Project	**IEEE 802.16 Broadband Wireless Access Working Group <http://ieee802.org/16>**
Title	**Multihop System Evaluation Methodology – Performance Metrics**
Date Submitted	**2006-05-01**
Source:	Gamini Senarath, Wen Tong, Peiying Zhu, Hang Zhang, David Steer, Derek Yu, Mark Naden, Dean Kitchener Nortel 3500 Carling Avenue Ottawa, On, K2H 8E9 Canada Gamini@nortel.com wentong@nortel.com
Re:	Response to a call for contributions for the Relay TG, see C80216j-06/001.pdf
Abstract	This document provides performance metrics proposals to be included in the evaluation methodology document.
Purpose	To propose performance metrics for multi-hop systems.
Notice	
Release	
Patent Policy and Procedures	

Multi-hop System Evaluation Methodology: Performance Metrics

Gamini Senarath, Wen Tong, Peiying Zhu, Hang Zhang,
David Steer, Derek Yu, Mark Naden, Dean Kitchener
Nortel

1 Introduction

This document provides a set of definitions and assumptions related to performance metrics for evaluating multihop relay systems (e.g. 802.16j, LTE relay extensions) to arrive at system wide voice, data, video or mixed data, voice, video performance on the forward and reverse links. This document also provides the descriptions required for modeling higher layers which include TCP parameters as well as network modeling.

The performance metrics are divided into two categories. They are:

- Single-user performance; and
- Multi-user performance.

Examples of single-user performance metrics are the link budget margins, C/I area coverage and data rate area coverage. These metrics are evaluated assuming that a single user is in a particular cell area utilizing all the resources in that cell and external interference may be evaluated assuming that at least a single active user is available in the external cell (for both forward and reverse link). These metrics are not end-to-end performance metrics and therefore, could be evaluated without modeling higher layer protocols and is independent of applications.

However, when multiple users are in the system the system resources have to be shared and the single-user rates are not available to users. Therefore, multi-user metrics indicates how a system behaves under a multi-user environment.

2 Single-user performance Metrics

Note that area coverage mentioned here is equivalent to percentage of users meeting a given requirement when the users are uniformly distributed in the interested geographical area.

2.1 Noise limited performance – single-cell performance

Since relays can be used to extend the range covered by a cell under noise limited environment (i.e. no interference from other cells but the limitation coming from the fact that the transmit power Is not enough to provide a sufficient signal strength above thermal noise) is a metric of importance in such cases.

Coverage range is defined as the maximum radial distance to meet a certain percentage of area coverage (x%) with a signal to noise ratio above a certain threshold (target_snr) over y% of time, assuming no interference signals are present. It is proposed that x be 99 and y be 95.

2.2 C/I Coverage – interference limited multi-cell performance

The C/I coverage is defined as the percentage area of a cell where the average C/I experienced by a stationary user is larger than a certain threshold (target_ci).

2.3 Data Rate Coverage – interference limited multi-cell performance

The percentage area for which a user is able to transmit/receive successfully at a specified mean data rate. No delay requirement is considered here.

3 Multi-user Performance Metrics

There are several important aspects of performance that need to be considered when relays are to be used as a viable solution to enhance the multi-user performance of a system and the defined metrics should reflect those aspects of performance.

First, as mentioned before, the multi-user performance metrics need to take into account the impact of the performance of a user due to the existence of other users in the same cell requiring same resources. For example, assume that there is a user who can be provided a maximum effective data rate of 2 Mbps using a shared TDM channel. If that user is to obtain a video streaming at 2 Mbps the user will be able to receive it, but no other user will be able to get any service during whole video session (which may extend for more than an hour). Therefore, in this example video service is not a viable service for the operator and performance of coverage need to be couple with the capacity in order to reflect viable service solutions and the metrics proposed in this document reflect those aspects of performance.

Second, instead of installation of relay solutions, new base stations could be installed to obtain the same performance. In order to compare the solution based on extra relays vs a solution based on extra base stations, the number of extra relays may be taken into consideration.

Third, because maximum system capacity may be obtained by providing low throughput to some users, it is important that all mobile stations are given service in a fair manner which should be reflected when evaluating the multi-user performance. Due to the near-far effect of a wireless system (i.e. mobile which are located far away from the base and relay receives low signal quality or data rate), the efforts to provide fairness means some users need to be provided with additional resources that may also impact system throughput since shared channel is not used at the peak data rate. Therefore, fairness is an important performance aspect and is considered at the beginning in a separate section.

