**İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY**

# A Control Plane for Prioritized Real-Time Communications in Wireless Token Ring Networks

**M.Sc. Thesis by**

**Mahmut Nezih YİĞİTBAŞI, B.Sc.**
**(504061520)**

**Date of submission : 5 May 2008**

**Date of defence examination: 11 June 2008**

**Supervisor (Chairman): Assoc. Prof. Feza Buzluca**

**Members of the Examining Committee Prof.Dr. Emre Harmancı**

**Assist.Prof.Selçuk Paker**

**JUNE 2008**

## FOREWORD

First of all I would like to thank my supervisor Assoc. Prof. Feza BUZLUCA for his help, support and invaluable guidance while writing this thesis. This could be impossible without his support.

And also I would like to thank my family for their love and support during my whole life in all the decisions I take.

Nezih YİĞİTBAŞI

May 2008

**TABLE OF CONTENTS**

**FOREWORD** ..... **ii**
**ABBREVIATIONS** ..... **v**
**TABLE LIST** ..... **vi**
**FIGURE LIST** ..... **vii**
**SYMBOL LIST** ..... **viii**
**ÖZET** ..... **ix**
**SUMMARY** ..... **x**

**1. INTRODUCTION** ..... **1**
1.1 Real-Time Communications ..... 2
1.2 Real-Time Communications in Wireless Networks ..... 4
1.2.1 Main Challenges ..... 6
1.2.2 Some Applications Requiring Real-Time Guarantees ..... 7

**2. TOKEN RING PROTOCOLS FOR WIRELESS NETWORKS** ..... **8**
2.1 Wireless Token Ring Protocol (WTRP) ..... 10
2.1.1 WTRP System Architecture ..... 11
2.1.2 WTRP Operation ..... 12
2.1.3 WTRP Finite State Machine ..... 13
2.2 Improved Wireless Token Ring Protocol (iWTRP) ..... 14
2.3 Wireless Dynamic Token Protocol (WDTP) ..... 15
2.4 Rether ..... 17
2.5 A Token Based MAC Protocol for Wireless Industrial Control Networks 17
2.6 High Frequency Token Protocol (HFTP) ..... 18
2.7 Enhanced Wireless Token Ring Protocol (EWTRP) ..... 19
2.8 A Token Passing MAC Protocol for Ad Hoc Networks (T-MAH) ..... 20
2.9 Virtual Token Passing CSMA (VTP-CSMA) ..... 21

**3. SYSTEM ARCHITECTURE** ..... **23**
3.1 The Network and Message Model ..... 24
3.2 Timed Token Protocol Parameters and Operation ..... 26
3.2.1 Timed Token Protocol Parameters ..... 26
3.2.2 Timed Token Protocol Operation ..... 27
3.3 Synchronous Bandwidth Allocation ..... 28

**4. THE PROPOSED CONTROL PLANE** ..... **35**
4.1 The Proposed System Architecture ..... 35
4.2 System Operation ..... 38
4.3 Finite State Machine ..... 47
4.3.1 Init State ..... 47
4.3.2 Idle State ..... 47

4.3.3 Reforming_Ring State ..... 48
4.3.4 Conn_Req State ..... 49
4.3.5 Conn_Accept State ..... 49
4.3.6 Conn_Reject State ..... 49
4.3.7 Disconn_Req State ..... 49
4.3.8 Disconnected State ..... 49
4.4 Sample Use Cases for the Control Plane ..... 50
4.4.1 Wireless Industrial Automation Networks ..... 50
4.4.2 A Military Application ..... 53

**5. SIMULATIONS AND RESULTS ..... 56**
5.1 Simulation Studies ..... 60
5.1.1 First Simulation Study ..... 60
5.1.2 Second Simulation Study ..... 65

**6. CONCLUSIONS ..... 71**
6.1 Suggestions for Improvement and Future Work ..... 72

**REFERENCES ..... 74**

**CURRICULUM VITAE ..... 77**

## ABBREVIATIONS

| | |
|---|---|
| **ITU** | : International Telecommunication Union |
| **QoS** | : Quality of Service |
| **Gbps** | : Giga Bits per Second |
| **WSN** | : Wireless Sensor Network |
| **WMSN** | : Wireless Multimedia Sensor Network |
| **CMOS** | : Complementary Metal Oxide Semiconductor |
| **OSI** | : Open Systems Interconnection |
| **MIB** | : Management Information Base |
| **IEEE** | : Institute of Electrical and Electronics Engineers |
| **TDMA** | : Timed Division Multiplexing |
| **FDMA** | : Frequency Division Multiplexing |
| **CDMA** | : Code Division Multiple Access |
| **CDMA** | : Code Division Multiple Access |
| **CSMA-CD** | : Carrier Sense Multiple Access Collision Detection |
| **CSMA** | : Carrier Sense Multiple Access |
| **MIB** | : Management Information Base |
| **MANET** | : Mobile Adhoc Network |
| **THT** | : Token Holding Time |
| **TRT** | : Token Rotation Time |
| **TTRT** | : Target Token Rotation Time |
| **MTRT** | : Maximum Token Rotation Time |
| **TXOP** | : Transmission Opportunity |
| **FDDI** | : Fiber Distributed Data Interface |
| **EMCA** | : Enhanced Minimum Capacity Allocation |
| **SBA** | : Synchronous Bandwidth Allocation |
| **LC** | : Late Counter |
| **NPA** | : Normalized Proportional Allocation |
| **MCA** | : Minimum Capacity Allocation |
| **BER** | : Bit Error Rate |
| **TC** | : Traffic Class |
| **FSM** | : Finite State Machine |
| **LAN** | : Local Area Network |
| **3G** | : Third Generation Networks |

**TABLE LIST**

**Page Number**

**Table 1.1:** Performance Targets for Different Types of Services ..... 3
**Table 5.1:** Connection Related Statistics For The Proposed Control Plane..... 61
**Table 5.2:** Connection Related Statistics For The Pure Timed Token Protocol ..... 62
**Table 5.3:** Connection Related Statistics For The Proposed Control Plane..... 66
**Table 5.4:** Connection Related Statistics For The Pure Timed Token Protocol ..... 67

**FIGURE LIST**

**Page Number**

**Figure 1.1:** Hard v.s. Soft Real Time Systems ..... 1
**Figure 2.1:** WTRP System Architecture ..... 11
**Figure 2.2:** WTRP Join Operation ..... 12
**Figure 2.3:** WTRP Finite State Machine ..... 14
**Figure 2.4:** TPQ Maintenance ..... 16
**Figure 2.5:** HFTP Ring Merging Mechanism ..... 19
**Figure 3.1:** Absolute and Relative Deadline ..... 25
**Figure 3.2:** A Wireless Token Ring Network ..... 26
**Figure 4.1:** System under Consideration ..... 38
**Figure 4.2:** Connection Request Accepted ..... 39
**Figure 4.3:** Connection Request Rejected ..... 41
**Figure 4.4:** Flowchart for Connection Handling ..... 43
**Figure 4.5:** Disconnect Request Processing ..... 46
**Figure 4.6:** Flowchart for Disconnect Request Handling ..... 47
**Figure 4.7:** Finite State Machine ..... 48
**Figure 4.8:** A Simple Wireless Industrial Automation Network ..... 51
**Figure 4.9:** A Simple Military Network ..... 54
**Figure 5.1:** The Class Diagram of the Simulation Program ..... 59
**Figure 5.2:** Total Network Utilization with the Proposed Control Plane ..... 63
**Figure 5.3:** Total Network Utilization for the Pure Timed Token Protocol ..... 64
**Figure 5.4:** Total Network Utilization ..... 65
**Figure 5.5:** Total Network Utilization with the Proposed Control Plane ..... 68
**Figure 5.6:** Total Network Utilization for the Pure Timed Token Protocol ..... 69
**Figure 5.7:** Total Network Utilization ..... 70
**Figure 6.1:** A Multi-Ring Network Topology ..... 73

**SYMBOL LIST**

| | |
|---|---|
| **TRT** | **:** Token rotation time |
| **R(n)** | **:** Receiving time of the $n^{th}$ token |
| **THT** | **:** Token holding time |
| **MTRT** | **:** Maximum token rotation time |
| $\mathbf{NA_i}$ | **:** $i^{th}$ member of the process group for VTP-CSMA |
| $\mathbf{AC_0}$ | **:** Local counter maintained by each station |
| **TTRT** | **:** Target token rotation time |
| **H** | **:** Synchronous bandwidth allocated to a node |
| $\mathbf{S_i}$ | **:** Synchronous message stream i |
| **C** | **:** Maximum time needed to transmit a message |
| **P** | **:** Period |
| **D** | **:** Deadline |
| **TC** | **:** Traffic class |
| **SA** | **:** Source address |
| **DA** | **:** Destination address |
| **U(M)** | **:** The utilization factor of the synchronous message set M |
| $\mathbf{LC_i}$ | **:** Late counter |
| $\mathbf{X_i}$ | **:** Minimum amount of time available for a node to transmit its synchronous messages in a time interval |
| $\mathbf{LC_i}$ | **:** Late counter |
| $\boldsymbol{\tau}$ | **:** Fraction of TTRT that is unavailable for synchronous message transmission |
| $T_{\square}$ | **:** Time needed to transmit one packet |

# ÖZET

Kablosuz ağlarda gerçek zaman kısıtlarını sağlamak zor bir araştırma problemidir. Jetonlu halka mimarisine sahip ağlar yanıt süreleri deterministik olduğundan ve gecikmelerin üst sınırının tahmin edilebilir olmalarından dolayı gerçek zaman kısıtlarını sağlamak için daha elverişli bir yapıya sahiptir. Bu tezde kablosuz jetonlu halka ağları için katı gerçek zaman kısıtlarını sağlamak üzere MAC katmanında zamanlı jeton protokolünü içeren merkezi bir denetim düzlemi önerilmektedir. Merkezi denetim düzleminin en önemli bileşeni yerel ağda bulunan bir yönetim istasyonudur. Ağa yeni katılmak isteyen istasyonlar ve ağdan ayrılmak isteyen istasyonlar yönetim istasyonu ile haberleşerek isteklerini bildirmektedirler. Bu denetim düzleminde üç tane fonksiyonu gerçekleyen bir algoritma tasarlanmıştır. Bu fonksiyonlar kabul denetimi, istasyon çıkarma ve trafik ayrımı fonksiyonlarıdır. Kabul denetimi fonksiyonu sayesinde yönetim istasyonuna gelen yeni bir bağlantı isteğinin kabul edilip edilmeyeceğine karar verilir. İstasyon çıkarma fonksiyonu sayesinde yeni gelen yüksek öncelikli bir istasyonun isteğinin karşılanması için gerekirse olası en düşük öncelikli istasyonun ağdan çıkarılabilmesi sağlanmaktadır. Son olarak trafik ayrımı fonksiyonu sayesinde ise kablosuz jetonlu halka ağında aynı anda düşük ve yüksek öncelikli trafikler beraber bulunabilmekte ve bu önceliklere ilişkin gerekli servis kalitesi sağlanabilmektedir. Bu algoritma sonucunda dinamik bir halka yapısı oluşturulmuş ve yüksek öncelikli trafiğin ağa girme şansı artmış ve düşük öncelikli trafik taşıyan istasyonların gerektiğinde yüksek öncelikli trafiğe yer vermeleri için ağdan çıkarılmaları sağlanmıştır. Gerçekleştirilen simülasyon sonuçlarına göre önerilen denetim düzlemi sayesinde yüksek öncelikli trafiğin düşük öncelikli trafiğe göre ağda daha fazla bant genişliğine sahip olduğu ve katı gerçek zaman kısıtlarının sağlandığı görülmüştür.

## SUMMARY

Providing real-time guarantees in wireless networks is a challenging research problem. Token ring networks are more suitable for real-time communications due to the fact that the response time is highly deterministic and also the upper bound of the latency in these networks is predictable. This thesis proposes a centralized control plane incorporating the timed token protocol in the MAC layer for providing hard real-time guarantees in wireless token ring networks. In the proposed control plane there exists a management station which takes care of the connection and disconnection requests of the stations that want to join or leave the wireless ring. In this control plane an algorithm that implements three important functions is designed. These functions are the admission control function, the station eviction function and a traffic differentiation mechanism. With the admission control function the management station decides whether a new connection request can be accepted or rejected. With the station eviction function the management station can remove the station with the least possible priority in order to accept a new high priority connection request. And finally with the traffic differentiation mechanism both low and high priority traffic can coexist in the wireless token ring and can get the required QoS levels. In this approach a dynamic ring structure is built where high priority stations have more chance of admittance and stations with low priority can be removed from the ring. Simulation results show that the proposed control plane ensures higher priority traffic more bandwidth than lower priority traffic and guarantees that deadline constraints of hard real-time traffic are satisfied.

## 1. INTRODUCTION

Real-time systems are those in which the correctness of the system does not depend only on the logical correctness of the computation but also on the time at which the results are produced. These two types of correctness are called the logical correctness and the temporal correctness. Logically correct systems produce correct sequence of outputs given the set of inputs and temporally correct systems produce output at the right time. The right time depends on the application and the temporal correctness of a real-time system can be checked by formal verification or timing analysis. [1]

Real-time systems generally need a higher degree of reliability and fault tolerance requirements. The most important issue in real-time systems is predictability. That means the worst case response time of real-time systems must be predictable. In other words predictability and determinism are two important requirements of real-time systems.

Real-time systems are categorized as soft and hard real-time systems. In hard real-time systems a late result may cause disastrous consequences or becomes worthless. However in soft real-time systems timely results are desirable but the usefulness of the results decreases as the response time increases.

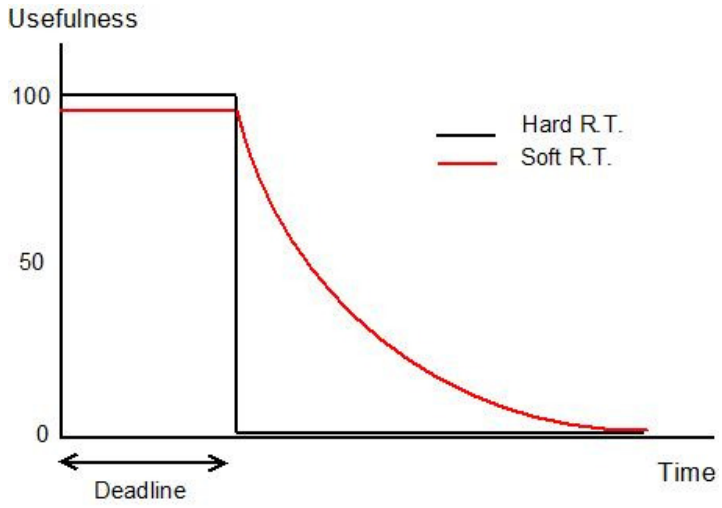


**Figure 1.1:** Hard vs. Soft Real Time System

To summarize, the following properties are desirable for real-time systems:

- *Timeliness*: The results produced by a real-time system must be temporally correct.
- *Predictability*: The behavior of the system should be easy to predict based on current inputs.
- *Testability*: It should be easy to test whether the system can meet all the deadlines.
- *Cost optimality*: The system should use resources optimally and should not waste resources unnecessarily.
- *Fault Tolerance*: The system should handle faulty and exceptional conditions carefully and should not crash in those situations.

## 1.1 Real-Time Communications

With the increase in bandwidth of networks and the processing power of the nodes, the need for real-time communications has become more important. Hence new class of applications with real-time requirements has been developed to be deployed on these powerful network architectures. Especially many kinds of real-time applications have been deployed on industrial control networks, military networks and on sensor networks where the infrastructure is wireless. In addition many real-time applications have also been developed for the internet. The traffic characteristics and requirements such as performance or timing requirements of these applications differ from those of conventional applications. And as a result, design and implementation of new network protocols that should provide the requirements of these real-time applications became necessary. Considering these wide ranges of applications, ITU-T has provided an indication of suitable requirements for different types of services as shown in Table 1-1 [2].

**Table 1.1:** Performance Targets for Different Types of Services

| **Data services** | | | | | |
|---|---|---|---|---|---|
| ***Application*** | ***Degree of symmetry*** | ***Typical data amount*** | ***Key performance parameters and target values*** | | |
| | | | **1-way delay** | **Delay jitter** | **PLR** |
| Web browsing | One-way | ~ 15 KB | Preferred <2 s<br>Acceptable < 4s | N.A. | zero |
| Interactive games | Two-way | < 1 KB | < 200 ms | N.A. | Zero |
| Bulk data retrieval | one-way | 10 KB-10 MB | Preferred < 15 s<br>Acceptable < 60 s | N.A. | Zero |
| **Streaming services** | | | | | |
| Audio streaming | One-way | 128 Kbps | < 10 s | << 1 ms | < 1 % |
| Video on demand | One-way | 480 Kbps | < 10 s | | < 1 % |
| **Interactive services** | | | | | |
| ***Application*** | ***Degree of symmetry*** | ***Typical data rates[3]*** | ***Key performance parameters and target values*** | | |
| | | | **1-way delay** | **Delay jitter** | **PLR[4]** |
| Voice over IP | Two-way | 64 Kbps | < 150 ms preferred<br>< 400 ms limit | < 1 ms | < 1 % |
| Video phone | Two-way | 384 Kbps | < 150 ms preferred<br>< 400 ms limit | | < 1 % |

The aim of real-time communications is to guarantee that the timing constraints of the messages are met which are sent by the nodes in the network. Also there are other objectives like increasing the overall system throughput and minimize the average transmission delay of messages. In real-time communication systems the principal performance consideration is rather the percentage of messages that are delivered within their time constraints [1, 3]. Other key performance metrics for users are [2]:

1. *Delay:* Delay has a direct impact on user QoS experience and includes the transmission delay, queuing delay, propagation delay and other delays in the intermediate and end nodes.
2. *Delay Variation:* This parameter is also called *jitter* and is very important at the transport layer for packetized multimedia data due to inherent variability in arrival times of the packets from the application layer. To reduce the delay variation applications usually implement buffering mechanisms.
3. *Packet Loss:* This parameter has also a direct impact on user QoS experience irrespective of the type of data whether it is voice, video or data. The fraction

of lost packets should be kept in acceptable bounds for better QoS experience.

There are three types of messages used in the networks which are designed for real-time communications. The first type of these messages is the *hard real-time messages.* These messages are critical messages that should be transmitted before their deadlines. In other words these messages need hard real-time guarantees. The second type of messages is the *soft real-time messages.* These messages are also time critical messages that can tolerate some tardiness or loss. In other words these messages need soft real-time guarantees. And the last type of messages is the *non real-time messages.* These messages do not have special timing constraints. It is enough for these messages to be transmitted in a best effort manner.

A real-time communications architecture should provide the necessary timing requirements that are needed by different kinds of traffic and also should be fair in the sense that non real-time messages should not be starved.

The following properties are some desirable properties for real-time communications [4]:

1. Jitter should be low
2. Latency should be low
3. It should be easy to integrate non-real-time and real-time traffic
4. It should be adaptable to dynamically changing network and traffic conditions
5. The performance should stay in acceptable bounds as the size of the network and the number of connections increases
6. The bandwidth utilization should be effective
7. The overhead in the headers of packets or cells should be low
8. The processing overhead per packet within the network or at the end nodes should be low

To provide these properties and necessary requirements for real-time applications, suitable protocols should be designed.

## 1.2 Real-Time Communications in Wireless Networks

Wireless links have a much higher BER (bit error rate) than wired links. As a result wireless networks are usually not the first choice for real-time systems. Due to tremendous research in the field of wireless networking, many new standards and technologies have been proposed in both the academia and the industry. These

technologies and standards made high data rate and more reliable wireless communications possible for both mobile and fixed devices. With these advances diverse communication requirements have emerged such as video or voice communication over wireless links. Quality of service guarantees such as delay, bandwidth, jitter and minimal packet loss are needed for these multimedia applications. And as a result, network architectures to provide these QoS guarantees should be designed and implemented.

The problem of real-time communications should be addressed in the MAC and physical layers in wireless networks. In physical layer suitable modulation or coding techniques should be designed to make providing real-time guarantees easier for upper layers. Actually most of the research to provide real-time guarantees is done for the MAC layer. Especially since the 802.11 standard is the most widely used standard, researches generally focus on providing better QoS guarantees and real-time guarantees for 802.11 networks.

And also another class of protocols has been proposed for real-time communications over wireless networks. These protocols are actually token based protocols. In token ring networks the stations are arranged in a ring topology and a token is circulated around the ring. Upon receiving the token a station can transmit for a predetermined period of time if it has data to send and then passes the token to the next station in the ring or passes the token directly if it has no data to send.

Actually wireless token ring protocols are more suitable for providing real-time guarantees in wireless networks. The most important advantage of token ring networks is that the response time is highly deterministic and also the upper bound of the latency in these networks is predictable. Since every station releases the token after it sends for a specified amount of time, every other station on the ring has a chance to transmit before a station can make its second transmission. And since the time that a station can hold the token is known, then a station in the ring can predict the worst case arrival time of the token. As stated at the beginning of this section both predictability and determinism are the most important properties of real-time systems as a result token ring architectures make wireless networks more suitable for real-time communications.

Another important advantage of wireless token ring networks is that the collision problem is solved since each station transmits only when it holds the token. So the penalty of retransmissions caused by collisions is not an issue in wireless token ring networks.

