of available data, and a lack of sophisticated inter-domain TE tools)
  - However, as the demand (QoS, Spam, DDoS, routing table growth etc.) placed on the networks continually increases, there is a growing need for smarter inter-domain TE
- Inter-domain, Distributed, (Off-line), Network-based, usually Availability maximizing
**TE method 4/9: IGP Fast Convergence**

- IGP Fast Convergence tries to make optimizations to IGPs that reduce the time it takes for routes to converge
- With clever optimizations it is possible to reduce the time from 10-30 seconds to as low as one second (or even below one second) for first 500 prefixes in a well designed backbone
- What is needed to achieve fast convergence without sacrificing stability is good damping algorithms which can separate unstable components from the stable components
- Intra-domain, Distributed, Network-based, Availability maximizing

**TE method 5/9: Fast Re-Route**

- Fast Re-Route (FRR) can be implemented with MPLS
- If correctly configured, it is usually slightly faster and more deterministic than IGP fast convergence, however, it can also perform a lot worse
- FRR uses local detection and protection at the point of failure
- May be needed for telephony users
- Intra-domain, Distributed, (Off-line), Network-based, Availability maximizing

**TE method 6/9: ECMP TE**

- Equal Cost Multipath (ECMP) balances load for all equal cost paths towards a destination
- Every router performs a hash on received packets to determine which one of the paths should be used for the packet
- The hash may be a function of source address, destination address, source port, destination port etc. and can be static or dynamic
- The hashing can be performed separately for each packet or based on flows
- Provides path redundancy
  - If some resource on one of the equal cost paths breaks, other equal cost paths are often still available
- Intra-domain, Distributed, (Off-line), Network-based, Has components of Performance Optimization as well as Availability maximization

**ECMP TE (continued)**

- ECMP is often considered to be sufficient TE method if both topology and traffic demands are symmetrical like in the figure
**TE method 7/9: MPLS TE**

- MPLS TE enables explicit (source) routing, which is quite significant, since it makes **unequal-cost load balancing** possible.
- In the figure, traffic load is divided evenly by means of two LSPs to unequal-cost router-paths.
- MPLS TE is somewhat complicated to deploy, but when correctly configured, it can be very effective and flexible.

**MPLS TE (continued)**

- In MPLS TE the Label Switched Paths are (iteratively or in the beginning) calculated (either at a centralized server or in network’s edge routers) using Constrained SPF-algorithm (parameters are topology and traffic matrix), and then the corresponding configurations (bandwidth reservations) are made to all routers.
- Medium term, Intra-domain, Centralized/Distributed, Online/Off-line, Network-based, Can be used in both Performance optimizing and Availability maximization (backup LSP).

**TE method 8/9: IGP Metric Based TE**

- Tactical IGP Metric Based TE tries to alter routing protocol (OSPF or IS-IS) link weights in some specific points of congestion:
  - Often only shifts the point of congestion.

**IGP Metric Based TE (continued)**

- Strategic IGP Metric-Based TE tries to find an optimal link weight setting for the network (using some optimization goal(s))
IGP Metric Based TE (continued)

- The main idea in strategic IGP Metric Based TE is to 1) get the topology and traffic information from the network to a centralized server, 2) feed this data to an optimization algorithm, and 3) send the optimal link weights back to the network’s routers.
- The routers calculate routes based on the optimal weights and traffic gets forwarded to near-optimal paths.
- IGP Metric Based TE is in general not as good in load balancing as MPLS TE, but it does not need the additional protocol layer of MPLS.
- Medium term, Intra-domain, Centralized/Distributed, On-line/Off-line, Network-based, (Performance optimizing)

TE method 9/9: Probing (TE-XCP)

- Probing in the context of traffic engineering basically means that edge routers send additional packets (probes) to the network in order to find out if certain paths towards a destination have less load than others.
- Based on the responses to these probes, routers can divide load to different paths that have been pre-configured to them.
- Very good results have been obtained by using the probing approach, however, this approach seems to require quite radical changes to current router software.
- Short term, Inter-domain/Intra-domain, Distributed, On-line, Network-based, Performance optimizing

Regular topologies enable simpler TE algorithms

- Single ring
  - 2 independent paths between nodes

- Double ring
  - 3 independent paths

- Torus grid
  - 4 independent paths

Regular topologies

- Benefits of regular topologies:
  - Faster forwarding due to faster route lookups
  - Fast route recovery
  - Reduction in routing traffic volume
  - Reduction in routing table size
  - Automatic load balancing if traffic matrix is uniform
    - Note that in a random topology central links and nodes tend to get more congested than ones near the edges
  - Easier to add links in an organized way (impact of new node is easier to analyze)
Multilevel topology

- Provide good properties in large-scale implementations
- Higher level nodes can be fully meshed

Summary

- Traffic Engineering (TE) methods try to make network operations more effective and reliable while at the same time optimizing resource utilization
- TE methods can be classified along many axis
- Many TE methods can be used at the same time
- Most suitable TE method(s) depend on the network’s (e.g. random topology, regular topology, multilevel topology) and users’ properties (applications, demands)