

Chair for Network Architectures and Services – Prof. Carle Department for Computer Science TU München

Master Course Computer Networks (IN2097)

Introduction to

Network Resilience

Dr. Ali Fessi Dr. Nils Kammenhuber Prof. Dr.-Ing. Georg Carle





- I. Terminology
- II. Challenges in the current Internet
- III. Resilience Mechanisms



I. Terminology

- II. Challenges in the current Internet
- III. Resilience Mechanisms



- 1. The "fault \rightarrow error \rightarrow failure" chain
- 2. Fault tolerance
- 3. Resilience
- 4. Dependability
- 5. Security
- 6. Availability vs. Reliability



□ Service:

- Sequence of the system's external state
- Correct service is delivered when the service implements the system function

Definition

- A service <u>failure</u>, or simply failure, is an event that occurs when the delivered service deviates from correct service
- i.e., at least one external state of the system deviates from the correct service state
- (de: Ausfall)



- Definition
 - The deviation of an external state of the system from the correct service state is called an <u>error</u>
 - Thus, an error is the part of the total state of the system that may lead to its subsequent failure
 - (de: Defekt)
- Definition
 - The cause of an error (adjuged or hypothesized) is called a <u>fault</u>
 - (de: Fehler)



Definition

 A system is fault-tolerant if it can mask the presence of *faults* in the system by using *redundancy*

- □ Redundancy means
 - 1. Replication of the same object (software or hardware) or
 - 2. Diversity
 - Design or implementation
 - Hardware or software



- D Origin
 - Latin verb: "resilire" ~ jump back
- Resilience definition in different fields
 - Physics
 - A material's property of being able to recover to a normal state after a <u>deformation</u> resulting from external forces;
 - Ecology
 - Moving from a stability domain to another under the <u>influence of</u> <u>disturbance;</u>
 - Psychology and psychiatry
 - Living and developing successfully when facing <u>adversity;</u>
 - Business
 - the capacity to reinvent a business model before <u>circumstances</u> <u>force to;</u>



Definition:

"Resilience is the persistence of <u>dependability</u> when facing <u>changes</u>."

J.-C. Laprie. "From Dependability to Resilience". In 38th International Conference On Dependable Systems and Networks. IEEE/IFIP, 2008.

□ Changes can be particularly *attacks*





- Availability
 - Readiness for correct service
- Reliability
 - Continuity of correct service
- □ Safety
 - Absence of catastrophic consequences on the user(s) and the environment
- □ Integrity
 - Absence of improper system alterations
- Maintainability
 - Ability to undergo repair and modification



- □ "CIA" model
 - <u>Confidentiality</u>, <u>Integrity</u>, <u>Availability</u>
- Confidentiality
 - Absence of unauthorized disclosure of information
- □ Availability
 - Readiness for correct service
- □ Integrity
 - Absence of improper system alterations
- □ Notes:
 - CIA model actually not sufficient to describe "security"
 - "Security" addresses all kind of possible <u>attacks</u> which may lead to the deviation from correct service



 The <u>reliability</u> of a unit at a point of time *t* is the probability that the unit is operational <u>until</u> *t*

R(t) = Pr [unit is operating <u>until</u> t]

 The <u>availability</u> of a unit at a point of time *t* is the probability that the unit is operational <u>at</u> *t*

A(t) = Pr[unit is operating <u>at</u> t]



□ Mean Time To Failure (MTTF)

- Mean time between
 - Point of time when a unit is put into operation
 - Point of time when the unit fails for the next time
- □ Mean Time To Repair (MTTR)
 - Mean time between
 - Point of time when a unit fails
 - Point of time when the unit is put into operation again
- □ This results into an average availability

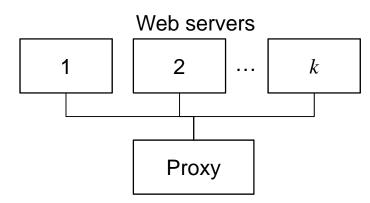
$$A_{avg} = \frac{MTTF}{MTTF + MTTR}$$



DNS lookup (stateless service)

