PROXCACHE: A NEW CACHE DEPLOYMENT STRATEGY IN INFORMATION-CENTRIC NETWORK FOR MITIGATING PATH AND CONTENT REDUNDANCY

ABDULLAHI IBRAHIM

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2016

Awang Had Salleh Graduate School of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

ABDULLAHI IBRAHIM

calon untuk Ijazah **PhD**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"PROXCACHE: A NEW CACHE DEPLOYMENT STRATEGY IN INFORMATION-CENTRIC NETWORK FOR MITIGATING PATH AND CONTENT REDUNDANCY"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : *18 Mei 2016.*
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
May 18, 2016.

Pengerusi Viva: *(Chairman for VIVA)* **Prof. Dr. Zulkhairi Md Dahalin** Tandatangan *(Signature)*

Pemeriksa Luar: *(External Examiner)* **Prof. Dr. R Badlishah Ahmad** Tandatangan *(Signature)*

Pemeriksa Dalam: *(Internal Examiner)* **Dr. Massudi Mahmuddin** Tandatangan *(Signature)*

Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)* **Dr. Ahmad Suki Che Mohamed Arif** Tandatangan *(Signature)*

Tarikh:
(Date) ***May 18, 2016***

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Salah satu paradigma yang menjanjikan perkongsian sumber adalah dengan mengekalkan asas semantik Internet ialah Rangkaian Maklumat Berpusat (ICN). Perbezaan ICN dengan Internet semasa adalah dari segi keupayaan merujuk kandungan melalui nama dan sebahagiannya dengan memutuskan amalan hos-ke-hos pada alamat-alamat protocol *Internet*. Tambahan pula, pengagregatan kandungan dalam *ICN* dilihat sebagai tindakan utama untuk mencapai rangkaian-kandungan bagi tujuan mengurangkan jumlah capaian pelayan. Amalan pengagregatan semasa dalam ICN menggunakan *Tinggalkan Salinan di Semua Tempat*, menjanakan masalah lebihan pemendapan kandungan yang dikenali sebagai kelewahan kandungan, kelewahan laluan, kurang kadar kenaan-cache dalam pelbagai rangkaian dan rendah kepelbagaian kandungan. Kajian ini mencadangkan strategi baru mengatur kedudukan cache yang dirujuk sebagai *ProXcache* untuk memperoleh hubungan nod menggunakan konsep hiper-pinggir daripada hiper-graf untuk menentukan kedudukan cache. Kajian ini merumuskan hubungan melalui anggaran laluan dan jarak untuk mengurangkan kelewahan kandungan dan laluan. Kajian ini mengguna pakai pendekatan Reka Bentuk Kaedah Penyelidikan (DRM) untuk mencapai objektif-objektif penyelidikan. ProXcache telah dikaji menggunakan penyelakuan pada topologi rangkaian Abilene, GEANT dan DTelekom untuk strategi pengagregatan *LCE* dan *ProbCache* dengan menggunakan taburan *Zipf* untuk membezakan pengkategorian kandungan. Hasil kajian menunjukkan kelewahan kandungan dan laluan keseluruhan dikurangkan dengan operasi pengagregatan kurang enam pengendapan untuk setiap permintaan berbanding sembilan dan sembilan belas bagi *ProbCache* dan *LCE* masing-masing. ProXcache menghasilkan nisbah kepelbagaian kandungan yang lebih baik iaitu 80% berbanding 20% dan 49% untuk LCE dan ProbCache apabila saiz cache diperbagaikan.Hal ini juga meningkatkan nisbah kenaan-cache melalui kedudukan proksi dalam *ProXcache*. Semua ini ada pengaruh yang signifikan pada pembangunan ICN untuk pengurusan kandungan yang lebih baik ke arah penggunaan dalam Internet masa depan.

Kata kunci: Internet Masa Depan, Kepelbagaian kandungan, Strategi pengagregatan, Hiper-graf

Abstract

One of the promising paradigms for resource sharing with maintaining the basic Internet semantics is the Information-Centric Networking (ICN). ICN distinction with the current Internet is its ability to refer contents by names with partly dissociating the host-to-host practice of Internet Protocol addresses. Moreover, content caching in ICN is the major action of achieving content networking to reduce the amount of server access. The current caching practice in ICN using the Leave Copy Everywhere (LCE) progenerate problems of over deposition of contents known as content redundancy, path redundancy, lesser cache-hit rates in heterogeneous networks and lower content diversity. This study proposes a new cache deployment strategy referred to as ProXcache to acquire node relationships using hyperedge concept of hypergraph for cache positioning. The study formulates the relationships through the path and distance approximation to mitigate content and path redundancy. The study adopted the Design Research Methodology approach to achieve the slated research objectives. ProXcache was investigated using simulation on the Abilene, GEANT and the DTelekom network topologies for LCE and ProbCache caching strategies with the Zipf distribution to differ content categorization. The results show the overall content and path redundancy are minimized with lesser caching operation of six depositions per request as compared to nine and nineteen for ProbCache and LCE respectively. ProXcache yields better content diversity ratio of 80% against 20% and 49% for LCE and ProbCache respectively as the cache sizes varied. ProXcache also improves the cache-hit ratio through proxy positions. These thus, have significant influence in the development of the ICN for better management of contents towards subscribing to the Future Internet.

Keywords: Future Internet, Content diversity, Caching strategy, Hypergraph.

Declaration Associated with This Thesis

Some part of the work presented in this thesis have been published and submitted to the following journals and conferences listed below:

[1] **Ibrahim Abdullahi**, Suki Arif, and Suhaidi Hassan. Survey on caching approaches in information centric networking. *Elsevier Journal of Network and Computer Applications, (56):48-56, July 2015. Rank Q1, ISI and Scopus Indexed.*

[2] **Ibrahim Abdullahi**, Suki Arif, and Suhaidi Hassan. Ubiquitous Shift with Information Centric Network Caching Using Fog Computing. *Advances in Intelligent Systems and Computing. Springer International Publishing, (331):327-335, January 2015. ISI and Scopus Indexed.*

[3] **Ibrahim Abdullahi**, Suhaidi Hassan, and Suki Arif. Prospective Use of Bloom Filter and Muxing for Information Centric Network Caching. *ARPN Journal of Engineering and Applied Sciences, 10(3):1169–1177, February 2015. Scopus Indexed.*

[4] **Ibrahim Abdullahi**, Suki Arif, and Suhaidi Hassan. Content Caching in ICN Using Bee-Colony Optimization Algorithm. *Advanced Science Letters, Vol. 21 (11), pp 3538-3542, 2015. ISI and Scopus indexed.*

[5] **Ibrahim Abdullahi** and Suki Arif. Cache-less Redundancy Using Hypergraph in Information Centric Network. *Advanced Science Letters,Vol. 21 (11), pp 3548-3551, 2015. ISI and Scopus Indexed.*

[6] **Ibrahim Abdullahi** and Suki Arif. Cache-Skip Approach for Information-Centric Network. *ARPN Journal of Engineering and Applied Sciences, 11(5):3413–3418, March 2016. Scopus Indexed.*

[7] Suki Arif, Suhaidi Hassan, **Ibrahim Abdullahi**. Cache Replacement Positions in Information-Centric Network. 4th International Conference on Network Applications, Protocol and Services (NETAPPS2015), Cyberjaya, Malaysia. 1-3 December, 2015.

[8] **Ibrahim Abdullahi**, Suki Arif and Mohd. Hasbullah Omar. *Positioning Cache in Information Centric Networking. Advanced Science Letters 2016. ISI and Scopus Indexed.*

[9] **Ibrahim Abdullahi**, Suki Arif . *Content Diversity in Information-Centric Network Caching . Advanced Science Letters 2016. ISI and Scopus Indexed.*

[10] Fatima Binta Adamu, Adib Habbal, Suhaidi Hassan, R. Les Cottrell, Bebo White, **Ibrahim Abdullahi**. A Survey on Big Data Indexing Strategies . 4th International Conference on Network Applications, Protocol and Services (NETAPPS2015), Cyberjaya, Malaysia. 1-3 December, 2015.

[11] **Ibrahim Abdullahi**, Sahalu Junaidu, U. M Gana, Fatima B. Adamu. Scheduling: Potential Criteria Evaluation Using Longer Job First. 4th International Conference on Network Applications, Protocol and Services (NETAPPS2015), Cyberjaya, Malaysia. 1-3 December, 2015.

Acknowledgements

In the name of ALLAH, Most Gracious, Most Merciful:

"*Glory be to Thee! we have no knowledge but that which Thou hast taught us; surely Thou art the Knowing, the Wise*". (The Holy Qur'an - *(Surah Al Baqarah 2:32)*)

All praise and glory be to Allah for granting me health, strength and knowledge to attain this stage of my life journey. The favors and mercy of preserving me are undeniable. My profound and sincere appreciation goes to my Supervisor, Dr. Ahmad Suki Mohamed Arif for granting me the space and scholarly guidance through this wonderful journey. It would have been almost impossible without you. Your professionalism and daily inspection on the work and my well-being was a key. Working in this area with you has broaden my understanding and improved my thinking; I say a very big thank you and may GOD increase you in wisdom, strength, health and wealth.

My special acknowledgment is to my super Supervisor, Prof. Dr. Suhaidi Hassan. Your guidance is enormous that words are short to describe. I appreciate the guidance and advices through presentation and objective criticism that binds us to the Laboratory. I want to thank you for providing us with the benevolent and most conducive research environment.

My unreserved gratitude goes to my Parents, brothers and sisters; your prayers, supports, love and understanding kept me going. The strength of you all motivated my journey to this length. You all showered your motivation to me that I could do it. I pray that Allah almighty continue to strengthen our bonds and bless the entire household.

To my first research Teammate of ICN caching; Ikram Ud Din. Your motivational force of working tirelessly through the difficult times of simulation, weather and descriptive analysis shall never be forgotten. May God reward you abundantly. My appreciation also continue to my International supports team specifically Ioannis Psaras, Saino Lorenzo for lending me a helping hand during my challenging time with the ICARUS caching simulator. Dr. Cesar Bernadini, your relentless responses to my emails are appreciated. Your support through guiding me during the simulation with SocialCCNSim shall forever remain in my heart.

I want to specifically appreciate the efforts of the entire InterNetWorks Research Laboratory, Universiti Utara Malaysia Team. You are my family forever and I shall live to be honored to have ever worked with you. IEEE, ACM, Institute of Neural Network Society, ICNRG and Internet Society has been too much of a help through out, Thank you.

To the esteem team of the dedicated thesis reviewers. The Chairman, Prof. Dr. Zulkhairi Md Dahalin, Prof. Dr. R. Badlishah Ahmad my external examiner from Universiti Perlis, Malaysia. Your professional suggestions and recommendations were golden in adding value to the thesis. My out most appreciation goes to Dr. Massudi Mahmuddin who critically guided me through constructive inputs; without your sacrifices of reading and grading my work, the worthiness would not be there; I pray to God to increase you in knowledge health and wealth. I also extend my appreciation to Prof. Dr. Salmah Bint Ayob, Prof. Dr. Zulaikha Jamaludeen and Associate Professor Osman Ghazali for their respective suggestions at the begining of my PhD journey.

I want to specially thank my University, Ibrahim Badamasi Babangida University, Lapai for deeming it fit to encourage my coming to study in UUM. Your support shall never be forgotten. To the entire members of the Department of Mathematics and Computer Science. My Masters degree Professor Sahalu B. Junaidu, Associate Professor S. E Abdullahi, Professor Adewale Olumide, and Associate Professor Buhari Seyed; I owe my research background to your scholarly coaching and mentoring.

A special appreciation goes to the wonderful orthopedic Dr. Charanjeet Singh for operating and constructing a new Anterior Cruciate Ligaments and Meniscus on my right leg after the damage.

My all-out appreciation, love and hearty cheers goes to my darling wife Fatima Binta Adamu, your understanding and sacrifices are second to none. Your motivation and supports kept me going even in the most difficult times of my study, my accidents and surgery. Your demonstration and care is truly a reflection of the angelic blessing you ferried into our young family; our little baby Hanan. Both of you are worthy of the status of heroines as you showed me LOVE, patience and sacrifices. May Allah bless you with Jannah and the best in this world. I LOVE you.

Dedication

To my family and TEACHERS at all capacity of my Knowledge pursuit.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AS	-	Autonomous System
CCN	-	Content Centric Network
COMET	-	Content Mediator Architecture for Content Aware Network
CS	-	Content Store
DHT	-	Distributed Hash Table
DONA	-	Data Oriented Network Architecture
DRM	-	Design Research Methodology
DS-I	-	Descriptive Study-I
DS-II	-	Descriptive Study-II
EU	-	European Union
FIB	-	Forwarding Information Base
ICN	-	Information Centric Network
ICNRG	-	Information Centric Network Research Group
IP	-	Internet Protocol
ISP	-	Internet Service Provider
LCD	-	Leave Copy Down
LCE	-	Leaave Copy Everywhere
MDHT	-	Mulitlevel Distributed Hash Table
NDO	-	Named Data Object
NetInf	-	Network of Information
NR	-	Name Resolution
NRS	-	Name Resolution System
PIT	-	Pending Interest Table
PS	-	Perspective Study
PRE	-	Path Redundancy Elimination
ProbCache	-	Probabilistic Cache
ProXcache	-	ProXy Cache
PSIRP	-	Publish-Subscribe Internet Routing Paradigm
PURSUIT	-	Publisher Subscriber Internet Technology

P2P	-	Peer-2-Peer
RENE	-	Rendezvous Network
RH	-	Resource Handlers
SAIL	-	Scalable Adaptive Internet Solution
SMA	-	Shared Memory Architecture
SSL	-	Secure Socket Layer
TSB	-	Time Since Birth
TSI	-	Time Since Inception
UGC	-	User Generated Content
URL	-	Uniform Resource Locator
VNI	-	Virtual Network Index
VoD	-	Video on Demand

CHAPTER ONE
INTRODUCTION

1.1 General Overview

Information dissemination has been the main idea that gave birth to the Internet. Its flexibility has since been widely acceptable due to the Internets benefits out-weighing the threats of security, privacy and other vulnerabilities associated to its practice. The current Internet was built on the architectural plan of host-to-host communication idea. However, benefiting users are tending to be less concerned about the host of these information with the quest of having to concentrate more on the content ahead of the host address. This practice on the Internet is therefore seen as the projection for the future Internet to Internet of Things (IoT), Internet of Everything (IoE), Cloud infrastructure and fifth generation (5G) technology. With the huge amount of data being requested and transferred over the Internet hitting the mark of about 1,000 *Exabytes* (*Zettabyte*) in year 2016 according to Cisco *Virtual Network Index* (VNI) [1, 2, 3]. Therefore, a need to subscribe into the Information-Centric Networking is deemed appropriate. As the name implies, Information Centric Network (ICN) [4, 5] usually described alongside Networking Named Content [6] is aimed at achieving the possibilities of bringing a new dimension and improved information dissemination on the Internet.

The major point of distinction between the traditional/conventional Internet will be its ability to use names (content-aware) and not the host address content like the conventional Internet Protocol (IP) addressing. Several researches and studies have received huge interests by proposing various architectural platforms in ICN to meet the yearned aspiration for a shift to content-centric networking. ICN projects has proven that the aspiration will soon be a success in line with the huge support it has been receiving through European Union FP7 projects, Internet Research Task Force (IRTF) and Internet Engineering Task Force (IETF) [6, 7, 8, 9]. The advantages of this new paradigm cannot be over emphasized as bandwidth consumption which has necessitated the need for its adequate usage will be reduced through neighbor node information/interest distribution (caching) and power consumption efficiency. These aims at minimal utilization of network services along path since high inference to the main server will be drastically reduced by caching (*on-path*) on the nodes.

ICN approach and practice provides an advantage of the multi-information sharing in a form identified as extension of the Content Delivery Network (CDN) [10]. Consequently, ICN has received several proposal and architectural projects. However, ICN popular projects are still faced by some challenges of using the named data alongside caching for its conformity to provide the basic Internet needed services and maintain the current state of Internet reach-ability. ICN has therefore been promised to relief some magnanimous load of data in the future [10]. With the nature of the ICN, its building block in terms of data structure needs to be modified to accumulate the built-in functionality of its proposal. ICN provides the network functionality of granting information to neighbor node when interests are sent. In ICN, users normally referred to as Subscribers send out requests for data or named data object (NDO). The NDO is then transferred according to the routing information in the dedicated routing base called pending Interest table (PIT) and forwarding information base (FIB) as used in some ICN projects. Once the NDO is matched in the network, the publisher of the NDO cache the request and sends out the resulting data without specifying its address.

ICN project has become so interesting as researchers have identified many issues in designing architectures, frameworks and algorithm to drive the quest to achieving Content Centric Networking [6, 11, 9, 12, 8]. The influencing factors can be further elaborated as: Frequency - in number terms, how many requests are posted or how frequent is an object requested for? Recency - the time an object or content was referred to or demanded for. Size - the size of content, cost of retrieval - the cost incurred to retrieve the content or object, time of update - a modification in the cache, placement and replacement - the best time content becomes less relevant in ICN.

The thesis is organized as thus: Section 1.1 provides the description of ICN, while in Section 1.2 detailed description of the ICN architecture and research issues are presented. Section 1.3 presents the motivation of the study and the following section, 1.4 provides the introduction to cache deployment strategies. Section 1.6 highlights the problem statement of the research and Sections 1.7 and 1.8 covers the research questions and objectives respectively. The research scopes and synopsis of the study is described in Sections 1.9 and 1.10 while the significance and organization of the research in Section 1.11 and 1.12 concludes the chapter.

1.2 Overview of ICN

Information-Centric Network (ICN) is a network driven on content, where the nodes serve as the point of reference and referral to the user and the publisher (host) [13, 14, 15, 4]. ICN built-in differ and provide the advantages of fast-look up, earlier delivery through the ability of serving contents when they are cached. In ICN, once a path has been visited, establishment of the content or content chunks are lodged in the nodes. A clear difference between the host-to-host Internet and the content-aware Internet is the less visitation to the host [16]. Subscribers (referred to as users) pose and send request in ICN without necessarily knowing where the data is lodged. The lack of specification of address in ICN benefits the architecture by answering the question of who has what data, against where do data lodge [17, 18, 6]. This is the flexibility to a user to have more concern and reservation for the content ahead of its location. A strong hold to this paradigm is the focus of data being transferred among users than among hosts. ICN premise is the actualization of lesser congestion, better scalability, speed and fast look-ups, flexibility of configurations and security [10, 19].

1.3 ICN Architectures and Research Issues

ICN yearns to change the Internet semantics and rule of fetching data from a recognized host destination to specifically transporting the exact named data to the user. A good related architecture of the ICN is the Named Data Network (NDN) [4, 20, 6, 21] which consist of three special data structures of Content Store (CS), Pending Interest Table (PIT) and Forwarding Information Base (FIB). In ICN practice, the data is driven by the subscriber who sends out interest to the network by only binding the name with a prefix. Once the interest leaves the Subscriber node, the interest is copied and recorded in the PIT which takes the count of serving the interest. In the occasion when the interest is found on-path in the network, the CS holding a chunk of the data feed back the data using the path where the interest traversed through. The interest status are always recorded in the PIT as granted or otherwise [14]. Once a node receives an interest, it is looked up in the PIT to serve. If the interest is not found in the node PIT, a pending mark is recorded for future delivery of the interest. The FIB serves the routing information provider by specifying the interest forwarding to a data granter (Publisher).

The Publisher in ICN is the producer of the data as referred by the Publisher and Interest from the subscribers perspective. As shown in Figure 1.1, a user sends an Interest through router A, the first router on the path checks its data structures (CS, PIT, FIB) to provide the data requested by Client 1. In the course, the Interest is forwarded to the neighbor router to check the availability of the data. Router D is visited until it reached the router holding a chunk of the requested interest. From the Publisher end, the Publisher publish the data into the network using the path A-B until it gets to the subscriber (Client 1). The operation continues in this form by adding the advantage of next iteration of request saving the long traversing operation. Once the data is released into the network by Publisher, the cache-ability of ICN routers provide the flexibility of Client 4 router to cache from Router D. This makes ICN fast look up improved and information dependability relieved from the host. Congestion and manageability is thus improved through sharing of chunks from neighbors. In ICN, this operations are all coordinated by the operations of name resolution, routing, forwarding and caching.

Caching provides the flexibility advantage of minimizing the cost of traversing through out the network. When an Interest is sent and fed, the subsequent demand for the information is easily fetched from a neighbor node as submitted by [6, 4, 9]. The studies in [6] and [22] argued that the advantages of dissociating host and IP will also improve fast delivery of services, better scalability, achievable time and improved due dates among others. These trend is therefore achievable by caching the contents on router nodes while data and Interests traverse. Caching in ICN is the ability of a node/router to predictably save and hold data chunk in the network for transfer and future referral. Recent studies and submission of caching practice in ICN has been faced with the problem of content flooding, un-referred data, data deposition leading to the path being crowded and content been redundant. Content redundancy can be said to be the unwanted deposition of data in a node/router (memory) as cache without necessarily needing the content over a given period of time.

Caching in ICN is therefore the predictor operation that defines the architectural composition of Information-centrism to guaranty future data manageability on the Internet. However, with the Leave Copy Down (LCD) and Leave Copy Everywhere (LCE) caching practices [23, 6, 24], the content availability would be improved but faced with large *data redundancy* as explained in the next sections.

