International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Optimal Power Allocation Policy based on Planned and Deferred Data Transmission in Energy Harvesting Wireless Devices

Anjana Murali[1], Belma Anna Kurian[2]

[1]*P.G Scholar Dept. of Electronics and Communication Engineering, Believers Church Caarmel Engineering College Perunad, Pathanamthitta, Kerala, India,*
[2]*Asst. Prof. Belma Anna Kurian, Dept. of Electronics and Communication Engineering, Believers Church Caarmel Engineering College, Perunad, Pathanamthitta, Kerala, India*

Abstract - *In this paper, system developing an optimal power allocation policy that maximizes a utility function defined over average transmission rates. Data transmission referred to computer mediated communication among system users. Energy harvesting nodes are the link of this system. Energy harvesting is the process by which energy is derived from external sources (like solar power, thermal power, wind energy, salinity gradient, kinetic energy etc.) captured and stored for wireless communication purpose. Energy harvesting capability of energy harvesting nodes are significant here. A machine learning algorithm based on Reinforcement learning and power allocation policy is the thread of this paper. Suppose a system used for planning is later used to generate and but might compose a message for deferred transmission to an inactive user. Throughput maximization and convergence time reduction are the main objectives of this multi-hop communication system.*

Key Words: Energy harvesting, Multi-hop communication, Two-hop communications, Reinforcement learning, Power allocation

1. INTRODUCTION

Energy Harvesting (EH) wireless nodes have the capability of Energy Harvesting. The communication system is multi-hop and the basic building block of this system is two-hop. Power Optimization can obtain maximum throughput. Here, EH nodes have the availability of local causal knowledge. In this scenario, a machine learning algorithm known as Reinforcement Learning (RL) algorithm along with Markov Decision Process (MDP) is applied. EH nodes are harvesting energy from natural sources and using this energy for data transmission in the multi-hop communication system. The system is time variant and channel fading coefficient is also considered in the entire system while transmission. Harvested energy level, battery buffer level and data buffer level are the energy states used for study in this communication system. The data rate of transmission occurrence is based on this study.

All the existing system design is based on non-causal knowledge. Past and present conditions are verified by the system and hence future predictions are taking place. All the assumptions may not be true, therefore probability of the occurrence of errors are happening in transmission time. Energy Harvesting time and channel coefficient between transmitter and relay are the main part of error formation in current communication system assumption. Realistic scenario-based functioning is not possible if we follow the same. Maximum rate of data transmission with less power consumption practically impossible in the existing research areas. Wireless networking system consisting of EH nodes begin from performance limits of information-theoretical function to transmission scheduling policies, resource allocation, networking issues and medium access. Energy cooperation and energy-information transfer model of sustainable EH wireless network is present in some research area [1]. The applications in the area of industry and academia is long term, uninterrupted and self-sustainable. Energy storage capability limitation, scarcity of energy and device complexity are the main challenges which leads to the design of intelligent energy management policies for EH wireless devices [5]. Future wireless system extension basic principle is Energy Harvesting radios. Efficient transmission strategies of rechargeable nodes and optimization problem are identified and maintained. Energy causality constraints are used to adopt optimum transmission policy [2]. Two-hop relying problem in EH nodes is considered first and an optimal offline transmission scheme operating with full-duplex mode is framed [4]. The solution for two-hop channel indirectly assumed by an infinite size buffer and proving that there is no need of data buffer for relay. Average throughput reduced by constant relay rate as peak energy harvest rate of source become high [3].

Transmitter, relay and receiver in our system design must consider harvested energy, battery level of energy, data buffer level, channel coefficient, and transmitting data. Reinforcement learning (RL) algorithm can be developed to find the Power allocation policy. Reduction of transmitter and relay power consumption at the time of transmission is the optimization in our machine learning algorithm. The data arrival process with feature functions are also considered. Our system model works based on causal knowledge, which means current and past knowledge of channel coefficient and harvested energy is known to the system. Lack of availability of future knowledge about the harvested energy and channel coefficient is a great challenge. EH multi-hop communication is another difficulty to make the system reality. Several intermediate EH nodes are functioning in the

