BEAM HOPPING- UTILIZATION PERFORMANCE ANALYSIS

Avraham Freedman
SatixFy Ltd., 12 Hamada St. Rehovot, Israel 74140, Tel: +972-89393203,
Fax: +972-89393223, avi.freedman@satixfy.com

Abstract

Beam-hopping was shown to provide a level of flexibility that makes it possible to increase served traffic, reduce areas of unmet demand while enabling the reduction of power consumption on-board. In [1] various issues regarding beam-hopping, from the terminal, payload and eco-system point of view, were addressed. In [2], we discussed the system considerations for implementation of beam-hopping in a multi-beam environment and the trade-offs required for different applications. In this paper we analyze the performance of beam-hopping system in terms of beam utilization, comparing various approaches to beam-hopping scheduling, for both transparent and regenerative payloads.

1. Introduction

Beam-hopping was shown to provide a level of flexibility that makes it possible to adapt the payload increase served traffic, reduce areas of unmet demand while enabling the reduction of payload power consumption. In [1] various issues regarding beam-hopping, from the terminal, payload and eco-system point of view, were addressed. In [2], we discussed the system considerations for implementation of beam-hopping in a multi-beam environment and the trade-offs required for different applications. Various references are listed in [2] analyzing various scenarios.

The effectiveness of a beam hopping system highly depends on the efficient use of the resources, and this efficient use is a function of the transmission strategy of the beam-hopping scheduling, and parameters such as dwell time, cycle time, number of cells covered by a beam, beam hopping transmission time plan and more.

However, the effectiveness should be measured in a real environment, where the data for transmission arrive randomly. In this paper we analyze the performance using a stochastic traffic model and compare how the transmission strategy affect the latency and utilization, in that realistic scenario.

In [2] several transmission strategies were presented these include:

- Strategy based on a pre-defined transmission plan, according to prior information on the expected load in each cell. This can be called the classical beam hopping strategy.
- A strategy based on fixed transmission time intervals. In this strategy the packets for each cell are queued. At constant instants, beams are transmitted to the cells with the longest queues, or, to cells for which the revisit time constraint has expired.

As a reference a pure "point and shoot" system, where no buffering per cell takes place. In this, rather ideal, strategy, each transmission is dedicated to a packet.

The classical, pre-defined time plan strategy provides the flexibility of allocation transmission time resources according to the demand per cell, however it still depends on a-priori knowledge of the demand distribution and needs to be modified as the demand changes. Additionality, it has some drawbacks:

- Queueing takes place for every cell, thus causing large queuing delays in case of loaded cells.
- Dwell times are fixed for a given beam-hopping time plan and might be under-utilized in case of light load.

The pure point-and-shoot strategy aims to overcome the queuing delay by transmitting every packet as it arrives to the system. This means, in essence, a packet by packet beam hopping to the destination cells. The resulting queuing delay is reduced considerably as now there is only a single queue serving all the users, and not separate queues for each cell, thus enjoying the advantages of statistical multiplexing.

On the other hand, implementation of this strategy involves providing routing information per transmission packet, and, as transmissions has granularity resulting from the need for minimal dwell time, allowing for switching time at the transmitter, as well as burst acquisition time at the receiver, the efficiency of the transmitted packet, with respect to the useful data reduces as the granularity increases.

The third strategy aims at increasing utilization, by accumulating the arriving per cell, but still allowing all the transmission channels to serve all the cells, resulting in a queuing system which still enjoys a level of statistical multiplexing. The allocation is still dynamic and follows the variations in the demand automatically.

It should also be noted that the latter two strategies result in random illumination pattern. In case of several transmission channels interference may occur if adjacent cells are illuminated at the same time instant. We ignore this limitation in the sequel, although it has to be taken care in implementation.

2. System Model and Analysis

Consider a satellite beam hopping systems, serving N_u users (terminals) located in N_c cells, with N_{tx} transmission channels.

