For Discussion
on 13 January 2003

Legislative Council
Panel on Information Technology and Broadcasting

Report on the Findings and Recommended Measures in relation to Telephone Network Congestion on 11 September 2002

Background

On 11 September 2002, the fixed and mobile telecommunications networks experienced serious congestion in the first few hours after the public announcement from the Hong Kong Observatory (HKO) of the imminent issue of the No 8 typhoon signal. At the meeting of the Panel on 17 September 2002, the Office of the Telecommunications Authority (OFTA) explained to Members the Government's current requirements for fixed and mobile telecommunications operators on their network capacities and the existing emergency arrangements, and undertook to study measures to be taken to improve the situation in future.

2. This paper briefs Members about the results of the review and the recommended measures to be taken to ease telephone traffic congestion and reduce traffic demand in similar situations in future.

The Review

3 OFTA has asked the major fixed network operators and all mobile network operators to provide detailed information on:

- emergency call handling
- the network statistics relating to the critical hours on 11 September 2002
- priority arrangements, if any

- network management actions taken on 11 September 2002
- capacity and interconnection design arrangements.

OFTA has also engaged Ovum as consultant to study and compare the technical capacity, performance and capability of local fixed and mobile networks with the international best practices under severe weather conditions.

4. On network capacity, generally speaking, the total capacity of a telecommunications network is determined by its capability to handle call traffic during busy hour and the spare capacity available to handle unexpected surge in traffic. The capability to handle the busy hour call traffic is measured most commonly as the Grade of Service (GoS), which is the probability of a call being blocked by lack of network capacity in the busy hour. According to the information supplied by the operators, the fixed networks in Hong Kong are designed to a 0.5 – 1.0% GoS, whilst mobile networks are designed to a 5% or better GoS. Ovum has compared these figures with international standards and design practice in other countries as follows:

Organisation / Country	*GoS for Fixed Network*	*GoS for Mobile Network*
Recommended by ITU-T[1]	2%	5 – 10%
Australia	0.2 – 1.0%	5 – 10%
Ireland	0.5%	not available
Switzerland	1%	not available
United Kingdom	0.8%	not available

It can be seen that the design targets of the networks in Hong Kong are better than those recommended by the ITU-T and are commensurate with the design practice in other countries.

5. The operators also provided information on spare capacity of their networks, which is in the order of 20% for both fixed and mobile networks. Ovum has advised that a figure of 10% to 30% would be typical and the networks in Hong Kong are dimensioned in accordance with accepted international best practice.

6. Ovum concludes that there are no fundamental shortcomings in the design of the networks in Hong Kong, which are designed in accordance with

[1] The Telecommunication Sector of International Telecommunication Union.

international best practice. Records from the operators indicated that the call attempts on 11 September 2002 went up to as high as six times the level of a normal day after the announcement of the Pre-8 typhoon message (ie No 8 signal would be issued within two hours). Ovum recognises that even with better management, the networks would not have been able to satisfy all the public's demand for communications under the circumstances that prevailed on that day.

7. The consultant has however identified some areas of improvement in respect of traffic management and recommended steps that should be taken by the operators to improve the traffic management of their respective networks in emergency or severe weather conditions in future. The operators are also advised to develop closer coordination among their network management centres. OFTA will adopt the consultant's advice and **coordinate with the fixed and mobile network operators to implement the measures to enhance the telephone handling capacity and capabilities of the telecommunications networks in emergency situations.**

Measures to Ease Congestion and Reduce Demands

8. Despite the fact that the design of telecommunications networks of Hong Kong are commensurate with international standards and practices, there is a need to consider measures which would reduce the traffic demand and ease the burden on telecommunications networks during severe weather conditions. In this connection, OFTA has held discussions with other frontline Government departments, including the Hong Kong Observatory (HKO), Education Department (ED), Information Services Department (ISD), Police and Security Bureau (SB), which all have a role to play in severe weather conditions, on measures that can be taken to ease telephone traffic congestion and reduce traffic demand in similar situations in future. OFTA sets out below the measures that are recommended to be adopted:

- **Priority access arrangements to fixed networks will be maintained for agencies (such as the police) that are likely to be involved in an emergency and also government officers and other personnel with genuine need;**

- **HKO offers to issue an Alert Message to a designated focal point assigned by OFTA for automatic onward dissemination to telecommunications operators to enable the operators to take necessary precautionary measures to handle possible upsurge of telephone traffic in anticipation of any probable issue of Pre-8 typhoon message.**

- **The Government agencies should explore the use of communications means other than through normal telephone lines to facilitate liaison among themselves during emergency situations.**

- **ED and HKO have reached a consensus to synchronize ED's announcement on closure of schools with that of HKO on the issue of Pre-8 typhoon message.**

- **The public will be encouraged to consider using alternative advanced telecommunications facilities for communications.**

- **Publicity programmes will be devised to educate the public on how to make best use of telecommunications facilities.**

- **If there is severe telephone congestion on a particular hour of a day, special messages may be broadcast on TV and radio to advise the public to make fewer telephone calls to ease the burden on the telephone networks.**

9. Annexed to this Paper is a report prepared by OFTA setting out in detail the findings of the review of telephone network congestion on 11 September 2002 and the recommended measures to ease telephone network congestion and reduce traffic demand in severe weather conditions. The report prepared by Ovum is appendixed to the OFTA's report.

Way Forward

10. OFTA will coordinate with the fixed and mobile network operators to implement the measures recommended by Ovum to improve traffic

management of the networks. OFTA will also liaise with other government departments to implement the recommended measures to ease telephone traffic congestion and reduce traffic demands on similar occasions in future.

Office of the Telecommunications Authority
3 January 2003

Annex

Report on the Findings of the Review of Telephone Network Congestion on 11 September 2002 and Recommended Measures to Ease Telephone Network Congestion and Reduce Traffic Demand in Severe Weather Conditions

Background

On 11 September 2002, the fixed and mobile telecommunications networks experienced serious congestion in the first few hours after the public announcement from the Hong Kong Observatory (HKO) of the imminent issue of the No 8 typhoon signal (Pre-8 typhoon message). In a paper prepared by the Office of Telecommunications Authority (OFTA) for the Legislative Council Panel on Information Technology and Broadcasting dated 16 September 2002, the Director General of Telecommunications has stated that he would conduct a review to look for areas for improvement to ensure that the public telecommunications networks of Hong Kong are well equipped to cope with the existing and future demands of our community.

2. The review is completed and this report examines the findings and recommends measures to be taken to ease telephone network congestion and reduce traffic demand on the telecommunications networks during severe weather conditions, such as when Hong Kong is under the threat of tropical cyclones.

The Sequence of Events on 11 September 2002

3. The following is a sequence of events in relation to notifications given to OFTA and the public on 11 September 2002:

Time	From	To	Message
1145	HKO	Public (via radio and TV, HKO's website, HKO's telephone	Pre-8 typhoon message was issued (ie No 8 Signal would be issued within two hours)

		answering systems)	
1146	HKO	Information Services Department (ISD) (by leased telephone line)	No 8 Signal would be issued within two hours
1149	ISD	OFTA (by telephone)	No 8 typhoon signal would be issued within two hours
1205	Education Department (ED)	Public (by issuing press release and announcement on radio/TV via ISD)	All schools would be suspended.

4. Upon receipt of the notification from the ISD, OFTA followed the procedure given in the "Guideline for the Dissemination of Telecommunications Network Congestion Warning" to alert major telecommunications operators of any possible network congestion, and requested them to take necessary precautionary actions. However, by that time, the upsurge in traffic had already commenced. In the study of the incident, OFTA received from operators detailed breakdown of traffic information. The total number of call attempts for fixed and mobile networks within the first two hours from 1100 hours on 11 September 2002 are given below:

Time	Total number of call attempts made to all fixed networks	Compared to the average normal peak hour	Total number of call attempts made to all mobile networks	Compared to the average normal peak hour
1100 – 1200	8.7 million	1.24 times	4.5 million	0.91 times
1200 – 1300	33 million	4.8 times	13 million	2.8 times

It was recorded that, during 1200 hours to 1300 hours, the peak hour call attempts[1] handled by the PCCW-HKT Telephone Limited (PCCW-HKTC)'s network had reached a peak of 6 times of the average normal peak hour call

[1] A call attempt is an attempt by the caller to establish a connection to the called party. The handling of a call attempt consumes processing resources in the switches of the public telephone network. The public telephone network is designed to handle a specified number of call attempts during the peak hour (peak hour call attempts)

attempts while some individual mobile networks had also experienced up to 4 times of the normal level. The sudden upsurge of the telephone traffic exceeded the maximum handling capacities of both fixed and mobile networks.

Findings on Network Capabilities and Management of Telephone Congestion

5. In a study prepared by Ovum (appendixed to this Report), the consultant engaged by OFTA, Ovum concluded that there are no fundamental shortcomings in the design of the networks in Hong Kong, which are designed in accordance with international best practice. The grade of services of the network in Hong Kong is comparable with that of the networks in other developed economies. They have adequate capacity to handle normally encountered upsurge of telephone traffic. However, it is impossible for these networks and indeed any other adequately designed networks to handle an upsurge of call attempts of the level encountered on 11 September 2002. However, the consultant has identified some areas of improvement in respect of traffic management at times of emergency. They include the following:

- to control call attempts at the point where they first encounter congestion so as to make optimum use of the network capacity;
- to improve the exchange of information and coordination among the network management centres of the operators, so as to enable a network operator that is experiencing problems to ask other network operators to reduce the call attempts across the interconnection links;
- to give a 15-minute notice to allow operators sufficient time to implement the necessary measures.

6. The events on 11 September 2002 also resulted in an increased number of calls to the "999" Emergency Reporting Centres. But according to the Police there were no signs of irregularities on the "999" system between 1200 hours and 1459 hours. However Ovum has recommended that further enhancement is possible in the following areas:

- all agencies that are likely to be involved in an emergency should have priority access to the fixed networks;

- as far as possible a consistent approach should be taken by all operators in routing emergency calls.

7. Currently, emergency calls originated from other networks are routed through PCCW-HKTC's network before reaching the "999" Emergency Reporting Centres. Some but not all operators have dedicated routes for connection with PCCW-HKTC. Although there was no problem experienced in the incident of 11 September 2002, OFTA agrees with the consultant that in order to further enhance the "999" traffic flow, a consistent approach is necessary and operators should deploy dedicated trunks for conveying emergency calls to the network of PCCW-HKTC. In the longer term consideration should be given to studying whether emergency calls originated from mobile phones should be routed through additional alternative paths to reach the "999" Emergency Reporting Centres.

8. OFTA has considered the findings and recommendations of the consultant and agrees to adopt the proposed measures to enhance the telephone handling capability and capacities of the telecommunications networks in emergency situations **(Recommendation 1**). These measures will be implemented well before the start of the next typhoon season. OFTA also recommends to maintain priority access arrangements to fixed networks for agencies (such as the police) that are likely to be involved in an emergency, as well as government officers and other personnel with genuine need. **(Recommendation 2)**

Alert Messages to OFTA and Operators

9. PCCW-HKTC is the major fixed telecommunications operator in Hong Kong and has provided details on its network and traffic conditions during the initial hours from before noon on 11 September 2002. PCCW-HKTC submitted that a 10-minute notice before HKO issues the Pre-8 typhoon message to the public would be necessary to enable it to implement the network management measures to better handle sudden traffic surge. In Ovum's report, the consultant, after considering the situations of various

networks, suggests that a 15-minute notice ahead of the issue of Pre-8 typhoon message would allow fixed and mobile operators sufficient time to initiate necessary network management measures.

