Abstract: The future IP networks need to bear multimedia services with diverse quality of service (QoS) requirements. To sustain the contracted QoS needs not only to commit network resources, but also to monitor and control QoS adaptively. The location of the network segment that affects QoS negatively is crucial for resources controlling. This paper analyzes the current QoS distributed monitoring scheme, and defines a new concept of key network segment (KNS). Two schemes, namely centralized locating scheme and distributed locating scheme, that can locate KNS are proposed, analyzed, and compared. After located the KNS, resources could be tuned to guarantee the contracted QoS.

Key words: Quality of service (QoS), QoS management, QoS monitoring, Key network segment (KNS)

1. Introduction

Future integrated services networks must support multimedia applications with diverse performance requirements. To provide quality of service guarantees and ensure the contracted QoS be sustained, it is far from sufficient to just commit resources, since QoS degradation is often unavoidable [2,3]. Any fault or weakening of the performance of a network element may result in the degradation of the agreed QoS. Thus, QoS monitoring is required to track the ongoing QoS, compare the monitored QoS against the expected performance of applications, and find possible QoS degradation.

The mechanisms for QoS monitoring can be classified into two categories by the QoS information where to obtain from: end-to-end QoS monitoring and distributed QoS monitoring [2]. In an end-to-end QoS monitoring scheme, only the end-to-end QoS between the sender and receiver of a real-time flow is monitored, so as it is hard to locate the network segments that cause the QoS degradation. However, the distributed monitoring scheme monitors not only the end-to-end QoS, but also the QoS distribution experienced by the flow in different network segments.

It is admitted that the goal of QoS management is not to monitor QoS, but to guarantee the contracted QoS by means of tuning and controlling network resources. The precondition of network resources tuning is to locate the network segments that affect the QoS negatively. We name these network segments as key network segments (KNS).

While a considerable amount of research has been reported within QoS provision [1-8], such as QoS architecture [1], QoS monitoring [2,3], and QoS controlling [4], none of these papers mentions the definition of KNS and how to locate KNS. The main purpose of this paper is to define the KNS, identify the challenges involved in locating KNS, and propose two schemes to meet these challenges. Adopting the two proposed schemes, network manager can locate the KNS of a given real-time flow. Consequently, network resources can be tuned and contracted QoS can be guaranteed based on the located KNS.

This paper is structured as follows. Section 2 introduces the current QoS distributed monitoring scheme. KNS definition, two schemes that can be adopted to locate KNS, and comparison of the two schemes are presented in section 3. Section 4 discusses the issues that remain open. And finally, section 5 concludes this paper.

2. QoS Distribution monitoring

The distributed QoS monitoring scheme [2] proposed by Yuming Jiang, Chen-Khong Tham, and Chi-Chung Ko is illustrated in figure 1. The real-time application name server (RTANS) provides a mechanism for monitoring application to locate relevant monitors of a real-time flow, which operates as follows.

Prior to monitoring, the RTANS is set up and made known to all monitors and the monitoring application in the system. During monitoring, traffic information is collected from relevant monitors that are metering the flow in different network segments. Once detecting a new real-time flow, the monitor registers its address and the traffic attributes of the monitored real-time flow with RTANS. The traffic attributes of a flow include the flow’s sources and destination addresses, the byte count of the flow’s traffic, and...
so on. Furthermore, according to the flow’s attributes, relevant monitors of each flow are sorted into a real-time application (RTA) table, so as the monitoring application could find out the addresses of a flow’s relevant monitors and retrieve traffic information of the flow from them. Table 1 is a RTA table corresponding to figure 2. RTANS is used as a bridge between the monitors and the monitoring application. This scheme is referred to as the relevant monitor based (RM-based) scheme. In this scheme, there is only one common RTANS and one monitoring application for the whole network. Monitoring application and all monitors are manually configured to know where the RTANS is.

Another scheme proposed in [2] is referred to as improved relevant monitor based (IRM-based) scheme. There are three differences between RM-based scheme and IRM-based scheme. 1) The IRM-based scheme has more than one RTANS, each for a different network management domain. 2) Each monitor can report traffic information to multiple monitoring applications that belongs to different network management domains. 3) Each RTANS registers its address with the monitoring application in the same management domain and all monitors in the system. Each monitor maintains a RTANS list, which stores the address of the RTANS in the system. During monitoring, the monitors use these addresses to locate and register with all the RTANSes in the list its own address and the monitored traffic attributes of real-time flows. And monitoring applications can locate a flow’s relevant monitors to retrieve the flow’s QoS information. Figure 3 illustrates the IRM-based QoS monitoring scheme.