Another important advantage of token ring networks is fairness. In token ring networks the token is usually passed linearly such that when station *i* holds the token it passes the token to (*i*+1) mod N where N is the number of stations in the network. So this makes token ring networks fairer.

### 1.2.1 Main Challenges

Providing real-time guarantees in wireless networks is not as easy as to provide these guarantees in wired networks. Actually this is because of two major differences between these networks. These are the *link characteristics* and *mobility* [5]. Generally wired networks have higher data rates in the order of Gbps (Giga Bits Per Second) and very low error rates in the order of $10^{-4}$ - $10^{-9}$. And also the propagation characteristics of wireless links are time-varying and interference is an important issue. Also the bandwidth of wireless links is limited and lower than wired links.

Another important property that makes providing real-time guarantees challenging in wireless networks is mobility. Since the nodes in wireless networks can be mobile, it is necessary to redesign or revise the infrastructure for wireless networks for QoS support. Also the quality of the link and connection should be kept consistent throughout the movement of a node. Another issue with mobility is the hand-off problem. The hand-off should be seamless and the quality of the negotiated connection should not be degraded.

In addition to these issues, with the emergence of wireless sensor networks, the issue of power saving has also been important. This power awareness is another issue that makes providing real-time guarantees challenging. In sensor networks, the nodes also have limited processing capabilities which makes the problem more challenging.

Another issue that makes providing real-time guarantees in wireless networks challenging can be heterogeneity of the types of traffic carried in the same network. For example a station in a wireless network can carry high priority traffic like voice and another station can carry best-effort traffic like e-mail data. The problem here is to provide the required guarantees to both stations. Neither of them should be starved. So a mechanism for traffic differentiation should be adopted for heterogeneous wireless networks.

The last issue that makes the problem more challenging is the problem of contentions that can occur while accessing the shared wireless medium. This contention results in a large amount of collisions and retransmissions, which can result in more packet loss and as a result of retransmissions the end to end delay will be higher. It is

usually desirable that high priority traffic should seize the wireless medium in the case of contentions.

### 1.2.2 Some Applications Requiring Real-Time Guarantees

Wireless networks are especially being used in industrial automation applications possibly in the areas where it is either difficult or expensive to wire. Usually wireless networks are extensions to the Ethernet segments in these networks. Also it became possible for devices to be mobile since there are no physical restrictions like wires. As a result of these, it can be said that usage of wireless networks will gain popularity in the future.

In these networks generally the carried data is a vital control or monitoring information. And much of the status or control information is carried in short bursts which generally require relatively little bandwidth and connection speed and also the traffic is usually periodic. Or sometimes large data like log files or other event logs can be carried over the medium. The key requirement for communications in these networks is the timely delivery without failure.

And also many wireless military networks have been proposed like wireless sensor networks or experimental cellular and short range 3G networks. What makes wireless communications in military applications attractive is that there are no physical restrictions like wires, technological advances can provide long range communication capabilities and also these wireless systems provide mobility. These wireless military networks are gaining popularity and probably will be of greater importance in the future. As a result number of wireless military network deployments will increase.

In wireless military networks the main concerns are actually the real-time guarantees of the stations in the network and secure communications. Especially due to the increase in the data rates of wireless communications, new applications are being considered. For example a robot in a battlefield can take photos of an enemy target and can transmit this information to a center to be processed for further action. These multimedia applications are being used today and will also be used extensively in the future. To provide the quality of service needed by the stations in the battlefield specialized real-time command and control software and quality of service aware network protocols should be designed.

Moreover in recent years there has been tremendous research in wireless sensor networking. WSNs (Wireless Sensor Network) are being deployed for various applications like habitat monitoring, object tracking, nuclear reactor controlling, fire

detection and traffic monitoring. With the availability of low-cost small-scale imaging sensors, CMOS cameras, microphones, which may capture multimedia content from the field, Wireless Multimedia Sensor Networks (WMSN) have been proposed and drawn the immediate attention of the researchers. WMSN applications, e.g., multimedia surveillance networks, target tracking, environmental monitoring, and traffic management systems, require effective communication of event features in the form of multimedia such as audio, image, and video. To this end, additional challenges for energy-efficient multimedia processing and communication in WMSN, i.e., heterogeneous multimedia reliability definitions, tight QoS expectations, and high bandwidth demands, must be addressed as well. So providing real-time guarantees in wireless multimedia sensor networks is a new challenge for researchers [6].

## 2. TOKEN RING PROTOCOLS FOR WIRELESS NETWORKS

Due to tremendous research in the field of wireless local area networking many new standards and technologies have been proposed in both the academia and the industry. These technologies and standards made high data rate and more reliable wireless communications possible for both mobile and fixed devices. With these advances diverse communication requirements have emerged such as video or voice communication over wireless links. Quality of service guarantees such as delay, bandwidth, jitter and minimal packet loss are needed for these multimedia applications. To meet these requirements quality of service aware protocols for wireless local area networks must be developed.

Enabling QoS in wireless local area networks requires QoS aware protocols in the data link layer and the physical layer of the OSI model since the wireless local area networks are actually defined in the physical and data link layer of the OSI model. According to the IEEE 802 family of standards the data link layer is decomposed into two sub layers namely the logic link control sub layer and MAC (medium access control) sub layer. The function of the MAC sub layer is to coordinate access to the wireless medium which is actually a broadcast medium to which any station can transmit and listen in their ranges. Its primary function is to minimize the collisions and achieve a reasonable utilization. Also MAC sub layer must support fair and prioritized access to the medium. QoS is generally addressed in the network layer focusing on routing protocols. But in the wireless domain it must also be addressed in the data link layer.

In order to understand the various wireless token ring networks it will be helpful to review the MAC techniques. MAC techniques can be classified as static channelization (TDMA, FDMA and CDMA) or dynamic medium access control. Dynamic medium access control is also classified as scheduled access (reservation and polling systems) and random access (ALOHA, Slotted ALOHA, CSMA and CSMA-CD). In reservation systems each cycle begins with a reservation period and each station makes a reservation request for accessing the medium. In polling systems each station in the network take turns to access the medium. So at any given time only one station can access the medium. Token ring medium access is actually an extension to the polling approach for networks with ring topology. Actually token ring access is a distributed technique for polling. The stations are arranged in a ring

topology and a token is circulated around the ring. Upon receiving the ring the station can transmit for a predetermined period of time if it has data and then passes the token to the next station in the ring or passes the token directly if it has no data to send.

Actually wireless token ring protocols are more suitable for providing real-time guarantees in wireless networks. The most important advantage of token ring networks is that the response time is highly deterministic and also the upper bound of the latency is predictable. Since every station releases the token after it sends for the token holding time, every other station on the ring has a chance to transmit before a station can transmit a second frame. And since the time that a station can hold the token is known, then a station in the ring can predict the worst case arrival time of the token. As stated in the first section determinism and predictability are two important properties of real-time systems, token ring architectures make wireless networks more suitable for real-time communications.

Another important advantage of wireless token ring networks is that the collision problem is solved since each station transmits only when it holds the token. So the penalty of retransmissions caused by collisions is not an issue in wireless token ring networks.

Another important advantage of token ring networks is fairness. In token ring networks the token is usually passed linearly such that when station $i$ holds the token it passes the token to $(i+1) \bmod N$ where N is the number of stations in the network. So this makes token ring networks fairer.

Unfortunately none of the proposed wireless token ring protocols are claimed to be appropriate for hard real-time communications; they are just trying to provide soft real-time guarantees or they just claim to guarantee an acceptable level of QoS. And they also do not incorporate a control plane or focus on the management aspect that makes the protocol more suitable for real-time communications.

## 2.1 Wireless Token Ring Protocol (WTRP)

"WTRP (Wireless Token Ring Protocol) is a medium-access-control (MAC) protocol for applications running on wireless ad-hoc networks that provide quality of service." WTRP provides QoS guarantees in terms of bounded latency and reserved bandwidth which are both crucial for real-time applications. WTRP is a distributed protocol which supports dynamic network topologies. It is robust against single node failures [7]. The WTRP system architecture is given in Figure 2.1:

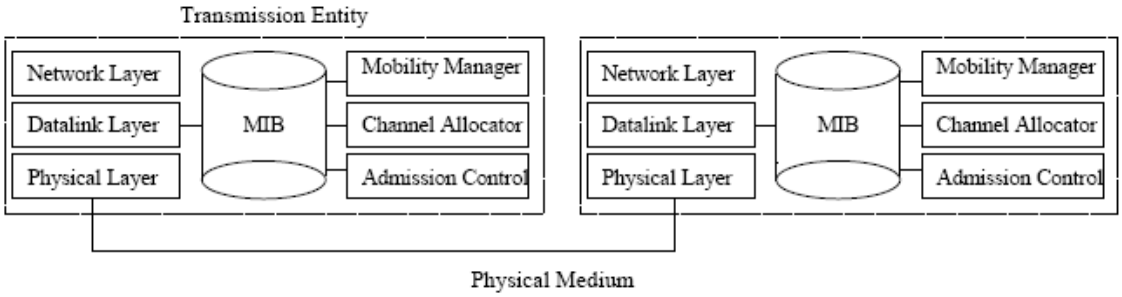


**Figure 2.1:** WTRP System Architecture [7]

### 2.1.1 WTRP System Architecture

#### 2.1.1.1 Medium Access Control

In WTRP the MAC layer performs the ring management and the timing of the transmissions. The ring management functionality includes ensuring the uniqueness of the ring address, ensuring the uniqueness of the token in the ring, ensuring that the rings are proper and management of the joining and leaving operations.

#### 2.1.1.2 Channel Allocator

In WTRP the channel allocator module is local to each station and chooses the channel on which the station can transmit. This module can access the network topology information from the MIB.

#### 2.1.1.3 Mobility Manager

This module is responsible for determining when the station must leave one ring and join another ring a.k.a the hand-off time.

#### 2.1.1.4 Admission Control

In WTRP there is an admission control manager on each ring which is responsible for ensuring that a QoS level is maintained in terms of bounded latency and reserved bandwidth. This module also determines whether a new station can be accepted to join this ring without violating the QoS guarantees of the existing stations.

#### 2.1.1.5 Policer

In WTRP actually the policer is the token holding timer which determines the maximum available time a station can hold the token.

#### 2.1.1.6 Management Information Base (MIB)

This module is responsible for holding the information necessary for the management module to perform the management of the protocol.

### 2.1.2 WTRP Operation

In WTRP each station has a connectivity table which holds the station's view of the network topology. The predecessor and the successor of a station determine the transmission order. The stations in the ring rely on an implicit acknowledgment mechanism to monitor successful transmission. An implicit acknowledgement is any packet heard after token transmission that has the same ring address as the station. Upon transmission the station starts its *idle_timer* and waits for an implicit acknowledgement. If the timer expires before receiving an acknowledgement than the packet transmitted is assumed as lost and retransmitted.

WTRP also supports flexible topologies by allowing the nodes to join and leave the ring dynamically. The admission control manager module is responsible for dynamic joining and leaving operations. Suppose that the admission control manager on station STA1 broadcasts *solicit_successor* token whenever there are enough resources in the ring. Upon hearing this *solicit_successor* another station STA2 which wants to join the ring transmits a *set_successor* token. After STA1 receives the *set_successor* token it transmits a *set_predecessor* token to STA2 and STA2 transmits this *set_predecessor* token to the old successor of STA1 namely STA3. This operation concludes the joining of STA2 to the ring.

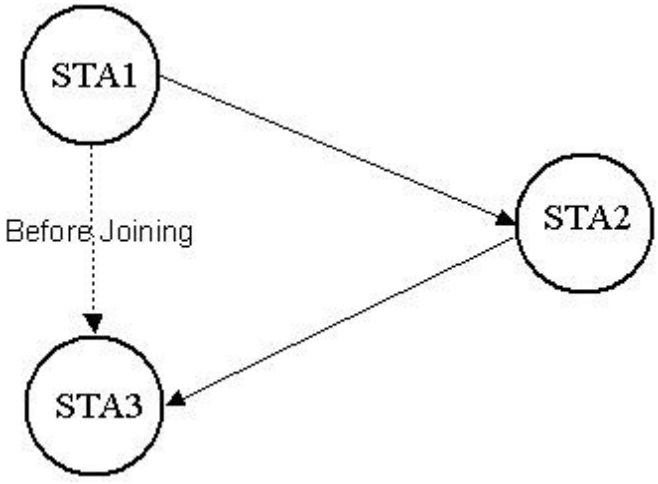


**Figure 2.2:** WTRP Join Operation

Figure 2.2 summarizes the join operation of the WTRP. The dotted arrow show the logical link between the STA1 and STA3 before the join operation is completed. The normal arrows are the logical links formed after the station STA2 is joined to the

ring. At the end of the join operation the new successor of STA1 is STA2 and the new predecessor of STA3 is STA2. A similar procedure is performed by a station that wants to leave the ring.

### 2.1.3 WTRP Finite State Machine

The finite state machine of WTRP is given in Figure 2.3. The states are summarized below:

- *Beginning*: The protocol initialization state. The machine goes directly to the floating state after initialization.
- *Floating:* The station resets its parameters and waits to join a ring in this state. If the station detects a ring in this state, it waits for an invitation. If an invitation is not received in a predetermined period then the station goes to *idle* state and forms a self ring. Self ring is defined for a station whose predecessor and successor are itself.
- *Offline:* A station goes to this state if it does not belong to a self ring or it detects a new ring around or it fails to pass the token to its successor. In this state the station clears its queues and sets its offline timer. The station does nothing until the offline timer expires.
- *Joining:* Upon receiving a *solicit_successor* and sending a *set_successor* message, the station goes into *joining* state. If the station received set predecessor message, then the join is successful and the station goes into the *monitoring* state.
- *Soliciting:* In this state the station initializes its *solicit_wait_timer*. If a *set_successor* message is received in this state then there is another station responding to this station's *solicit_successor* message. And so the station sends a *set_predecessor* message to complete the join operation and goes to the *monitoring* state. If no response is received and the *solicit_wait_timer* expires then the station goes to the *idle* state if it is a self ring or to the *monitoring* state if it is not a self ring.
- *Idle:* In this state when the station passes the token to its successor it makes a transition to the monitoring state in order to receive the implicit acknowledgement. If the station receives a token with a higher priority and the packet queues are not empty, then the station goes to *have token* state.
- *Monitoring:* In this state the station monitors the wireless medium to see whether the transmission was successful or not.

- *Have Token:* In this state the station initializes its token holding timer to maximum token holding time. To make a transition to this state from the *idle* state, there must be some data in the packet queues of the station. The station makes a transition to the *monitoring* state after the token holding timer expires or there is no packet left for sending.

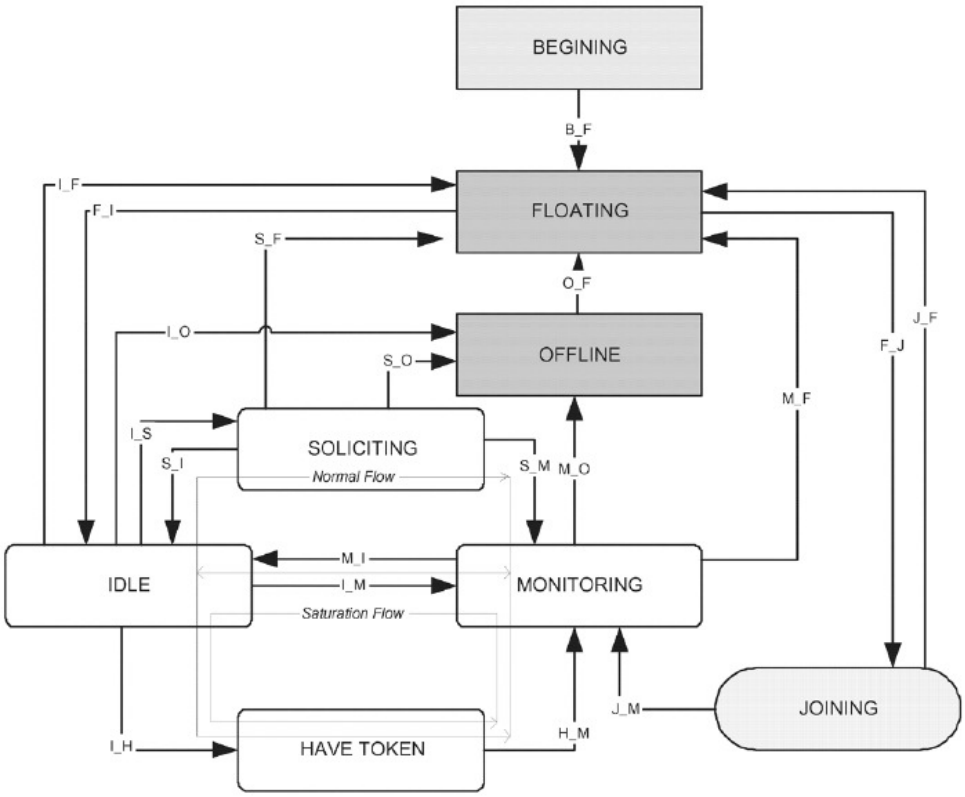


**Figure 2.3:** WTRP Finite State Machine [7]

## 2.2 Improved Wireless Token Ring Protocol (iWTRP)

IWTRP provides a novel method to allow multiple tokens and as a result it allows simultaneous transmissions. IWTRP is designed for wireless metropolitan area networks (WMANs). "IWTRP employs both the request to send/clear to send (RTS/CTS) handshaking and the network allocation vector (NAV) updating techniques to resolve collision resolution and token-elimination problems resulted from multiple tokens" [10].

Some assumptions are made for the WMAN. First assumption is that each station in the ring has two radio interfaces one of which is for clockwise and the other is counterclockwise transmission. The second assumption is that the stations are not mobile. But if there are mobile nodes the join/leave procedure of WTRP can be used.

IWTRP uses a token re-generation idea to generate multiple tokens. But this multi token operation can cause more collisions in the network. To prevent these collisions iWTRP uses request to send/clear to send (RTS/CTS) handshaking and the network allocation vector (NAV) updating techniques. Actually iWTRP is a hybrid solution based on WTRP and 802.11.

To enable token re-generation a new frame is introduced to WTRP namely the Generating Token Sequence (GTS) frame. Stations can use GTS frame to generate a new token and as a result enable parallel transmissions in the token ring network. As a result of adopting the spatial reuse and the token re-generation idea, iWTRP enhances the total throughput of the network.

### 2.3 Wireless Dynamic Token Protocol (WDTP)

WDTP is a MAC protocol for mobile adhoc networks (MANET) which is actually based on WTRP. “In WDTP, the dynamic token transfer algorithm, which is similar to the traversal of graph by depth first search, is proposed to substitute the method of token transfer in WTRP to control the token transfer. With this proposed algorithm, the path that the token passing through can be adjusted automatically corresponding to the subnet’s dynamic topology and has not to be a ring.” [15].

In WDTP all stations are clustered in the network as subnets and transmitting on different channels. Actually these subnets correspond to the logical token rings in WTRP. In each subnet there is an owner which is responsible for token maintenance. In order to determine its successor, a station uses the network topology information and some history that their neighbor stations held the token.

To make the path of the token traversal shortest, WDTP proposes the following algorithm:

- Each station in the subnet maintains a Token Passing Queue (TPQ) to hold its neighbors in this subnet in order.
- After a station has done with the token it passes the token to the station which is the head of the TPQ.
- The stations that do not have the token listen to the channel. If they discover that any station has processed the token transfer then they push this station to the rear of their TPQs. And if this node is already present in TPQ then this record is removed from the TPQ.

As a result of this algorithm the head of the TPQ is actually the station which has not hold the token for longest time. The order of token passing is the order of the nodes in the TPQ.

The maintenance of the TPQ is given in the Figure 2.4.

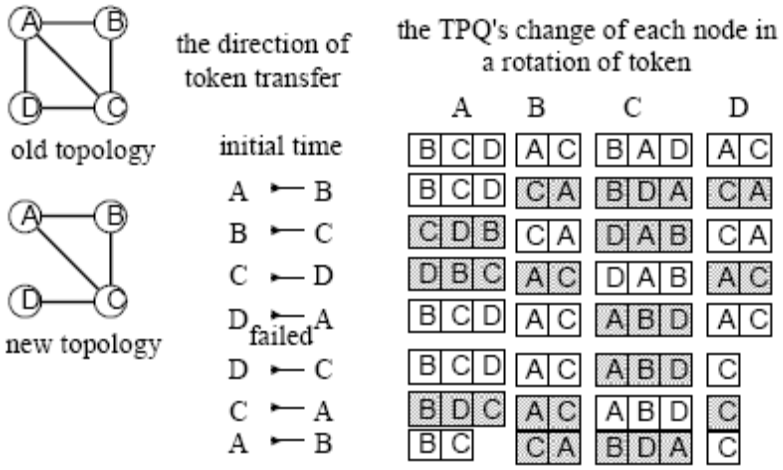


**Figure 2.4:** TPQ Maintenance [15]

Initially the TPQs of the stations consist of their neighbors. First of all station A looks at its TPQ and determine that station B is the head of TPQ. So station A passes the token to station B. The stations B, C and D hear this transmission and according to the algorithm they push the station A to the rear in their TPQs. As a result their TPQs are shown as shaded in the Figure. Similarly after station B determines that the head of its TPQ is station C, it passes the token to station C. All stations in the subnet hear this transmission and push station B to the rear of their TPQs. After an interval when station D has the token it looks at its TPQ and determines that station A is the head of its TPQ, so it transmits the token to station A. But when the transmission fails, station D discovers that the link between it and station A is broken so it deletes station A from its TPQ and the new head of its TPQ become station C. As a result station D passes the token to station C. And other stations hearing this transmission pushes station D to rear in their TPQs. When station A holds the token again it discovers that station D has never transferred the token in the last traversal and finally it removes station D from its TPQ.