10. Taking into account the concerns of PCCW-HKTC and the recommendation of Ovum, OFTA considers there is room to improve the timing of notifications to the frontline Government agencies and essential public services providers before the issue of the Pre-8 typhoon message. To address this information need, HKO offers to issue a message to OFTA to serve as an early alert for a probable issue of Pre-8 typhoon message (Alert Message). The Alert Message will be sent directly from HKO to one designated focal point, assigned by OFTA, for automatic onward dissemination to telecommunications operators to allow them sufficient lead time to take precautionary actions against any likely telephone traffic upsurge. **(Recommendation 3)**

Means of Notifications among Government Agencies

11. It was reported that some communications among HKO, ISD and ED were affected by the congestion in the telephone systems. To ensure that the Alert Message and any subsequent related messages can be disseminated to the frontline Government agencies without disruption and to facilitate subsequent liaison among the agencies during emergency situations, OFTA recommends and these agencies agree that they shall explore the use of communications means other than through normal telephone lines. **(Recommendation 4)**

School Suspension Notice

12. We notice that 11 September 2002 was a school day, and the Pre-8 typhoon message was issued before noon time. The timing clashed with the midday rush hours during which the morning class students were soon to end their school, and afternoon class students might already be on their way to school. Parents were anxious to get in touch with the schools and their children regarding school closure arrangements.

13. For comparison purpose, telephone operators were asked to supply to

OFTA telephone traffic statistics relating to similar previous events. There were two incidents of typhoon signal No 8 hoisted in 2001. The first one was between 1930 hours on 5 July 2001 to 1340 hours on 6 July 2001, and the other one between 0030 hours to 1940 hours on 25 July 2001. There was also a black rain storm warning signal issued at 2245 hours on 1 September 2001. On all these occasions, there was no network congestion or failure reported to OFTA[2]. It is noted that the typhoon signals were all issued in the evening or at night. And in the first two cases, the schools were still on summer vacation.

14. Recognising that communications among parents, children and schools potentially make up a significant proportion of telephone calls in severe weather situations, OFTA considers that there is room to improve the timing of issuing school suspension notice. OFTA noticed that on 18 October 2002 there was an announcement of red rain warning at 1035 hours. ED made an announcement to close the afternoon school at 1040 hours. In this case the announcement was made 5 minutes after that of HKO's. Records of telephone operators indicated that there was no surge of call attempts. OFTA believes that the short time gap between the two announcements was the main reason for not causing upsurge of telephone calls. OFTA reports that HKO and ED have reached a consensus to synchronize ED's announcement on closure of schools with that of HKO on the issue of Pre-8 typhoon message. OFTA welcomes the new arrangement and expects if the school closure arrangement is announced simultaneously with the Pre-8 typhoon message, the number of call demands will reduce, and a lesser burden will be imposed on the telecommunications networks. **(Recommendation 5)**

Use of Alternative Telecommunications Facilities

15. The reports submitted by the network operators show that on 11 September 2002, the public relied heavily on the voice channels of the fixed and mobile networks to communicate with each other. It was the networks that carried voice traffic that experienced serious congestion. For example, as section 5.7 of the report of Ovum shows, although the mobile operators

---

[2] For details please refer to "A Note on Previous Cases of Telephone Traffic Congestion due to Typhoon or Heavy Rain" dated 5 November 2002 prepared by OFTA for the information of Legislative Council Panel on Information Technology and Broadcasting.

reported an increase in short message service (SMS) traffic when compared to normal peak day loads, the increase was not as dramatic as the increase in telephone calls attempts, and the increase in traffic was well within the capacity of the SMS systems.

16. Hong Kong has developed an advanced telecommunications infrastructure. Household penetration for connection to the broadband networks has reached 40% and is increasing. As for the mobile networks, inter-operator SMS was available from December 2001, allowing customers of different networks to send short messages to each other. During the critical hours on 11 September 2002, these alternative network facilities were not utilised to their fullest extent. OFTA considers that the public should be encouraged to employ non-voice alternative facilities for communications in emergency situations wherever possible. The following two paragraphs give some precise recommendations from OFTA. **(Recommendation 6)**

17. SMS, e-mail or web service of Internet has the characteristic of being able to disseminate a single message to multiple parties. Thus, OFTA recommends that schools and organisations that need to dispatch identical notices or broadcast announcements in emergency conditions to consider making use of non-voice alternative means for notification, by developing an e-mail or SMS contact list, and posting messages on website.

18. OFTA is aware that on 11 September 2002, some restaurants and shops experienced difficulties in getting connected with credit card centres to obtain validation codes for customer purchases by credit cards. The reason was that some connections to the card centres were through voice telephone lines. Major banks and financial institutions have already used dedicated lines or data networks and did not experience any problem caused by the congestion. OFTA recommends that sizeable restaurants and shops may consider establishing connection with card centres by dedicated lines or data networks to improve the situation during telephone traffic congestion.

Efficient and Proper Use of Resources

19. Our recommendations are not intended to dissuade the public from using the voice channels for communications in emergency situations.

Nonetheless, OFTA would like to highlight the fact that all network facilities have their capacity limits. There was evidently an over reliance on the network facilities for voice communications on 11 September 2002. As pointed out in the Ovum's report, even with better management, the networks would not have been able to satisfy all the public's demand for communications under the circumstances that prevailed on that day. Thus, OFTA would encourage the public that, wherever the circumstances allow, they should consider using alternative means for communications. With more telephone traffic diverted to the alternative facilities, there can be more efficient and balanced use of the available resources, and the burden on the voice channels of the fixed and mobile networks can be reduced.

20. Further, the public should be advised how to make best use of the telecommunications networks in times of emergency and congestion. In addition to encouraging them to use alternative means of communications, OFTA would adopt the suggestions made in the Ovum's report that the public should be educated to:

- limit the number of calls to those that are essential;
- keep the calls short;
- avoid making frequent call attempts. (For fixed line telephones, they should wait for dial tones rather than immediately hang up and re-dial. For mobile phones, they should wait a while before pressing the "redial" button.)

These measures will also help ease the demand on the networks. **(Recommendation 7)**

Publicity

21. Advice to the public and concerned institutions (such as schools) on how to make best use of the telecommunications networks in times of emergency, severe weather conditions and congestion will be included in OFTA's publicity plans and programmes. Further, during severe weather situations, such as when HKO has issued a Pre-8 typhoon message, and if the

operators report that their systems are severely congested, OFTA will consider whether special messages should be broadcast on TV and radio to advise the public that the telephone system is congested and that they should try making only important calls and keeping the calls short. **(Recommendation 8)**

Summary of Measures to be Taken

22. In summary, to improve the traffic handling capability and capacities of the telecommunications network operators, reduce the traffic demand, and ease the burden on telecommunications networks during severe weather conditions, the following measures will be taken:

- OFTA will co-ordinate with the fixed and mobile network operators to implement the measures to enhance the telephone handling capacity and capabilities of the telecommunications networks in emergency situations. (Recommendation 1)

- Priority access arrangements to fixed networks will be maintained for agencies (such as the police) that are likely to be involved in an emergency and also government officers and other personnel with genuine need.. (Recommendation 2)

- HKO offers to issue an Alert Message to a designated focal point assigned by OFTA for automatic onward dissemination to telecommunications operators to enable the operators to take necessary precautionary measures to handle possible upsurge of telephone traffic in anticipation of any probable issue of Pre-8 typhoon message. (Recommendation 3)

- The Government agencies should explore the use of communications means other than through normal telephone lines to facilitate liaison among themselves during emergency situations. (Recommendation 4)

- ED and HKO have reached a consensus to synchronize ED's announcement on closure of schools with that of HKO on the issue of Pre-8 typhoon message. (Recommendation 5)

- The public will be encouraged to consider using alternative advanced telecommunications facilities for communications. (Recommendation 6)

- Publicity programmes will be devised to educate the public on how to make best use of telecommunications facilities. (Recommendation 7)

- If there is severe telephone congestion on a particular hour of a day, special messages may be broadcast on TV and radio to advise the public to make fewer telephone calls to ease the burden on the telephone networks. (Recommendation 8)

Office of the Telecommunications Authority
3 January 2003

Appendix

Comparison of the Technical Capacity, Performance and Capability of Local Fixed and Mobile Networks in Hong Kong with the International Best Practices under Emergency and Severe Weather Conditions

A Report to OFTA

20 December 2002
Project CYI85

Contents

1. Executive Summary

On 11 September 2002 a combination of events occurred that resulted in major congestion of the telecommunications networks in Hong Kong. An announcement of the imminent issue of a “No. 8” Typhoon Warning Signal was given unexpectedly by the Hong Kong Observatory during the network busy hour at 11.45 am. The announcement also coincided with the changeover of the morning and afternoon school shifts. This generated a massive surge in traffic on all telephone networks as the population of Hong Kong began to make arrangements to return home, and to coordinate travel and other arrangements.

A significant part of the surge in traffic was apparently focused on calls to schools by parents wishing to confirm whether or not schools would close for the day, and to make arrangements for the collection of their children. Since the schools would have been able to accept only a small fraction of the calls, this created a huge increase in call attempts that could not be satisfied. However, there was also a surge in calls associated with businesses and consumers in general making alternative arrangements due to the sudden change in circumstances.

The increase in call attempts caused automatic overload protection to operate in a number of switches in the networks. Automatic overload protection and some manual measures taken by operators reduced the interconnection capacity such that the volume of calls over some interconnection links was lower than normal. This reduction in network capacity further increased the volume of unsuccessful calls attempts.

The actions by the operators were uncoordinated and the operators did not inform each other about the state of their networks nor the measures that they were taking. Although such coordination is desirable, better coordination alone would not have reduced the problems significantly.

The best practice approach in these circumstances would be for the networks to limit the call attempt capacity at the network edges and maintain the interconnection capacity. This approach requires the operators to be warned in advance before the traffic starts to grow rapidly.

Even with better management, the networks would not have been able to satisfy all of the public's demand for communications under the circumstances that prevailed on 11 September. A significant part of this demand related to the need for parents to contact schools with limited capacity for handling large numbers of incoming calls. Consequently other methods need to be considered for the dissemination of information to parents to reduce the number of calls.

Hong Kong is unusual compared to other parts of the world in that the region is small, very densely populated, and subject to severe weather events that affect the whole territory. Larger territories would have more scope for relieving congestion in particular localities through using network capacity from areas of the territory where traffic demand had not increased to the same degree.

The fact that there are many interconnected networks in Hong Kong, as a result of liberalisation and competition, did not make the problems experienced on 11 September materially worse. The loss of capacity that occurred on the interconnection links between networks would probably have occurred on links within a single monopoly network, particularly if the switches applied automatic overload protection as a result of call attempts not being adequately limited at the network edges. The management of the situation, however, might have been better if there had been a coordinated approach to traffic management between the operators.

From the information provided by network operators we consider that there are no fundamental shortcomings in the design of the networks in Hong Kong, which are designed in accordance with international best practice. It is our assessment that no comparable networks anywhere in the world would have been able to handle the level of call attempts experienced in such a short period on 11 September 2002.

Our main recommendations to reduce the effects of similar events in the future are:

1. OFTA should be informed well in advance of the issue of announcements to the public concerning No. 8 Typhoon Warning Signals. In this way, OFTA will be able to inform the operators who can then apply methods to restrict call attempts at the network edges, thus preventing switches going into automatic overload conditions and reducing trunk capacity. A 15 minutes pre-warning would allow operators sufficient time to implement the necessary measures

2. OFTA, in conjunction with the operators, should undertake a long term program to educate the public about how the telephone network will behave in the event of excessive demand and how best to use the network, eg limit the number of calls to those that are essential, keep calls short, avoid making frequent call attempts and to use alternative form of communication such as SMS wherever possible

3. all networks that are vulnerable to events such as occurred on 11 September should have the capability to reduce the level of call attempts to specific numbers or number ranges . This means the introduction of call gapping for all networks that are likely to experience high volumes of call attempts that would result in congestion or processor protection measures. Before introducing this requirement, OFTA should discuss with the operators the cost involved and assess the overall risks, benefits and costs

4. the network operators should develop closer coordination between their network management centres under suitable confidentiality agreements. Such communications should use facilities such as a private network that would not be affected by congestion on the PSTN

5. schools should consider establishing a range of methods for distributing information to parents in times when normal schooling is disrupted due to adverse weather conditions. This could include group e-mailings and SMS messages to parents and carers, and the posting of information to school web sites. It could also include the provision of information via a dedicated call centre number when lines to schools are likely to be congested.

2. Background

The telecommunications infrastructure in Hong Kong comprises multiple telecommunications networks operated on a competitive basis. As at the end of September 2002, the Telecommunications Authority (TA) had licensed:

- six local wireline fixed networks
- eleven mobile networks, plus four mobile operator licences for the operation of upcoming third generation mobile services
- a large number of service providers licensed for the operation of external telecommunications services (ETS) which are mainly international direct dialling services based on International Simple Resale or Internet telephony and Internet Service Providers (ISP).