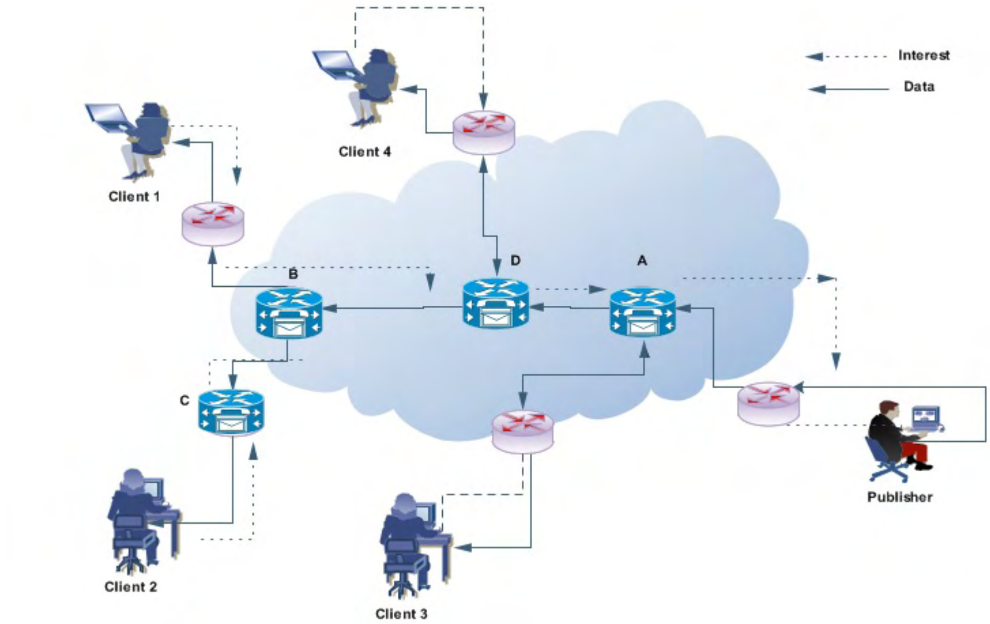

Figure 1.1. ICN Overview

1.4 Research Motivation

Internet usage has been fundamental to all users in different folds [25, 26]. Its incorporation has been widely acceptable to the turn of new faces and direction such as cloud, ubiquitous [27, 28], fog [28], pervasive, fly [29] computing amongst others. However, the need to sustain its satisfactory use is very important. Therefore, since the global IP traffic is exponentially increasing [3, 30, 31, 32, 2, 26, 25], this has motivated this study to adopt the new paradigm of ICN to curtail and mitigate the global data tsunami predicted through cache deployment. Figure 1.2 shows a strong hold of how the current host-to-host Internet is overloaded with high demands on data.

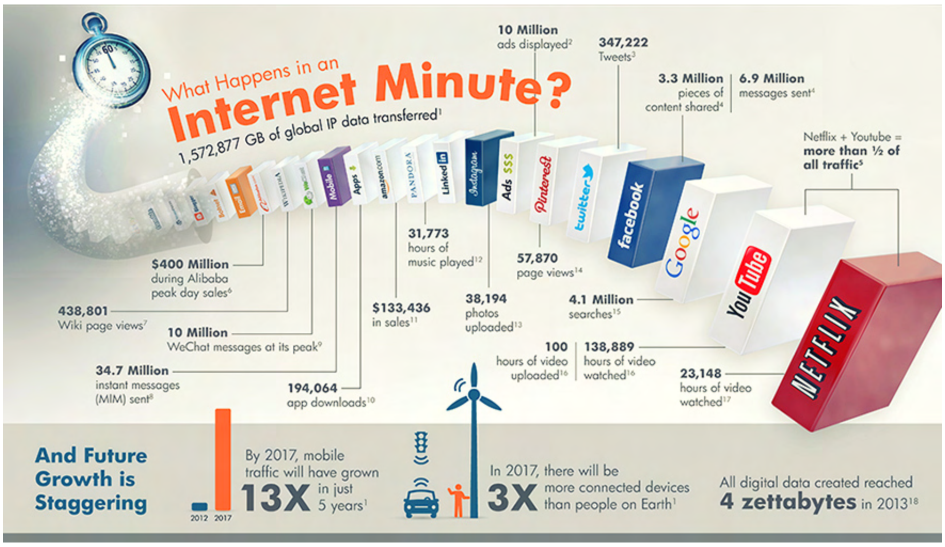

Figure 1.2. Global Internet Traffic in a Minute [32]

According to the latest data traffic and information shared/accessed on the Internet in every minute [33, 32, 3], an approximated over 34.7 Million instant messages are sent over the current Internet [32]. With user diversification tending towards multimedia data, high demand for videos shared on YouTube have recorded a staggering 138,888 hours of video watched against an approximate 100 hours of video uploaded every single minute. This makes up more than half total of the Internet traffic when NETFLIX video and YouTube videos are actively engaged. With these magnitudes of data mostly multimedia, it is worthy to dissociate host-centrism of data to an Information/content-centric paradigm. This would be easily manageable by intelligently deploying cacheable devices in the network through ICN concept.

Additionally, Google being used as one of the leading search engines, earned a high hits of 4.1 million searches in an Internet minute; an average of 4.1 million in every minute of an hour search [32]. ICN have therefore offered the uniqueness of relieving high infer and travel hectic of constant overloading (caching on-path) a host in all Internet activities. Part of the motivation is that, when queries are served, a cache-enable node/router should have a copy in its local content store for further deliverance in lesser time, saved bandwidth, less redundancy on the path and content to mitigate the high cost in traffic.

Determinately, all data created have reached four Zettabytes in 2013 and a staggering 28 billion network devices will be connected to the Internet by 2019 [1, 32, 3]. Thus, making it possible to project that by 2017, there would be a three times fold of devices outnumbering the world population and a triple fold of 2017 traffic in 2019 [32, 1]. According to the Information-Centric Network Research Group (ICNRG) [8], information distribution and manipulation have been the business of the Internet today. Therefore, the need to curtail the challenges posing threats to the future Internet is essential. Among the objectives of the ICNRG [8] according to recent drafts/baseline [34, 35] includes: to propose and create a name resolution, routing and scalability schemes in ICN, storage managements (caching) and performance evaluation of the paradigm. These along other issues makes ICN caching strategy worthy of study to conform with the recent drafts and proposals of ICNRG.

Interestingly to note is the recent report by Cisco in its Visual Network Index (VNI) 2014 [32, 3] that submitted that there are huge amount of information demanded every minute. Users now tend to be more interested in largely multimedia-driven data in lieu of the ordinary file transfers. Another strong motivation behind the paradigm can be based on the report by [2] on the rapid growth of contents on the Internet of about 500 Exabytes in 2008, and thus Internet Protocol (IP) traffic will increase exponentially thus causing heavy blockage in 2009 to 2018 which could be a big problem as the years passes on. It further opened the need to find alternative to the current information dissemination platform on the Internet community benefits. Supporting VNI [3], however, authors in [4, 7] identified other issues that need quick apprehending to have an acceptable framework which include: confidentiality, data integrity, accountability (Owner/Publisher authentication and identification), availability, storage and control access (read, write, execute)[6].

Therefore, special funding and research groups have formed several projects and architectures to fund ICN. Among the leading groups are the European Union Research and Innovation (EU FP7) under the FP7 projects are the 4WARD project, Network of Information (NetInf), Publish Subscribe Internet Routing Paradigm (PSIRP), Publish-Subscribe Internet Technology (PURSUIT)[36, 37], SAIL project [17], COMET, CONVERGENCE [8, 38], COMBO, COAST, GreenICN among others. The US-based ICN projects that amount to billion of dollars ranges from the Content Centric Network (CCN) project, Named Data Networking (NDN) [20] and the Data-Oriented Network architecture (DONA) [21]. A recently added ICN project called Active Content Management at Internet Scale "COMIT" at the University College London (UCL) was launched in January, 2014 to the sum of GBP 486,939. A strong motivation for the need to prepare for the future data tsunami has been the objective of these projects. Accordingly, each approach and proposal differing in its architectural deployment and functionality.

Decisively, the sole purpose of these projects are to develop a network that will be driven by named-data content distribution to manage disruptions, flash-crowd effects, denial of service, lesser utilization of limited resources and improved energy management along neighbor node inference.

Caching can be seen as a probable agent of actualizing the ICN manageability. Due to cache nature of storing chunks of content *on-path* and *off-path*, it will be fair to explore the techniques of *cache deployment*, *content placement* and enhance *content* *retrieval* through various techniques. Appropriate cache is mostly an influencing factor to achieving some of its advantages which include: good use of the bandwidth by reducing wastage thus enhancing the prompt delivery of information (reduced delay) and reducing the overall loads on the main source.

1.5 Cache Deployment Strategies in ICN

The cache deployment strategy can be defined as the methods, operation and cache management in ICN as will be fully explained in Chapter Two. Cache deployment strategies define the forms by which content chunks (data) are carefully saved as cache on nodes and routers. In the early proposal of the ICN, a popular web caching technique known as LCE [23] was used. LCE is the cache management that records a *hit* (found content) and stores a chunk of the data on all traversing nodes to serve the user. As mentioned in Section 1.1, ICN benefits from the node neighbor data transfer. However, in recent researches, LCD was proposed to curtail some of the research problems inherited from the LCE. LCE deposits contents on all nodes leading to content over-flooding the network also referred to as *content redundancy*. LCD cache management defines an extended form of content deployment. In LCD, only immediate node cache the chunk as against the LCE; this practice yearned to solve the problems of over-flooding but challenged with less content hit ratio (see Section 2.1). Thus, cache deployment strategy is said to include the cache positioning of all cache-able nodes, approaches and algorithms involve in cache content management in and out of the network.

1.6 Problem Statement

Information-Centric Network (ICN) as a new paradigm that focuses on naming content ahead of its identification by IP addresses benefits from caching data on router nodes it crosses. However, with this earlier seen as an advantage using the *LCD* and *LCE*

[6, 23, 24] caching forms, unwanted content caches are flooding the network. Thus, holding data on routers has necessitated careful re-organization of the cache positions known as *deployment strategies* [39, 40, 41, 42, 43, 44, 45, 46]. Caching contents on all routers is faced with a high rate of delays in network, excessive wastage of bandwidth resulting to the problem of *content redundancy* and *path redundancy* [24, 47, 48, 49]. Content redundancy is the deposition of chunk data on all routers crossed with lesser and inessential referral to the data in the network. Thus this makes the availability of the content in a router irrelevant and as such resulting into excessive wastage of bandwidth (high caching operation). Path redundancy on the other hand in ICN is the exhaustion of memory spaces for data path delivery through consuming the CS spaces as a result of excessive caching.

ProbCache by [24, 47] was seen as a strategy that has gained high popularity in ICN cache management. The sole objective and strength of the ProbCache was the elimination proposed of the LCE and LCD due to their one position caching scheme except for the LCE that caches everywhere. In ProbCache, cache chunks are probabilistically cached to mitigate content over loading on routers that serves as the Publisher. However, with the probability in place, caches are mostly observed to be cached in routers with probability values *p*. ProbCache used the concept of determining an ideal position which according to the probability nature sometimes increases distances between the hops (known as network stretch) when cache hits occur far from the Subscriber. Using ProbCache sometimes actualize the main goal of the Subscriber when the content increases in popularity but suffers the reverse and deletion when a content become least recently used (LRU). However, as submitted by other researchers [50, 11, 51, 52], proxy caching at the edges of the Subscriber and Publisher would benefit the entire network performance thereby improving the cache-hit ratio [53]. This study therefore propose to tackle some of the issues inherited from the current stage of ProbCache deployment strategy to mitigate content and path redundancy. Proportionally the propose concept will also improve bandwidth efficient usage as submitted in [54].

Therefore the need to have a flexible cache deployment strategy is of essence to mitigate the aforementioned challenges faced by the current ICN cache deployment strategies through acquiring node relationships in the network, distance approximation and enhance content diversity to accommodate the heterogenous state of the Internet. Content-centric networking [4, 6, 7] agreed and predicted the new paradigm of ICN as a major solution to foresighted problems that the Internet may face in the future. ICN like any other technology is faced with providing a good cache deployment idea so as to successfully implement the content-centric communication. The purpose of this study is to explore and propose new and effective caching deployment concept that would mitigate the problems (content and path redundancy) faced by the current ICN design for the future Internet. Proper caching data has therefore been agreed to reduce the challenges but faced with information loss, unauthenticated users, fraud data and information monopoly. Enhancing caching therefore becomes the predictor practice to make ICN a success to overcome these challenges.

1.7 Research Questions

To actualize a new cache deployment strategy in ICN, the following research questions are raised:

1. How would cache proxy approximation be achieved to the Subscriber and Publisher to aid better caching communication?
2. How would cache deployment position be placed for cache effectiveness in a network using hypergraph and obtaining ideal mid-cache positions?
3. How would cache enabled-routers be effectively deployed in ICN to mitigate

path and content redundancy?

1.8 Research Objectives

The objective of this research is to propose a new cache deployment strategy for ICN to achieve better cache-hit ratio and mitigate content and path redundancy.

1. To propose a new network cache mechanism that would acquire the node relationships using an algorithm that will improve content placement in ICN.
   a) Formulate the relationships between edge nodes to obtain cache-time windows (the time a content needs to be cached).
   b) Design an algorithm for path length estimation to aid in determining positions of selecting cache-able routers in the network.
2. To propose a new distance approximation between the Subscriber and Publisher to obtain the network caching for Interests and data.
   a) Formulate the relationship to get the centrality between Subscriber and Publishers at each interest time.
   b) Propose the cache positions (proxy) in the network to enable Subscriber and Publisher close communications using hypergraph arcs to mitigate content and path redundancy.
3. To verify, validate and evaluate the performance of the proposed cache model in ICN using SocialCCNSim caching simulator.

1.9 Research Scope

The entirety of this study is focused on *on-path* caching techniques in ICN. The study implored the data traffic generated using Poisson distribution to model the demands of Name Data Objects (NDO). Caching operations in the research adopted the Least Recently Used (LRU) replacement policy and the resulting strategy focused and compared the findings with LCE, ProbCache caching strategies. Other strategies such as LDC and MCD were only covered to expose the literature connection between the proposed strategy. In ICN caching, edge nodes were seen to be more significant in network and metric comparison for better performance evaluation. Therefore the study used the technique to model the resulting approach in the edge node perspectives with little emphasis made on the intermediary nodes. The study thus covered cache deployment strategies, content fast look-ups from cache deployed routers.

1.10 Research Steps

The following research steps would be carried out to fulfill the goal of this research:

1. Survey of existing ICN cache deployment and content placement strategies to identify the problems.
2. Propose the deterministic and continuous case scenarios of content generation using adequate statistical distribution.
3. Explore the properties of Hypergraph [55] and its manageability to ICN.
4. Develop hypergraph clusters and edge/node selection using an algorithm.
5. Comparative performance evaluation of present cache deployments of ICN and the proposed hypergraph strategy to explore their strengths and weaknesses.

1.11 Significance of the Research

This study will be significantly important in promoting and conceptualizing the infrastructure of the ICN. Its workability through proportioning cache positions in an ICN will save the future designs of physical network devices, time and less bandwidth consumption. When explored effectively, in other networking domains such as Grid computing and Cloud, the submitted ProXcache will enhance due date realization of the former in resource delivery and enhance trust in the latter when using the Infrastructure as a service (IaaS), software as a service (SaaS) among other services. The study will also be significantly helpful in actualizing Internet of Things (IoT) and the Internet of Everything (IoE) due to the architecture of providing the Internet ubiquity. IoE and IoT information needs duplication (caching) to meet up with updates and request fulfillment.

The resulting findings from this research will be used to aid information dissemination on the Internet considering the relevancy in quick data sharing on the Internet. Caching affects the nature of high information reach-ability to end-users of the Internet. Current impact of users demanding high media-driven data would mitigate the problems of over redundancy and flooding of data in locations that are not needed. When caching concepts are improved and standardized, the entire Information-centric network research community shall benefit grossly on its deployment through:

1. Coming up with new caching scheme in ICN that will enhance the content manageability through careful node selection in the network.
2. Propose new caching algorithms that would be used in ICN approach designs with lesser overhead in configurations.
3. Designing and extending the available simulation tools by adding the proposed concepts through programming.

1.12 Organization of the Thesis

The thesis has been organized into seven chapter. The organization of the chapters are as follows:

Chapter One gives the general introduction of the study, its relevancy through importance, justification, the objectives, and research questions set out to be answered by the research.

Chapter Two of the thesis focuses on literature related along facts, arguments and contributions in the fields of Internet, data communication and Information-Centric Network caching but not limited to other web caching algorithms, frameworks and concepts.

Chapter Three itemize the series of steps, methodologies and approach taken by the researcher in terms of answering all questions of the research to tally out with the objectives of the study. In the Chapter, the research shall provide ways of carrying out the objectives in reach describing the taxonomy and simulations used.

Chapter Four describes the full deployment and techniques of achieving the first objective. This answers the question of proposing the Path Redundancy Elimination (*PRE*) model and also reviewed the popular theories compared in the study.

Chapter Five presents the study in cache manageability and the proposed ProXcache strategy. The models and theories of proxy-cache strategy; the result of simulation and topology variations are covered in the chapter.

Chapter Six covers the results of the simulation in wider perspectives and varying parameter to measure the effectiveness of the study on different network topologies.

Chapter Seven concludes the study by covering and suggesting the future research directions along the contribution.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Internet was invented after series of needs and quest to provide better forms of information dissemination [56, 57, 6]. The advantages derived on the Internet have given rise to human inter-personal, socio-economic, academic and interactive benefits which cannot be dissociated from its usage. However, even with the immense efforts placed at providing the shared information, Internet is projected to face huge challenges of data volumes and traffic due to the quest by users to manage media-driven documents more [10, 58, 59, 60]. Host-centric traditional Internet is becoming heavily demanding by media data out numbering ordinary operational and textual information [10]. Therefore a need to adopt a shift of the Internet from the host-centric nature to a content-centric network seem of the essence.

The need to address the issue of information management globally posed a challenge such as: IP traffic will cause heavy blockages thus limiting addresses on the Internet platform by 2015 [2]. An estimated 64 Exabytes could be added in each month as the years pass on from 2014 [3, 33]. Usage of video and sharing will take over the mere sharing of files and lighter files [12]. It is predicted that by 2015, video messaging will be the most practice. Conflict of addresses shall be the order of the day as IP addresses faces limit sizes. Predictions by Cisco submitted the total number of mobile devices connected to the Internet will be more than the worlds' population [3, 1]. However, the current Internet platform is challenged with unbearable demands of multimedia to users if content are not well managed. Information retrieval and access cannot be fully guaranteed in expected time in the nearer future.

This chapter presents the literature review including relevant areas of research that have been carried out by other researchers. Section 2.2 presents the literature overview in higher details about ICN, thereby itemizing the leading research issues such as Naming, Routing, Name Resolution and Security. Section 2.3 presents various ICN projects and proposal architectures. Section 2.4 covers caching ICN and Section 2.5 discussed the cache managements. While Section 2.6 highlights the cache deployment strategies.

ICN as earlier mentioned in Chapter One will be covered in more details in this chapter. The network that yearn to replace and dissociate the location from the contents of data and information. ICN concept of Information-centrism aims at providing the advantages of neighbor node data and information delivery without necessarily connecting with location. ICN has some properties that constitutes its building block. This include the Content Store (CS), Pending Interest Table (PIT) and the Forwarding Information Base (FIB) as described in some of ICN approaches [4, 6, 22].

CS handles the information storage and data residential. When Interests are sent into the network, the CS is the first visitation point to provide the Interest. Chunks and whole data as Interests are look-up in the CS [61]. Caching contents therefore becomes practicable in the CS. PIT on the other hand is responsible for account and record taking. The PIT can be related to the Process Control Block (PCB) in operating system process [62]. Interests launched into the network are recorded into the table of PIT. In an event that the Interest has been served previously, the Interest is saved on a PIT as granted. However, if the Interest has never been served, the PIT keeps the Interest so that the same interest can be served and cached on the event that it passes the router in question. Additionally, FIB is the base that provides the mechanism holding information on forwarding the data/Interests. This operation is similar to the routing

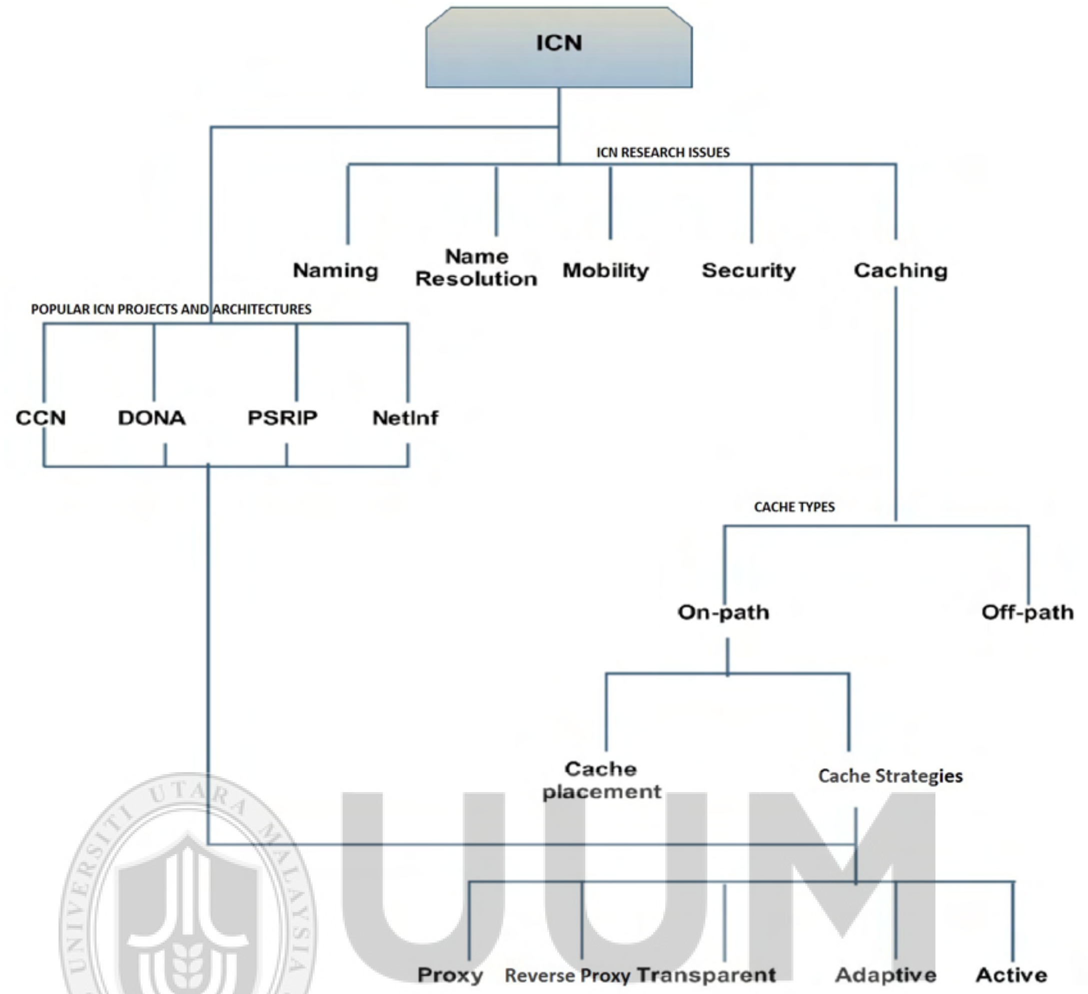

Figure 2.1. Research Scope

operation in the host-centric networking. The advantages of dissociating location and data cannot be over emphasized [4, 6, 21, 63]. Furthermore, fairly to note is that all ICN approaches depend on caching with each differing in practice and techniques. Caching can be said to be the idea or medium of storing data or chunk of data in a location for future referral based on placement, replacement and eviction scheme. An illustration of the relationship in the literature review is presented in Figure 2.1.

