MINIMIZATION OF INTERNET SERVICE PROVIDERS' INABILITY TO CONTROL PEER-TO-PEER TRAFFIC BY USING PEER-TO-PEER AND SOFTWARE DEFINED NETWORK BASED WEB SEARCH ENGINE

CHANG CHOONG CHING

OPEN UNIVERSITY MALAYSIA
2014

MINIMIZATION OF INTERNET SERVICE PROVIDERS' INABILITY TO CONTROL PEER-TO-PEER TRAFFIC BY USING PEER-TO-PEER AND SOFTWARE DEFINED NETWORK BASED WEB SEARCH ENGINE

CHANG CHOONG CHING

A Master's Project submitted in partial fulfilment of the requirements for the degree of Master of Information Technology

Centre for Graduate Studies
Open University Malaysia

2014

DECLARATION

Name: CHANG CHOONG CHING

Matric Number: CGS 00705203

I hereby declare that this Master's Project is the result of my own work, except for quotations and summaries which have been duly acknowledged.

Signature:

Date:

15th December 2014

MINIMIZATION OF INTERNET SERVICE PROVIDERS' INABILITY TO CONTROL PEER-TO-PEER TRAFFIC BY USING PEER-TO-PEER AND SOFTWARE DEFINED NETWORK BASED WEB SEARCH ENGINE

CHANG CHOONG CHING

December 2014

ABSTRACT

The Software Defined Network works together with the modified search engine as a single system so that software routers located throughout the autonomous network will guide the network traffic in searching and downloading of content and avoid network congestion. The network congestion could occur since the path of the P2P system is unpredictable and it uses its own methodology in choosing a path. The search engine acts as a central directory so that updated content could be immediately broadcasted to all interested parties as the P2P engine will use its own algorithm to propagate information to peers on its own pace and depending on which peer joined the network (which is unpredictable). It is also used as the initial point of initiating a search for content and locating the said content directly. The P2P engine within the search engine itself is also programmable as to how many hops that the propagation will go and therefore this eliminates unnecessary propagation that is likely to cross over multiple boundaries and thus prevent the incurrence of large inter-boundary network traffic costs.

(178 words)

Keywords:
Peer-to-Peer,
Inter-AS traversals,
Transiting networks boundaries,
Networking,
Telecommunications
(Not more than 5 words/phrases)

PENGURANGAN KEBUNTUAN PARA PENYEDIA PERKHIDMATAN INTERNET UNTUK MENGAWAL TRAFIK PEER-TO-PEER MELALUI ENGIN PENCARIAN WEB BERLANDASKAN PEER-TO-PEER DAN SOFTWARE DEFINED NETWORKING

CHANG CHOONG CHING

Disember 2014

ABSTRAK

Software Defined Network bersama engin pencarian web, berfungsi sebagai satu system, agar routers berbentuk perisian di dalam rangkaian jaringan dapat mengawal trafik yang mencari dan memuat turun kandungan tanpa mengalami kesesakan. Kesesakan di dalam rangkaian jaringan kerap berlaku hasil dari halatuju perjalanan P2P yang tidak menentu dan kepenggunaan kaedah tersendiri untuk menentu halatuju tersebut. Engin pencarian web berfungsi sebagai pusat rujukan dan mengakses kandungan agar kandungan terkini dapat disebarkan kepada pihak yang berminat serta-merta manakala P2P engin yang terkandung didalamnya akan menyebarkan maklumat dengan kaedahnya tersendiri dan juga bergantung kepada komputer yang menjadi ahli kepada sistem P2P pada ketika itu. Sistem P2P ini juga boleh mengawal bilangan lompatan yang dibenarkan untuk mendapatkan maklumat dan ini akan mengurangkan kebarangkalian lompatan yang terkeluar dari sempadan rangkaian jaringan penyedia perkhidmatan internet yang berkaitan dan mengakibatkan kos yang tinggi.

Kata Kunci:
Peer-to-Peer,
Melalui sistem-sistem autonomous,
Melangkaui sempadan jaringan,
Jaringan,
Telekommunikasi
(Tidak melebihi 5 perkataan/rangkai kata)

ACKNOWLEDGEMENTS

I wish to express great thanks and appreciation to Dr. Jalil Md. Desa, Ms. Jaspal Kaur, my other lecturers and the staff in Open University Malaysia for their dedication and commitment in getting the best out of each of their student and in believing in me.