The terminology, following that of [2], is:

- *Beam*: the directional radio signal transmitted from a satellite
- *Cell*: an area on the ground illuminated by a beam
- *Transmission channel or just Transmitter*: The power amplifier and additional components handling the transmission, and sh
- *Cluster*: A set of beams served by one transmitter
- *Dwell time*: the time duration in which a given transmission channel is allocated to a given beam.
- *Off time*: the time duration in which a given cell is not illuminated
- *Transmission Packet*: The transmission that occurs during the dwell time
- *Beam Hopping Transmission plan*: the absolute transmission times and dwell times allocated for each beam
- *Revisit time*: The maximal time -period in which a terminal is revisited
- *Cycle*: The period of time during which a transmitter covers all the beams within its allocated cluster.

It should be noted that although the terminology and the analysis below refer to the forward direction, the analysis may very well be applied to the return direction, where the satellite reception beams are hopping, and the reception channel resources serve a different cell each hop. However, in the sequel we consider a single gateway transmitting the data to the users via the beam-hopping satellite.

In this section we analyze the performance of the system in various hopping strategies. For the sake of simplicity, we assume all users have the same traffic data rates, R_d (bps), the same SNR condition, and the same bandwidth hence the data is transmitted at the same symbol rate, R_s (baud). We assume the data arrive in packets, of which the sizes (in bits) and arrival times are stochastic. For this analysis we take the arrival process to be Poisson, with packet arrival rate $\lambda_p = N_u \lambda_u$, where λ_u is the packet arrival rate for a single user. The probability that the inter-arrival time between packets is given by:

$$P_r\left(T_{arr} < t\right) = \lambda_p e^{-\lambda_p t} \tag{1}$$

The users are non- uniformly distributed among the cells, such that:

$$\sum_{c=1}^{N_c} n_u\left(c\right) = N_u \tag{2}$$

Where $n_u(c)$ is the number of users in cell c.

The satellite uses N_{tx} transmission channels to serve those users according to the specified illumination strategy. Each channel can support a given transmission rate, with spectral efficiency, η, of which the expected rate is equal to a serving rate of μ_p packets per second, such that:

$$\frac{\lambda_p}{\mu_p} = \frac{R_d}{R_s \eta} = \rho \tag{3}$$

In the following analysis we compare different strategies in terms of the total inherent latency and beam utilization, as a function of two variables: the system load and the transmission granularity.

As a “black box” the system can be seen as a queuing system with packets arriving at a rate of λ_p packets per second and leaving at a rate of $N_{tx}\mu_p$ packets per second, with general queuing regime M/G/N_{tx} according to Kendall’s notation [3]. Unfortunately, most performance metrics for this model are still an open problem. We focus on some special cases, according to the strategies specified above.

2.1. Pure Point-and-Shoot

We will analyze this strategy for continuous and discrete service times. It is presented first, as it is simpler and more general than the other.

2.1.1. Continuous Service time

The service time is simply the transmission time of the packet, which we take as distributed exponentially with mean $1/\mu_p$. The resulting system can be analyzed as a classical Erlang C, [4], or a M/M/c queue, according to [4]. According to this queuing model, the probability that a packet would be queued is given by:

$$C\left(N_{tx},\rho\right)=\frac{\dfrac{\rho^{N_{tx}}}{N_{tx}!\left(1-\rho/N_{tx}\right)}}{\sum_{k=0}^{N_{tx}-1}\dfrac{\rho^k}{k!}+\dfrac{\rho^{N_{tx}}}{N_{tx}!\left(1-\rho/N_{tx}\right)}} \tag{4}$$

Equation (4) applies, as long as $\rho<N_{tx}$.

The waiting time in the system, W, namely the latency experienced by a packet until it is transmitted is exponentially distributed:

$$P\{W>t\}=C\left(N_{tx},\rho\right)\exp\left[-\left(1-\frac{\rho}{N_{tx}}\right)N_{tx}\mu_p t\right] \tag{5}$$

Thus, the average total latency is:

$$T_L=E\{W\}+\frac{1}{\mu_p}=\frac{C\left(N_{tx},\rho\right)}{\left(N_{tx}-\rho\right)\mu_p}+\frac{1}{\mu_p} \tag{6}$$

The mean transmitter utilization would be given by:

$$E\left\{U\left(N_{tx},\rho\right)\right\}=\sum_{j=0}^{N_{tx}-1}\frac{j}{N_{tx}}P_j+C\left(N_{tx},\rho\right)=\frac{\rho}{N_{tx}} \tag{7}$$

Where P_j is the probability of the system to have *j* transmitters active, and is given by:

$$P_j=\frac{\dfrac{\rho^j}{j!}}{\sum_{k=0}^{N_{tx}-1}\dfrac{\rho^k}{k!}+\dfrac{\rho^{N_{tx}}}{N_{tx}!\left(1-\rho/N_{tx}\right)}}\quad ,j=0,..,N_{tx}-1 \tag{8}$$

The utilization, in average, equals to the total system load, and reflects how busy the transmitters are.