2.2 Information-Centric Networking

This section highlights some issues in ICN design and architectures; the issues explain the principles in naming, name resolution, mobility and security. These listed issues would be better achieved when an optional cache description and form is in place. However, the caching in ICN now becomes so relevant and important as to aid the implementation of the future Internet. ICN concepts generally has predicted the step by step manifestation of the current Internet by its wide advantage of enabling the closest nodes specifying and granting request that are similar when interests are sent. The earlier mentioned issues as explained in [6, 64, 65] can be depicted and illustrated on the Figure 2.2 .

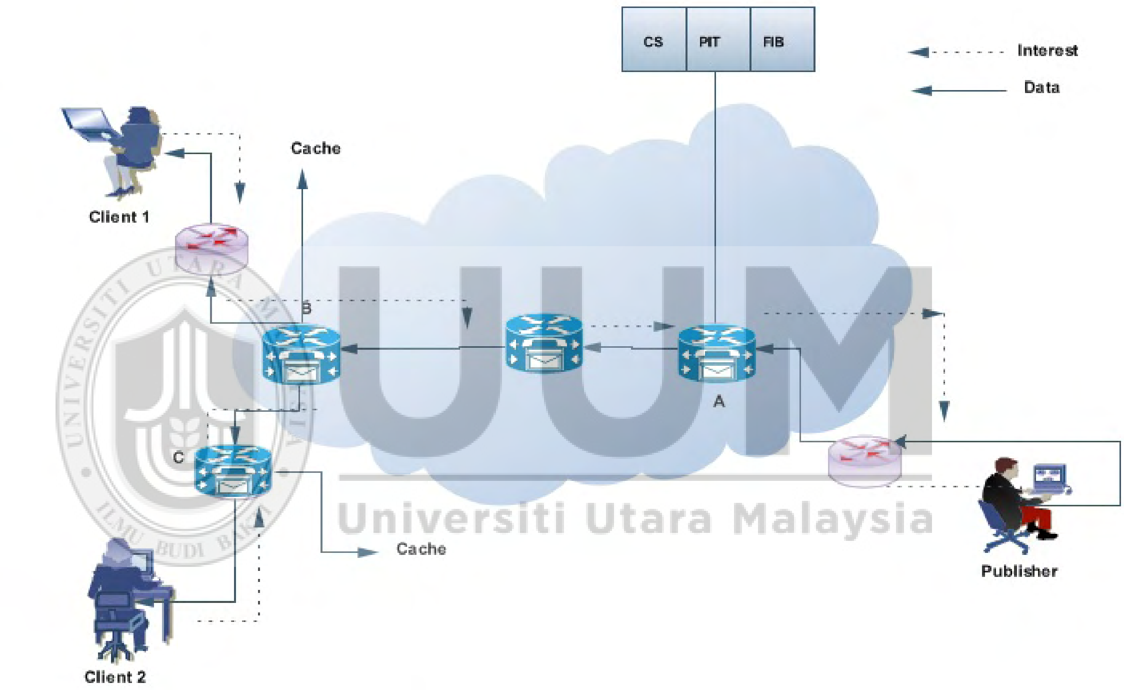

Figure 2.2. ICN Concept

However, the advantages through caching are that any time a request is posed by Client 2 (see Figure 2.2), for a similar interest, the closer nodes that has previously initiated such request offer the data requested for. This practice is only possible when a chunk of the data is cached in the CS of the former thus enhanced delivery of service along side lower upstream and downstream is also incurred [6, 14]. From the Figure 2.2, Client 1 initiates the request channeled through its closest router node B. The routers in ICN have incorporated the CS, FIB and the PIT as seen on router A. Router B cache the data obtained from router A and thus passes it on to router C when similar data is requested. To actualize the positive contribution of ICN deployment, one would notice that all information (content) needs to be called via interest by name, which only can be served after caching. Name resolution therefore becomes paramount because when interest is sent, there has to be resolution from the content store as in the case presented in Figure 2.2 above when Client 2 sent out its interest. Mobility [66, 67] and security also have gained lot of attention in the ICN research community. Other section describe some issues as it relates to effective caching mechanism to actualize the goals of ICN [6] as below:

- Naming
- Name Resolution
- Mobility
- Security
- Caching

2.2.1 Naming

Naming in ICN has been presented as the idea of trying to answer the questions related to the huge amount of information on the traditional Internet . This is due to the fact that in ICN, the clients and nodes are less concerned about the host location address but rather the request of their interest [6]. ICN as a proposed leaf of the current Internet will have to borrow some ideas of information retrieval and naming from the traditional Internet Protocol (IP) concept. Naming now becomes a vital point of the underlying design of ICN. Most practices have adopted the hierarchical naming and Distributed Hash Tables (DHTs) schemes as a form of identification as suggested in [6, 67, 68, 69]. Naming is thus seen as the way the requester (Client 1 and 2) sends interests. The information is then received by the closest node/router A or station (as depicted in Figure 2.3) which if it has in its Content Store (CS), it forwards through the function defined in the Forwarding Information Base (FIB) otherwise the interest is placed in a special table known as the Pending Interest Table (PIT) as presented in [6]. The motivation here is that the content is then cached alongside the interest for further request of the same interest. For the purpose of this research, we shall discuss the caching in remaining sections. An advantage also incurred in ICN is that information are not automatically deleted like in the IP forwarding, thus decreasing the upward streaming of bandwidth and making information fast available for any user that subsequently send an interest of the same information.

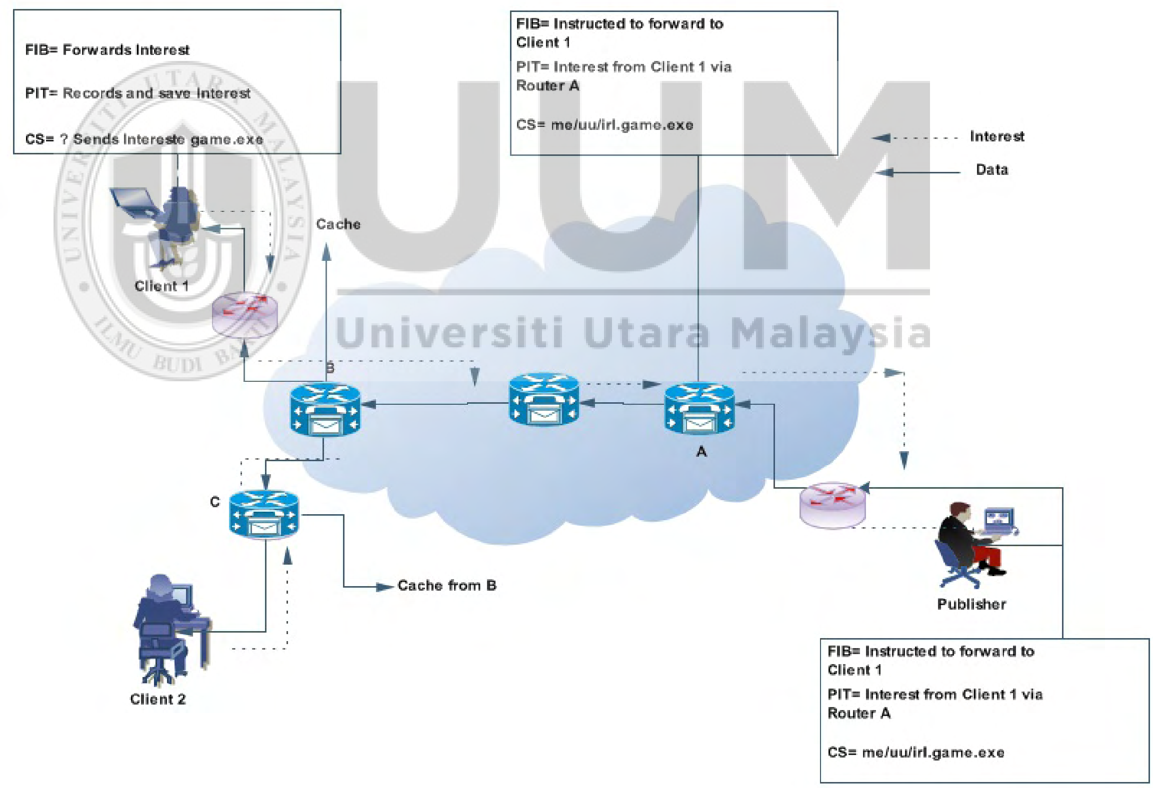

Figure 2.3. Naming in ICN

2.2.2 Name Resolution

Name Resolution action normally occurs at a router-like station for Data Oriented Network Architecture (DONA) [21], Network of Information (NetInf) [18], Scalable and Adaptive Internet Solutions (SAIL) [4, 15], 4WARD [7], Content Mediator Architecture for Content-Aware Networks (COMET) [10], CONVERGENCE and the rendezvous as in the case in Publisher-Subscriber Internet Routing Paradigm (PSIRP/PURSUIT) [36]. The rendezvous in PSIRP architecture work in a fashion of a negotiator and a settler by receiving the interest from the subscriber (Interest). When interests are posed, then a match is found through an opposite provider (Publisher) of the said interest and data. So many algorithms have since been proposed and discussed in as [6] and [7], which are yielding results in different ICN project and research implemented test [4, 38, 70].

Consequently unlike IP, the request and resolution could be settled by a node that has the information and also belong to a close name resolution station. These flexibility of the advantage of retrieving from multiple nodes span Name Resolution System (NRS) which add the advantage of high speed delivery and thus reducing the congestion and bottle neck of scheduling. A resolution of name will also prove tough without accurately providing a chunk memory handler at the NRS as cache. From Figure 2.4, a Subscriber (User) sends Interest into the network by registering the required interest with an NRS. The NRS checks its memory to match interests earlier published by a Publisher (Producer). If the name matches the requested interest, a forwarding action is then taken through referral. Since the content-centrism does not care about the location, subscriber is then provided with NRS authorities of delving into the network for the data supply. This makes ICN challenging in dealing with the large data expected either through subscription or publishing.

Name resolution then becomes the main section of ICN that handles how the routing will be achievable. Since the absence of routing table is felt, new medium of reactive, proactive and hybrid routing are introduced. In the reactive routing, a subscriber sends his interest as the initiator of the operation; thus making the operation possible and successful through caching. This interest is further routed to an appropriate router or node thereby getting a publisher or repository that could serve the request; therefore, adequate caching medium becomes so paramount and highly efficient. This research therefore further suggests ways to inculcate the caching part into the ICN architectures as mentioned in [4, 7, 15]. In the proactive routing technique, the subscriber sends interest into the network without necessary having the interest request. This is more or less like broadcasting what the Publishers advertise. The name is then cached and copied into locations for further forwarding when requests are made of similar object. In the hybrid, just as the name implies it combines the both approaches.

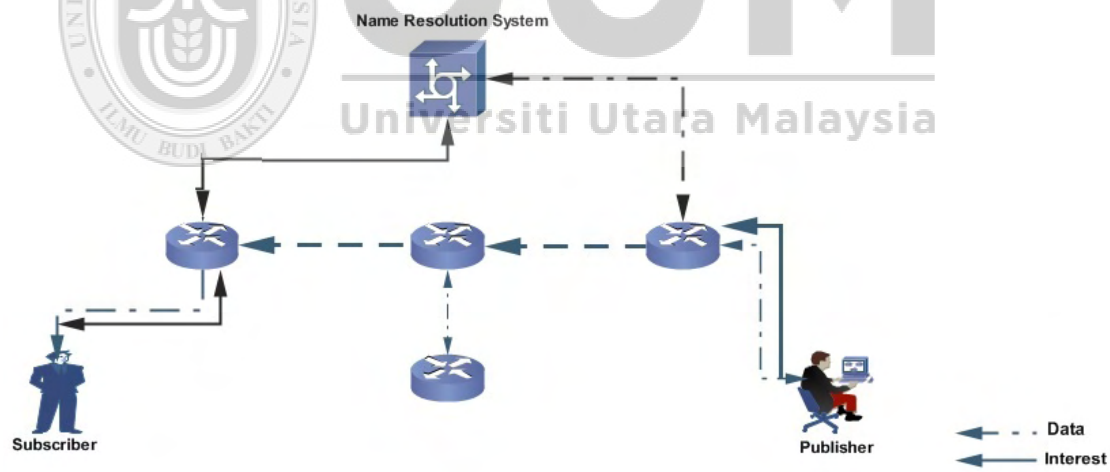

Figure 2.4. Name Resolution

2.2.3 Mobility in ICN

ICN solidly claim the mobility of the host and Publisher gives an edge over the usual IP networks [67, 71]. Advantages are that when a mobile node moves in and out of the network, there is always a node that has a copy (cache) of the information therefore reducing disruption and providing ideas on scalability [4]. The concern remains that when mobility in fully in practice; How genuine can a replica station be? Could the content be trusted? These questions are still not clearly answered by research to gain authority or the best frame work to tackle this issue. Mobility also will reduce the cases of information monopoly, censorship and reduce the privacy right. One may argue that of what benefit the ICN mobility would add to usability if the questions of content reliability are not fully answered? In a way to tackle such assertion, ICN provide high advantage for disaster areas that cannot normally retrieve information from its original home (i.e. multihoming is an advantage [4]).

2.2.4 Security in ICN

Information on the Internet needs to be trusted thereby provision of adequate security measures are issues in ICN. Looking at the Internet from this perspective would create the gap between why do we have firewalls, and other security features like encryption? As such, ICN will also not be excluded in having to provide high mechanism to control the menace of misuse and unguided information. A major advantage ICN yearn to have in terms of security would be the fact that no information is given unless asked; this to a certain extent could reduce the trouble of bad data transmission and a good way to have a strong security holding responsibility of the node or publisher for bad data publication [7, 25, 72]. However, since ICN has lots of forwarding nodes, an intruder could as well be a forwarding node for some time and decide to be a nuisance at the next iteration. These problems suggest open and holistic forms of checking and prevention against such. However, other effective means of data security has been established such as key distribution, encryption, digital signing, stale timing, binding, designed strong key hash computations using trees etc. [4, 7, 73].

2.2.5 Caching in ICN

With the immense increase in the information shared on the Internet, a motivational idea behind ICN paradigm is its property for a neighbor node to supply request without referring to the main host [74, 75, 76, 77, 52] . Caching has therefore become the underlining techniques that presents the benefits of reduced bandwidth network usage, reduction in user-perceived delays, and monopoly of overloading the main server of information [78]. Caching can therefore be seen in ICN as the linchpin of the content centrism network. Several ICN caching issues has been studied ranging from the careful selection on the data to place in placement, eviction, positioning and not caching [39, 40, 42, 74, 76, 77, 79, 80, 81, 82]. Examples include, Collaborative Cache, Most Popular Cache, Cooperative Cache, Assign Cache, MAGIC (See 2.2).

2.3 ICN Architectures

In this research study, some popular ICN architectures [4, 9, 18, 11] are presented alongside their different features. The advantages of scalability, routing, naming and data retrieval cannot be maximumly attained without a better caching option in place. However, for the Publish Subscribe Information Routing Paradigm (PSIRP) architecture, a server-based known as autonomous systems (AS) included in the design needs a chunk or copies of the said interest and also the data (requested) before other functions of forwarding, and granting request would be possible. This section therefore, presents the various popular ICN architectures and highlights their distinct functionality and other areas as related to caching. Every node in ICN holds a chunk of information through caching and this serves as part of the basis of selecting these popular ICN architectures [7].

2.3.1 Content Centric Networking (CCN)

CCN [6] is a popular and most easily understandable ICN architecture. It operations is in the form of client making request into the network without necessary specifying the address of who gives what information and how? This is solved by routing the request/interest sent by Client 1 (see Figure 2.5) through a router/node that communicates with its neighbors (client 2 and 3) in search of an answer to that request. Subsequently, if the requested interest is identified at the first router, it is then routed back to the requester; see Figure 2.5. A good point to mention is that, in CCN operation, all forwarding nodes cache the data being asked in their memory for future related request for on-path caching while the off-path does the caching at the main repository. Unlike the IP networks, ICN does not need the Uniform Resource Locator (URL) to locate the information, nor the file structure to name the file and the IP to name the interface of the storage node [4]. Independent object identifiers are used to name the information objects [9, 12, 83].

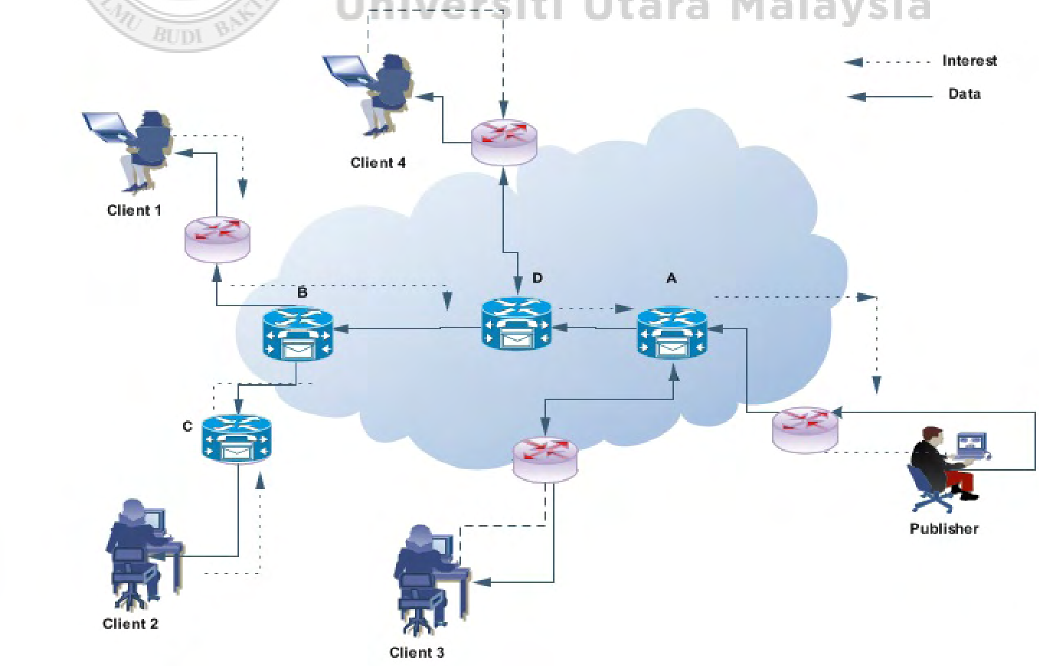

Figure 2.5. Content Centric Networking (CCN)

2.3.2 Data Oriented Networking Architecture (DONA)

Data Oriented Network Architecture (DONA) [21] also a popular ICN driven platform uses the flat names instead of the hierarchical naming [20, 22, 84, 85]. This architecture could be viewed as an architecture with added advantages which include the ability to cache information at the network layer using the Autonomous System (AS) introduced in its design. The inclusion of the AS allows information to be easily retrievable even when the nodes are not available (mobility) since the information is cached in a node added to the AS. Its strict cryptography ability enables users to authenticate the genuine state of their request. However, this architecture uses IP and routing address which are exhibited locally and globally [7]. The inclusions of unique servers known as Resolution Handlers (RH) are charged with resolving names as operations unfold on the network as shown in Figure 2.6. RH and routers are enclosed into the AS thereby functioning according to the routing policies enforced on the network. DONA operation functions in a fashion similar to the above discussed CCN with its major different in operation via sending a REGISTER message alongside its object to an in-network RH. The RH works in a caching fashion and stores a pointer so as to enable easy re-broadcast to other tiers housing RHs. This operation of caching continues as mapping takes place in each RH in the AS, subsequently, a Publisher node also issues REGISTER on to the network informing all RH hierarchy that the request can be provided. DONA engages in its operation using two approaches, namely the coupled and de-coupled operational approach. DONA caching remains most interesting as further discussion on its caching will be covered in the subsequent sections.