- MTTF: 30 min
- MTTR: 1 ms
- A_{avg} = 0.998
- One can achieve
 - high availability
 - with <u>low</u> reliability (low MTTF)
 - if MTTR is sufficiently low
- Conference bridge (statefull service)
 - Each time, the bridge fails, participants need to re-dial
 - Even if MTTR is sufficiently low, it has to be guaranteed that the MTTF is sufficiently high to assure service quality





$$R_{system}(t) = R_{proxy}(t) \cdot R_{webserver\ pool}(t)$$

$$R_{webserver\ pool}(t) = 1 - (1 - R_{webserver}(t))^{k}$$

□ Same holds for the availability

$$A_{system}(t) = A_{proxy}(t) \cdot A_{webserver\ pool}(t)$$

$$A_{webserver\ pool}(t) = 1 - (1 - A_{webserver}(t))^{k}$$



I. Terminology

II. Challenges in the current Internet

III. Resilience Mechanisms



- 1. Topology Failures
- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. Domino Effects



1. **Topology Failures**

- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. Domino Effects



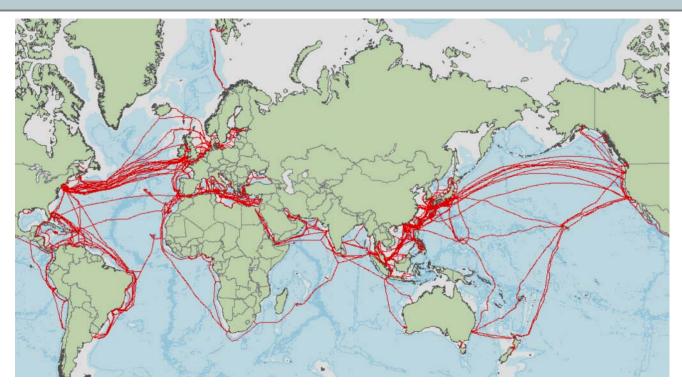
□ Failures in the "network graph"

□ Network graph

- Physical topology
- Logical topology including service dependencies, e.g., DNS

Dependency graphs





- □ ~99% of inter-continental Internet traffic (less than 1% using satellites)
- □ High redundant
- But vulnerable to
 - Fishing and anchoring (70% of sub-marine cable failures)
 - Natural disasters (12%)
 - Cable theft



□ Hengchun earthquake (December 2006)

Bloomk	BloombergAnywhere Updated: New York, Oct 27 1									
?	QUOTE SEARCH NEWS SYMBOL LOOKUP	🖵 Live TV								
HOME NEWS MARKET DATA PERSONAL FINANCE TV and RADIO										
news	Technology Currencies Forex Trading Videos	ETFs CEO Commodi								
 Exclusive Worldwide Regions 	Asian Internet, Phone Services Hit by Taiwan Quakes (Update2) Share Email Print A A A									
□ Markets □ Industries □ Economy □ Politics □ Law	By Tim Culpan and Andrea Tan Dec. 27 (Bloomberg) Internet and telephone services across Asia were disrupted, hampering financial transactions, after earthquakes near Taiwan damaged undersea cables.									
 Environment Science Opinion Spend 	``The repairs could take two to three weeks," said Leng Tai-feng , president of Chunghwa Telecom Co. 's international business. The Taipei-based company, Taiwan's largest phone operator, said two of its undersea cables were cut.									



Submarine Cables; Natural Disasters

□ Hengchun earthquake (December 2006)

□ Impact

- Affected countries: China, Taiwan, Hong Kong, Philippines
- China's Internet connectivity reduced by 70%
- Hong Kong's Internet access completely disabled

□ Recovery

- BGP automatic re-routing helped to reduce disconnectivity
- But resulted into congested links
- Manual BGP policy changes + switch port re-configuration were necessary
- Hong Kong's Internet users were still experiencing slow Internet connections 5 days after the earthquake



□ In Jan. + Feb. 2008, 3 successive events

□ Impact

- Affected countries: Egypt, Iran, India and a number of other middle east countries
- Disruption of
 - 70% in Egypt
 - 60% in India



- In March 2007, pirates stole an 11 kilometers section of the submarine cable connecting Thailand, Vietnam and Hong Kong,
- □ Impact: significant downgrade in Internet speed in Vietnam.
- □ Intention: The thieves wanted to sell 100 tons of cable as scrap.



