Dynamic Channel Sharing Strategies through Game-theoretic Reinforcement Learning

Eric Masiga Ayienga, Elisha Toyne Opiyo, William Okello-Odongo
School of Computing and Informatics
University of Nairobi
Nairobi, Kenya
Email: Ayienga {at}uonbi.ac.ke

Bernard Manderick
Computational Modeling (COMO) Lab
Vrije Universiteit Brussel (VUB)
Brussels, Belgium

Abstract—**Random access protocols are used by multiple nodes in wireless networks to effectively share a wireless channel for data transmission. While competing for the channel, the nodes seek to achieve an individual or group objective. Game theory, can thus be applied to analyze and model individual or group behavior of nodes in random access networks. It can also be used as an 'engineering' application and subsequently re–engineer the system. In this paper, the current CSMA/CA mechanism was analyzed using game theory. Based on the analysis, the strategy space available to individual nodes was increased so that the optimal strategies for different situations learnt using reinforcement learning. From the analysis it was determined that the Nash equilibrium was not Pareto optimal. Simulation experiments yielded better results for the modified algorithm pointing to moving the Nash equilibrium towards being fair and Pareto optimal.**

Keywords-Distributed Channel Sharing; Random Access; Medium Access Control; Game Theory; Strategy; Reinforcement Learning; Q-learning.

I. INTRODUCTION (HEADING 1)

A network of computers that use a multi--access medium requires a protocol for effective sharing of the media. In this broadcast mode, the main problem is how different nodes get control of the medium to send data, i.e. "who transmits next?" The protocols used for this purpose are known as Medium Access Control (MAC) protocols. The key issues that the MAC addresses are where and how control is exercised. Control can be exercised in a centralized} or distributed manner. In a centralized system a master node grants access of the medium to the other nodes. Although this scheme has a number of advantages and is easier to implement, it is vulnerable to the failure of the master node leading to reduction in efficiency and limited scalability. On the other hand, in a distributed approach all the nodes collectively perform the medium access control function which dynamically grants them access. This thus makes the approach more attractive to networks that are characterized by a distributed, dynamic, self-organizing architecture such as ad hoc, mesh and sensor networks. How control is exercised is constrained by the topology and trade off between cost, performance and complexity. Many formal approaches to medium access control have been devised as [1] cited in [3], but are we concentrate on random access protocols and in particular the carrier sense multiple access with collision avoidance (CSMA/CA) as shown in Fig. 1

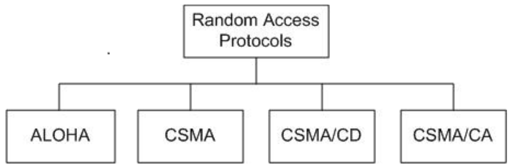

Figure 1. Taxonomy of fundamental Random Access Protocols

Each node running the CSMA/CA mechanism must make individual decisions that depend on the conditions in the wireless environment. The nodes independently set parameters like contention window (cw), backoff time, transmit power, packet forwarding etc. In making these decisions, the nodes seek to optimize one or more of the following [6]:

- The global optimum network performance, e.g. the network operating efficiently through fair distribution of bandwidth, reduced delay, etc.
- The local optimum by acting only in their self--interest.
- Malice, by seeking to ruin network performance.

Designing a medium access mechanism in which nodes behave in desirable ways gives the network appealing features and hence motivates our application of game theory. In the first case, game theory can offer some useful insights because, even when nodes have shared objectives, they will each still have a unique perspective on the current network state, leading to possible conflicts regarding the best course of action [6]. In second and third cases, the application of game theory is straightforward, since game theory traditionally analyzes situations in which player objectives are in conflict. Hence game theory can be used in two ways [13]:

- Direct application in the analysis of the CSMA/CA MAC mechanism.

This research was supported in part by the University Cooperation for Development of the Flemish Interuniversity Council (VLIR-UOS-IUC) Project and the School of Computing and informatics, University of Nairobi.

- An "engineering" application to reverse engineer the CSMA/CA MAC mechanism using mechanism design.

In this work we have used both to try and ensure that the global efficiency and fairness of the CSMA/CA is maximized under all the circumstances above.

This paper is organized as follows: Section II gives background of the CSMA/CA mechanism. Section III explains the use of game theory in analyzing the CSMA/mechanism. Section IV introduces the idea of learning in the mechanism. Section V presents the proposed model. Some simulation results are presented in Section VI. Analysis of the simulation result is done in Section VII. Section VIII concludes the paper.

