ECSE-4963: Experimental Networking

Informal Quiz
Shivkumar Kalyanaraman: shivkuma@ecse.rpi.edu
TCP

- TCP can re-assemble IP fragments
- Path-MTU refers to the procedure of finding the minimum MTU of the path to reduce the probability of fragmentation.
- The IP header checksum field is the 16-bit two’s complement of the one’s complement sum of all 16-bit words in the header.
- TCP provides reliability only at a packet-level.
- Transport protocols are minimally required because IP does not provide application multiplexing support.
- TCP is called “self-clocking” because the source sends traffic whenever it likes.
- TCP by default uses a selective retransmission policy.
- The RTT estimation algorithm in current can only tolerate variances of up to 30%.
- The TCP congestion control algorithm is stable because it detects congestion reliably and its rate of window decrease is faster than its rate of window increase.
- TCP’s use of cumulative acks reduces the need for any timeout/retransmission of acks.
- Delayed-acks are good for bulk traffic, but bad for interactive traffic.
- A two-way handshake is sufficient for the robust setup of a half-duplex connection, but a three-way handshake is necessary for the robust setup of a full-duplex connection.

Shivkumar Kalyanaraman
TCP

- If timeouts are not used, burst packet or ack-losses cannot be recovered from.
- A duplicate ack gives the same information as a NAK, but it presumes the notion of a sequence number.
- Sequence numbers allow the detection of duplicate packets, but the sequence number space must be sized sufficiently large compared to the window size depending upon the retransmission algorithm (go-back-N or selective-repeat) used.
- In a lossless network, window-based transmission can achieve full utilization.
- TCP sets its RTO to an average RTT measure + 4*mean deviation of RTT, based upon Chebyshev’s theorem.
- Retransmission ambiguity would not occur if timestamps were used on packets.
- Self-clocking of TCP can be a liability in asymmetric networks where the reverse path can artificially constrain the forward path.
- Self-clocking can also lead to burstiness if the reverse path is congested, and/or the receiver uses a delay-ack time to suppress ACKs.
- The end-to-end congestion control model is the only one that can guarantee avoidance of congestion collapse.
- In equilibrium, TCP attempts to conserve packets and operate at high utilization.
- TCP does not guarantee low queueing delays because it depends upon packet loss for congestion detection.

Rensselaer Polytechnic Institute

Shivkumar Kalyanaraman
TCP/Congestion Control

- Fast retransmit refers to the procedure of using three duplicate acks to infer packet loss.
- TCP Tahoe sets its window to 1 after every loss detection.
- TCP Reno may timeout quickly in a multiple packet loss scenario.
- TCP SACK uses selective retransmit, and like NewReno, it does not reduce its window more than once per window of packets.
- With a 28kbps reverse link, 1500 byte packets are regular TCP behavior, the forward link throughput is at most around 2 Mbps.
- FIFO+droptail provides service isolation among the participating TCP flows.
- Synchronization occurs because DropTail leads to bursty and correlated packet losses amongst flows; and flows react to same events.
- Dropping packets early has the risk that transient burstiness may be mistaken for true overload (demand > capacity).
- RED determines random drop probability by comparing the average queue size to a max and min thresholds.
- Random dropping/marking with a bias in RED helps break synchronization.
A probability density function (PDF) is a generalization of a histogram for the continuous random variable case.

A random variable (R.v.) models a measurement, whereas probability models an experiment, and r.v. is used when the measurement does not necessarily captures the set of all possible outcomes of the experiment.

In the experiment of tossing a die, the set $X = \{0, 1, 2\}$ which denotes the possibility of the outcomes being 0, 1 or 2 is a random variable.

A mean of a random variable is also known as the first moment or centroid of a distribution.

A median is the 50th percentile element, found using the inverse of the CDF with an argument of 0.5.

A mean is the preferred central tendency measure in a skewed distribution.

A mode (or the most probable element) is usually used with categorical random variables instead of mean or median.

C.o.V. and SIQR are measures of central tendency.

Covariance, a measure of dependence between random variables, always lies between $-1$ and $+1$. 

Shivkumar Kalyanaraman
Probability/Statistics

- If \( E(XY) = E(X)E(Y) \), the random variables \( X \) and \( Y \) are independent.

