On the State of ECN and TCP Options on the Internet

PAM 2013, March 19, Hong Kong

Mirja Kühlewind <mirja.kuehlewind@ikr.uni-stuttgart.de>
Sebastian Neuner <sebastian.neuner@ikr.uni-stuttgart.de>
Brian Trammell <trammell@tik.ee.ethz.ch>

Universität Stuttgart
Institute of Communication Networks
and Computer Engineering (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

Overview

- Motivation: Measurement Study on TCP Extensions
- Measurement Methodology and Results of ECN and TCP Option Deployment
  - Active Probing of Webservers
  - Deployment in IPv6
  - Passive Measurement of ECN Adoption
- Burst Loss Study
- Conclusion

A Measurement Study on TCP Extensions

Goal Monitoring of deployment of ECN and most common used TCP options

Extensions to improve TCP throughput and congestion control characteristics

Reference	Date	ECN	SACK	TSOPT
Medina *et al.*	2000	1.1%	28%	-
Medina *et al.*	2004	2.1%	68%	30%
Langley *et al.*	2008	1.06%	-	-
Bauer *et al.*	2011	17.2%	-	-

ECN: Explicit Congestion Notification, SACK: Selective Acknowledgement Option, TSOPT: Timestamp Option

Explicit Congestion Notification (ECN)

- Hosts negotiate ECN support in TCP handshake (using TCP header flags)
- Routers can mark ECN-capable packets in case of congestion

  → *Congestion Experience* (CE) codepoint in IP header

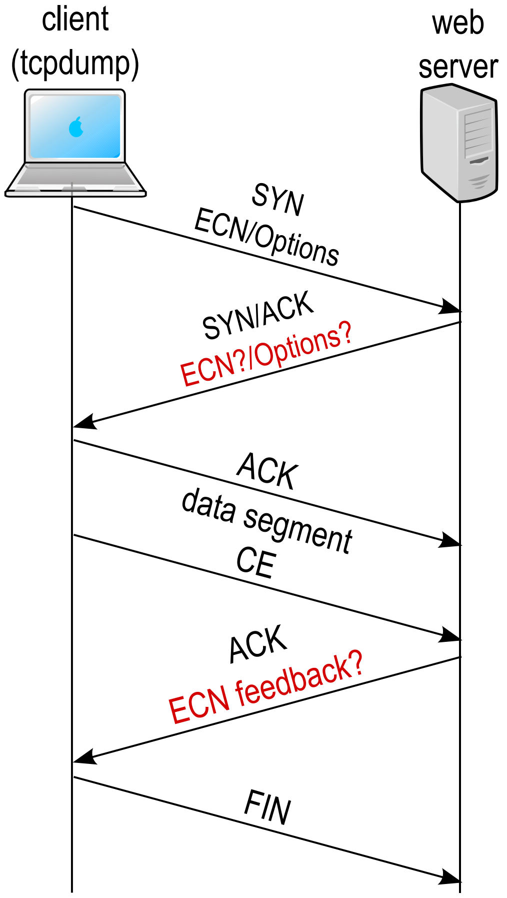
  → Avoid buffer overflow/packet loss

Deployment Problems

- Endsystem and routers need to support ECN
- Middleboxes and firewalls clearing ECN marks or dropping ECN-capable packets

Active Probing of Webservers

Goal Probing of the ECN-readiness and TCP option support of webservers from *Alexa Top 100,000* webserver list



Measurement Methodology

1. Sending of a TCP SYN to target server
   - ECN negotiation (flag in TCP header)
   - Selective Acknowledgement Option (SACK)
   - Timestamp Option (TSOPT)
   - Window Scaling Option (WSOPT)
2. Capturing of the resulting SYN/ACK (using tcpdump)
3. Sending of one data segment with CE codepoint (if target server is ECN capable)
4. Capture ACK and send FIN

