## MINIREVIEW ARTICLES

# Latency Assurances for Multi-Class Traffic at the QoS-Capable Home Network Wireless Access Point (Tutorial)*

**Chaiwat OOTTAMAKORN[1] and Dennis BUSHMITCH[2]**

*School of Engineering and Resources[1], Walailak University, Thasala, Nakhon Si Thammarat 80160, Thailand.*
*Panasonic Information and Networking Technologies Laboratory[2], NJ, USA.*

## ABSTRACT

Emerging broadband networks and technologies have changed the way of home network provision. Provisioning the broadband services has merely accelerated the complexity of home networks to grow. The services demanded by applications are still inadequate under current home network infrastructures. These services require different Quality of Service (QoS) requirements from the networks. As wireless-based communication within the home (e.g. 802.11×) gains wide acceptance because of its mature protocol, becoming the predominant technology for home networking, the question of efficient QoS-aware wireless access point resource allocation and scheduling function implementation arises. In this article we present a resource management and scheduling algorithm for a QoS-aware wireless access point for use in enterprise and home network solutions. The presented algorithm provides delay and packet loss service guarantees to a diverse set of application traffic classes, produced as a result of convergence of home entertainment and data networks. The algorithm is simple, controllable, and scalable for implementation with the support of both absolute and relative QoS guarantees to multi-class traffic.

**Key words:** Home networks - Quality of services - Wireless communications

## INTRODUCTION

Recently, emerging broadband networks and technologies have changed the way of home network provision. For years, the well-known "last mile" problem hindered customers from accessing broadband services such as high-speed digital Internet access, Internet telephony, smart homes or buildings (Remote maintenances), surveillance systems, power monitoring/management, etc. However, with the novel network infrastructures such as the high-speed data transfer-thru-telephone techniques

---

*This paper is the tutorial version of the paper which is considering to be published in Journal of Communications and Networks, 2005.

(ADSL, SDSL, and VDSL), the fiber-to-the-home (FTTH), and even the wireless local loop, the problem is alleviated. On the other hand, provisioning the broadband services has merely accelerated the complexity of home networks to grow. In addition, the services demanded by applications are still inadequate under current home network infrastructures.

To support new broadband services at home, future digital home networks are expected to provision Quality of Service (QoS). These services require different QoS requirements from the networks. One important parameter of QoS requirements is the delay through the network experienced by the services. Based on a delay requirement, services can be classified into two categories: non-real-time and real-time. Non-real-time service such as e-mail, file transfer (FTP), and web services (HTTP) do not explicitly specify bounds on the delay through the network other than the traffic should be sent with no loss and error. The traditional Internet provides best-effort service and does not explicitly guarantee any QoS to connections. This network works well with non-real-time service. On other hand, popular devices at home networks generate real-time multimedia data such as High Definition TV (HDTV), video conferencing, high speed data transfer, and video on demand that have different constraint requirements in bandwidth, packet delay, jitter, and packet loss. In order to meet this demand, the networks need to provide certain QoS guarantees. Therefore, the key to ensuring the co-existence of diverse applications in such IP-based home networks is QoS provisioning of networking resources within the home networking domain.

As wireless-based communication within the home (e.g. 802.11×) gains wide acceptance because of its mature protocol, becoming the predominant technology for home networking, the question of efficient QoS-aware wireless access point resource allocation and scheduling function implementation arises. There are several research groups working on the specification of the IEEE 802.11e, new wireless LAN architecture, in order to support additional QoS provisions (1,2). Initially, IEEE 802.11 wireless networks operate at the 2.4 GHz band and provide data rates up to 2 Mb/s without QoS provisioning standard. Later, the demand of real-time applications has made the QoS requirement the key to success for future multimedia wireless-based home networks. Besides the QoS support, the unlicensed-5-GHz wireless standard (IEEE 802.11a) has the ability to provide up to 54 Mb/s using the orthogonal frequency-division multiplexing (OFDM) technique at the physical layer (3).

The key question in the design, performance evaluation and management of a QoS-capable IP-based wireless network is the implementation of resource management and scheduling algorithms at wireless access points. Addressing the following problem: given a set of different QoS requirements from a number of traffic classes, how should we implement resource management and scheduling algorithms in order to assure QoS requirements without the knowledge of arriving traffic characteristics? In addition, the implemented algorithm has to be able to provide controllable differentiated QoS and be simple, reliable, and scalable for implementation reasons. Our solution has two key techniques:

1) Due to the lack of arrival traffic knowledge, the passive monitoring mechanism is exploited to monitor the arriving traffic and obtains its predictive characteristics.
2) For resource management, the service rate at the output link of wireless access point is adaptive depending on the characteristics of predictive performance parameters and service requirements.