II. BACKGROUND

The use of distributed coordination function (DCF) is mandatory in ad-hoc LANs and hence we shall focus on it and specifically DCF with CSMA/CA. Using CSMA/CA, there are several parameters that are used to control the waiting time before a node can access the medium. The values of the parameters depend on the type of modulation being used by the physical layer. They are defined in relation to a slot time. This is derived from the propagation delay of the medium and delays in the transmitter, and is 20 µs for direct sequence spread spectrum (DSSS) and 50 µs for frequency hopping spread spectrum (FHSS). The medium is idle if a signal called the clear channel assessment (CCA) is present. The two most important time parameters are shown in Fig. 2

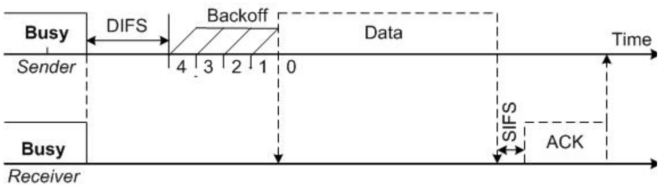

Figure 2. The CSMA/CA Basic Access Mechanism showing the timing diagram of two nodes competing for a channel.

- Short inter-frame spacing (SIFS)}: this is the shortest waiting time – and hence highest priority – and is used for control messages such as data acknowledgments (ACK).
- DCF inter-frame spacing (DIFS)}: this is the longest waiting time – and hence lowest priority – and is used for asynchronous data service.

$$\text{DIFS} = \text{SIFS} + 2(\text{slot times}). \tag{1}$$

Using the CSMA/CA mechanism, if the medium is sensed idle for at least the duration of DIFS, the node can access the medium at once and transmit. Upon successful transmission, the receiver sends back an acknowledgment (ACK) frame after a SIFS period. This allows for short access delay under light load. But, as soon as more and more nodes try to access the medium, it becomes overloaded and additional mechanisms are needed to mitigate collisions.

A busy channel indicates that the node has lost the cycle and has to wait for the duration of DIFS, and then enter a contention phase. In this phase, each node now calculates a backoff time (r) and additionally delays medium access for this

$$r = \text{Random}[0, cw] \times \text{slot time}. \tag{2}$$

- Random is a random number generator function which randomly selects a number from a uniform distribution [0, cw].
- Contention window (cw): A number computed using the equation:

$$cw = 2^{\{BE\}}\text{-}1. \tag{3}$$

- Backoff exponent (BE): Used for the computation of the cw value. Some of the values used include 3, 5 etc. to give cw values of 7, 31 etc. The cw is bounded by $cw_{min} \leq cv \leq cw_{max}$. After each successful transmission, cw is reset to the minimum contention window size cw $_{min}$.

If the randomized additional waiting time for a node is over and the medium is still idle, the node can access the medium immediately. If the channel is sensed "busy" during the backoff, the backoff is paused until the channel is sensed idle again, and then resumed. The CSMA/CA mechanism is summarized by Algorithm 1.

Algorithm 1. The Basic DCF MAC Protocol:

```
If medium is free for ≥ DIFS
        Transmit
Else backoff
        Wait for medium to be free for DIFS
        Choose a backoff time (r)
While r > 0:
        Sense medium for one slot time
        If medium free throughout slot → r := r – 1
Transmit frame
```

Analysis of the CSMA/CA mechanism shows that there are a number of MAC parameters at the data link layer that can possibly be engineered to obtain optimal network behavior. The parameters in question include:

- Interframe space (IFS) i.e. SIFS and DIFS. But these parameters are governed by the laws of physics at the PHY layer.
- cw sizes. These are set by the mechanism at the MAC layer and hence are more practical in terms of manipulation.

III. USING GAME THEORY TO ANALYZE THE MAC PROTOCOL

Game theory is a bag of analytical tools designed to help us understand the phenomena that we observe when decision--makers interact [12]. Its purpose is to predict what will happen when a game is played. It is relevant to random access networks because of the following features:

- Decentralized operation of the nodes.

- Node self-configuration.
- Power/energy awareness of the nodes.
- Each node in the network runs the distributed CSMA/CA protocol, making decisions that affect other nodes.

We model the random access network environment running the CSMA/CA mechanism as an incomplete information, static, non-cooperative, n-person repeated game. A typical mapping of random access network components to a game is as shown in Table 1 below:

TABLE I. TYPICAL MAPPING OF AD HOC NETWORK COMPONENTS TO A GAME

Components	Elements of a Random Access Network
Players	Nodes in the network
Actions	Actions related to the functionality being modeled (e.g. setting the value of the cw or Backoff)
Utility/Payoff	Improved performance (e.g. increase in efficiency and fairness)

A. The Normal--form CSMA/CA Game

In a random access network, nodes carry elastic traffic. In these networks, the fraction of channel time that is used for successful packet transmissions is called effective capacity (T). The rest of the channel time is consumed by control packets, the backoff process or packet collisions which constitutes the inefficiency of the network. The effective capacity should ideally be shared out fairly, with each node receiving a fraction of T i.e. t. This environment becomes a game once nodes knowingly or unknowingly exploit protocol imperfections for self gain. We illustrate this through an example.

Let us consider an ad hoc network model, in which there are 2 nodes repeatedly competing for the channel as shown in Fig. 3

Figure 3. A 2-node Random Access Network used to demonstrate the Prisoner's Dilemma.

In this environment, a transmission is successful if the nodes do not access the channel simultaneously. We assume that the two nodes are homogeneous and can use any of two moves that are defined by the cw size as shown in Table 2.