I also with to thank my mother, Wong Sin Kwai and my sister, Chang Choong Yeen who had supported my efforts in obtaining my Masters Degree in Information Technology and to my late father, Chong Chee Wai, who believes that continuous education is a way towards improving oneself.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF ABBREVATIONS

Initials	*Full Term Represented*
AS	Autonomous System
IP	Internet Protocol
ISP	Internet Service Provider
MPLS	Multi-Protocol Label Switch
P2P	Peer-to-Peer

CHAPTER 1

INTRODUCTION

1.1 Background to the Study

The background facts detail the mechanics of the Peer-to-Peer systems and how its internal workings have created tensions between them and the Internet Service Providers.

1.1.1 Peer-to-Peer (P2P) Systems

A peer-to-peer ("P2P") network is whereby each computer in the network acts as a client to use the resources located in other computers in the network but the said computer also acts as a server to provide access to locally located resources (such as files in its file system and storage) to other computers on the network.

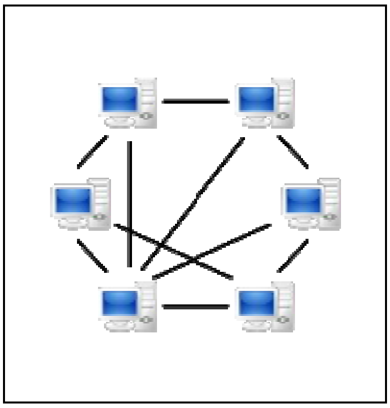

Figure 1.1: Network diagram of a Peer-to-Peer ("P2P") network

Recent developments have indicated that P2P applications are now using decentralized systems with unstructured networks without a centralized system directory (keeping track of peers/nodes and the resources that they hold) to avoid legal implications especially when resources being shared whose copyrights do not belong to these resource holders.

BitTorrent is an example of this new breed of P2P applications. Decentralized systems with unstructured networks would use Distributed Hash Tables ("DHT") to store routing and resource location information. Each peer/node in the network and each shared resource/file within the DHT are mapped to an address space ring, approximately the size of 2^{160}. The peers are placed on the address space ring via a node ID that is a mathematical hash of the Internet Protocol address of the node. The object to be shared is mapped to the address space via a key (which is a hash of the shared object's name or a reference to the shared object) (FOR001).

The direct method of storing an object in the DHT is by keeping an object on a node whose node ID is closest to the hashed key of the object. E.g. 2 peers with each node having ID of 34 and 46 and the key of the object is 38, the object will be stored at the node ID 34 as it is the nearest. Nearest is subject to interpretation of each peer application's protocols (FOR001). The indirect method of storing an object in the DHT is by allowing an object's owner to store the object but a reference of the object is stored in the node whose node ID is closest to the key of the object. E.g. there are 3 peers on the network with node ID of 15, 20 and 25 with node ID 25 being the owner of the file (and where the physical file is kept). The key of the object is 19 and therefore a copy of the reference of the file object is stored on node ID 20 as it is nearest to the key. (FOR001)

The BitTorrent protocol uses a set of policies to reinvent the overlay topology based on the available network performance as new peers join the network. To avoid overloading the peers and paralyzing the system, each node is limited to four concurrent connections to its neighbours. A peer flags its neighbours (peers in the network) as choked (the list of peers that it is currently not connecting to but may connect to in the future), unchoked (the list of peers that it is currently connecting to and performs uploading and downloading of files with them), interested groups (peers that hold items of interest) and uninterested groups (peers that do not hold items of interest to this peer) (FOR001). Every 10 seconds, a peer will try out a peer from the choked but interested group to determine whether it offers better bandwidth and performance. If there is a more efficient peer, the efficient peer is included in the unchoked group and a less efficient peer from the unchoked group is swapped into the choked group in place of the efficient peer. In this way, peers will be able to find other peers with better performance and better data rates (FOR001). To enable information to propagate to new peers without anything interesting to share, a peer will periodically place a peer from the choked group into the unchoked group regardless of its uploading rates. The BitTorrent protocol also tries to create circulation of the latest information by using the rarest first strategy in which the information with the fewest copies among the neighbours is downloaded first (FOR001).