The users using relay stations to send and receive data uses additional resources of the system. This need to be taken into consideration when algorithms are developed to control fairness. Therefore, for relay systems special consideration may be given to incorporate the impact of these when evaluating capacity and fairness.

3.1 Fairness Criteria

The fairness is evaluated by determining the normalized cumulative distribution function (CDF) of the user throughput, which meets a predetermined function in two tests (seven test conditions). **The same scheduling algorithm shall be used for all simulation runs. That is, the scheduling algorithm is not to be optimized for runs with different traffic mixes.** The owner(s) of any proposal are also to specify the scheduling algorithm.

Let $T_{put}[k]$ be the throughput for user k. The normalized throughput with respect to the average user throughput for user k, $\tilde{T}_{put}[k]$ is given by

$$\tilde{T}_{put}[k] = \frac{T_{put}[k]}{\operatorname{avg}_{i} T_{put}[i]}.$$

Since one of the primary objectives of the introduction of relays is to address this issue by providing uniform throughput across all the users, the system should be able to provide different levels of fairness as desired by an operator. Therefore, when comparing different designs, features or configurations the throughput may be compared under the same fairness level. A measure of fairness is introduced for this purpose as defined by the following expression.

Fairness Index = 1 / standard deviation of the normalized per user throughput.

Therefore, the system performance may be compared under the following three scheduling disciplines. Depending on the service type and test case the evaluation methodology may specify what fairness requirement has to be met.

(1) Equal Throughput Scheduling:

(2) To have a reasonably compromise fairness as specified in [1].

(3) To meet a given fairness index.

3.1.1 Equal Throughput or Full Fair:

Under a full-queue simulation, this means all the users who are admitted to the system gets equal throughout if they have same amount of traffic to send.

3.1.2 Moderately fair solution as specified in [1].

We propose that the fairness requirement specified in [1] which is appended below may be used for this purpose.

Another way of achieving reasonable fairness is that the throughput is made proportional to the data rate capability of a user.
The CDF of the normalized throughputs with respect to the average user throughput for all users is determined. This CDF shall lie to the right of the curve given by the three points in Table 1.

Table 1 Criterion CDF

Normalized Throughput w.r.t average user throughput	CDF
0.1	0.1
0.2	0.2
0.5	0.5

3.1.3 Fairness to meet a specified fairness index

In this case, the fairness index should be lower than a target value. This target value may be specified under each test case.

3.2 Specific Metrics to reflect coverage and performance of multi-hop networks

3.2.1 Combined Coverage and Capacity Metric

Unlike single-user scenario, although a user may be covered for a certain percentage area (e.g. 99%) for a given service, when multiple users are in the system the resources are to be shared with other users. It can be expected that a user's average data rate may be reduced by a factor of N, compared to a single user rate when N users are active in the system. However, if the system could exploit multi-user diversity [1], the rate could be increased depending on the speed of the mobile.

Therefore, it can be deduced that increasing % area coverage itself does not give an operator the ability to offer a given service to the customers because the operator should be able to provide the service to multiple users in the same time.

Therefore, the number of users that can be supported under a given coverage captures actual coverage increase for a given service with a viability point of view. For example, an operator may offer a service to the public if the operator can support at least 10 users in average in a sector.

Therefore the viability of a service with a certain coverage captures both coverage and capacity aspects of the problem and be a good metric to evaluate the systems such as multi-hop systems which are provided to enhance coverage.

The combined coverage and capacity index is defined as the maximum number of simultaneous users (N) that can be supported for a given service, with a specified level of area coverage.

3.2.2 Multihop system spectral efficiency

This metric reflects the overall performance of a multihop system as a combined function of the number of base stations and number of relays so that the multihop solutions for coverage enhancement can be directly compared with alternative base station solutions.

Please refer to the contribution by Nortel [3].

3.2.3 Data Services and Related Output Metrics

It is recommended that the statistics provided in [1] be generated and included in the evaluation report. They are listed below for consideration.