Offline Evaluation

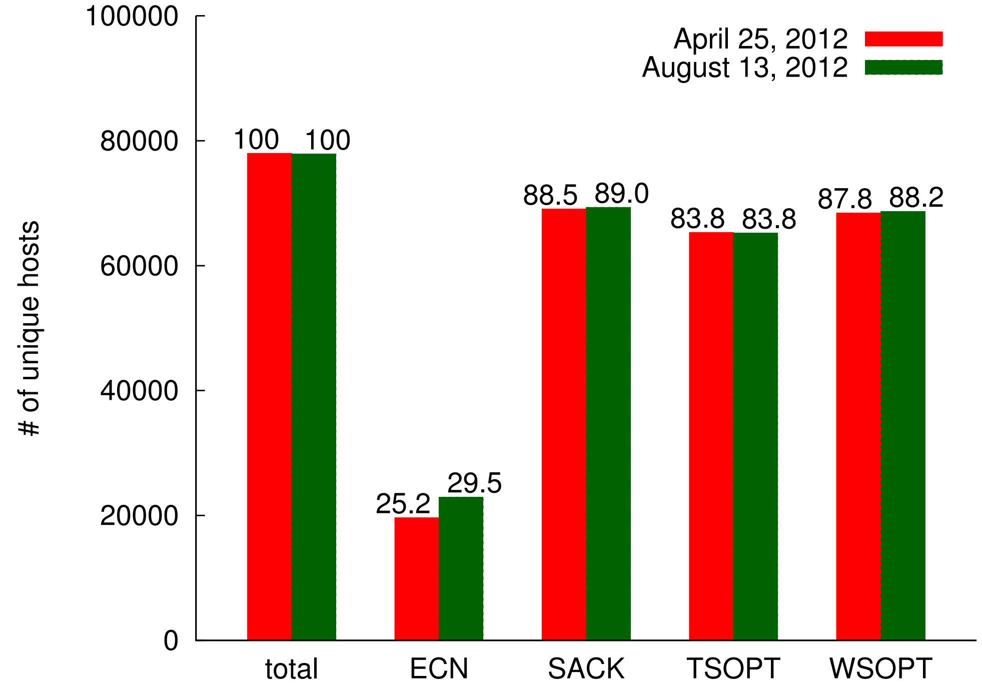
- Check ECN negotiation and TCP options in SYN/ACK
- Check ECN feedback in data ACK to determine whether ECN is usable on a path
- IP Time-to-Live (TTL) as an estimate of the operating system

Active Probing of Webservers

Results (1)

ECN and TCP Option Deployment

Measurement (April/Aug 2012): Alexa Top 100,000 webservers



→ Increase in ECN deployment (29.5% in August 2012)

ECN is better supported on Linux (38.5%) than on Windows (4.9%) hosts

Active Probing of Webservers

Results (2)

ECN usability on the path

- Measurement (Aug 2012): 22487 ECN-capable hosts (CE set on data segment)
  - 90.9% sent correct ECN feedback in response to an congestion notification
  - 8.2% replied with an ACK without ECN feedback
  - 0.9% sent no ACK at all
- UMTS networks (two German mobile networks)

  100% ECN support but 0% ECN feedback (ECN-capable proxies, clearing of CE)

→ Middleboxes still significantly affect the end-to-end usability of ECN in the network

Active Probing of Webservers

Results (3)

General responsiveness

Measurement (Sep 2012): Alexa top 100,000 servers (probing without ECN negotiation or any TCP options in SYN)

– 429 more responses than when probing with ECN negotiation and TCP options in SYN
– 1.1% responding with SACK in SYN/ACK (even though not requested)
– 0.4% responding with WSOPT in SYN/ACK (most presumably Windows hosts)
– None responding with TSOPT or ECN in SYN/ACK