### 3. Schemes for Locating Key Network Segment

In the QoS distributed monitoring schemes described in section 2, the traffic information of real-time flow can be retrieved from the relevant monitors that are metering the flow. Comparing detected traffic information against the contracted QoS requirements, if any of the relevant monitors finds the flow’s QoS can’t meet the SLA [1], the network resources must be tuned.

![Figure 3: IRM-based QoS monitoring scheme](image)

In fact, not all of the network segments that transfer a flow affect the flow’s QoS negatively. We name the network segments that affect a real-time flow F’s performance negatively as flow F’s QoS key network segments (KNS). Only after the KNSes are located, can the corresponding network resources be tuned and the agreed QoS be sustained. We will define KNS firstly and introduce two schemes to locate it thereafter.

Prior to define KNS, we define “immediate predecessor” and “immediate successor” firstly. Assuming M_i and M_j are flow F’s relevant monitors. M_i observes traffic packets of flow F before M_j and there is no any other monitor M_k which prior to M_i and posterior to M_j detects flow F’s packets. M_j is referred to as the immediate predecessor of M_i. And M_i is called M_j’s immediate successor accordingly. The immediate predecessor of the monitor that observes flow F’s packets firstly is flow F’s sender, and the immediate successor of the monitor that find F’s packets last is flow F’s receiver.

Now we can define KNS. As illustrated in figure 4, assuming that there are N monitors M_1, M_2, M_3, … , M_N are metering flow F’s traffic, if monitor M_i finds flow F’s QoS can’t meet F’s performance requirement and M_i’s immediate predecessor M_{i-1} observes a higher level QoS of flow F, we define the network segment (NS) NS_i that between M_{i-1} and M_i as flow F’s QoS key network segment. According to this definition, if a network segment is F’s KNS, it affects F’s QoS necessarily. However, if a network segment is not a KNS of F, we are not sure whether it can provide the contracted QoS. For example, assuming 1 is the highest and contracted QoS level, and NS_{i-1} can support level 3 QoS, if NS can only support 3 or lower level QoS, the above definition of KNS determines NS_{i-1} is not a KNS. Therefore, it is impossible to locate all KNSes for one times, so we define first KNS (FKNS) and most KNS (MKNS) concepts.

FKNS is the first KNS that can’t support flow F’s contracted QoS. MKNS is the KNS that affects F’s QoS most heavily. We are interested in FKNS

<table>
<thead>
<tr>
<th>Real-Time Flow</th>
<th>Relevant Monitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>M_i, M_j, M_k, …</td>
</tr>
<tr>
<td>F2</td>
<td>M_h, M_j, M_l, …</td>
</tr>
</tbody>
</table>

Table 1: Figure 2’s RTA Table
because, firstly, if flow F’s performance degradation is observed by any of the monitors M_{i+1} to M_N, it may be caused by NS_i indeed. Secondly, if there is any other network segments posterior to NS_i, affect Flow F’s QoS, NS_i’s problem must be solved firstly. On the other hand, we concern MKNS for having solved the problem of MKNS, the F’s QoS can be improved in some degree. In this paper, we will primarily concentrate on arguing about FKNS problem.

![Figure 4: A simple real-time flow](image)

To locate a real-time flow F’s FKNS, the sequence of F’s relevant monitors must be determined, which requests the F’s relevant monitor to be synchronized. A common scheme to determine the sequence of F’s relevant monitor is to synchronize all monitors to a certain reference time such as the universal coordinated time (UCT). This can be achieved by using the Global Positioning System (GPS)[4], the Internet network time protocol [NTP], a global time base (GTB)[9], or scheme proposed and realized by Yuming Jiang [3]. Using these synchronizing scheme, different arriving time of F’s packets at different monitors are observed and relevant monitor’s sequence is determined. We improve the RTA table of [3] by adding monitor order information to RTA.

The following two subsections will propose two schemes to locate FKNS respectively: centralized locating scheme and distributed locating scheme, according to their different mechanisms and different component to perform locating.

### 3.1 Centralized locating scheme

Centralized locating scheme is also referred to as monitoring application based locating scheme (MABLS). As illustrated in Figure 5, MABLS improve the monitoring application’s function illustrated in figure 1 by adding FKNS locating component. The arrow 1 in figure 5 indicates network resources tuning function. In MABLS, monitoring application locates the FKNS by consolidating QoS distribution information collected from the relevant monitors, which operates as follows.