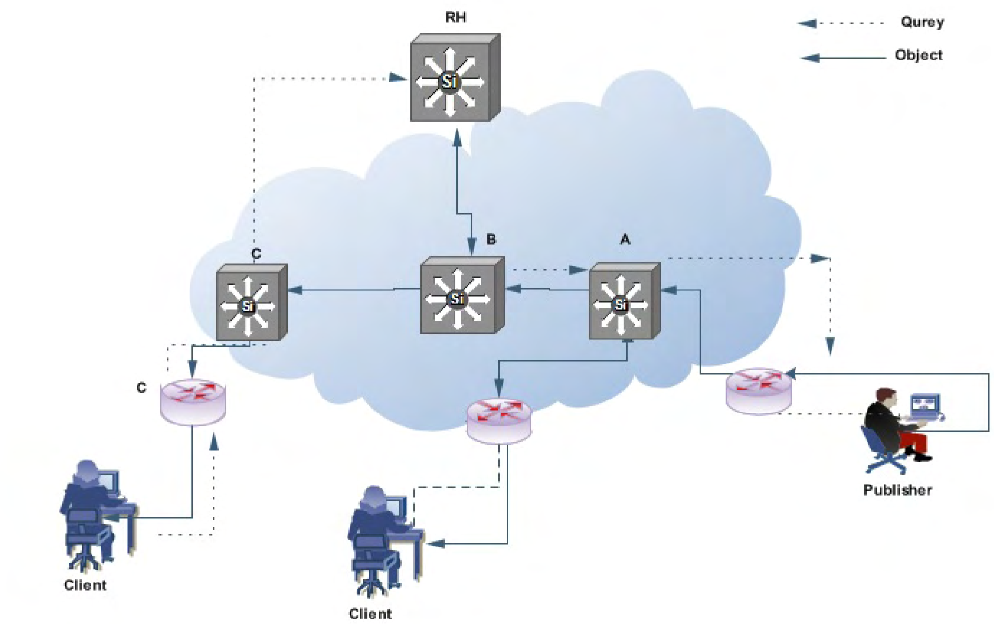

Figure 2.6. Data Oriented Networking Architecture (DONA)

2.3.3 Publish-Subscribe Internet Routing Paradigm (PSIRP)

Publish-Subscribe Internet Routing Paradigm [36, 37] is also worthy to be mentioned among the architecture of ICN. In the PSIRP, information objects are retrieved by subscribers after the information objects are fully published into the network by the producing node. A new matcher and negotiator known as rendezvous system handle the matching. PSIRP initiates its operation by the subscriber sending its interest/request to a rendezvous at the client side (see Figure 2.7). The operation is then followed by a subscription through the closest router station. A follow-up of an opposite operation is usually performed by an origin node of the object requested for. It follows similar fashion of operation by registering its data/ information to its rendezvous in the Publisher closed area. Here the server-based rendezvous thus, resolves with the client registered rendezvous at the publishers side to obtain a match. When these operations are done, the data can then be routed through the inner router using the same registered

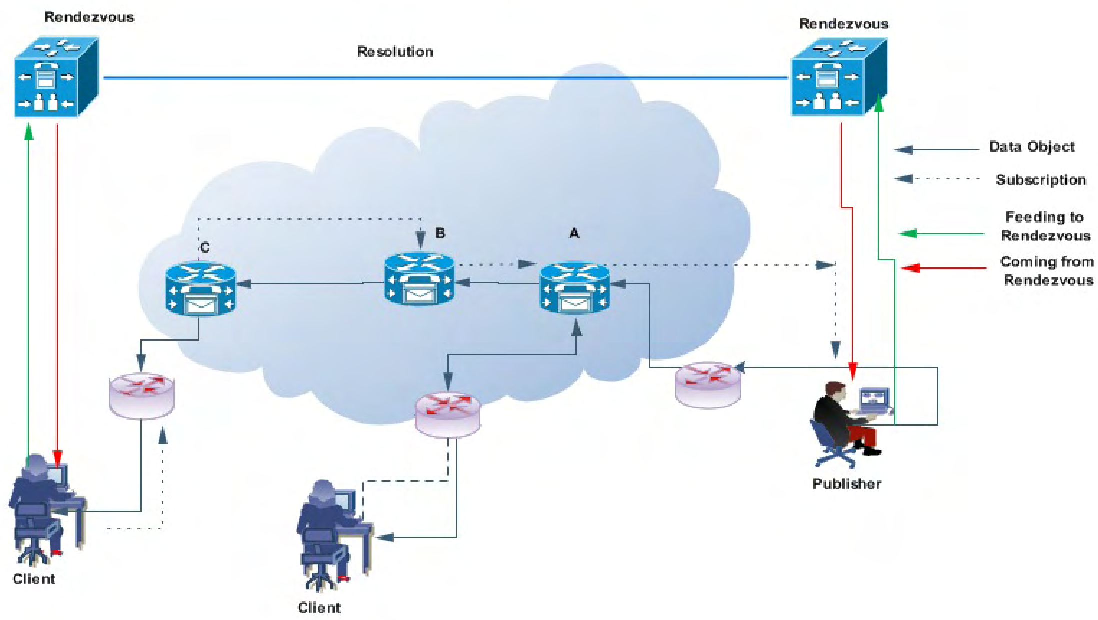

Figure 2.7. Publish-Subscribe Internet Routing Paradigm (PSIRP)

path to supply the request to the client. The PSIRP is also known as a continuation research work by the EU Framework 7 as Publish Subscribe Internet Technology (PURSUIT). Naming in this architecture is done uniquely by some statistical identities in pair, PURSUIT share the same naming structure as DONA (flat names) but managed by the rendezvous and further routed by a Distribution Harsh Table [85, 45, 86]. For the benefit of this research, rendezvous is the concern as it relates to how caching is done on PSIRP/PURSUIT [70].

2.3.4 Network of Information (NetInf)

The NetInf [18, 38, 79] architecture on the other hand works in a Publisher/Subscriber way even though it has a major difference in terms of operation based on the techniques of NRS as the guide. Subscriber gets their information through a look ups initiated by the Publishers as Named Data Objects (NDO) which are incorporated as some key search aiding attributes as seen in Figure 2.8. The look ups are started by the consumer engaging in the closest node (proxy) search before expanding if data is not found. Resolutions are less stretchful as Multilevel Distributed Hash Tables (MDHT) are used [87]. Concurrently, every named object has a self-identifier to a location. NetInf possess issues of mobility, scalability and naming as seen in all previously mentioned ICN architectures. However, it has the ability to cache on-path even though requester nodes make iterations during the look-up until a match is served from the NR service. NetInf architecture can be said to be intelligent as the consumer nodes are always provided with the requested-prone list using the in-direction send and receive through the use of the MDHT [67].

For the benefit of our quest to examine the caching techniques in the above selected ICN architecture, some techniques in caching shall be proposed thereby predicting the possibilities of some borrowed caching algorithms and data flow concepts to attain better ICN performance along caching.

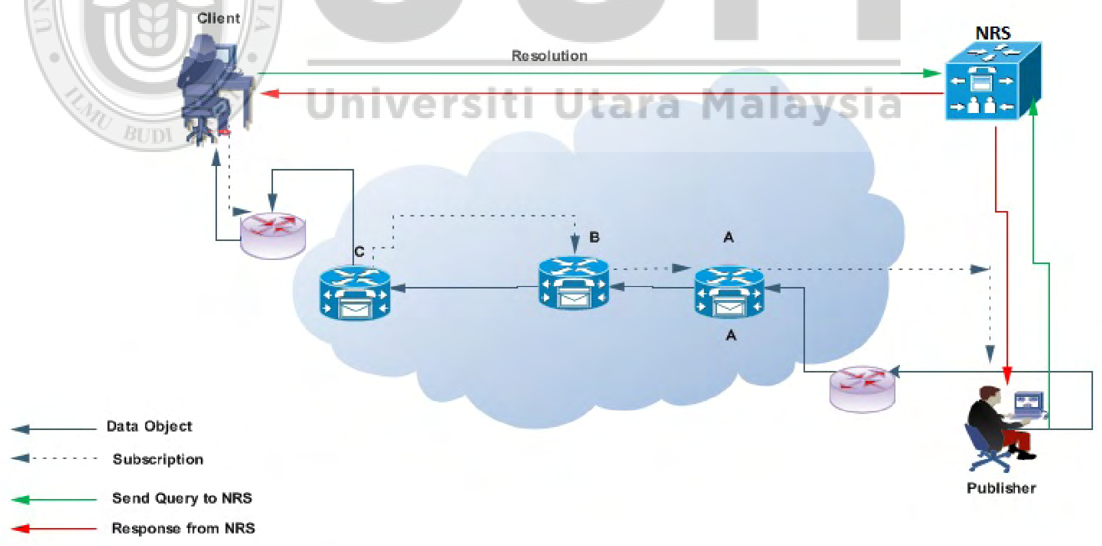

Figure 2.8. Network of Information (NetInf)

2.4 Caching in ICN

Caching can be defined as a medium of having information, data and object temporarily saved in a location for predictive usage on frequent or closely related interval. Content caching as it relates to ICN is our concern in this work; caching therefore will help in reducing the high cost in up streaming and down streaming of data, information and interest in ICN [44, 74, 88]. Previously discussed ICN architectures (CCN, DONA, PSIRP and NetInf) itemized the various forms of caching [13, 24, 45, 46, 61, 89, 90, 54, 91]; thereby identifying the on-path or in-path and the off-path or node caching as the two distinct forms of ICN caching. Various studies such as in [49, 92, 50] and [78] suggested ways to fulfill the ICN dream of being promising and advantageous through adequate cache deployment. However, caching is still a common issue in all aforementioned ICN architecture. Some major advantages of caching as mentioned earlier will reduce traffic, redundancy of information, bottle neck queuing and competitiveness on frequently accessed or visited domains [16, 93, 94, 95, 96]. Some benefits enjoyed when cache is in place will include good utilization of bandwidth, thereby reducing information wastes and high garbage collection (misuse of memory). Bottleneck queue and congestion, improved time of response, bandwidth over heading will serve efficient also [47, 78, 50].

However, to benefit from the advantages of caching, one may wonder when and how can caching be done and why? A study in [91] however, suggested some ways of collaborative approach to propose some additions of caching into the ICN architecture. The question still remains very challenging particularly as it relates to ICN; caching will be treated in the context of ICN through types and forms of caching that can be borrowed from the traditional Internet in the subsequent sections.

Caching as vital as it sound could be classified based on deployment structures into the following viz: proxy, reverse proxy, transparent, adaptive, push and active [50, 23]. Each type differs in its operation through some sets of properties and characteristics. For the ICN, one will definitely need to be predictive on the option to implore and at the best or most suitable topology/architecture in question. ICN caching are further categorized according to their nature of position, location and deployment. These classes are the on-path cache and the off-path caching.

2.4.1 Off-Path Caching

In off-path caching as described and summarized in [7, 9], CCN/NDN, DONA, PSIRP, NetInf all have an the capability of the On-path caching using different forms. However, the Off-path caching in all the mentioned architectures requires additional processes registrations in DONA and PSIRP, additional routing information in NDN/CCN, NetInf has already got additional advantages through cache-aware transport facility and name Resolution System NRS [7, 9]. However, in a recent proposed intra-autonomous caching idea via cooperative mechanism in [94] argued that since caching in CCN is inefficient because of data redundancy, an autonomous system in cooperation needs to tackle this issue. This was submitted by the efforts of the neighboring node in AS charged to serve and take care of the request. This unarguably was demonstrated through the collaboration of all nodes in neighboring terms servicing each request concurrently as an AS. In accordance to the submission by [94, 95], AS serve as an off-path caching idea which differ from the on-path caching. In similar fashion, Wang *et al.,* in [97] in another study based on collaborative caching using hash-route included an extension in the content store called ingress link. The ideas was based on the issues of limited cache sizes by eliminating the problem of caching via referral. Their algorithm *AssignCache* was said to have better cache hits. The study in [63, 95], further concentrated on the issues of better performance of caching even though the major questions still exist as when would the off-path stations be more reliable? How legitimate can the information be?

2.4.2 On-Path Caching

In this approach, contents are basically structured and cached according to some practices based on history of contents and frequency which are often referred to as recency and frequency in some literature [44, 48, 78, 65, 63, 98, 99]. In a recent study conducted in [48], in-network caching was used in two distinct forms namely: Coordinated and non-coordinated with most interest in performance and coordination costs as the metrics. One will agree that just like in other computer paradigms, a trade-off is always reached to sub-do a factor to enhance performance or vise varsa. Coordinated in-network cache according to [48], described how objects are cautiously picked with the intention of avoiding duplicates in the store; this practice definitely improve the latency and thus makes neighboring node hold non-similar content thereby improving subsequent interest on the path without the need to access the source. In actualizing the above coordinated cache, load on the origin (source) was reduced, routing hop count was minimized, and storage coordination was absent. The load on origin and hop counting was used as the cadence for network performance while storage coordination was used to measure the capability of storage cost in the network. One will wonder how the coordinated caching could be performed to achieve the objectives mentioned [42, 74, 100]. Mechanism is provided by the designer to check and make sure the nodes don't save same objects using the compare content algorithm [52, 101]. The task therefore inherit some cost of scanning a wide range of the topology to check and share the objects. Whereas in non-coordinated in-network cache, this cost of spanning around the topology becomes absent as the need to safely guide and protect against duplicates is not of essence as described in the former. Non-coordinated cache can thus be concluded to have a higher hop counts when the objects subscribed need to be fetched from the main source (Publisher). In submission we can therefore conclude to say coordinated cache as described in [48] and [24] earns more comfort to the end subscriber. However, a focus towards cache deployment strategy shall overcome the problems of content and path redundancy faced in coordinating the cache. Better deployment will improve time delivery, lesser bandwidth usage and above all lower hop counting. This research shall therefore adopt the on-path caching and improve content look-ups through autonomous sites to mitigate some of the challenges faced in ICN caching.

2.5 Cache Management Strategies

ICN cache management has been the most active area of content dissemination for better scalability, security and fast reach-ability. Several cache managements have been proposed while a good number of the managements are borrowed from the current Internet caching [102, 103, 104]. With the cache management strategies in ICN still in its infancy through development and research. Cache management strategies describes the form of cache operations as related to the position, the size, time and manner in which content chunks are placed, replaced and evicted in ICN nodes. Below are some of the cache management strategies for ICN.

2.5.1 Leave Copy Everywhere (LCE)

Cache management in LCE [23] is defined by it operation of caching chunks or data in every node crossed. Part of the practice of the ICN is the ability to make information readily and easily accessible as describe in ICN initial proposal [6]. As a user sends out a request using LCE, the nature of the network serves the interest using hierarchical search and ordering of nodes to acquire a cache-hit. Given a network with a hierarchy ordered nodes n_1, n_2, ..., n_m, once a request for a data is sent in the order 1,..., m (from the subscriber to Publisher R5-R1) and the data gets a hit (cache-hit) at a node n_i, the copy of the data (cache) is cached at all immediate nodes on the route of content delivery in the order: $\{n_{i-1}, n_{(i+1)-1}, ..., n_{(1-m)}\}$. This as a result of having to save a copy of the content in all single leveled node causing the high *content redundancy* (see Figure 2.9).

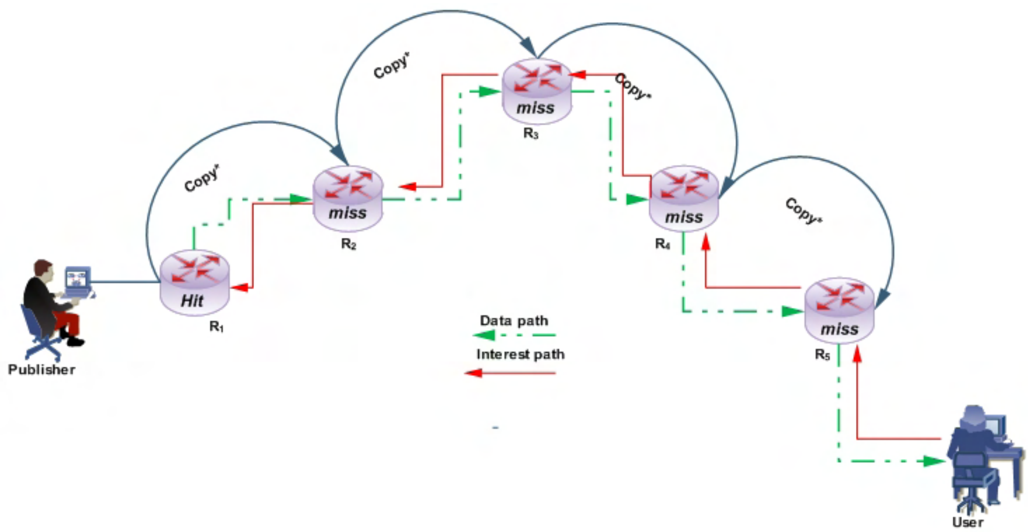

Figure 2.9. Mechanism of Leave Copy Everywhere

2.5.2 Probabilistic Cache (ProbCache)

The Probcache [23, 24] is assumed as a probabilistic LCE in a randomized manner. Its main distinction to the LCE is the ability to keep its cache copy in an immediate neighbor node (n) with probability *p*. Probcache therefore does not keep a cached copy in a node with *q (1-p)* probability. Once a cache is hit, a probability computation is acted on depending on the chosen distribution from the subscriber to Publisher R_5- R_1 (see Figure 2.10).

2.5.3 Leave Copy Down (LCD)

The Leave Copy Down (LCD) [23, 24] is a cache management strategy that defines the form and manner of content caching on nodes. Its operation works in a similar fashion as the popular "drop at the first neighbor" process. In LCD once a request for an interest is posed, a path link is created in the form of $\{n_1, n_2, ..., n_m\}$ to the Publisher of the data or the content holding node. As the content records a hit on

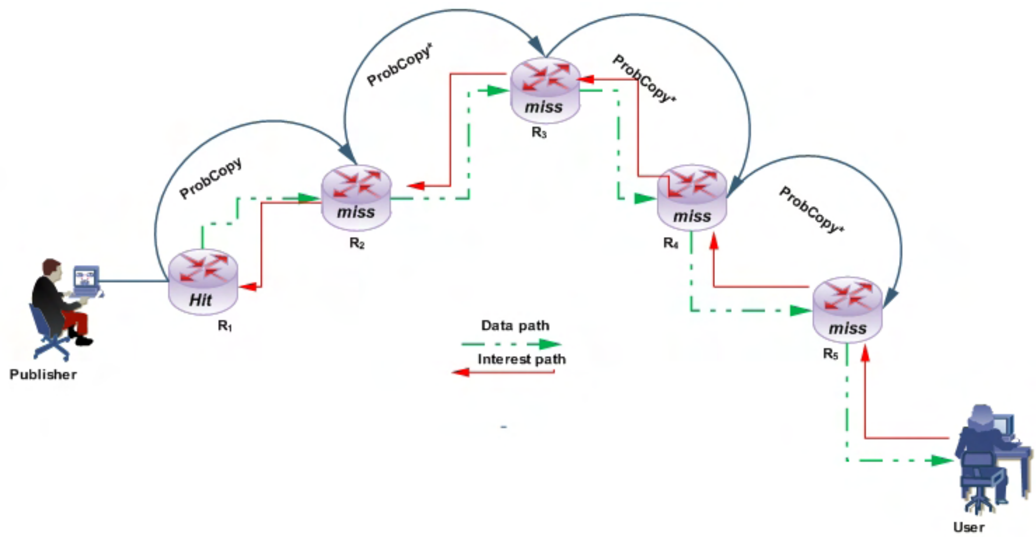

Figure 2.10. Mechanism of ProbCache

R_1, it only stores the copy in its direct neighbor node using the traversing path to the subscriber R_2. For LCD to gain the high hits of availing contents in almost all nodes, it needs a content-hit at almost all levels of $\{n_1, n_2, ..., n_{m)}\}$ (see Figure 2.11).

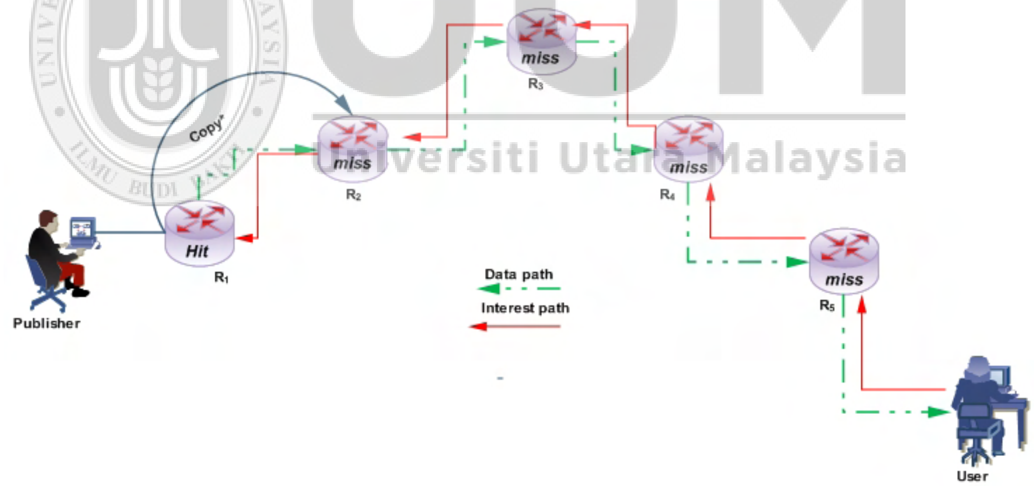

Figure 2.11. Mechanism of Leave Copy Down

2.5.4 Move Copy Down (MCD)

Move Copy Down (MCD) defines an extension of the LCD through it caching algorithm. However, its big difference in its form of operation is down streaming ability to the cached content copy delivery by deleting a local copy (from $R_1$and copy cached in R_2). In MCD, its hierarchical ordering can be thought as the multilevel queue in operating systems [62]. In MCD, once a node-hit is matched, the copy is moved down in the hierarchical route manner to the available slot and the local cache node R_1 (n_{i-1}) copy is deleted. The idea of the copy deletion is to mitigate flooding and keeping the replicas to improve the diversity rate on the path of traversing as depicted in Figure 2.12.

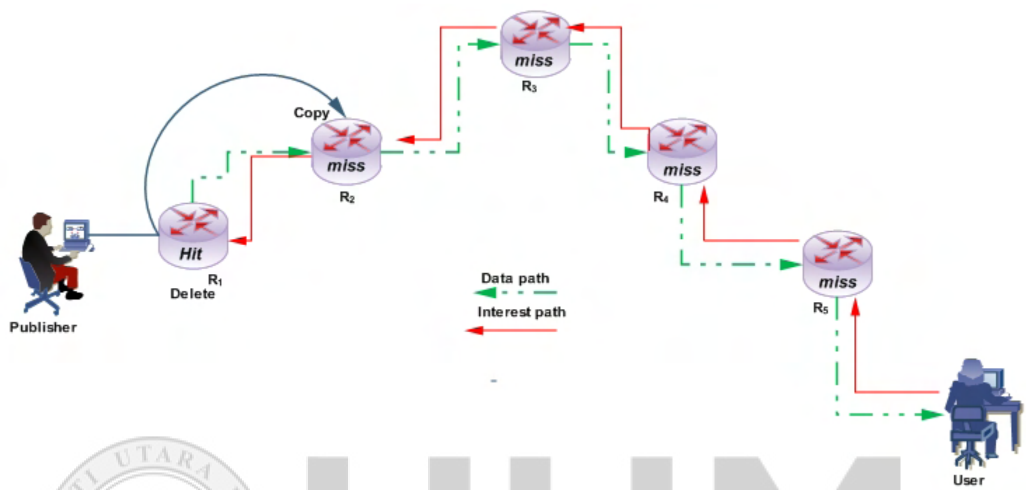

Figure 2.12. Mechanism of Move Copy Down

2.6 Cache Deployment Strategies

Cache deployment and positioning [23, 50, 26] has raised quite a number of interest in the ICN research community. The need to adequately place a cache-enabled node or router could influence the effectiveness of the network. Placing a cache-enabled router increases the hit rate depending on the used algorithm. Studies in [51] presented few cache deployment strategies for P2P networks ISP. Deployments of cache-enable routers and nodes improves the overall performance of the network. The positioning of the cache-enabled routers or nodes predict high cache and content hit ratio. According to [50], the position aid in better hop crossing and time efficiency of the data. Caching as vital as it sound could be classified based on deployment structures into the following viz: proxy, reverse proxy, transparent, adaptive, push and active [50]. Each type differs in its operation through some sets of properties and characteristics. For the ICN, one will definitely need to be predictive on the option to implore and the most suitable topology/architecture in question.