1.1.2 Tensions between ISPs and P2P systems

The Palo Alto report has stated that bandwidth consumption of P2P file-sharing (particularly BitTorrent) had frequency of use of 63% to 68% globally and composed of 3% to 6% of total bandwidth consumed globally despite worldwide ISP's attempts to control it (PAL001). The report also said that “it is like a weed that continues to return, despite repeated control efforts. It raises the question of whether or not it can ever be controlled" (PAL001). The ISP's attempts to control it by targeting P2P traffic bandwidth consumption had received adverse response from its customers and this incident triggered an investigation by the Federal Communications Commission in United States (COM001)(NYT001)(EFF001). Furthermore, P2P applications have a tendency of using non-standard designated ports specific to their protocol which creates difficulty in traffic engineering via deep packet inspection, in analyzing the packets and then in introducing routing policies to govern them. Although, the more popular P2P applications' (that make up the bulk of the bandwidth consumption are just a few) ports are known, using this method of control has caused end consumers to take their business elsewhere and therefore not a popular solution (SHE001). As there may be more P2P applications emerging, it would also be administratively cumbersome to always have to update the ports to cater for these emerging popular applications and hence not a feasible solution.

The tensions prevailing between Malaysian ISPs, namely such as Maxis, Digi, Telekom Malaysia, Celcom, Jaring, Time Telekom and others are no different. Each ISP have specifically mentioned the use of bandwidth would be subjected to constrain within their individual Fair Usage policies governing the subscription of the ISPs' services as noted on their corporate websites. Complaints have been made to the Malaysia Communications and Multimedia Commission ("MCMC") that Deep Packet Inspection techniques were used to analyze and neutralized network packets by Telekom Malaysia to control Internet content sharing, as reported by the Malaysia Chronicles on 20th May 2013. The MCMC itself has requested for a ban on traffic to 10 P2P websites as reported by the Star newspaper on 11th June 2011 to prevent the sharing of copyrighted content.

The ISPs had until to-date used Multi-Protocol Label Switched Network ("MPLS") for traffic engineering ("MPLS-TE") such that they are able to control traffic by creating a tunnel in which packets having a pre-defined label are able to travel through whilst other packets would follow the normal open shortest path first algorithm (CAR001). MPLS is part of the Network Layer protocol in the TCP/IP stack.

However, the P2P protocol is an Application Layer protocol in the TCP/IP stack and its algorithm is based on a Distributed Hash Table ("DHT") which stores the directory of computer nodes ("peers") that are part of the P2P and pointers to peers holding the files to be shared. The P2P uses a network overlay approach in which the DHT is superimposed over the available infrastructure of the ISPs. The P2P uses its own algorithm in testing the efficiency of each connection to link up to peers of same connection speed and it constantly tests these connections and changes its linkage to other peers if faster connection speeds are detected. (FOR001)

The ISPs would notice that traffic would be heavy in certain links and will use MPLS-TE to redirect flow of traffic to underutilized network links to load balance network traffic. The P2P protocol would notice that connection speeds in certain links have changed and will change its connection patterns in response to these faster links. Again, the previously faster traffic (the tunnel created by the ISP) link will now succumb to heavy traffic load from the P2P as a result of P2P's network overlay algorithm. This causes a pendulum effect in the traffic management, making it uncontrollable. (AGG001)

The P2P algorithm also deals with pointers to resources and propagating of resources/shared files in response to connection speeds and works in an arbitrary manner independent of Internet routing and topology considerations (as described above). As a result of this algorithm, connections pass through international lines multiple times in reaching the requested content (AGG001). A node in Kuala Lumpur will end up requesting for content from a node in London when a node in Petaling Jaya holds the same content. The local ISPs are then charged for the use of these trans-international network backbones. In some cases, if the trans-international network backbone is subjected to joint maintenance by 2 ISPs, they will incur additional capital expenditure to upgrade the network infrastructure to cope with this in-transit traffic. Adding to the woes of the ISPs is that most bottlenecks are at the access networks (where users' homes connect into the ISPs' switches at the gateway) and at the links between ISPs (which are subjected to cost sharing arrangements with carriers) and not in the backbone infrastructure of the ISPs themselves (AGG001). This means that efforts at infrastructure improvement to accommodate the traffic are difficult and complicated as more external parties are involved for discussion as compared to the ISPs own infrastructure.

Therefore, a gap exist such that a local ISP should do as much possible to isolate network traffic such that available content are fetched once and stored within its own P2P such that outgoing traffic is reduced. This also allows the ISP to orchestrate its MPLS-TE and its own P2P's DHT into a single seamlessly integrated network engine. A local ISP's search engine which keeps track of popular contents and re-designate contents to nodes for storage (among its peers) whilst allowing it to control traffic flows via MPLS-TE, would offer local users more robust surfing experience whilst minimizing cost of outbound connection and chaotic traffic flows.

1.2 Problem Statement

Although many studies have proposed solutions to resolving the tensions, none of them have been able to offer a solution that could claim to be able to effectively reduce the tensions. The reason as to this occurring is detailed below.