The rest of the article is organized as follows: We first introduce the QoS models and its generalized requirements. We provide an overview of our system model. We then discuss on how to achieve the QoS provisioning.

**The QoS Models**

Providing the requisite QoS guarantees to different types of applications in the Internet has been an important problem in research and commercial communities in the last decade. The explosive growth of the Internet has put more requirements on QoS architectures. In order to provide Quality of Service (QoS), there are two proposed architectures by the Internet Engineering Task Force (IETF). First, the Integrated Services (IntServ) architecture developed in the mid 1990s provides *per-flow* QoS guarantees where each flow is guaranteed and isolated, and *absolute* QoS, such as packet delay or packet loss, with a traffic policer to enforce traffic flow to conform to a set of prescribed traffic parameters. IntServ architecture assures that each flow will receive its isolated service by expressing its QoS guarantee in term of absolute bounds (e.g. delays of packets from a flow $\leq 3$ msec). In order to achieve the per-flow QoS guarantees, IntServ uses a signaling protocol to set up resource reservations for each flow at all network nodes along the path. Since this approach requires all nodes to have a per-flow state and to process per-flow resource reservations, it has known scalability problems. Differentiated Services (DiffServ) is another approach to provide QoS, but without the restrictions on scalability. DiffServ architecture can resolve the scalability problem by providing the QoS guarantees on the traffic aggregate level (*per-class* QoS). This network traffic is grouped into a service class. This classification is used to provide differentiated distributed services among network traffic. This network nodes then only need to implement per-hop scheduling and buffering mechanisms and apply them based on the class index presented in the header of arriving packets. The end-to-end QoS service across several domains is built by combining the per domain behaviors (PDBs) of individual DiffServ domains. These PDBs, in turn, are constructed by an appropriate deployment of the per-hop forwarding behaviors (PHBs). Several DiffServ domains can exist within a single autonomous system. In DiffServ architecture, the QoS guarantees are usually relative QoS which do not specify any exact quantitative guarantees or any absolute bound on services but are expressed in terms of qualitative differentiation between classes. In other words, guarantees for traffic classes are specified in terms of guarantees received by other classes, e.g. Class-1 delay $\approx$ 0.2Class-4 delay. In IEEE 802.11e standard, QoS provisioning is realized by adding a new medium access mechanism. Stations are categorized into eight classes to allow differentiated distributed access to wireless mediums. In addition, IEEE 802.11e defines four channel access classes for traffic from each station to support service differentiation among the traffic. The channel access classes are shown in **Table 1** (2).

From the two proposed architectures, it appears that there is a trade-off between the complexity of implementation and the strictness of the QoS guarantees. In IntServ, the *strong* QoS support can be provided while the complexity of the implementation prohibits the system to become scalable. On the other hand, DiffServ provides *weak* differentiated service solution to the scalability problem with its simplified implementation. For some applications, the relative QoS guaranteed by DiffServ may not be suitable because of the dependency of the services among traffic classes, that is, the deteriorated service for a class will affect the services for the other classes. The challenge of networking at home is to support different types of QoS for different applications in a way that is simple, reliable, and inexpensive. To achieve this goal of supporting various types of QoS guarantee, we provide technical descriptions on the simple, scalable, and controllable resources management algorithm that can be built at a wireless access point with both absolute and relative service differentiations.

**Table 1.** Channel access classes for traffic from each station in IEEE 802.11e (2).

| Access Class | Designation |
|---|---|
| 0 | Best Effort |
| 1 | Video probe |
| 2 | Video |
| 3 | Voice |