2.6.1 Proxy Deployment Strategy

As the name implies, this deployment architecture serves as an aid in the middle. It works is based on the principle of a node/router as seen in Figure 2.13 which keeps a content on the path before accessing the main Publisher [53]. In that way, response time is enhanced and bandwidth is drastically minimized since the re-interest of the same data set will not require going back to the Publisher. On-path as it may seem in this architecture means, if cache content is not found, then a forward operation to the server/publisher is then initiated. Great benefits in this architecture places the cache proxy closer to the client nodes but could suffer in terms of updated contents and future data corruption when altered [51].

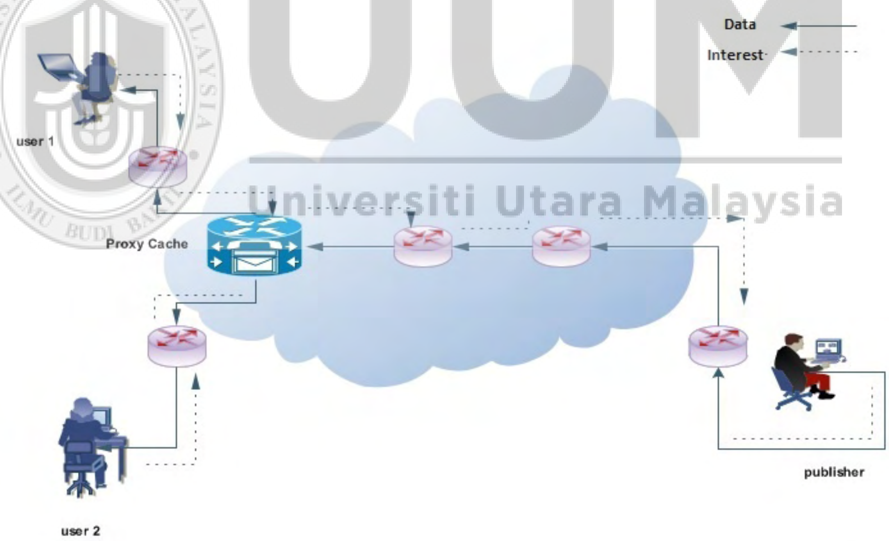

Figure 2.13. Proxy Cache Deployment Strategy

2.6.2 Reverse Proxy Deployment Strategy

Cache center is positioned closer to the Publisher in reverse proxy as against the mode in proxy. For this architecture, we hope and predict its efficacy and better use to reduce the access of malicious and suspected harmful information when employed on a ICN architecture. It has the advantage of making the content almost fully autonomous to the client as the cache center is placed closer to the Publisher against the previously mentioned ordinary proxy cache. As mentioned earlier, the advantage could prevent a client forwarding a malicious content to any neighbor that request same data. However, if the Publisher fails the chances of keeping a smooth publish request could be grossly reduced (see Figure 2.14).

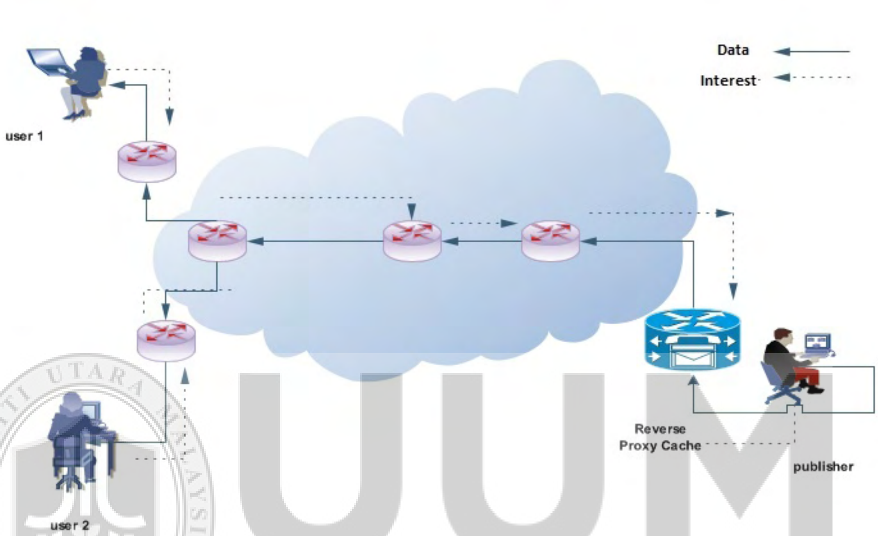

Figure 2.14. Reverse Proxy Deployment Strategy

2.6.3 Transparent Deployment Strategy

This offers architecture similar to the entire concept of ICN. It performs caching based on a cache supervisory system that hold the initial content cached, thereby reducing and balancing loads through sub routed caches connected. It improves the caching by reducing overall work force to other routed caches using the cache-to-cache connection. Advantages could be seen that the caching arrangement is similar to the Shared Memory Architecture (SMA) in computer architecture and organization. Disadvantage could sometimes be recorded when a node cache attached to a parent cannot re-fetch from the directly routed. As shown on Figure 2.15, a cache coordinator directs via instruction the router node that caches contents. Part of its drawbacks as described by Li *et al.* in [48] is the high cost incurred during coordination.

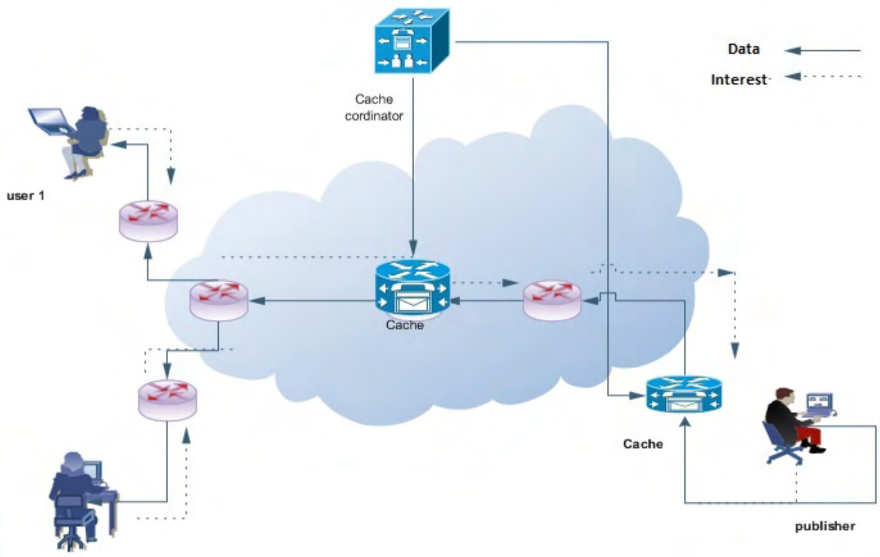

Figure 2.15. Transparent cache Deployment Strategy

2.6.4 Adaptive Deployment Strategy

An adaptive cache architectural option works in a way similar to a mesh network topology. The cache information is sent to groups which are created overlapping each other with the intention of serving highly requested data or interest. The intelligence in the adaptive comes to play as it re-organizes the data into types and forming groups. Thereby, reducing high demand from the main source or Publisher. This could be related to how the Autonomous System in PSIRP and rendezvous network (RENE) systems work in the ICN. Figure 2.16 depicts the operation of the active cache deployment. From the Figure 2.17, each coordinator belongs to a set of group. When interests are served, the coordinator maintains the fair cache enability of the nodes on the network. Each treated as individual as possible.

2.6.5 Active Deployment Strategy

In active, request or interest will be dynamically cached thereby allowing also for personalization of the said request. Users often prefer their request to be personalized

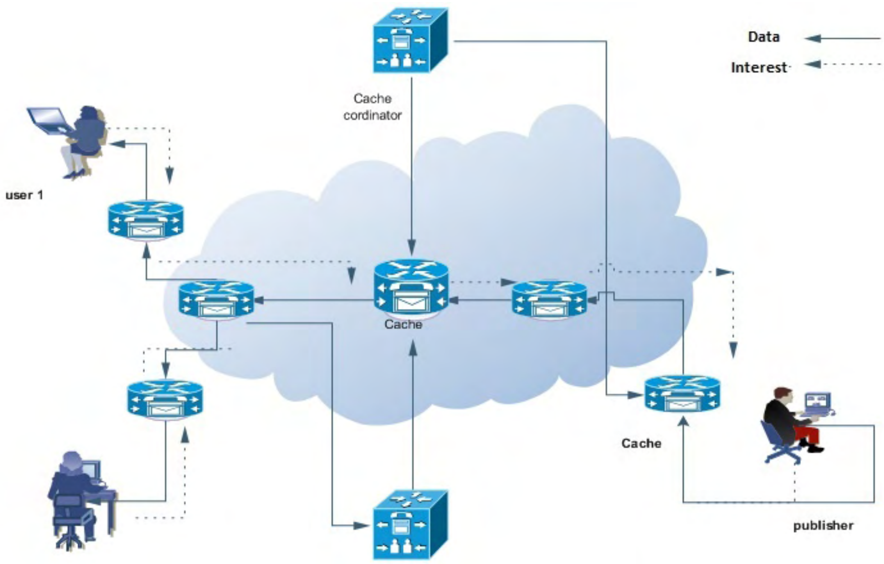

Figure 2.16. Adaptive Deployment Strategy

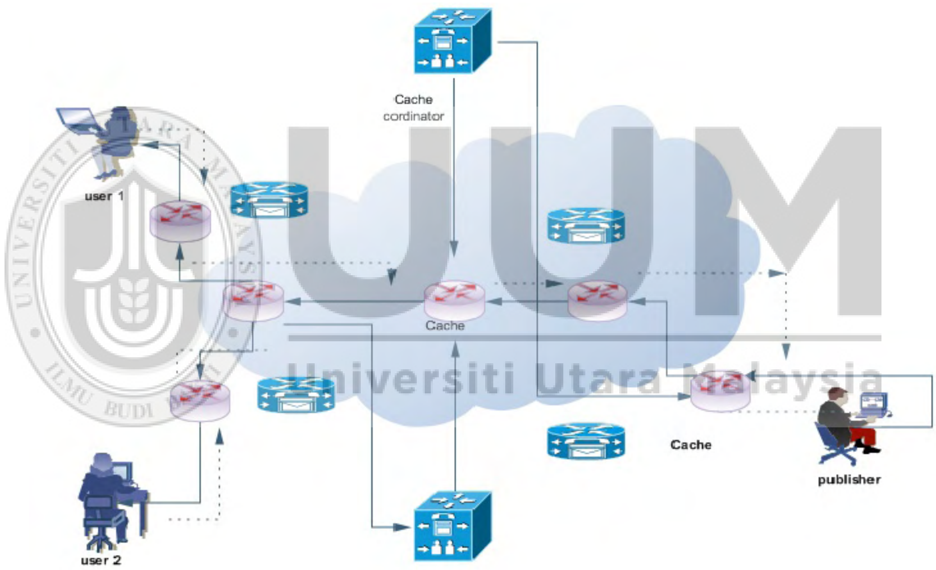

Figure 2.17. Active Deployment Strategy

therefore an intelligent section of such codes needs to be added to the caching component to perform the task of personalization. In this type, the cache gives the flexibility of the server to move its request to a closer proxy. This shares the proxy to the user at the same time sharing the proximity to the server (See Figure 2.17). The coordinator nodes makes the cache routers available to the nodes bearing the Subscriber and the server. Coordinating cost posed a major challenge to this deployment strategy when adopted in ICN.

2.7 Cache Deployment in ICN Architectures

Cache deployment are mediums of cache positioning on and off-path network. The deployment is grossly effective in position and thereby influence the route of information sharing. The closer the proxy, the more dependent the rate of cache hits. ICN architectures have used some of the cache deployments strategies in place. Among the popular ICN studied, are CCN, DONA, NetInf, and PSIRP.

2.7.1 Cache Deployment in CCN

In CCN, since every router/node on-path has the facility of caching chunk [6], it becomes easier to meet due dates. Thus, feedback seems to be faster with lesser delay. Contrary to CCN deployment advantage of fast delivery, the recent work by [24, 47] has submitted the weakness and drawbacks of using this strategy. Some obvious problems are the challenges pointed out in this study in Chapter One: Path redundancy and content redundancy. When all on-path routers are cache-enable, the maximum distance interests and data needs to cross becomes an issue. The utilization of bandwidth and power results in wastage as compared to when selective cache-enable routers are deployed.

2.7.2 Cache Deployment in DONA

In DONA, resolution handlers (RH) serves as coordinators of the interests and data. RH serves as the universal cache proxies and deployment aides. The probable challenge in DONA project can be envision to suffer from over flooding of interests and data, single point of failure and crowded requests. This can be seen as one of the leading challenging issues submitted by Li *et al.* in [48] as coordination cost. Li *et al.*, suggested ways to mitigate the cost and content redundancy by employing the popularity based caching concept.

2.7.3 Cache Deployment in PSIRP

The strategies of deployment in PSIRP can be said to adopt the transparent form of strategy. The special feature of Rendezvous act as the supervisory cache coordinator. However, the supervisor rendezvous network (RENE) upon gaining requests, it populates the results by specifying the route for content and data delivery [45, 85, 86]. A predicted problem faced by PSIRP is the coordination cost overhead, the path calculation in terms of forwarding and the content redundancy.

2.7.4 Cache Deployment in NetInf

Deployment strategy used in NetInf is based on proxy. A name resolution system (NRS) is dedicated to resolve queries sent out by the Subscribers. Convincingly, according to [18, 38], multilevel distribution hash table (MDHT) [67] are used in NetInf. Once a data object is obtain, a cache-all action is taken by all routers. NetInf cache deployment strategy also suffers from the content redundancy and path redundancy presented by Psaras *et al.*[24, 47].

2.8 Hypergraph

The study has extensibly scout for network modelling approaches, among which are Bloom filter prospects and Bee-Colony optimization algorithms to achieve the setout objectives in Chapter One. Hypergraph concept selection in this study was appropriate due to its advantages of clear node relationship formulation using the hyper arcs and edges. Several studies used the concept to model and achieve the network relationship and assertain traffic dependencies between network devices [105, 106, 55, 107].

2.8.1 Properties of Hypergraph

Hypergraph salient features justifies the potentials and strength of using hypergraph in ICN. Among the properties are:

1. Isomorphism: This defines the hypergraph properties of a graph H such that the vertices and edges of another graph G implies the same permutation results. Thus, ICN graph nodes are used in the study as cacheable devices and routers.
2. Equality: This defines a property in structural relationship of vertces and edges. A graph H is said to equivalent to G, and writes H equiv., G if the condition of isomorphism such that $H \equiv G$.
3. Symmetrism: It defines the degree of edge interconnection also refered to as degree. In ICN, this is refered to as the node relation of the highest outgoing interface.
4. Transversal properties: A hypergraph related to the network interrelation according to the defination by [55, 106] has nonempty node and edge relationship. Thus, this made hypergraph an adequate model for ICN cache manageability.
5. Pertitioning: In network science, the partitioning properties of hypergraph make it posible for cluster, set membership and autonomous partitioning.

2.8.2 Advantages of Hypergraph

Hypergraph advantages have been widely studied in different reseach domains of applied and natural science. The areas of its applicability include: System modelling and Engineering for Algebraic theory, concept of duality, complex and real analysis [105]. The advantages of combinatorics and relationships in modellings are used in hypergraph. In organic science and chemical engineering, hypergraph relationships are used to represent the molecular inclusion of atoms and hyperedges. In telecommunication models, vertices are used as cells and hyperedges form the edges.

2.8.3 Hypergraph Relationship and Size

Hypergraph is a graph method that describe the relationship between 2 entities, Vertices and Edges [105, 106, 55, 107]. As described by Gallo *et al.* [55],

Let a network be represented as an hypergraph :$H = (V, E)$, there exist the vertices and edges

where $V = \{v_1, v_2, v_3, ..., v_n\}$ represents the vertices and Edges $E = \{E_1, E_2, E_3, ..., E_m\}$

.

Such that there exists a direct relationship between the vertices and edges (E and V) as

$E \subseteq V$ for $i = 1, 2, 3, ..., m$.

In hypergraph, a forwarding arc defines the continuous relationship between a node and a preceding node. Backwards arc defines a link that results to a leaf or edge node.

According to several researchers in the field of graph modeling, the size of the graph is defined as n and m for the standard graph [105, 106, 55]. Size in this study has great significance. As such the size is given as:

$$Size\,H = \sum_{E_1 \varepsilon E|E_i|} \tag{2.1}$$

Thus, this research propose to extend the use of Hypergraph in order to achieve the objectives submitted in Chapter One.

2.9 Related Work

Cache deployment strategies has long been proposed and analyzed in ICN based on different criteria. As presented on Table 2.2 (Cache Summaries), a recent study by Psaras *et al.,* [47] and Ren *et al.*, [54] produced some improved cache management results against the previous LCE which ICN was built upon but research issues of content over flooding and high traffic redundancy were inherent. However, an attempt to coordinate the caching in ICN was nicely presented and reduced delay when a neighbor node is selected [16] in Selective neighbor node caching. In the work by Vasilakos *et al.*,[16] the cache complexity of acquiring a neighbor node was seen as an issue in the design. A major trade-off in the study was against the delay and the cache cost. An optimal cache management was not achieved as coordination cost, delay in neighbor selection and mobility were the challenges in their work. Consideration for a mobility affected node was covered but lacks an ideal form of tackling a delayed Interest that could not be served to a subscriber. In the study, the importance of proxy was highlighted however, it was not achieved due to the complexity and cost of proactively caching in a neighbor node after mobility in and out of the network. The approach used the prediction, neighbor information and hop distances and listed proxy at closer nodes in ICN future research direction as related to different topologies.

In another study by Sourlas *et al.,* [108], Distributed cache management was used in cache deployment to mitigate server loads. The study focused mainly on autonomic intelligent caching scheme in ICN, with coordinated cache managers edging the positions of nodes and routers in the network. The attempt to mitigate the server overload was hit with the research challenges of configuring the algorithm almost at intervals. The scheme in the study is an example of the transparent cache deployment strategy explained in Section 2.6 where supervisory caching is done. Communication and computational complexity were the drawbacks in the research as caches are always reconfigured and thereby adding to the delays of content delivery. Part of the weaknesses recorded was the complexity in handling real-time operations, volatility and popularity of contents and the need for the cache supervisors to be online always.

Recent studies in ICN caching as the Max-Gain In-network cache (MAGIC) cache strategy in [54]. The study stretched the need for saving bandwidth in ICN caching, hop reduction and the problems faced in replacing least recently used contents in ProbCache and other caching strategies in ICN. As presented in the study, several ICN tend to trade bandwidth consumption against cache-hits on nodes. However, the extended algorithm of ProbCache presented in MAGIC introduced cache replacement cost known as penalty to reduce content redundancy. The weakness in MAGIC was the un-automated algorithm that demonstrate ways of mitigating the transparent cache deployment used in the content chunk saving. Part of the weakness in the study was also lack of considering a true network scenarios in evaluation.

ProbCache by [24, 47] was seen as a strategy that has gained high popularity in ICN cache management. The sole objective and strength of the ProbCache was the elimination proposed of the LCE, LCD and MCD due their one position caching scheme except for the LCE that caches everywhere. In ProbCache, cache chunk are probabilistically cached to mitigate content over loading on routers that serves as the Publisher. However, with the probability in place, caches are mostly observed to be cached in routers with probability values *p*. ProbCache used the concept of determining an ideal position which according to the probability nature sometimes increases distances between the hops (known as network stretch) when cache hits are acquire far from the Subscriber. Using ProbCache sometimes actualize the main goal of the Subscriber when the content increases in popularity but suffers the reverse and deletion when a content become least recently used. However, as submitted by other researchers

[51, 48, 50, 23], proxy caching at the edges of the Subscriber and Publisher would benefit the entire network performance thereby improving the cache-hit ratio. This study therefore propose to tackle some of the issues inherited from the current stage of ProbCache deployment strategy to mitigate content and path redundancy. Proportionally the concept will also improve bandwidth efficient usage as submitted in [54].

Experimental Comparison of Related Work

In order to strengthen our critical analysis of the previous works, we performed some empirical experimental comparison of the common proposed cache management strategies used in ICN. The experiments reveals the strengths and weaknesses of some of popular cache deployment strategies. In the results obtained, the cache-hit ratio was the main concern for the benefit of this stage of the study. To simulate an acceptable scenario of ICN, social network traffic generator by [109, 110] was adopted to test the simulations of some ICN topologies. Abilene network simulation topology and Duetsch Telecom (DTelekom) were simulated. The details of the values and parameters set are presented on Table 2.1 as thus; using a network research traffic from a large sample users known as Facebook [111, 112, 113, 114], 4,039 nodes were involved in the traffic generation. Each User sends and request different or similar data depending on the traffic set. Modeling ideal network traffic is best tested with approximate probability distribution with exponential or uniformly placed time intervals.

This work thus adopted the nature of traffic as with the following parameters. The simulation was ran for forty (40) occurrences to attain an acceptable rates (almost stable) of instances in ICN traffic manageability.

This section presents the computational and simulation comparisons of LCD, LCE and ProbCache. For easier analysis, the section focuses on the three popular strategies. In

Table 2.1
Simulation Parameters

Parameters	Value
Number of Nodes	4039
Number of files	10,000
Mapping Algorithm	Random
Time limit	86,400
Social graph	Facebook [111, 112]
Number of communities	5
Cache replacement	LRU

the results obtained from simulation, ProbCache is seen as a leading strategy using the same parameters and network topologies.