1.2.1 Issues in cooperation due to political and legal reasons

Despite the efforts of the researchers in the creation of frameworks to resolve the tension between the ISP and P2P systems, cooperation is an essential part of these frameworks. The cooperation needed is in turn subjected to many political factors and challenges which need to be resolved in order for the cooperation to work. The challenges involved are (DAN001):

- Non-cooperative behaviour where one party (either the ISP or the P2P system) may try to exploit the information provided by the other party without providing any valuable/real information requested by the other party.
- Incentive mechanisms need to be created such that the ISP could encourage P2P systems to maintain any cooperation established and similarly P2P systems need to offer incentives to ISPs to obtain the ISPs' support.
- Information exchange mechanisms need to be created to support agreed-upon semantics and scalable protocols for purposes of authentication, integrity and enable protection from denial of service attacks.
- Implication into illegal content as ISPs would be seen as aiding the illegal distribution of copyrighted content through its caching.
- Capital outlay for the purposes of creating infrastructure and incurring operating expenses for maintenance programmes to implement these frameworks and would prove to be beneficial only if proximity aware operations result in larger cost reductions than these costs.
- Application requirements of the P2P systems must be made known to the ISPs for QoS purposes but yet needs to be accommodative enough to enable future applications to be catered for.

- Maintaining principles of network neutrality whereby each packet in the network is routed and forwarded impartially regardless of its contents.

If network neutrality is defined as ISPs should not be able to favour or punish any specific types of traffic, some of the frameworks proposed by these researchers would violate these principles as they would provide quality improvements for certain types of traffic and could even be used to put the P2P systems at the mercy of the ISPs. This still applies even if network neutrality is defined as classes of traffic without specific reference to types of traffic or protocols within a class.

Network topology challenges are created (in terms of customer's physical location) when customers could be located physically nearby within the same infrastructure but fall under the jurisdiction of different ISPs for internet access.

Proximity can be difficult to discover even if the ISPs cooperated. Modern telecommunication structures are complicated due to the division of the roles of network owners and network service providers and the manner in which jurisdiction could be separated to many parties (a connection from the last mile up to the core network infrastructure could be serviced by several different parties). Without proper mapping out of the topology, the relevant parties' jurisdiction and ISPs' IP address assignment policies, a discovery or proximity search in the first, third and fourth class of framework is likely to fail. Neighbouring ISPs may also have multiple peering points or bilateral peer agreements with each other. This would complicate situations when ISPs try to maintain proximity based peer selection schemes to favour a peer within the same ISP when a physically nearby peer from a different peering ISP is in fact nearer and would better optimize network utilization.

1.2.1.1 Issues in agreeing to type of information to be disclosed

Furthermore, information exchange between the ISP and the P2P systems are jeopardised by the fact that many of these information are highly sensitive and present critical success factors to the survival of the ISP or P2P system. Therefore, both parties put their business models and survival at risk by divulging these pieces of information.

Examples of sensitive information which ISPs need to provide in order to obtain the P2P overlay provider's cooperation:

Network topology and state:

- Such as Network capabilities, Optimal low latency routes, Optimal high-throughput routes and Network performance

ISP policies:

- Such as Routing policies, Preferences on port ranges and protocols and Network policies

Examples of sensitive information which P2Ps need to provide in order to obtain ISPs' cooperation:

- Such as Location of the peers in the overlay
- Popularity of content.

A reveal of the above information from the ISP may be damaging commercially as its competitors could react to, by offering better products and services. Hackers could also use the information revealed to undermine the infrastructure. Similarly, information from the P2P in revealing the location of the peers may cause users' content interest to be profiled, hackers to pinpoint vulnerabilities in the overlay networks or the users' content interest profile may be sold by the underlay providers to direct marketing companies.

As a result of this, it would appear that having a model that integrates both the elements of ISPs and P2P overlay network provider and traffic engineering appears to be more feasible - namely a modified web search engine integrated with P2P engine for overlay network management and traffic engineering for underlay network management.

1.2.2 Unwillingness of ISPs for capital investment amidst dropping ARPU

The complications in the handling of the P2P traffic are further complicated by the fact that Average Revenue Per User ("ARPU") of Internet Service Providers ("ISPs") worldwide have been on a downward trend (GSM001).