data transmission path of transmitter and receiver. If the channel condition is not good, then a node can defer the utilization of energy harvested and save the power. Energy harvesting and Power allocation go hand in hand during transmission. Point to point communication in multi-hop can form the problems and solutions of data transmission. The two-hop communication can be considered as the basic building block of the multi-hop system design. EH process and channel fading process are studying alternately. As a part of power consumption, non-causal knowledge and assumptions are required. To avoid such requirements, a Reinforcement Learning (RL) to develop Power allocation policy is adapting. Realistic conditions can form a practical system with optimum power. Transmitter has to harvest energy multiple times, whereas relay needs only single harvest at a transmission. Randomly situated nodes are participating in data transmission function. The EH nodes find out strategy for transmission by continuous evaluation and make decisions to obtain maximum throughput.

The non-causal knowledge centered ideal system for EH communication in the current research area is not realistic [2]-[9]. Since the node harvesting energy depends upon time as well as the energy source used for harvesting. The point to point problem solution can create half-duplex mode relay which leads to the formation of full-duplex mode relay. The distributed space time coding mentioned in [8] points the presence of multiple relays in the transmission path. Transmission time minimization through harvest-then-decode-and-forward algorithm in relays of research [9] also require an absolute non-causal knowledge availability. Data arrival and channel fading process can form this system model only an imaginary thought.

The causal-knowledge based study begin from the data arrival process of transmitter onwards. The online setting approach instead of offline setting [10-12] for EH process also exist, an on-off mechanism at transmitter explain here. Dynamic programming in [11] and [12] tries to achieve maximum throughput. Time invariant EH process is the powerful assumption in these all researches. But RL application in EH point-to-point communication system can create a time varying form [11]. Q-learning in RL algorithm mentioned by the authors can maximize the throughput function with channel coefficient and transmit power. Linear function approximation along with RL algorithm state-action-reward-state-action can overcome the defects of Q-learning by using the causal-knowledge available. The power allocation policy must need to learn by EH nodes, but they are not aware of the channel fading process and EH process of other EH nodes.

In this paper, EH multi-hop communication scenario is concentrated as a continuation of EH two-hop communication. Data arrival process at the transmitter and data transmission within the available conditions are taken for study. Local causal knowledge-based energy harvesting, data arrival proves and channel fading process are the real facts in the system. Transmitters and relays know their current and past amount of battery level, data buffer level, channel coefficient and incoming energy. Main objective of the system design is to create an optimum power allocation policy to form maximum throughput with less convergence time.

RL approach-based machine learning algorithm for the optimal power allocation policy in the multi-hop communication is generated. According to [13], linear function approximation and RL algorithm are applied for point-to-point communication to find power allocation policy in EH multi-hop. Feature functions for data arrival process and avoidance of data buffer overflow situation are proposed. Two problems are formulated in two-hop communication, which is the basic building block of multi-hop communication system and applied RL to the formulated problem scenario.

The remaining part of the paper is structured as follows. System design introduced in second section. In Section (3), EH two-hop communication and power allocation problem for throughput maximization is proposed. Two point to point problem for communication is formulated in the Section (4). While in the Section (5), RL is applied to the system by comparing it with the Markov decision process to obtain power allocation policies. In Section (6), design part of the system is taken to study for other facts. Performance of the designed system and result analysis are explained in Section (7) and the paper is concluding in Section (8).

2. SYSTEM DESIGN

This paper first mention about the two-hop communication system. Three nodes of communication units are taken for the transmission function. For the multi-hop communication,N_l number of nodes are considered. Initially assumes $l \in \{1, 2, 3\}$, then two-hop communication system is generated. N_1 Must transmit data to N_3. If they are far apart, there is a need of relay for transmission. Hence considering N_2 in between the two nodes. Direct communication between the nodes are not at all possible, so full-duplex decode-and-forward relay communication is establishing here. Interference inside the system onwards can be avoided due to this remedial measure. A planned and deferred data transmission link can be formed by joining such type of many two-hop communication blocks. Data arrival process assumed at the first node by R_0 data reception at the time t_i. N_2 node acts as relay for the transmission of this data towards the other node. Available data transmission at the nodes N_1 and N_2 are D_{1max} and D_{2max} respectively. $N_2$is storing the data received fromN_1. The data frames have no limits as the objective of the system is to obtain maximum throughput.