Abilene Network Topology

Sometimes referred to Intetnet2 network topology. Abilene topology was designed with the sole object of transferring large size data across nodes. Its building was ideal to test and compare our simulation parameters to record the cache-hits and observe the efficacy of our cache managements. Abilene heterogeneous nature depicts eleven (11) station sites as nodes. Its hierarchical structure provides the flexible neighbor composition needed to test ICN cache deployment. This forms the basis of our topology selection to test the performance of the strategy for heterogenous Internet.

From Figure 2.18, we present the results obtained from simulations using the socialccnSim [109]. The results obtained show the hit-ratio of the content cached. On Abilene topology using the parameters, it was observed that the hit-ratio was good as noticed in the graph with a cache-hit ratio hitting the 100 mark at the 24 run of the simulation. The correlation of the runs of simulation was not far from granting a user request almost at good runs between 20-25. A declination observed in the simulation is attributed to the exhaustion of router (path) space earlier mentioned as path redundancy. LCE cache a requested data on all router making contents reside redundantly on nodes.

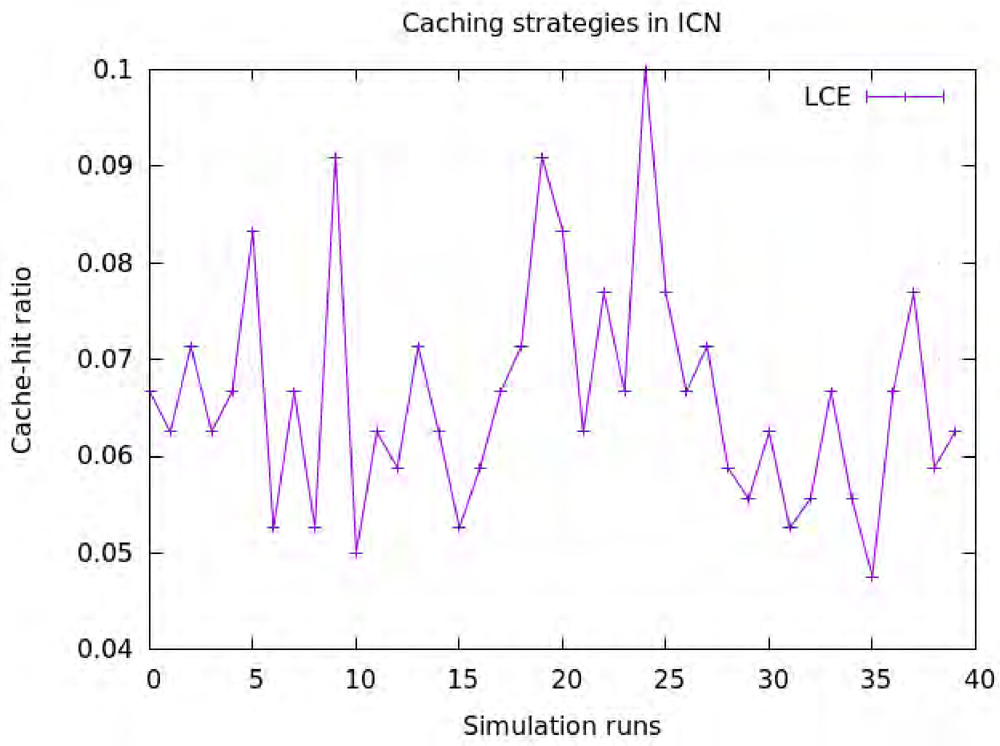

Figure 2.18. LCE Abilene

This shows a waste of bandwidth with a trade-off of a cache-hit. Comparing with the LCD, the cache-hits differ in their form of acquiring a requested data on the Abilene topology.

From Figure 2.19, the results obtained from the simulation, the LCD performs better in terms of hit-rate as compared to the LCE. However, it hits the high mark at simulation 13 run. It shoots high to 0.11 cache-hit ration due to its advantage of caching at the immediate nodes without necessarily deleting its preceding node. However, in LCD, the stretch was high, but a trade-off of bandwidth utilization was obtained against hit-ratio. It is thus right to say a lesser cache redundancy was obtained while ProbCache on Figure 2.20, has been seen through the simulation to be more effective in terms of redundancy in path and content. Its only negative transformation was observed between runs 25-31. It was noticed that no cache-hit was obtained as the runs and demand for the increased data. This shows the probability of the content being requested

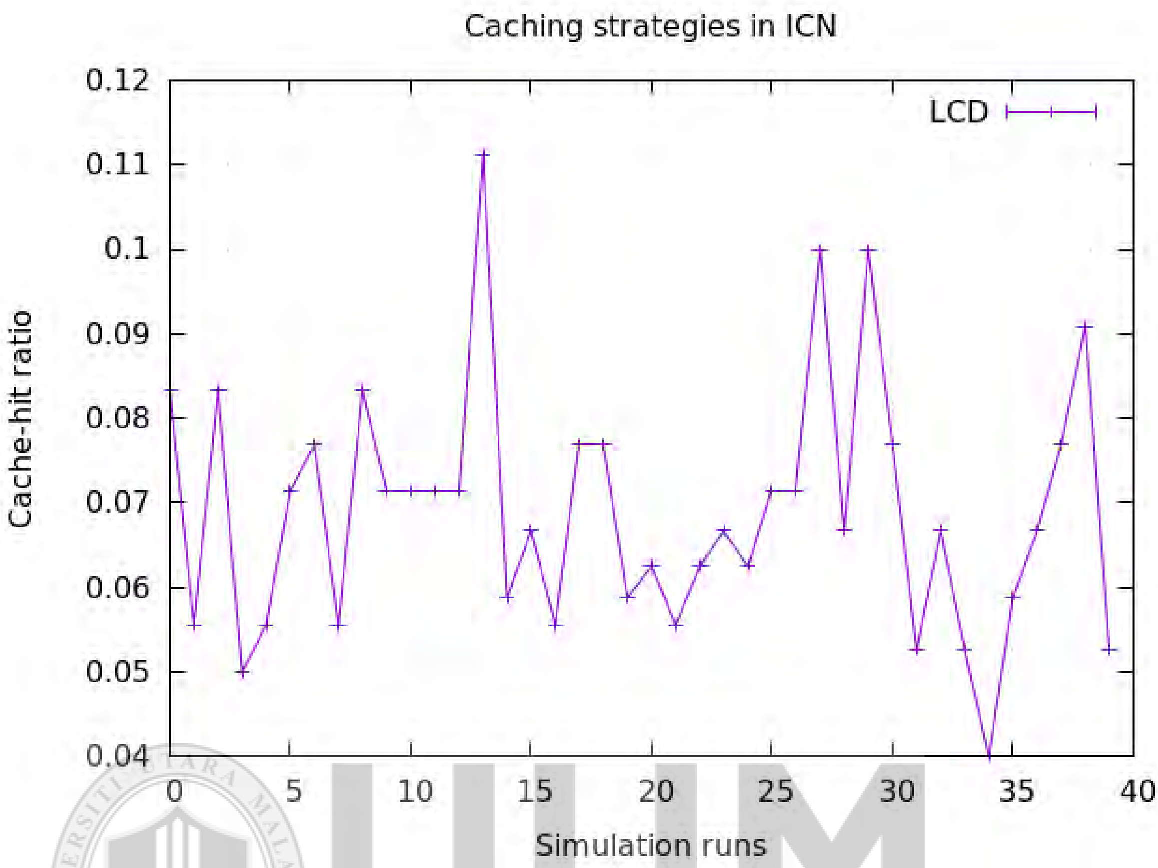

Figure 2.19. LCD Abilene

was (1-p) = q at those instances. However, it is fair to conclude that the trade-off was high in terms of keeping the path less congested, better stretch probabilistically against LCD.

Therefore, the ProbCache won in terms of avoiding all routers/nodes consumed. ProbCache can thus be extended to improve stretch, and cache-hit ratio through proxy positioning.

DTelekom Network Topology

To test fully the heterogeneity in the ordering of nodes and stations, we used the DTelekoms topology to investigate the cache-hits and stretch in the network of content. We found out through the simulation that the values of the stretch was quiet different

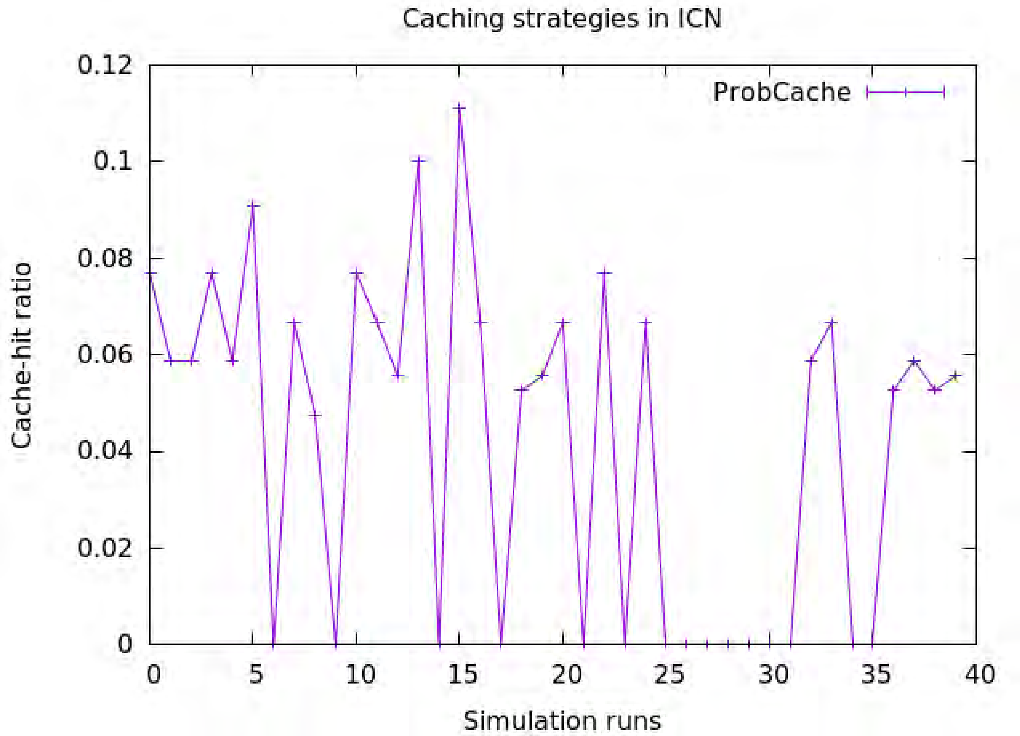

Figure 2.20. ProbCache Abilene

compared to what we obtained in Abilene topology. Using the same network traffic parameters as presented above, the DTelekom cache-hit was almost the same as the hit ratio was observed to be more at the 0.072 between the 2-3, 5-6, 14-15, 20-22 and continuously in that order of increment. However, the highest hit ratio was noticed this time around on the 33 run simulation mark on the x-axis. This has proven that the amount of data flooded is high and as such affecting the hit ratio. Consequently, in LCD run on DTelekom, the results appear to be the opposite as more hit were constant or related at about 0.077. As previously mentioned, the LCD saves the memory waste and implies a hit on DTelekom due to the heterogeneous ordering of the topology. Conclusively, it is fair to note that LCD saves bandwidth and recorded a good hit using the DTelekom topology (See Figure 2.21)

ProbCache seems the most flexible cache management technique and suited to a DTelekom architectural structure of a network. It is worthy to mention that once a

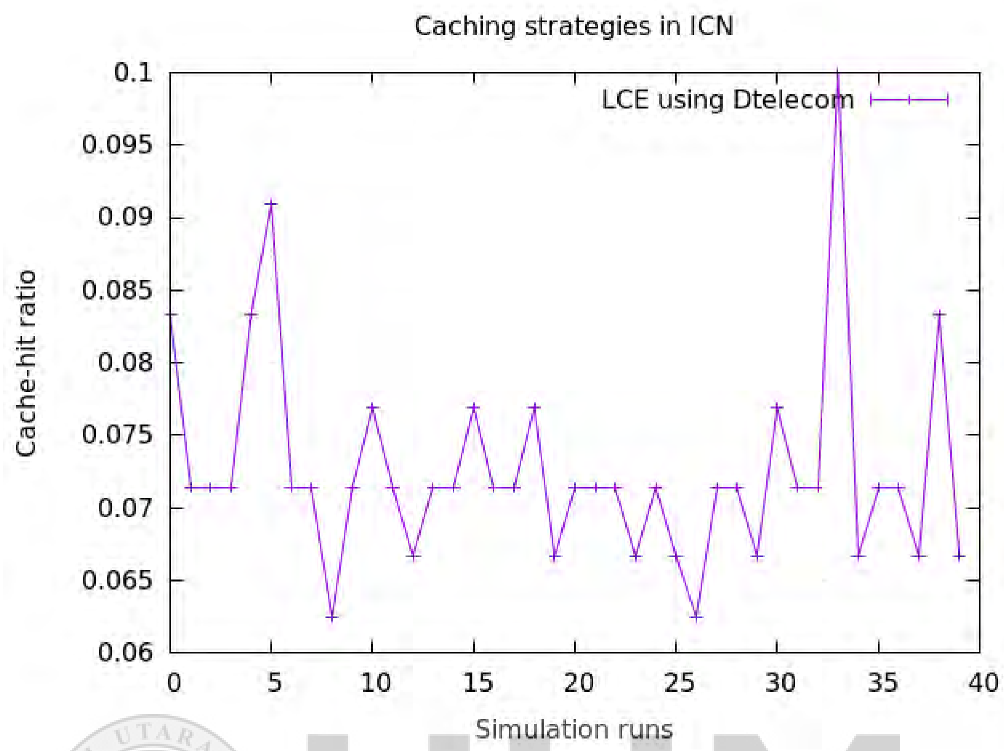

Figure 2.21. LCE DTelekom

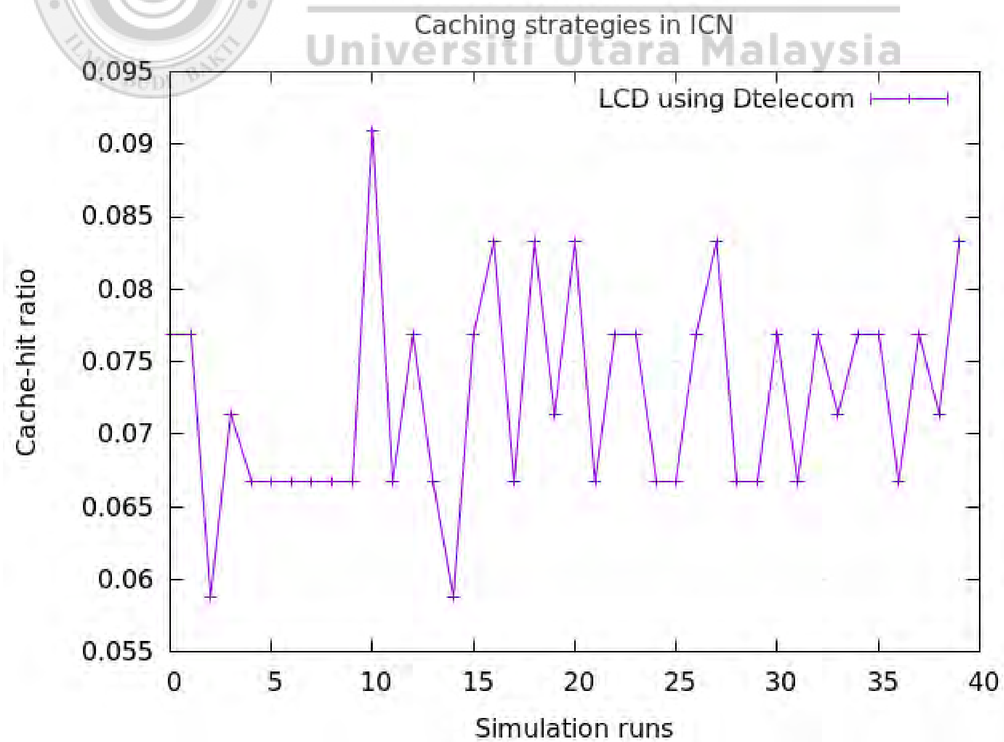

Figure 2.22. LCD DTelekom

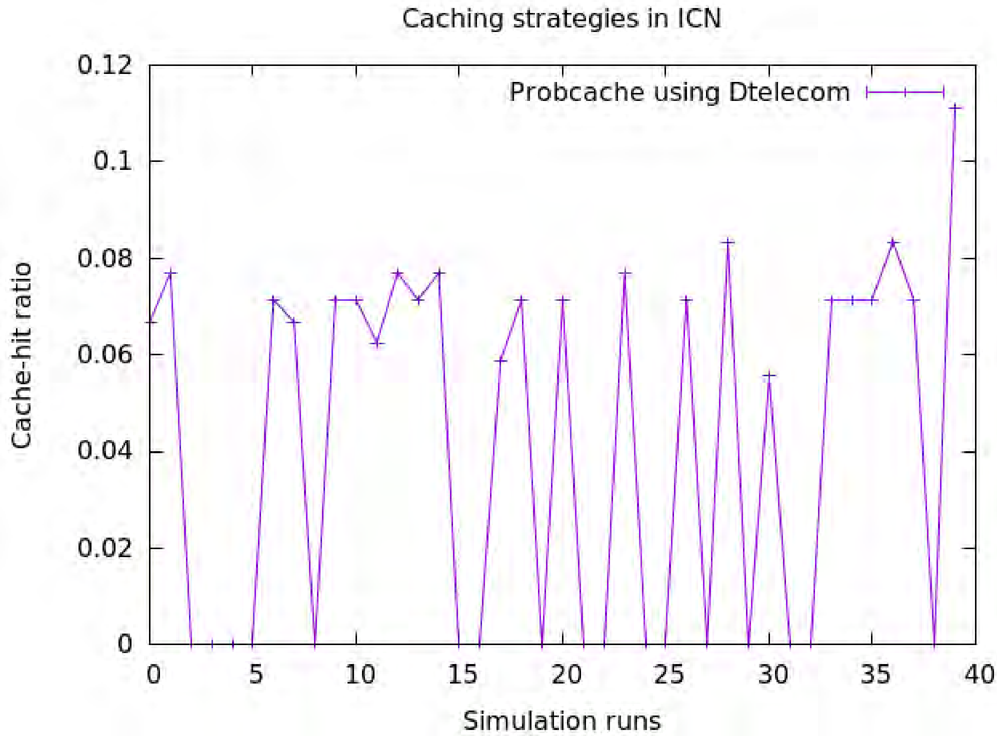

Figure 2.23. ProbCache DTelekom

data has been found (cache-hit), probabilistically copies of the data is replicated as caches in the next availing spaced nodes. This thus, increase the chances of a data gaining better hits and likely proxy positioning to the subscriber (See Figure 2.23).

Simulation results (see Figure 2.24) has given the bases for an extension of the Prob-Cache to enhance data availability with lesser path and content redundancy. In ICN, the yearn advantage is to avail data almost at all nodes on the path. But the excessive requests of data driven activities on the Internet posed a threat to the traffic and upward and downward streaming as submitted by Cisco [3].

2.10 Cache Size and Cache Location

From the summaries presented on Table 2.2 (Cache Summaries), several researches in ICN have shown that cache sizes are relatively small as compared to the predictive requests that move in and out of nodes or stations. As explain in [46, 115, 116],

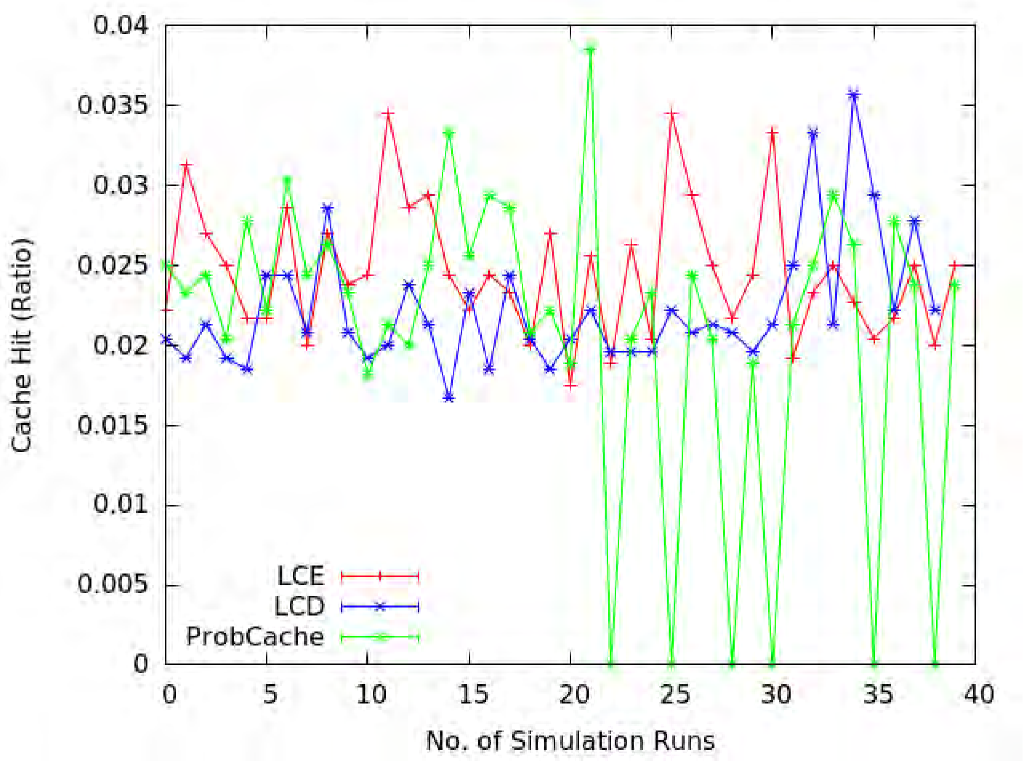

Figure 2.24. Composite LCD, LCE and ProbCache

cache type is on-path. Meaning the location of the cache-enabled router is within the route/path of interest and data access. Xylomenos *et al.,* [9] in their work exposed the need to have additional mechanism for off-path caching in ICN. Therefore, for closer node and information to the user, proxy strategy was adopted by four authors on the summary table in 2.2. From the traditional Internet perspective, caches are meant to hold few information for some limited duration. Placement and cache location deployment has gained a lot of attention in the ICN research community. The deployment and location can be an influencing factor for bandwidth utilization, mitigating high hop crosses, minimizing traffic and overall network delay. Studies by [47, 54, 108] have identified the need for proxy, reverse proxy, active deployments of caches with each related to the localization of the cache-enabled node/router. Therefore, cache size would be efficiently utilized based on the deployment scheme and algorithm of placement, eviction or replacement of contents in ICN.