Table 1.1: Average annual ARPU growth from 2001 to 2011 (extracted from bar chart as published by GSMA Wireless Intelligence)

Region:	Average Revenue Per User Percentage:	Average Revenue Per Connection Percentage
Africa	-5.5%	-10.0%
Americas	-0.5%	-4.5%
Asia	-8.0%	-13.0%
Europe	-2.9%	-5.9%
Oceania	1.8%	-0.2%

With declining average revenues per user, ISPs are hesitant to invest in infrastructure and instead look towards enhancing efficiency instead. This conclusion is supported by a study made by the Insight Research Corp titled "Telecommunications and Capital Investments: Impacts of the Financial Crisis on Worldwide Telecommunications, 2012-2017" (CAB001).

At the same time, user demand for data traffic is on the increase as indicated below. (KPM001)

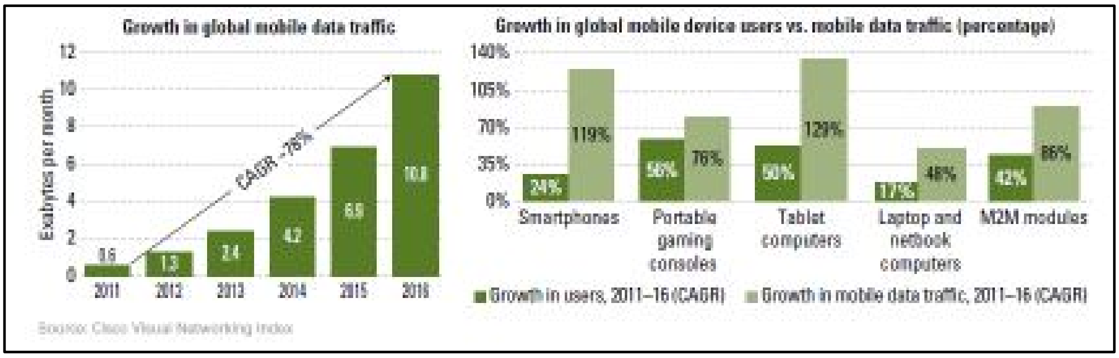

Figure 1.2 Growth in global mobile data traffic and growth in global mobile device users against mobile data traffic (KPM001)

Therefore, there is a desperate need to create efficiencies in the network to cope with the growing data demand whilst at the same time resist investments into capital infrastructure to the latest moments possible.

Compounding this situation is the fact that no ISPs work in isolation in the connectedness of the Internet. Connections between ISPs are subject to peering agreements and transit agreements with the majority being transit agreements. The transit agreements occurs when an ISP's traffic flow into another ISP's network en-route to other destination on the Internet. Peering agreements are meant to apply a direct route into another ISP's network as most network traffic's final destination are in the latter's network. Peering agreements tend to be in the form of barter whereby one ISP exchanges outbound traffic to another ISP for inbound traffic from that ISP to its network (normally without incurring any major fees). Transit agreements are calculated based on the 95% percentile whereby a billing period's traffic is analyzed by sorting the traffic from the largest to the smallest and the top 5% of the traffic is then discarded and the remaining 95% is billed (DUN001). Network traffic tends to be in bursts and therefore this computation would ignore the impact of any abnormal traffic bursts.

The pendulum effect between the ISPs and the P2P systems and the P2P systems' natural tendency in causing multiple cross-boundary traversals from an ISP into other ISPs (resulting from the P2P's hashing algorithms of its DHT), causing high transit traffic and transit costs. The situation also causes difficulties in getting peering agreements since no one wants to be a partner in a lopsided peer agreement or create peer agreements when the traffic ultimately uses an ISP to transit into another ISP. The P2P systems' tendency to cause multiple connections within an autonomous system also causes nuisance to the ISP's internal networks at a time when the ISP's are trying to minimize investments.

1.2.3 Hybrid web search engines are being researched but are not being deployed as part of the solution

Whilst the idea of an integrated module (composing of a web search engine with P2P capabilities and Traffic Engineering capabilities) is novel, progress has already been made towards that direction in the form of distributed web search engines which are used to enhance the robustness of the web engine's performance but not as part of solutions to resolve the tensions between the ISPs and P2P systems.

In a study of a P2P based web search engine (WAN001) which aims to achieve better scalability by not having a centralized engine to prevent a single point of failure. GALANX, the name of the P2P based web search engine, has 2 sets of peers which are consistently online acting as information providers and the other sets of peers are those who join and leave the P2P systems as consumers of the search services. The data indexes which index the data storage (data is not moved around) and the peer indexes which keep track of the peers who hold the documents are managed by the Distributed Hash Tables.