**The System Model**

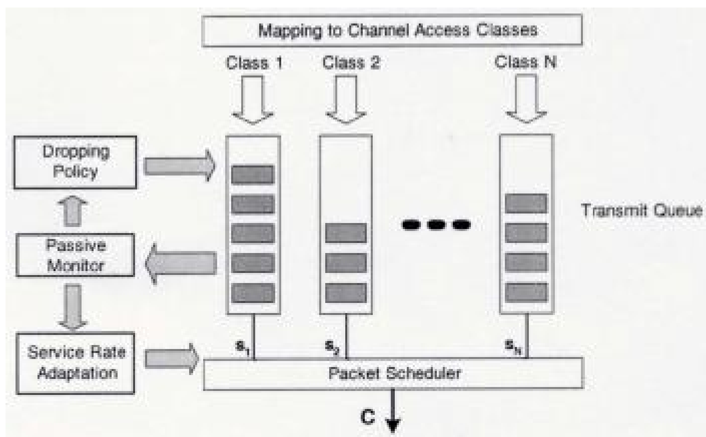



**Figure 1.** An output link of a wireless access point.

We focus on the resource management and scheduling algorithm at an output link of a wireless access point shown in **Figure 1**. At the wireless access point's input, arriving traffic is considered to belong to one of the traffic classes, labelled Class 1, Class 2, ..., Class N. For example, in IEEE 802.11e's differentiation service architecture, $N = 4$. In our notation, the lower the class index is, the better the quality of services provisioned by the access point. This means that traffic is classified and differentiated according to its QoS requirements. We consider the QoS requirements most appropriate for the multimedia traffic: delay. The QoS requirements of each class can be absolute or relative. The framework consists of N prioritized *First-Come-First-Served* (FCFS) queues, each with a finite buffer size. Traffic in each queue $i$, will be served with a service rate $s_i$, where the sum of all $s_i$ = the total link capacity $C$. As shown in **Figure 1**, arriving packets are first queued in a corresponding QoS class queue or dropped if the queue is full. In order to provide QoS guarantees, there are two key components: Passive Monitoring and Service Rate Adaptation. The system passively monitors arriving traffic for each queue to capture its characteristics and calculate the predictive system performance. The system then performs observation-based control of the service rate adaptation algorithm using measured performance and QoS requirement parameters. The goal of the adaptation algorithm is to determine how much service rate is needed for each queue to meet its latency target. If it is not possible to determine, the dropping policy will kick in.

## RESOURCE MANAGEMENT

In this section we give an overview of the details in achieving QoS differentiations (both absolute and relative). Let us consider a transmission queue in **Figure 1**. In **Figure 1**, we assume that the packet scheduler is a work conserving scheduler whose property is never idle if there is traffic waiting in the transmission queue. We consider a discrete system where time is measured in slots of equal length numbered 1, 2, …. QoS guarantees will be performed just when there is a backlog in the system. Therefore we consider each busy period independently, where there is traffic backlogged in the transmission queue. The beginning of a busy period of the queue is assumed to be one, that is, the queue is empty at the beginning of busy period. Upon a packet arrival at a transmission queue, the arrival curve is the aggregate arriving traffic in the time interval since time slot one. The queuing curve is the total amount of traffic in the queue after dropping some traffic over an interval since time slot one. Traffic in the queue is served with a service rate. The service curve is the total amount of traffic service over a time interval. It is noted that the slope of the service curve is the service rate. **Figure 2** illustrates the representations of the curves.

From **Figure 2**, the delay of traffic arriving in time slot $n$ can be determined graphically by the horizontal distance between the arriving and the service curves the backlog at the beginning of time slot $n$ is determined by the vertical distance between the arriving and the service curves. Therefore if we know all arrival curves for all queues, given a set of delay QoS requirements, we can easily determine a set of adaptive service rates which can guarantee not only latency of a particular class but also to satisfy interclass relationship.

***Passive Monitoring*** - Since there is no admission control and no policing mechanism for the arriving traffic in the proposed framework, such traffic's characterization is not previously known. In order to characterize the aggregate arriving traffic, we propose to implement a simple online moving-window measurement scheme at each queue input. The measurement can be done in the following way. Again we divide time into slots with equal length. For each slot, we measure the amount of arriving traffic. At the beginning of a slot, we calculate a predictive arrival curve using the statistical traffic characteristics measured and determined from previous time slots (4). In a straightforward approach, based on the predictive arrival curve, delay QoS requirements, and the system's constraints, we can determine an effective service rate for satisfying each class's QoS requirements. We develop a service rate adaptation algorithm by forming a set of linear equations. The equations can easily be solved in order to obtain the service rate for each class. As seen, one of the aspects of our scheme is to predict the QoS parameters in advance so that the queuing system can prepare sufficient resources to serve each traffic class to assure its QoS requirements.