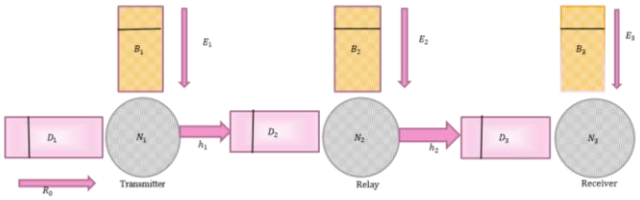


Fig -1: EH two-hop communication units

Let the amount of energy be E_l and it will be a positive real number with l = {1, 2}. Like this way in multi-hop if N_l does not harvest energy att_i, E_l = 0. The energy can be harvested from any natural resources and at theN_l, maximum amount of energy harvested beE_{max}. The harvested energy E_l will be stored in a finite rechargeable battery and the capacity maximum isB_{max}. The time interval for EH process is assumed as $[t_i; t_{i+1}]$ and the time duration T_i is taken as a constant T.The channel fading coefficient from N_1 to N_2 is h_1 and h_2 is the channel fading coefficient between N_2 and N_3 respectively. An identically independent noise reception is assumed among the nodes. Therefore, zero mean additive white Gaussian noise having variance $\sigma^2{}_1=\sigma^2{}_2=\sigma^2{}_3=\sigma^2$ is identically distributed. The power transmission ofN_l is constant at time interval i and it is p_l.Thus within the local causal knowledge availability N_l node know about its battery level B_l, data buffer level D_l, channel coefficient h_l and harvested energy level E_l (where all parameters belongs to positive real numbers), according to this N_l choose the power p_l and transmit data.

3. Problem detection

The first problem is to find power allocation policy for throughput maximization. The throughput means amount of data reached at N_3 within the time interval i in bits. R_1 is the data from N_1 to N_2 and R_2 is from N_2 to N_3 respectively. That is the throughput which is in the transmission path (which means transmitter to relay path and relay to receiver path). Throughput obtained during time interval i is given follow,

$$R_l = T log_2\left(1+\frac{|h_l|^2}{\sigma^2}\right), \quad l=\{1,2\} \tag{1}$$

Node N_l harvest energy and store in battery, thereafter utilize some part for EH process. Thus, the power allocation policy can be used for data transmission. Hence energy causality limitation is given by,

$$Tp_l \le B_l, \quad l=\{1,2\} \tag{2}$$

to avoid the energy wastage situation, finite capable battery is considered and as a result the energy flow limitation,

$$B_l - Tp_l + E_l \le B_{max}, \quad l=\{1,2\} \tag{3}$$

The data arrival process and transmission through nodes are determine the data buffer as given below,

$$D_l = \sum_{n=1}^{i-1} R_{l-1,n} - \sum_{n-1}^{i-1} R_{l,n}, \tag{4}$$

The data causality limitation can create a deadline for throughput as follow,

$$R_l \le D_l, \qquad l=\{1,2\} \tag{5}$$

Thus, the retransmission of data can be possible only after the reception.

Data buffer overflow situation can also be avoided as battery buffer overflow avoidance in (3). Thus, the information overflow limitation is given by,

$$D_l - R_l + R_{l-1} \le D_{max}, \tag{6}$$

Throughput maximization problem for EH two-hop communication condition is obtained by considering equations (2), (3), (5) and (6) with time interval i as shown below,

$$\left(P_{l,i}{}^{opt}\right)_{l,i} = \underset{\{P_{l,i}, l=\{1,2\} i=\{1,\ldots,I\}\}}{argmax} \sum_{i=1}^{I} R_{2,i} \tag{7a}$$

Prone to, $\sum_{i=1}^{M} Tp_{l,i} \le \sum_{i=1}^{M-1} E_{l,i}, \forall, M = 1\ldots, I,$ **(7b)**