In a separate study (MIC001), a P2P based web search engine which distributes both data and query processing. The data posting algorithm uses an order-preserving hash function and performs load-balancing by migrating data. Query processing is by means of a framework of highly distributed versions of top-k algorithms by using simple distributed top-k algorithms to manage vertical and horizontal data replication. Replication is made for popular search terms so that delays in searching are minimized as the searched items are available locally as opposed to the normal chaining process in which the peer first searches locally and then proceed to continually search each neighbouring peer until a peer which holds the necessary data index on the subject is located. Fast and efficient sequential access of index list's data which requires a maximum of one hop communication and the use of algorithms which exploit data replicas.

In line with progression towards open architectures, ODISSEA (SUE001) which stands for Open DIStributed Search Engine Architecture has a 2 tiered architecture. The lower tier architecture has a global index structure against which searches can be made. The upper tier has update clients which are used to update the top tier's global index and has query clients which search the global index. Searches could be performed on a node or span across a set of nodes. Therefore, queries with multiple keywords would require combination of results of the various individual keyword searches over the network. The underlying global namespace uses a DHT structure with each object being identified by a hash of its name.

The other model, which is similar to a simplified search engine but is actually a publish/subscribe service for unstructured P2P communities, is the PlanetP (CUE001) with content search and ranking capabilities. It requires peers to propagate information using periodic, randomized, point-to-point message exchange with each other. It has a globally content-ranked data collection, a local index data structure on each peer that describes content held by that peer and a global index data structure of each peer's shared content.

PlanetP uses the gossiping algorithm to propagate the global index data structure across the peer nodes. The gossiping algorithm is composed of a rumouring algorithm, anti-entrophy algorithm and a partial anti-entrophy algorithm.

The rumouring algorithm causes new information (such as new shared content) to be propagated by a peer node to a finite number of randomly chosen peer node in the global namespace every few seconds. Theses chosen nodes will then propagate the information to other peer nodes in similar fashion.

The anti-entrophy algorithm is used to ensure that the process of change propagation does not die out before all of the peer nodes have been updated. A test is made by having a peer randomly choose another peer every few seconds and request from the chosen peer a copy of is global index. The requesting peer then updates its global index with the information new to it.

The partial anti-entrophy algorithm is meant to cut the time spent in propagation of changes. It does this by piggybacking on top of the rumouring algorithm such that the pushing node updates a pulling node with updates to the global index, the pulling node sends a small list of the most recent changes that it has received back to the pushing node and the pushing node then updates its global index with changes that it has not received previously.

The JXTA framework on which the proposed modified search engine of this research uses a similar mechanism as the rumouring algorithm of the PlanetP's distributed web search engine.

1.3 Objectives of the Study

The main objectives of the research project are:

- To recommend a solution that I believe is better able to minimize the tensions between the ISPs and P2P systems as compared to other alternatives proposed.
- To provide supporting evidence of the effectiveness of the solution proposed which uses Software Defined Networks, JXTA P2P components, Apache Solr web search engine and customized programming.
- To document any lessons learnt during the research that could lead to further work in future towards betterment of the solution proposed or in the testing methods.

1.4 Research Questions and Hypotheses

The main research question and hypothesis in whether the systems proposed is able to offer a better alternative to current practices and other solutions proposed by other researchers.

The search for an answer to this question has directed the methodologies in Chapter 3, the analysis of data and findings in Chapter 4 and the conclusion in Chapter 5.

1.5 Significance of the Study

The primary concern of the study is the effectiveness of the solution proposed in actualizing cost savings resulting from savings of transiting network traffic between autonomous networks but still providing reasonable performance in downloading of content.

The cost savings could then be invested into enhanced network efficiency and expansion into other services and enabling better customer satisfaction and retention of customer loyalty.

A full analysis of the cost and benefits are provided in Appendix A, using Telekom Malaysia Berhad as a potential client for the solution. This illustrative Business Case seeks to provide an insight to the potential value of the systems developed. Based on a projected network efficiency of 10%, the annual returns in terms of operating profits could be as much as RM101 Million over the 5 years useful life of the systems. Based on this same projection, additional cash inflows from savings in capital expenditure and operating expenditure could be as much as RM 230 Million per annum over the useful life of 5 years.

1.6 Scope of the Study

The search for a plausible solution and the validation of the said solution in response to the pressing concerns caused by the tensions between the ISPs and P2P systems and gaps in the research (as elaborated in sections 1.1 and 1.2) are the main areas of concern in this research.

The scope of the research is primarily centered on:

Firstly, a literature review is made of the previous attempts made by other researchers in finding a method towards minimizing as documented in Chapter 2 and a deep analysis is made as to the appropriateness of these solutions in minimizing the tensions.