- Coefficient of Variation (C.o.V) and Correlation Coefficient (\( \rho_{XY} \)) are normalized measures of spread and dependence respectively.

- The C.o.V would be a useful metric to measure the unfairness of rate allocations to TCP flows passing through a single bottleneck.

- The correlation coefficient would be a useful metric to measure the degree of traffic and window synchronization between a pair of TCP flows competing at a bottleneck.

- Given 50 RTT samples, one can estimate the 95% confidence interval of the path RTT and a good estimate of maximum RTT (to set the timeout value in TCP).

- A Bernoulli distribution can be studied by considering a sequence of \( N \) Bernoulli trials, and counting the number of successes in \( N \) trials.

- Taking a large bet with a probability of success 0.5 in a single experiment (like a lottery, without regard to cost) is superior to taking smaller bets (with probability 0.01 each) in 50 repeated, identical experiments. (Hint: probability of success in latter case is \( 1 - (0.99)^{50} \)).

- The Poisson distribution is a continuous-time approximation of the binomial distribution, derived by assuming np = \( \lambda \), and \( n \) is very large.

- In a Poisson arrival process, the average time since the occurrence of the last arrival is the same as the average time for the next arrival.
The Chebyshev bound for spread of a random variable is a very loose bound, especially for the normal distribution.

The distribution of sample means from any distribution (I.e. sampling distribution, assuming random sampling) tends to a normal distribution.

Confidence interval gives less information compared to the notion of “statistical significance” and “null hypothesis”.

A t-distribution is an approximation of the normal distribution with n-1 degrees of freedom that can be constructed with n samples from a normal population & the approximation is good when n is at least six.

The confidence interval is constructed from a normal or normal-like distribution (eg: t-distribution) of a random variable (eg: the sample mean) by excluding the tails of the distribution based upon the given confidence level.

Pairing and randomized experiments are ways of ensuring the random sampling assumption and reducing correlations between experiments.

If two confidence intervals for an estimate of a mean overlap and the means also lie in the CIs of each other, the means cannot be declared to be different at that level of confidence.
TCP (SOLNS)

- TCP can re-assemble IP fragments (✓)
- Path-MTU refers to the procedure of finding the minimum MTU of the path to reduce the probability of fragmentation (✓)
- The IP header checksum field is the 16-bit two’s complement of the one’s complement sum of all 16-bit words in the header. (✓)
- TCP provides reliability only at a packet-level. (✓)
- Transport protocols are minimally required because IP does not provide application multiplexing support (✗)
- TCP is called “self-clocking” because the source sends traffic whenever it likes. (✓)
- TCP by default uses a selective retransmission policy. (✓)
- The RTT estimation algorithm in current TCP can only tolerate variances of upto 30%. (✓)
- The TCP congestion control algorithm is stable because it detects congestion reliably and its rate of window decrease is faster than its rate of window increase. (✓)
- TCP’s use of cumulative acks reduces the need for any timeout/retransmission of acks (✗)
- Delayed-acks are good for bulk traffic, but bad for interactive traffic. (✓)
- A two-way handshake is sufficient for the robust setup of a half-duplex connection, but a three-way handshake is necessary for the robust setup of a full-duplex connection. (✓)

Rensselaer Polytechnic Institute

Shivkumar Kalyanaraman
TCP (SOLNS)

√ □ If timeouts are not used, burst packet or ack-losses cannot be recovered from

√ □ A duplicate ack gives the same information as a NAK, but it presumes the notion of a sequence number

√ □ Sequence numbers allow the detection of duplicate packets, but the sequence number space must be sized sufficiently large compared to the window size depending upon the retransmission algorithm (go-back-N or selective-repeat) used.

√ □ In a lossless network, window-based transmission can achieve full utilization

√ □ TCP sets its RTO to an average RTT measure + 4*mean deviation of RTT, based upon Chebyshev’s theorem

√ □ Retransmission ambiguity would not occur if timestamps were used on packets.

√ □ Self-clocking of TCP can be a liability in asymmetric networks where the reverse path can artificially constrain the forward path.

√ □ Self-clocking can also lead to burstiness if the reverse path is congested, and/or the receiver uses a delay-ack time to suppress ACKs.