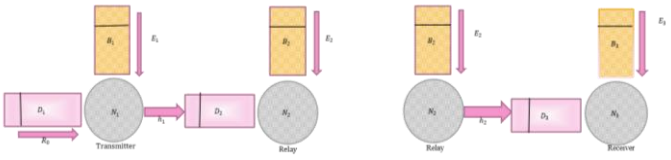
$$\sum_{i=1}^{M} E_{l,i} - \sum_{i=1}^{M} Tp_{l,i} \le B_{max,l}, \forall, l, M, \tag{7c}$$

$$\sum_{i=1}^{I} R_{l,i} \le \sum_{i=1}^{M-1} R_{l-1,i} \; \forall \; l, M, \tag{7d}$$

$$\sum_{i=1}^{I} R_{l-1,i} - \sum_{i=1}^{M-1} R_{l,i} \le D_{max,l} \; \forall \; l, M, \tag{7e}$$

$$p_{l,i} \ge 0, \forall, i = 1\ldots I, \tag{7f}$$

The local causal knowledge availability leads to the reinforcement learning approach. EH time interval, harvested amount of energy, future channel and future data buffer level cannot be predicted by assumption. Every node must find substitution for utilization and storage of energy in the current time interval to avoid overflow conditions and obtain better channel conditions.



(a) N_1 to N_2 link (b) N_2 to N_3 link

Fig -2: Two point to point communication problem (Reconstruction of two-hop communication)

Optimal power value is selected by considering maximum throughput according to equation (7a). So, expected throughput value can be taken with a discount factor γ, which is used to weight for the achievement of maximum throughput within the present interval of time. The value of γ must be in between 0 and 1 for the higher throughput achievement. As the value approaches 0, current time interval throughput maximization considered. If the value tends to 1, next time interval condition selected, and the larger value is found, future time interval is considered. Hence throughput is calculated as,

$$R = \lim_{I\to\infty} E\left[\sum_{i=1}^{I} \gamma^i R_{2,i}\right] \quad (8)$$

4. Throughput maximization problem improvisation

Energy harvesting process is independent in the two point to point links, but the power allocation policy is coupled. Throughput limitation by considering multi-hop system for communication is $R_{l,i}$. EH process, data arrival or channel process, as well as power allocation policy are not dependent to each other among the nodes. But data buffer overflow situation must be avoided. Hence the suitable measure is to adopt power allocation policy for throughput maximization by each links and it is calculated as,

$$p_{l,i}^{opt} = \underset{\{p_{l,i,i=1,2,\ldots,I}\}}{\operatorname{argmax}} \lim_{I\to\infty} E\left[\sum_{i=1}^{I} \gamma^i R_{2,i}\right] \quad (9a)$$

prone to, $\sum_{i=1}^{M} \mathrm{T} p_{l,i} \leq, \sum_{i=1}^{M-1} E_{l,i}, \forall\, M=1\ldots, I,$ (9b)

$$\sum_{i=1}^{M} E_{l,i} - \sum_{i=1}^{M} \mathrm{T} p_{l,i} \leq B_{max,1}, \quad (9c)$$

$$R_{l,i} \leq D_{l,i}, i=1,\ldots, I, \quad (9d)$$

$$D_{l,i} - R_{l,i} + R_{l-1,i} \leq D_{max,l} \forall, i \quad (9e)$$

$$p_{l,i} \geq 0, \forall, i \quad (9f)$$

Since the relays are full duplex transmitting units, data buffer overflow limitation may not be rectified all time. To avoid these problems, RL approach is taken for study and implementation.

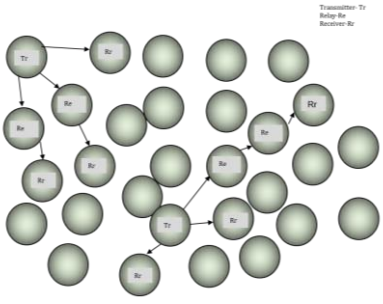


Fig -3: EH multi-hop communication system

5. Reinforcement learning method

Comparison of point to point communication problems with Markov decision process (MDP) and findings of RL for power allocation policy with throughput maximization is occurring here. State, action and reward-based functioning as mentioned in [13] is selected for study. The set of states S_l, set of actions A_l and set of rewards R_l as in [15] for N_l parameters can be considered. S_l Is substituted by energy states $B_{l,i}, h_{l,i}, E_{l,i}$ and $D_{l,i}$, (battery energy, channel coefficient Which utilized energy, harvested energy and data buffer). A_l By (optimal power allocated) $p_{l,i}$ and R_l as (throughput) $R_{l,i}$. The energy states are infinite. According to each $S_{l,i}$ and $p_{l,i}$, $R_{l,i}$ is obtained in time instant. As MDP solution [15], πl mapping can form action valued function $Q_l^{\pi l}$. The optimal values are Q^* and π^*.