□ Failures in the IP topology graph

- Failures of routers (nodes)
- Failure of links between routers
- Failure of links between routers generally caused by disconnection at lower layers
- □ Failure of routers
 - DoS attacks
 - Failures due to software bugs
 - Examples of reported bugs
 - Vulnerability to too long AS (BGP <u>Autonomous Systems</u>) paths
 - Long passwords to login to the router
 - Overflow of connection tables in some commercial firewalls



□ Time to Recovery

- Intra-domain routing (OSPF, RIP, IS-IS, EIGRP): up to several 100ms
- Inter-domain routing (BGP): up to several minutes



□ Other reasons

Misconfiguration which leads to false modification of the Internet topology

a	ars) rs techni	ica						
All Apple	Business	Gadgets	Gaming	Hardware	Microsoft	Open Source	Science	Tech Policy
features	content							Subsci
📃 Old Co	ontent							
system, wl		ectively p	otect itse	lf against th		ut into the Inte istake?or attac		ing
cutesy kitte to Pakistan video sharir YouTube's routing infor	en-with-screwe . Which was ng site. What servers are in mation escap	driver page, curious, be apparently to a "black bed from Pa	nothing. V cause the happened hole" as a kistan Tele	Vhat happene Pakistan gov is that Pakis simple meas acom to its IS	ed was that p ernment had tan Telecom sure to filter a SP PCCVV in	nternet. No "sorr just instituted a routed the addre iccess to the se Hong Kong, wh in Pakistan Tel	YouTube we ban on the ess block th rvice. Howe ich propaga	ere flowing popular nat wer, this ted the



- 1. Topology Failures
- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. Domino Effects



- □ Topology failures are binary (link or node is up or down)
- But equipment in the network (routers, servers, etc.) have limited capacity
 - Queue length
 - CPU power
 - etc.

Overload (congestion) is not rare



Lack of Congestion at the Network Layer

- □ Routing protocols react to the failure of a link or a router.
- □ But not to network congestions
- □ ARPANET had some mechanisms to react to congestions
- But they resulted into oscillations
- Congestion control was introduced in the Internet as enhancement of TCP
- But TCP has
 - no knowledge about the network topology
 - no way of re-wiring the traffic path in case of congestion



□ Big challenge

- Ambiguous differentiation between DoS attacks and <u>flash</u>
 <u>crowds</u>
- <u>Flash crowds</u>: unusual but legitimate traffic
- Even if attacks are identified as such, it remains difficult to separate between malicious and legitimate traffic and to eliminate the malicious traffic

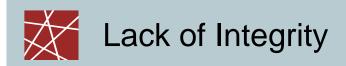


□ Some DoS attacks have a political or ethnical reasons





- 1. Topology Failures
- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. Domino Effects



Majority of Internet traffic (signaling and data) is not integrityprotected

- □ This leads to several security vulnerabilities
 - ARP poisoning
 - Forged BGP announcements
 - Forged DNS responses

 - etc.



- 1. Topology Failures
- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. Domino Effects



- Developments faults
 - Introduced during the development phase
- Configuration faults
 - Introduced during the deployment phase



Examples

- Buffer overflows in server or router implementation
- BGP Youtube misconfiguration
- On Jan. 31st 2009, Google search engine marked every search result with "This site may harm your computer"; Root cause: Database of suspected sites was mistakenly extended by ,/'
- Software update of the Authentication Server (Home Location Register HLR) of T-Mobile on April 21st 2009
 - Impact: phone calls and text messaging were not possible for 4 hours