Secondly, a secondary research study is made of the technologies and open source components that could be used to build the proposed solution.

Thirdly, detailed methods used through computer simulation are used in testing the validity of the solution in comparison to other alternatives presented by other researchers as documented in Chapter 3.

Thirdly, detailing findings and analysis thereof resultant of the procedures carried out in testing the validity of the solution in Chapter 4.

Fourthly, a conclusion and summary of salient facts is noted in Chapter 5 based on the analysis and work done from Chapter 3 and Chapter 4.

Lastly, documentation is made of the various factors that could be improved upon to enhance the efficiency of the modified search engine and factors that could better aid the testing process are looked at in Chapter 5 as potential future research work.

CHAPTER 6

REFERENCES

AGG001 Aggarwal, Feldman and others (2007) Can ISPs and P2P users Cooperate for Improved Performance? ACM SIGCOMM, Computer Communication Review Volume 37,Number 3, July 2007

AHN001 Ahn and Chun (2001) Overview of MPLS Network Simulator: Design and Implementation

BIG001 www.bigswitch.com (2013) Big Network Controller Datasheet http://www.bigswitch.com/sites/default/files/sdnresources/bncdata sheet_0.pdf (page 2)

BIN001 Bindal, Cao and others (2006). Improving Traffic Locality in BitTorrent via Biased Neighbour Selection. Proc. 26th IEEE International Conference Distributed Computing Systems 2006

CAB001 Cabling I. & M. Magazine (2012) Report: Global Telecom spending to hit $223.3 billion by 2017Cabling Installation & Maintenance Magazine 17 Dec 2012

CAI001 www.caida.org (2013) www.caida.org/data/overview (TraceRoute Probe Method 2008- Aug). http://data.caida.org/datasets/topology/trmethod-200808/README

CAR001 Cariden Technologies Inc. (2012) (www.Cariden.com) IP-MPLS Traffic Engineering.pdf. Cariden is a leading service provider to telco carrier networks

CAS001 Castro, Druschel and others (2002) Topology-aware routing in structured peer-to-peer overlay networks. Tech. Rep. MSR-TR-2008-82

CHO001 Choffnes and Bustamante (2008) Taming the Torrent. Proceedings of ACM SIGCOMM Aug 2008

COM001 Comcast Corporation (2008) WC - Docket No 07-52. Comments of Comcast before the Federal Communications Commission Washington DC

DAN001 Dán,Hosfeld and others (2011) Interactions Pattern between P2P Content Distribution Systems and ISPs. IEEE Communications Magazine May 2011

DAN002 Dán, György (2013) Cache-to-Cache: Could ISPs Cooperate to Decrease Peer-to-Peer Content Distribution Costs? IEEE Transactions of Parallel and Distributed Systems

DUN001 Dunaytsev, Roman (2012) ISP Interconnection and Traffic Exchange : An Overview. Space Internetworking Center, Democritus University of Thrace

EBR001 Ebrahim, Khan and others (2012). Peer-to-Peer Network Simulators: an Analytical Review. Asian Journal of Engineering, Sciences & Technology 2.1 (2012)

EFF001 Electronic Frontier Foundation (2007) Packet Forgery By ISPs: A Report of The Comcast Affair

EGE001 Eger, Hosfeld and Kunzmann (2007) Efficient Simulation of Large-Scale P2P Networks : Packet-Level vs Flow-Level Simulations. 15th EUROMICRO Int. Conf. On Parallel, Distributed and Network-based Processing 2007

FLO001 www.projectfloodlight.org (2013) Floodlight Architecture. http://docs.projectfloodlight.org/display/floodlightcontroller/Architecture

FOR001 Forouzan, Behrouz A. (2013) Data Communications and Networking 5th Edition Global Edition extracted from Chapter 29 Peer-to-Peer Paradigm, pages 1024 to 1029

GRA001 Grainger and Potter (2013) Solr in Action. Chapter 1 Introduction to Solr, Manning Publications

GRA002 Gradecki and Gradecki (2002) Mastering JXTA : Building Peer-to- Peer Applications. (Java open Source Library) Wiley & Sons Publication

GSM001 GSMA Intelligence (2013) Traditional ARPU distorting consumer mobile spending trends. GSMA Intelligence February 2013

GUM001 Gummadi, Dunn and others (2003) Measurement, Modelling and Analysis of a Peer-to-Peer File- Sharing Network. Proceedings of 19th ACM Symposium on Operating Systems Principles 2003