Linear function approximations with feature function f can be used for finding alternate variations of state and action. State, action, reward, state, action is the order of RL learning. Reward at the optimal power value,

$$R = Q^{\pi^*}(S, P) \quad (10)$$

The action value function if infinite energy state is considered with pair of energy and optimal power is,

$$Q_l^{\pi l}(S_{l,i}, p_{l,i}) \quad (11)$$

The updates based on weight w_l is,

$$\Delta w_l = \propto_i [R_{l,i} + \gamma \widehat{Q_l}^{\pi l}(S_{l,i+1}, p_{l,i+1}, w_l) - \widehat{Q_l}^{\pi l}(S_{l,i}, p_{l,i}, w_l)] \nabla_{w_l} \widehat{Q_l}^{\pi l}(S_{l,i}, p_{l,i}, w_l) \quad (12)$$

Action value function with probability 1-ϵ is calculated as,

$$P_r[p_{l,i} = \underset{P_{l,k}\in A_l}{max} \widehat{Q_l}^{\pi l}(S_{l,i}, p_{l,k})] = 1\epsilon,\ 0<\epsilon 1 \quad (13)$$

while dealing with overflow condition, feature function is given as

$$f_1(S_{l,i}, p_{l,i}) = \begin{cases} 1, & if\,(B_{l,i} + E_{l,i} - \mathrm{T} p_{l,i} \leq B_{max,l}) \wedge (\mathrm{T} p_{l,i} \leq B_{l,i}) \\ 0, & else \end{cases} \quad (14)$$

6. Requisition to other situations

EH multi-hop communication systems are of different types. Those having single transmitter and receiver, those having multiple transmitters and receivers, those which has amplify and forward two-hop building blocks and so on. In every condition, a machine learning algorithm is used for RL

functioning as a part of self-study and analysis for the adaptations. The algorithm is given below,

6.1 Machine learning algorithm [13]

Step 1: Initialize γ, α, probability and w_l
Step 2: Observe $S_{l,i}$
Step 3: Select $p_{l,i}$ using the probabilistic greedy policy
Step 4: **While** node N_l is harvesting energy **do**
Transmit using the selected $p_{l,i}$
Calculate corresponding reward $R_{l,i}$
Observe next state $S_{l,i+1}$
Select next transmit power $p_{l,i+1}$ using
Probabilistic greedy update w_l
Set $S_{l,i}$=$S_{l,i+1}$ and $p_{l,i}$=$p_{l,i+1}$
End while

For the first condition, independent point to point communication problem is solved and find the data transmission policy by using the algorithm. In the second condition, reward is the weighted throughput and so various weights are selected for the data transmission purpose by transmitter with the algorithm. Third condition is resolved by forming a combined transmission policy by transmitter and relay for the achievement of maximum throughput.

7. Performance and result analysis

EH multi-hop communication system generated with 30 nodes and used the system for RL approach. Self-learning and adaptation of the system is significant. Machine learning algorithm has two stages, learning stage and exploration stage. In the learning stage, optimal power value is selected for throughput maximization. Here, different power values are given to the state for actions to obtain maximum reward within the available conditions. Thus, exploration probability value is selected. While in the exploration stage, execution of operations with optimal power value is occurring. Spectrum channel Wi-Fi having IEEE802.11a standard is used in the simulation. The rate control function is done by Remote station manager in the Wi-Fi mac layer. Before the transmission of a frame, the Remote station manager check all the status like power level, rate, number of antennas, channel condition etc… of the multi-hop system and record selfly as transmission vector. Whether any obstruction or unsuitable condition occur in the system, it will be reported as shown in Fig -7.