2.1.2. Granular Service Time

When introducing minimal dwell time limitations, or granularity (namely, transmissions of only an integer number of frames/ super-frames), the service time distribution will no longer be exponential, but rather discrete (geometric). For a positive integer n:

$$\Pr\left\{D_g = x\right\} = \begin{cases} \dfrac{e^{-\mu_p x}}{1-p} & x = nD_{SF} \\ 0 & Otherwise \end{cases} \tag{9}$$

Where $D_g = \left\lceil D_p / D_{SF} \right\rceil$ is the transmitted packet size, D_p is the stochastic variable indicating the data packet size (in symbols), D_{SF} is the size of a frame or superframe, which is the minimal unit that can be transmitted. We denote $p = e^{-\mu_p D_{SF}}$. In this case an exact expression for the waiting time does not exist, but one can use a good approximation:

$$E\left\{W^{M/G/c}\right\} = \frac{1+C_v^2}{2} E\left\{W^{M/M/c}\right\} \tag{10}$$

Where the expected value for the waiting time for a general M/G/c queue can be approximated by a correction factor of the expected waiting time for the M/M/c queue, with C_v being the ratio of the service time standard deviation to its mean. In the case of the distribution in (9):

$$E\left\{D_g\right\} = \frac{D_{SF}}{1-p}, \ \mathrm{var}\left\{D_g\right\} = \frac{p}{\left(1-p\right)^2} D_{SF}^2 \tag{11}$$

$$C_v^2 = \frac{\mathrm{var}\left\{D_g\right\}}{E^2\left\{D_g\right\}} = p = e^{-\mu_p D_{SF}} \tag{12}$$

while for the M/M/c queue the value of C_v is unity, here it is less than 1, resulting in a somewhat smaller waiting time. However, the expected service time is at least D_{SF} and converges to that value with growing service rate or growing frame (superframe) size.

As for utilization - we have to add the fact that the transmission time is granular, namely there is always some “unused space” in each transmission.

The probability that a packet length would require $n+1$ frames is:

$$\Pr\left\{nD_{SF} < d_p < \left(n+1\right) D_{SF}\right\} = p^n \left(1-p\right) \tag{13}$$

Where *p* is as defined above. The utilization of that frame would then be:

$$U \left| \left(nD_{SF} < d_p < \left(n+1\right) D_{SF}\right) \right. = \frac{d_p - nD_{SF}}{\left(n+1\right) D_{SF}} \tag{14}$$

The expected value of the utilization of that packet:

$$E\left\{U \left| nD_{SF} < d_p < (n+1) D_{SF} \right.\right\} = \left[\frac{1}{\mu_p D_{SF}} - \frac{p}{1-p}\right] \frac{1}{n+1} \tag{15}$$

The complete probability is then:

$$E\left\{U\right\} = \left[\frac{1}{\mu_p D_{SF}} - \frac{p}{1-p}\right] \frac{1-p}{p} \ln \frac{1}{1-p} \tag{16}$$

2.2. Pre-defined Transmission Plan

In this case, the queue is not common to all users. Transmission for users in each cell are queued and served for a pre-defined time. Obviously, as known from queuing theory, this is not as efficient as a common queue case, as it doesn't make use of the statistical multiplexing advantage. With our simplifying assumptions, the queue for a cell is of packets arriving at a rate of $\lambda_c = n_u(c)\lambda_u$. The service rate can be described as follows:

- When the beam illuminates the cell- the service rate is, as above, equals μ_p
- When the beam illuminates another cell, the service rate is 0.

During the illumination time, we can model the queuing system as a M/M/1 queue. Otherwise there is no service.