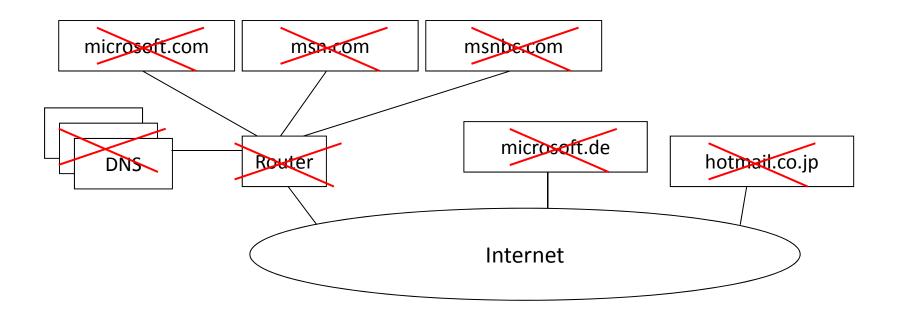
- 1. Topology Failures
- 2. Overload
- 3. Lack of Integrity
- 4. Software Faults
- 5. **Domino Effects**



- Any kind of challenges mentioned above may lead to other challenges
 - E.g., failure of a server in a server pool may lead to overload of neighboring servers
 - Router failures may lead to congestion of neighboring links and routers
 - DNS failure may lead to unavailability of other services,



E.g., DoS attack on Microsoft router on 24th + 25th Jan. 2001 lead to unavailability of DNS and thus of services located in other MS sites





- I. Terminology
- II. Challenges in the current Internet
- **III. Resilience Mechanisms**



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



1. **Topology Protection**

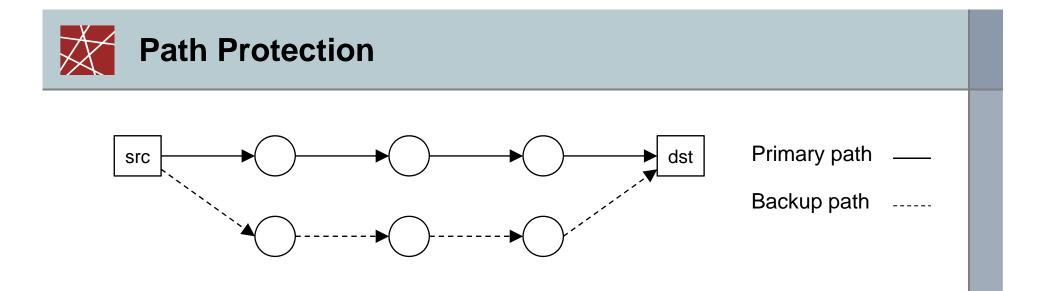
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



- Several metrics exist
- But not all are useful

Definitions

- <u>k-link (edge) connectivity</u> is the minimal number of links whose removal would disconnect the graph
- <u>k-node (vertex) connectivity</u> is the minimal number of nodes whose removal (including removal of adjacent links) would disconnect the graph
- A <u>k-regular graph</u> is k-node-connected if there are k nodedisjoint paths between any pair of nodes.



□ Traffic is forwarded using backup path in case of failure

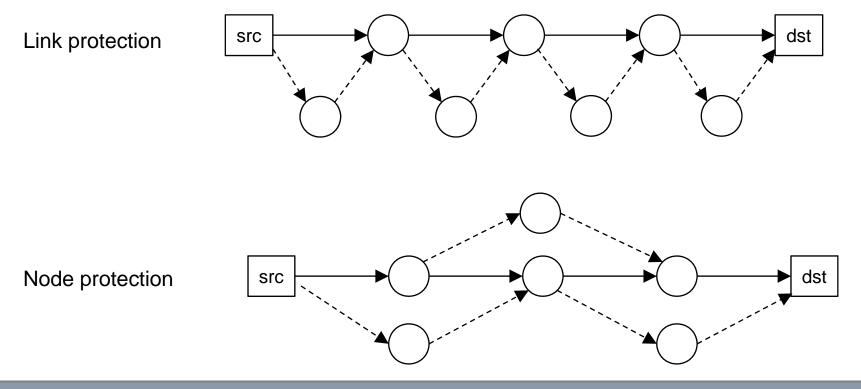
□ Source needs to monitor the operation of primary path