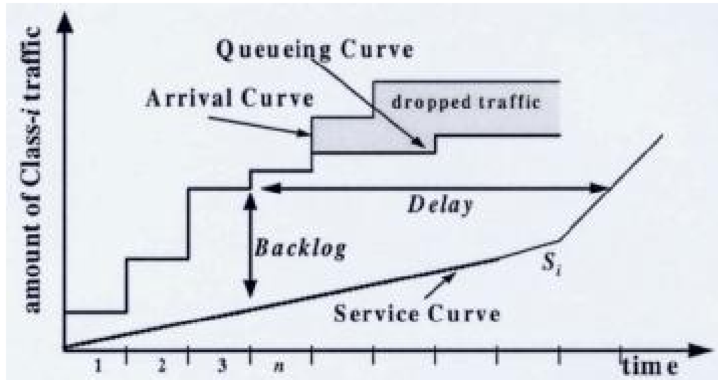
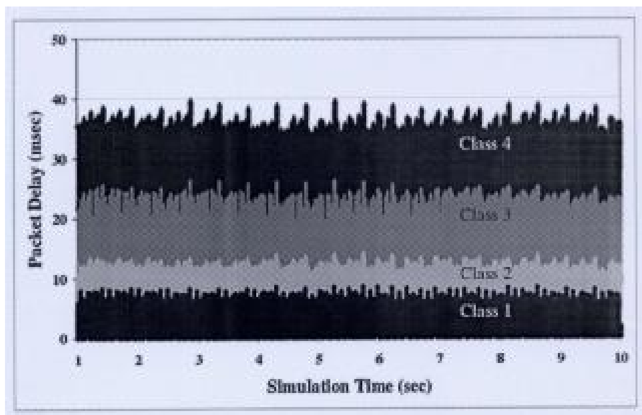



**Figure 2.** Traffic curves and system performance parameters.

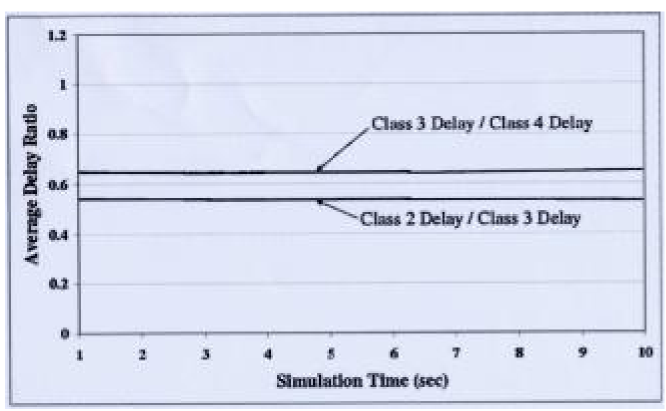
***Service rate adaptation and packet scheduling*** - With a set of service rates for each class needed to be adaptive over time, we will need to implement the following packet scheduling to meet the control. Again each Class-*i* FCFS Queue is served at the rate $s_i$. Traffic from each queue is scheduled with its respective service rate $s_i$ at the scheduler with a link capacity *C*. The algorithm can be described as follows: Initially, for each queue ($i$ = 1, 2, ..., $N$), the accumulative service duration ($\psi_i$) is determined by $\psi_i = L_i / s_i$, where $L_i$ is the length of the first packet in Queue *i*. The server will serve a packet from a non-empty Queue *i* with minimum $\psi_i$. After the transmission of the packet, the $\psi_m$ for the next packet in the queue is recalculated by $\psi_m \longrightarrow \psi_m + L_m / s_m$, where $L_m$ is the length of the next packet in Queue *m*. Again the server will serve a packet from a non-empty Queue *m* with minimum $\psi_m$ and so on. According to this algorithm, traffic stored at each queue will be served relative to the accumulative service duration ($\psi_i$) that is, the traffic in Class-i queue will be served with the rate C if the queue is not empty and $\psi_i$ is the smallest among all classes. As the result, each queue *i* will be served with an effective rate $s_i$.