TABLE II. THE POSSIBLE MOVES THAT CAN BE USED BY THE TWO NODES

Action	Description
c	Complying with the standard protocol by truthfully using the specified cw parameters.
v	Violating the standard protocol by for example by maintaining a small fixed cw_{min}.

We also assume for purposes of this example, that the utility (u) of a node is defined by throughput (t). The outcomes of the combination of various moves are summarized in Table 3.

TABLE III. OUTCOMES OF NODES USING DIFFERENT ACTIONS

Action	Outcome
Both nodes use v	u = t_{vv}: The probability of collision increases considerably, reducing u.
Both nodes use c	u = t_{cc}: The two nodes have an almost equal chance of accessing the channel. The probability of collision is minimized and t is shared almost equally.
One node uses v, the other uses c	The node using v unfairly gets most of the capacity t at the expense of the node using c.

We can now illustrate this strategic scenario as a normal-form game where:

- Players: N = {1, 2}
- Actions: A = {c, v}
- Utilities/Payoffs: $u_i = \{0, t_{vv}, t_{cc}, t\}$ for $i \in N$

The game can be summarized by the payoff matrix shown in Fig. 4

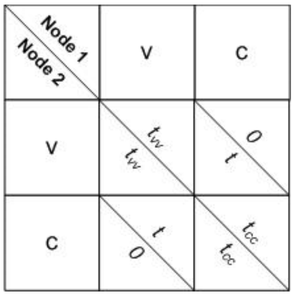

Figure 4. Payoff Matrix for a Two Node Random Access Network.

$$0 < t_{vv} < t_{cc} < t$$

Analysis using game theory shows that, there is a unique strict Nash equilibrium} [10], which is (v, v). v dominates the c. v is a stable equilibrium while c is an unstable one. Assuming one node is playing v, the best choice is to play v as well. Similarly, assuming a node is playing c, the best choice is to play v. If both players commit to play the dominated strategy c, they will be better off. The normal-form game runs into a Prisoner's Dilemma [14] that is neither fair nor Pareto-optimal.

This simple example can be extended to more nodes. This can be formally defined as:

- Game: A finite n-person game (N, A, u)
- Players: $N = \{n_1, n_2 \ldots n_k\}$.
- Actions: A = {c, v}

It can be concluded that the CSMA/CA mechanism works well if all nodes follow the predefined rules. But violating the protocol promises greater rewards. Nodes do not have information about each other and if they all decide to individually violate the protocol there is network collapse. This leads to a phenomenon referred to as tragedy of the commons [4]. An issue that arises from this is: How do we make the Nash equilibrium Pareto--optimal and fair?

B. From a Single Strategy to Multiple Strategies

In gaming, players' actions} are governed by laid down rules and procedures that the players need to follow. A sequence of actions is called a strategy. A player's strategy in a game is a complete plan of action for whatever situation that might arise; this fully determines the player's behavior. A strategy must specify what action will happen in each contingent state of the game.

Strategies in game theory may be deterministic (pure) or random (mixed). Pure strategies can be thought of as a special case of mixed strategies, in which only probabilities 0 or 1 are assigned to actions. Situations that involve interdependent decisions are everywhere, potentially including almost any endeavor in which self--interested agents cooperate and/or compete as shown by the following examples:

- Soccer: It is only simple in the rules and the basic game play. The strategy of the game can be quite complex and is what determines the outcome. Example strategies used include 4-2-4, 4-3-3, etc.
- Chess: The strategy consists of formulating a plan and arranging the chess pieces to accomplish this plan in view of the opponent's best response.

From the two examples, one notes that by having different combinations of actions, it is possible to come up with many strategies. A optimal strategy is a sequence of moves that results in the best outcome.

Using the CSMA/CA mechanism as currently implemented, nodes are hard-wired to use a single strategy. This strategy only relies on the randomness generated by the probability space of the cw to ensure efficiency and fairness. Based on its variation, it is possible to come up with many strategies. The role of analysis and reverse engineering is the identification of strategies that can be used to get an optimal outcome. Current approaches have concentrated on experimentation to get the optimal parameters. But the random access network environment is dynamic and the parameters should change with the environment, hence the need to make the medium access mechanism adaptive. The objective of each node is to maximize it payoff or utility function in the network through maximizing its total throughput t_i, reducing medium access delay d_i and increased network utilization p_i. If we denote the payoff of node i by u_i, its payoff function is written as follows:

$$u_i = f(t_i, d_i, p_i). \quad (4)$$

This should also have the effect of increasing overall network fairness f_N. The total utility is defined as the sum of the achieved utilities of all of the nodes on channel c, given by:

$$u_c = \sum_i u_i \quad (5)$$

This is a non increasing function of the number of nodes deployed on c. If the CSMA/CA mechanism is perfect, u_c is independent of the number of nodes on c. In practice however, the cw and r values used in the CSMA/CA implementation are not optimal and owing to this, u_c becomes a decreasing function of the number of nodes on c.