Thousand independent random channels are chosen, and infinite data transmission is also considered here. N_l Node is harvesting energy within an interval of time. According to Rayleigh fading process, channel coefficient is assumed zero mean and unit variance. The step size is taken as 0.02 times maximum battery buffer level. The learning rate α and probability is for probabilistic greedy policy and α=1/*i*, so probability=1/*i*. Discount factor γ=0.9.

First source and harvester energy is compared and plot it as in Fig -4. Then the remaining energy in the system, amount of energy harvested and consumed are generated. Initial system is not an advanced type, so the variations of parameters are not considered widely. Here, time is taken in x-axis and energy is taken in y-axis respectively.

Unit for energy is Joule and time is second as the interval for energy harvesting process. In Fig -4, harvesting energy is varying and battery source of energy also varies according to harvested energy, the increasing energy is not in linear manner.

After fully charged, there will be no variation and hence constant. The$S_{l,i}$, has a 4 dimensional space as$D_{l,i}$, $B_{l,i}$,$E_{l,i}$and$h_{l,i}$.Battery size can be set as $B_{max,1}$=$B_{max,2}$=B_{max}=2E_{max} and here the time instant *I*=100 times EH instant time. Data buffer level D_{max} as 5 times$R_{l,i}{}^{B_{max}}$, it means that throughput can be obtained if $p_{l,i}$ = B_{max} / T and $|h_{l,i}|$ = 1. Performance progresses according to the energy harvesting process. Unrealistic assumptions can be avoided by the Machine learning algorithm. Data buffer size effect isE_{max}/ $2\sigma^2$=5dB and for N_2 buffer size is D_{2max}=$\beta R_{l,i}{}^{B_{max}}$, β is tunable constant. If the value is less than 1, buffer overflow condition cannot be avoided. Thus, for problem (7) has no solution. This was offline non-causal assumption. The system can give highest value and hence buffer overflow condition is reduced, saturated value is 3.

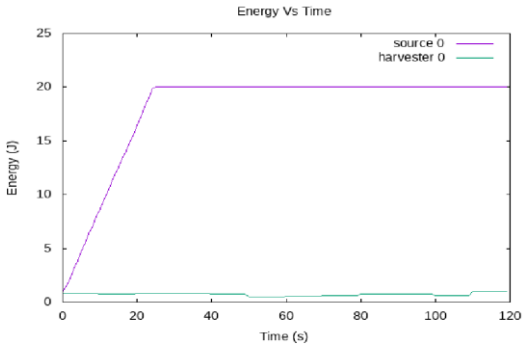


Fig- 4: Source and harvester comparison

The function call diagram in Fig- 7 explains about mac layer functioning. State, action and reward-based system transmission has Remote station manager for controlling all the performance. At the receiver section, checks whether the data reached safely and reports. In the transmitter section, data, request to send and all the parametric conditions like energy level, channel etc… are verified and record in transmission vector or matrix form. Nodes with internet protocol address and transmission is shown in the Fig. 5. The channel considered as first slow flat fading channel. Since AWGN, with matched filter and sampler discrete-time model created.

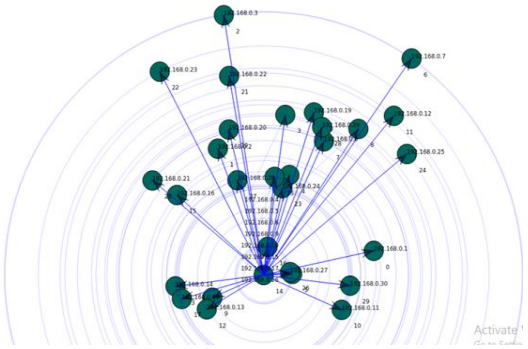


Fig- 5: EH multi-hop transmission simulation

The battery size factor is taken as μ. For maximum throughput, this value will be less than 2. To reduce convergence time, tunable constant value is taken as 5. The feature functions used in the system can find the learning rate. Learning stage really update each working status and modify the changes. Fig- 6 shows the frame transmission in 30 nodes. The size of frame packets in bytes are explained with respect to time intervals in second.