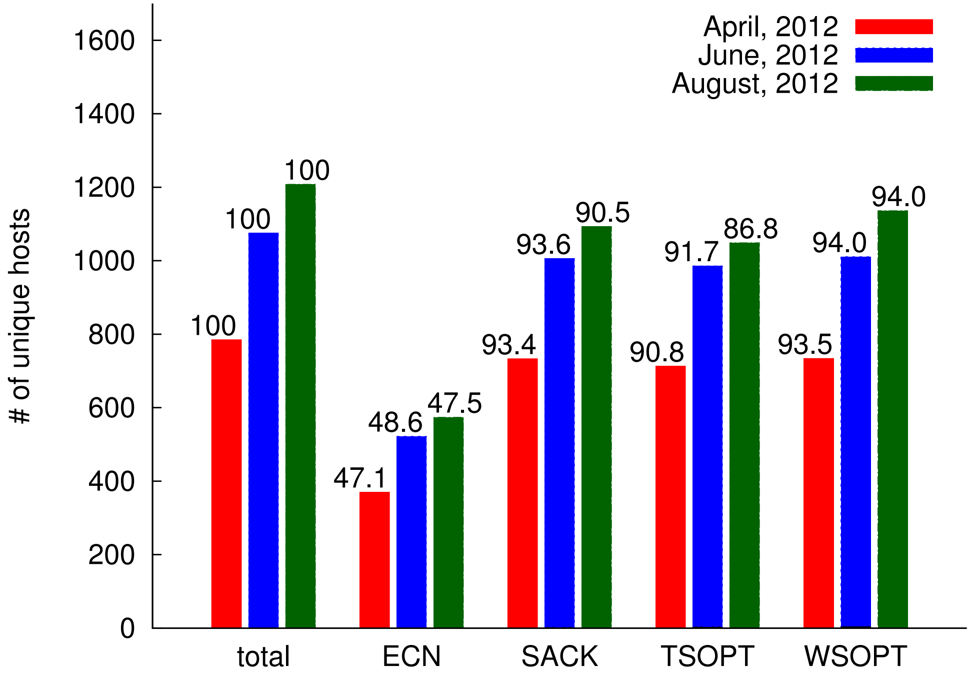
→ Small number of non standard conform hosts

Active Probing of Webservers

Deployment on IPv6

ECN and TCP option support of IPv6 enabled webservers

Measurement (April/June/Aug 2012): Alexa Top 100,000 webserver



→ Increase after World IPv6 Launch event on 6 June 2012

but still only 2.3% support IPv6 as of August 2012

→ Larger ECN (47.5%) and TCP options (~ 90%) support than in IPv4

but no increase over time

Passive Measurement of ECN Adoption

Analysis of Aggregated Flow Data

Goal Analyze the actual use of ECN in the network

Measurement Methodology

- Flow data (NetFlow version 9) from the border of SWITCH (Swiss national research and education network)
  - 2.4M IPv4 addresses
  - Daily traffic volume on the order of 100 TB
- Data of four (out of six) border routers

Evaluation

Check ECN field in the IP header of continued flows

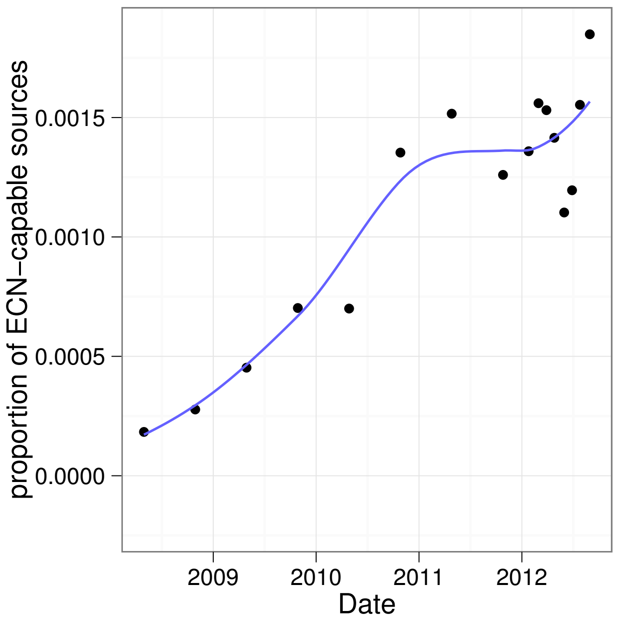
- IP header fields are recorded only for the first packet of each flow record
- SYN should never be ECN-capable
- Continued flows: more than one record (records get exported on active timeout of 300s)

→ Only flows that are longer than the active timeout are regarded

→ Lower bound for ECN-capable sources

Passive Measurement of ECN Adoption

Results on ECN Deployment



Historical trend in ECN capability

Data set: 13:00 - 14:00 UTC of the last Wednesday of the month

Increase from 0.02% in 2008 to 0.18% in 2012

ECN usage

Data set: full day Wednesday, August 29, 2012

- 0.16% (11,039) distinct ECN-capable IPv4 sources
  - 0.77% only of continued flows
  - 3.01% (1.77TB of 58.84TB) of measured TCP traffic volume
  - 19 public web servers (13 appear in the Alexa list) in top 50 ECN-capable sources
- 24,580 sources use erroneously ECN IP bits in TCP SYN packet