For an M/M/1 queue, we can use equations (4) - (8) with $N_{tx} = 1$, arrival time rate can be determined as

$$\lambda_c = n_c \lambda_u = \frac{n_c}{N_u} \lambda_p \tag{17}$$

And the mean service rate given by:

$$\mu_c = N_{clus} \mu_p \frac{T_c}{T_{cyc}} \quad , \tag{18}$$

where T_c is the time allocated to cell *c*, T_{cyc} is the cycle time, and we assume all the transmission channels are loaded uniformly. The mean load is then determined as:

$$\rho_c = \frac{\lambda_c}{\mu_c} = \frac{n_c T_{cyc}}{N_{tx} N_u T_c} \rho \tag{19}$$

As for the average quantities, such as mean service rate, mean queue length, mean waiting time etc. the queue can be modelled as a M/M/1 queue with the load described above, the actual probability distributions are not the same.

As the arrival process is Poisson, and is independent of the time axis, we can assume that any arrival can occur in any time interval within the cycle with the same probability. Thus, we can calculate the probability that a packet will be queued as:

$$P_q = \frac{T_c}{T_{cyc}} \rho_c + \left(1 - \frac{T_c}{T_{cyc}}\right) = 1 - \frac{T_c}{T_{cyc}}(1 - \rho_c) \tag{20}$$

The waiting time (sojourn time that includes the actual transmission time) is given by:

$$T_{Lc} = \frac{C(1, \rho_c)}{(1-\rho_c)\mu_c} + \frac{1}{\mu_c} = \frac{\rho_c T_{cyc}}{(1-\rho_c) N_{tx} \mu_p T_c} + \frac{T_{cyc}}{N_{tx} \mu_p T_c} \tag{21}$$

Assuming the beam hopping time plan is well matched, namely, for every cell:

$$\frac{n_c}{N_u} = \frac{T_c}{T_{cyc}} \tag{22}$$

Then, from (19), $\rho_c = \rho / N_{tx}$ for every cell, (21) is the average delay for the whole system.

As for utilization, each transmission channel hops from cell to cell, as the average utilization of the server of each cell is ρ_c. In this case, utilization is not a question of granularity, but rather a question to what extent the dwell time is full. In our simplified model, this value, averaged over all the cells is ρ.

In the case of a pre-defined time plan, the granularity affects the matching of the hopping plan to the actual user density distribution. Large superframes granularity may make it impossible to match the hopping pattern to that distribution. Furthermore, if there are several transmission channels covering the area in parallel, the cycles may not match, which would make it difficult for planning the hopping pattern without illuminating adjacent cells in the same time, thus causing interference.

2.3. Fixed Transmission time interval

With this strategy, packets are queued per cell as for the pre-defined transmission plan, however, transmissions take place each T_f sec, from the most loaded queues. The goal in this strategy is to enjoy the advantages of the fact that each server can serve any queue, while accumulating enough packets in a transmission as to increase utilization. The exact derivation of the statistics involved is beyond the scope of this paper, however, we can point out a few observations:

- In average – each cell gets served at a rate proportional to its packet arrival rate, automatically following any rate change that might occur.
- As the arrival rate at each cell is stochastic the waiting time is not controlled, thus a re-visit time constraint is essential to ensure that no cell is starved.

3. System Simulation

3.1. Test Scenario

The comparison between the different strategies are made against the following test scenario:
200 users distributed among 16 cells, in pseudo exponential manner (see Figure 1). The average SNR per user is 10dB. In the simulations the SNR per user is selected from a log-normal distribution with standard deviation of 3dB.
The users are served by 4 beams, each of a 100MHz bandwidth. Shannon’s law would yield, total system capacity (for the average SNR, taking a 2dB penalty) of 2.3 bits/Hz resulting in 1.4 Gbps in this particular scenario, namely 5.7Mbps per user. In the simulations data rates were tested for the range between 10% and 100% of this value.
The data rate per user as well as the arrival rate per user were selected randomly around the average data rate, and an average arrival rate of 50 packets/sec, correlated (80%) to the data rate.
The packet size (in symbol) per user was then determined by the specific data rate and spectral efficiency of this particular user. All cases were also checked for different granularities, for frame sizes of 100, 30000 and 500000 symbols.
While in the pre-determined time plan, all cells are visited during a cycle, the other schemes are stochastic in nature and do not assure that. As the re-visit time might be necessary for terminals for keep-alive and maintenance, in those schemes a revisit time constraint of 30msec was applied. Namely, if a cell is not visited within this time limit a special “dummy” frame is transmitted for it.