Info about node or link failure needs to be propagated back to src

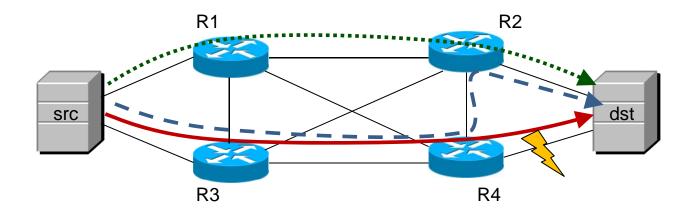


Node or link failures are detected locally and backup paths are used until routing re-converges

- This can reduces the MTTR by the order of a magnitude compared to path protection
- Contra: higher signaling and equipment overhead



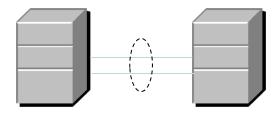




- □ Location protection at IP layer
- □ Routing protocol: OSPF
- □ Local protection according to IP Fast Reroute (IPFRR) (RFC 5714)
- 1. Normal operation: Routing from src to dst via R3 and R4
- 2. After failure of link between R4 and dst: Rerouting from R4 to dst via R2
- 3. Then, info is propagated in the network, OSPF routing converges and a new path is used from src to dst via R1 and R2.



- □ IEEE Link Aggregation allows for bundling
 - several physical Ethernet connections
 - into a logical one
- Connection between
 - Two hosts
 - Two Ethernet switches
 - Host and switch



- □ IEEE Link Aggregation allows for increasing bandwidth
- But is also a fault tolerance mechanism
 - If a cable is plugged out,
 - e.g., for maintenance reasons,
 - the two layer-2 devices remain connected.



Multihoming refers to a network setup where a host or a network is connected to the Internet via more than 1 connection

□ It can be applied in various contexts

- Host Multihoming
 - An IP host connected via multiple network interfaces
 - Each network interface might be connected to a different access network
- Multihoming at the transition point between networks
 - An enterprise network connected to the Internet via multiple ISPs
 - BGP peering with multiple providers



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks

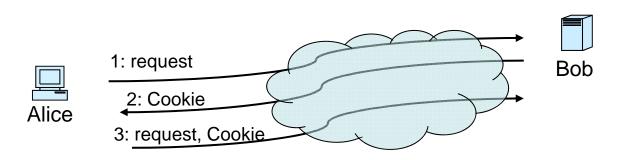


- □ TCP congestion control
- □ Traffic Engineering
- □ Protection again DoS attacks
 - Rate limiting: vulnerable to
 - "false positives", i.e., legitimate traffic is classified as malicious
 - "false negatives", i.e., malicious traffic is classified as legitimate
 - Cookies

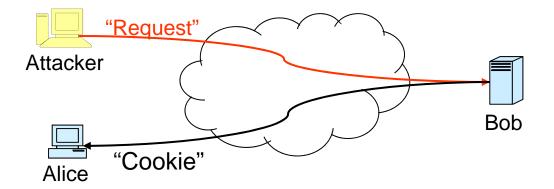


- □ Addresses network congestion at the network layer
- Goals
 - Optimize network throughput, packet loss, delay
- Input
 - Network topology
 - Traffic matrix (may change over time, e.g., daily patterns)
- Output
 - (Eventually modified) link weights used to compute routing tables





- □ Upon receiving a request from Alice, Bob calculates a Cookie and sends it to Bob.
- □ Alice will receive the Cookie and resend the request with the Cookie together.
- Bob verifies that the Cookie is correct and then starts to process Alice's request.
- An attacker that is sending requests with a spoofed (i.e. forged) source address will not be able to send the Cookie.





- □ Cookies discussion:
 - Advantage: allows to counter simple address spoofing attacks
 - Drawbacks
 - Requires CPU resources
 - In some applications, e.g., DNS, it might be easier to respond to the request than generating the cookie
 - Requires one additional message roundtrip.
 - Network may remain congested



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



Manual configuration, e.g., ARP messages with wrong matching (IP to MAC) are discarded

Too costly

□ IPv6 SEcure Neighbor Discovery (SEND) (RFC 2461 and 2462)

 Uses a Cryptographically Generated Address (CGA)

Routing prefix	Hash62(Host public key)
----------------	-------------------------



□ Protects DNS responses with cryptographic signatures

□ In a dedicated DNS record: the RRSIG record (RFC4034)