1. **Data throughput per sector**. The data throughput of a sector is defined as the number of information bits per second that a sector can deliver and are received successfully by all data users it serves, using the scheduling algorithms validated under a specified fairness requirement.
2. **Averaged packet delay per sector**. The averaged packet delay per sector is defined as the ratio of the accumulated delay for all packets it delivers to all users and the total number of packets it delivers. The delay for an individual packet is defined as the time between when the packet enters the queue at transmitter and the time when the packet is received successively by the mobile station. If a packet is not successfully delivered by the end of a run, its ending time is the end of the run.
3. **The histogram of data throughput per user**. The throughput of a user is defined as the ratio of the number of information bits that the user successfully receives during a simulation run and the simulation time. Note that this definition is applicable to all data users.
4. **The histogram of packet call throughput for users with packet call arrival process.** The packet call throughput of a user is defined as the ratio of the total number of information bits that an user successfully receives and the accumulated delay for all packet calls for the user, where the delay for an individual packet call is defined as the time between when the first packet of the packet call enters the queue for transmission at transmitter and the time when the last packet of the packet call is successively received by the receiver. If a packet call is not successfully delivered by the end of a run, its ending time

is the end of the run, and none of the information bits of the packet call shall be counted. Note that this definition is applicable only to a user with packet call arrival process.

5. **The histogram of averaged packet delay per user**. The averaged packet delay is defined as the ratio of the accumulated delay for all packets for the user and the total number of packets for the user. The delay for a packet is defined as in 2. Note that this definition is applicable to all data users.
6. **The histogram of averaged packet call delay for users with packet call arrival process**. The averaged packet call delay is defined as the ratio of the accumulated delay for all packet calls for the user and the total number of packet calls for the user. The delay for a packet call is defined as in 4. Note that this definition is applicable only to a user with packet call arrival process.
7. **The scattering plot of data throughput per user vs. the distance from the user's location to its serving sector.** In case of SHO or sector switching, the distance between the user and the closest serving sector shall be used. The data throughput for a user is defined as in 3.
8. **The scattering plot of packet call throughputs for users with packet call arrival processes vs. the distance from the users' locations to their serving sectors.** In case of SHO or sector switching, the distance between the user and the closest serving sector shall be used. The packet call throughput for a user is defined as in 4.
9. **The scattering plot of averaged packet delay per user vs. the distance from the mobile's location to its serving sector.** In case of SHO or sector switching, the distance between the user and its closest serving sector shall be used. The averaged packet delay per user is defined as in 2.
10. **The scattering plot of averaged packet call delays for users with packet call arrival processes vs. the distance from the mobiles' locations to their serving sectors.** In case of SHO or sector switching, the distance between the user and its closest serving sector shall be used. The averaged packet call delay per user is defined as in 4.
11. **The scattering plot of data throughput per user vs. its averaged packet delay**. The data throughput and averaged packet delay per user are defined as in 3 and 2, respectively.
12. **The scattering plot of packet call throughputs for users with packet call arrival processes vs. their averaged packet call delays**. The packet call throughput and averaged packet call delay per user are defined as in 4.

Appendix A provides formulas of the above definitions.

3.2.4 VOIP Related Output Matrices [2]

It is proposed that the performance metrics for VOIP in [2] may be used for this purpose. The relevant section from [2] is provided below.

VoIP performance shall be compared on the basis of the cdf of the R values generated as a result of simulating voice traffic. R values with the corresponding impairment factors shall be obtained for the forward and reverse links.

The following metrics shall be evaluated for each VoIP user.

Mean Delay (T_a): On the forward link, the delay is measured from the point the voice packet is generated at the wired origin point to the time it is delivered at the mobile station. On the reverse link, the delay is measured from the point the voice packet is generated at the mobile station to the point it is delivered to the wired termination point.

Packet Loss Probability (Ppl, measured in percents): The packet loss probability is measured separately on the reverse and forward links.

The following set of formulas (as defined in G.107) shall be used to compute a R-factor as a function of the delay and packet loss probabilities. Proposals with better R-factors shall be considered to have better performance.