## SIMULATION RESULTS

All the simulations are implemented using the OPNET modeler simulation framework. In simulations, we consider the topology consisting of two wireless access gateways, each with an output link capacity C = 54 Mbps (e.g. IEEE 802.11e), and four traffic classes ($N$ = 4). We lay emphasis on highly bursty multimedia traffic loads by using four MPEG-4 compressed video traces as a model for traffic arrivals (5). It is assumed that all the flows from the same class are homogenous, that is all the flows from the same class are of the same video trace. The frame rate for all traces is 25 frames per second. The starting time of each trace was randomly chosen over a frame time period. The total traffic load is between 60 - 120 percent of the link capacity. It is also noted that there is cross traffic from Class 1 and 2 about 30% injected into the second access point. We used the simulation parameters as follows: an absolute delay bound, 10 msec, for Class-1 requires, the ratio packet delay between Class 2 and Class 3 of 0.5 and 1.0 at the first and second access points respectively, and the ratio packet delay between Class 3 and Class 4 of 0.67 at both access points. The simulation results are shown in **Figure 3 and 4**. From the plots we observe that 1) the individual delays of Class 1 packets are not greater than the delay bound, i.e. 10 msec (see **Figure 3a and 4a**). With a certain number of Class-1 flows the queue can assure the delay bound requirements. It is noted that the delay bound requirement may not be satisfied when one increases the number of Class-1 flows, since there may not be enough service rate to provision the QoS requirements. 2) the queue can also efficiently guarantee the relative delay requirements during load fluctuation. In **Figure 3b**, the average delay ratio of Class-2 and 3 is 0.52, which is close to the target delay requirement. And the graph also shows the ratio of Class-3 and 4 meets the target, which is 0.67. In addition, **Figure 4b** shows the delay differentiations among the classes, i.e. the average delay ratio of Class-2 and 3 at the second access point is about 1.0, close to the target delay requirement. We also found that the traffic at the second access point is less bursty, that is, the arriving traffic has already been shaped by the first access point.

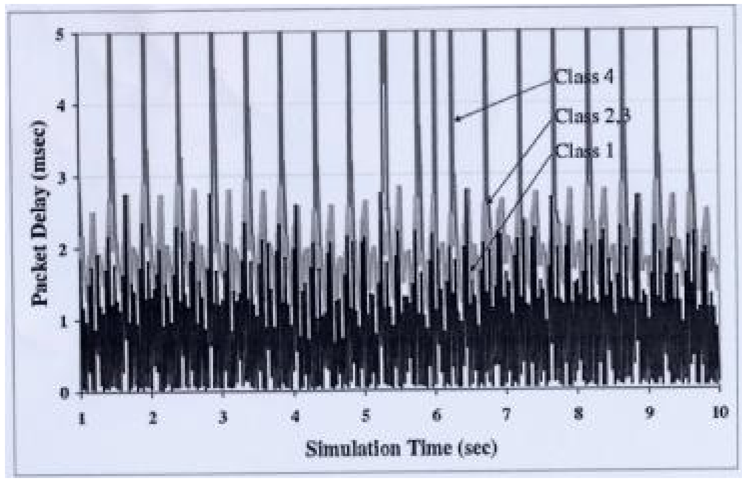



**(a)**

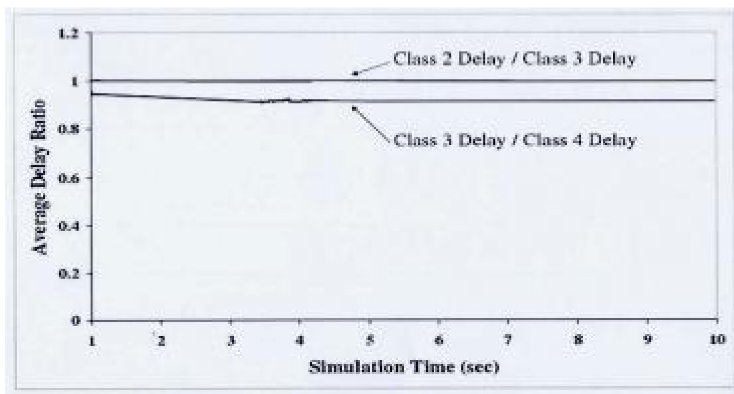



**(b)**

**Figure 3.** Delay differentiations at the first access point: a) packet delays, b) ratios of delays.




**(a)**




**(b)**

**Figure 4.** Delay differentiations at the first access point: a) packet delays, b) ratios of delays.

## CONCLUSION

In this article we present the resource management and scheduling algorithm of a Quality of Service (QoS)-aware wireless access router, which can be used for QoS provisioning in home networks. We developed a queuing system with a service rate adaptation algorithm based on a set of predictive traffic performance parameters. Simulation results illustrate that the proposed algorithm can provide both absolute and relative delay guarantees. At the same time, the algorithm is simple and scalable for its implementation at an IP-based wireless access point. In addition, the proposed approach does not require a prior specification of the arriving traffic and, therefore, does not require any traffic policing mechanism.