Currently, there are about 3.9 million telephone lines and 5.9 million mobile phone users including pre-paid SIM users in Hong Kong.

2.1 Network congestion on 11 September 2002

From before noon to 2 pm on 11 September 2002, after the public announcement of the Pre-8 message at 11.45 am, all fixed and mobile telecommunications networks experienced serious congestion. A significant proportion of telephone calls could not be connected on both the mobile and the fixed networks. Information provided by the network operators indicated that measured call attempts during the period in question was some six times over the busy hour loading on normal days.

The incident aroused public interest and concern about whether fixed and mobile networks had adequate capacity in place to handle the upsurge of telephone traffic during emergency and severe weather conditions. Some critics questioned whether the problem was caused by poor design or insufficient margin in the fixed and mobile networks in Hong Kong. They also questioned whether:

- the cause is the extensive liberalisation of telecommunications
- the incumbent fixed network operators are reluctant to invest further while the new operators are relying too much on the existing telecommunications infrastructure, or
- the local telecommunications networks can cater adequately for Hong Kong's future development and demand and maintain Hong Kong as a pre-eminent telecommunications centre to support economic development.

OFTA briefed the members of the Legislative Council on 17 September 2002 on the Government's current requirements for fixed and mobile network operators, on their network capacities, existing emergency arrangements and measures to be taken to improve the situation in future. OFTA was further required to produce a review report in three month's time on the issue.

2.2 Terms of reference for this report

This report has been prepared by Ovum in response to OFTA's Consultancy Brief dated 26 September 2002.

The Consultancy Brief required Ovum to address the following core questions regarding whether or not the capacity of fixed and mobile networks:

- is adequate to meet existing and future demand
- includes sufficient spare capacity to handle a traffic upsurge, in accordance with widely adopted best practices in other developed economies/countries.

The report will be used by OFTA as a key input in the development of its review report to the Legislative Council.

In considering the questions outlined in the consultancy brief, Ovum was asked to make reference to international standards and best practices with regard to network design, quality and safety margins. Factors such as Grade of Service (GoS), network availability and loading, call handling capability, blocking probability and spare capacity were also to be considered

The study findings were required to address issues and concerns outlined in Section 2.1 above. These concerns are summarised as:

- concern about the ability of the network to handle the upsurge in traffic during emergency and severe weather conditions
- scepticism about whether the problem was caused by poor or inadequate design
- concern about the impact of extensive telecommunications liberalisation and possible degradation
- concern about whether the Hong Kong network could adequately continue to cater for future development and demand

Ovum was asked to make recommendations and suggestions on any precautionary actions and measures for operators to implement in order to maintain the service level and protect the networks from failure under similar mass calling conditions.

2.3 Methodology and approach

The study commenced with a briefing from OFTA. At the briefing, Ovum was provided with replies to a series of questions previously circulated by OFTA to each of the operators. The questions in summary sought detailed information on:

- emergency services call handling
- the network statistics relating to the hours of the incident
- priority arrangements, if any
- network management actions taken during the incident

- interconnect design arrangements.

Responses were provided from:

- Hong Kong Broadband Network Limited (Fixed Network)
- Hong Kong CSL Limited (Two Mobile Networks)
- Hutchison Global Communications Limited (Fixed Network)
- Hutchison Telephone Company Limited (Three Mobile Networks)
- New World PCS Limited (Mobile Network)
- New World Telephone Limited (Fixed Network)
- PCCW-HKT Telephone Limited (PCCW) (Fixed Network)
- Peoples Telephone Company Limited (Mobile Network)
- SmarTone Mobile Communications Limited (two Mobile Networks)
- Mandarin Communications Limited (Sunday) (Mobile Network)
- Wharf New T&T Limited (Fixed Network)

Note: *The responses from the operators were quite varied in the level of information provided. Most operators insisted that the information provided should remain confidential and not be published.*

Ovum was also provided with the monthly statistical reports submitted to OFTA by the operators. While these reports contain substantial detail from the mobile networks, the fixed networks information is more limited.

A series of interviews were conducted by Ovum and OFTA with each of the operators that had provided written responses, except Hong Kong Broadband. During the course of the interviews the written responses were reviewed and supplementary questions were asked to obtain clarification or further information on issues not addressed in the written responses. Suggestions from the operators were also explored.

International Standards, such as ITU-T Recommendations relevant to the issues of concern were examined, as were the practices in other developed markets including the UK, Europe, North America and Australia. The results of this part of the investigation are given in Section 6.

3. The circumstances on 11 Sept

From 11.45 am on 11 September 2002, following the public announcement of the imminent issue of a “No. 8” Typhoon Warning Signal by the Hong Kong Observatory, the telephone network of Hong Kong experienced a surge in call attempts well beyond the ability of the network to respond. The overload of the network continued until 3 pm and all parts of the network were affected.

This situation caused a very large increase in call demand, in particular:

- parents needed to call the schools to find out if they needed to collect their children to bring them home or if the children would be kept at school
- businesses and people in general needed to re-arrange their planned activities for the afternoon.

3.1 Network congestion

During the period of network congestion, callers experienced a high rate of call failure, and many call attempts were met with busy tone, a delay in receiving dial tone or no dial tone.

Operator reports indicate that the rate of measured call attempts was two to ten times the normal daily busy hour figures. Several operators suggested that these statistics are an under estimate as they do not capture the full extent of call attempts due to the imposition of call restrictions or, in the case of calls originating on mobile networks, a loss of calls at the air interface before reaching the mobile switch. In any event, once such a level of repeat attempts is reached, the network statistics are no longer an accurate measure of demand.

Once the capacity of the network to carry traffic (or complete calls) is reached, further calls will be rejected, leading to repeat call attempts. These very high rates of call attempts indicate that the network was under considerable stress, itself necessitating action by the operators both to protect their networks from possible failure and to optimise the use of available resources.

In assessing network performance against international best practice, the level of call attempts is not a meaningful measure of network performance for the reasons indicated above. The true benchmark of performance is the level of spare capacity above the normal busy hour traffic loading and the efficiency with which this spare capacity is utilised under unusual load conditions.

The level of traffic carried during the period in question was mostly well above the normal peak traffic level - operators reported traffic increases of 20% to 50%. However, on some interconnection routes, problems and outages were experienced and so the traffic level between some networks was lower overall than normal. The call success rates dropped significantly for calls between other network operators and PCCW. This resulted from a combination of congestion in the PCCW network, as a result of call attempts being rejected to avoid processor overload, and unusually high levels of call attempts meeting called party busy.

3.2 Notification of the announcement to operators

We understand that the normal practice of the Hong Kong Observatory, in making announcements of its intentions to issue a “No. 8” Typhoon Warning Signal, is to make a public announcement of its intentions two hours ahead of the signal being issued. OFTA, along with other Government Agencies, are given notice of the public announcement by the Information Services Department (ISD) so that OFTA can immediately advise the operators and prepare their networks for a likely surge in traffic demand. In the case of PCCW, this preparation includes the manual application of call gapping at the local exchanges to limit the effects of high levels of call attempts on the switches. This ensures that the network can operate to the maximum of their capacity.

OFTA is required by the “Guideline for the Dissemination of Telecommunications Network Congestion Warning” to advise telecommunications network operators within five minutes of receipt of a weather warning message. On this particular occasion, OFTA was notified of the announcement at 11.49 am. This was after the public announcement was posted on the Observatory web site at 11.45 am. As a consequence, the upsurge in traffic had already commenced before OFTA was able to notify the operators, and the operators had no time to make their normal preparations. The delay in making the announcement to operators was a contributing factor to the severe overload.

Statistics supplied by operators relating to similar previous events:

- the “No. 8” Typhoon Warning Signal of 5-6 July 2001 and 25 July 2001 and
- the Black Rain Signal of 1 September 2001,

showed a much lower level of call failures. For example, the maximum traffic volume of one fixed network operator was only 2.4 times the normal peak level, and that of a mobile operator was only 1.44 times the normal peak level. These events occurred outside normal office hours.

4. The Hong Kong networks and their traffic handling

4.1 Introduction

There are five fixed network operators in Hong Kong:

- PCCW
- Hong Kong Broadband Network
- Hutchison Global Communications
- New World Telephone
- Wharf New T&T[1]

and there are six mobile network operators:

- Hong Kong CSL
- Hutchison Telephone Company
- New World PCS
- Peoples Telephone Company
- SmarTone Mobile Communications
- Mandarin Communications (Sunday)

There are direct interconnection links between most pairs of networks but some links are not dimensioned to take all the expected traffic and rely on other indirect routes, such as routes through fixed operators such as PCCW, to take overflow traffic.

Each of the fixed networks has its own number portability database which includes information on both fixed and mobile ported numbers, and is therefore able to route calls directly to the correct terminating network. However, the mobile networks only have information on the ported mobile numbers and normally route calls to fixed numbers to the networks that were originally allocated the number block that contains the called number.

[1] Note: Hong Kong Cable Television Ltd, a sister company of Wharf New T&T, also has an FTNS licence from OFTA and is using the Hybrid Fibre Coaxial cable network for provision of cable modem service, i.e. for Internet access. This company has not been included in the analysis as it does not have a PSTN network as such.

4.2 Network design and traffic handling techniques

The following parts of Section 4 explain the basic approach to network design and traffic handling and are therefore act as an introduction to these issues.

Section 5 describes how the networks in Hong Kong handled the events of 11 September and analyses the observations of the different operators.

Section 6 examines the practices in other countries to see how well the Hong Kong networks are designed and operated in comparison with the networks in other countries and the best international practices.

Basic techniques

International practice is for networks to be designed and dimensioned to handle a certain level of traffic with observed statistical variations. The main design parameters are:

- the number of circuits between the switches (the transmission capacity)
- the processor capacity of the switches which determines primarily the number of call attempts per second that the switch can handle.

The total capacity of the network in normal circumstances is determined by the transmission capacity since the switches are designed to be capable of switching calls on all the transmission circuits connected to them.

The switch processor determines primarily the number of call attempts that the switch can handle in a given unit of time since the main processor load occurs during the establishment of a call and there is little load on the processor during the course of a call. The processor capacity is designed to handle significantly more than the expected number of call attempts so that maximum use can be made of the transmission capacity.

The network capacity is designed from observation of the telephone traffic in Hong Kong and is increased to ensure that during the busy hour (the hour when the network is used most heavily) the probability of a user not being able to set up a call because the network is congested will be less that a small percentage, which is called the Grade of Service (GoS). The target GoS is typically 1 – 2% for a fixed network.

In practice, networks normally have greater capacity than is needed to achieve the design GoS. This spare capacity enables them to handle more calls than occur in normal busy hours. The factors that contribute to spare capacity are discussed in more detail later in this section.

The following parts of this section examine these design factors in more detail.

Transmission capacity

Network transmission capacity is designed to achieve a given GoS based on knowledge of the statistics of the traffic carried. This means that the networks require more capacity than the average busy hour traffic volume so that they achieve a specified probability of being able to handle a new call attempt. Figures of 1% and 2% are typical GoS targets for transmission, which is the percentage of new call attempts that will fail because the network capacity is already fully used in the busy hour. The number of circuits needed to achieve these GoS targets on a medium sized link between switches is such that normally some 5% more circuits are required than would be necessary to carry the average number of simultaneous calls in the busy hour.

Figure 1 shows the design GoS figures claimed by the operators for calls in their networks. In practice when traffic is growing, operators need to procure extra capacity in advance and make allowances for higher growth than predicted and therefore normally have more spare capacity than is needed for the GoS. The figure also includes the operators' assessment of the typical usage of their transmission capacity and switch processor capacity in a normal busy hour.