$R_{MBWA} = 93.2 – I_d – I_{e\text{-}eff}$

The quantity I_d is defined as given below

$Id = Idd$

For $Ta < 100$ ms:

$$Idd = 0$$

For $Ta > 100$ ms:

$$Idd = 25\left\{\left(1+X^6\right)^{\frac{1}{6}} - 3\left(1+\left[\frac{X}{3}\right]^6\right)^{\frac{1}{6}} + 2\right\}$$

with:

$$X = \frac{\lg\left(\frac{Ta}{100}\right)}{\lg 2}$$

Further, $I_{e\text{-}eff}$ is defined as shown below, with Ie=11 and Bpl = 19% (note that Bpl is measured in percents based on random packet loss).

$$Ie - eff = Ie + (95 - Ie)\frac{Ppl}{Ppl + Bpl}$$

The results shall include a histogram of R values for VoIP users in the system. Additionally, a histogram of packet delays may be included.

Explanation

ITU-T G.107 defines an objective model known as E-Model based on Network, Speech, Terminal/Device parameters to estimate/predict the perceived quality of VoIP session. The primary output of the E-Model is transmission rating factor R (Total Value Index) that can be mapped one-to-one to an estimated MOS.

The E-model defines 20 different parameters each with a default value and their ranges of values are defined. If all parameters are set to the default values, the calculation results in a very high quality with a rating factor of $R_{default}$ = 93.2, which is also defined as the intrinsic quality of a voice call with a mouth-to-ear delay of 0 ms. The intrinsic quality of a packetized voice call transported without packet loss in the G.711 format corresponds to this $R_{default}$ = 93.2.

However, for MBWA system specific impairments such as Packet Loss, Delay etc considered, the effective R factor for such system needs to be estimated by incorporating equipment impairment factor, delay factor. The effective R factor is

$R_{MBWA} = 93.2 – I_d – I_{e\text{-}eff}$

Here I_d, the impairment factor representing all impairments due to delay of voice signals Talker Echo, Listener Echo and I_{dd}, a loss of interactivity, represents the impairment caused by too-long absolute delay Ta, which occurs even with perfect echo canceling. Here we assume perfect Jitter operation resulting no packet loss and additional delay introduced by jitter.

The packet-loss dependent Effective Equipment Impairment Factor Ie-eff is derived using the codec specific value for the Equipment Impairment Factor at zero packet-loss Ie and the codec specific Packet-loss Robustness Factor *Bpl*

I_e represents the effect of degradation introduced by CODECs, Packet Loss. G.113 –Appendix (2002) provided provides parameters for use in calculating Ie from CODEC type and Packet Loss rate. For a G.729 Codec Ie = 11, and for random packet loss Bpl = 19.

4 Service Delay and outage requirement

Each service to be used in the simulation is associated with a service quality requirement, which is usually a delay requirement, jitter requirement and outage requirement. These need to be taken into consideration in the simulations and they are defined below.

4.1.1 Outage Criteria for HTTP, FTP and Near Real Time Video [1]

It is proposed that the outage criterion used in [1] be used for HTTP, FTP and real time video services. This is included below.

Except for VOIP and Live Video services, all the packet data (HTTP, FTP, or near real time video) users shall satisfy the following delay criterion: no more than 2% of the users shall get less than 9600 bps throughput (goodput). The throughput will be the user's packet call throughput, except in the case where there is no arrival process (FTP users are persistent) in which case it will be the throughput averaged over the simulation time. Neal real time video users shall also satisfy the performance criteria defined below.

4.1.2 Additional Performance Criteria for Near Real Time Video

It is proposed that the following description in [1] be included as a near real time video performance requirement.

"Video playout buffers introduce a delay between receipt of frames and the frame playout. This absorbs variations in the data arrival pattern and permits a continuous playout of the frames. The actual design of these playout buffers involves a number of factors (including reset policies when the buffer runs dry) and is specific to the mobile. To avoid modeling such implementation details, we focus on what the BS scheduler must do to generally accommodate this continuous playout. Therefore, the scheduler should transmit an entire video frame within 5 second of receipt of the entire frame (i.e., receipt of the last octet of the last slice of the frame). If a frame exceeds the 5-second requirement, the scheduler discards the remainder of the frame that has not yet been transmitted.

.

Therefore, the performance requirement is that the fraction of video frames that are not completely transmitted within 5 seconds of their arrival at the scheduler shall be less than 2% for each user. All users shall meet the above performance requirement."

References

[1] 3GPP2/TSG-C.R1002, "1xEV-DV Evaluation Methodology (V14)", June 2003.