2.11 Summary

This chapter covers the new paradigm shift of ICN by introducing the techniques of its operation. Various projects and proposal such as CCN, DONA, NetInf and PSRIP were identified as they relate to caching. Caching as one of the leading issues that receive most of the attention in the ICN research community was critically analyzed. Caching management includes it deployment in relation to positioning, content placement and replacement in relation to content usage, replacement and eviction of contents to mitigate the excessive bandwidth wastes.

However, the deployment strategies was seen as a variable that determines the flexibility of the ICN approaches. Therefore, adequate and intelligent cache deployment strategy is needed to efficiently manage Interests and data traverses in and out of the network.

Table 2.2
Cache Summaries

Authors	**Cache Type**	**Gains**	**Issues**	**Algorithm**	**Cache Strategy**	**Future Work**
Li et al., 2012 [48]	On-Path	Mitigate Excess Bandwidth	Coordination Cost	Popularity Concept	Proxy	Coordination
Vasilakos et al., 2012 [16]	On-Path	Reduced Cost and Delay	Cost and Delay	Selective Neighbor	Reactive	Coordination and Delay costs
Sourlas *et al*., 2012 [117]	On-Path	Improved Complexity & Execution time	Optimality	Greedy Algorithm	Active	Coordination cost
Psaras *et al*., 2012 [24]	On-Path	Hit Ratio	AS Caching	Popularity	Reverse Proxy	Cost
Galluccio et al., 2012 [118]	On-Path	Mitigate Excess Bandwidth	Distance	SatCache	Selective	Coordination cost
Wang *et al*., 2013 [97]	On-Path	Byte Hit Ratio	Hit Ratio	AssignCache	Adaptive, Transparent	Forwarding
Suksomboon et al., 2013 [119]	On-Path	Time & Cache Distribution	RTT	Popularity	Proxy	Averaging Download Time
Saha et al., 2013 [120]	On-Path	Reduced Content Redundancy	Cooperative Cache	WaxMan	Proxy	Redundancy
Sourlas et al., 2013 [108]	On-Path	Minimized Network Traffic cost	Overhead, Complexity	Cooperative, Holistic & Myopic	Active, Transparent	Extended AutoCache
Chen et al., 2013 [74]	On-Path	Minimized Traffic, Delay, Less hop crosses, Cache hit	Overhead, Delay, Traffic, Route Distance	Cooperative Cache by LUV-Path	Proxy, Reverse Proxy (Upstream & Downstream)	Continuous & Undetermined cases
Psaras et al., 2014 [47]	On-Path	Reduced traffic & cache redundancy	Resource Mgt.	ProbCache	Active, Transparent	Cache pollution, Extension. Deterministic
Ren et al., 2014 [54]	On-Path	Minimized Network Bandwidth, Cache Penalty	Hop Counts	LCE, RAC, MFU	Active, Transparent	Extended ProbCache

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents the full details in designing the proposed cache deployment strategies for Information-Centric Network. Cache deployment as described in previous chapters influence the performance of the entire network. Several deployment strategies according to the positioning of the cache-enabled routers affects the content centrism. A unified positioning of cache routers are needed to mitigate the content and path redundancy as submitted by Psaras *et al.*, [47] and [54]. The chapter shall adopt the methods described in Design Research Methodology (DRM) by clearly explaining each steps of the approach.

A detailed research approach, which will cover the areas where the literature analysis link the gaps will be covered. Section 3.2 shall cover the Research approach and the stage of Research Clarification (RC) in 3.3. A general description of the RC, methods adopted, stage and deliverable shall be presented. Furthermore, Section 3.5 called the Descriptive Study-I (DS-I), shall handle some literature analysis and experimental approach of the cache strategies. Steps of actualizing the conceptual models, current study stages and the designs of the proposed strategies shall also be covered in DS-I. Section 3.6, known as the Perspective Study (PS) shall present the methods and approach of the proposed cache deployment strategies for ICN. This stage shall be dependent in terms of deliverable with the DS-I. Conclusively, methods of cache deployment using content placements in ICN shall be tested to measure the performance evaluation of the network using some popular ICN metrics such as cache hit ratio, path redundancy, diversity and delays in the last phase of the DRM method.

3.2 Research Approach

The main objective of this research is to design a cache deployment strategy that will mitigate the amount of content and path redundancy in ICN. Deployment proximity to the subscriber and publisher has corresponding benefits in two ways. The proximity of a cache-enabled router to the user benefits the subscriber closer nodes and stations. Since ICN performs forwarding and data sharing between neighbors, subscriber proximity would reduce high traffic delay of revisiting the main publisher node. Step-by-step elaboration of the Design Research Methodology (DRM) as described by [121] concludes that all steps in the DRM should complement each other in order to produce better and reliable results to the research problems. The authors [121], however relates that research design should include scientific representation and both theoretical and practical sense.

The DRM is an approach, guideline and set of supporting methods that fully itemized the phases of adequately doing research. Blessing and Chakrabati states that "it helps in making designing research more rigorous, effective and efficient; its outcomes are academically and practically more worthwhile" [121]. Due to DRM wide acceptance in the research community, DRM by Blessing and Chakrabati [121] has been adopted in this research as used by [122]. Some of the objectives of the DRM are as follows:

- it guides in providing adequate framework for individual researchers to conduct and design research
- serves as a platform that helps in identifying research areas that are academically and practically worthwhile and realistic in conduct
- it allows a variety of research methods in addition to helping in selecting methods and combining methods
- in providing guidelines for rigorous research and systematic planning.

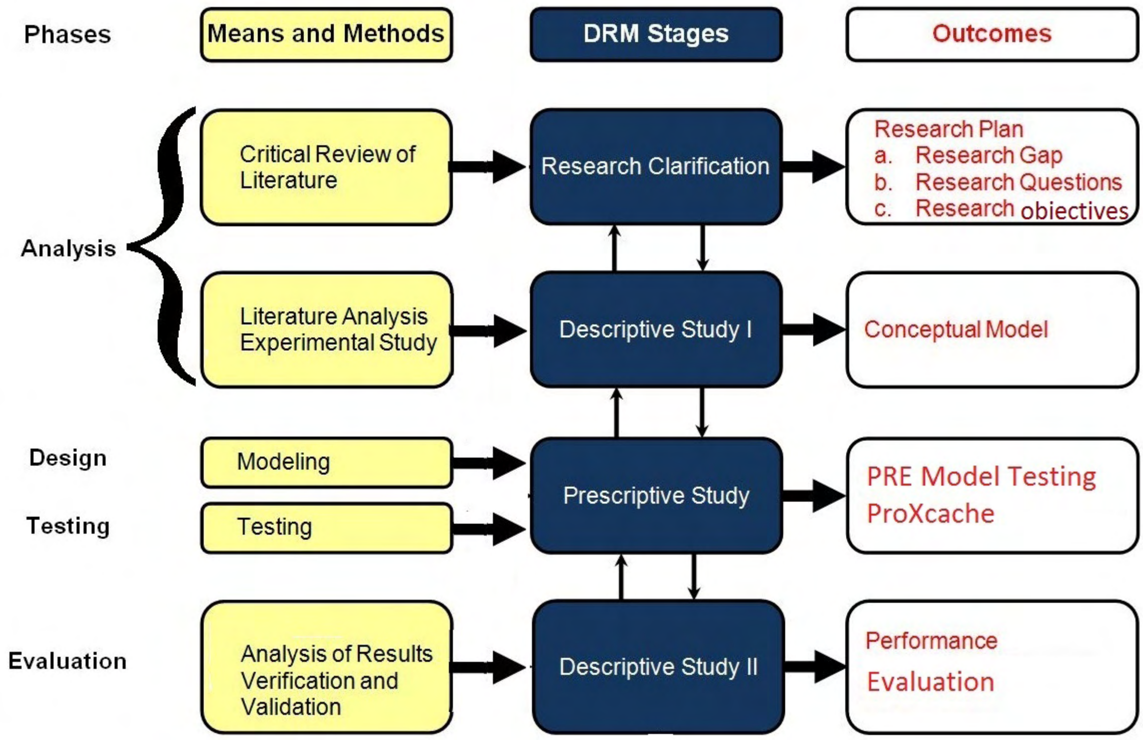

Figure 3.1. Research Approach

DRM stages can be classified into four main parts, namely Research Clarification (RC), Descriptive Study-I (DS-I), Prescriptive Study (PS), and Descriptive Study-II (DS-II).

Figure 3.1 describes the framework of the research using the stages of the DRM. The illustration provides the methods employed, the respective stages and the deliverable. The light arrows between the stages depicts the main process while the bold arrows in each stage demonstrate methods used and deliverable of the particular stage.

3.3 Research Clarification (RC)

Research clarification (RC) presents the in-depth understanding and clarification of the objectives of this research. The phase also presents six interconnected stages: Identifying the overall topic of interest, clarifying the current understanding and expectation, clarifying the main questions and objectives, research type selection, determining ar-

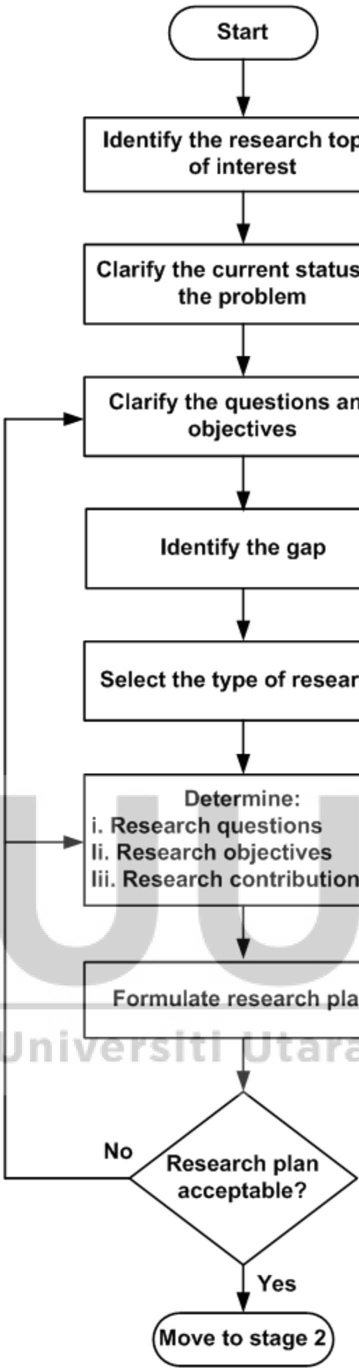

Figure 3.2. Research Clarification Steps [122]

eas of contribution and relevancy by formulating the research plan (See Figure 3.2).

Conclusively, the deliverable in the RC is chapter one. These present the entire deliverable and plan as the following:

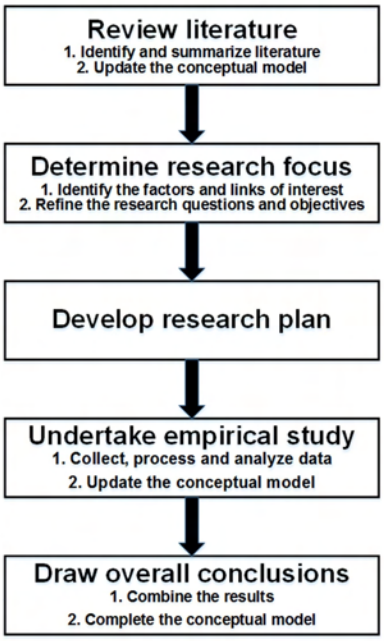

Figure 3.3. Descriptive Study-I [122]

- Research focus and motivation
- Research problem and research questions
- Relevant areas of the topic
- Approach (scope, research type, stages and methods)
- Areas of contribution

3.4 Descriptive Study-I (DS-I)

The next stage after the RC is the DS-I. The DS-I was used to gain a deep understanding of the current situation and it involves the critical literature reviews in Information Centric Network (ICN) caching as well as the empirical studies. Inline with the current works and proposals by other researchers, enhanced current contributions related to this study were reviewed. These provided better understanding of the area. Like the RC, the DS-I consist of five sub-stages. The stages are reviewing current literature, determining research focus, developing research plan for DS-I, undertaking empirical study and overall conclusion as depicted in Figure 3.3.

3.5 Conceptual Models

In this section the overall description of the research concept will be identified. The cache deployment strategy would be divided into two phases. The first phase handling the advantageous proxy to the subscriber and the publisher. Path distance, hops and position acquiring are needed in placement of position to deploy subscriber-proxy node. A subscriber sends out an interest I_0 at an instance of time t_0 to the network. The proposed concept dedicates and calculates the path due to the factors of network weight, load on servers, mobility and predictor history about popularity as used in [24, 119, 46]. The concept as depicted on Figure 3.4 shows the relationship between the distance and proximity between nodes and cache-enable routers.

The action is started at point 1; however at the phase t_0, the routers are assumed to be free as related to memory. Action point 2, traverses the request via its neighbor according to the usual ICN practice [6] by searching the nodes for the match as data. The action continues in that fashion until action point 6 where the subscriber interest matches the data of the publisher at t_0. According to this concept of ICN, the corresponding data is shipped through the routes from action point 7 until the 15 mark. As the data traverse on-path in the network, a logical computation is made to cache at intermediary node(s) to mitigate the problems identified in recent studies (path redundancy and content redundancy) [24, 46, 120]. When a request is further posed into the network, the cache manageability will have to be computed again as the former iteration in mathematical and logical form. Thereby, aimed at mitigating overloading, high cost of coordination and path exhaustion. From the Figure 3.4, Subscriber n sends interest at time t_n, of an interest I_n. The data is displaced and calculated into the network using the logical mathematical concepts proposed.

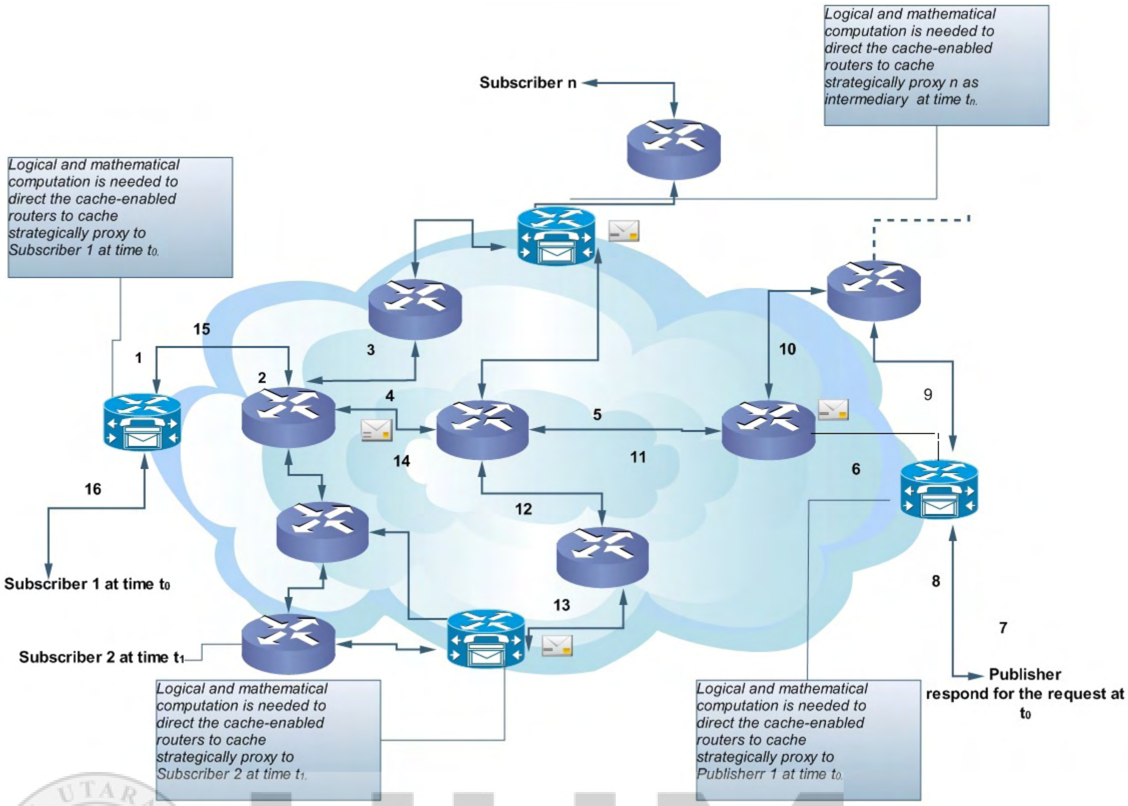

Figure 3.4. Conceptual model for cache deployment strategy

3.5.1 Proposed Cache Deployment Strategy

ICN cache manageability have been seen as the ultimate factor of content-centrism. With the huge number of researches focusing on different forms of managing chunks (data) in the network; recent studies demonstrated the flexibility of some cache management strategies as described in Table 2.2 in Chapter Two. However, the initial underlying technology proposal of ICN is faced with the problems of content overflooding, deposition on all routers on-path (resulting into content redundancy). Path redundancy has also been termed as memory wastage problem due to the limited sizes in caches and catalog (presented in Section 2.7).

ICN proposal focuses more on the designing phase in naming. ICN optimal objective is its ability to deliver information (Interest/data) to users without necessary caring about the location of the provider. This has thus been referred to a man-in-the middle benefits. To actualize the ICN goal, cache deployment strategies need to be thoroughly researched upon. Due to the nature of LCE in ICN, an extension of the LCE was the LCD. Several researches also try to extend the strategies such as the ProbCache [24], MultiCache [45], Proactive Selective Cache [16], PropCache [119], Intra-AS cooperative cache [94] etc.

However, content proxy positions from the earlier web cache studies [50] promised better cache-hit ratio and less content redundancy. The need to propose a flexible cache deployment strategy using hypergraph is appropriate to enhance cache-hit ratio, better bandwidth manageability (through lesser number of cache operations), improved content diversity and lesser *network stretch* (the distance in hop betweeness). Hypergraph model of edges and vertices could benefit the ICN cache standardization by incorporating node proxy positioning to the Publisher and Subscriber.

Studies using different content management have suffered from the aforementioned problems of excessive deportation of the chunks on router. Some content placement and replacement algorithms have been tested on the cache management. However, the need to improve content management will improve the overall network performance. Consequently, to mitigate these problems, this research proposes the hypergraph cache deployment strategy. In this study, careful contents and data management will be introduced. To test and propose the workability of our strategy, LRU content replacement policy was adopted. The LRU adoption is inline with most researches believing LRU to be the optimal replacement strategy to benefit ICN [23, 123]. This strategy aims to curtail the weakness in the current ICN deployment strategies. Its functionality can be summarized as follows: In the hypergraph cache deployment strategy, a point to note would be its interest in computing cache positions, cache-able routers for each time a request is posed to the network.

In the proposed hypergraph strategy, a content is requested by the subscriber who attaches a header address to the interest data as the usual practice in IP network [6, 124]. The forwarding plane and routing information [6, 4, 15] takes care of the traversing operation of the data chunk. On checking the Interest in the designated PIT and CS to satisfy the request, path computation, cache time span and hop distances are acquired based on the proposed algorithm in the next subsection. Once a cache-hit is recorded, in return of the data packets (chunk), a proxy position to the publisher is obtained and first deposition of data is cached as an extension of the LCD. The distinction in our proposed strategy differs from previous studies through the continuous traversing of the data to deliver the chunk to the subscriber using path distance, cache time window, hop distance after which the data is probabilistically cached in the network mid-routers using the centrality computation algorithm. Based on the description of the algorithm, the data chunk is thus cached and moved to the immediate (proxy) node to the subscriber. This operation is performed through the PIT record on the direct intermediate node. Similar researches have submitted the advantage of proxy-caching to subscriber which results in better cache-hits and thereby mitigating excessive distances between hops. Bandwidth is thus saved and effectively used in the network. The flexibility in the hypergraph aids in path computation through differentiating between edge nodes and mid-points nodes.

In continuous network traffic and requests cases, the path capacity, cache time windows and proxy positions of the network are obtained in every iteration of the requests.

3.6 Perspective Study (PS)

In this section, the proposed cache deployment strategy shall be covered. Logical and mathematical models needed for the decision of cache deployment are vital to implement the events. Scenario setting alongside parameters used shall be describe in the

PS. For the benefit of this research for wider understanding, modeling and simulation as described in the chapters of [125] by Guizani *et al.,* shall be used. For the above conceptual model of cache deployment strategy, it can be said to be discrete or continuous. A discrete event can be seen as a network model that is static like without active mobility (mobile network). In discrete event models, occurrences of the model items happen as fixed time interval or stipulated mark. However, in the continuous case scenario like a typical ICN environment, requests are flooded and posed at exponential intervals. This case can be said to be continuous. For a typical caching model, a deterministic or probabilistic models could be considered. In deterministic models, repeated runs of data sets several times yield the same output while in probabilistic, probably occurrences are used and observed; hence when the runs are made the outputs may differ. Therefore, probabilistic seems more related to the packet and interests flooded into the network.

3.6.1 Verification and Validation

For the purpose of the proposed cache deployment strategies in this study, it is required that the approach is verified. The verification exercise would be done in conformance with the model and mathematical computations added to the codes of the simulation. However, the verification of the codes is vital to remove and correct all code and bug errors resulting from the computation. The first stage will be to represent the models into flowcharts for better understanding. This will promote the development of pseudo codes and modules to be included into the simulation tool. Verification is said to be the internal representation to have consistency with the models while the validation is the mimic and true representation of the reality with the models [125]. The intriguing question of "is the system built right?" is posed to the validation and the question of "is the built system right?" is answered by verification. However, another process known as calibration checks to verify if the data generated is in accordance to the real expected data [125].