By the proposed algorithms, the WDTP increases the channel efficiency and the token path maintenance is simplified. As a result WDTP improves the adaptability to network topology and increases the throughput.

## 2.4 Rether

Rether is a QoS mechanism that provides bandwidth guarantees to individual flows over 802.11 networks. Rether does not make any changes to the MAC layer. It is implemented on top of the data link layer and below the network layer.

Rether has client-server architecture. There is a server namely the Wireless Rether Server (WRS) which is responsible for admission control and coordination. The WRS is located near the access point. Rether uses a different token passing mechanism than other token passing approaches. Rether uses a centralized token passing method in which a center is responsible for token passing and maintenance. This central node is WRS. Rether makes a distinction between real-time and non-real-time stations. In a cycle the token visits the stations that make bandwidth reservations (real-time stations) first and then the non-real-time stations are visited. If all stations cannot be visited in this single cycle, it continues to visit non-real-time stations in the next cycle from where it leaves off in the previous cycle.

Rether uses an implicit bandwidth reservation approach which is based on port signaling. This decision is made to be compatible with legacy applications which cannot make explicit reservation requests. Upon intercepting a flow the WRC looks its policy file and determines the reservation for this flow and makes a reservation request to WRS. With this application Rether is able to provide bandwidth guarantees for both new and legacy applications in a simple and effective manner.

Rether also supports mobility. The WRS periodically broadcasts beacons on the network and new hosts can register with this WRS with register messages. When a station moves from one subnet to another one, its network card detects the new access point and performs a hand-off to the new subnet [12].

## 2.5 A Token Based MAC Protocol for Wireless Industrial Control Networks

The improvements in wireless networking made it possible to construct Wireless Industrial Control Networks (WICN). WICN is a single hop adhoc network with the following properties [13]:

- Frame lengths are shorter.
- Multi path fading is observed.
- The network has a small radius and so mobility is limited in a small region.
- The network should be integrated with upper layers in the intranet.

In order to meet the requirements of the WICN, the protocol should provide QoS guarantees in terms of bounded delay and high reliability. Since the traditional

wireless MAC protocols cannot meet the requirements of WICN, a new MAC protocol based on a single token control is proposed.

In the WICN the station which is the gateway the intranet is selected as the host/master station and the other stations are referred as slave stations. The master is responsible for forming the ring. Each station in the ring has a unique address and each station knows its predecessor and successor. After the master establishes the logical ring it builds a connectivity table which reflects the topology of the current ring. The transmission is deferred until the station gets the token. And after a predetermined time the station must pass the token to its successor.

The new MAC protocol also supports dynamic topologies by the join/leave procedures of the WTRP. This protocol is also robust against single node failures and recovers gracefully from multiple simultaneous failures like WTRP. It is appropriate for small industrial control networks where there are strict timing requirements. And the research shows that the proposed protocol performs better than CSMA/CA in terms of end to end delay and throughput.

### 2.6 High Frequency Token Protocol (HFTP)

HFTP is designed for surface-wave high frequency radio communications. HFTP is based on WTRP but supports two new operations one of which is token relaying and the other is ring merging. The token relay mechanism is used whenever there is a problem in the wireless link between two stations. After a transmission fails the station broadcasts a SOLICIT_RELAY message. Upon receiving this message, the receiving stations' responds carry a flag indicating whether they have sensed a transmission from the destination station in the SOLICIT_RELAY request. This relay information is kept at the source of SOLICIT_RELAY request until it overhears a packet from its successor. From then on the station sends RELAY_TOKEN (destination) messages to the relay node in order to reach the destination [14].

Another new mechanism introduced in HFTP is the ring merge operation. This mechanism is shown in Figure 2.4. This operation can be used for partitioned networks which regains its connectivity. To enable this new operation new states and messages are introduced into WTRP. Whenever a station hears a token with a higher priority from a foreign ring (Station F) it enters into "Want To Merge" state. And sends a MERGE_RINGS message to the station from which it heard the transmission (Station A). The MERGE_RINGS message carries the ID of the successor (Station D) of the station that sent the MERGE_RINGS message, station F in the figure.

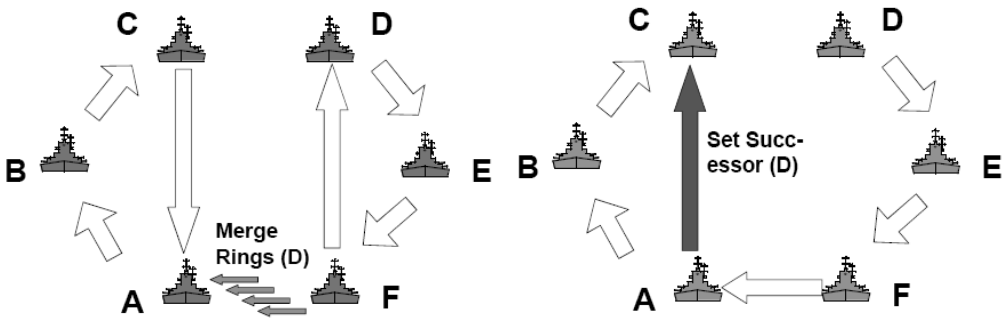


**Figure 2.5:** HFTP Ring Merging Mechanism [14]

These two mechanisms try to solve the connectivity problems as soon as they are detected.

## 2.7 Enhanced Wireless Token Ring Protocol (EWTRP)

EWTRP is designed to meet the QoS requirements of small scale adhoc networks. It introduces three mechanisms into WTRP which are the preemption mechanism, hibernation mechanism and contention mechanism [17].

- *Preemption Mechanism:* In WTRP stations have the opportunity to transmit for THT (Token Holding Time). EWTRP dynamically adjusts this THT value according to the network load possibly at each cycle. Whenever a station receives the token it calculates TRT (Token Rotation Time) with the formula

$$TRT = R(n) - R(n-1) \quad \textbf{(2.1)}$$

In this formula R(n) is the receiving time of the nth token. If the calculated TRT is less than the MTRT (Maximum Token Rotation Time) by a threshold δ and if the node has packets to send in its buffers after THT then it updates its THT according to the formula

$$THT = THT + T_{\square} \quad \textbf{(2.2)}$$

where $T_{\square}$ is the time needed to transmit one packet. This approach actually resembles the additive increase in TCP.

After the station receives the token, if it transmits for a period of time shorter then THT / 2 when its buffers become empty or its TRT is greater than MTRT then the station adjusts its THT according to the formula

$$THT = \max(THT / 2, T_{\square}) \quad \textbf{(2.3)}$$

This approach actually resembles the multiplicative decrease in TCP.

- *Hibernation Mechanism:* This mechanism allows stations to save power. When a station estimates that its buffers will be empty for some time then this station notifies its predecessor how long it will hibernate. By this notification the predecessor node updates its tables and bypasses the hibernated node until its hibernation period ends. This results in power consumption for hibernated nodes.
- *Contention Mechanism:* EWTRP introduces contention periods to WTRP. A station using EWTRP may choose not to join to a network if the traffic load of the station is relatively low. The owner of the ring determines the contention period whenever the TRT is less than MTRT - ɣ. This contention mechanism works the same as the contention period of the 802.11.

With the introduction of these new mechanisms EWTRP produces higher throughput and consumes less power and as a result is more suitable to operate in small scale wireless ad hoc networks.

## 2.8 A Token Passing MAC Protocol for Ad Hoc Networks (T-MAH)

T-MAH is a distributed MAC protocol for multi-hop networks. The stations are organized in clusters with each cluster having a leader. Clusters are also called token groups in T-MAH. By hierarchically clustering stations the number of collisions reduces and as a result the network resources are utilized more effectively. Each token group has a token group identifier. Stations can be part of different token groups at the same time. It is also possible for new stations to enter a token group [18].

T-MAH assumes that the nodes in the network have the same networking capabilities and each node can transmit an out of band Reception Busy Tone signal that is turned on when the node begins receiving from its radio interface. Upon collision event two different transmitters are detected and the receiver turns its Reception Busy Tone off and the transmission is stopped immediately.

A T-MAH network consists of clusters called token groups. Each token group has a unique identifier and each node can be a member of different token groups simultaneously. Any node in the token group knows the address of its predecessor and successor.

The T-MAH network operation has different phases [18].

- ***Normal Operating***: Nodes only transmit user data in this phase
- ***Token Group Head Election***: Token groups are identified in this phase and for each of the token groups a token group head is elected and a transmission frame is defined.
- ***Node Death Detection and Recovery***: In this phase the token group adapts both itself and the transmission frame to the death of this group's some member node.
- ***New Node Entrance***: In this phase the token group adapts both itself and the transmission frame to the join of a new node to this group.

And as a result with the combination of a controlled MAC protocol and a clustering scheme, the number of errors caused by packet collisions is reduced and the channel is utilized more effectively.

## 2.9 Virtual Token Passing CSMA (VTP-CSMA)

The VTP-CSMA architecture tries to solve the problem of heterogeneous devices sharing the same wireless medium in fully meshed communication scenarios. It is based on a virtual token passing procedure that circulates a token around the real-time devices. In addition to this, a traffic separation mechanism is incorporated in this architecture in order to prioritize the real-time traffic over non-real-time traffic [19].

The proposed VTP-CSMA architecture is based on a forcing collision resolution mechanism which is able to prioritize real-time traffic over non-real-time traffic by controlling only the real-time traffic. The VTP procedure sees the network as a process group G with *np* members. The membership is represented as

$$L = \{NA_1, NA_2 \ldots NA_{np}\}$$

Where $NA_i$ denotes the $i^{th}$ member of the group G and used as the identifier of the station *i*. Each station in the group maintains a local counter $AC_0$. This local counter simulates the actual token passing hence the name "Virtual Token". A station is said to capture the token when the local counter value equals the station identifier, mathematically

$$NA_i = AC_0 \qquad \textbf{(2.4)}$$

If the station has data to send whenever it captures the token then it is allowed to transmit for a period of time which is defined by the transmission opportunity period (TXOP).

In the VTP-CSMA architecture an enhanced ring management procedure is also proposed allowing the architecture to be an open group such that a station can dynamically join or leave the group.

The simulation studies also verify that the average packet delay and the average queue size are nearly constant irrespective of the network load. This actually means that the non-real-time traffic has negligible impact on the timing requirements on the real-time traffic. And also with the help of the enhanced ring management procedures dynamic communication scenarios are also supported.

## 3. SYSTEM ARCHITECTURE

As stated in the previous sections to provide real-time communications capabilities, suitable protocols must be designed and implemented. In the proposed system the timed token protocol is used as the MAC (Medium Access Control) sub layer protocol since this protocol is incorporated into several network standards such as FDDI and SAFENET due to its special timing properties. And for efficient operation of the timed token protocol, the EMCA (Enhanced Minimum Capacity Allocation) algorithm is used to allocate synchronous bandwidth.

The timed token protocol was first proposed by Grow in 1982 [20]. The timed token protocol is able to provide hard real-time communication capabilities because of its special timing properties. This protocol was studied with only two message classes namely the synchronous and asynchronous messages.

What makes the timed token protocol suitable for hard real-time communications is that the upper bound for the delay is known and hence the system becomes predictable which is the primary requirement for real-time systems. Sevcik and Johnson have proved that the maximum token rotation time in FDDI is 2.TTRT [21,22]. Later Chen and Zhao generalized this theorem and proved that the maximum time that can elapse between ν consecutive token arrivals at a node *i* is bounded by

$$(v-1).TTRT + \sum_{h=1}^{n} H_h - H_i + a \qquad \textbf{(3.1)}$$

where H is the synchronous bandwidth allocated to the node and a is the fraction of TTRT that is unavailable for synchronous message transmission [23,24].

Briefly, at the initialization phase of the network, the stations negotiate a value called target token rotation time (TTRT) which is the time the token is expected to rotate the ring. This TTRT value must be small enough to satisfy the most stringent timing requirements in the network. And also each station is assigned a $H_i$ value which is a fraction of the TTRT value and determines the maximum time a station is allowed to transmit its synchronous messages every time it receives the token. This $H_i$ value must be allocated such that it must be adequate for a station to transmit its synchronous messages before their deadlines. And also allocating too much synchronous bandwidth to a station may cause other stations to miss their message

deadlines. So the synchronous bandwidth allocation problem is actually a difficult optimization problem.

In this section first the network and message model assumed by the timed token protocol is presented. Then the timed token protocol operation and protocol parameters are reviewed. After reviewing the basics, important timing properties and mathematical relations are analyzed and a challenging problem namely the synchronous bandwidth allocation problem is presented. Finally the EMCA (Enhanced Minimum Capacity Allocation) bandwidth allocation algorithm is presented which is used in this thesis.

### 3.1 The Network and Message Model

The messages generated in the network are classified as synchronous and asynchronous messages. There are *n* streams of synchronous messages at a certain moment

$$S = \{S_1, S_2, ..., S_n\}$$

where stream $S_i$ originates at node *i*. Also each synchronous message stream $S_i$ can be characterized as

$$S_i = (C_i, P_i, D_i, T_i)$$

where

- $C_i$ is the maximum amount of time required to transmit a message in the stream.
- $P_i$ is the interarrival period between messages in the stream.
- $D_i$ is the relative deadline of messages in the stream, that is, the maximum amount of time that can elapse between a message arrival and completion of its transmission.
- $T_i$ is the traffic class (priority) of the stream.

Each synchronous message stream places a certain load on the system. We define the *effective utilization*, $U_i$ of stream $S_i$ as follows:

$$U_i = \frac{C_i}{\min(P_i, D_i)} \qquad \textbf{(3.2)}$$

Each station can transmit its synchronous messages as much as the synchronous bandwidth allocated to it namely $H_i$. To ensure stable operation of the timed token protocol, the total bandwidth allocated to synchronous messages must be less than the available network bandwidth. This protocol constraint is

$$\sum_{i=1}^{n} H_i \leq TTRT - \tau \quad \textbf{(3.3)}$$

The relative deadline is an interval of time that starts at the arrival of the message and ends at the absolute deadline. The notion of the difference between absolute and the relative deadline is given in Figure 3.1.

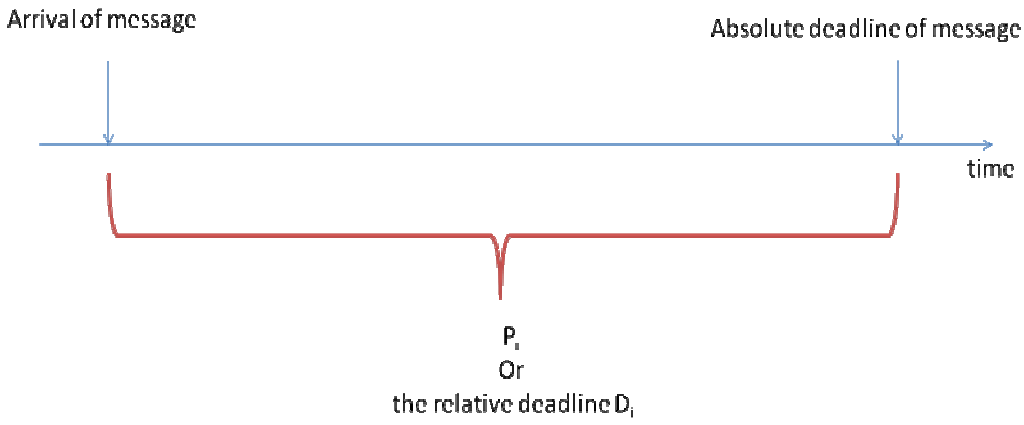


**Figure 3.1:** Absolute and Relative Deadline

The length of each message in the stream is $C_i$ which is the maximum amount of time needed to transmit a message in the stream.

The utilization factor of the synchronous message set M is the fraction of time the network is used to transmit the synchronous messages and denoted as U(M).

$$U(M) = \sum_{i=1}^{m} \frac{C_i}{\min(P_i, D_i)} \quad \textbf{(3.4)}$$

Asynchronous messages are non-periodic as opposed to the synchronous messages. And also the asynchronous messages do not have real-time constraints like synchronous messages.

The network is a token ring network consisting of *m* stations. A special frame called token is circulating around the ring. This token determines which station should use the shared medium. The token may circulate around the ring in the order 1, 2 … m, 1, 2 … m which is actually the case for FDDI. But in the wireless medium since there are no physical restrictions, the token may circulate in any order. Consider a wireless token ring network consisting of 5 stations given in Figure 3.2 and suppose that they are all in the range of each other. Considering the wireless links in the given example, this network is a fully connected mesh where any station can reach any other station.

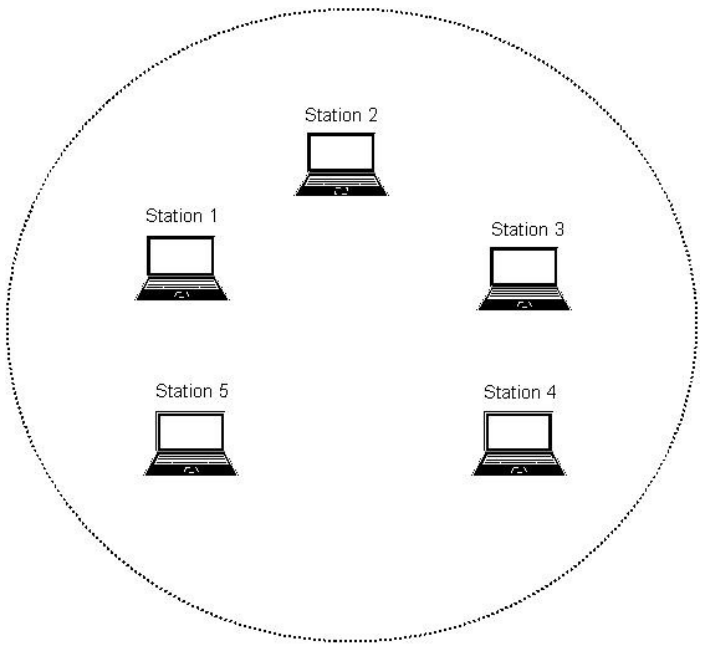


**Figure 3.2:** A Wireless Token Ring Network

In this network the token may circulate in the linear order such as 1,2,3,4,5,1… But since there are no physical restrictions in the wireless medium, the token may circulate in any order like 1,3,2,5,4,1,2,4,3,5… This arbitrary token passing order can be determined by using some metrics like synchronous message load, message deadline tightness, message periods, network load and asynchronous message load. By using such metrics and token passing mechanisms other than the conventional approach (linear token passing) the network utilization may be increased.

## 3.2 Timed Token Protocol Parameters and Operation

### 3.2.1 Timed Token Protocol Parameters

The following parameters and timers are used by the timed token protocol:

- TTRT (Target Token Rotation Time): This parameter is the expected rotation time of the token and negotiated at the initialization of the network. In other words, this parameter determines how frequent the token will visit a station.
- $H_i$: This parameter is the amount of time allocated to a station to transmit its synchronous messages every time it receives the token. This is the most crucial parameter that affects whether the real-time guarantees of the synchronous messages will be satisfied.
- $TRT_i$ (Token Rotation Timer): This timer determines the interval of time that elapsed between the last two visits to the station. Initially the TRT timer is initialized with TTRT value and it begins to count down. When the station

gets the token, it stops the timer and checks its value. If TRT is 0 then it means that the token arrived later than expected, otherwise the token is assumed not to be late. After that TRT timer is again initialized with TTRT and continues to count down.

- $THT_i$ (Token Holding Timer): This counter determines the amount of time a station can transmit its asynchronous messages.
- $LC_i$ (Late Counter): This counter counts the number of times token rotation timer has expired since the last token arrival of this station.

### 3.2.2 Timed Token Protocol Operation

At the initialization phase the following parameters are initialized:

- $THT_i \leftarrow 0$
- $LC_i \leftarrow 0$
- $TRT_i \leftarrow TTRT$

After the initialization the TRT timer starts counting down. During the operation of the timed token protocol the conditions that can occur and the corresponding actions are given below [25]:

- $TRT_i$ counter reaches zero and the value of the late counter is also zero. This means that the token is not late. Then the following actions are taken:
  - $TRT_i \leftarrow TTRT$
  - $LC_i \leftarrow 1$
- $TRT_i$ counter reaches zero and the value of the late counter is one. This means that the token is late. This is an indication that the token rotation took more than 2.TTRT and there is an abnormal condition in the ring. So the ring must be reinitialized.
- The token arrives at a station and the value of the late counter is also zero. Since the $TRT_i$ counter has not reached zero yet, the token arrived earlier than expected and the load on the networks is light. Then the following actions are taken:
  - $THT_i \leftarrow TRT_i$
  - $TRT_i \leftarrow TTRT$
- Synchronous messages are transmitted without exceeding the $H_i$ value. After transmitting the synchronous messages, if there is some time left the asynchronous messages are also transmitted.

- The token arrives at a station and the value of the late counter is one. This means that the token arrived later than expected. Then the following actions are taken:
    - $LC_i \leftarrow 0$
    - $TRT_i$ continues to count down. It is not initialized to TTRT when the token is late.
    - Synchronous messages are transmitted without exceeding the $H_i$ value. Asynchronous messages are not transmitted.