Scaling factor for the size is 512 bytes. Energy used by transmitters are more than relay power consumption and hence if transmitter and relay are closer, power consumption will be very less. If they are far apart, relays are needed for transmission, so power consumption will be little more than the above case. System design automatically maintains the power consumption rate. Nearest nodes can transmit data directly with less power and nodes at long distance can transmit only through relays.

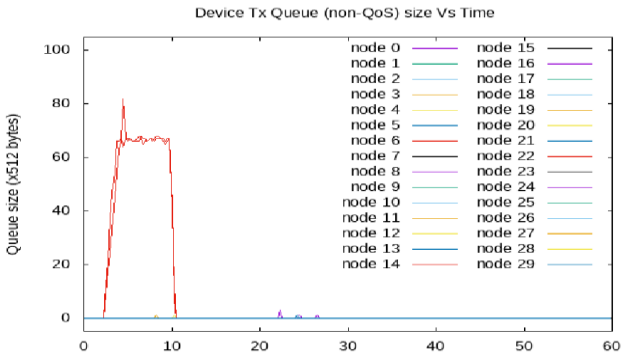


Fig- 6: Energy harvesting and energy consumption

At the starting time of transmission, power consumption will be little more, but gradually the consumption rate decreases. Relays in the system have forwarding function. Data size and transmission are different for different nodes, hence optimal power allocation policy adaptation is useful to the system. Reinforcement learning method is basic functioning approach for modifications of the designed system. The multi-hop communication system introduces here have the capability to transmit multiple data with less time consumption, as well as with small amount of energy usage.