[2] IEEE P 802.20™ PD-09 Version 1.0, September 23, 2005.

[3] IEEEC802.16j-06/024, "Multihop System Evaluation Methodology: Traffic Models**"** May 2006

Appendices

Appendix A. Formula to define various throughput and Delay Definitions

For each fixed simulation condition (e.g., voice load, distribution of different traffic, number of data users, etc.), simulation is run for multiple independent runs using Monte Carlo approach. Let

- T = simulation time.
- M = total number of independent runs (for a specific configuration).
- N= total number of data users in each run (for a sector).
- m= index of the simulation runs, i.e., m = 1,2,3, …, M.
- n = index of a data user within a simulation run, i.e., n = 1,2,…, N.

Therefore, the n-th data user in the m-th simulation run can be specified by *user(m,n)*.
Let

- K(m,n) = total number of packet calls generated for user(m,n).
- k = index of packet calls for a user. For user(m,n),, k = 1, 2, …, K(m,n).
- L(m,n,k) = total number of packets generated for the k-th packet call of user(m,n).
- l = index of packet within a packet call. For the k-th packet call of user(m,n), l = 1,2, …, L(m,n,k).
- B(m,n,k,l) = number of information bits contained in the l-th packet of the k-th packet calls for user(m,n) . If the packet is not successfully delivered by the end of the simulation run, B(m,n,k,l) = 0.
- TA(m,n,k,l) = arrival time of the l-th packet of the k-th packet calls for user(m,n). it is the time when the packet arrives at the transmitter side and is put into a queue.
- TD(m,n,k,l) = delivered time of the l-th packet of the k-th packet calls for user(m,n) . It is the time when the receiver successfully receives the packet. Due to fixed simulation time, there may be packets waiting to be completed at the end of a simulation run. For these packets, the delivered time is the end of the simulation.
- PCTA(m,n,k) = arrival time of the k-th packet call for user(m,n), it is the time when the first packet of the packet call arrives at the transmitter side and is put into a queue.
- PCTD(m,n,k) = delivered time of of the k-th packet call for user(m,n). It is the time when the receiver successfully receives the last packet of the packet call. Due to fixed simulation time, there may be packet calls waiting to be completed at the end of a simulation run. For these packet calls, the delivered time is the end of the simulation.

The arrival time of a packet call is the time when the first packet of the packet call arrives at the transmitter side and is put into a queue, and the delivered time of a packet call is the time when the last packet of the packet call is successfully received by the receiver, i.e., PCTA(m,n,k) = TA(m,n,k,1) and PCTD(m,n,k) = TA(m,n,k,L(m,n,k)). Due to fixed simulation time, there may be packet calls waiting to be completed at the end of a simulation run. For these packet calls, the delivered time is the end of the simulation. Figure D-1 demonstrates the arrival and delivered times for a packet and a packet call.
With the above notation, we can now define various throughputs and delays as follows.

Data throughput per sector $= \dfrac{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K(m,n)}\sum_{l=1}^{L(m,n,k)} B(m,n,k,l)}{MT}$, (D1)

Averaged delay per sector $= \dfrac{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K(m,n)}\left(PCTD(m,n,k)-PCTA(m,n,k)\right)}{\sum_{m=1}^{M}\sum_{n=1}^{N} K(m,n)}$, (D2)

Data throughput for user(m,n) $= \dfrac{\sum_{k=1}^{K(m,n)}\sum_{l=1}^{L(m,n,k)} B(m,n,k,l)}{T}$, (D3)

Packet call throughput for user(m,n) $= \dfrac{\sum_{k=1}^{K(m,n)}\sum_{l=1}^{L(m,n,k)} B(m,n,k,l)}{\sum_{k=1}^{K(m,n)}\left(PCTD(m,n,k)-PCTA(m,n,k)\right)}$, (D4)

Averaged packet delay per sector $= \dfrac{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K(m,n)}\sum_{l=1}^{L(m,n,k)}\left(\mathrm{TD}(m,n,k,l) - \mathrm{TA}(m,n,k,l)\right)}{\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K(m,n)} L(m,n,k)}$, (D5)

Averaged packet delay for user(m,n) $= \dfrac{\sum_{k=1}^{K(m,n)}\sum_{l=1}^{L(m,n,k)}\left(TD(m,n,k,l)-TA(m,n,k,l)\right)}{\sum_{k=1}^{K(m,n)} L(m,n,k)}$, (D6)