Figure 1: Grade of service and percentage usage of capacity in a normal busy hour

Operator	Design GoS	Usage of transmission capacity	Usage of switch processor capacity
Fixed			
A	0.5%	70-90%	63%
B	Low	Low	Low
C	1%	80%	Not given but was only 60% at peak of surge
D	0.5%	70%	48% tandem 23% local
E	0.5%	80%	80%
Mobile			
F	5% radio 1% network	55%	Not given
G	1%	80%	Not given
H	2%	70-80%	Not given
J	2%	80%	Not given
K	Refers to regular report	Not given	Not given
L	2% cell level	80%	75%

The figure shows that the fixed networks in Hong Kong are designed to a 1% or better GoS and spare capacity is in the order of 20% since 80% of the transmission capacity is typically used in the busy hour. Mobile networks are designed to a 5% or better GoS and spare capacity is in the order of 20%. These figures indicate that the networks in Hong Kong are dimensioned in accordance with accepted international best practice, as discussed further in section 6.

There are three main reasons why networks may have spare capacity above the level needed to achieve the GoS:

- the first is to add extra capacity as a commercial decision to provide for unforeseen surges. It is unlikely that any network provider in the world makes such provision except for specific contingencies such as new services where the demand may be hard to predict. The spare capacity in this circumstance could typically be 10%-20% in a specific area.
- the second is to allow for traffic growth and uncertainty about when new capacity will be introduced. Operators will add to the forecasts of future traffic demands a small factor of safety to reflect the risk in the forecast (typically 2% to 5%) and schedule the building of the network a short time in advance of the actual demand. The extent of this capacity will depend on the growth rate of a network. Designers are generally pleased when the actual traffic falls slightly short of the forecast. A very new, fast growing operator may have actual traffic loads, which are well short of capacity, giving a spare capacity of 30% or more. (A mature or declining network will not make this provision but may have excess capacity that has either not been recovered or is uneconomical to recover.)
- the third is the modularity of circuit provision which is supplied in units of 30 circuits (E1 links) or 24 circuits (T1 links). Thus a design for 10 circuits will result in 30 being provided, and a design for 310 will provide 330.

Across the entire network a figure of 10% to 30% would be typical and the operators have reported figures in this range to OFTA.

Newer and smaller networks have fewer switches and fewer tandem switches than older larger ones. Consequently they are less tuned to the normal geographical distribution of traffic and may have more flexibility to handle unexpected traffic loads in particular areas. In the case of the typhoon incident, however, it appears that the extra load did not vary significantly from the normal distribution across the region.

On September 11 most operators were able to use their spare capacity to carry more calls and in some cases over 50% more calls were handled.

Switch processor capacity

The switch processor handles a variety of tasks including:

- receiving, processing and sending signalling information related to calls and call attempts, including analysing the called number and determining which route the call should take
- managing the allocation of the switching elements in relation to the calls and their signalling

- reporting management information including charging information and responding to management commands.

The main load on a switch processor occurs during call set up and so the key parameter for the design of the processor is the number of call attempts in the busy hour (Busy Hour Call Attempts – BHCA).

In normal operation, the capacity of the switch processor will be more than sufficient to handle all the calls. However, when congestion occurs and calls are failed, the callers may make many repeated call attempts so that they increase the loading of the processor towards its limit.

Techniques for preventing processor overload

There are two techniques for controlling the processor load under conditions of repeat call attempts:

- ***dynamic or automatic traffic control****,* which is activated automatically when the load exceeds a threshold and de-activated when the load falls below a threshold. Under this condition, new originating calls including 999 calls receive delayed dial tone and new incoming calls are refused and receive network busy tone.
- ***manual originating traffic control (call gapping)****,* where a specified proportion of new originating calls from ordinary lines to specified numbers or number ranges are failed. Priority lines and 999 calls are unaffected. Incoming calls are also unaffected.

In both cases, calls that have already been connected are not affected.

Automatic traffic control is implemented on all switches. Call gapping requires specific software enhancements.

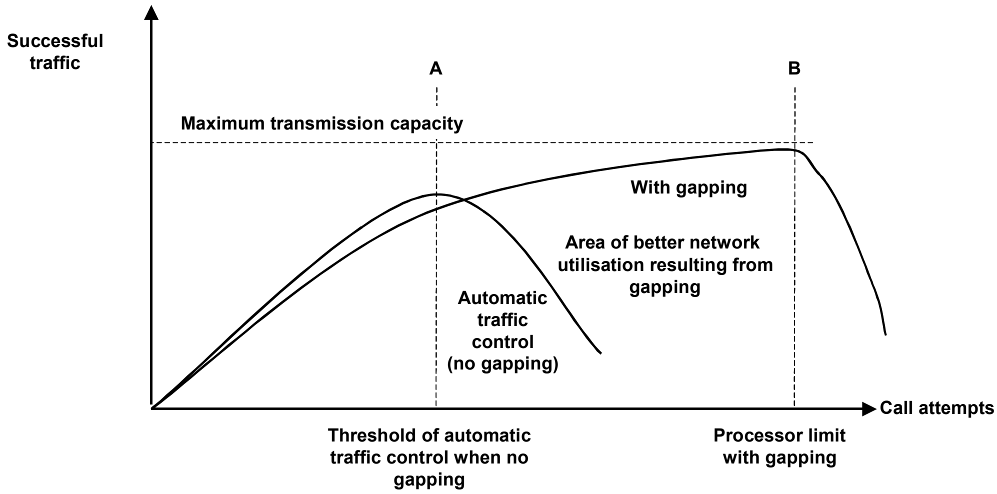
Figure 2 compares automatic traffic control and call gapping.

Figure 2: Comparison of automatic traffic control and call gapping

	Automatic traffic control	Call gapping
Main objective	To protect the switch processor	To protect “normal” traffic from being affected by a surge in demand to particular numbers and so to maximise the continuing use of the network
Applies to	All calls	Specific numbers
Used when processor load is	Typically above 90%	Below the threshold for automatic traffic control
Affects emergency calls	Yes	No
Applies at which switches	All	Normally only at local switches

Figure 3 illustrates qualitatively the distinction between automatic traffic control and call gapping.

Figure 3: Effects of call gapping on successful traffic carried



Without call gapping as the rate of call attempts increases the successful traffic peaks and then falls rapidly because the processor rejects calls indiscriminately to retain control.

With call gapping, as the rate of call attempts to the destinations that are attracting the surge of demand is limited well before the processor reaches the threshold for automatic traffic control. Thus the successful traffic can continue to grow towards the maximum network capacity determined by the transmission capacity. Eventually the load on the processor from applying the gapping algorithm grows to the point where the automatic traffic control starts to apply. Overall the use of gapping gives significantly better network usage shown in the area between A and B. Since the gapping algorithm uses processor capacity, the distance that B is from A increases with processor capacity.

There is a trade-off in the specification of call gapping. More complex gapping algorithms consume more processor power but can focus more precisely on the numbers that are attracting the increased number of call attempts and so can achieve better network utilisation.

Call gapping may be applied at either the local switches or the transit switches. The basic principle is that call gapping needs to be applied at the switch nearest to the origination of the call attempts that is likely to be overloaded by, or experience congestion from the call attempts.

Local fixed exchanges

In a fixed network, subscribers are normally connected to remote concentrator units (RCUs) which are connected by E1 or T1 lines to the main switch. Each RCU serves typically more subscribers than there are transmission circuits to the switch. If all the transmission circuits are busy then a caller will hear nothing when they go off-hook. Their call attempt will be queued and they will hear dial-tone only when a circuit becomes available because one of the other calls has ended. Failed call attempts in these circumstances would not be recorded by the switch. If there are free transmission circuits to the switch, but the switch processor is overloaded, the switch may:

- delay dial tone, ie queue call attempts, so that the rate of attempts falls to a level that it can handle, or
- reject a proportion of call attempts and return the network busy tone to callers (this is called call gapping)

Under these conditions, the switch may only record those call attempts that it tries to handle. Under extreme overload, the switch may cease to record any management or performance information in order to reduce its load further.

Call gapping is implemented on all PCCW local switches. It is not implemented on other networks.

Call gapping is normally always applied by PCCW to its local exchanges when it is notified of the impending announcement of a “No. 8” Typhoon Warning Signal.

Call gapping may be applied to one or all switches by the PCCW network management centre. Its application to all switches would normally take 15 minutes with batches of command instructions being sent to the switches. PCCW is able to specify both the proportion of calls that should be failed and the range of called numbers which should be subject to gapping. This enables PCCW either to apply the technique to a wide range of calls or to focus it on calls to a small block of numbers or a single number. This technique is used to limit the call attempts to numbers used for phone-ins.

Mobile networks

In a mobile network, there is no dial tone. When the caller presses the "send" button after the called number has been dialled, the terminal requests a signalling channel from the radio access network. If no channel is allocated by the base station controller within a given time the terminal will display "network busy". If the base station controller allocates a channel but the mobile switch cannot handle the request, "network busy" would also be displayed.

Operator L explained that they could manually inhibit call attempts at the base station controller to reduce the load seen by the switch, and that they used this manual intervention on 11 September because of the rapid growth in processor load on the switches.

Transit switches

A transit switch under overload would either reject a proportion of calls or signal to the preceding switch that it cannot accept more call attempts, ie it will close a whole E1 or T1 unit of capacity to new attempts for a period.

Where a processor is ceasing to operate in a controlled manner, it may fail to respond to call attempts and the timers in the switch that is sending the attempts may expire. In these circumstances the sending switch will record an alarm. Several operators observed alarms in connection with some of the interconnections to PCCW. Typically the switch may keep trying unsuccessfully to send more calls on the same route. One fixed network operator commented that this is a weakness of the control software since it would be better to attempt a different route. In practice a network manager may notice these events and manually intervene to re-route the traffic.

Analysis

Call gapping has been developed over many decades of network operating experience as the best technique for maintaining network integrity. A network operator has an obligation firstly to ensure the network continues to operate as normally as possible under difficult and emergency conditions and secondly, to ensure that when capacity is limited or overloaded, that the network is rationed equitably. The principle of call gapping is to discourage the unnecessary use of network facilities on calls which are highly likely to fail and conversely to ensure the processing capacity is directed to calls most likely to succeed. This is most important for calls passing through several switches, where the inefficiencies will multiply.

Most large fixed networks (typically the incumbents) are built with a large number of local exchange switches and a number of tandem switches. Other switches perform gateway switching functions to networks in other countries and other regions (not applicable in Hong Kong) and to competing networks in the same country.
Thus calls will often be passed from the calling customer's switch to another switch to another switch and ultimately to the called customer's switch. Most connections will pass through two or more switches. During call setup all these switches are called on to do processing and if the probability is high that the call will fail, there is a huge amount of unproductive processing being undertaken, to the possible exclusion of other calls.

Once identified the rate of calling to a difficult location or number can be controlled within the capacity by suppressing a percentage of calls or imposing a gap between calls, hence call gapping. Network Management monitoring is able to detect a high failure rate in particular parts of the network and reduce the rate of calling to these destinations. In the case of a typhoon warning it has been practice in Hong Kong to gap calls to all destinations. This process is well understood by PCCW as the incumbent operator with a large fixed network and a complex tandem hierarchy.

Some operators have argued that PCCW should apply call gapping in its interconnection gateways and that this would be a cheaper alternative to introducing gapping in most of the other networks. Mobile operators have also argued that their limited access capacity makes gapping unnecessary. We doubt the validity of these arguments for the following reasons:

- interconnected networks should operate in the same way as a single monopoly network, since PCCW implements gapping at the local exchanges then so should the interconnected networks. For the optimal use of the network, it is clearly desirable for all operators to employ a consistent strategy.
- mobile networks are as pervasive as fixed networks and need call gapping for the same reasons
- gapping by the PCCW gateways would cause congestion and processor overload from the resulting repeat call attempts in the interconnected networks

This analysis suggests that if gapping is required then the only exemptions should be:

- networks that are significantly over dimensioned
- networks with a customer base that is unlikely to call numbers that would result in frequent repeat attempts.

The effective use of call gapping across networks will, however, require much improved cooperation between the operators in traffic management.

Call gapping will require additional investment by the operators. We recommend, therefore, that OFTA review the overall risks, benefits and costs before imposing any requirement for call gapping. We do not exclude the possibility that overall it may be acceptable for operators not to make the investment in call gapping, and to live with the risk of another serious congestion problem, since the current level of risk will be reduced by the implementation of other recommendations in this report.