### 3.3 Synchronous Bandwidth Allocation

The synchronous bandwidth is referred as the amount of time allocated to a station to transmit its synchronous messages every time it receives the token. A synchronous bandwidth allocation scheme is an algorithm that takes the message parameters such as $P_i$, $C_i$, $D_i$ and TTRT as input and produces the synchronous capacities $H_i$ allocated to station i as output.

Together with the TTRT value, this H value must be properly selected to satisfy real-time constraints of the stations in the network.

If the TTRT is negotiated smaller, then the token arrives at a station more frequently. But this causes the overheads to become more dominant.

If the TTRT is negotiated as a larger value, then the token arrives at a station less frequently. Hence this will cause some messages to miss their deadlines.

Similarly if the H value is allocated larger then this will cause to violate the protocol constraints whereas if the H value is allocated smaller then some messages may miss their deadlines.

As it is seen determining the H and TTRT value are crucial to provide real-time guarantees. As a result too much research is done for the allocation of H values and determination of the TTRT value [25].

There are two constraints that must be satisfied while allocating synchronous bandwidth to the stations:

- Protocol Constraint: The amount of synchronous bandwidth allocated to all stations cannot exceed the usable portion of TTRT. The usable portion denotes the fraction of TTRT that can be used for transmitting synchronous messages. Mathematically:

$$\sum_{i=0}^{n} H_i \leq TTRT - \tau \quad \textbf{(3.5)}$$

Where $\tau$ is the amount of time that cannot be used to transmit synchronous messages, in other words it is the unusable portion of TTRT.

- Deadline Constraint: The synchronous messages must be transmitted before their deadlines. This is the main constraint for providing real-time guarantees. If $X_i$ is the minimum amount of time available for a node to transmit its synchronous messages in the time interval $(t, t + P_i)$ then

$$X_i(H) \geq C_i \quad \textbf{(3.6)}$$

Note that $X_i$ is actually a function of H.

There are two classes of SBA schemes; global and local. A global SBA scheme can use the network wide information while assigning the bandwidths. It may find better assignments but it is not suitable for dynamic environments. Because a change in the message stream of a station causes to recalculate the synchronous bandwidths of all stations in the network. Another scheme is the local scheme. In local SBA schemes the stations use only the local information to calculate their synchronous bandwidths.

Many proposals have been made for the optimal allocation of the synchronous bandwidth to stations. In [26] Zhang and Lee examines three SBA schemes namely the NPA (Normalized Proportional Allocation), MCA (Minimum Capacity Allocation) and EMCA (Enhanced Minimum Capacity Allocation) and show with numerical examples that none of these schemes are optimal. Hence finding optimal SBA schemes is still a research challenge.

Currently one of the best known approaches is proposed by Feza Buzluca which is a semi-local SBA scheme [27]. This scheme uses only local information available to each stations and the sum of all the synchronous bandwidths assigned to other stations in the network as global information. Hence this scheme is more suitable for dynamic environments.

But in this thesis due to its simplicity and ease of implementation the Enhanced Minimum Capacity Allocation (EMCA) algorithm is used for allocating synchronous bandwidth to the stations in the network.

As its name suggests, EMCA is an enhanced version of MCA which allocates the minimum required synchronous bandwidth to the stations.

EMCA adopts a more exact deadline constraint expression than MCA and is able to provide some allocations which are impossible with MCA. Before presenting the EMCA algorithm it will be helpful to review some theorems and corollaries [28].

**Corollary:** Let I(v) be the tight upper bound on the maximum time which could possibly elapse in the worst case before a station uses its next v (v > 1) allocated synchronous bandwidth $H_i$ then

$$I(v) = v.TTRT + \sum_{j=1}^{n} H_j + \tau - \left\lfloor \frac{v}{n+1} \right\rfloor . \left[ TTRT - \left( \sum_{j=1}^{n} H_j + \tau \right) \right] \quad \textbf{(3.7)}$$

To find the minimum amount of time for a station to transmit its synchronous messages during its period the following steps must be performed:

- An integer $m_i$ must be chosen which satisfies the following inequality

$$I(m_i - 1) \le P_i \le I(m_i) \quad \textbf{(3.8)}$$

According to the Corollary, during the first $I(m_i - 1)$time interval, i.e. the interval (t, t+ $I(m_i - 1)$], station *i* can use its allocated synchronous bandwidth $H_i$ at least $(m_i - 1)$ times. So

$$X_i \ge (m_i - 1).H_i \quad \textbf{(3.9)}$$

To use the remaining part of the allocated bandwidth $H_i$ in the worst case during the remaining time interval (t + $I(m_i - 1)$, t +$P_i$], if the following inequality holds

$$I(m_i) - H_i < P_i < I(m_i) \quad \textbf{(3.10)}$$

- Minimum amount of time available for station *i* to make synchronous transmission during the remaining period can be calculated by the formula

$$\max \left\{ P_i - \left\{ m_i.TTRT + \sum_{j=1}^{n} H_j + \tau - \left\lfloor \frac{m_i}{n+1} \right\rfloor . \left[ TTRT - \left( \sum_{j=1}^{n} H_j + \tau \right) \right] - H_i \right\}, 0 \right\} \quad \textbf{(3.11)}$$

By combining the previous results the minimum available time $X_i$ for station *i* to send its synchronous messages during $P_i$ in the worst case:

$$X_i(\vec{H}) = (m_i - 1)H_i + \max \left[ P_i - \left\{ m_i.TTRT + \sum_{j=1}^{n} H_j + \tau - \left\lfloor \frac{m_i}{n+1} \right\rfloor . \left[ TTRT - \left( \sum_{j=1}^{n} H_j + \tau \right) \right] - H_i \right\}, 0 \right] \quad \textbf{(3.12)}$$

So to find the $X_i$ vector the vector $M = \{m_1, m_2 \ldots m_n\}$ for a given synchronous message set must be determined. To determine the range of values $m_i$ can take, the following two lemmas are used.

**Lemma 1:** For a given synchronous bandwidth allocation vector $H = \{H_1, H_2 \ldots H_n\}$ the integer $m_i$ $(i = 1,2 \ldots n)$ which can make the inequality of $I(m_i - 1) \leq P_i$ hold must satisfy

$$m_i \leq \left\lfloor \frac{P_i.(n+1) + n.(TTRT - \sum_{j=1}^{n} H_j - \tau)}{n.TTRT + \sum_{j=1}^{n} H_j + \tau} \right\rfloor \quad \textbf{(3.13)}$$

**Lemma 2:** For a given synchronous bandwidth allocation vector $H = \{H_1, H_2 \ldots H_n\}$ the integer $m_i$ $(i = 1,2 \ldots n)$ which can make the inequality of $P_i \leq I(m_i)$ hold must satisfy

$$\left\lceil \frac{(n+1).P_i - \sum_{j=1}^{n} H_j - \tau - n.TTRT}{n.TTRT + \sum_{j=1}^{n} H_j + \tau} \right\rceil \leq m_i \quad \textbf{(3.14)}$$

**Theorem 1:** Assume that at time t, a synchronous message with period $P_i$ arrives at station. Then in the interval $(t, t + P_i]$ the minimum amount of time $X_i$ available for station i to transmit this synchronous message is

$$X_i(\vec{H}) = (m_i - 1).H_i + \max\left[ P_i - \left\{ m_i.TTRT + \sum_{j=1}^{n} H_j + \tau - \left\lfloor \frac{m_i}{n+1} \right\rfloor . \left[ TTRT - (\sum_{j=1}^{n} H_j + \tau) \right] - H_i \right\}, 0 \right] \quad \textbf{(3.15)}$$

where $m_i$ is an integer $(m_i \geq 2)$ which makes the inequality of $I(m_i - 1) \leq P_i \leq I(m_i)$ hold and must be either

$$m_i = \left\lfloor \frac{P_i.(n+1) + n.(TTRT - \sum_{j=1}^{n} H_j - \tau)}{n.TTRT + \sum_{j=1}^{n} H_j + \tau} \right\rfloor \quad \textbf{(3.16)}$$

or

$$m_i = \left\lfloor \frac{P_i.(n+1) + n.(TTRT - \sum_{j=1}^{n} H_j - \tau)}{n.TTRT + \sum_{j=1}^{n} H_j + \tau} \right\rfloor - 1 \qquad \textbf{(3.17)}$$

Theorem 1 gives the deadline constraint expression for the EMCA scheme.
The main EMCA procedure uses some helper procedures to allocate the synchronous bandwidth to stations. The procedure Find_X given below calculates the vector X and m as defined by the theorem and the corollary.

```
procedure Find_X
begin
        for i = 1,2,…,n
        begin
                m_i = ⌊ (P_i.(n+1) + n.(TTRT − Σ_{j=1}^{n} H_j − τ)) / (n.TTRT + Σ_{j=1}^{n} H_j + τ) ⌋
                Calculate I(m_i−1) as defined in Corollary
                if I(m_i−1) > P_i
                        m_i = m_i−1
                calculate X_i as defined in Theorem
        end
        return (X⃗, m⃗)
end
```

**Theorem 2:** For a synchronous message set to be guaranteed the synchronous bandwidth allocated to a station *i* must be in the range

$$\frac{C_i}{\left\lfloor \frac{P_i.(n+1)}{n.TTRT} \right\rfloor + 1} \le H_i \le \frac{C_i}{\max\left[ \left\lfloor \frac{P_i}{TTRT} \right\rfloor - 1, 1 \right]} \qquad \textbf{(3.18)}$$

Using Theorem 2 another helper procedure Min_H tries to allocate the minimum values for the synchronous bandwidth to stations. It first tries to allocate the minimum value of H which is the lower bound given in Theorem 2 and tries to refine this value iteratively. The Min_H procedure is given below.

**procedure** *Min_H*

**begin**

**for** i=1,2,…,n

$$H_i = \frac{C_i}{\left\lfloor \frac{P_i.(n+1)}{n.TTRT} \right\rfloor + 1}$$

**repeat**

**if** $P_{\min} \le TTRT + \sum_{i=1}^{n} H_i + \tau$

**begin**

deadline_constraint_violated = 1;

**return** (fail, deadline_constraint_violated);

**end**

**call** Find_X to calculate $\vec{X}$ with return( $\vec{X}, \vec{m}$ )

**for** i=1,2,…,n

**begin**

$\Delta_i = C_i - X_i$

**if** $\Delta_i > 0$

$$H_i = H_i + \frac{\Delta_i}{m_i - 1}$$

**end**

**until** none of $\Delta_i$ s are larger than zero;

deadline_constraint_violated = 0;

**return** ( $\vec{H}$ ,deadline_constraint_violated)

**end**

At each iteration the procedure Find_X is used to calculate the vector X and m. After calculating these values a deficiency value Δ, i.e. the difference between the minimum transmission time $X_i$ and the message length $C_i$, is then calculated. All the $H_i$ values with positive deficiencies are refined by an amount no more than the deficiency. So as a result the value of $m_i$ gets smaller at each iteration whereas the value of $H_i$ gets larger.

And finally by using the helper procedures Find_X and Min_H the main procedure EMCA is given below.

```
procedure EMCA

begin
      call procedure Min_H to obtain  H^min
      if  Σ_{i=1}^{n} H_i^min ≤ TTRT − τ and deadline_constraint_violated=0
            return (success, H^min );
      else
            return (fail,nil);
end
```

First the EMCA procedure tries to allocate the minimum synchronous bandwidth to the stations by calling the procedure Min_H. And then checks whether this vector satisfies the protocol constraint and the deadline constraint by checking the flag *deadline_constraint_violated*. If this allocated vector satisfies both the protocol constraint and the deadline constraint then procedure EMCA returns this allocation as the result since the allocated synchronous bandwidth is feasible.

If the allocated vector does not satisfy the constraints then this allocation is not a feasible allocation and the procedure EMCA returns nil as a result as well as indication of failure to the caller.

By using the given EMCA procedure the stations in a network can be assigned synchronous bandwidths to guarantee the transmission of their synchronous messages.

EMCA is chosen in this thesis for synchronous bandwidth allocation because it is not a complicated algorithm and simple to implement. It can find allocations in some cases where other bandwidth allocation algorithms cannot make feasible allocations.

# 4. THE PROPOSED CONTROL PLANE

The main problem this thesis tries to solve is to design an architecture that makes the wireless token ring networks more suitable for real-time communications. To achieve this, a management sub layer or in other words a control plane is implemented on top of the MAC sub layer. This control plane resides in the logical link control sub layer and performs critical management functions to make the network more suitable for real-time communications.

## 4.1 The Proposed System Architecture

The proposed architecture is designed for wireless networks with token ring topology. While designing the architecture that will provide the stated requirements, some simplifying assumptions are made. Although these assumptions simplify the analysis and design of the system, they are considered carefully so not to violate real-time guarantees of the stations in the network.

The control plane is actually designed for controlled wireless local area networks such as industrial automation networks or wireless local area networks for military applications. These types of networks are usually single hop adhoc networks where real-time requirements are crucial to system operation. An example use case will be given at the end of this section.

In the architecture designed, each station in the network is assumed to use the timed token protocol in the MAC sub layer. The timed token protocol has special timing properties and has solid mathematical foundations. It is designed to be used for hard real-time communications. And also extensive amount of research has been done on this protocol especially on SBA schemes and as a result it has been well understood in the research community.

Since the timed token protocol is used in the proposed architecture, a SBA scheme to allocate synchronous bandwidth to the stations should also be chosen. Many SBA schemes have been proposed by researchers. But finally Zhang and Lee showed that the most famous of these allocations namely the NPA (Normalized Proportional Allocation), MCA (Minimum Capacity Allocation) and EMCA (Enhanced Minimum Capacity Allocation) are not optimum SBA schemes [26]. It is shown in their paper

that these SBA schemes could not allocate the optimum synchronous bandwidth in some cases.

Buzluca has also proposed a semi local SBA scheme that uses only a little amount of global information such as the total synchronous bandwidth allocated to all stations in the network. This scheme produces better allocations than the well known schemes and also it is more suitable for dynamic environments. For the sake of simplicity and ease of implementation, the proposed architecture uses the EMCA scheme which can allocate synchronous bandwidth to the stations in the network pretty well.

Another simplifying assumption made while designing the system is that there is a management station in the network that makes the actual management. This management station is the heart of the system. This makes the system a centralized system. One can think that a centralized approach like this can suffer from scalability and being a single point of failure. Scalability is not an issue for our system because the system is actually designed for small scaled wireless local area networks such as industrial automation networks or networks used for military purposes. Also being a single point of failure is also not an important issue because in these types of networks generally a super administrator exists who takes care of the crucial parts of the systems. The target networks can be thought as simple hop ad-hoc networks with relatively small number of wireless stations. The most important property for these networks is that quality of service guarantees and real-time communication requirements are of crucial importance. Usually real-time applications such as critical control applications are deployed in these types of networks. In these types of networks, the characteristics and requirements can be stated as follows:

- Frame lengths are short
- Mobility is limited. Generally the radius of the network is less than 100m
- Bit error rate is generally low such as BER $\leq 10^{-6}$
- Since the radius of the network is relatively small, the propagation delay is also small such that $t_p \leq 200$ ms
- Data rate $\geq 2$ Mpbs

To fulfill such requirements the architecture must provide guaranteed QoS in terms of high reliability and bounded delays. Conventional MAC protocols are not suitable to provide the requirements stated above [13].

To simplify the analysis of the proposed architecture in terms of advantages and disadvantages, it is assumed that stations in the token ring network send connection requests to the management station to join the ring. After a connection is set up a

connection identifier is generated by the MAC layer and returned to the upper layer. This connection identifier will be used to tear down the connection later on by the source of the station.

A connection request consists of the following parameters with their corresponding definitions:

- $C_i$ : The length of the messages that will be carried in the traffic stream for this connection i.e. the maximum amount of time needed to transmit the messages in the stream.
- $P_i$ : The period of the messages that will be carried in the traffic stream for this connection
- $D_i$: The deadline of the messages that will be carried in the traffic stream for this connection
- $TC_i$ : The class of the traffic that will be carried over this connection
- SA: The address of the source station
- DA: The address of the destination station

Whenever a station wants to establish a connection to another station in the ring it must prepare a connection request message with the source address set to its MAC address, the destination address is set to destination station's MAC address, $C_i$ is set to maximum amount of time needed to transfer the messages in the traffic stream that will be carried by this connection, $P_i$ is set to the period of the messages in the traffic, $D_i$ is set to the deadline of the messages in the traffic and finally the traffic class is set to the class of the traffic that will be carried. The following traffic classes which are defined in 802.11e can be used in the proposed architecture:

- Voice
- Video
- Best effort
- Background

The traffic classes are given in the order of decreasing priority such that voice has the highest priority and background traffic has the lowest priority.

Another assumption made is that the connections in the system have a fixed lifetime. Each connection that has been requested has the same fixed lifetime. After the lifetime of a connection ends, a disconnect request is sent from the source station that has set up the connection and as a result the connection is torn down. Similar to connection requests, disconnect requests are also sent to the management station. A disconnect requests only consists of a connection identifier that was generated by the MAC layer during the connection setup phase.

## 4.2 System Operation

The control plane is implemented in a management station in the network which can bring the problems of scalability and single point of failure with it. But since the system is designed for wireless industrial automation networks and wireless networks for military applications, these problems are not of serious concern. Actually one can design the same control plane as a distributed architecture and can get rid of these problems but the algorithms get more complex and more challenging to implement and verify.

The control plane is implemented in a distinct management station in the network as shown in Figure 4.1. As it is stated before the control plane resides in the logical link control sub layer. This assumption seems reasonable when looking the system from the upper layers. The upper layers' real-time communication requirements are provided by the algorithms running on the logical link control sub layer and the management station.

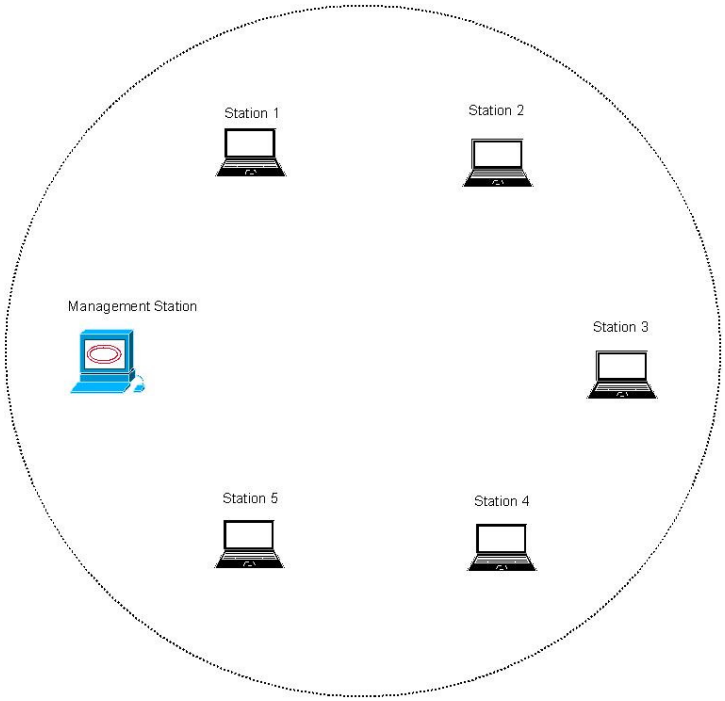


**Figure 4.1:** System under Consideration

To understand the operation of the proposed architecture Figure 4.2 will be used.

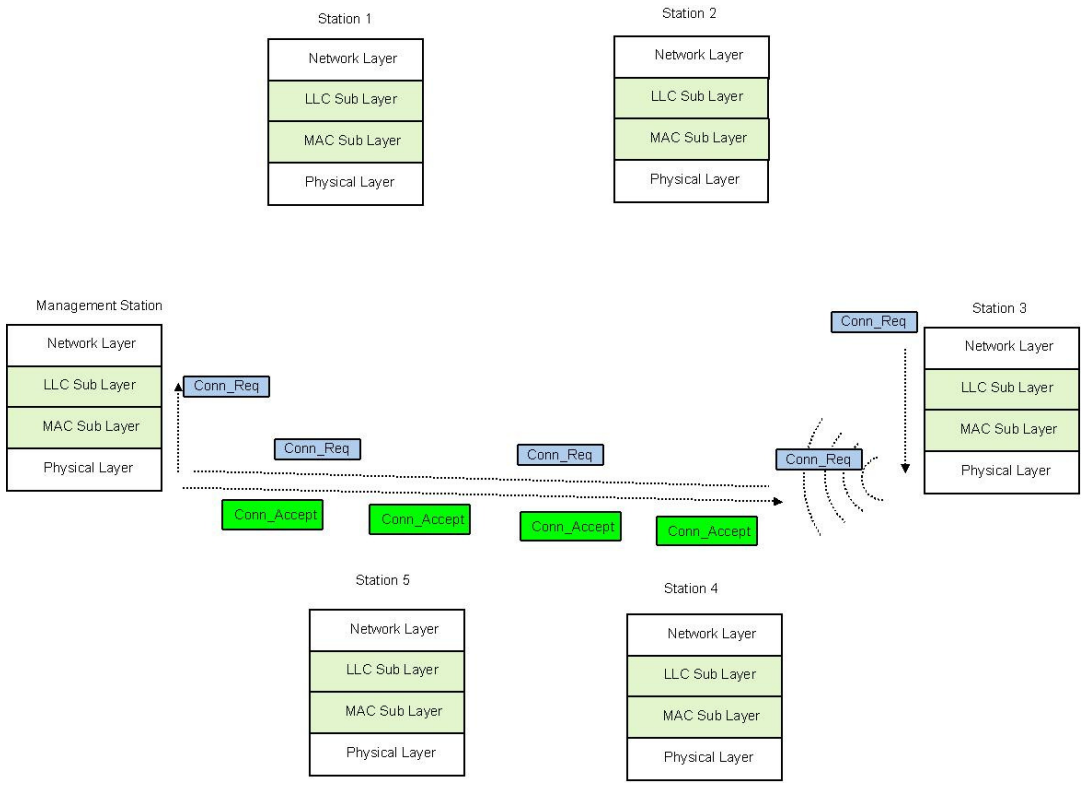


**Figure 4.2:** Connection Request Accepted

In Figure 4.2 each wireless station in the network is shown with the OSI layers, but for the simplicity of the figure the layers above the network layer namely the transport and application layers are omitted.