Averaged packet call delay for user(m,n) $= \dfrac{\sum_{k=1}^{K(m,n)}\left(PCTD(m,n,k)-PCTA(m,n,k)\right)}{K(m,n)}$. (D7)

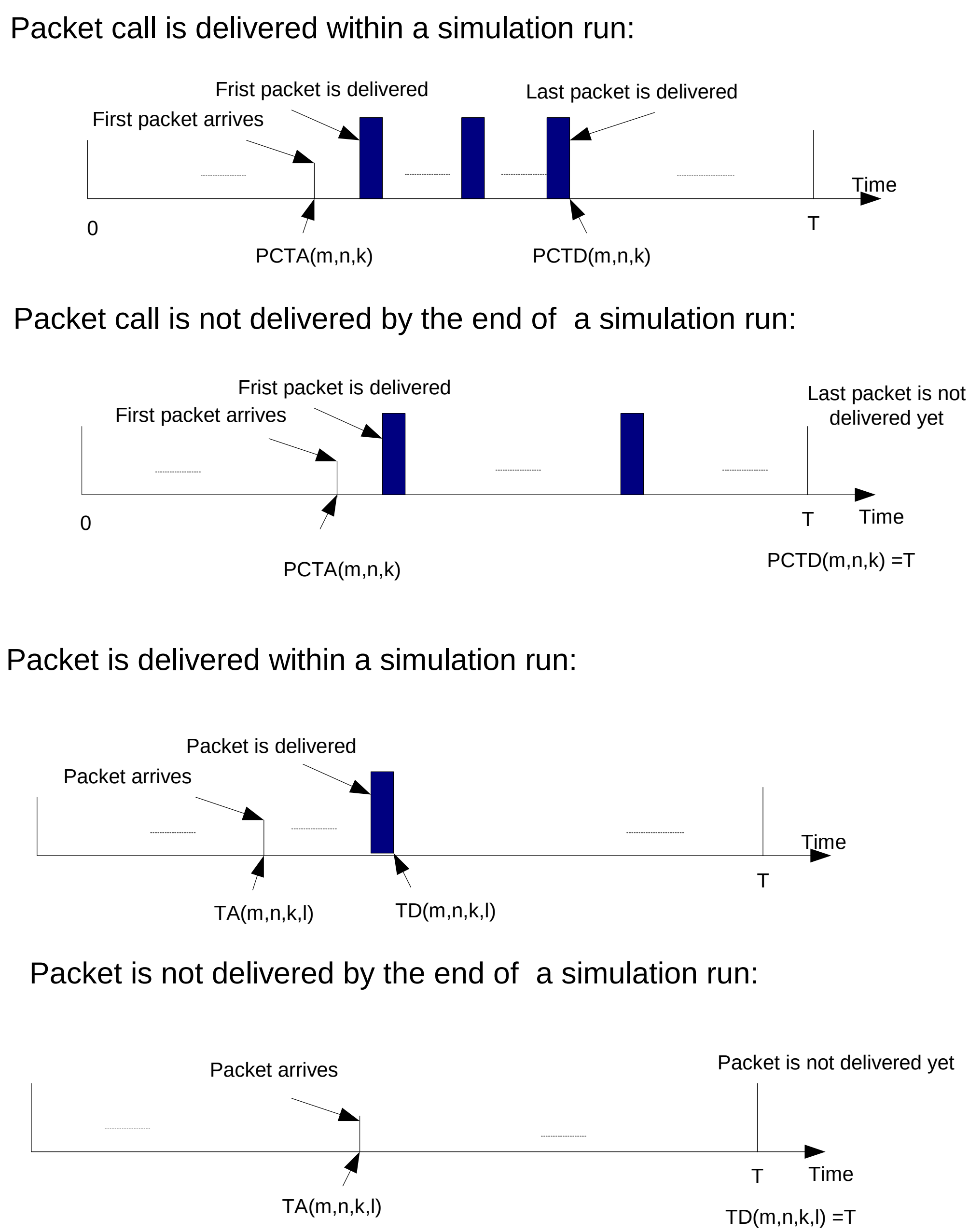

Figure A-1: Description of arrival and delivered time for a packet and a packet call.