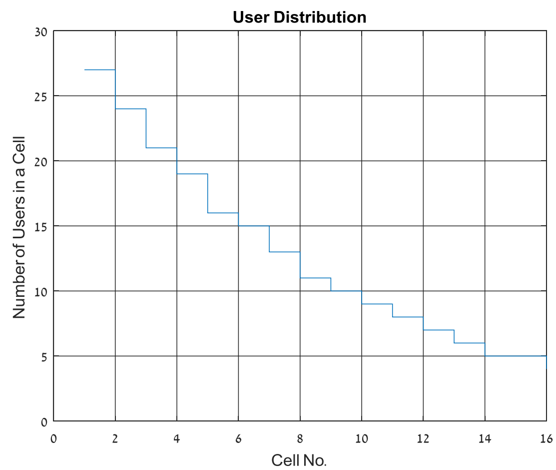


Figure 1: Users Distribution

3.2. Simulation Results

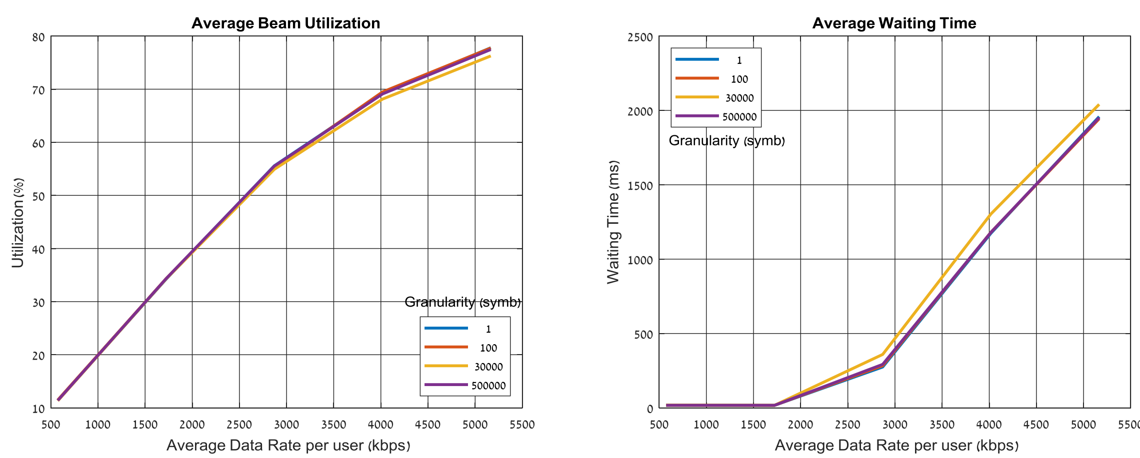
3.2.1. Pre-Determined Time Plan

In the simulation described here, a cycle of about 20ms was used. The cells were allocated to the 4 transmission channel in an almost uniform way, such that the number of users per transmission channel are equal (as much as possible). The dwell time per cell within each beam was set according to the number of users in each cell, within the granularity provided by the superframe size.

The pre-determined time plan scheme performance (waiting time and utilization, as a function of total average data rate and the "superframe" size) is given in Figure 2 (Fig. 2(a)) and average waiting time (Fig. 2(b)). For low data rate, the waiting time demonstrates a plateau of around 20msec, resulting from the fact that a large number of packets has arrived at no-transmission time, which, together with the queuing and service time, reached an almost constant level. As the system load gets higher with increased data rate, the waiting time grows almost linearly.

Utilization also grows monotonously with the load, as expected by theoretical analysis.

Superframe size had minor effect on the waiting time and utilization. Its main effect is over unmet demand in loaded cells vs. exceeding capacity in lightly loaded cells, as described in [2].