In the figure wireless station 3 wants to set up a connection with station 1. Initially station 3 is out of the ring and wants to join the ring by sending a connection request to station 1. So station 3 prepares a connection request frame and sets the source address field as the MAC address of station 3, sets the destination address as the MAC address of station 1 and sets $C_i$, $P_i$, $D_i$ and the $TC_i$ parameters according to the requirements of the application. It is stated in the requirements of industrial automation networks and military networks that in these types of networks the frames are relatively short hence in this figure it is assumed that no fragmentation is necessary while sending the frames over the wireless medium.

The connection request is passed from the network layer to lower layers on station 3 and finally it is transmitted by the physical layer. Although wireless medium is a broadcast medium, the communication path is shown with a dotted arrow as if it was a unicast communication.

After the connection request arrives at the management station, the frame is delivered to the logical link control sub layer where the management station needs to determine whether the connection should be accepted. This procedure is called *admission*

*control* procedure. Since the system is a centralized system the management station has the knowledge of the global state of the system. The global information includes all connections currently active in the system with their corresponding parameters. First of all the management station checks whether the connection request can be accepted without disrupting other stations' real-time guarantees. At this stage the EMCA SBA scheme is used for determining the synchronous bandwidths for the stations in the network. As a result of this check the new $H_i$ values are determined if possible and broadcasted to the ring members and station 3 joins the ring. After station 3 joins the ring station 2 updates its connectivity tables so that it can pass the token to station 3 in the next cycle.

If the management station cannot accept the connection request due to the fact that accepting it will disrupt other stations' real-time guarantees, it executes the *station eviction* procedure. This procedure is used to evict the best possible station from the ring such that the new connection could be accepted. Here the traffic class parameter in the connection request comes into play. Suppose that the connection request of station 3 could not be accepted and the management station began executing the station eviction procedure. The management station checks the stations other than the destination since to set up the connection the destination station must stay in the ring. And then the management station tries to find the station with the least possible traffic class such that removing this station from the ring will make the system handle the new connection without disrupting the real-time guarantees of other stations. In the example the management station tries to remove the stations other than station 1 since it is the destination. The management station checks stations 4 and 5. Suppose that the application on station 4 has a traffic stream with class best effort and the application on station 5 has a traffic stream with class video. And also suppose that the traffic class contained in the connection request is voice which is the class of the traffic that the application on station 3 manipulates. Two distinct decisions can be made here such as to choose the *first* possible station such that the removal of it will make the network handle the new connection or to choose the station with the *least* possible traffic class to remove and make the network handle the new connection. Generally in system design, the designers face these kinds of decisions. Actually this decision is to choose between the first fit station and the best fit station. The best fit operation is more costly than the first fit because the management station must try all the possible stations and choose the one with the least possible traffic class. But the best fit heuristics actually yield better results. Whereas when adopting the first fit approach the management station chooses the first station that satisfies the given conditions and evicts it.

After the management station determines the station to evict from the ring, the management station calculates the new $H_i$ values and updates the TTRT appropriately. Actually TTRT should be updated according to the deadline parameter in the connection request. According to the Johnson & Sevcik's Theorem [Timing properties of the timed token MAC protocol] the maximum amount of time that can pass between two visits of the token is 2.TTRT. And hence if D < 2TTRT the deadline can be missed. So D must be greater than or equal to 2.TTRT to meet the deadline. As a result management station should update TTRT if the deadline in the connection request is smaller than 2.TTRT [25].

After the necessary parameter updates, these parameters and the MAC address of the evicted station is broadcasted to all ring members so that they can update their connectivity tables and parameters appropriately. By updating the connectivity tables it is guaranteed that other stations in the ring do not pass the token to the evicted station hence leaving that station is out of the ring. And as a result of all these operations a new ring is set up such that the real-time requirements of all the members are guaranteed. And the ring continues to operate normally.

After all these procedures, if the management station still cannot accept the connection request, it sends a connection reject to the station that sent the connection request. And the application in the source station is notified such that the connection is rejected by the system and as a result the source station cannot join the ring. This connection rejection procedure is shown in Figure 4.3.

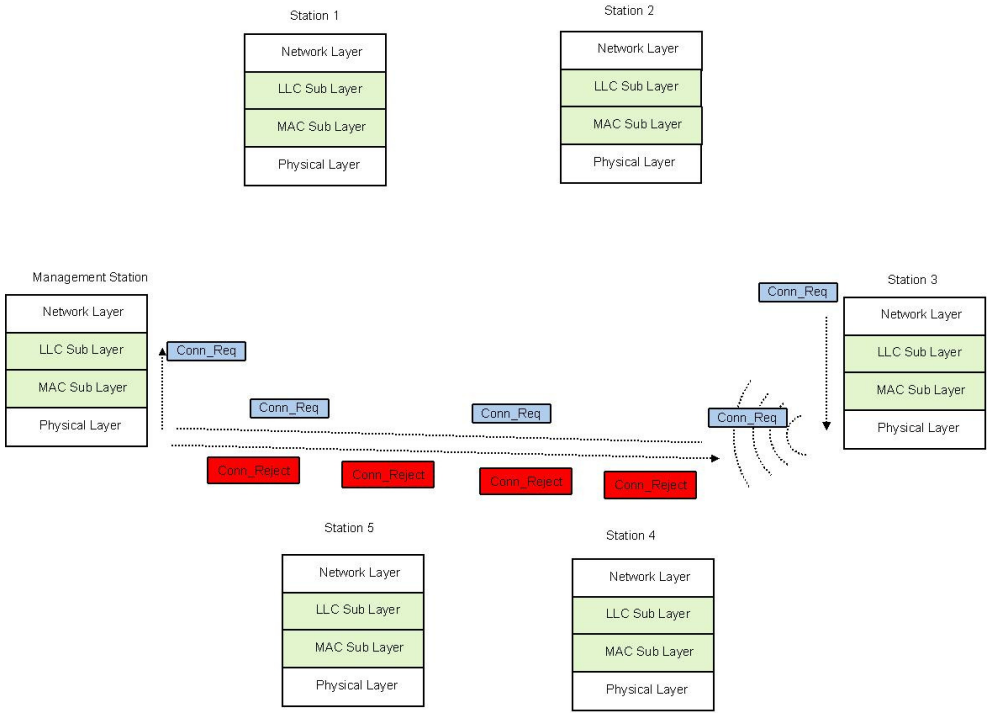


**Figure 4.3:** Connection Request Rejected

For the station eviction procedure there are two design alternatives that can be adopted after the station eviction procedure. After the eviction of the station, the connection of the evicted station is torn down and the application on this station is notified of disconnection. Here after a disconnection request is received from any other station, the management station could give priority to evicted stations when setting up a new connection for the sake of fairness. Another alternative is that the management station simply does not care about the evicted stations. The first alternative comes with the complexity of the implementation and management but provides a more fair system. Although the second alternative is simpler, it is also less fair. In this thesis the second alternative is chosen for the sake of simplicity and ease of management.

The overall process of connection requests are given in Figure 4.4.

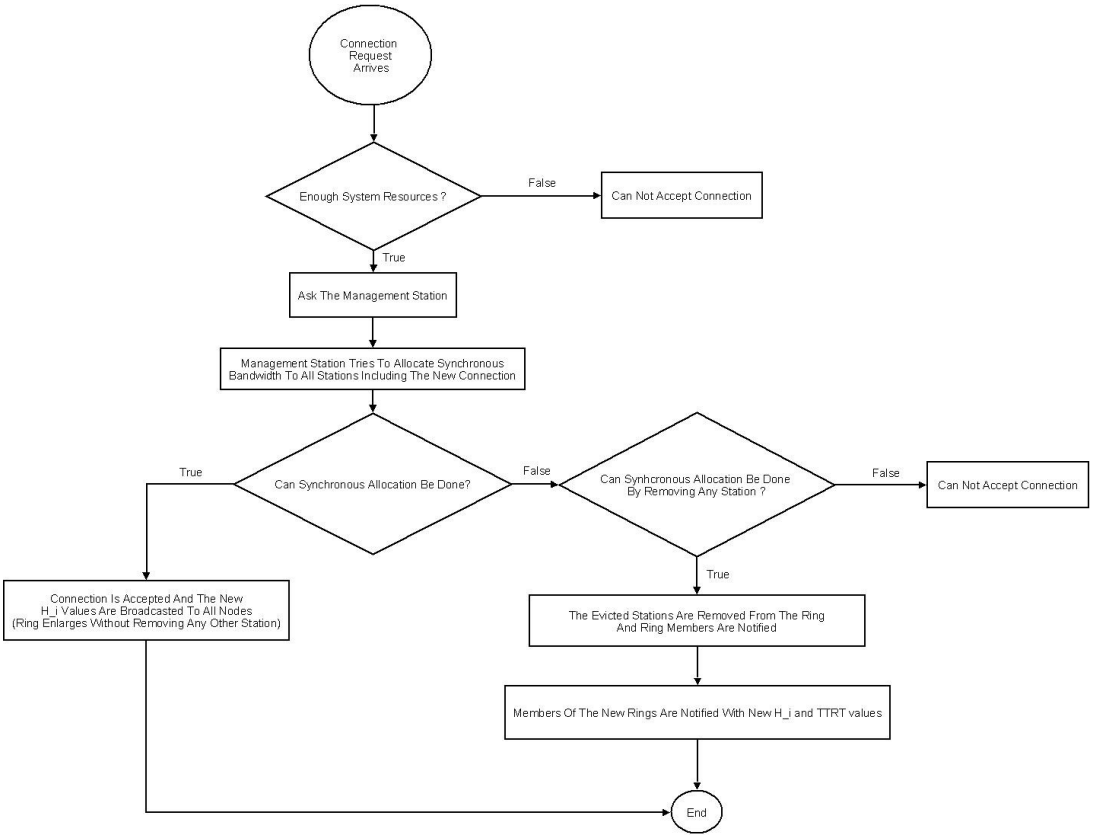


**Figure 4.4:** Flowchart for Connection Handling

Also to understand the implementation details the pseudocode will also be helpful which is given below. Source station $p$ that is not currently in the ring tries to establish a connection to destination station $q$ and transmit a synchronous stream $S_p$ where $S_p \notin$ S. $H$ is the set of synchronous bandwidths of stations that is calculated by the EMCA algorithm.

**Algorithm**: Dynamic Ring Management Algorithm

```
procedure DRM(Connection_Request[Sp])
p:Source, q:Destination, Ti: Traffic Class of Si
begin
  satisfied = EMCA(S+Sp,TTRT,τ,H);
  if satisfied then
   broadcast(H,TTRT);
   return ACCEPT;
  end if
  staToRemove = NULL;
  for ∀si ∈ S
   if q != i AND Ti < Tp then
    satisfied = EMCA(S+Sp-Si,TTRT,τ,H);
    if satisfied then
      if staToRemove = NULL
         OR TstaToRemove > Ti then
         staToRemove = i;
  end for
  if staToRemove != NULL then
   broadcast(staToRemove,H,TTRT);
   return ACCEPT;
  end if
  return REJECT;
end
```

To make the example more concrete suppose that stations 1, 2, 4 and 5 have the following message streams which are carried on already established connections:

- *Station 1*: The period of the messages in the stream ($P_i$) is 100 ms and the maximum amount of time needed to transmit the messages ($C_i$) is 8 ms where the class of traffic is background.
- *Station 2*: $P_i$ is 120 ms and $C_i$ is 5 ms where the class of traffic is best effort.
- *Station 4*: $P_i$ is 150 ms and $C_i$ is 20 ms where the class of traffic is voice.
- *Station 5*: $P_i$ is 140 ms and $C_i$ is 9 ms where the class of traffic is video.

Suppose that the connection request received from station 3 contains the parameter $P_i$ with value 112 ms and $C_i$ with value 20 ms and also suppose that the traffic class is voice. For all the stations let $P_i = D_i$.

First of all the management station tries to allocate synchronous bandwidth for the set of message streams with the EMCA scheme:

$$M = \{(100, 8), (120, 5), (150, 20), (140, 9), (112, 20)\}$$

Unfortunately EMCA fails to determine a feasible synchronous bandwidth allocation for this set of message streams. So the management station seeks for the station with the least possible traffic priority whose removal will make the system provide the

real-time guarantees for all the remaining stations in the ring. In this case the evicted station is 1. So the set of message streams M becomes:

$$M = \{(120, 5), (150, 20), (140, 9), (112, 20)\}$$

Finally the EMCA scheme can provide a feasible synchronous bandwidth allocation for this set of message streams where $H_1$=5 ms, $H_2$=10 ms, $H_3$=5 ms and $H_4$=20 ms. After the parameters are calculated the management station broadcasts these values together with the evicted station's MAC address to the ring members. And at the end the station 3 joins the ring and the ring continues to operate normally.

To disconnect an already established connection, a station must send a disconnection request to the management station. The disconnect request contains the following parameters:

- Connection Id: This is the identifier of the connection that uniquely identifies it on the station. This identifier is generated while setting up the connection.

Suppose that the station 3 wants to disconnect the connection established before. First the application layer on station 3 sends a disconnect request to the lower layer. Then the lower layers prepare the disconnect request message frame and send it over the wireless medium to the management station. Upon receiving the disconnect request the management station notifies the ring members that the station 3 will leave the ring. As a result of this notification all the ring members update their connectivity tables in order not to pass the token to station 3. And then the management station sends a confirmation to station 3 and the station 3 leaves the ring.

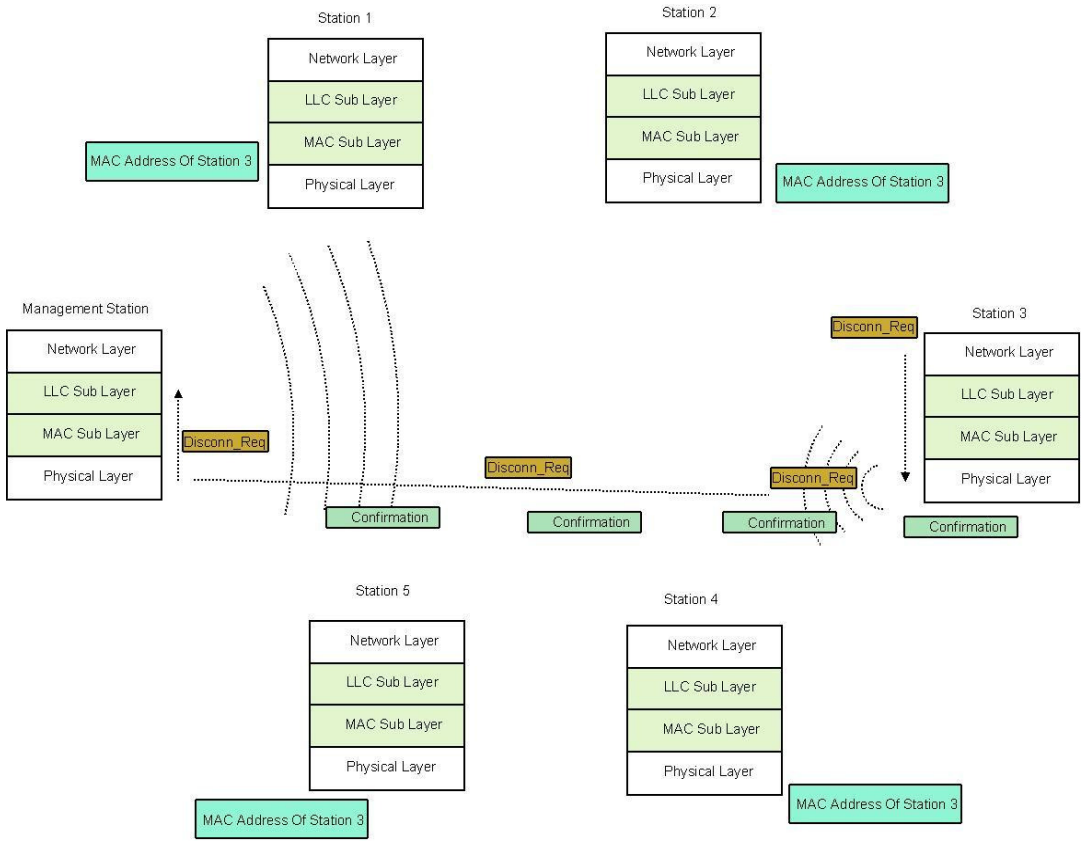


**Figure 4.5:** Disconnect Request Processing

There is an important step after the confirmation is sent in response to the disconnect request. The management station can recalculate the synchronous bandwidth values. Since a station has left the ring, the new values for the synchronous bandwidth will increase and as a result this increases the probability of meeting the real-time guarantees of the stations. The flowchart for the disconnect request processing is given in flowchart in Figure 4.6:

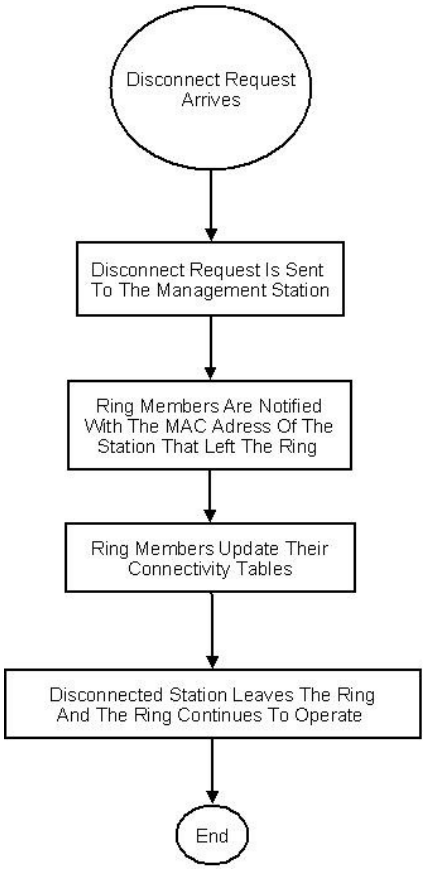


**Figure 4.6:** Flowchart for Disconnect Request Handling

## 4.3 Finite State Machine

The control plane is best understood by analyzing the finite state machine in Figure 4.7 where the arcs represent the actions that cause the transitions.

### 4.3.1 Init State

This state represents the initialization of the wireless station. In this state necessary allocations are made and necessary protocol parameters are initialized. Also some protocol parameters like TTRT are negotiated with other stations. After this state the station directly goes to the *Idle* state.

### 4.3.2 Idle State

This state represents the main loop of the protocol where the protocol waits for events from the upper layers or from the lower layers. Whenever a connection request is received in this state, the station makes a transition to the *Conn_Req* state. Or if a disconnect request is received then the station makes a transition to the *Disconn_Req* state.


**Figure 4.7:** Finite State Machine

The most important two transitions are the transitions labeled with *I_am_evicted* and *another_station_disconnected* . The *I_am_evicted* transition occurs whenever this station is notified that it is evicted by the management station because the management station determined that this station is the most suitable station to evict in order to accept the new connection. The *another_station_disconnected* transition occurs whenever a station sends a disconnect request to the management station and the management station sends a notification to this station such that another station has disconnected and left the ring. Upon receiving these events the station makes a transition to the *Reforming_Ring* state.

### 4.3.3 Reforming_Ring State

In this state the station updates its connectivity tables which it uses for determining the successor for this station. And also in this state the station may update some protocol parameters if the management station sends new values for these parameters. For example the management station can recalculate the values of the synchronous bandwidths and broadcasts these values along with the MAC of the

disconnected station and the receiving stations can update their synchronous bandwidth parameters appropriately. After the necessary action is taken then the station goes to the *Idle* state.

### 4.3.4 Conn_Req State

The station makes a transition to this state after a connection request is received. In this state if there are not enough resources to establish the connection, then the station makes a transition to the *Conn_Reject* state. If there are enough resources, then the station sends a connection request message to the management station and goes to the *Idle* state.

### 4.3.5 Conn_Accept State

The station can go into this state only after the connection request is accepted by the management station in the *Idle* state. In this state the station can perform some maintenance operations like setting up timers, allocating enough memory to maintain the connection or initialize some connection related parameters. After this state, the station goes directly to the *Idle* state.

### 4.3.6 Conn_Reject State

The station can go into this state only after the connection request is rejected by the management station in the *Idle* state. The upper layer is notified of the connection rejection. After this state, the station goes directly to the *Idle* state.

### 4.3.7 Disconn_Req State

The station makes a transition to this state after a disconnection request is received. In this state the station sends a disconnect request message to the management station and goes into the *Idle* state to wait for the confirmation of the disconnect request.