4.3 Operational interaction between the operators

Each network operator normally has a network management centre with detailed real-time information on the status of their network, the levels of traffic and the occurrence of various network alarms. The operators were asked about the extent of the communications between these centres and the answers showed that there is less communication and sharing of information about traffic overloads than is needed and than occurs in some other countries with multiple interconnected networks.

The best approach to handling processor overloads is to reduce call attempts at the origination of the calls by using call gapping. With interconnected networks this means that the operator of a network that is experiencing an overload should request the other operators to apply call gapping, and this requires openness and regular communications. Several operators have indicated a wish to improve the cooperation but a problem may be that if an operator is open about the performance of their network then the other operators may use this information in their marketing to show that the other network is inferior. Operators should sign a confidentiality agreement to protect against this possibility.

Only the fixed network operators use dedicated lines for communications between network management centres, others such as the mobile operators use the normal

PSTN. Since network problems such as occurred on 11 September could prevent management communications over the PSTN, we recommend that all operators use dedicated lines or a private network for these communications.

4.4 Access to emergency services and priority systems

The events on 11 September resulted in an increased number of calls to the emergency services but, according to a report by the Commissioner of Police to OFTA, there were no signs of irregularities on the 999 system between 12.00 and 14:59.

The average number of calls terminated and answered on the 999 system during this period were 85% and 63% respectively above those on a normal day. By comparison, calls to emergency services during the Black Rain Signal of 1 September 2001 showed the number of calls answered and terminated on the 999 system to 33% and 24% respectively above those on a normal day.

Figure 4 shows the information obtained from the operators. All emergency calls are directed via the PCCW network to the Police Emergency Centre. The fixed operators all use dedicated routes for the emergency calls at their interconnections to PCCW but of the mobile operators only operator G does this; the others route the calls with their normal traffic.

Figure 4: Handling of 999 calls

Operator	Calls routed to	Handling in own network	Handling at interconnection
Fixed			
A	Police emergency centre	Dedicated routes for 999. Call attempts not affected by controls if dial tone is given	Dedicated routes from other fixed operators Normal routes from mobile operators
B	Police emergency centre via PCCW	No information	Dedicated routes to PCCW
C	Police emergency centre via PCCW	Dedicated routes	Dedicated routes to PCCW
D	Police emergency centre via PCCW	Dedicated routes	Dedicated routes to PCCW
E	Police emergency centre via PCCW	Dedicated routes	Dedicated routes to PCCW
Mobile			
F	Police emergency centre via PCCW	No priority	Handled with other traffic
G	Police emergency	Dedicated routes	Dedicated routes to

	centre via PCCW		PCCW
G	As above	As above	As above
H	Police emergency centre via PCCW	Priority at the air interface	Handled with other traffic, but being reviewed
J	Police emergency centre via PCCW	Priority routing for 112 not 999	Handled with other traffic
K	Police emergency centre via PCCW	Priority handling	Handled with other traffic
L	Police emergency centre via PCCW	Priority at the air interface	Handled with other traffic

Operator A has a “priority system” at present. Normally, a priority class user has priority over other users for obtaining dial tone and a call from a priority user will not be blocked under automatic traffic control measures from receiving dial tone. In general, priority lines will have a higher chance than ordinary lines to receive dial tone to make originating calls when the local switch is under heavy load conditions and the traffic control mechanism is activated, provided that the system resource is not exhausted.

It is important to note that if a large number of users are given priority class that this will limit the effectiveness of the priority system. Operator A indicated that the number of priority class lines should be limited to no more than [ ]% on any single switch in order to maintain an acceptable GoS. Currently we understand that there are approximately [ ] priority users in total in Hong Kong.

Some essential departments and hospitals have priority lines and there was no service interruption on these priority lines in the September 11 incident.

It is technically feasible for mobile operators to provide a priority service but they do not provide such services to their customers commercially.

All GSM mobile operators provide priority access to channels at the air interface for calls to the international emergency number (112). In addition, a number of mobile operators have programmed their networks to give priority access to 999 calls.

We recommend that OFTA consider reviewing the arrangements for emergency calls, with a view to:

- enabling all authorities that are likely to be involved in an emergency to have priority access to the fixed networks
- seeking a consistent approach by all operators to providing priority to emergency calls. It should be noted, however, that due to different mobile systems and handset limitations, it may not be possible to implement a fully consistent approach across all networks.

If the authorities require such priority access as part of their procurement of telecommunications services, then OFTA will not need to direct the operators to offer it as they would be expected to do so on commercial grounds alone.

We do not think that each network should have separate direct access to the Police Emergency Centres provided that emergency calls have priority. However, we do consider it to be desirable for these calls to have dedicated routes to at the interconnections to PCCW. This is particularly the case if the switches do not give priority or are vulnerable to automatic overload protection measures. Where both priority and call gapping are implemented then dedicated routes may not be needed as it should always be possible to connect the calls reasonably quickly.

4.5 Preference schemes

Preference schemes are different from priority schemes. Preference schemes enable access to and use of the network to be limited to certain identified users in times of major disasters when the network is damaged and the emergency authorities have much greater demand themselves for telecommunications. Thus preference schemes serve the “post-disaster” scenario that is quite different from the “warning” or “false alarm” scenario that occurred on September 11.

Different countries provide different schemes with different levels of preference. For example one level would limit calls to those between emergency organisations, hospitals and governmental organisations, whilst another would include essential private businesses and the private telephones of key employees. The preference scheme was used in the New York area immediately after the Twin Towers disaster.

ITU-T Recommendation E.106 describes an international emergency preference scheme. ETSI and ITU-T are studying the development of preference schemes for next generation networks

The networks in Hong Kong do not provide preference schemes at present.

The information provided as part of this study is not sufficient to make a recommendation about the need for a preference scheme in Hong Kong. More detailed study is needed that takes account of:

- the possible disaster scenarios
- the availability of private and dedicated Government networks to handle traffic between emergency organisations
- the cost of implementing such a scheme
- the cost of maintaining such a scheme, in particular maintaining an accurate and up-to-date list of the numbers that require preference.

5. How the networks responded on 11 September

This section describes and analyses how the networks handled the traffic demand on 11 September. PCCW and most other operators were fully cooperative and provided useful information. However, a minority of operators were less cooperative and provided only limited information. We are, nevertheless, confident that the information collected is adequate to support the main conclusions and recommendations.

5.1 The response of networks to congestion

On September 11, all operators reported a reduction in the Answer Seizure Ratio (ASR) or call completion rate as would be expected in conditions of high call failure. Nevertheless all operators also reported carried traffic to be well above normal busy day levels, indicating that network spare capacity was being utilised. In addition the higher general level of calling will increase the probability of the called party being busy, again contributing to more call attempts and the lower success rate.

There is no evidence that call failures were more apparent in any particular geographic areas of the network. The peak of the congestion was variously reported as 12:00 to 13:00 or 13:00 to 14:00 and possibly varying according to the location or simply by switch. Most operators report traffic back to normal levels by 15:00. It appears and is perhaps not surprising that the peak demands moved around as people moved, with some evidence that the peak moved from the office areas to the residential areas during the few hours of congestion.

5.2 The Hong Kong Network

The interconnect arrangements of the networks in Hong Kong are particularly complex, the number of individual routes between operators being very high. This results from the large number of networks arising from the advanced state of competition.

There are five fixed and six mobile network operators, each potentially seeking interconnection with each other. Consequently there could be up to 30 one way interconnection paths. Not all these paths will be established as calls between two networks may be carried through a third network acting as a transit switch for the two operators. However, as each operator has implemented a number of gateway exchanges and most of these operator-to-operator paths comprise several routes.

PCCW remains the dominant network in Hong Kong. All other operators send and receive the bulk of their traffic to and from PCCW, and so a very large proportion of those calls that involve other operators have one end in the PCCW network. Furthermore operators who have direct routes between each other also send overflow traffic via PCCW as a transit operator, making PCCW's network particularly vulnerable to peaks in demand. Another large proportion of calls remain entirely in the PCCW network. From this we can conclude that the overall network performance is heavily, and to some extent disproportionately, weighted by PCCW performance and that other operators will be highly dependent on PCCW's performance.

All operators will have particular concerns about the gateway performance, since they lose control of the call when it passes across this boundary. Most attention from the operators other than PCCW was focused on the perceived performance of the PCCW gateway exchanges and in particular on a few particular locations. These locations were not consistent across all operators indicating that there is a varying dependence on particular gateway switches. A number of the new operators indicated that they observed a higher success rate on calls passed on direct interconnection (where this was available) when compared to calls through PCCW, although this would be influenced by the nature of the traffic as well as the existence of overflow traffic to PCCW. One operator reported a drop in circuit utilisation rate to one PCCW interconnect switch for about one hour suggesting call processing was severely limited in this case. A number of instances of problems with signalling links were also raised.

5.3 Network dimensioning

The networks in Hong Kong are dimensioned for the normal busy hour loads and normally have some capacity to carry additional traffic. This spare capacity is in both circuits and call processing resources, including the signalling. Most operators report spare capacity in the range of 10%-30% for circuits and 50% to 100% in call processing capacity, when compared to normal peak loadings.

Nevertheless, the ability of a network to carry increased traffic loading is very much dependent on the distribution of that increased load. In the case of this incident, it appears that the distribution of increased load was network wide and that in this instance much of the spare capacity was utilised, resulting in a significant increase in carried traffic over the normal peaks.

Some operators reported a reduction in successful Mobile Number Portability queries, possibly indicating difficulties with signalling processing.

5.4 Local fixed exchanges

Figure 5 summarises the techniques used in the fixed local exchanges, the typical increase in call attempts and whether or not call attempts were limited at the local exchanges. All the fixed networks experienced a substantial increase in the number of call attempts and several of the networks employed delayed dial tone to reduce the rate of call attempts. The figures for the increase in call attempts are likely to be underestimates as the statistics relate only to the events that the switches could detect and record.

Figure 5: Traffic at the local exchange for the fixed networks

Operator	Processor protection Technique	Is gapping supported	Increase in call attempts recorded by switch at peak	Technique used
A	Automatic protection that produces delayed dial-tone, and manual traffic control (call gapping)	Yes	5x	Automatic protection was triggered in most local switches before manual traffic control could be applied. Manual traffic control was applied but its application took up to 30 minutes in the most heavily used switches
B	None	No	Not given	None
C	None, since the processor is dimensioned to handle all the attempts that can be sent from the RCU and the signalling link capacity limits the number of attempts	No	Not given	None, but congestion did occur in the local access
D	Automatic protection that produces delayed dial-tone	No	4x	Delayed dial-tone for 10-20% of callers
E	Automatic protection that produces delayed dial-tone	No	2.5x	Delayed dial-tone Switches had 12% overload

A full description of the interactions between the PCCW local switches and the networks is given in section 5.6.

5.5 Access to mobile networks

Figure 6 summarises the situation for mobile access. There was a substantial increase in the rate of call attempts and several of the switches triggered their automatic overload protection. Operator L reported also applying manual methods to limit call attempts at the base station controllers before they reached the switches. The statistics are very probably substantial underestimates because only those attempts that were signalled to the switches would be included. Some callers clearly experienced great difficulty in making calls.

Figure 6: Traffic at the mobile access networks

Operator	Is gapping supported	Increase in call attempts recorded by switch at peak	Overall effect for callers
F	No	4x	Inability to get signalling channel so network busy displayed
G	Overload control mechanism was triggered automatically to maintain call processing capability. Emergency call was given priority treatment under such condition	3x	Overload mechanism was triggered to maintain call processing capability under high load condition No signalling failures Reported no congestion on signalling or control channels
G	No	5x	Call setup failure rate due to own network issue was very low
H	No	2.5x at radio access 4x at switch	Overload mechanism rejected some call attempts
J	Probably not. Information not explicit	<4.5x	Automatic overload protection is provided
K	No	10x at radio access 4x at switch	Automatic overload protection is provided
L	No, but there is a manual technique used at base stations	4x at radio access 1.5x at switch	Automatic overload mechanism rejects some call attempts at the switch. Manual methods were also

			applied to base station controllers to reject call attempts There were two peaks when up to 50% of calls attempts were rejected, mostly at the BSC. Blocking on the control channel was low, ie there were enough radio resources

5.6 Network and interconnection congestion

Figure 7 summarises the situation within the networks, ie between the switches and over the interconnection links. The figure includes both fixed and mobile networks.