To characterize stability in the random access game, we introduce the concept of the Nash Equilibrium. The strategy profile $s^* = s^*_1, \ldots, s^*_n$ defines a Nash Equilibrium (NE), if for each node i, and its strategy $s'_i \in s$ we have $u_i(s^*_i, s^*_{-i}) \geq u_i(s_i, s^*_{-i})$. This means that in a Nash Equilibrium, none of the nodes has an incentive to change their strategy to increase their payoffs. This is achieved by in-cooperating learning to ensure that nodes have the capability to adapt their strategies based on the network environment.

IV. Learning in the Random Access Network Mechanism

Building on ideas from the El Farol Bar Problem (EFBP) [1] and Marimon [8][7] we introduce the concept of a strategy space $[s_1, s_2, ..., s_n]$ to the nodes. The existing CSMA/CA protocol, is modified so that all the nodes have access to any known strategies. When the modified mechanism is in operation, nodes assign a score to each strategy used based on its immediate payoff. Each node uses the strategy with the highest score. This increases its probability of re-use in future. If strategies have the same score, then during the selection process, one is randomly selected. An unsuccessful strategy is unchanged so that it can be ignored in future. This is done through the use of reinforcement learning [9]. The rationale behind this is that random access networks are dynamic environments; hence nodes need to be adaptive if the network is to perform optimally. If the strategy a node is using wins a contention under the prevailing circumstances, the node continues using it hence exploitation. If it loses the contention, the node tries other strategies hence exploration. It is notable that when a node uses the v strategy, it only has an advantage when other nodes don't use it. By introducing multiple strategies, by other nodes changing strategy, they eventually end up with the v strategy. This forces the violating node to also change strategy and shift the operating point. The changing of strategies is a way of looking for suitable operating points so that system performance is maximized under the dynamic conditions. This enables the nodes adapt strategies so that either the strategies played converge to a robust equilibrium or they circle around a set of correlated strategies.

A. Q-Learning

To reinforce the strategies we use a variant of reinforcement learning known as Q-learning. Though some issues arise with using the basic Q-learning algorithm for nodes using the CSMA/CA protocol

- The traditional Q-learning is effective for a single learner in a stationary environment. The random access network has many nodes and the environment is generally non--stationary due to adaptation of other agents.

- In the traditional Q-learning, the Q-matrix is only used once there is convergence. In our case, the end of the game cannot be pre-determined and so is assumed to be infinite. The convergence of the Q-matrix for the nodes is therefore only theoretically possible.

We extend the Q-learning algorithm to the multi--node non--static game setting by having each agent simply ignore the other agents and pretend that the environment is passive.

V. PROPOSED MODEL

A. *Reverse Engineering the MAC Protocol*

Our game theoretic analysis has shown the Nash equilibrium of the current mechanism to be sub-optimal. From this analysis, the network is vulnerable to manipulation since there is a higher payoff for nodes if they use the v strategy. With the new programmable radio adapters and the newer IEEE 802.11e protocol [2], this is becoming a reality. The moves made by the nodes are not premeditated. They are simply a reaction to the current state of the environment with no adaptivity. This can lend itself to suboptimal behavior and/or abuse.

Optimization of the network performance involves seeking a stable operating point based on the above stated tunable parameters. Thus there is need to find a socially optimum way in which the cws of the various nodes can be set to maximize the performance of the system. So every node, without any coordination, should behave in a synchronized fashion to achieve the socially optimum network performance through setting their cw. We use game theory and learning to propose an enhancement that makes the mechanism more dynamic by nodes having more strategies that make them adapt to the network environment conditions.

B. *Model Design*

Our scheme alters the way nodes contend for the medium using the CSMA/CA method. Formally this can be defined as:

- Players: $N = \{n_1, n_2, \ldots, n_k\}$
- Strategies: $S = \{s_1, \ldots, s_m\}$, where s_i is a combination of different actions.
- Utilities/Payoff: $u_i(s)$ for $i \in N$ and $s \in S$

When the network is set up, all strategies s_i in the strategy space are initialized to have the same rank i.e. 0. Nodes randomly select any strategy from the strategy space. If the selected strategy wins the contention and the node accesses the medium or if the strategy fails a node in accessing the medium it is positively or negatively reinforced. The reward scheme is based on the formula:

$$\text{Reward} = (N_s - N_c). \tag{6}$$

Where

- N_s: Number of successful transmissions
- N_c: Number of collisions

Intuitively, one can see that winning a contention will result in a positive reinforcement while losing the contention will result in a negative reinforcement. Nodes learn by using this simple reinforcement learning scheme as shown in Fig. 5

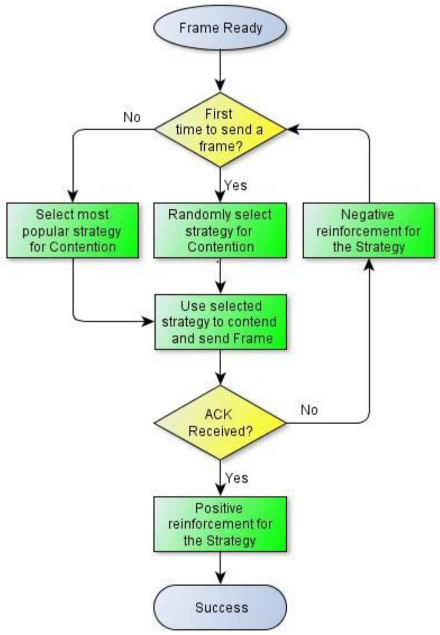

Figure 5. The Working of the new scheme.