### 4.3.8 Disconnected State

The station can go into this state only after the confirmation for the disconnect request is received from the management station. In this state the station performs some housekeeping operations like deallocating memory or stopping some timers used for maintenance. After performing these housekeeping operations the station goes directly to the *Idle* state.

## 4.4 Sample Use Cases for the Control Plane

As it is stated at the beginning of this section, the control plane designed is suitable for small scale and possibly one hop adhoc networks where the radius of the network is relatively small.

Wireless industrial automation networks and small scale networks for military applications are good examples for these kinds of networks. In these kinds of networks real-time service guarantees are of crucial importance so designing and implementing protocols for guaranteed real-time operation is an important research topic.

### 4.4.1 Wireless Industrial Automation Networks

Wireless networks are especially being used in industrial automation applications possibly in the areas where it is either difficult or expensive to wire. Usually wireless networks are extensions to the Ethernet segments in these networks. Also it became possible for devices to be mobile since there are no physical restrictions like wires. As a result of these, it can be said that usage of wireless networks will gain popularity in the future.

In these networks generally the carried data is a vital control or monitoring information. And much of the status or control information is carried in short bursts which generally require relatively little bandwidth and connection speed and also the traffic is usually periodic. Or sometimes large data like log files can be carried over the medium. The key point for communication in these networks is the timely delivery without failure.

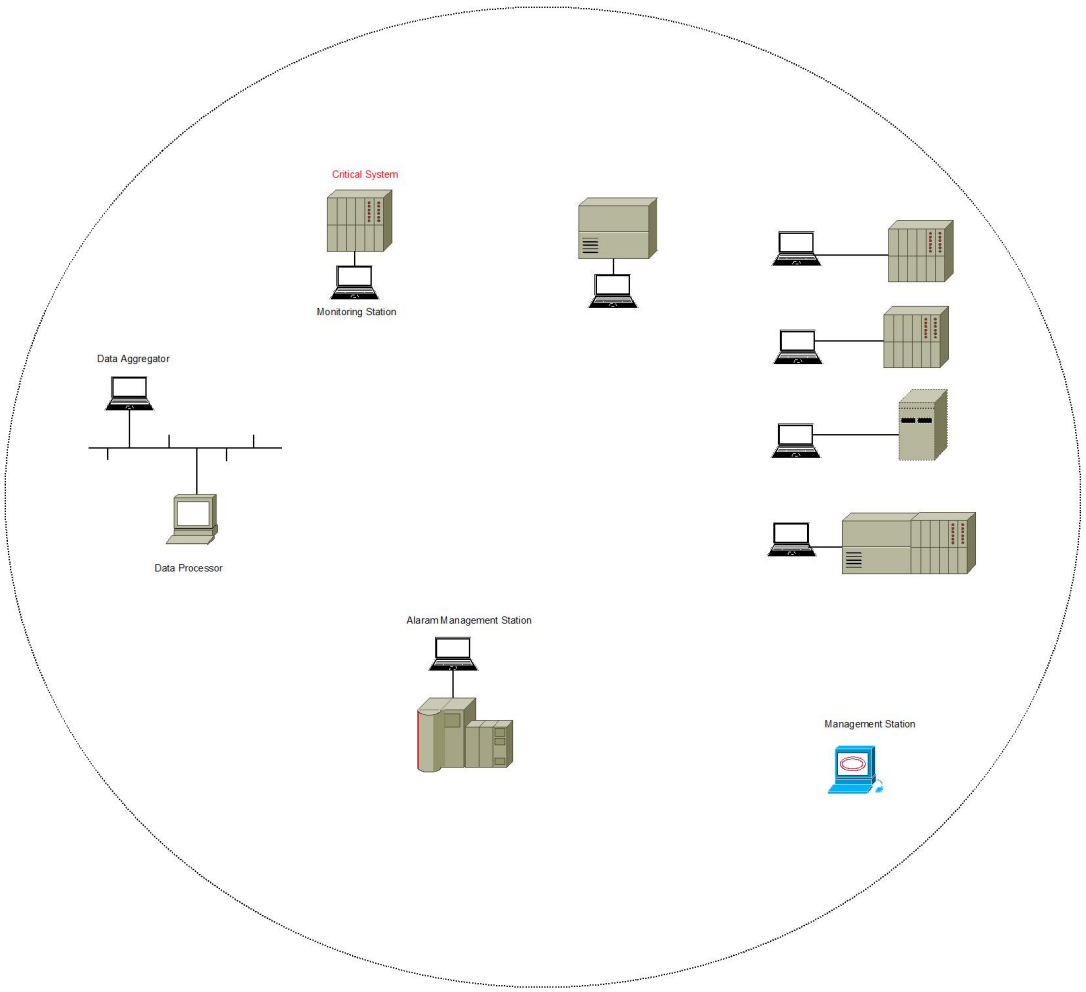


**Figure 4.8:** A Simple Wireless Industrial Automation Network

Consider the simple wireless industrial automation network in Figure 4.8. Initially suppose that the ring consists of all the stations except for the monitoring station. The monitoring station monitors a critical system which is a crucial entity in the network. Some stations gather data from other entities in the network and transmit to the data aggregator station which aggregates the data from the wireless stations and pass these data to be processed over a wired LAN to the data processor station. There is also an alarm management station which handles exceptional events received from the stations in the network. This alarm related traffic can be alarms caused by a physical event such as a fire or flood or something like exceeding the temperature threshold which can be measured by the deployed sensors. Or the alarm can be related to the system for example the size of the empty space in the file system of the data aggregator has fallen under some configured threshold. This kind of traffic has a higher traffic class than other traffic in the network. Actually in some deployments there can be wireless stations transferring voice or video traffic over the established connections. Similar to the alarm traffic, this multimedia traffic has higher priority

than other traffic in the network. Actually higher priority traffic has real-time guarantees such that the timely delivery of this traffic is crucially important. If the packets arrive at the destination after the deadline they become useless or can cause problems.

Suppose that the ring operates normally; the stations transmit the gathered data or logs and events to their destinations. At some time, suppose that there was a critical failure in the critical system which will cause the network to malfunction. To report this exceptional situation to the alarm management station, the monitoring station wants to send an alarm event to this station. To perform this it needs to establish a connection to the alarm management station and needs to join the ring. The monitoring station sends a connection request to the alarm management station by setting the source address as its MAC address, destination address as the alarm management station's MAC address, the traffic class as the highest priority traffic class and also sets the period, deadline and the maximum amount of time to transmit a message in its stream ($C_i$) appropriately and sends this request to the management station. First of all the management station checks whether it is possible for the monitoring station to join the ring without degrading the QoS guarantees of other stations. If so then the connection request is accepted. But if it degrades the QoS of other stations then the management station needs to execute the station eviction procedure. By performing the station eviction procedure the management station evicts the best possible station in the ring with a lower traffic class such that removing it from the ring will make the management station accept the new connection. Finally the management station evicts a station and removes it from the ring and accepts the new connection and the alarm can now be reported to the alarm management station.

Suppose for the worst case that the network consists of stations transmitting only high priority traffic and a new connection request with also high priority traffic arrives at the management station. Then if accepting the new connection will degrade the QoS guarantees of the stations in the network, then the management station must evict some station. But unfortunately all station have connections established that carry high priority traffic and none of them can be evicted. As a result the connection will be rejected. According to the station eviction procedure if there is a station having a lower traffic class than the new connection request then that station will be evicted and the connection will be accepted. This is the actually what makes the system more responsive to the high priority traffic and makes the system more suitable for real-time communications.

### 4.4.2 A Military Application

Due to advances in wireless communications technology many wireless military networks have been proposed like wireless sensor networks or experimental cellular and short range 3G networks. What makes wireless communications in military applications attractive is that there are no physical restrictions like wires, technological advances can provide long range communication capabilities and also these wireless systems provide mobility. These wireless military networks are gaining popularity and probably will be of greater importance in the future. As a result number of wireless military network deployments will probably increase in the future.

In wireless military networks the main concerns are actually the real-time guarantees of the stations in the network and the security of the communications. Especially due to increases in the data rates of wireless communications, new applications are being considered for deployment on this wireless networks. For example a robot in a battlefield can take photos of an enemy target and can transmit this information to a center to be processed for further action. These multimedia applications are being used today and will also be used extensively in the future. To provide the quality of service needed by the stations in the battlefield requires specialized real-time command and control software and quality of service aware network protocols.

Consider the simple military network deployed on a battleship in Figure 4.9.

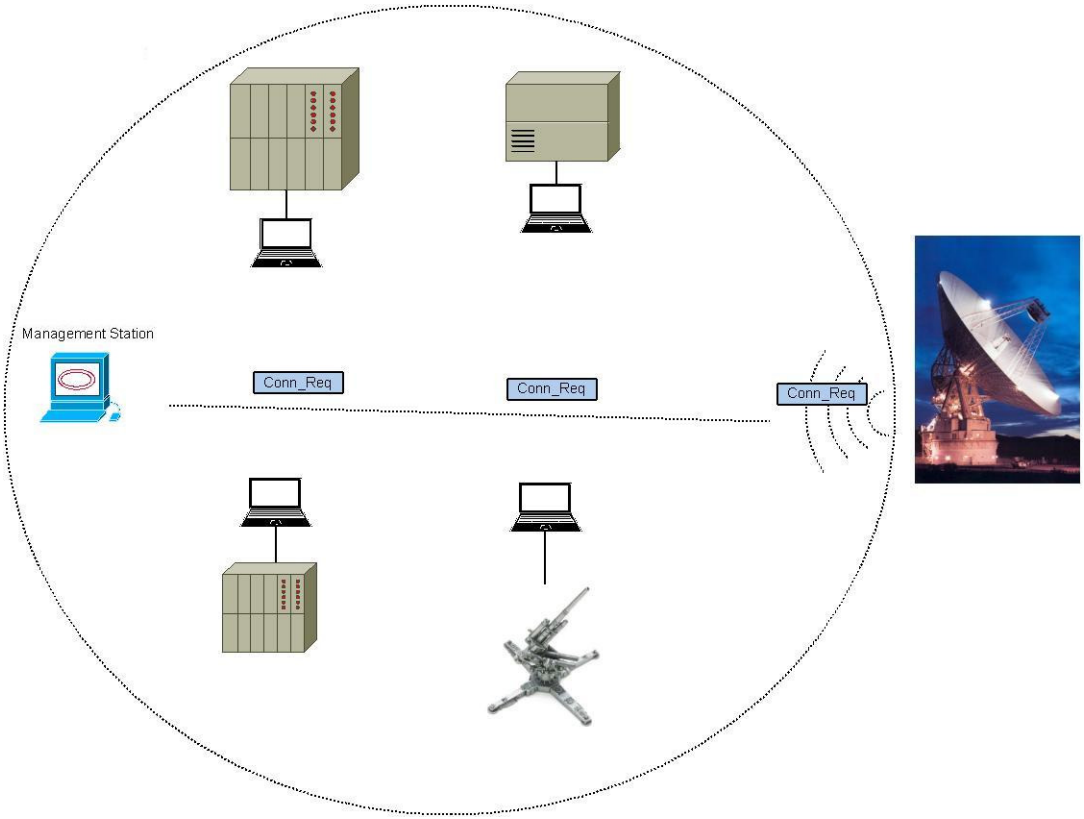


**Figure 4.9:** A Simple Military Network

There are some stations performing routine tasks like monitoring, controlling or logging some events. There may also be people using the wireless network for example using the internet. And also there is a wireless station controlling a gun. Initially the network consists of all the stations except for the radar which is at the rightmost edge of the Figure. The radar tracks for the existence and the motion of the possibly enemy objects in the air. Suppose that the ring operates normally and at some instant the radar detects a critical object in the air and needs to position the guns to that possibly enemy object. To perform the positioning, the radar must send the coordinates of the object to the gun controller. So it must send a connection request to the management station and must join the ring. The coordinate data that will be carried during this connection is high priority traffic which has strict real-time guarantees or in other words hard real-time guarantees. If the guns cannot be positioned to the foreign object then the results of this can be disastrous.

To join the ring, the radar sends a connection request to the management station as shown in the Figure 4.2. The source address in the connection request is set to the radar's MAC address, the destination address is set to the station's MAC address controlling the gun, the traffic class to the highest priority class and also sets the message related parameters like P, D and C appropriately. After preparing the connection request, the radar sends this request to the management station.

First of all the management station checks whether it is possible for the radar to join the ring without degrading the QoS guarantees of other stations. If it doesn't degrade then the connection is accepted. But if it degrades the QoS of other stations then the management station needs to execute the station eviction procedure. By performing the station eviction procedure the management station evicts the best possible station in the ring with a lower traffic class such that removing it from the ring will make the management station accept the new connection. Finally the management station evicts a station and removes it from the ring and accepts the new connection and radar joins the ring and can send the high priority coordinate data to the guns and then instruct the controller to position the guns.

These two examples show the necessity of control planes for real-time systems such that a well designed control plane can make a wireless token ring network more responsive to real-time traffic and as a result the network can become more suitable for real-time communications. And also with the help of the control plane the throughput of the network can increase and the network resources can be utilized more efficiently. With the design and implementation of real-time communication protocols and control planes designed for these protocols, better results can be obtained for providing real-time guarantees in wireless networks.

## 5. SIMULATIONS AND RESULTS

As stated in previous sections the problem of real-time communications should be addressed in the MAC and physical layers in wireless networks. In the physical layer, suitable modulation or coding techniques should be designed to make providing real-time guarantees easier for upper layers. Actually most of the research to provide real-time guarantees is done for the MAC layer. Especially since the 802.11 standard is the most widely used standard, researches generally focus on providing better QoS guarantees and real-time guarantees for 802.11 networks.

In addition many token ring protocols for providing real-time guarantees have been proposed. These token ring protocols for wireless networks were reviewed and analyzed in section 2. Nearly all the proposed protocols are based on the WTRP which is actually implemented on top of 802.11 MAC.

Unfortunately none of the proposed protocols is claimed to be appropriate for hard real-time communications and none of them incorporates a control plane that makes the protocol suitable for hard real-time communications. They are just trying to provide soft real-time guarantees or they just claim to guarantee an acceptable level of QoS.

In this thesis we adapted the timed token protocol to wireless networks to support hard real-time communications and proposed a centralized control plane for wireless timed token ring architectures. The primary goal of the control plane is to manage the dynamic wireless token ring network and provide sufficient bandwidth to higher priority traffic in order to satisfy their hard-real time constraints. The reason to use the timed token protocol is that, the timed token protocol has special timing properties and has solid mathematical foundations. It is designed to be used for hard real-time communications and also extensive amount of research has been done on this protocol especially on SBA schemes and as a result it has been well understood in the research community. The main goal of the proposed control plane is to make wireless token ring architectures more suitable to provide real-time guarantees. In order to reach these goals, three important functions are implemented in this control plane, namely the *admission control procedure*, the *station eviction procedure* and a *traffic differentiation* mechanism. The admission control procedure determines whether a connection request should be accepted. This decision is actually based on

the current network state such as the current load, the number of connections established and the class of traffic carried over these connections. If the connection cannot be accepted, the management station executes the station eviction procedure. With this procedure the management station tries to evict the best possible station in the ring with a lower traffic class than the requested connection such that removing it from the ring will make the management station accept the new connection. Finally the management station evicts a station and removes it from the ring and accepts the new connection. The last contribution of the proposed control plane is the traffic differentiation mechanism. The traffic classes as proposed by 802.11e are supported in this control plane. And as a result of the traffic differentiation mechanism and the station eviction procedure the system is expected to be more responsive to the high priority traffic and also the system is expected to be more suitable for real-time communications.

For the simulation studies a discrete event simulation program is designed and implemented with the object oriented programming language C++. The aim of the simulations is to study the impact of the proposed control plane on the wireless token ring networks. To do this both the pure timed token protocol and the timed token protocol with the proposed control plane are studied. And as a result of the simulation studies the effect of the control plane in terms of the system responsiveness and throughput is analyzed.

The simulation program implements the EMCA SBA scheme and the connection request handling in the management station as given in Figure 4.4 and the disconnect request handling in the management station as given in Figure 4.6. And also a traffic generator has been implemented to generate connections according to the Poisson probability distribution. The simplified class diagram of the simulation program is given in Figure 5.1 in UML notation. The *TrafficGenerator* class is responsible for generating connection requests that arrive according to the Poisson probability distribution. These connection requests are then processed by the management station according to the connection request handling algorithm given in Figure 4.4. It is assumed that each connection has a fixed lifetime. The management station keeps statistics with the *SimulationStatistics* class like the total utilization, utilization per traffic class or the number of accepted and rejected connection requests. At the end of the simulation the statistics are dumped to text files for further analysis. The initial members of the ring is read from a text file with their corresponding traffic parameters like period (P), deadline (D), maximum amount of time needed to transmit a message in the stream (C), the initial synchronous bandwidth of the station (H), the load of the synchronous traffic in the station in Mbps, the load of the

asynchronous traffic in the station in Mbps and the period of the asynchronous traffic in the station in Mbps. After reading the initial topology from the file the simulation begins and connection requests are generated according to the Poisson distribution and a disconnection request is generated after a fixed amount of time representing the connection lifetime for each generated connection request.

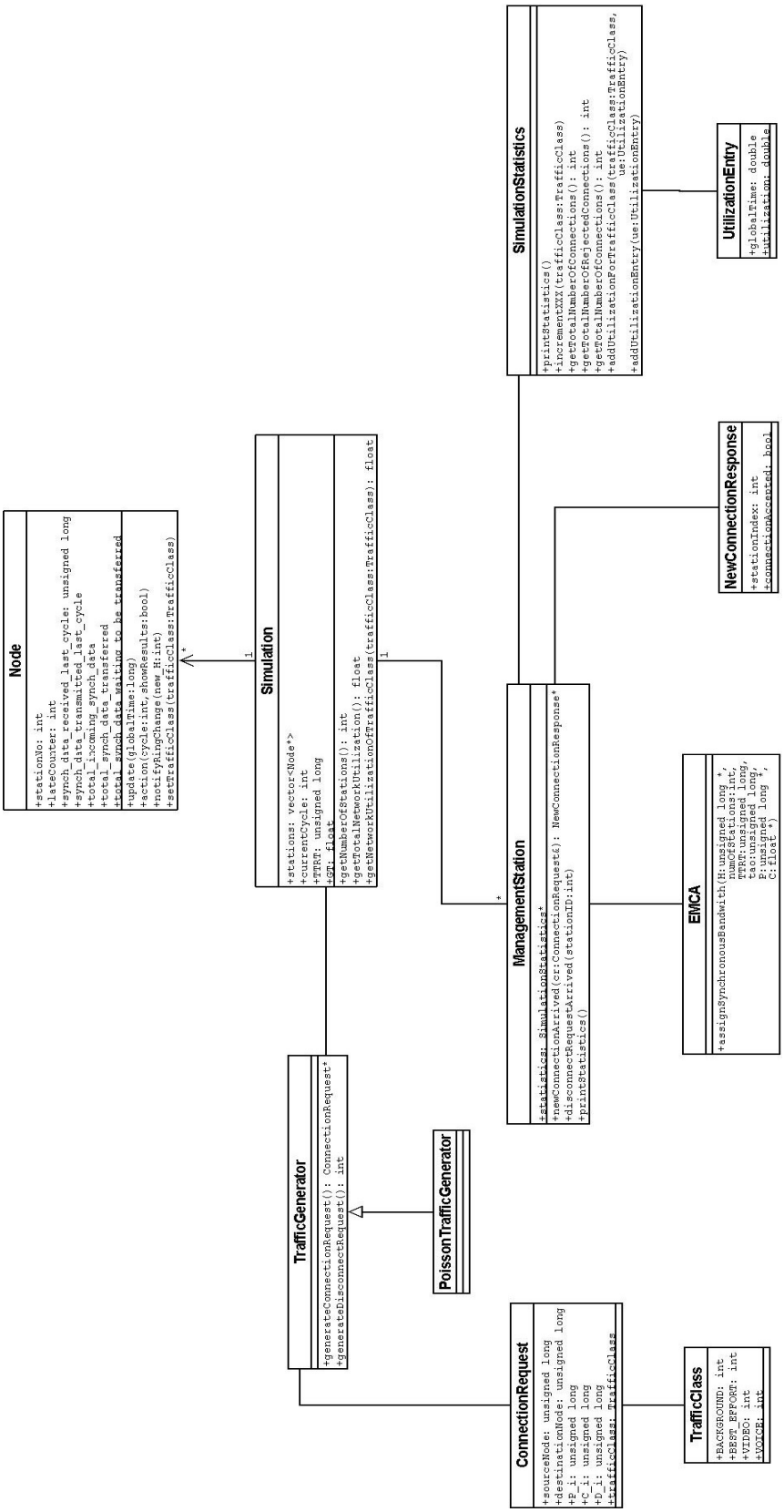


**Figure 5.1:** The Class Diagram of the Simulation Program

## 5.1 Simulation Studies

Two set of simulations have been performed for the simulation studies. In the first set of simulations the connection requests are generated according to the Poisson probability distribution and each connection has the *same* period (P), deadline (D), maximum amount of time needed to transmit a message in the stream (C). In the second set of simulations the connection parameters P, C and D are set according to the traffic that will be carried by the connection. The TTRT is taken as 85 ms and $\tau$ is assumed as 0 during the simulation. In both of the simulations initially the network consists of 4 stations whose traffic parameters are read from a given file. A traffic generator is designed to generate the same number of connections for each of the traffic classes namely BACKGROUND, VOICE, BEST-EFFORT and VIDEO to see the effects of the proposed control plane. After the connections are generated, the management station assigns synchronous bandwidth to these connections with the EMCA algorithm if possible. At the end of the simulations the number of connection requests generated, the number and percentage of accepted and rejected connections, the number of connections accepted by the station eviction procedure and the average percentage of the connection lifetime used per traffic class are analyzed. In addition to this, the network utilization per traffic class is also examined to see the effect of the proposed control plane.