Figure 7: Network congestion experienced by operators

Operator	Own network calls	Interconnect to PCCW	Other interconnects
Fixed			
A	Affected by automatic protection on most local switches until manual traffic control was applied	N/a	Some gateway exchanges applied automatic protection reducing the volume of calls handled by the gateways even on routes that did not carry excessive call attempts
B	No problems	No problems for o/g but answer ratio dropped dramatically Noted drop in i/c traffic from operator A Fault detected at PKT and TMN but not at Jordan. No connection to KGR	No problems with operator E
C	No problems	High volumes of alarms triggered by absence of response to signals, especially on Jordan, also on KGR and TPC	They are dimensioned to be full under normal loads and overflow to operator A . No problems noted
D	No problems	Links to KTN were down for 2 hours mostly affecting transit traffic from operator H	No problems

E	No problems	Congestion on links and call success rate dropped	No problems
Mobile			
F	Processors were overloaded	Processors were overloaded and blocked 30% of incoming capacity	No problems reported
G	No problems	Some congestion and call success rate dropped from 3% to 50% but may be due to called party being busy Would not say which interconnects were affected	Some congestion and call success rate dropped from 3% to 50% but may be due to called party being busy Would not say which interconnects were affected
G	No problems	No problems reported	No problems reported
H	No problems	Call success rate dropped to low figure Incoming trunks were not fully loaded suggesting that operator A was restricting calls Outgoing trunks were not congested because call success was low Signalling problem with PKT only	No problems reported
J	The successful answer rate in the air interface dropped by 60% during 12:00 to 13:00 compared with the same period one week earlier	Successful answer rate for traffic through operator A dropped by more than 80% during 12.00 to 13.00 compared with the same period one week earlier. C7 links to KGR were faulty (100%) from 12:00 to 15:00 1/3 of circuits to Jordon were blocked between 12:13 and 16:09	No problems reported
K	No problems	Problem with PCCW	No problems reported

	reported	KGR	
L	All switches were overloaded. Call success rate dropped to 64%	Lengthy signalling outages were experienced to PCCW KGR and PKT Call success rates to these exchanges dropped to around 10-12% at worst hours	There were problems with links to operator C

Because the operators were not given advance notification of the announcement of the "No. 8" Typhoon Warning Signal before the public became aware of the announcement, the volume of call attempts grew very rapidly and triggered automatic protection in most of PCCW's local exchanges. PCCW and the mobile networks probably experienced a greater surge in demand than the other fixed networks because they carry the highest proportion of local calls.

The automatic protection caused the switches to delay dial tone.

Automatic protection was also applied in the PCCW tandem exchanges and interconnection gateways, blocking incoming calls from other switches, reducing the success rate for incoming calls and reducing the useful traffic carried over the interconnection links. The worst affected interconnection gateways were Jordan (JDN) and Kings Road (KGR). The frequent occurrence of automatic protection on some interconnection gateways gave other operators the impression that they had failed completely.

The increase in call failures was experienced by all networks, but especially by the mobile networks, since most calls were to PCCW. The increased failures led to increase attempts that in turn led to automatic protection being applied in other networks. Some other operators then compounded the problem by blocking the incoming routes from PCCW thus reducing the overall capacity of the networks and the call success rate thereby aggravating yet more call attempts from PCCW customers. These operators should instead have asked PCCW to apply gapping to their number ranges but the situation probably developed too quickly for them to do so and they may not have been an adequate climate of cooperation for them to do so.

PCCW started to apply manual traffic control (call gapping) but the overload of the switches was sufficiently severe that some had disabled their management functions and it took 20-30 minutes to establish manual control and start to bring the situation under control.

Relatively few problems were experienced with the interconnects between operators other than PCCW. Most operators have direct interconnections to most other operators, but some dimension these interconnections to be fully loaded and overflow traffic via PCCW making PCCW especially vulnerable to high peaks in traffic demand.

Several operators noted the sharp reduction in the success rate for calls including the answer-seize ratio[2]. We believe that this was due to two factors:

- Some interconnect gateways in PCCW rejecting call attempts as part of the overload protection mechanism that would have been triggered partly as a result of the peak of overflow traffic and the absence of call gapping in the mobile networks
- Many callers attempting to call the same numbers, eg schools, and receiving busy signals

Most operators have their own number portability databases and query these databases before routing calls so that they route them to the destination where the caller is currently served. The fixed network operators include both the fixed and mobile ported numbers in their databases, but the mobile operators only include mobile ported numbers because they are not given access to the information of ported fixed number. They are dependent, therefore, on the fixed network operators for this service. Where a mobile network operator has direct connections more than one fixed network operator, calls are routed to a particular fixed network operator according to:

(a) the original network operator of the called number

(b) the traffic volume in the possible paths

(c) the interconnection charge.

After database interrogation by the selected fixed network operator, the call is routed directly to the final destination by that operator.

Since most ported fixed numbers have been ported from PCCW to the other operators, congestion or outages on the interconnection links to PCCW will prevent the routing of calls to numbers that have been exported to their networks. This problem would have been overcome if the mobile network operators had access to the information on ported fixed numbers.

The result of these arrangements is that calls from mobiles to ported numbers on fixed networks are more susceptible to congestion, especially between the mobile operators and PCCW, than calls to non-ported (native) numbers on the same destination network. Whilst in principle we think that the mobile operators should have access to the same information on ported numbers as the fixed operators have, we note that:

- the proportion of ported numbers on the fixed networks is not high
- the measures recommended elsewhere will go a long way to avoid problems of congestion on the interconnection links re-occurring.

[2] The answer seize ratio is not a good measure of performance in the extreme overload conditions experienced, but it is a measurement quoted by several operators.

We therefore recommend that OFTA should take this limitation into account when it next reviews the number portability requirements to see if the limitation can be lifted.

We do not think that the limitation is such a problem that is requires an immediate review of the number portability requirements.

5.7 Short Message Service

All Mobile operators reported an increase in SMS traffic, when compared to normal peak day loads, although not as dramatic (typically 1.5 to 2 times) as the increase in telephone call attempts. This increase in traffic was within the capacity of the SMS system, which allows messages to be queued and so does not suffer from repeat attempts in the same way that voice calls do. Because the SMS messages use the signalling channels of the mobile networks, SMS messages can be sent even when all the voice channels are in use.

Figure 8 shows the increases in SMS traffic detected by the operators and their assessment of how their networks performed. The assessments of performance apply the networks ability to deliver messages to its own customers or to the SMS gateway through which the SMSCs are interconnected. Operators do not have accurate information on whether messages sent to other networks were delivered satisfactorily in those networks.

Figure 8: Performance of SMS service

Operator	Increase in load	Performance
F	2x	Peak demand was within capacity available but delivery delay increased to 1 hour
G	1.5x – 2x	No significant decrease in successful rate
H	1.5x	Normal operation was maintained with delivery times under 1 sec on own network (plenty of spare capacity)
J	1.6x	Services performed normally
K	No information	No problem reported
L	1.6x	Senders would have experienced problems on radio access Overall performance measurements not available

The SMS processing system is somewhat independent of the voice call processing. SMS messages are transmitted to and from mobile phones using a signalling channel. The signalling channel is used to register the phone (so the network knows where the owner is to be located) and to setup and take down calls. This signalling channel is always available, when a mobile phone has signal even though the voice channels at the local base station may be occupied.

During congestion such as experienced on September 11, there are many more call attempts being made than is normal. Once all the voice channels are occupied any other call attempt will fail as will call attempts to mobiles in that base station area. The signalling channel is designed to have far more capacity than is needed for signalling and normally capacity exists for the SMS messages. Even when the signalling does become congested as it may for short periods, SMS messages being of lower priority, simply queue until they can be delivered. The SMS system is a distinct processor or server, which has its own capacity limits but is not involved in call processing. As a caution, it should be noted that delivery of SMS messages is not guaranteed by the network, although in practice messages are not normally lost.

SMS is a simple and effective form of communication that is well suited for meeting much of the traffic demand that occurred on September 11 and its design as a packet based store and forward system handles congestion well and avoids the problems of call failure and repeat call attempts. The use of SMS could, therefore, be considered as an alternative method of communications between members of the public in times of emergency and network congestion.

However, SMS is much better suited to communications in languages that use the Roman alphabet than those that use characters such as Chinese. For example, the time taken by the user to input Chinese characters may be longer than for the Roman alphabet. In addition, the local SMS is limited to 160 alpha-numeric characters and 70 for Chinese characters. The potential value of SMS for emergency purposes could, therefore, be limited in Hong Kong.

6 International practice

OFTA asked Ovum to compare the design and traffic handling techniques of the networks on Hong Kong with those in other countries and with the requirements of international standards to see if the networks follow the best international practice. In order to make this comparison, Ovum has:

- investigated the standards on network design, including both transmission capacity and switch processor capacity
- investigated the practices for managing increased traffic loads that are used by operators in other countries
- investigated to see if similar events have occurred in other countries from which lessons can be learned

6.1 Network design

Transmission capacity

Transmission capacity is measured most commonly as the Grade of Service, which is the probability of a call being blocked by lack of network capacity in the busy hour.

The Telecommunication Sector of the International Telecommunication Union (ITU-T) has published a number of Recommendations on the design of networks in terms of their transmission capacity.

Fixed networks

The most relevant recommendation for the Hong Kong networks is Recommendation E.721, which gives "Network grade of service parameters and target values for circuit-switched services in the evolving ISDN." Table 2/E.721 lists the target values for a range of GOS parameters of which the key parameter is the probability of end-to-end blocking. The target values are shown in Figure 9.

Figure 9: Target blocking probabilities specified by ITU-T for end-end connections in fixed networks

	Normal load	High load
Local connection	2%	3%
Transit connection	3%	4.5%
International connection	5%	7.5%

The performance of an end to end connection will depend on the number of switches that the connection passes through. The local connection assumes up to 4 switches, whereas the transit connection assumes 5-7 switches. Connections that traverse more switches will experience a worse grade of service and higher probability of blocking, than calls that traverse fewer switches because the blocking effects are cumulative. Consequently the design targets for individual links will be lower than the end-end targets.

In terms of the ITU-T Recommendation and the number of switches involved, the Hong Kong network is essentially a large local network, so the design target values of 0.5% - 1% quoted by Hong Kong operators are better than the target of 2% given in this Recommendation.

Mobile networks

For mobile networks, ITU-T E.771 "Network grade of service parameters and target values for circuit switched public land mobile services" specifies target values.

For mobile networks, the probability of end-to-end blocking, is dominated by the probability of not having a radio channel available. Due to the statistical nature of telephone traffic, the relatively small number of channels provided at each base station and the priority given to hand-over calls, it is usual to allow a higher blocking rate of perhaps 5% in this part of the network.

Recommendation E.771 states that blocking in existing networks is typically:

- 5-10% on radio channels
- 1% on connections to fixed networks.

It sets an objective for mobile networks to improve to the following blocking levels:

- 1% on radio channels
- 0.5% on connections to fixed networks.

Australia

The Australian Communications Industry Forum (ACIF) specifies in Section 7.11.5 of G502:

- access network: GoS should be 0.2 – 1.0%
- transit network: GoS should be 0.2 – 0.5%
- mobile access radio channels: GoS should be 5 – 10%, but the long term goal should be radio channel blocking of less than 1%
- mobile to fixed: GoS should be 1%, but the long term goal should be 0.5%

Ireland

An early draft of the Eircom reference Interconnection Offer required the interconnection links to be dimensioned to a GoS of 0.5%.

Switzerland

The Swisscom Manual of Operational Procedures specifies a 1% GoS for the interconnection.

United Kingdom

The BT Standard Interconnection Agreement requires traffic forecasts to be based on a GoS of 0.8% for each route with 2% at 10% traffic overload and 5% at 20% overload.

The C&W Interconnection agreement, Document 6644-1, requires the traffic routes to give a GoS of better than 0.8%.

In its determination on interconnection in 1985, Oftel drew attention to the need for adequate dimensioning and Grade of Service but did not specify any target figures.

Conclusion

The design targets of the networks in Hong Kong and of their interconnection arrangements are marginally better than those recommended by the ITU-T and are commensurate with the design practice in other countries.