Our modified algorithm is a combination of the Win-Stay, Lose-Shift (WSLS) strategy [11] and Q-learning [16] as shown in Algorithm 2.

The modified win stay-lose shift strategy with Q-learning

Algorithm 2. Modified Q-Learning algorithm

Begin: For every contention
    Consult the Q-matrix
    Use the best action from the stored Q-values
    Evaluate the reward
If Positive
    Update the Q-matrix
Else
    Randomly Explore another action from the stored Q-values
If there is another round, go to Begin.

The interaction of the node running the algorithm in an environment is as shown in Fig. 6

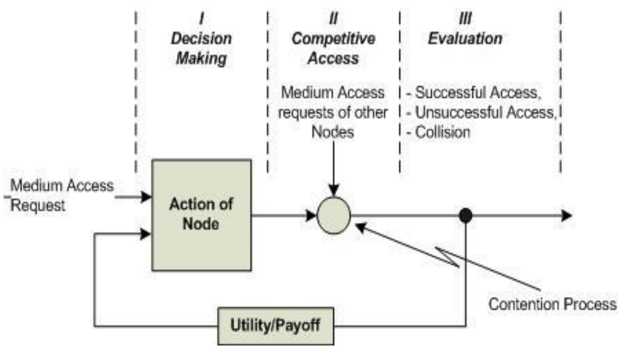

Figure 6. Working of the node

As the node learns and updates its Q-matrix, the matrix reflects the direction of convergence. During each episode, the Q-matrix remains the same or gets better compared to the previous one. Therefore the Q-values for each state-action pair represent how each move benefits a particular node at a particular time.

Nodes keep doing this as they try to settle on an optimum strategy. There is a possibility of learning the best strategy through reinforcement. Equilibrium is reached when the aggregate throughput of all the contending nodes oscillates around the bandwidth capacity of the network.

VI. Experimental Design

A. Game Design

We assume a single IEEE 802.11 basic service set (BSS) with a set of N nodes running heterogeneous applications and operating in the distributed coordination function (DCF) mode. Each node selfishly tries to transmit whenever it has data, assuming the existence of other nodes. Formally this can be defined as:

- Players: $N = \{n_1, n_2, \ldots, n_{20}\}$
- Strategies: $S = \{s_1, s_2, s_3, s_4\}$

  The four strategies explored in our work are taken from the IEEE 802.11e parameters and one where the cw_{min} is kept constant as shown in Table 4:

- Utilities/Payoff: $u_i(s) = f(t_i, d_i, p_i)$

  For $i \in N$ and $s \in S$

TABLE IV. Using IEEE 802.11e CSMA/CA parameters to create different strategies.

Strategies	**Parameters**
s_1	The standard DCF where r is calculated from the range $[0, cw_{min}]$
s_2	Calculates r from the range $\left[\left(\frac{(cw_{min}+1)}{2}-1\right), cw_{min}\right]$
s_3	Calculates r from the range $\left[\left(\frac{(cw_{min}+1)}{4}-1\right),\left(\frac{(cw_{min}+1)}{2}-1\right)\right]$
s_4	A small fixed cw size having a value of 8 slots

An improved mechanism should display higher efficiency through increased throughput, network utilization and increased fairness, reduced delay and retransmission attempts.

B. Simulation Characteristics

The proposed protocol was compared to the existing protocol by simulation through the use of the Pamvotis WLAN simulator version 1.1 [15]. The simulation characteristics were as shown in Table 5.

TABLE V. Typical mapping of ad hoc network components to a game

Characteristic	**Type**
Network Size	20 Nodes
Bandwidth	1 Mbps
Radio Type	DSSS
Applications	FTP Source Generic Source HTTP Source Video Source

The topology used for 20 nodes is as shown in Fig. 7

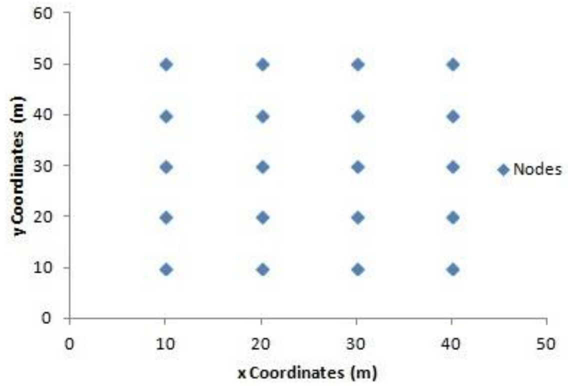

Figure 7. A twenty node topology used in the experiments. All the nodes are within range of each other