### 5.1.1 First Simulation Study

In this set of simulations the connection requests are generated according to the Poisson probability distribution and each connection has the *same* period (P), deadline (D), maximum amount of time needed to transmit a message in the stream (C). During the first set of simulations 377 connection requests are generated. The statistics about the generated connections is given in Table 5.1. As stated above the traffic generator is designed to generate nearly equal number of connections per traffic class to see the effects of the designed control plane. The column "*Traffic Class*" shows the class of traffic for which the statistics are collected, the "*Total # of Connection Requests*" column is the total number of connection requests generated for the given traffic class. The "*Accepted*" column shows the *total* number of connections accepted. The "*Accepted By Eviction*" column shows the number of connections accepted by performing the station eviction procedure in other words to accept those connections the management station has evicted a station from the token ring network. The "*Rejected*" column shows the number of connections rejected by the control plane. The *"# Of Stations Evicted*" column shows the number of stations which are evicted by the management station as a result of the station eviction

procedure. The "*Average % of Connection Lifetime Used*" column shows the percentage of the connection lifetime that is used by the stations of the given traffic class. This statistic shows how much of the dedicated connection lifetime a station from a given traffic class can use before being evicted. The last two columns show the percentage of connections accepted and rejected respectively.

**Table 5.1:** Connection Related Statistics For The Proposed Control Plane

| Traffic Class | Total # Of Connection Requests | Accepted | Accepted By Eviction | Rejected | # Of Stations Evicted | Average % of Connection Lifetime Used | % Accepted | % Rejected |
|---|---|---|---|---|---|---|---|---|
| BG | 95 | 36 | 0 | 59 | 37 | 19.41 | 37.90 | 62.10 |
| BE | 95 | 42 | 13 | 53 | 38 | 39.42 | 44.21 | 55.79 |
| VIDEO | 95 | 68 | 37 | 27 | 29 | 88.09 | 71.58 | 28.42 |
| VOICE | 92 | 92 | 54 | 0 | 0 | 100 | 100 | 0 |

In the traffic class column the abbreviation BG represents background traffic and the abbreviation BE represents the best-effort traffic.

The connection related statistics shown in Table 5.1 shows that the control plane accepts as much VOICE traffic as possible. 54 of these connection requests are accepted by executing the station eviction procedure. As a result 100% of all the VOICE connection requests are accepted. This result verifies that the system became more responsive to the high priority traffic. Since the VOICE class is the highest priority traffic class, the management station cannot evict any station having a connection carrying voice traffic hence the "*Stations Evicted*" column for this clas is 0. And by looking at the "*Average % of Connection Lifetime Used*" column it seen that since none of the VOICE stations can be evicted these stations can use all of their connection lifetime.

Since the VIDEO traffic is the second highest traffic class, 37 VIDEO connections caused the management station to evict lower priority stations and as a result nearly 72% of the VIDEO connections has been accepted by the control plane. And looking at the "*Stations Evicted*" column we can say that 27 stations carrying VIDEO traffic have been evicted by the management station to accept VOICE connections. Again since the VIDEO class is evicted less, those stations can use much of their connection lifetime before being evicted.

And for the BEST-EFFORT connections the system has evicted 13 stations with BACKGROUND traffic. And to accept higher priority connections, 38 stations carrying BEST-EFFORT traffic have been evicted. As a result 45% of the BEST-EFFORT connections have been accepted. Since this class is evicted by VIDEO and VOICE traffic classes nearly they can use 40% of their connection lifetime before being evicted.

For the BACKGROUND traffic, since this traffic class is the least priority traffic class none of the stations have been removed from the ring to accept these connections hence the "*Accepted By Eviction*" column is 0 for this class. So the percentage of the connections accepted for this class is lower than other traffic classes and is nearly 38 %. And to accept higher priority traffic 37 stations carrying BACKGROUND traffic have been evicted. Since this class can be evicted by all the higher priority classes on average 20% of the connection lifetime can be used by these stations before being evicted. This table verifies that the proposed conrol plane makes the wireless token ring network respect higher priority traffic classes over lower priority traffic classes and as a result makes the system more responsive for high priority traffic classes.

**Table 5.2:** Connection Related Statistics For The Pure Timed Token Protocol

| Traffic Class | Total # Of Connection Requests | Accepted | Rejected | % Accepted | % Rejected |
|---|---|---|---|---|---|
| BACKGROUND | 95 | 16 | 79 | 16.84 | 83.16 |
| BEST-EFFORT | 93 | 16 | 77 | 17.20 | 82.80 |
| VIDEO | 96 | 15 | 81 | 15.62 | 84.38 |
| VOICE | 93 | 17 | 76 | 18.28 | 81.72 |

Table 5.2 shows the connection related statistics for the pure timed token case, in other words the wireless token ring network does not have the proposed control plane. This table shows that higher priority traffic class is treated same as the lower priority traffic classes hence there is no traffic differentiation mechanism to provide the needed QoS guarantees to the wireless stations. And also since there is no station eviction procedure, the high priority traffic cannot evict lower priority traffic classes so after the network saturates no connections can be accepted. These results show that the station eviction procedure makes the wireless token ring network more

responsive to high priority traffic and hence more suitable for providing real-time guarantees.

The network utilization per traffic class with the proposed control plane is shown in Figure 5.2. This graph shows that the utilization of the higher priority traffic classes namely the VOICE and VIDEO traffic classes have higher throughput than the lower priority traffic class. This result also justifies the argument that the control plane makes the system respect higher priority traffic over lower priority traffic classes and as a result of the traffic differentiation mechanism and the station eviction procedure the system tries to admit as much high priority traffic into the network as possible and becomes more responsive to the high priority traffic.

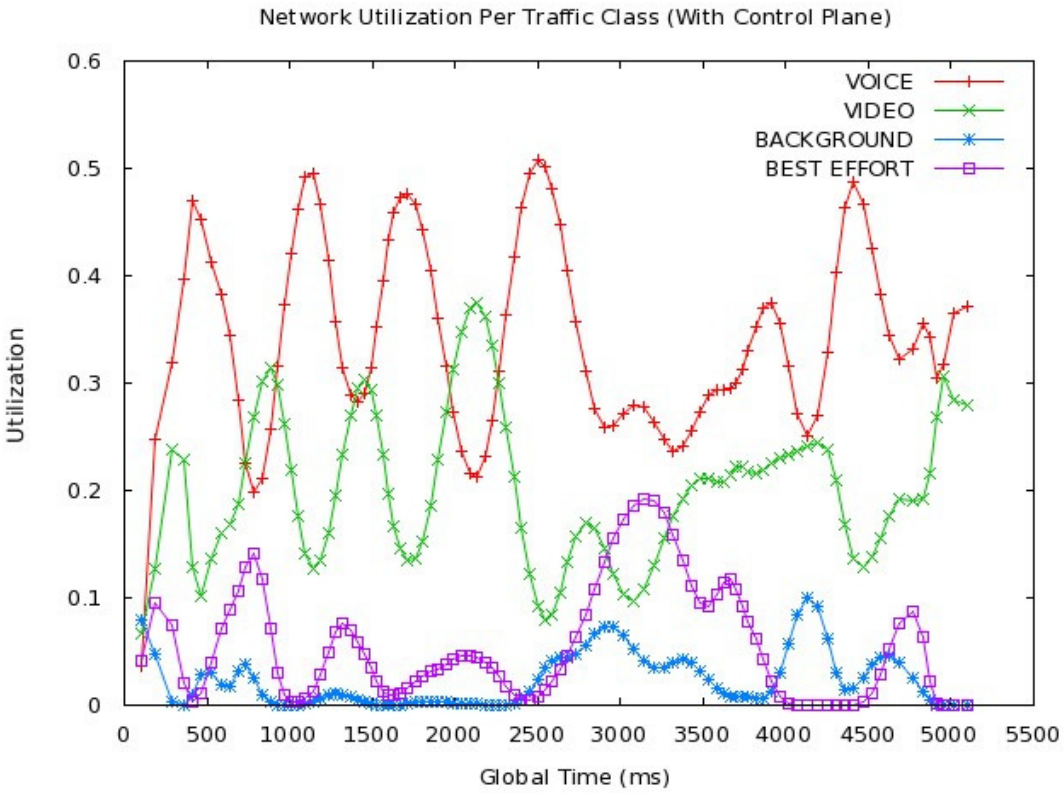


**Figure 5.2:** Total Network Utilization with the Proposed Control Plane

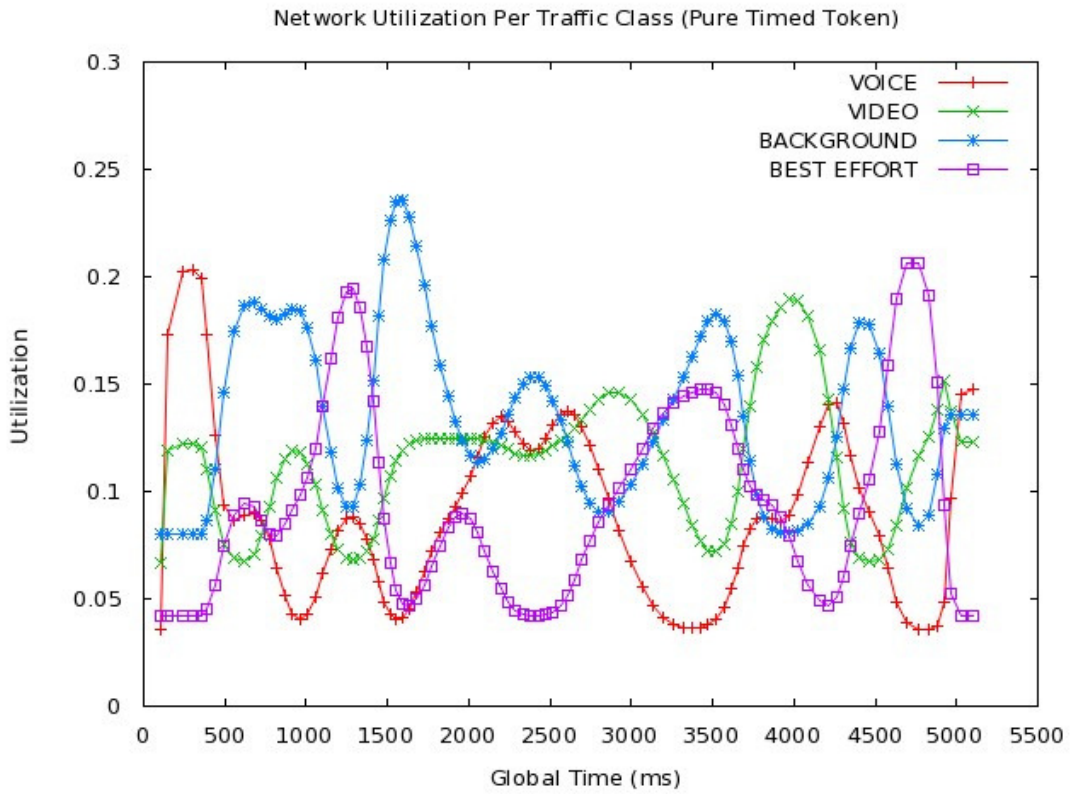


**Figure 5.3:** Total Network Utilization for the Pure Timed Token Protocol

The network utilization per traffic class with the pure timed token protocol is shown in Figure 5.3. The wireless token ring network without the control plane does not respect high priority traffic and hence the system is less suitable for providing QoS guarantees for high priority traffic. The pure timed token protocol does not differentiate among the traffic classes resulting in less throughput for high priority traffic.

Finally the total utilization of the network can be seen in Figure 5.4. As stated at the beginning of this section that there are initially 4 stations in the network in both of the simulations whose effective utilizations ($U_i$) are smaller than the effective utilizations of the new connections. And since the system with the control plane can evict stations, the stations having low effective utilizations can be evicted by the management station and new connections with higher utilizations are admitted into the network. But since the system for the pure timed token case cannot evict the stations with low effective utilization the total network utilization cannot be as high as the system with the control plane.

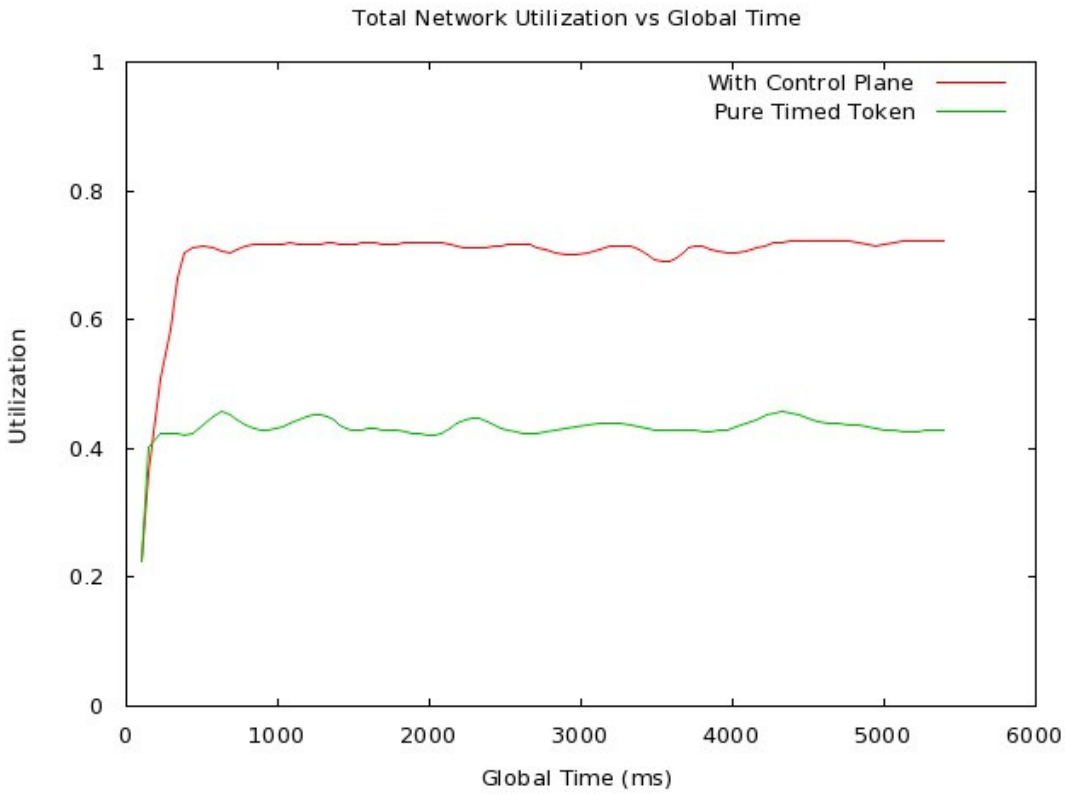


**Figure 5.4:** Total Network Utilization

By looking at the graphs it can easily be said that the control plane admits as much high priority traffic class as possible into the network making the system more responsive to high priority connections and also the total utilization of high priority traffic increases in the network.

### 5.1.2 Second Simulation Study

In this set of simulations the connection requests are generated according to the Poisson probability distribution and each of the traffic parameters of the connections namely P, C and D are set according to the traffic that will be carried by this connection. During the second set of simulations 377 connection requests are generated. The statistics about the generated connections is given in Table 5.3. As stated above the traffic generator is designed to generate nearly equal number of connections per traffic class to see the effects of the designed control plane.

**Table 5.3:** Connection Related Statistics For The Proposed Control Plane

| Traffic Class | Total # Of Connection Requests | Accepted | Accepted By Eviction | Rejected | # Of Stations Evicted | Average % of Connection Lifetime Used | % Accepted | % Rejected |
|---|---|---|---|---|---|---|---|---|
| BG | 90 | 31 | 0 | 59 | 32 | 11.33 | 34.44 | 65.56 |
| BE | 94 | 44 | 9 | 50 | 36 | 38.08 | 46.81 | 53.19 |
| VIDEO | 98 | 66 | 27 | 32 | 20 | 91.79 | 67.35 | 32.65 |
| VOICE | 95 | 95 | 52 | 0 | 0 | 100 | 100 | 0 |

In this set of simulations the period and deadline of the lower priority class are set as larger than the high priorty traffic classes since high priority traffic classes need better real-time guarantees. And also the maximum amount of time needed to transmit a message in the stream (C) in lower priority traffic classes are set as larger than the higher priority traffic classes. This sounds reasonable when we think a voice application and a monitoring application in an industrial control network. Suppose that the monitoring application sends event logs periodically to another station where the messages will be probably larger in size hence having larger C and do not have strict deadline constraints hence larger P and D. The voice packets generated by the voice application will be smaller in size hence smaller C and will have smaller period and deadlines and hence smaller P and D.

The connection related statistics shown in Table 5.3 shows that the control plane accepts as much VOICE traffic as possible with the help of the station eviction procedure. The management station has evicted 52 stations to accept VOICE connections and as a result 100% of all the VOICE connection requests are accepted. This result verifies that the system became more responsive to the high priority traffic. Since the VOICE class is the highest priority traffic class, the management station cannot evict any station having a connection carrying voice traffic hence the "*Stations Evicted*" column for this clas is 0. And by looking at the "*Average % of Connection Lifetime Used*" column it seen that since none of the VOICE stations can be evicted these stations can use all of their connection lifetime.

Since the VIDEO traffic is the second highest traffic class, 27 VIDEO connections caused the management station to evict lower priority stations and as a result nearly 68% of the VIDEO connections has been accepted by the control plane. And looking at the "*Stations Evicted*" column we can say that 20 stations carrying VIDEO traffic

have been evicted by the management station to accept VOICE connections. Again since the VIDEO class is evicted less, those stations can use much of their connection lifetime before being evicted.

And for the BEST-EFFORT connections the system has evicted 9 stations with BACKGROUND traffic. And to accept higher priority connections, 36 stations carrying BEST-EFFORT traffic have been evicted. As a result nearly 47% of the BEST-EFFORT connections have been accepted. Since this class is evicted by VIDEO and VOICE traffic classes nearly they can use 39% of their connection lifetime before being evicted.

For the BACKGROUND traffic case, since this traffic class is the least priority traffic class none of the stations have been removed from the ring to accept these connections hence the "*Accepted By Station Eviction*" column is 0 for this class. To accept higher priority traffic 32 stations carrying BACKGROUND traffic have been evicted. This table also verifies that the proposed conrol plane respects higher priority traffic classes over lower priority traffic classes and as a result makes the system more responsive for high priority traffic classes. Since this class can be evicted by all the higher priority classes on average 12% of the connection lifetime can be used by these stations before being evicted.

**Table 5.4:** Connection Related Statistics For The Pure Timed Token Protocol

| Traffic Class | Total # Of Connection Requests | Accepted | Rejected | % Accepted | % Rejected |
|---|---|---|---|---|---|
| BACKGROUND | 95 | 23 | 72 | 24.21 | 75.79 |
| BEST-EFFORT | 98 | 33 | 65 | 33.67 | 66.33 |
| VIDEO | 94 | 12 | 82 | 12.77 | 87.23 |
| VOICE | 90 | 14 | 76 | 15.56 | 84.44 |

Table 5.4 shows the connection related statistics for the pure timed token case, in other words the wireless token ring network does not have the proposed control plane.

Since low priority traffic classes will have larger P, D and C the probability that they will be admitted to the network increases. As a result connections that will carry lower priority traffic have more accepted connection requests. And also since the network does not have the proposed control plane, higher priority traffic class is

treated same as the lower priority traffic classes. And due to the fact that there is no station eviction procedure, after admitting the lower priority connections the network cannot admit higher prority connections.

The network utilization per traffic class with the proposed control plane is shown in Figure 5.5. This graph shows that the utilization of higher priority traffic classes is more than the lower priority traffic classes. This result justifies the argument that the control plane makes the system respect higher priority traffic over lower priority traffic classes.