GUM002 Gummadi,Gummadi and others (2003). The Impact of DHT Routing Geometry on Resilience and Proximity. Proceedings of ACM SIGCOMM 2003

GUP001 Gupta, Ghonge and others (2013) Open-Source Network Simulation Tools: An Overview. International Journal of Advanced Research in Computer Engineering & Technology Vol 2 April 2013

HAF001 Hafeeda and Saleh (2008) Traffic Modelling and Proportional Partial Caching for Peer-to-Peer Systems IEEE/ACM Transactions on Networking Vol.16 No. 6 Dec 2008

HAL001 Haldar and Chen (2002) .Network Simulator (ns) Tutorial 2002 – Introduction. http://www.isi.edu/nsnam/ns/ns-tutorial/tutorial-02/

HOS001 Hosfeld, Hausheer and others (2009). An Economic Traffic Management Approach to Enable the TripleWin for Users, ISPs and Overlay Providers. Proceedings of 2nd Future of Internet Conference May 2009

IET001 Internet Engineering Taskforce (IETF) (1998) RFC 2475 : An Architecture for Differentiated Services

IET002 Internet Engineering Taskforce (IETF) (2006) RFC 4594 : Configuration Guidelines for DiffServ Service Classes

IET003 Internet Engineering Taskforce (IETF) (2008) RFC 5127 : Aggregation of DiffServ Service Classes

IET004 Internet Engineering Taskforce (IETF) (2007) RFC 4195 : Multi-Topology (MT) Routing in OSPF

IET005 Internet Engineering Taskforce (IETF) (2008) RFC 5120 : M-ISIS Multi-Topology Routing in Intermediate System to Intermediate System

IET006 Internet Engineering Taskforce (IETF) (2001) RFC 3031 : Multiprotocol Switching Architecture

IET007 Internet Engineering Taskforce (IETF) (1999) RFC 2702 : Requirements for Traffic Engineering over MPLS

IOS001 Iosup, Garbacki and others (2005) Correlating Topology and Path Characteristics of Overlay Networks and the Internet. http://www.pds.ewi.tudelft.nl/~iosup/

ISH001 Ishiguro and others (2013) Quagga 0.99.22 A routing software package for TCP/IP networks. http://www.nongnu.org/quagga/docs.html

KPM001 KPMG International (2012) Telcos scrambling for Spectrum. Issues Monitor Sept 2012 Vol. 11

LEI001 Leibowitz, Bergman and others (2002) Are File Swapping Networks Cacheable ? Characterizing P2P Traffic . Proceedings of 7th International World Wide Web Caching Workshop (WCW-7) August 2002

LIU001 Liu, Cui and others (2009) Locality-Awareness in BitTorrent-like P2P Applications. IEEE Transactions of Multimedia Vol 11 No 3, 2009

MAS001 Maslow and Sneppen (2002) Specificity and stability in topology of protein networks. Science (2002) and arvix.org

MAX001 Max Planck Institut Informatik (2014). Network Analyzers' interpretations and formulas (http://med.bioinfo.mpi-inf.mpg.de/netanalyzer/help/2.7.

NAS001 Nascimento, Rothenberg and others (2010). QuagFlow : Partnering Quagga with OpenFlow. ACM SIGCOMM 2010, 30 August - 3 September 2010

NYT001 The New York Times (2007). Comcast: We are Delaying, Not Blocking, BitTorrent Traffic

PAL001 Palo Alto Networks (2013). The Application Usage and Threat Report April 2013

PAP001 Papafili, Soursos and other (2009). Improvement of BitTorrent Performance and Inter-Domain Traffic by Inserting ISP-owned Peers. Proc. of 6th International Workshop of Internet Charging and QoS Technologies May 2009

QUA001 www.quagga.net (2013) About Quagga. Website will redirect to www.nongnu.org/quagga

SEE001 Seeley, Yonik (2013) people.apache.org/~yonik/presentat ions/solr_architecture.ppt. Yonik Seeley is a Lucene/Solr Committers team member, contributing developer

SHE001 Shen, Wang and others (2007) HPTP: Relieving the Tension between ISPs and P2P. IPTPS conference 2007

SNA001 Leskovec, J. and Krevl, A. (2014). SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data

XIE001 Xie, Yang and others (2008). P4P: Provider Portal for Applications. Proceedings of ACM SIGCOMM Aug 2008

YUJ001 Yu and Li (2008). CBT: A Proximity-Aware Peer Clustering System in Large Scale BitTorrent-like P2P Systems. Computer Communications Vol 31 No 3, 2008