C. Performance Metrics

The proposed model was evaluated against the existing MAC protocol based on:

- Throughput (packets/sec): The number of packets that a node successfully transmitted in a specific time interval.
- Utilization: The percentage of the channel capacity the node occupied. The utilization is the node's throughput in bits per second divided by the node's data rate.
- Media access delay: The delay of a packet from the time it is picked up from the transmitter until it is successfully received from the receiver. This statistic contains the delay due to retransmission attempts and the transmission delay.
- Fairness: How well the system shares bandwidth among multiple users. It is an important consideration in most performance studies especially in distributed systems where a set of resources is to be shared by a number of users. Assuming that fair implies equal and that all paths are of equal length, Raj Jain [5] proposed the following fairness index. Given a set of a set of flow throughputs, $(x_1, x_2, \ldots, x_n)$ the fairness index $f(x_i)$:

$$f(x_i) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i^2}. \tag{7}$$

The fairness index always results in a number between 0 and 1, with 1 representing greatest fairness.

VII. Results and Analysis

A. Prisoner's Dilemma and Tragedy of the Commons

To demonstrate the Prisoner's Dilemma, two nodes in the network i.e. node 1 and node 3 were configured to have a smaller cw_{min} than the rest of the nodes. On running the simulation, the results obtained are as shown in Fig. 8].

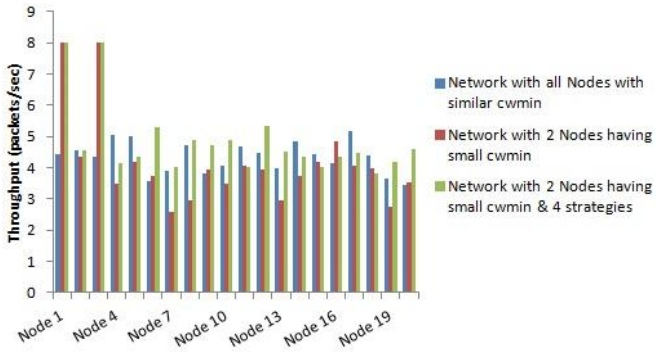

Figure 8. Comparing performance when two nodes set a small cw_{min} and use one or 4 strategies.

The nodes with a smaller cw_{min} get an advantage over the other nodes as exhibited by the throughput of node 1 and node 3. Increasing the number of strategies available to the nodes, in most cases slightly increased the individual node and system throughput.

All the nodes are then configured with a standard cw_{min} followed by a smaller cw_{min}. The respective throughputs are as shown in Fig. 9.

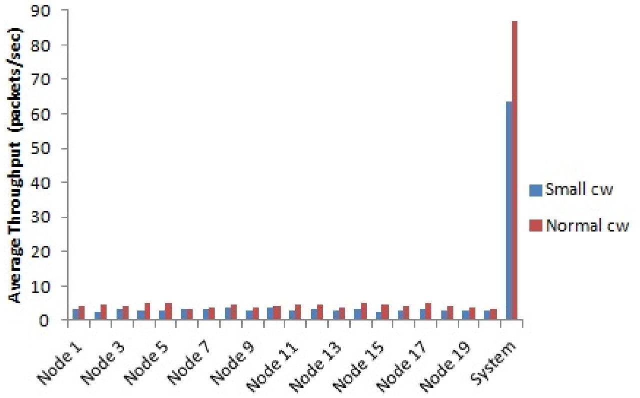

Figure 9. Demonstrating the tragedy of commons when all nodes use the normal cw_{min} and a smaller cw_{min}

By using a smaller cw_{min}, than the standard cw_{min}, both the individual and system throughput are reduced. This demonstrates the tragedy of the commons.

B. Performance of Individual Strategies and Combinations of various Strategies

All the nodes are configured with each of the four strategies and various combinations of the strategies in turn. The results are summarized in Fig. 10.

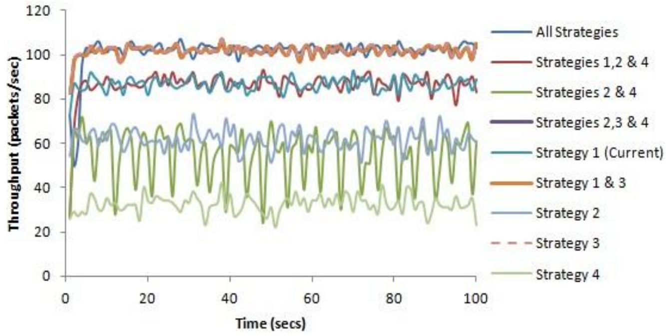

Figure 10. Comparing the throughput of individual strategies and various combinations of strategies.

The averages of throughput are also summarized in Fig 11.