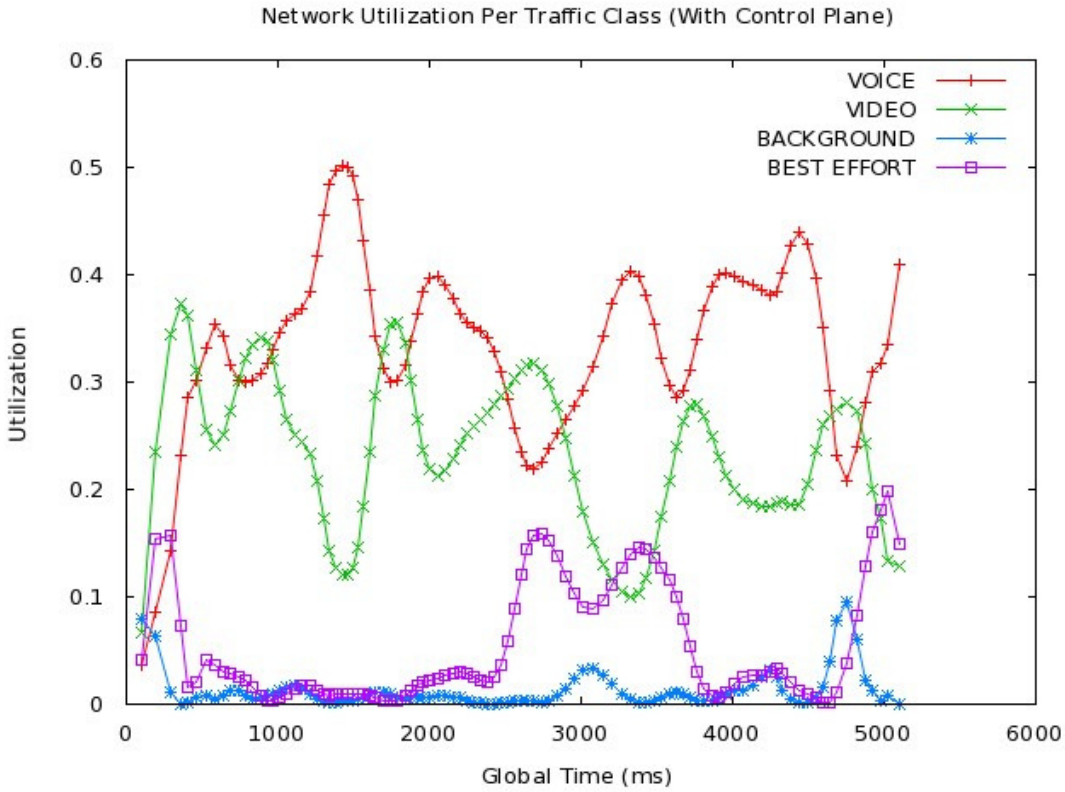


**Figure 5.5:** Total Network Utilization with the Proposed Control Plane

The network utilization per traffic class with the pure timed token protocol is shown in Figure 5.6. The wireless token ring network without the control plane does not respect high priority traffic and treats all requested connections equally and hence the higher priority traffic could not have throughput values as much as lower priority traffic.

By looking at the two graphs in Figure 5.5 and 5.6 it can easily be said that the control plane admits as much high priority traffic class as possible into the network making the system more responsive to the high priority connections and also the total utilization of high priority traffic increases in the network.

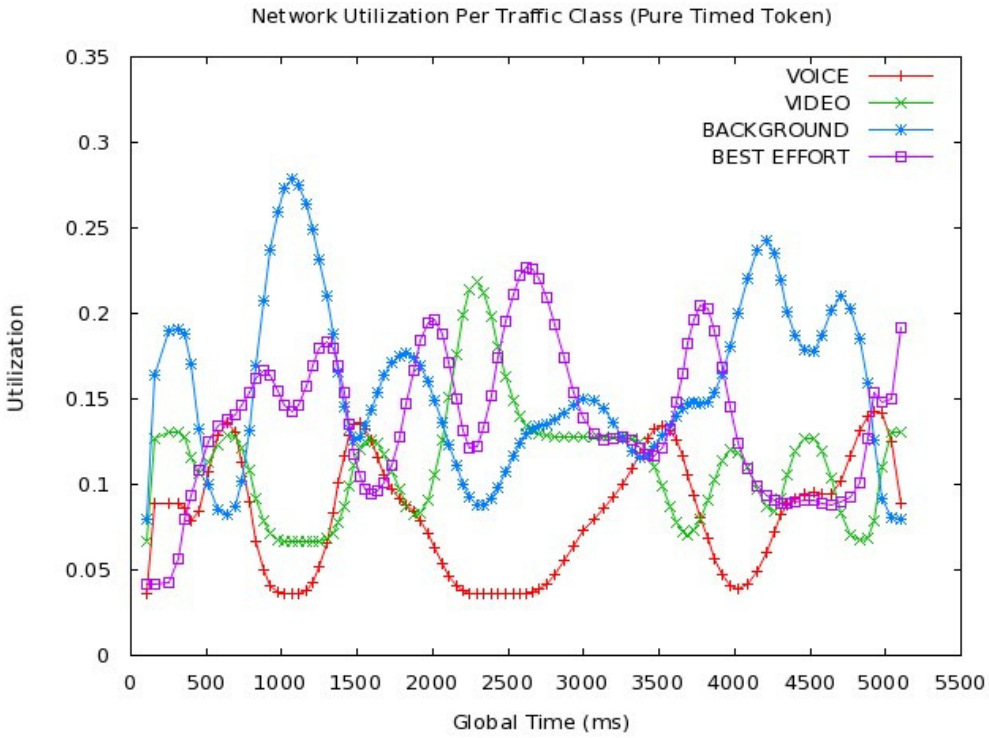


**Figure 5.6:** Total Network Utilization for the Pure Timed Token Protocol

Finally the total utilization of the network can be seen in Figure 5.7. As stated at the beginning of this section that there are initially 4 stations in the network in both of the simulations whose effective utilizations ($U_i$) are smaller than the effective utilizations of the new connections. And since the system with the control plane can evict stations, the stations having low effective utilization can be evicted by the management station and new connections with higher utilization are admitted into the network. But since the system for the pure timed token case cannot evict the stations with low effective utilization the total network utilization cannot be as high as the system with the control plane. The total network utilization is lower than the total utilization for the first simulation study since during the second simulation studies the connections have variable P, D and C values and so the effective utilizations of high priority connections are less than lower priority connections. As a result of the control plane these higher priority connections are admitted into the network as much as possible making the total network utilization less than the first simulation studies.

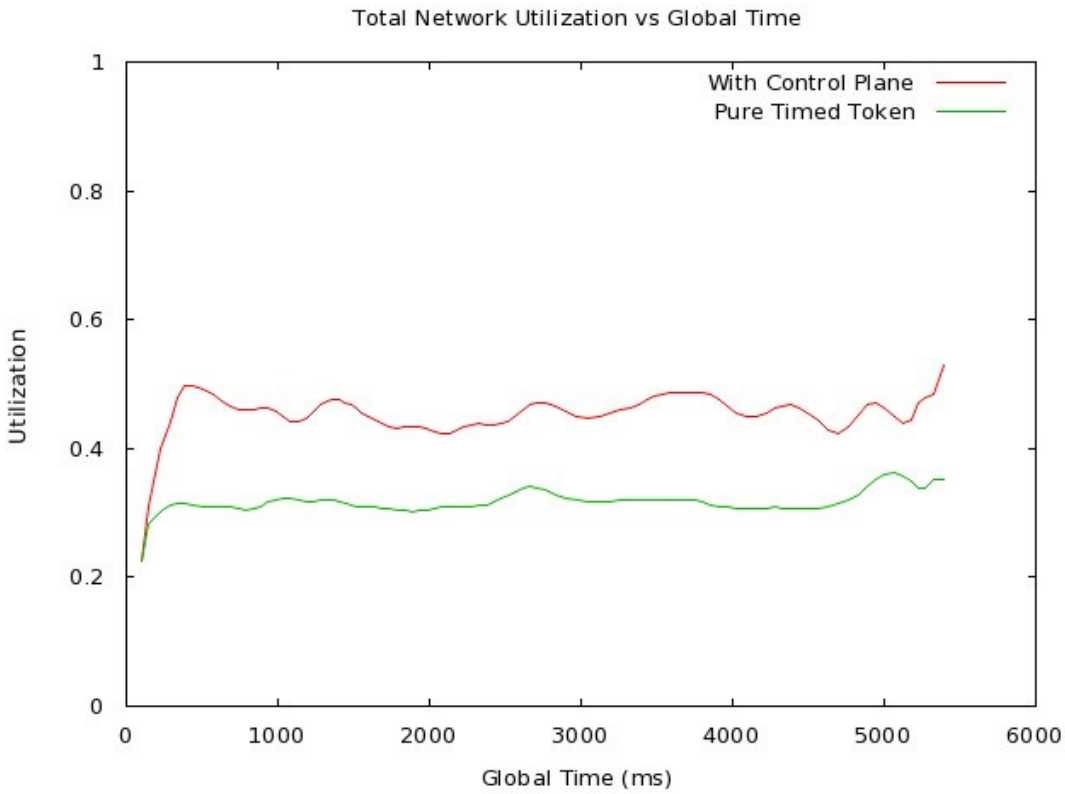


**Figure 5.7:** Total Network Utilization

## 6. CONCLUSIONS

In this thesis we adapted the timed token protocol to wireless token ring networks to provide hard real-time guarantees and also proposed a centralized control plane for wireless timed token ring architectures. The primary goal of the control plane is to manage the dynamic wireless token ring network and provide sufficient bandwidth to higher priority traffics in order to satisfy their real time constraints. And the reason to use the timed token protocol is that, the timed token protocol has special timing properties and has solid mathematical foundations. It is designed to be used for hard real-time communications and also extensive amount of research has been done on this protocol especially on SBA schemes and as a result it has been well understood in the research community. And also the upper bound for the delay is known and hence the system becomes predictable which is the primary requirement for real-time systems. Sevcik and Johnson have proved that the maximum token rotation time in FDDI is 2TTRT [21,22]. Later Chen and Zhao generalized this theorem and proved that the maximum time that can elapse between ν consecutive token arrivals at a node i is bounded by

$$(v-1).TTRT + \sum_{h=1}^{n} H_h - H_i + a \qquad \textbf{(6.1)}$$

where H is the synchronous bandwidth allocated to the node and a is the fraction of TTRT that is unavailable for synchronous message transmission [23,24].

For the timed token protocol, a SBA scheme must also be adopted to allocate synchronous bandwidth to the stations in the network. In the proposed control plane, EMCA is used as the SBA scheme.

Three important functions are implemented in the proposed control plane, namely the *admission control procedure*, the *station eviction procedure* and a *traffic differentiation* mechanism. The admission control procedure determines whether a connection request should be accepted. This decision is actually based on the current network state such as the current load, the number of connections established and the class of traffic carried over these connections. If the connection cannot be accepted, the management station executes the station eviction procedure. With this procedure the management station tries to evict the best possible station in the ring with a lower traffic class than the requested connection such that removing it from the ring will make the management station accept the new connection. Finally the management station evicts a station and removes it from the ring and accepts the new connection.

The last contribution of the proposed control plane is the traffic differentiation mechanism. The traffic classes as proposed by 802.11e are supported in this control plane. And as a result of the traffic differentiation mechanism and the station eviction procedure the system becomes more responsive to the high priority traffic and makes the system more suitable for real-time communications.

The simulation results justify that the proposed control plane ensures higher priority traffic more bandwidth than lower priority traffic and guarantees that deadline constraints of hard real-time traffic are satisfied. The wireless token ring network becomes more responsive to the connection requests carrying high priority traffic. As a result of the simulations it is also seen that the proposed control plane increases the throughput of higher priority traffic in the network. All of these results show that by adapting the timed token protocol and utilizing the proposed control plane the wireless network becomes suitable for real-time communications.

## 6.1 Suggestions for Improvement and Future Work

The proposed control plane utilizing the timed token protocol is a new research topic targeting wireless token ring networks. So the proposed system can be improved in various ways. By enhancing the control plane, wireless token ring networks can be made much more suitable for providing real-time guarantees.

The proposed system has a central management station to make admission control decisions and execute station eviction procedure. So to make the system more scalable the proposed algorithms can be made decentralized. By making the system a decentralized one, larger token rings can be supported easily. And also since it is assumed that the control plane is designed for small scaled ad-hoc networks, by making the system decentralized it will be possible to support large scale token ring networks.

Another problem that is not addressed in this thesis is the mobility problem. Making the system suitable for supporting mobility is an open challenge. Probably the most challenging part of supporting mobility is to keep the guaranteed quality of service level intact during hand off procedures. The mobility scenario will include more than one token ring and hence this will also bring the problem of multiple ring management. Consider the multi ring topology in Figure 6.1. There is a gateway connecting the two wireless token rings. This gateway can be a base station or another dedicated network element. Suppose that the STATION A in the figure wants to establish a connection to the STATION B in another ring. The problem gets more challenging because in order to make an admission control decision,

knowledge of the state of both wireless token rings is required. And the control plane should not degrade any station's real-time guarantees in any of the token ring networks by accepting a connection spanning more than one wireless token ring. And also in these multi-ring network topologies latency guarantees are harder to determine.

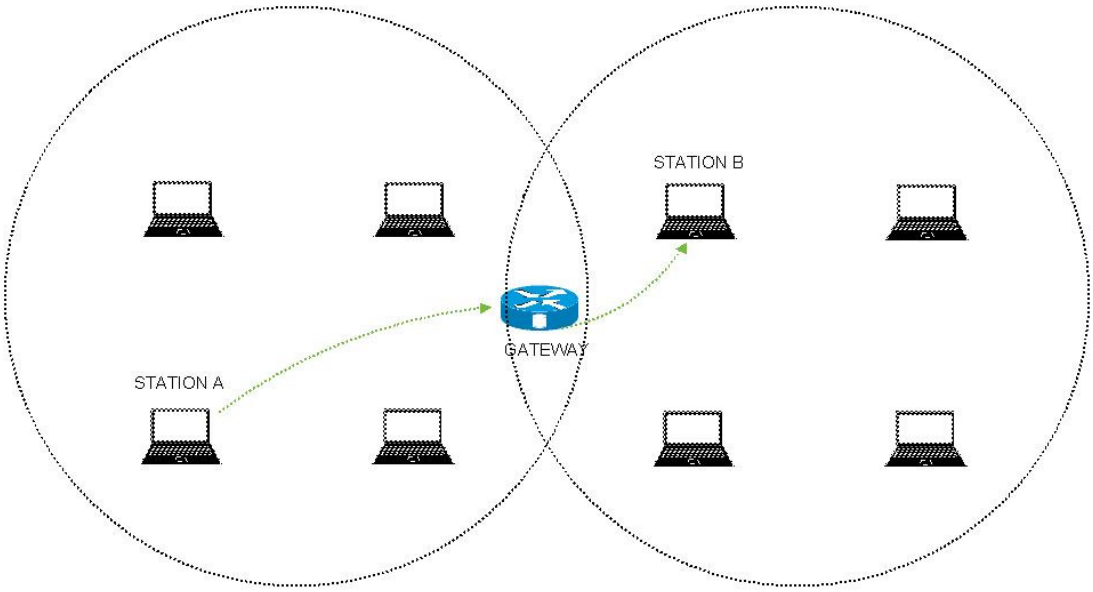


**Figure 6.1:** A Multi-Ring Network Topology

Moreover in recent years there has been tremendous research in wireless sensor networking. WSNs (Wireless Sensor Network) are being deployed for various applications like habitat monitoring, object tracking, nuclear reactor controlling, fire detection and traffic monitoring. With the availability of low-cost small-scale imaging sensors, CMOS cameras, microphones, which may ubiquitously capture multimedia content from the field, Wireless Multimedia Sensor Networks (WMSN) have been proposed and drawn the immediate attention of the research community [6]. So it is an open problem to make the proposed control plane more suitable for wireless multimedia sensor networks. To do this the system should be made decentralized and the proposed algorithms should be made power aware.

## REFERENCES


[1] **Stankovic, J.**, 1958. Real-Time Computing Systems: The Next Generation, *Proceedings of Computer Communications and Networks,* February 1958.

[2] **ITU-T G.1010 Series G**, 2001. Transmission Systems and Media, Digital Systems and Networks, Quality of Service and Performance End-User Multimedia QoS Categories.

[3] **Kurose, J., Schwartz, M. and Yemini, Y.**, 1984. Multiple-Access Protocols and Time-Constrained Communication, *ACM Computing Surveys (CSUR),* **16**, 43-70.

[4] **Aras, C.M., Kurose, J.F., Reeves, D.S. and Schulzrinne, H.**, 1994. Real-time Communication in Packet-Switched Networks, *Proceedings of IEEE,* **82**, 122-139.

[5] **Sajal K. and Chatterjee, M.**, 2001. Challenges in Wireless Multimedia Networks, *Proceedings of the International Conference on Information Technology for the New Milennium (IConIT),* 2001.

[6] **Akyıldız I., Melodia, T. and Chowdhury, K.**, 2007. A Survey on Wireless Multimedia Sensor Networks, *The International Journal of Computer and Telecommunications Networking*, **51**, 921-960.

[7] **Ergen, M.**, 2002. WTRP-Wireless Token Ring Protocol, *MSc Thesis,* University of Berkeley, California.

[8] **Ergen, M., Lee, D., Varaiya, P. and Sengupta, R.**, 2004. Wireless Token Ring Protocol, *IEEE Transactions on Vehicular Technology*, **53**, 1863-1881.

[9] **Ergen, M., Lee, D., Puri, A., Varaiya, P., Sengupta, R., Attias, R. and Tripakis, S.**, 2002. Wireless Token Ring Protocol, SCI Orlando, July, 2002.

[10] **Cheng, R., Chang, R. and Hua, K.**, 2007. IWTRP: Spatial-Reuse Enhancement of the Wireless Token Ring Protocol, *IEEE Communications Letters,* **11**, 701-703.

[11] **Sun, X., Zhang, Y. and Li, J.**, 2007. Wireless Dynamic Token Protocol for MANET, *Proceedings of the 2007 International Conference on Parallel Processing Workshops,* 2007.

[12] **Sharma, S., Gopalan, K., Zhu, N., De, P., Peng, G. and Chiueh, T.**, 2001. Quality of Service Guarantee on 802.11 Networks, *The Ninth Symposium of High Performance Interconnects,* September.

[13] **Hou, W., Liu, W. and Fei, M.**, 2006. A Token-Based MAC Oriented Wireless Industrial Control Networks, *Proceedings of the 2006 IEEE International Conference on Information Acquisition,* August.

[14] **Johnson, E., Anaya, G., Tang, Z., Balakrishnan, M., Zhang, H. and Sreepuram, S.**, 2004. Performance of the HF Token Protocol, *Proceedings of MILCOM 2004,* November.

[15] **Johnson, E., Tang, Z., Balakrishnan, M., Zhang, H. and Sreepuram, S.**, 2003. Robust Token Management for Unreliable Networks, *Proceedings of MILCOM 2003,* October.

[16] **Lee, D.**, 2001. Wireless Token Ring Protocol, *MSc Thesis,* University of Berkeley, California.

[17] **Deng, Z., Lu, Y., Wang, C. and Wang, W.**, 2004. EWTRP: enhanced wireless token ring protocol for small-scale wireless ad hoc networks, International Conference of Communications, Circuits and Systems, **1**, 398-402.

[18] **Maniezzo, D., Pau, G., Gerla, M., Mazzini, G. and Yao, K.**, 2002. T-MAH: A Token Passing MAC Protocol for Ad-Hoc Networks, *MedHocNet2002,* September 2002.

[19] **Moraes, R., Vasques, F., Portugal, P. and Fonseca, J.A.**, 2007. VTP-CSMA: A Virtual Token Passing Approach for Real-Time Communication in IEEE 802.11 Wireless Networks, *IEEE Transactions on Industrial Informatics,* **3**, 215 – 224.

[20] **Grow, R.**, 1982. A Timed-token Protocol for Local Area Networks, *Electro 82,* May 1982.

[21] **Johnson, M.**, 1987. Proof That Timing Requirements of the FDDI Token Ring Protocol Are Satisfied, *IEEE Transactions on Communications,* **35**, 620-625.

[22] **Johnson, M.**, **and Sevcik, K.**, 1986. . Cycle Time Properties of the FDDI Token Ring Protocol, *Joint International Conference on Measurement and Modeling of Computer Systems,* 1986.

[23] **Chen, B. and Zhao, W.**, 1992. Properties of the Timed Token Protocol, *Real-time Systems Group Technical Report,* **92-038,** Texas A&M University.

[24] **Hanssen, F. and Jansen, P.**, 2003. Real-time Communication Protocols: an overview, *Centre for Telematics and Information Technology Technical Report,* **TR-CTIT-03-49,** University of Twente.

[25] **Buzluca, F.**, 1997. Design of an Efficient Real-time Communication Structure for FDDI Based Network System, *PhD Thesis,* İTÜ The Institute of Science and Technology, Istanbul.

[26] **Zhang, S. and Lee, E.**, 2000. The Nonoptimality of Synchronous Bandwidth Allocation Schemes for the Timed Token Protocol, *IEEE Communications Letters,* **4**, 101-103.

[27] **Buzluca, F. and Harmanci, E.**, 2001. Dynamic Synchronous Bandwidth Allocation Scheme for Hard Real-Time Communication in FDDI Networks, *IEE Proceedings of Computers and Digital Techniques,* **148**, 15-22.

[28] **Zhang, S. and Burns, A.**, 1995. An Optimal Synchronous Bandwidth Allocation Scheme for Guaranteeing Synchronous Message Deadlines With the Timed-token MAC Protocol, *IEEE/ACM Transactions on Networking,* **3**, 729-741.

[29] **Zhang, S. and Burns, A.**, 1995. Timing Properties of the Timed Token MAC Protocol, *Proceedings of Computer Communications and Networks,* September 1997.

[30] **Markowski, M.**, 1995. Design and Analysis of Wireless Real-Time Data Link Layer Protocols, *PhD Thesis,* University of Delaware, USA.

## CURRICULUM VITAE

Mahmut Nezih YİĞİTBAŞI (Oct. 04,1984) has received his BSc degree from the Computer Engineering Department of Istanbul Technical University with honors in 2006. He has been working as a software engineer in a private company responsible for the development of telecommunications software. His research interests are in the areas of computer networking, distributed systems and grid computing.