DNS Records can be verified with a "chain of trust"

- Public key of the DNS root zone must be known by clients
- □ Authority delegation is restricted to sub-domains
 - e.g., system administrator of "net.in.tum.de" can not sign records for "lrz.de"
 - Note: this is not the case for PKIs currently used in the web



Not trivial

Can not be solved by simply adding message integration protection of BGP announcements

- E.g., what is if "Pakistan Telecom" signs BGP announcements for a Youtube prefix?
- Integrity of BGP announcements needs to be validated by a combination of
 - ∽ topology authentication,
 - BGP path authentication and
 - ∽ announcement's origin authentication



Domain Keys Identified Mail (DKIM)

- Allows for validation of a domain name associated with an email address
- An organization takes responsibility for a message in a way that can be validated by a recipient
- Prominent email service providers implementing DKIM
 - Yahoo, Gmail, and FastMail.
 - Any mail from these organizations should carry a DKIM signature



□ Spammers can still sign their outgoing messages

- ☞ DKIM should be used with reputation:
 - Email messages sent by a domain that is known for signing good messages can be accepted
 - while others may require further examination.



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



□ Server redundancy as a *fault tolerance* mechanism

- □ Servers instances may be
 - in the same LAN or
 - different sub-networks
 Geographic diversity

- □ Supporting mechanisms
 - IP Takeover
 - NAT Takeover
 - DNS



□ Simple redundancy mechanism

Backup server receives periodic "keep alive" messages from master server, e.g., every 10ms

 $\hfill\square$ In case of no response

- Backup server broadcasts an ARP message in the LAN
- From now on, all IP traffic is forwarded to the backup server

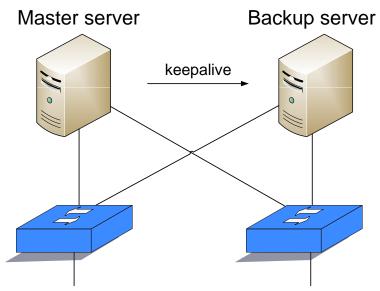
Drawbacks
 Existing session state gets lost
 Ethernet switch is a single point of failure



Server Redundancy; IP Takeover with 2 Switches

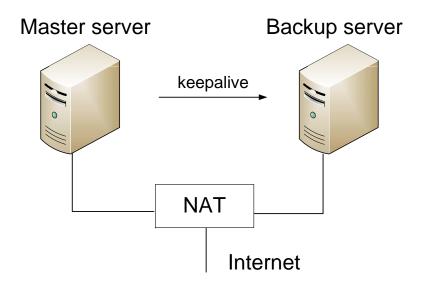
□ Both master and backup servers are connected to 2 switches

- $\hfill\square$ Same procedure with ARP
 - Incoming requests from both switches is forwarded to the backup server
- Any component (server or switch or cable) can be removed, e.g., for maintenance reasons, while the service keeps on being available





- □ Similar to IP Takeover
- □ "Keep alive" messages from backup to master server
- □ Change NAT binding upon lack of response from master server
 - ☞ Incoming requests are forwarded to the backup server



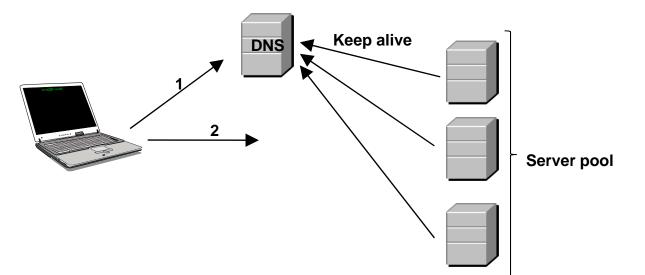
□ Note: Master and backup server do not have to be in the same LAN



□ DNS can provide several IP addresses for the same name

□ By monitoring the availability of servers from a server pool,

unavailable servers can be removed from DNS responses



Moreover, DNS responses can be adjusted according to the current load

∽ See, e.g., Content Distribution Networks (CDN)



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



Different virtualization techniques, e.g., KVM, Xen, etc.