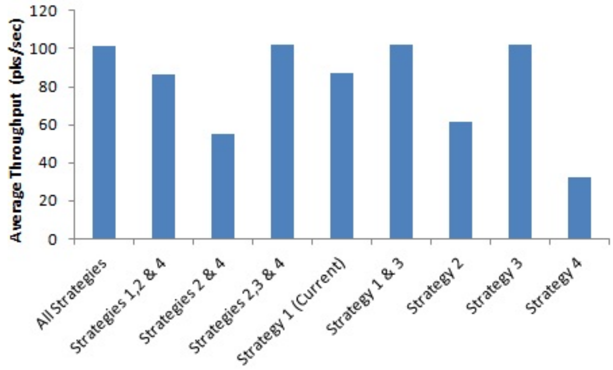

Figure 11. The average throughput when different strategies and different combinations of strategies are used.

Various observations can be made. Strategy 3 is the best performing strategy while strategy 4 is the worst performing. Strategy 1 which is employed in the current protocol is outperformed by strategy 3. As long as strategy 3 is used in combination with other strategies the network always yields the highest throughput. The line graphs of the better performing strategies are also smoother indicating stability. This indicates that with learning, nodes always settle for the best strategy.

C. Fairness

The fairness of the individual strategies and a combination of strategies was calculated using Jain's fairness index. The summary is as shown in Fig.12.

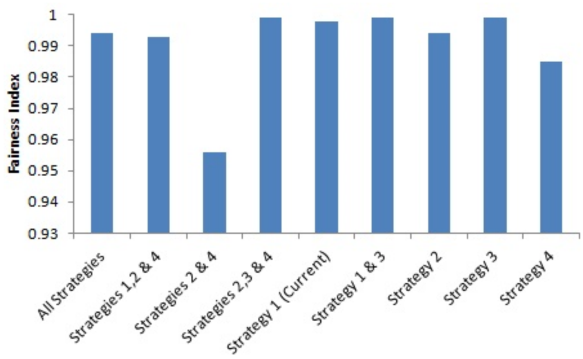

Figure 12. Comparing fairness when different strategies are used

Strategy 3 and all the strategy combinations having it performed better than than all the others including strategy 1.

D. Shorter Contention Window

The cw was made shorter by having a smaller cw_{min} and cw_{max}. The network performance is evaluated using strategy 1, 3 and a combination of all the strategies as shown in Fig.13.

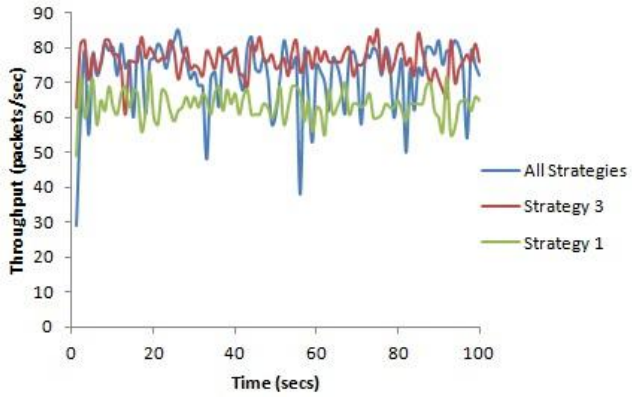

Figure 13. Comparing the throughput of one and 4 strategies when standard and short cw_{min} are used

From the figure, strategy 3 and a combination of all the strategies outperform the standard protocol. But when the combination of strategies is used, the network becomes unstable as displayed by the perturbation of the line graph profile. This happens because the cw values are not optimized. As a result the nodes keep changing strategies to try and get a stable operation point. However, the usage of strategy 3 which is the best performing strategy is more than the others as shown in Fig.14.

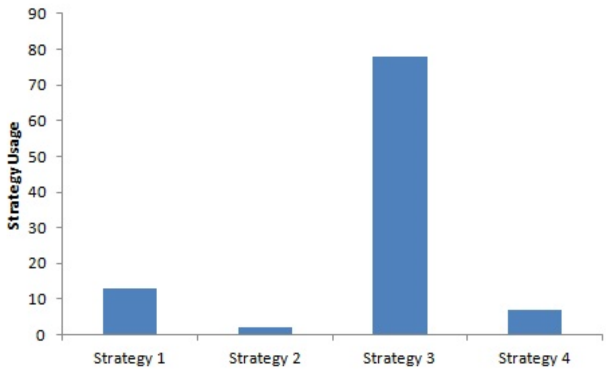

Figure 14. Strategy selection when nodes use 4 strategies and a short cw.

E. Summary of other evaluation Parameters

Other evaluation parameters were also considered. These include the mean network utilization on a scale of 1 and the media access delay in microseconds. Fig. 15 shows the mean network utilization.

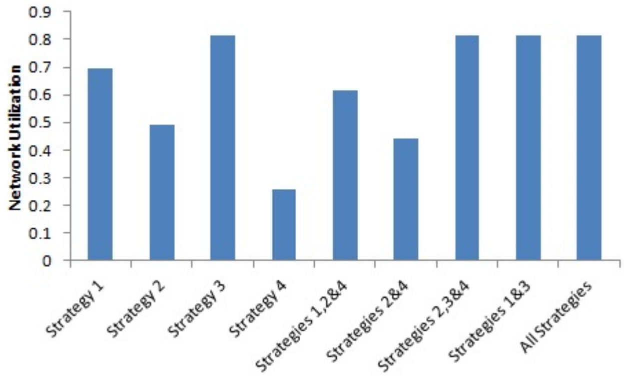

Figure 15. Comparing network utilization

From the figure, one notices that when the best performing strategy is used individually or in combination with other strategies, the network is utilized maximally.