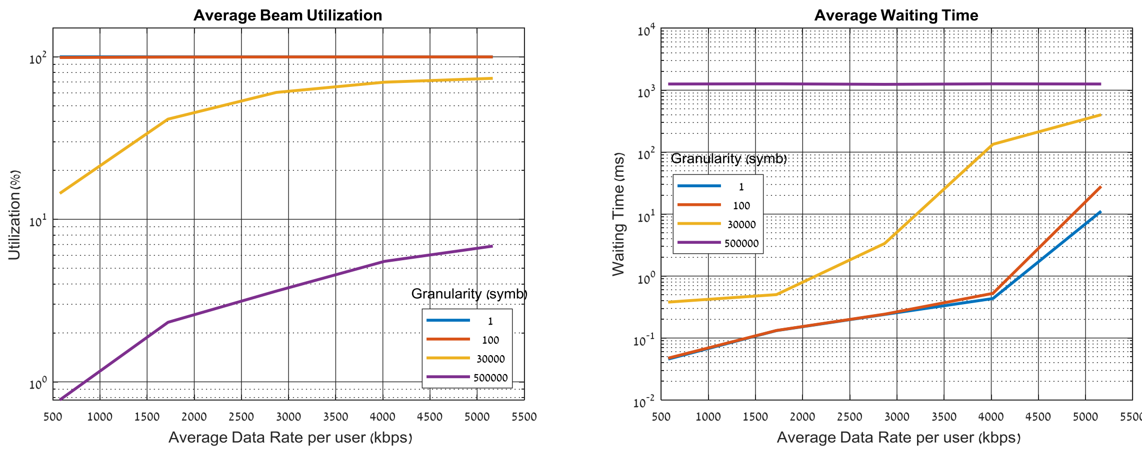
(a) **(b)**

Figure 2: Predefined Time Plan

(a) Average Beam Utilization, (b) Average Waiting Time

3.2.2. Point and Shoot

Figure 3 shows the average beam utilization for a point and shoot system. Note that we used logarithmic scale to show the high range of values better. It shows that with minimal dependency on the data rate (and hence on system load), the waiting time is minimal and the utilization is nearly 100%, as long as the granularity is small (100 symbols), For larger superframes, those two parameters are directly affected. For our simulation the waiting time grows from less than 0.1 msec at small granularity up to 100msec for a 500k symbols superframe. The utilization drops from 100% at small granularity to 5% for the large superframe.



(a) **(b)**

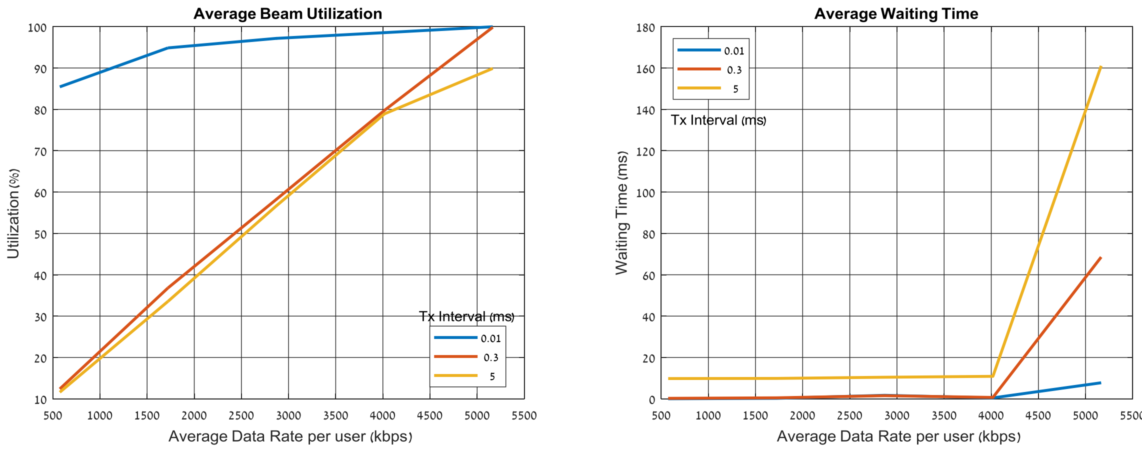
Figure 3: Pure Point and Shoot Strategy

(a) Average Beam Utilization, (b) Average Waiting Time

3.2.3. Fixed Transmission Times

This strategy, designed to enable flexible allocation of transmission resources, thus enabling good matching of the transmissions to the user density distribution. The superframe size, in this case plays the role of the total transmission packet size. Figure 4**: Fixed Transmission Times** shows the average beam utilization and the average waiting time acheived in the simulation for this strategy. Good beam utilization is achieved for small superframe size, but, as it grows, the utilization becomes linear as a function of data rate, as the transmission packet is not filled up during the relatively long time of transmission.