□ Can be used to enhance resilience of network services

- Start new servers from existing images on demand, e.g.,
 - To address overload situations
 - In case servers in other locations crash



- 1. Topology Protection
- 2. Congestion Control
- 3. Signaling Integrity
- 4. Server Redundancy
- 5. Virtualization
- 6. Overlay and P2P Networks



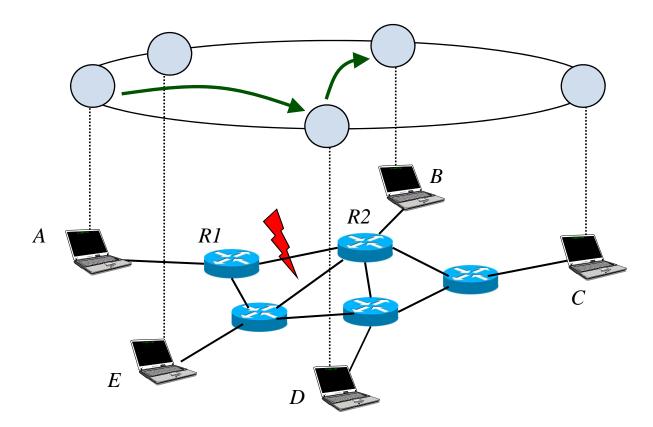
Overlay networks

- Are networks built on top of existing networks
- They typically provide additional functionality not provided at the "underlay" network
- Overlay routing
 - End hosts can organize themselves in a P2P network
 - and provide routing using the overlay in case the underlay routing fails



□ Example

- Upon link failure between R1 and R2
- A can reach B via D or C





□ Typical reasons for lack of connectivity in the underlay

- Misconfigured middleboxes (firewalls, NATs)
- Slow BGP convergence
- □ Systems supporting overlay routing
 - Tor
 - while it is actually designed with anonymization in mind, it provides overlay routing and can be useful in case of network partial failures
 - Skype
 - Skype supernodes typically provide connectivity for Skype clients behind firewalls or NATs



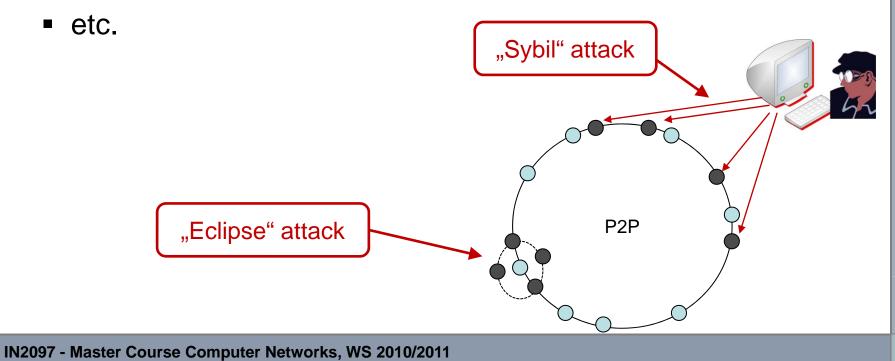
Resilience properties

- Decentralization
- Geographic diversity
- Ability to cope with "churn"
 - "Churn" means that peers join and leave at any time
 - ∽ Replication of each data item on several peers
 - ∽ Autonomic recovery from stale P2P routing tables



Drawback: several attacks are possible

- Sybil attacks:
 - Attacker participate with several fake identities
 - In order to control a portion of the network
- Eclipse attacks,
 - Attacker control the neighborhood of a peer or content
 - In order to make unavailable for other participants in the P2P networks





Common approaches

- Managed P2P networks (or supervised P2P networks)
- ☞ E.g., Google File System (GFS), Skype



I. Terminology

- □ The "fault → error → failure" chain
- □ Fault tolerance, Resilience, Dependability, Security
- □ Availability vs. Reliability

II. Challenges in the current Internet

- D Topological Failures, Overload, Lack of Integrity
- Software Faults, Domino Effects

III. Resilience Mechanisms

- D Topology Protection, Congestion Control, Signaling Integrity
- □ Server Redundancy, Virtualization, Overlay and P2P Networks