The media access delay is shown in Fig. 16

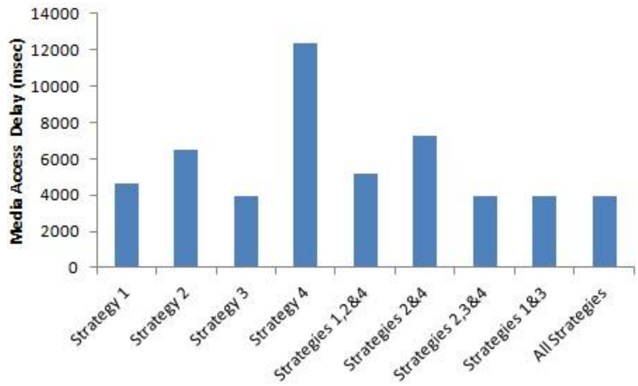

Figure 16. Comparing network media access delay

The best performing strategy or when other strategies are used in combination with it again lead to minimal media access delay. In both cases, this means that the best performing strategy has reduced number of collisions. As a result, nodes with data to send don't have to queue the packets for long periods of time and the retransmission attempts are minimized.

VIII. CONCLUSIONS AND FUTURE WORK

Analysis of the simulation results indicates that the enhanced mechanism outperforms the existing mechanism in terms of throughput, dropped packets and fairness. This is more defined as the network size increases.

- When many strategies are used performance is comparable to the best strategy.
- When a particularly bad strategy is combined with a good strategy, performance improves.
- When the best strategy is removed, performance drops and the network is unstable since the nodes keep trying the other strategies as they seek a global stable operating point.

- When current optimized cw values are changed arbitrarily, the overall system performance drops when compared to the optimized parameters. However the best performing strategy still outperforms the other strategies. If the number of strategies is increased, the performance is better than the currently implemented strategy but the network is unstable. This is also because the nodes keep trying all the other strategies as they seek a stable global operating point.

There is still a lot of work which could be done that comprises future work. This involves improving on the strategies to be adopted and the size of the strategy space. Additionally experiments need to be done to determine the optimal reward scheme and learning rate.

ACKNOWLEDGMENTS

We also thank Simon Ndunda for his help in the coding and useful discussion and insights. We also thank the Computational Modeling (COMO) lab, Free University of Brussels (VUB) for the support they accorded us.

REFERENCES

[1] W. B. Arthur. Complexity in Economic Theory: Inductive Reasoning and Bounded Rationality. The American Economic Review, 84(2):406–411, May 1994.

[2] LAN/MAN Standards Committee. IEEE Standard for Information technology - Telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999), pages C1–1184, 12 2007.

[3] Behrouz A. Forouzan. Data Communications and Networking. McGraw-Hill, 2007.

[4] Garrett Hardin. The Tragedy of the Commons Science, 162(3859):1243–1248, 1968.

[5] Rajendra K. Jain, Dah-Ming W. Chiu, and William R. Hawe. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems. Technical Report DEC-TR-301, Digital Equipment Corporation, 1984.

[6] A.B. MacKenzie and L. DaSilva. Game Theory for Wireless Engineers, volume 1 of Synthesis Lectures on Communications, pages 1–86. Morgan & Claypool Publishers, 2006.

[7] Ramon Marimon. Adaptive learning, evolutionary dynamics and equilibrium selection in games. European Economic Review, 37(23):603– 611, 1993.

[8] Ramon Marimon and Ellen McGrattan. Learning and Rationality in Economics, Chapter On Adaptive Learning in Strategic Games. Basil Blackwells, 1995.

[9] Tom M. Mitchell. Machine Learning. The McGraw-Hill Companies, Inc, New York, 1997.

[10] John Nash. Non-cooperative Games. Annals of Mathematics, 54(2):286–295, 1951.

[11] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature, 364, 1993.

[12] Martin J. Osborne and Ariel Rubinstein. A Course in Game Theory. MIT Press, USA, 1994.

[13] Stuart Russell and Peter Norvig. Artificial Intelligence, A Modern Approach. Prentice-Hall, Inc, New Jersey, 1995.

[14] T. Sandholm and R. Crites. Multiagent Reinforcement Learning in the Iterated Prisoner's Dilemma. Biosystems, Special Issue on the Prisoner's Dilemma, January 1995.

[15] Dimitris El. Vassis and Vassilis Zafeiris. Pamvotis - IEEE 802.11 WLAN Simulator. 2007. http://pamvotis.org/ – accessed on 22/10/2013.

[16] Christopher John Cornish Hellaby Watkins. Learning from delayed rewards. PhD thesis, King's College, London, 1989. Available at http://www.cs.rhul.ac.uk/ chrisw/newthesis.pdf. Accessed on 22/10/2013.