Prior to QoS monitoring and FKNS locating, the RTANS is set up and made known to the monitoring application and all monitors in the system, which can be done by network manager manually or by a directory server automatically.

During monitoring, once detecting a new real-time flow, the monitor registers with the RTANS its address, the time it detecting the flow, and the traffic attributes of the flow. Relevant monitor sequences of each flow are sorted and determined according to the flow’s attributes and the time that the flow is first detected. Then, the monitoring application can start to collect relevant monitor sequences that are metering the flow in different network segment from RTANS and the flow’s traffic information from the relevant monitor sequences.

![Figure 5: Centralized locating approach architecture](image)

The monitoring application collecting both the relevant monitor information from RTANS and the QoS distribution information from relevant monitors can adopt either “pulling scheme” or “pushing scheme”. As for the former, pulling scheme is that the monitoring application polls the RTANS periodically for the relevant monitor information, meanwhile pushing scheme is that the RTANS reports only these relevant monitors of some kinds of flow the monitoring application is interested in. When a relevant monitor has detected a real-time flow and registered its address and the flow’s attributes with RTANS, if the flow is the monitoring application interested in, RTANS reports the monitor’s address to the monitoring application. Collecting QoS information uses the same mechanism as collecting relevant monitoring.

If a monitor M_i metering the real-time flow F observes F’s QoS degradation and reports to the monitoring application, the monitoring application can determine if there is any QoS degradation observed by M_i’s immediate predecessor M_{i-1}. If the answer is “no”, the network segment between M_{i-1} and M_i is the FKNS of flow F. Otherwise, the FKNS is another network segment prior to M_{i-1}.

### 3.2 Distributed locating scheme

Distributed locating scheme is also named as monitor based locating scheme (MBLS). MBLS locates FKNS by the collaborating of relevant monitors. As illustrated in figure 6, MBLS place communication and FKNS determining functions to relevant monitors in figure 3, which operates as follows.

Similar to MABLS, MBLS made all
RTANSes known to the monitoring application in the same management domain and all monitors in the system prior to monitoring and FKNS locating. During monitoring, monitors register their addresses and attributes of monitored flows with all RTANSes. The RTANSes will determine the monitor’s sequence basing on the monitors synchronizing information and broadcast the

sequence information to all monitors. Thus, all monitors know their immediate predecessor and immediate successor.

If monitor $M_i$ observes QoS degradation of flow $F$, it sets an inner flag $f_0$ to indicate that $M_i$ has found the QoS degradation, and reports the degradation to its immediate predecessor $M_{i-1}$. While $M_{i-1}$ receives the QoS degradation report, it tests the inner flag $f_0$ firstly. If $f_0$ is set, $M_{i-1}$ discards the QoS degradation report. Otherwise, the $M_{i-1}$ knows that the network segment between it and $M_i$ is FKNS and reports it to the monitoring applications. As all relevant monitors monitoring the traffic information simultaneously, if there is any QoS degradation, FKNS can be located quickly.

3.3 Comparison

Both MABLS and MBLS can be used to locate FKNS, but there are 4 primary differences:

1) MABLS enriches the function of monitoring application with FKNS locating function. MABLS is suitable to RM-based monitoring scheme. MBLS, on the other hand, improves the relevant monitor’s function, and the FKNS locating function of relevant monitors is transparent to the monitoring application. MBLS is suitable to IRM-based QoS monitoring scheme.

2) MABLS requires that all relevant monitors are known by the monitoring application. It is used in one management domain usually. Whereas MBLS needs that the relevant monitors know their immediate predecessor, regardless whether or not the monitor and its immediate predecessor belong to the same management domain or not. It can be used in multi-management domain.

3) For MABLS, if the monitoring application determines that a network segment is FKNS, it must assure that all QoS distribution information observed by monitors prior to this network segment has been obtained. While in MBLS, when a monitor $M_i$ receives a QoS degradation report from its immediate successor, it can determine if the network segment between it and its immediate successor is FKNS immediately. $M_i$ needs to know the QoS information observed by other monitors. As a result, to locate FKNS in the same network, MBLS may be faster than MABLS.