  68.9% of a single ISP which sets the CE codepoint on 99.1% of its outgoing traffic
- Not a single continued flow with CE set (first packet only; except erroneous sources)

→ Hosts and devices supporting ECN are seeing increased deployment, but ECN is mostly not used :-(

Burst Loss Study

Identifying Conditions of Congestion

Goal Analysis of loss pattern in typical Internet usage scenarios in the absence of ECN

Measurement & Evaluation

- Active measurement and offline evaluation based of estimation of TCP retransmissions
- Metric: burst loss (= # of losses within one RTT after the first loss)

  → Capture loss measurements in a congestion control-aware way

Results

- Regular burst loss pattern of FTP download and YouTube traffic (large burst losses)

  Few probes showed a larger number of small burst losses (presumably caused anomalies in the network or at the server side)

- Irregular but few, small burst losses when web browsing

→ Understanding loss dynamics in relation to application behavior

Conclusion

ECN and TCP Option Deployment

- Active Probing of Webservers

  → Deployment of ECN-capable hosts in the Internet continues (about 30% of Alexa top 100,000 webservers in September)

  → Earlier problems with ECN deployment are not completely solved (failure rate of 9%)

- Deployment in IPv6

  → General increase in IPv6 support over the IPv6 Launch Day (only 47.5% ECN-capable)

- Passive Measurement of ECN Adoption

  → Lower bound for actual ECN usage is two orders of magnitude less common than ECN capability

  → Even worse: twice as many observed sources misused the CE codepoint

Burst Loss Study

Problem: Identifying conditions and sources of congestion

→ Approach: Loss model of transport and application behavior to differentiate losses

caused by the host (transport and application behavior) from losses

caused by the network or influenced by cross traffic

Thank you for your attention!

Questions?

Backup

Active Probing of Webservers

ECN and TCP Option Deployment on IPv4

April 25, 2012, 77969 unique hosts (of 93573 responding hosts)

	All	TTL < 64	64 < TTL < 128	TTL > 128
hosts	77969 (100.00 %)	57610 (100.00 %)	12794 (100.00 %)	7590 (100.00 %)
ECN	19616 (25.16 %)	18954 (32.90 %)	521 (4.07%)	143 (1.88%)
SACK	69037 (88.54 %)	52409 (90.97 %)	11506 (89.93 %)	5145 (67.79 %)
TSOPT	65307 (83.76 %)	49667 (86.21 %)	10729 (83.86 %)	4928 (64.93 %)
WSOPT	68419 (87.75 %)	53137 (92.24 %)	10047 (78.53 %)	5258 (69.28 %)

August 13, 2012, 77854 unique hosts (of 93756 responding hosts)

	All	TTL < 64	64 < TTL < 128	TTL > 128
hosts	77854 (100.00 %)	57651 (100.00 %)	12471 (100.00 %)	7769 (100.00 %)
ECN	22948 (29.48 %)	22193 (38.50 %)	616 (4.94%)	145 (1.87%)
SACK	69334 (89.06 %)	52783 (91.56 %)	11226 (90.02 %)	5353 (68.90 %)
TSOPT	65220 (83.77 %)	49749 (86.29 %)	10379 (83.23 %)	5112 (65.80 %)
WSOPT	68684 (88.22 %)	53420 (92.66 %)	9846 (78.95 %)	5446 (70.10 %)

Active Probing of Webservers

ECN and TCP Option Deployment on IPv6

	IPv4 Aug'12	IPv6 April'12	IPv6 June'12	IPv6 Aug'12
responding hosts	93573	980	1819	2132
unique hosts	77854 (100.00 %)	785 (100.00 %)	1075 (100.00 %)	1208 (100.00 %)
ECN	22948 (29.48 %)	370 (47.13 %)	522 (48.56 %)	574 (47.52 %)
SACK	69334 (89.06 %)	733 (93.38 %)	1006 (93.58 %)	1093 (90.48 %)
TSOPT	65220 (83.77 %)	713 (90.83 %)	986 (91.72 %)	1049 (86.84 %)
WSOPT	68684 (88.22 %)	734 (93.50 %)	1011 (94.05 %)	1136 (94.04 %)