## ACKNOWLEDGMENTS

This work is supported by the Thailand Research Fund through grant MRG4780198.

## REFERENCES

1) Castelluccia A Castelluccia C. Differentiation mechanisms for IEEE 802.11. *Proc. IEEE Infocom 2001*. Apr. 2001.
2) IEEE 802.11e draft/4.0. Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS). Nov. 2002.
3) IEEE 802.11a. Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High Speed Physical Layer in the 5 GHz Band. 1999.
4) Oottamakorn C Bushmitch D. Resource management and scheduling for the QoS-capable home network wireless access point. *Proc. IEEE CCNC 2004.* Jan. 2004.
5) Fitzek F Reisslein M. MPEG-4 and H.263 traces for network performance evaluation (extended version). *Technical Report TKN-00-06*, Technical University Berlin, Dept. Electrical Eng., Germany, Oct. 2000.

## บทคัดย่อ

ชัยวัฒน์ อุตตมากร[1] และ Dennis BUSHMITCH[2]

**การส่งข้อมูลแบบประกันคุณภาพความล่าช้าในเครือข่ายคอมพิวเตอร์ภายในบ้าน**

ในปัจจุบัน การให้บริการบนอินเตอร์เน็ตมีการขยายตัวเพิ่มขึ้นอย่างมาก ทำให้การเชื่อมต่ออินเตอร์เน็ตความเร็วสูงเป็นที่ต้องการของผู้ใช้บริการทั้งในภาคธุรกิจ สถานศึกษา หรือแม้แต่การติดต่อสื่อสารส่วนบุคคล การขยายตัวเพิ่มขึ้นนี้ทำให้การเชื่อมต่อจากเครือข่ายภายในบ้านผู้ใช้ไปยังอินเตอร์เน็ตด้วยความเร็วสูง มีความต้องการเพิ่มมากขึ้น การสนับสนุนการสื่อสารข้อมูลเพื่อรองรับเครือข่ายอินเตอร์เน็ตสถาปัตยกรรมใหม่ คือการให้บริการที่มีคุณภาพที่ดี และแตกต่างระหว่างข้อมูลประเภทต่างๆ ดังกล่าวบนเครือข่ายไร้สายยังอยู่ในระหว่างการวิจัยและพัฒนา แต่เครือข่ายภายในบ้านผู้ใช้ยังไม่สามารถรองรับการให้บริการสถาปัตยกรรมใหม่ได้ทั้งหมด โดยเครือข่ายยังขาดในส่วนของการให้บริการรับส่งข้อมูลที่แตกต่างที่บริการบนอินเตอร์เน็ตเหล่านี้ต้องการ การเชื่อมต่อผ่านเครือข่ายไร้สายเป็นทางเลือกหนึ่งสำหรับเครือข่ายภายในบ้าน บทความวิจัยนี้นำเสนอกลไกที่ใช้ในการจัดการตารางการให้บริการแก่และการบริหารจัดการที่จุดเชื่อมต่อของระบบเครือข่ายไร้สาย เพื่อรองรับการให้บริการบนอินเตอร์เน็ตแบบต่างๆ เหล่านี้บนเครือข่ายภายในบ้าน กลไกที่ใช้สามารถรับประกันความล่าช้าในการรับส่งข้อมูลและการสูญหายของข้อมูลสำหรับการให้บริการบนอินเตอร์เน็ตแบบต่างๆ กัน ผลของการวิจัยแสดงว่ากลไกนี้สามารถรองรับเครือข่ายอินเตอร์เน็ตสถาปัตยกรรมใหม่สำหรับเครือข่ายภายในบ้านและยังสามารถนำไปสร้างบนเครือข่ายขนาดใหญ่ได้อีกด้วย

---

[1] สำนักวิชาวิศวกรรมศาสตร์และทรัพยากร มหาวิทยาลัยวลัยลักษณ์ อำเภอท่าศาลา จังหวัดนครศรีธรรมราช 80160

[2] Panasonic Information and Networking Technologies Laboratory, NJ, USA