Switch Processor Capacity The design objective for the switch is that the processor should be adequate to:

- handle the maximum volume of traffic that the network can handle and that is determined by the transmission capacity of the routes into and out of the switch
- handle a reasonable volume of additional call attempts such as could occur during a short burst of peak demand

The switch processor is therefore not normally the factor that determines the maximum capacity of the networks and therefore the processor capacity is not specified in the ITU-T Recommendations. ITU-T Recommendation Q.543 does, however, give a general description of how processor loading should be handled and it gives the following guidelines:

- *give preference to the processing of terminating calls;*
- *give preference to priority class lines, calls to priority destinations based on digit analysis and incoming calls with priority indications in, for example, the Initial Address Message of a call using CCITT Signalling System No. 7, if an essential service protection capability has been invoked;*
- *defer some or all activities non-essential to handling offered traffic; examples are some administration and maintenance processes in the exchange. (Nevertheless the man-machine communications essential for priority operational tasks should always be preserved. In particular, network management terminals and functions associated with interfaces to network management support systems should be*

afforded high priority, since network management actions can play an important role in reducing exchange overloads);

- *maintain normal charging and supervisory functions, and established connections until the receipt of the appropriate release signal;*
- *assign priorities to specific exchange measurements, such that low priority measurements cease at a predetermined level of congestion. Higher priority measurements may be ceased at a higher level of congestion, or may be run continuously, depending on their importance to the call handling functions;*
- *give preference to calls already being processed, before accepting new calls.*

The spare processor capacity in switches during the busy hour will vary significantly from switch to switch depending on local circumstances because processor capacity is procured in comparatively large units and each procurement will consequently take account of the traffic growth that is forecast for that switch for the next 2-4 years. Thus a relatively new switch in an area where significant traffic growth is forecast might have say 50% spare capacity whereas an older switch might have only 15-20%. Furthermore the spare capacity needed to handle peaks in call attempts will depend on the burstiness of the traffic and the quality of the network management. A well organised and managed network can handle peaks of demand with less spare capacity than a poorly managed one. Facilities such as call gapping also enable networks to operate with less spare capacity. It is therefore impossible to make very meaningful comparisons with networks in other countries.

We have discussed the issue of processor capacity with C&W who have studied processor management over some period to establish the best way to operate their network. They say that the maximum normal busy hour loading for a switch varies significantly depending on the traffic type and the switch architecture, eg whether allowances need to be made for CPU failures in switches with a cluster of processors, but they suggest that:

- 70% is a typical safe loading for a local switch processor
- 75% is a typical safe loading for a transit switch processor

The tasks of a transit switch are less varied and it is protected to some extent by the local switches and so a higher loading can be used safely.

The figure quoted by operator A as a typical switch processor loading is 63%, which is comfortably within the limits of good practice, however the loading will vary significantly from switch to switch and we cannot exclude the possibility that some switches are operated with insufficient capacity margin.

The approach to switch processor loading is broadly in line with established international practice and the methods for handling overloads are in line with the guidance given by the ITU-T. However the ITU-T is not prescriptive about the margin of spare capacity to be carried and the margin in the various switches, especially in the operator A network will vary from switch to switch and we cannot exclude the possibility that the margin is insufficient in some switches.

6.2 Network Management

Cooperation between operators

Interconnected networks need to cooperate to manage traffic in a way that makes best use of the available network resources. If there is not close cooperation and ready sharing of information, the performance of the ensemble of interconnected networks will be poorer than if the networks were operated by a single entity. Thus there is a risk that the introduction of competition and interconnection will result in poorer performance.

Cooperation involves the following elements:

- sharing advanced information about expected increases in traffic demand so that routing can be adjusted if necessary to meet the demand or call gapping be applied in advance to specific numbers or number ranges
- giving advanced warning of maintenance that will result in network outages so that re-routing can be arranged
- discussing in real time the management of unforeseen problems

Experience in other countries shows that it takes time for operators to establish the degree of trust needed for good cooperation in network operation between competitors. One of the biggest problems is that the staff concerned do not understand how these issues should be handled.

We have discussed the extent of cooperation with the operators in Hong Kong. Our findings are that:

- there is adequate cooperation about planned maintenance
- there is not enough cooperation abut increases in traffic and the management of unforeseen problems
- the cooperation is reportedly better between the fixed operators than between the mobile and fixed operators.

Operator A has informed us that they use dedicated lines for exchanging network management information between the fixed operators but not between the fixed and mobile operators who use normal dialled calls that are themselves subject to congestion.

Australia

The degree of cooperation between the operators in Australia is similar to that in Hong Kong. In Telstra, there is an inter-carrier group within the Network Management area, which liases with other operators. While they work together, there has not been much need to co-operate in real time, apart from reporting faults to each other and general advice of the situation.

Ordinary PSTN calls are used for communication between the network management centres.

Germany

Whilst no detailed information is available, we have been informed that the German operators cooperate extensively at the technical and operational level to ensure a good service to the public and this suggests that there is better cooperation than in Hong Kong.

United Kingdom

According to C&W there is very good cooperation between the operators in UK with information on expected traffic loads being exchanged freely. This is especially important as many of the traffic terminations for mass calling events are on the C&W network whereas most of the callers are on BT. The network management centres aim to avoid staff changes and to develop good personal relationships with each other in order to facilitate growth in experience and good cooperation.

Ordinary PSTN calls are used for communication between the network management centres.

Singapore

SingTel continues to be dominant operator in Singapore. Historically other networks have had little impact on traffic patterns during mass calling events but there is cooperation between operators as required. Interconnect circuits are monitored and communications and co-ordination will take place between the parties if a problem arises. Any action is normally left to the discretion of individual operators.

Canada

The Canadian Radio-Television and Telecommunications Commission (CRTC) publish Network Management Guideline for interconnecting carriers in Canada. The guidelines address management of events or conditions such as:

- network congestion due to abnormal calling
- mass calling events
- network congestion due to facility failures.

Specific network management procedures require pre-planning and negotiations between interconnecting carriers.

Conclusion

The cooperation between the network management centres in Hong Kong needs to be improved and is not fully up to international best practice. Lack of coordination resulted in some operators taking actions on September 11 that were counter productive. The introduction of call gapping will make good cooperation even more important.

The introduction of dedicated communications links between the network management centres, which we are recommending, will put Hong Kong ahead of the current best practice.

Techniques for managing increased traffic loads

ITU-T Recommendation Q.542 gives general guidance on techniques for handling increases in traffic demand including call gapping.

The most common causes of short-term surges in traffic demand are organised mass calling events such as phone-ins that are organised in conjunction with television programs. These events have caused various problems with network overloads since the mid 1980s, and operators and manufacturers have developed call gapping techniques as the best means to manage the resulting traffic surge and to prevent the networks from becoming overloaded.

PCCW uses call gapping within its network with gapping applied at the local switches. Call gapping is not used in other networks in Hong Kong at present.

Typical switch vendors support a range of capabilities under Network Management Features in local, transit and gateway switches. Their actual definition and mode of operation will vary in the detail. For example:

- **Call Gapping:** Sets a limit on the number of call attempts to a specified destination code in a particular period of time. So for 10 calls per minute, the gap would be 6 seconds and only one call would be processed each 6 second window. The gap can be varied from 0 (no blocking) to minutes or no calls at all. Callers who fail are given busy signal or possibly an announcement.
- **Code Blocking:** Uses a percentage technique to block from 0% to 100% of calls to a destination in steps. Blocked calls are given busy or an announcement. There may possibly be thousands of destination codes controlled in this way.
- **Calling rate control:** Sets a limit on the calling rate to a destination. By setting a rate of so many calls per second, the control will not apply if the calling rate is instantaneously low and is therefore not as rigid as the gapping or blocking.

These are all similar features and are not all necessarily available from every vendor. These features are normally activated by MML (Man Machine language) commands over a data link to the switches in a network. The normal procedure would be for the Network Management Centre to put these controls in action in the early stages of an event if at all possible.

Australia

Telstra, the dominant operator, maintains a robust network with a high level of inbuilt re-routing capability. In the event of natural disasters, they apply call gapping outside the area affected to ensure that local traffic is given the best chance to complete, with priority given to calls originating in the area affected. Manual call limits are also applied to the exchanges in the affected area to protect the processors from overload (in addition to the inherent overload protection in the switch).

Telstra use both call-gapping and circuit blocking to control traffic, along with many other alternatives such as re-routing through under-utilised capacity.

All these techniques are applied for mass calling events such as quiz lines and telethons.

United Kingdom

Call gapping is the normal approach to controlling traffic in the UK network for mass calling destinations and for real time unforeseen events such as disasters.

C&W Communications specialises in the market for terminating mass calling services. They are in regular contact with fellow operators to schedule call gapping to known mass calling numbers, typically 10 to 20 events per day. Information is exchanged as a formal request using e-mail. For emergency action for an unforeseen event, requests are made by phone and followed up in writing. Communication between Network Management Centres is by ordinary PSTN lines.

Messages to apply call gapping are usually prepared in advance as a batch process for the scheduled events and are sent to the switches remotely..

Whilst both BT and C&W can apply gapping, some smaller operators cannot.

C&W Communications has the ability to block circuits but they do not do this for mass calling as it is likely to provoke overloads in the signalling.

BT applies call gapping at the local switches and can also apply it at the DMSU transit switches.

Singapore

SingTel employs call gapping as the preferred technique for controlling overload. For pre-planned mass calling events such as prize lines and televoting, SingTel has mass calling tandems, whereby calls to the associated special number range will be choked when the capacity to these tandems is filled. For unforeseen events or where they get short warning for increased load on ordinary numbers, then call-gapping or "code blocking" is applied as soon as possible or just before the event.

Canada

The CRTC Network Management Guidelines address a range of procedures to be employed during conditions of network overload. These include:

- code-blocking
- call-gapping

- choke trunk arrangements.

USA

The USA has a structural separation between local carriers and long distance (transit) carriers. There is competition at both levels.

AT&T is the oldest and one of the largest transit operators and it provides call gapping at each of its points of presence where local carriers hand calls to AT&T. This enables AT&T to control traffic at the edge of its network.

Conclusion

Call gapping is the best technique available in other networks for handling surges in traffic demand, and is preferable to the provision of delayed dial tone alone. It is widely used within networks and is used between networks in some countries such as the UK and USA.

In circumstances where priority access is provided to emergency numbers then call gapping arrangements should also allow for delayed dial tone to be extended to priority calls whenever possible.

Hong Kong is slightly behind best practice in that call gapping is currently available only within the PCCW network. Some other fixed network operators in Hong Kong provided delayed dial tone during peak hours but this is not an adequate substitute for call gapping under the conditions that prevailed on 11 September.

6.3 Comparable events

We have tried to identify comparable events in other countries to see what lessons can be learned from them, but with little success.

There are frequent occurrences of traffic surges from mass calling events to individual numbers and most operators have developed methods for handling them.

There are occasional disasters which result in calls to a particular area and increased needs for calls involving various authorities. The New York World Trade Centre disaster is an example. Where preference schemes are available, they are commonly brought into force and limit the extent to which the public can use the network.

The main characteristics of September 11 in Hong Kong were:

- greatly increased but “unfocussed” demand for calls between members of the public
- demand spread across the whole network
- no particular increase in demand from the authorities, since no disaster materialised

Similar events are likely to occur only where:

- there are conditions that can generate widespread greatly increased demand, and the only events that are in this category seem to be warnings of impending disasters that necessitate changes to arrangements. Warnings of many possible disasters would generate large population moves but these moves would not necessarily be accompanied by widespread increased demand for telecommunications. It was the orderly nature of the population move and the unfortunate timing that generated the huge increase in demand. A panic situation would probably have generated lower demand than occurred in Hong Kong because people would be looking after themselves and not telephoning. Natural rather than terrorist disasters are more likely to affect larger areas. Thus similar events are unlikely in large parts of the world where the weather is less extreme and natural disasters are rare
- the population is well informed and well equipped with telecommunications. This is not the case in many of the areas that are most affected by weather extremes such as the Caribbean
- the population is large and dense so that there are many switches affected.

The nearest similar event that we have been able to identify occurred in Trinidad in 1990 as a result of a political coup. There was a very large increase in call demand and despite the application of call gapping and the existence of priority lines the network “collapsed”. Because of the continued high demand in some areas, certain switches were turned off and could not be brought back into normal service for some time. Trinidad differs from Hong Kong in that although it is also a small geographical area it has a much lower penetration of lines and at the time had a single monopoly operator.

Conclusion

The events in Hong Kong on September 11 were the result of a rare occurrence of events. Surges in demand for calls between the public across the whole network are very rare but a similar surge occurred in Trinidad in 1990 resulting in substantial and prolonged disruption to the network, which we understand was more severe than was experienced in Hong Kong.

7. Conclusions from the investigation of network performance

The following are our main conclusions from the investigation into the performance of the networks.

1. networks are reasonably dimensioned, taking account of international standards and practices, and many have 20-30% spare capacity for calls
2. the surge in demand existed in all directions and also occurred on SMS
3. the surge in demand for calls affected the whole of Hong Kong and was so large that substantial congestion could not be avoided. The surge started in the business district but later spread to the whole region
4. the switch logic is designed correctly to try to protect processors and handle excessive call attempts in a controlled manner, and most networks operated in accordance with their normal control procedures. Protection of the processors resulted in call attempts being suppressed such that the full transmission capacity available was not used in all circumstances (eg in some PCCW gateways)
5. network management is not as interactive as in UK and in some cases is dependent on ordinary PSTN calls that are themselves subject to congestion. Operators need to agree to exchange information and warnings of possible problems more frequently and openly under a confidentiality agreement that prohibits this information from being used for marketing purposes. Dedicated communications links need to be used for these communications
6. the most appropriate approach to the large rise in traffic demand is to limit the rate of call attempts at the switch nearest to the origination of the call attempts that is likely to be overloaded by the call attempts. This approach currently can only be applied manually. The automatic overload protection is valid but can result in the capacity to accept calls from other networks being reduced, which leads to more call attempts and aggravates the situation. PCCW's network suffered from the absence of gapping in other networks, especially the mobile networks. Correct management (call gapping) that restricts calls at the first switch that could be overloaded would avoid a situation where the automatic overload would be used. Since the typhoon announcements tend to produce very rapid rises in traffic demand, it is essential that the operators are warned in advance of public announcements so that they can apply manual call gapping
7. only PCCW has call gapping at present. Although precise information has not been obtained from all mobile operators it seems that none has a call gapping capability in its switches. Call gapping should be introduced in every network, including mobile networks, that is vulnerable to congestion or processor loading that requires protection measures as a result of high volumes of call attempts. Since call gapping will require significant investment, OFTA, in conjunction with

the operators should assess the risks, benefits and costs before introducing such a requirement.

8. several operators took unilateral measures to reduce call attempts such as disabling interconnection links. These measures were not coordinated and in some cases were probably counter productive as they may have stimulated yet more call attempts. Better exchange of information and coordination is needed

9. the available information suggests that, due to the nature of the traffic, the most significant congestion occurred in the PCCW network. The processors of some local switches had to take self protective action to the extent that some interconnection transmission capacity was under used. This occurred, despite operator A having dimensioned its network, including interconnect links, to meet a 0.5% GoS level. We consider that it would be beneficial if PCCW, in conjunction with other operators affected, reviewed the traffic on particular interconnect gateways to establish whether or not a better load balancing could be achieved

10. the inability of mobile operators to access information on ported fixed numbers resulted in mobile operators attempting unsuccessfully to pass calls to PCCW when they could have routed the calls direct to the recipient operator over less congested interconnections

11. it is unrealistic to expect the networks to be dimensioned to handle the traffic demand experienced on 11 September. Therefore, measures should be taken to:

    - educate customers to make more efficient use of the networks at times of high demand, including keeping calls short, avoiding making unnecessary calls and large numbers of repeat call attempts and using SMS messaging wherever possible

    - use alternative means as much as possible to disseminate information intended for groups of people and reduce so to reduce the peak demand.

8. The options to prevent similar problems re-occurring

There are three approaches to preventing similar occurrences taking place n the future:

- improving the warning system
- improving the networks
- improving the public's knowledge of how best to use the networks.

This section reviews various options under these different headings.

8.1 Improving the warning system

The purpose of improving the warning system is to provide an early alert to telecommunications operators to enable them to implement measures to cope with the potential upsurge in traffic, and thereby reduce the peak of demand for telecommunications.

The problems of 11 September were compounded for three reasons:

- the announcement of the "No. 8" Typhoon Warning Signal was relatively unexpected compared to other previous situations
- the network operators were not given advanced notice of the announcement . As a result, PCCW could not implement its manual call gapping procedures before switches commenced automatic overload protection. This resulted in the surge in demand not being handled as well as it could have been
- the announcement occurred at a time when the call demand was already at its normal busy hour peak.

Whilst an announcement may always generate large demand, steps should be taken to reduce the likely demand as far as possible and to enable the networks to make best use of their facilities for traffic management.

It would be desirable to avoid making warning announcements at times when they are likely to cause widespread disruption to the plans of the citizens. Instead, announcements should be given earlier, wherever possible, so that citizens can make contingency plans more easily.

The operators need at least 15 minutes advanced notice of a surge in traffic demand so that they can make necessary preparations. We recommend that OFTA should always be informed in advance of the issue of announcements to the public so that they can give the operators sufficient notice.

A significant part of the demand on 11 September came from parents trying to contact schools to find out about the local arrangements for their children. We recommend that the Education Department should examine other means for disseminating this information that avoids the need for parents to telephone the schools. Possibilities are:

- posting the information on web pages
- sending e-mails to exploder lists of parents
- sending SMS messages
- making more information available on the radio and television
- making information available via a dedicated call centre number when calls direct to schools are likely to be congested
- organising information trees to distributing information between parents (eg the school calls 5 parents and each of them call another 5).

None of these methods is a sufficient solution, but a combination of these methods should be effective in reducing demand and in improving the flow of information.

8.2 Improving the networks

The following possibilities have been identified:

- ***introduce call gapping in all switches nearest to the origination of the call attempts that are likely to be overloaded by the call attempts.*** This will require investment in software upgrades but would give the operators the ability to control call attempts at the point where they first may cause problems and so to make optimum use of the network capacity. Careful consideration should be given to deciding which switches should introduce call gapping. The main criterion should be the likelihood of receiving significant volumes of call attempts that would cause congestion or affect the performance of the switch. This requirement should be considered for mobile networks as well as fixed networks because mobile networks were seen to experience high volumes of call attempts and account for a substantial proportion of calls. *(Recommended)*
- ***improve the exchange of information between the network management centres of the operators****.* This is an essential complement to call gapping as it will enable a network that is experiencing problems with processor load to ask other networks to reduce the call attempts across the interconnection links. Because network quality is a sensitive commercial issue, we recommend that the operators should make such information confidential and agree to prohibit its use for marketing. The communications facilities used between the network management centres should be such as not to be affected by congestion in the PSTN. *(Recommended)*
- ***review the processor power of switches****.* The switch processors need to have sufficient power both to handle the traffic in progress, the call attempts and any call gapping algorithm in use. With the introduction of call gapping, the operators need to check that the processor power is still adequate and they may need to

add processor capacity for some switches. This is particularly relevant for PCCW gateway switches. *(Recommended)*

- ***increase the transmission capacity in the networks***. The transmission capacity appears to be generally satisfactory and since voice traffic levels are dropping in the fixed networks, it would be inappropriate to increase capacity just to handle rare increases in demand. *(Not recommended)*
- ***introduce a capability into the switches to limit call duration during congestion***. This would help to ration the use of the networks during high demand. It would be a very effective measure but the operators would have to procure additional software to provide this function and they would have to add announcements to explain to callers that the limit is in progress. Overall we are not convinced that the benefits outweigh the costs, especially if more is done to educate users and advise them to make only essential calls during congestion. *(Not recommended)*
- ***block calls that are likely to be non-essential.*** The problem is to identify non-essential calls. Dial-up Internet access is likely to be less essential or non-essential, and blocking these calls would release capacity in the local exchanges. It would not, however, release capacity at the interconnect gateways because these calls are handled on separate circuits. Blocking at the local exchange would be possible with call gapping provided that the processors could handle the algorithms. We think that this possibility is adequately covered by the recommendation for call gapping. *(No action needed)*

8.3 Improving the public's knowledge of how best to use the networks

The demand for telecommunications could be reduced and the problems of network management mitigated if users were advised how to make best use of the telecommunications networks in times of emergency and congestion. We recommend that OFTA and the operators seek to educate users to:

- avoid unnecessary calls
- keep calls short
- avoid rapid re-dialling
- use SMS and e-mail where possible

These messages should be broadcast after weather related announcements and warnings are given on the radio and television.

9. Conclusions and recommendations

11 September produced a most unfortunate combination of events. The typhoon announcement was given unexpectedly at the time during the changeover of the school shifts, which was likely to generate the maximum level of traffic demands with the traffic to some extent focused on calls to schools. Since the schools would have been able to accept only a small fraction of the calls, this created a huge increase in call attempts that could not be satisfied. The increase in call attempts caused automatic overload protection to operate in a number of switches. The overload protection and some manual measures taken by operators reduced the interconnection capacity so the volume of calls over some interconnects would have been lower than normal. This reduction increased the unsuccessful calls and aggravated the situation further.

The actions by the operators were uncoordinated and the operators did not inform each other about the state of their networks nor the measures that they were taking. Although such coordination is desirable, better coordination alone would not have reduced the problems significantly.

The best practice approach in these circumstances would be for the networks to limit the call attempt capacity at the switch nearest to the origination of the call attempts that is likely to be overloaded by the call attempts. This approach requires the operators to be warned in advance before the traffic starts to grow rapidly.

Even with better management, the networks would not have been able to satisfy all the public's demand for communications. A significant part of this demand relates to the need for parents to contact schools who have very limited capacity for handling large numbers of incoming calls. Consequently other methods need to be used as much as possible for the dissemination of information to parents to reduce the traffic demand.

Hong Kong is unusual compared to other parts of the world in that the region is small and very densely populated and also subject to reasonably frequent severe weather events, and these events affect the whole territory. Larger territories would have more scope for using capacity from areas where traffic demand is not increased.

The fact that there are several different interconnected networks, as a result of liberalisation, did not make the problem materially worse. The loss of capacity that occurred on the interconnection links would probably have occurred on links within a single monopoly network if the switches applied automatic overload protection as a result of call attempts not being adequately limited at the network edges. The management of the situation, however, might have been marginally better.

From the information provided by network operators we consider that there are no fundamental shortcomings in the design of the networks in Hong Kong, which are designed in accordance with international best practice. It is our assessment that no comparable networks anywhere in the world would have been able to handle the level of call attempts experienced in such a short period on 11 September 2002.

Our main recommendations to reduce the effects of similar events in the future are:

1. OFTA should be informed well in advance of the issue of announcements to the public concerning No. 8 Typhoon Warning Signals. In this way, OFTA will be able to inform the operators who can then apply methods to restrict call attempts at the network edges, thus preventing switches going into automatic overload conditions and reducing trunk capacity. A 15 minutes pre-warning would allow operators sufficient time to implement the necessary measures
2. OFTA, in conjunction with the operators, should undertake a long term program to educate the public about how the telephone network will behave in the event of excessive demand and how best to use the network, eg limit the number of calls to those that are essential, keep calls short, avoid making frequent call attempts and to use alternative form of communication such as SMS wherever possible
3. all networks that are vulnerable to events such as occurred on 11 September should have the capability to reduce the level of call attempts to specific numbers or number ranges. This means the introduction of call gapping for all networks that are likely to experience high volumes of call attempts that would result in congestion or processor protection measures. Before introducing this requirement, OFTA should discuss with the operators the cost involved and assess the overall risks, benefits and costs
4. the network operators should develop closer coordination between their network management centres under suitable confidentiality agreements. Such communications should use facilities such as a private network that would not be affected by congestion on the PSTN
5. schools should consider establishing a range of methods for distributing information to parents in times when normal schooling is disrupted due to adverse weather conditions. This could include group e-mailings and SMS messages to parents and carers, and the posting of information to school web sites. It could also include the provision of information via a dedicated call centre number when lines to schools are likely to be congested.