Active Probing of Webservers

Number of Hops

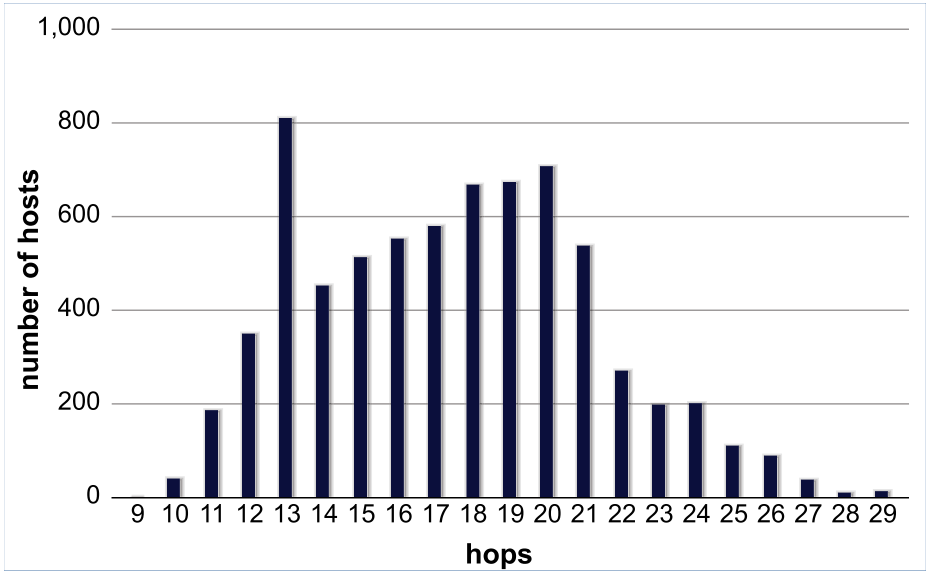
Measrement Methodology

- Top 10,000 servers
- Based on ICMP traceroute

→ Ensure less than 64 hops and thus validate the start TTL estimates

Results

9 hops within BelWü network



Identifying Conditions of Congestion

Burst Loss Study

Measurement from September 10, 2012

	mean # of packets	mean # of retrans-missions	mean loss rate	mean # of burst losses	mean # of packets per burst loss	time between burst losses
Web-browsing	80779	533.96	0.66 %	227.88	2.37	-
Download (all)	58643	703.04	1.2 %	535	2.10	2.88
Download (21 of 24)	58639	76.14	0.13 %	34.29	2.23	3.28
YouTube1 (11.59MB)	8469.2	176.29	2.08 %	5.58	31.72	27.31
YouTube1 (18 of 24)	8469.4	159.83	1.89 %	5	31.97	29.40
YouTube2 (4.62MB)	3386.2	34.04	1 %	1	34.04	-

Identifying Conditions of Congestion

Burst Loss Study

Motivation Understanding loss dynamics in relation to application behavior in the absence of ECN

Goal Analysis of loss pattern in typical Internet usage scenarios

Measurement Methodology

- Traffic
  - Web-browsing: request 33 common websites with a 12 second delay
  - Download: FTP download of a 80.56 MB file from a host using cubic congestion control
  - YouTube: Two videos with 4.62 MB and 11.59 MB
- Measurement setup
  - Residential access network with a maximum measured datarate of 5.7 MBit/s
  - Traces of 24 trials over a single day

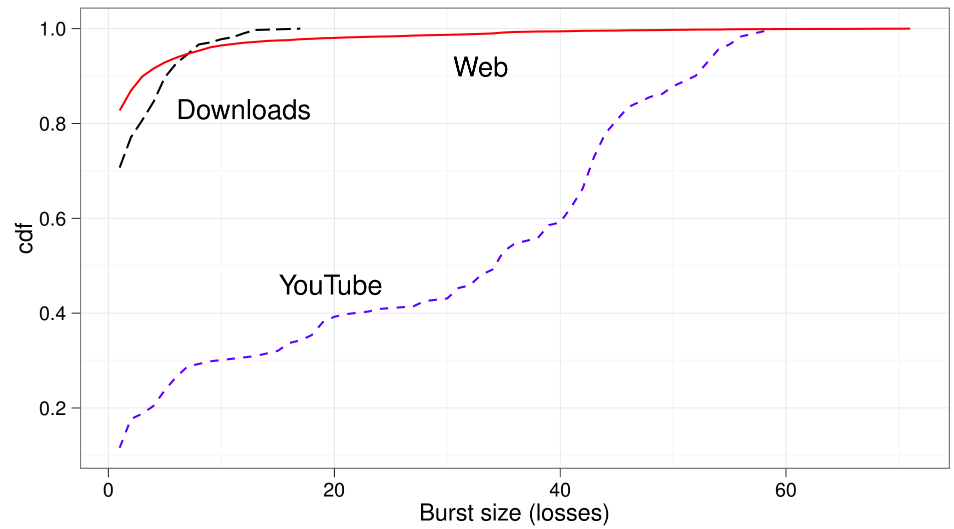
Evaluation

- Offline evaluation based of estimation of TCP retransmissions
- Metric: burst loss (= # of losses within one RTT after the first loss)