√ □ The end-to-end congestion control model is the only one that can guarantee avoidance of congestion collapse.

√ □ In equilibrium, TCP attempts to conserve packets and operate at high utilization.

√ □ TCP does not guarantee low queueing delays because it depends upon packet loss for congestion detection
TCP (SOLNS)

√  □ Fast retransmit refers to the procedure of using three duplicate acks to infer packet loss

√  □ TCP Tahoe sets its window to 1 after every loss detection

√  □ TCP Reno may timeout quickly in a multiple packet loss scenario

√  □ TCP SACK uses selective retransmit, and like NewReno, it does not reduce its window more than once per window of packets

√  □ With a 28kbps reverse link, 1500 byte packets & regular TCP behavior, the forward link throughput is at most around 2 Mbps

□  √  FIFO+droptail provides service isolation among the participating TCP flows

√  □ Synchronization occurs because DropTail leads to bursty and correlated packet losses amongst flows; and flows react to same events

√  □ Dropping packets early has the risk that transient burstiness may be mistaken for true overload (demand > capacity)

√  □ RED determines random drop probability by comparing the average queue size to a max and min thresholds

√  □ Random dropping/marking with a bias in RED helps break synchronization
A probability density function (PDF) is a generalization of a histogram for the continuous random variable case.

A random variable (R.v.) models a measurement, whereas probability models an experiment, and r.v. is used when the measurement does not necessarily captures the set of all possible outcomes of the experiment.

In the experiment of tossing a die, the set \( X = \{0,1,2\} \) which denotes the possibility of the outcomes being 0, 1 or 2 is a random variable.

A mean of a random variable is also known as the first moment or centroid of a distribution.

A median is the 50th percentile element, found using the inverse of the CDF with an argument of 0.5.

A mean is the preferred central tendency measure in a skewed distribution.

A mode (or the most probable element) is usually used with categorical random variables instead of mean or median.

C.o.V. and SIQR are measures of central tendency.

Covariance, a measure of dependence between random variables, always lies between –1 and +1.
If \( E(XY) = E(X)E(Y) \), the random variables \( X \) and \( Y \) are independent.

Coefficient of Variation (C.o.V) and Correlation Coefficient \( (\rho_{XY}) \) are normalized measures of spread and dependence respectively.

The C.o.V would be a useful metric to measure the unfairness of rate allocations to TCP flows passing through a single bottleneck.

The correlation coefficient would be a useful metric to measure the degree of traffic and window synchronization between a pair of TCP flows competing at a bottleneck.

Given 50 RTT samples, one can estimate the 95% confidence interval of the path RTT and a good estimate of maximum RTT (to set the timeout value in TCP).

A Bernoulli distribution can be studied by considering a sequence of \( N \) bernoulli trials, and counting the number of successes in \( N \) trials.

Taking a large bet with a probability of success 0.5 in a single experiment (like a lottery, without regard to cost) is superior to taking smaller bets (with probability 0.01 each) in 50 repeated, identical experiments. (Hint: probability of success in latter case is \( 1 - (0.99)^{50} \)).

The Poisson distribution is a continuous-time approximation of the binomial distribution, derived by assuming \( np = \lambda \), and \( n \) is very large.

In a Poisson arrival process, the average time since the occurrence of the last arrival is the same as the average time for the next arrival.
The Chebyshev bound for spread of a random variable is a very loose bound, especially for the normal distribution.

The distribution of sample means from any distribution (i.e. sampling distribution, assuming random sampling) tends to a normal distribution.

Confidence interval gives less information compared to the notion of “statistical significance” and “null hypothesis”.

A t-distribution is an approximation of the normal distribution with n-1 degrees of freedom that can be constructed with n samples from a normal population & the approximation is good when n is at least six.

The confidence interval is constructed from a normal or normal-like distribution (e.g: t-distribution) of a random variable (e.g: the sample mean) by excluding the tails of the distribution based upon the given confidence level.

Pairing and randomized experiments are ways of ensuring the random sampling assumption and reducing correlations between experiments.

If two confidence intervals for an estimate of a mean overlap and the means also lie in the CIs of each other, the means cannot be declared to be different at that level of confidence.