4) In MABLS, the monitoring application knows all QoS distribution information of flow $F$, so MABLS can be used to locate MKNS also. If a monitor $M_i$ observes the lowest QoS level and its immediate predecessor $M_{i-1}$ has a higher QoS level, the network segment between $M_{i-1}$ and $M_i$ is the MKNS. MBLS, however, can’t used to locate MKNS directly. To locate MKNS, a MBLs monitor needs not only to know $N_{i}$ is KNS, but also to determine that all following NS of $N_{i}$ is not KNS. This can be done by comparing the monitored QoS with the QoS observed by receiver, if $M_{i-1}$ knows that $N_{i}$ is KNS and the QoS level observed by $M_{i-1}$ is equal to QoS level observed by the receiver, $M_{i-1}$ can determine that $N_{i}$ is MKNS.

3.4 Discussion

So far, we have introduced and compared two schemes for locating KNS. There are several issue need to be discussed further.

1) Pulling and pushing scheme: Monitoring application using either pulling or pushing scheme can obtain QoS distribution information from relevant monitors. To collect QoS information from relevant monitors, the pulling scheme has a fixed delay. The maximum delay is a polling period. Thus, the polling period must be determined carefully. In practice, the polling period can be determined dynamically according to the network situation. For example, when flow $F$’s QoS has stabilized for a long time, or $F$’s QoS level is much higher than agreed QoS level, the polling period can be made longer. Whereas, while $F$’s QoS fluctuates acutely, or $F$’s QoS level is close to the contracted QoS, the polling period is better to be shorter. Adopting variable polling period can minimize the QoS information traffic and keep the real time requirement of monitoring simultaneously, but with disadvantage of the system complexity.

The pushing scheme has the reporting period issue also. In practice, we can set a lower limit, so as when QoS level is over the lower limit, the monitor don’t report QoS information to monitoring application. This reduces the traffic of QoS information, but the monitoring application can’t obtain the QoS information in real time. The monitoring application knows that the QoS is fine, but don’t know how fine it is actually.

![Figure 6: Distributed locating approach architecture](image-url)
Two Schemes for Network QoS Degradation Locating

2) The loss of QoS information Packet: We use the following scheme to solve the problem of QoS information packet loss in pushing scheme. The QoS information receiver must confirm every QoS packet that a monitor sends to it. If a monitor has not sent any QoS information packet to the receiver for a fixed-time that we named as “report period”, it must send a “Live” message to the receiver. “Live” message indicates that the monitor functioning properly and the network is normal. If the receiver has received neither QoS information packets nor “Live” message, it can assume that the monitor had faulted or the network congested. If a monitor didn’t receive confirming message of a QoS information packet, it must resend the packet before a predefined max resend times reached.

Pulling scheme also has packet loss problem. Both query packet loss and response packet loss are treated similarly, that is, If the puller can’t receive response from the relevant monitors, it queries again before reach a predefined times.

3) Multi-KNS problem: In all above sections, we assume QoS degradation is caused by FKNS or MKNS. If there are more than one network segments that affect the flow’s QoS, adopting the two FKNS locating schemes proposed in this paper can locate the first network segment that we defined as FKNS. After the FKNS’s problem has been solved, another network segment becomes new FKNS and can be located.

Additionally, by analyzing other flow’s QoS information on the same network segment, we can determine whether or not a network segment is a flow’s KNS. For example, if all flows on a network segment have degraded in some degree, we can determine that this network segment is KNS. Otherwise, if only one flow F’s QoS degrade on this network segment, this network segment may not be the KNS of F.

4. Questions remain open

We have introduced two schemes for locating KNS and discussed a few related issues further. However, several issues remain open.

1) After the KNS has been located, the network resources need to be tuned. How to tune the KNS’s network resources is a difficult problem. We can achieve the contracted QoS level by adopting schemes such as changing routing policy, queuing algorithm, priority of the service, and buffer management policy. Traffic engineering can be used also. But, in general, which scheme is adopted requires further research.

2) To analyze relative of different flows on the same network segment is another difficult question. How these different flows interplaying and affecting mutually, and how to deduce a flow’s QoS information from other flow’s QoS information require future research.

3) QoS parameters normally include bandwidth, delay, jitter, and packet loss [1]. Different network segments have different impact on different QoS parameters. To locate KNS, we may locate on different network segments for different QoS parameters of the same flow. But the mechanism needs further study.

5. Conclusion

In integrated services networks, providing sustained QoS is crucial. Sustained QoS requires tuning network resources when QoS degrades. This paper’s primary contribution is that, for the first time, it defines KNS and introduces two schemes to locate KNS. As for how to tune network resources is a question will be discussed in another paper.

References