  captures loss pattern in a congestion control-aware way

Identifying Conditions of Congestion

Results



- Web browsing (many short flows)
  - Only 5.8% of flows experiencing any loss
  - 82.7% of bursts consist of a single loss
  - Bursts of up to 71 losses occurred
- FTP download (one single, long flow)
  - Very regular loss pattern due to cubic congestion control
  - 70.7% of bursts with only a single loss plus also frequent bursts of up to 12 losses
  - Three probes (at 3am, 10am and 5pm) with a very large number of small burst losses (4058, 3905, and 4157, respectively)

    → Presumably an anomaly in the network or at the server side
- YouTube (block sending)
  - Regular, larger burst loss
  - 18 of 24 trials for the longer video show exactly five bursts
  - Always one burst for the smaller video
  - Mean burst size around 33 (in both cases)

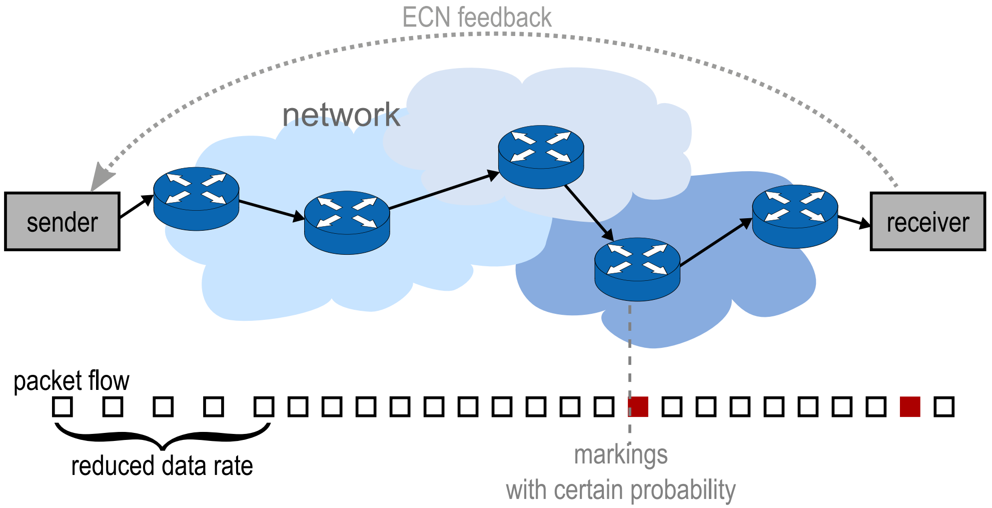
Explicit Congestion Notification (ECN)

Goal Signal congestion to the sender before buffer overflow to avoid packet loss

ECN Codepoints in IP header (two bits)

- 00: Not-ECN capable
- 01/10: ECN-Capable Transport (ETC(0)/ETC(1))
- 11: Congestion Experience (CE)

→ Routers can mark packets as congested



Flags in the TCP header (two bits)

- ECN-Echo (ECE) is set in the ACK when CE is seen by the receiver
- Congestion Window Reduced (CWR) is set by the sender when ECE is seen
- Negotiation ECN support in TCP handshake using ECE and CWR

→ Receiver provides congestion feedback to the sender

Measurement Methodology

1. Active probing of webserver
   - ECN-readiness of a large set of popular web-servers
   - TCP options: Selective Acknowledgment (SACK), Timestamps (TS), Window Scale (WS)
2. Passive measurement of ECN adoption

   ECN usage from flow data collected on a national-scale research and education network
3. Burst loss study

   Analysis of loss pattern in typical Internet usage scenarios (web-browsing, YouTube, FTP)

   → Understanding loss dynamics in relation to application behavior in the absence of ECN