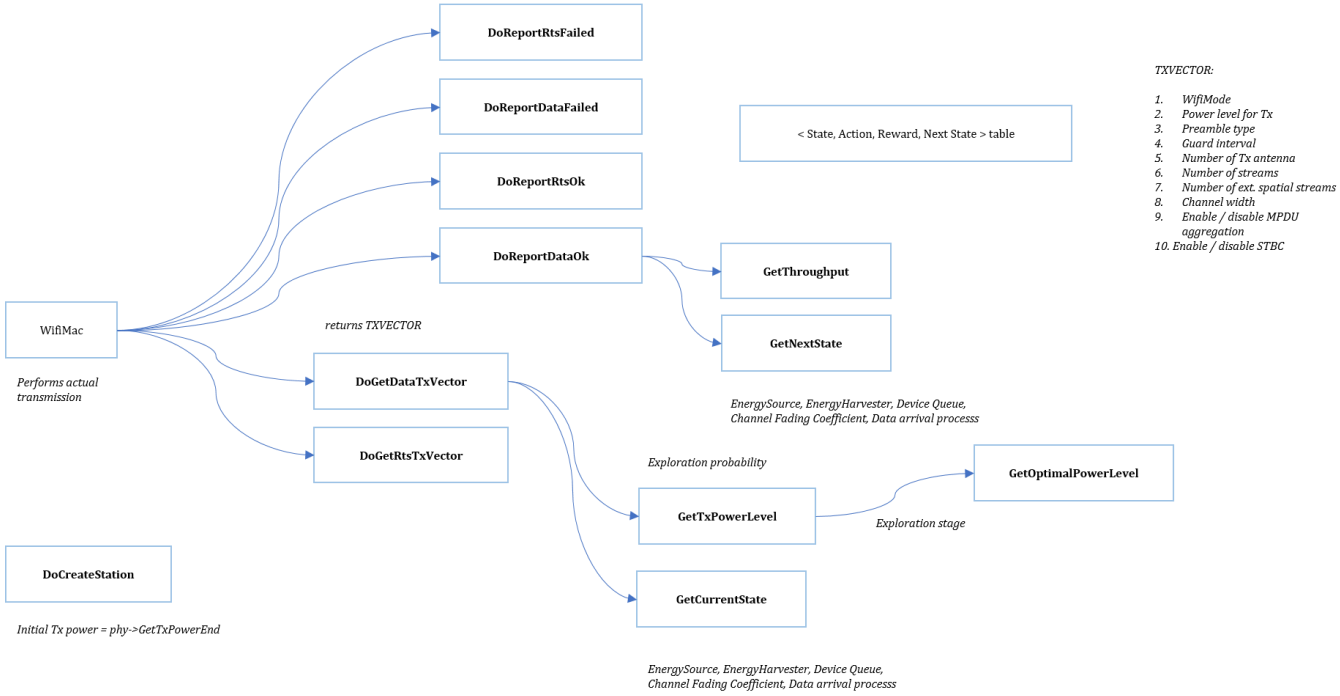


Fig- 7: Function call diagram

8. CONCLUSION AND FUTURE SCOPE

An optimal power allocation policy based on planned and deferred data transmission in energy harvesting wireless devices studying and executing broadcast transmission in multi-hop communication system. Convergence time reduction and fragmentation frame modification are also designed in the system model. Local causal knowledge based EH process, data arrival process and channel conditions of transmitters and receivers are selected for study. Reinforcement learning and Markov decision process with linear function approximation is used in the Machine learning algorithm generation. Data arrival process are explained with feature function and the performance are verified by Wi-Fi mac layer unit. Thus, the limitations of non-causal knowledge-based system are rectified by the causal knowledge based multi-hop communication system. Thus, the performance of system is improved through its realistic approach. In future, each parameter can be taken for study by the algorithm and can improve system model.

REFERENCES

[1] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communication: A review of recent advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 360– 381, March 2015.

[2] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," IEEE Trans. Wireless Commun., vol. 11, no. 3, pp. 1180–1189, March 2012.

[3] B. Varan and A. Yener, "Two-hop networks with energy harvesting: The (non-)impact of buffer size," in Proc. IEEE Global Conf. Signal Inform. Process. (GlobalSIP), Austin, December 2013, pp. 399–408.

[4] D. Gu¨ndu¨z and B. Devillers, "Two-hop communication with energy harvesting," in Proc. IEEE Int. Workshop Comput. Advances Multi- Sensor Adaptive Process. (CAMSAP), San Juan, December 2011, pp. 201–204.

[5] D. Gu¨ndu¨z, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing intelligent energy harvesting communication systems," IEEE Commun. Mag., vol. 52, no. 1, pp. 210–216, January 2014.

[6] O. Orhan and E. Erkip, "Energy harvesting two-hop communication networks," IEEE J. Sel. Areas Commun., vol. 33, no. 12, pp. 2658–2670, December 2015.

[7] A. Zanella, A. Bazzi, and B. M. Masini, "Analysis of cooperative systems with wireless power transfer and randomly located relays," in Proc. IEEE Int. Conf. Commun. (ICC), Workshop Green Commun. with Energy Harvesting, Smart Grids and Renewable Energies; London, June 2015, pp. 1964–1969.

[8] Y. Liu, "Wireless information and power transfer for multirelay-assisted cooperative communication," IEEE Commun. Lett; vol. 20, no. 4, pp. 784–787, April 2016.

[9] L. Tang, X. Zhang, and X. Wang, "Joint data and energy transmission in a two-hop network with multiple relays," IEEE Commun. Lett; vol. 18, no. 11, pp. 2015–2018, September 2014.

[10] J. Lei, R. Yates, and L. Greenstein, "A generic model for optimizing single-hop transmission policy of replenishable sensors," IEEE Trans. Wireless Commun., vol. 8, no. 2, pp. 547–551, February 2009.

[11] P. Blasco, D. Gu¨ndu¨z, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," IEEE Trans. Wireless Commun., vol. 12, no. 4, pp. 1872–1882, April 2013.

[12] I. Ahmed, A. Ikhlef, R. Schober, and R. K. Mallik, "Power allocation in energy harvesting relay systems," in Proc. IEEE Veh. Technol. Conf. (VTC Spring), Yokohama, May 2012, pp. 1–5.

[13] A. Ortiz, H. Al-Shatri, X. Li, T. Weber, and A. Klein, "Reinforcement learning for energy harvesting point-to-point communications," in Proc. IEEE Int. Conf. Commun. (ICC), Kuala Lumpur, May 2016, pp. 1–6.

[14] A. Geramifard, T. J. Walsh, S. Tellex, G. Chowdhary, N. Roy, and J. P. How, "A tutorial on linear function approximators for dynamic programming and reinforcement learning," Foundations and Trends in Machine Learning, vol. 6, no. 4, pp. 375–454, December 2013.

[15] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2010.