An interesting effect can be observed regarding the waiting time. It is kept almost constant as the function of the load, until some break point, where the waiting time becomes large.



(a) **(b)**

Figure 4: Fixed Transmission Times

(b) Average Beam Utilization, (b) Average Waiting Time

3.3. Analysis and Comparison

Table 1 below compares the waiting time and utilization for the three strategies at different operation conditions: low and high data rate, small, medium and large transmission frame size.

Table 1: Comparison of Illumination Strategies

		Waiting time (ms)			Utilization (%)		
Date rate per user (Mbps)	**Tx Frame size (symbols)**	**Pre-defined**	**Point and Shoot**	**Fixed Tx Times**	**Pre-defined**	**Point and Shoot**	**Fixed Tx Times**
1	100	20	0.07	2	20	100	90
1	30k	20	0.5	2	20	20	20
1	500k	20	1000	10	20	1.5	20
5	100	1700	8	10	77	100	100
5	30k	1700	500	60	77	80	95
5	500k	1900	1000	140	75	8	89

As can be observed from the table, and the graphs at the previous sections, the classical pre-defined beam hopping time plan both waiting time and beam utilization behave almost linearly with the load. With the point and shoot strategy waiting time can be decreased significantly and the utilization improves, as long as the granularity constraint is low enough. The fixed time strategy provides a good and flexible compromise between the two- low waiting time and linear utilization with the load.

4. Summary and Conclusion

Beam Hopping was shown to provide high flexibility in adapting the satellite resource to traffic demand and its distribution in time and space. In reality, the demand is not constant but rather stochastic. This stochastic nature results in queuing delay and the effectivity of beam- hopping in terms of beam utilization may be compromised.

In this paper we investigate the effectivity of beam hopping in view of the stochastic nature of the demand, in terms of its utilization and the additional waiting time it imposes. The conventional way beam hopping is treated is by making, for a given scenario of demand distribution, a pre-defined time plan by which every transmission channel hops from cell to cell, in a given cycle time. This time plan can be modified but only over a relatively long time scale, by instructions from the satellite operator. In addition to the fact that this pre-defined plan may deviate from the actual demand distribution, it also introduces some inefficiency in view of queuing delays, and reduced utilization. In this paper we analyzed the performance of this scheme in comparison to more dynamic strategies.

In contrast, point-and-shoot strategy, which transmits each packet as it arrives, avoids the queuing delay, enables statistical multiplexing advantage and adapts to the demand automatically. However, considering the fact that transmission is granular in time, it may introduce delays and under-utilization when the transmission frame size is too small.

A third scheme suggested here, is to use fixed transmission times as a basis for transmission, and transmit at fixed times, to the cells for which maximal traffic has accumulated. This scheme proved to be quite effective in reducing the waiting time, while the utilization increases linearly with the load, as observed in the pre-defined beam case.

Beyond this paper, other strategies, which present other trade-off can be investigated. Furthermore, analysis of scheme with random beam illumination, like the two presented here, should also be analyzed for different deployment scenarios and in conjunction with other techniques (e.g. pre-coding) to increase total system capacity.

References

[1] A. Freedman, D. Rainish, Y. Gat: Beam Hopping – How to Make It Possible, *22nd Ka and Broadband Communications Conference,* (Ka-2016), Cleveland, OH, Oct. 2016

[2] A. Freedman, D. Rainish, System Design Consideration, *23rd Ka and Broadband Communications Conference,* (Ka-2017), Triest, Italy, Oct. 2017

[3] Kendall, D. G.: "Stochastic Processes Occurring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain". The Annals of Mathematical Statistics. 24 (3): 338. 1953

[4] R. B. Cooper: "Introduction to Queueing Theory", 2nd Edition, North Holland, 1981

: