CONEX
Internet-Draft
Intended status: Informational
Expires: April 22, 2010

H. Tschofenig
Nokia Siemens Networks
A. Cooper
Center for Democracy & Technology
October 19, 2009

Congestion Exposure Problem Statement
draft-tschofenig-conex-ps-00.txt

Status of this Memo

Copyright Notice

Abstract

The availability of broadband connections together with flat-rate pricing has made new types of peer-to-peer applications possible. From an Internet evolution and end user value point of view this is very exciting. As a consequence, an increase of user-to-user traffic was observable all around the world over the last few years. With the usage of p2p systems the observation can be made that a certain group of users, called high-consuming users, decided to use their flat-rate contract excessively. This in turn seems to have caused network operators to take actions.

This document illustrates a couple of techniques used by operators today to deal with excessive bandwidth usage. More information can improve the decision making process.

Table of Contents

1. Introduction 3
2. State-of-the-Art Building Blocks 5
2.1. Means of Identifying the Causes of Congestion 5
2.2. Potential Actions Operators might take in Response 6
3. New Activities 8
4. Summary 9
5. Security Considerations 10
6. IANA Considerations 11
7. Acknowledgments 12
8. References 13
8.1. Normative References 13
8.2. Informative References 13
Appendix A. Example Policy Statement 15
A.1. Fair Usage Policy 15
A.1.1. What is the Fair Usage Policy? 15
A.1.2. How do I know I'm a very heavy user? 15
A.1.3. I have Contract Option 3, does the Fair Usage Policy apply to me? 15
A.1.4. Peer to Peer (P2P) 15
Authors' Addresses 17

1. Introduction

In 2006 K. Cho et al. [traffic] published a paper about the growth of residential user-to-user traffic in Japan that indicates '... a small number of users dictate the overall behavior; 4 % of high-consuming users account for 75 % of the inbound volume, and the fiber users account for 86 % of the inbound volume.'. The same paper also indicates a substantial increase in traffic growth, namely 37 % per year according, and not just a different distribution of traffic among the users. At that time 63 % of the residential traffic volume is contributed by user-to-user traffic.

These numbers itself do not represent a problem as by themself and do not necessarily lead to congestion. However, some operators very likely had different expectations about the growth rates and traffic consumption of individual users and statistics (used for their pricing models) did not work out too well for them. The profit margins for Internet access are quite slim due to fierce competition. This puts a lot of pressure on operators to deal with these high-consuming users who cost them a lot of money. Finally, some broadband networks may not have the ideal characteristics (such as the topology for routing traffic) for user-to-user traffic (e.g., Cable Networks).

> Congestion is often mentioned in this context and as stated in RFC 5594 [RFC5594] "... congestion can be viewed merely as a manifestation of cost. An ISP that invests in capacity could be considered to be paying to relieve congestion. Or, if subscribers are charged for congesting the network, then cost and congestion could be viewed as one and the same. The distinction between them may thus be artificial.".

To summarize in a simplistic way, those who produce a lot of traffic cost a lot.

Operators are now facing a range of options, see sections below, that

can be taken and there is a tradeoff between what is allowed (legally and from a public relation point of view) and what is useful from a performance point of view. The latter aspect can be seen from the point of view of a device performance (as many of the mechanisms actually slow down the forwarding performance quite a bit) and consequently a cost challenge.

The existence of flat rate pricing contributes to some of the problems since the bandwidth usage in total needs to be covered by the money obtained from broadband customers but the usage of individual users is not reflected in the amount. As such, users that rarely utilize the network pay the same amount as someone who uses

P2p filesharing excessively.

However, from a psychological point of view humans tend to strive to avoid uncertainty and hence offerings that reduced uncertainty. For a user there are essentially two aspects to worry about

- Uncertainty in the bill: unpredicable costs that make planning difficult.
- Uncertainty in the performance: performance degradation as part of the actions being taken

True flat rate pricing avoids uncertainty in the bill. Unfortunately, most of the solutions described below lead to some uncertainty and thereby increase unhappyness of customers.

2. State-of-the-Art Building Blocks

2.1. Means of Identifying the Causes of Congestion

RFC 2975 [RFC2975] describes accounting as "The collection of resource consumption data for the purposes of capacity and trend analysis, cost allocation, auditing, and billing.".

Over the years the number of information elements that can be sent from an accounting client to an accounting server using standardized protocols, such as RADIUS (see [RFC2866] and [RFC2865]) and Diameter [RFC3588], has increased. The fact that standardized protocols have been available allowed different AAA networks to be interconnected and their usage can be found in almost every enterprise and operator network. The initial accounting mechanisms envisioned a rather non-real time nature in reporting resource consumption but with mechanisms like like Diameter Credit Control [RFC4006] allowed real-time credit control checks.

It has to be noted that RADIUS and Diameter are not the only protocols that can be used to collect usage information and to trigger certain actions, even they are fairly popular. Other

approaches, as documented in [I-D.livingood-woundy-congestion-mgmt], lead to similar results.

Deep packet inspection refers to inspecting traffic that passes through the operators networks up to the application layer. Depending on the configuration of the device traffic shaping, packet dropping/blocking and other usages might be applied. For example, content sharing p2p applications maintain many simultaneous TCP connections with other nodes for the purpose of simultaneous downloads. An operator may, for example, limit the number of connection setups from a single subscriber. Certain end user contracts may also allow operators to ban servers from residential access.

Determining the type of application that a subscriber is running was seen as necessary to throttle only certain applications, instead of impacting the full range of traffic a subscriber is using. A side-effect is the additional investment for the device and operational costs. The process of inspecting traffic is performance intensive and continous software updates are necessary to ensure that the detection engine recognizes the latest protocol variants. Additionally, the attempt to selectively deal with applications (even though these applications might be the reason for the high traffic volume) has not been received well by the users.

2.2. Potential Actions Operators might take in Response

What actions are taken based on the collected information and in what time frame is largely left to the choice of those who run the infrastructure. In the context of this discussion the collected information may be used to charge the user per volume, per time and in various different combinations. Additionally, the RADIUS and Diameter allow the server side with a server-initiated message (see Change of Authorization in [RFC3576], and the functionality provided in the Diameter Base specification [RFC3588]) to push decisions to the AAA clients, typically edge nodes, acting as enforcement nodes. These decisions include actions like shaping or packet marking.

Shaping: End user contracts often offer a combination of 'flat-rate' scheme whereby a fixed price tariff is used up to a certain usage

volume (typically quite high for regular usage). Subsequently, if the consumption goes beyond a certain threshold then the entire traffic is given lower priority and potentially shaped.

In many countries operators have to offer a clear description of the service they offer and since the term 'flat-rate' is already associated with a certain meaning the term 'Unlimited Data Rate' is often used for this type of service. Contracts typically contain statements that allow certain actions to be taken. An example of such a fair use statement can be found in Appendix A.

Note that traffic shaping is often only applied to high-consuming users (since they are known based on the accounting procedures) or the effect becomes only visible during peak hours when the network fills up.

Class-Based Assignment: In this technique users are classified into a set of classes depending on their past behavior. Subsequently, the traffic is treated according to the associated class. It is ensured that the traffic of lightweight users, users that really rarely use their Internet connection, cannot be pushed away by heavy users. This mechanism again requires some form of DiffServ marking to be in place.

Charging for Excessive Traffic: As a possible action a user might get charged differently for traffic that exeeds a certain threshold compared to the traffic that falls within the agreed limits.

Discontinuing Contracts: In some rare cases ISP also decided to cut connectivity under certain condition. In fact this might be justified in certain cases. For example, in case of new botnets malware distribution when the operator recognizes an infected machine and hotlines the entire traffic of that particular machine to a separate network and, like in WLAN hotspots, HTTP traffic is intercepted to display further information. In some cases the same technique has been applied with excessive usage of P2P

traffic, either intentionally or due to a false alarm caused by a statistical traffic analysis technique.

In France the HADOPI law adopted in parliament that allowed an 'independent authority' to punish copyright violators with a temporary suspension of their Internet access has raised discussions within Europe about the fundamental right to 'communicate and express' and its applicability to the Internet access. Although this discussion is still ongoing the French Supreme Court had striked down portions of the law arguing that any restriction of such a right can only be decided by a judge.

3. New Activities

In response to the P2P infrastructure workshop in 2008 (with a summary documented [RFC5594]) two working groups and one research group has been created that focus on a certain area of the application space:

LEDBAT (Low Extra Delay Background Transport) [ledbat] is designed to allow to keep the latency across the congested bottleneck low even as it is saturated. This allows applications that send large amounts of data, particularly upstream on home connections, such as peer-to-peer application, to operate without destroying the user experience in interactive applications.

LEDBAT is a promising approach when applied widely in P2P clients. This solution has been focused P2P applications, and its applicability to other applications, such as video using H.264, is unclear.

ALTO (Application-Layer Traffic Optimization) [alto] aims to design and specify mechanisms that will provide applications, typically P2P applications, with information to perform better-than-random initial peer selection to increase their performance and at the same time to avoid excessive cross-domain traffic that tends to be more expensive for the operator. For legal content ALTO mechanisms with the ability for ISPs to deploy proxies appear to be a viable solution. However, a lot of the content being distributed in P2P filesharing networks today is not legal and caching such content by operators could turn out problematic for them.

Peer to Peer Research Group [p2prg] aims to provide a discussion forum for resarchers related to all sorts of challenges of P2P systems in general, such as P2P streaming, interconnecting distinct P2P application overlays, security and privacy, etc. [I-D.irtf-p2prg-mythbustering] provides a number of literature references to support some of the claimed benefits of ALTO solutions mechanisms, such as the expected decrease in cross-domain traffic.

4. Summary

Heavy users are a reality. Operators that would like to counteract the impact of heavy users on their networks have a fair number of tools at their disposal. These tools may allow operators to identify heavy users, collect performance and usage indications, and choose from a variety of mitigating steps depending on the operator's preferred business practices. Subscriber-specific information, including policies, resource consumption information, and details about the current network attachment point, may be available in accounting servers. Information about the network topology and the state of particular topology elements may be available in the network management infrastructure. Solution approaches similar to [I-D.livingood-woundy-congestion-mgmt] have demonstrated one way of taking congestion information into consideration.

The currently available mechanisms for identifying and mitigating congestion largely run wholly within an operator's network and without a lot of information exchange about congestion information to or from end hosts or other network operators. Exposing this information may allow end devices to make more informed decisions (although policy enforcement would still be required by the operator).

The collection of congestion information poses the challenge of deciding where in the network to put the metering agents to ensure that enough information is collected at the right point in time. Distributed collection and the correlation of the information across different nodes is a complex task. An approach that collects this congestion information along the path of the data packet (via inband signaling) would simplify this task. Regardless of the technical solution utilized for collecting information, certain users will undoubtedly observe the effects of decisions that operators make about how to handle congestion. Allowing users to understand these decisions will be crucial and having a channel to send feedback to the end device and/or subscriber would be a helpful step towards increased transparency.

5. Security Considerations

This document highlights approaches for dealing with heavy network usage and all of them raise security and privacy concerns. This document does, however, not introduce new mechanism and hence the reader is referred to the description of the respective mechanism.

6. IANA Considerations

This document does not require actions by IANA.

7. Acknowledgments

The authors would like to thank Alan DeKok, Jens-Peter Haack, Jouni Korhonen, Tommy Lindgren, Lars Eggert, for their time to discuss the topic. Additionally, we would like to thank Marcin Matuszewski for his help with the P2P infrastructure workshop paper (as it was used as a starting point).

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

8.2. Informative References

[I-D.irtf-p2prg-mythbustering]
Marocco, E., Fusco, A., Rimac, I., and V. Gurbani, "Improving Peer Selection in Peer-to-peer Applications: Myths vs. Reality", draft-irtf-p2prg-mythbustering-00 (work in progress), August 2009.

[I-D.livingood-woundy-congestion-mgmt]
Bastian, C., Klieber, T., Livingood, J., Mills, J., and R. Woundy, "Comcast's Protocol-Agnostic Congestion Management

System", draft-livingood-woundy-congestion-mgmt-01 (work in progress), September 2009.

[RFC2865] Rigney, C., Willens, S., Rubens, A., and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

[RFC2866] Rigney, C., "RADIUS Accounting", RFC 2866, June 2000.

[RFC2975] Aboba, B., Arkko, J., and D. Harrington, "Introduction to Accounting Management", RFC 2975, October 2000.

[RFC3576] Chiba, M., Dommety, G., Eklund, M., Mitton, D., and B. Aboba, "Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)", RFC 3576, July 2003.

[RFC3588] Calhoun, P., Loughney, J., Guttman, E., Zorn, G., and J. Arkko, "Diameter Base Protocol", RFC 3588, September 2003.

[RFC4006] Hakala, H., Mattila, L., Koskinen, J-P., Stura, M., and J. Loughney, "Diameter Credit-Control Application", RFC 4006, August 2005.

[RFC5594] Peterson, J. and A. Cooper, "Report from the IETF Workshop on Peer-to-Peer (P2P) Infrastructure, May 28, 2008", RFC 5594, July 2009.

[alto] "", <http://www.ietf.org/dyn/wg/charter/alto-charter.html>.

[ledbat] "", <http://www.ietf.org/dyn/wg/charter/ledbat-charter.html>.

[p2prg] "", <http://www.irtf.org/charter?gtype=rg&group=p2prg>.

[traffic] Cho, K., Fukuda, K., Kato, H., and A. Kato, "The impact and implications of the growth in residential user-to-user traffic", SIGCOMM Comput. Commun. Rev. 36, 2006.

Appendix A. Example Policy Statement

A.1. Fair Usage Policy

A.1.1. What is the Fair Usage Policy?

The Fair Usage Policy is designed to ensure that the service received by the vast majority of our customers is not negatively impacted because of extremely heavy usage by a very small minority of customers. This is why ISP X continuously monitors network performance and may restrict the speed available to very heavy users during peak time. This applies to customers on all Options. Note if you are a heavy user we will only restrict your speed, service will not be stopped so ability to upload and download remains. No restrictions will be imposed outside of the peak times. Only a very small minority of customers will ever be affected by this (less than 1 %).

A.1.2. How do I know I'm a very heavy user?

There is no hard and fast usage limit that determines if you are a heavy user as the parameters that determine heavy use vary with the demands placed on the network at that given time. If you have a query about fair usage related restrictions on your line please call us.

A.1.3. I have Contract Option 3, does the Fair Usage Policy apply to me?

Yes, the Fair Usage Policy applies to all customers on all Options, including Option 3. Option 3 allows unlimited downloads and uploads inclusive of the monthly rental price, so you will not be charged for over-use, however this does not preclude ISP X from restricting your speed at peak times if you are a heavy user. If you are an Option 3 heavy user this does not prevent you from continuing to use your service, nor does it cost you any more but it ensures that you do not negatively impact the majority of our customers who share the available bandwidth with you.

A.1.4. Peer to Peer (P2P)

A.1.4.1. I'm noticing slower P2P speeds at peak times even though I'm not a very heavy user, why is this?

P2P is the sharing and delivery of files amongst groups of people who are logged on to a file sharing network. P2P consumes a significant and highly disproportionate amount of bandwidth when in use even by small numbers of users.

This is why we have a peak time policy where we limit P2P speeds to manage the amount of bandwidth that is used by this application in particular.

Without these limits all our customers using their broadband service at peak times would suffer, regardless of whether they are using P2P or not. It's important to remember that P2P isn't a time-critical application so if you do need to download large files we advise you to do this at off-peak times when no restrictions are placed, not only will you be able to download faster but your usage will not negatively impact other users.

A.1.4.2. Does this mean I can't use Peer-to-Peer (P2P) applications?

No, we are not stopping you from using any P2P service, P2P will just be slowed down at peak times. Again, P2P is not generally a time-sensitive application.

Authors' Addresses

Hannes Tschofenig
Nokia Siemens Networks
Linnoitustie 6
Espoo FIN-02600
Finland

Phone: +358 (50) 4871445
Email: Hannes.Tschofenig@gmx.net
URI: http://www.tschofenig.priv.at

Alissa Cooper
Center for Democracy & Technology
1634 I Street NW, Suite 1100
Washington, DC
USA

Email: acooper@cdt.org