To juxtapose the ideal cache practice during validation, the results obtained shall be compared with other results obtained form current studies related to the proposed system. Below on Figure 3.5 is the phases of validation as it relates to the description of the deployment strategies.

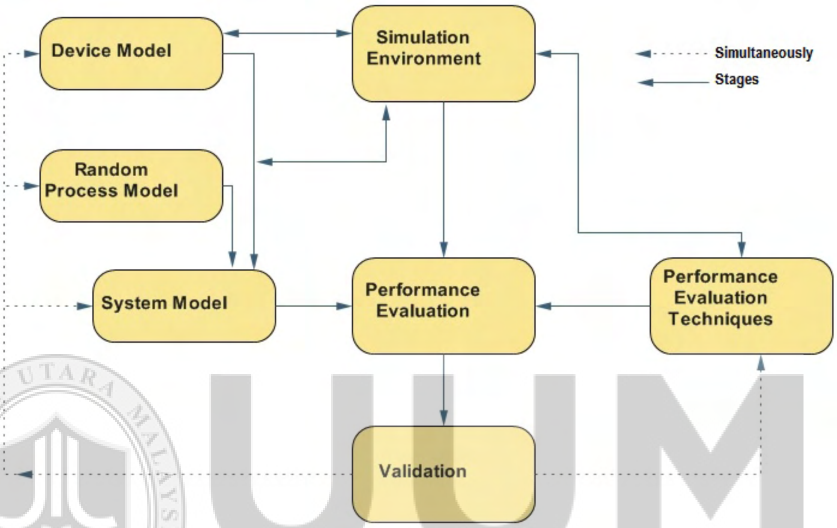

Figure 3.5. Validation Phases [125]

The phase handles several models as the source, device models, randomly generated models and system models. These are then included into the simulation, the simulation is then run on several scenarios with altered environmental parameters. Once the phase is completed, it is then checked through the performance and validated. The cycle continues until an acceptable mark of simulation is reached from the results obtained.

3.7 Descriptive Study-II (DS-II)

This section explains the various ways of evaluating the entire work. From Guizani *et al.*, in [125], evaluation can be described as the sets of assumptions that correlate with the models for an estimated measure of performance of any system. Performance evaluation could be done using two broad approaches namely: Direct measurements and Models. Analytical/Mathematical modeling and simulation approaches are seen as the model performance evaluations techniques while measurements which is mostly seen as the most reliable posed a heavy challenge when dealing with networks or system that involve costly configurations to come-by.

Each of the aforementioned methods of performance evaluation is dependent on the scenario and system. Measurements are highly recommended when a system is time driven and requires lesser repetition. Testbeds are usually used for this kinds of environment where the system is readily available. However, for the benefit of this study, models seem more of the viable option. In terms of cost and time required to perform evaluation; direct measurements is at the high side compared to the models. Furthermore, direct measurement will require more experts thereby causing a huge tradeoff to be taken for convenience. Therefore, simulation based performance evaluation will be adopted for the study.

3.7.1 Simulation

Simulations are widely seen as a form of representing real behavior and nature of systems. Simulations are computed-driven models and application that allows computer logic and decisions to be included. Its wide acceptability and cost effectiveness has made it possible to represent costly and time consuming operations. This makes analysis, protocol performance, extensibility, scalability controllable in repetitions. For the purpose of this study, simulations will be adopted in lieu of direct measurement faced in setting up a test bed. From several studies such as [126, 127] investigated each potentials. However, SocialCCNSim [109] and ICARUS [128] are seen as the leading open-source simulators that were mainly designed to model ICN caching scenarios. Their flexibility will enable the study to in-cooperate the logical and computational extensions to the modules. However, other ICN simulators are ICNSim [129, 130] and ndnSIM [131, 132] are also a viable option for ICN-based simulations.

Some of the advantages envisioned in choosing the SocialCCNSim simulator include:

- A stationed computer or a small number of interconnected computers.
- The simulators would provide the flexibility to the researcher to perform several runs within a short time frame.
- Topology and different network parameters would be tested which are advantageously possible in simulation as an advantage over direct measurements (testbeds).
- Statistical assumptions are easily tested to mimic and validate the natural environments.
- A good incorporation of the traffic generator SONETOR.

3.7.2 SocialCCNSim cache Simulator

A descriptive representation of the SocialCCNSim simulator has the CS, PIT and FIB as shown on Figure 3.6 presents the full inner components of the simulation events for clearer understanding. The figure shows the relationship between the CS, FIB and PIT as described throughout the research. The strategy layer, core layer and the client requests arranged in chunks for Interests traversing on each cache-able device.

3.7.2.1 Simulation Steps

Simulation setups involves some series of phases. These phases include: Defining the problem, designing the network models and selecting network parameters, performance metrics selection, variable parameter settings, model constructions, configuring the simulator, data running and simulation execution, result presentation and interpre-

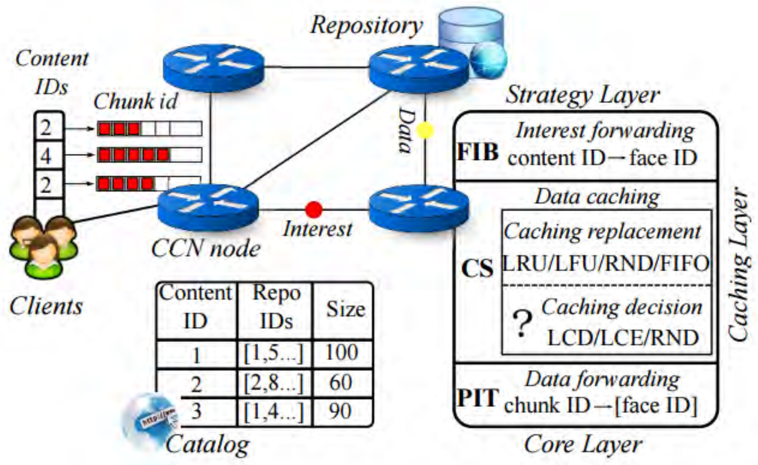

Figure 3.6. CCNSim Architecture [126]

tation for analysis.

The aforementioned phases can be illustrated on Figure 3.7 according to Hassan and Jain [133].

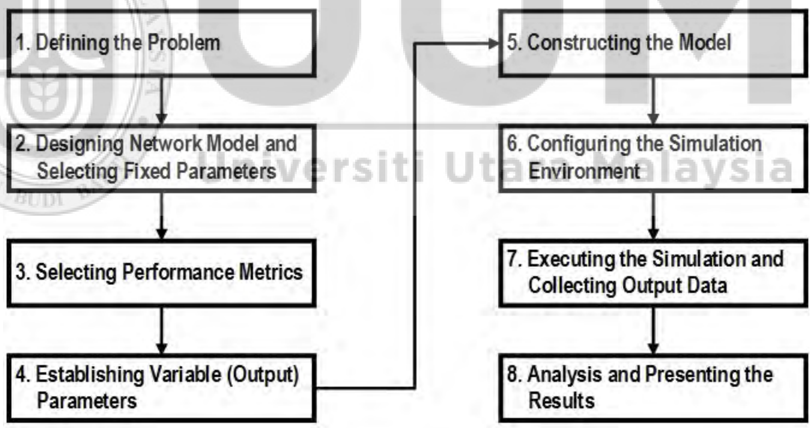

Figure 3.7. Simulation setup

3.7.2.2 Performance Metrics

Metrics to be tested on this study are determinant on the effectiveness of the cache deployment strategies. Adequate caching will result to better hit ratio, lower latency, throughput and efficient utilization of the bandwidth which are in relation to the following metrics.

Redundancy

In the context of ICN, redundancy means the number of content copies cached at more than one location in the network with lesser referral or unsubscribing. The redundancy of content is the excessive caching on traversing nodes which result into less or no referral to the content. Redundancy is therefore presented for each network setup and accordingly during interests and data dissemination as described by Van Den in [134]

$$R_{ci} = \mu_i * \frac{1}{m_c} * N_i \tag{3.1}$$

and this further defines the Total content redundancy as

$$R_c = \sum_i R_{ci} \tag{3.2}$$

where R_{ci} is the i^{th}content that was cached in the network at a given phase of time. R_c defines the total content redundancy value. m_c represents the overall number of content and μ_i is the i^{th} content item using its index categorization and N_i defines the overall content stored on an i^{th} node position.

Diversity

This defines the number and rate of caching distinct contents in the heterogeneous network. The resulting rate is computed to mitigate high replica. Diversity defines the performance of cache replication. It provides and compute the ratio of distinctly storing unique contents in the cache-able devices. It is expressed in ratio terms given as $(1, \frac{1}{|N|})$. Where $\frac{1}{|N|}$ presents the total ratio of caches with the same elements in the nodes and 1 is the value that returns the occupation of all nodes having unique content chunks.

$$Diversity = \frac{len(\cup_{n=1}^{|N|} C_n)}{\sum_{n=1}^{|N|} |C_n|} \tag{3.3}$$

Bandwidth

These include the distribution in terms of knowing the content to be cached in the local and distributed manner. Caching tradeoff memory for bandwidth utilization through the demand of contents.

The coordination tradeoff can be represented as :

$$Bandwidth = Demand * CacheMissRate \tag{3.4}$$

Accordingly, Cisco's annual report [3, 1] submitted that 96% of traffic is caused by content transfer (media-driven data). If many users request a content from the server at the same time, maximum bandwidth is consumed and ultimately content delay occurs. A cache deployment management would improve fair utilization of the bandwidth.

Hit-ratio

Hit ratio means the portion of content requests satisfied by caches implemented within the network. Cache Hit ratio returns the amount in average of available content hit (found) when requests are sent [135]. Several studies in database and networking domain have emphasized on the performance of the network through better cache-hit ratio. This metric shows the potential of the caching strategy to moderate the number of redundant content copies in terms of searching, how often is the data found.

$$Cache\_Hit = \frac{\sum_{i=1}^{|N|} hit_i}{\sum_{n=1}^{|N|} (hit_i + Miss_i)} \tag{3.5}$$

3.7.3 Topology

Subjecting ICN research approaches need verification and validation on the universal widely used network topologies. Studies from [24, 136, 16, 137] have observed that the ideal structure of the topologies can affect the overall performance of the network. This is seen as a result of the overall structure or the hierarchical relationships that exist in the network topologies. This study shall be subjecting the proposed ProXcache to four distinct and carefully selected topologies for the implementation and simulation comparisons. Several studies have also submitted the benefit of making cache-enable devices closer to edges [138, 53]. In the literature, leading researchers in ICN have used several topologies to test the efficiency of each strategy. Studies by Psaras *et al.*, [24], Saino *et al.*, [128, 139], Ren [54], Li *et al.*, [140, 46, 141], Chai *et al.*, [75, 43] and Cho *et al.*, [142] used the ISP-Level topologies with different hierarchical relationships and structures. Popular among the topologies are GEANT (Pan-European and Educational network) [143], used in the study by Saino *et al.*, [139, 128]. WIDE (Japanese Educational network), TISCALI and GARR were also tested on the study by [139, 128]. Psaras *et al.*, [24, 47] used a 6-level binary tree to test the performance of their study. Li *et al.*, in [140, 46, 141], and Chai *et al.*, [75, 43] used tree topologies for performance measures. Rossi *et al.*, [144, 145] used a total of 5 topologies namely: Level3 topology, TIGER, DTelekom and Abilene. Recent study by Bernadini in [146, 147] used the TIGER, GEANT, Abilene, DTelekom and TREE topologies.

The representation of the tested topology are presented as in Figure 3.8. The Abilene

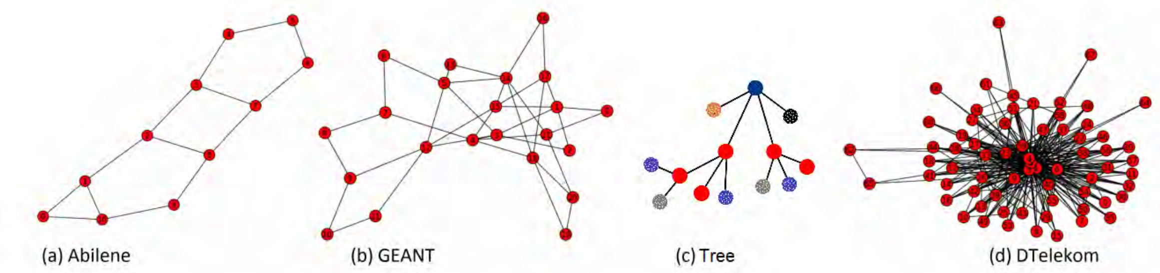

Figure 3.8. ISP-level Network Topologies[146]

topology, GEANT, Tree and DTelekom topologies. Each of the simulated result was tested in the following topology to suit the designed PRE model. Several studies have suggested the need to have flexible path detection forms in ICN design [148].

Consequently, this study shall be exploring and subjecting the proposed concept to 3 distinct topologies for the ProXcache strategy developed and presented in Chapter 5 for performance evaluation. It was noticed from the simulation results obtained, the higher the network topology inherent relationship, the more diversified a network is supposed to be. GEANT and DTelekom therefore, mimicked more ideal Internet setting as against ordinary Tree and Abilene network topologies.

3.7.3.1 Abilene Network Topology

Sometimes referred to Intetnet2 network topology. Abilene topology was designed with the sole object of transferring large size data across nodes. Its building was ideal to test and compare our simulation parameters to record the cache-hits and observe the efficacy of our cache managements. Abilene heterogeneous nature depicts eleven (11) station sites as nodes. Its hierarchical structure provides the flexible neighbor composition needed to test ICN cache deployment. This forms the basis of our topology selection to test the performance of the strategy for heterogeneous Internet.

3.7.3.2 GEANT Network Topology

GEANT network topology is seen as a full blown project network topology launched in the year 2000. Its launching has received enormous usage and benefits ahead of its initial project objectives. It is Funded through various phases of the Project, Its sophistication has made it one of the most advanced and well connected research and education network in the world. GEANT's pan-European research and education network interconnects Europe's National Research and Education Networks (NRENs). Together we connect over 50 million users at 10,000 institutions across Europe.

3.7.3.3 Tree Network Topology

Popularly also referred to as star bus topology, the tree topology among the most common and widely used network setups. It holds high similarity to bus topology and star topology. A tree topology connects one star network to other star networks. Figure 3.8(c) depicts tree topology, with a simple head-node (parent child relationship) using the star topology, connected to another network using the star topology.

3.7.3.4 DTelekom Network Topology

Deutsche Telekom is agreed as Europe's largest communications company, also one of the largest communications providers worldwide according to the 2004 revenues generated to the ton of 57.9 billion Euro. The company is effectively involve in the strategic areas of Broadband/Fixed-network, Mobile TeleCommunications and Business Customers. Its complex nature amount to the larger number of about 57.2 million access lines (2004). T-Mobile, Deutsche Telekom's mobile telephony subsidiary, Deutsche Telekom group today serves worldwide far above 77.4 million mobile telephony customers worldwide.

3.8 Summary

The chapter presents the full research methodology adopted for the purpose of the research. Cache deployment strategies for Information-Centric Network have been discussed extensively to be a key influencing area for better network performance. From the chapter, using the DRM by [121], the phases were covered in more detail. This shows that the objectives set can be achieved. From the description of the stages of RC, DS-I, PS and DS-II, the deliverable and outcome of each stage was itemized and covered. The outcome of the RC being the initial stage was to point out the research plan. This sets the peddle for the research proper. The research problems as related to objectives, questions and the motivation behind the research clarifies the path to follow in order to achieve the set out goals of the study. DS-I stage presents the description on how to achieve the first objective of the research. Measures and computational logic needed for the study were stated and analyzed. A good coverage of the challenges posed at the current situation, models and proposed conceptual approach was described. While the third activity highlights the methods adopted in designing the proposed cache deployment strategy for ICN, named Perspective Study (PS). The last activity named DS-II focuses on the evaluation of the cache deployment strategy.

CHAPTER FOUR
PATH REDUNDANCY ELIMINATION

4.1 Introduction

In line with the aforementioned methodology in Chapter Three and the slated objectives in Chapter One, this chapter presents the proposed *Path Redundancy Elimination (PRE)* in Information-Centric Network (ICN) caching. The proposed concept would be carefully considering some criteria as a base for proper mitigation of the redundancy in the path of an ICN. A cost model for data redundancy would be considered along the path accumulation of a subscriber to a Publisher. In the current ICN cache manageability, LCE which is used, deposits all requested data on the path of information traversal. As earlier mentioned, this practice results in overall delay and redundancy on paths. Further problems in efficient cache management strategy and variables to be considered include the content placement algorithm, the nature of traffic and behavior patterns of the contents, selection of cache contents and the flexibility of adapting to the traffic nature [149].

These needs also motivated the consideration of evaluating overlay of content caches. To solve the problems of cache deployment strategies, adequate and efficient information is thus needed to mitigate the above relating caching conditions. This chapter proposes a path redundancy elimination by reviewing the theoretical contributions of other cache forms (specifically LCE, LCD and ProbCache). Subsequent section of the chapter highlights the Mathematical and analytical relationship for achieving the *PRE.* Major cost of overhead through flooding excessive data on the path are covered in our proposed path redundancy elimination model. Section 4.2 presents the existing concepts in theory of various caching forms. This aids in fully understanding the basic operations of the concepts. Section 4.3 discussed the path redundancy elimination model; mathematical relationship and algorithms are presented in the section. The PRE model in ICN is covered in Section 4.4 while the chapter summary concludes the Chapter in Section 4.5.

Conclusively, a validation of the proposed PRE through design and simulation will juxtapose our submission for mitigating high path redundancy in ICN.

4.2 Theoretical Analysis

This section shall critically present and analyze the general description of theory in other studies. This include the thoeries followed and submitted in previous studies of ICN cache management. Among the studied theories are: LCE, LCD and ProbCache. ICN implementation as proposed by [6] exhibits the path delivery operation using the LCE [23] caching strategy. However, its acceptance was highly noticed in the Internet of names and centrism. With the predicted demand of media data that would engrave the overall Internet, basic LCE proves insufficient to curtail the traffic (Virtual Networking Index [3]).

ICN implementation through early submission of its framework include the three building blocks as explained in Chapter Two (Section 2.1). The Content Store (CS) which houses the data and content serves as a grand repository for caching. Once a request is posed into the network, a checker known as a Pending Interest Table (PIT) scans through the block to intimate the CS of the content. In this form, a copy is thus deposited into the CS and the following operation of forwarding/routing is initiated by the special designated Forwarding Information Base (FIB).

The following subsections shall present the theoretical relationship of the current strategy by carefully analyzing each operation in LCE, LCD and ProbCache.

4.2.1 Leave Copy Everywhere Strategy

The LCE strategy theory is based on the mid-node interaction. In LCE, copies of the demanded data are replicated along the path in an in-network operation. This form of strategy enables high-spread of content caching by accruing high path redundancy and route/path exhaustion in memory. This has been seen as a form that enhances neighbor intermediary but suffers from high traffic delays and lower cache-hits of heterogeneous data request. Redundancy is a challenging issue but necessary to mitigate overloads and better network utilization [150]. High impact studies in [151] proved that adequate redundancy elimination in networks can guarantee an increase in efficiency in a network by improving about 10% to 60% mitigation of redundant data in various network-driven setting for university accesses and enterprises.

Thus this study aims to eliminate path redundancy in an ICN caching scheme by introducing variables that control the deposition of unpopular or less frequently used data in-network path. This would improve the entire network performance, reduce delays and improve the throughput thus saving high wastage of bandwidth among other resources. A detail algorithm analysis of LCE is presented with variable relationships of each crossed node deposition of the data in ICN. A major challenge to this strategy is the cost of generating a new path and replacements when heterogeneous contents are demanded (See Algorithm 4.1[109]).

Algorithm 4.1 LCE Algorithm Analysis

```
#Get the parameters in the entire network (SourcePosition, time, DestinationPosition,
content, logRecord)
    Set SourcePosition = Publisher (content)
    PathDescription = Publisher acquire shortestPath (Subscriber, Publisher)
# Route acquiring aid the request to the destination to deposit caches on all path
  Begin Request (Interest) (time, DstinationPosition, content, logRecord)
           for i, j,k in path_links(PathDescription);
                 Subscriber forward_request via nodes (i, j,k)
                  if node_1_has content (cache j);
                        Publisher(i) = node_1 (j);
                        break
          If cache is not obtain on first check, get content from Publisher
                 Subscriber forward_request via (k)
                 Publisher(j) = node_2 (k)
Return data
    PathDescription = list (reversed(nodes_shortestPath(Subscriber, Publishers)))
    for i,j,k in path_links(PathDescribtion);
         Subscriber forward_request via nodes (i,j,k)
         if node_1_has content (cache j);
         Cache (i,j,k)
all_nodes_path = content(i)
LCE_end_sesson()
```


The presented algorithm shows the branches of operation in the LCE strategy. Critical analysis through simulation were carried out to ensure a real replication of the strategy. From the start of the subscription operation, when a Subscriber pose a request, paths are computed based on the traversing actions of the content on all paths of reaching the Publisher. Algorithm 4.1 describes the full usage of the path links *i, j, k.* When Interests traverse all the links, it deposits the data on all the links as it crosses and traverse back as depicted on the *#Return data object* in Algorithm 4.1. This therefore adds in aggravating the problem of path and content redundancy.

4.2.2 Leave Copy Down Strategy

LCD [23] strategy is based on the first neighbor node content caching. Copies of content known as interests are flooded into the network as requests. The resulting information of Interest is handled by the special designated data structures of PIT, CS and FIB as in the case of Named Data Network (NDN) of ICN. However, according to LCD form of caching, only a copy of the requested data (content) is cached as a replica of the data by the Publisher. Following a unique algorithm in LCD, a route/path of the Interest is computed in reverse hierarchical manner where the data is lodged. On a cache-hit operation, the content is immediately deposited in the first reversed traversing node. LCD therefore aims at minimizing the excessive replication of caches on-path with lesser degree of cache-hits due to fewer replication of data. This has thus motivated the need to make more data (content) available with lesser delays and less traversing distance between the Publisher and Subscriber. A detail theoretical representation of the strategy is presented in Algorithm 4.2 [109].

Algorithm 4.2 LCD Algorithm analysis

```
#Get the parameters in the entire network (SourcePosition, time, DestinationPosition,
content, logRecord)
    Set SourcePosition = Publisher (content)
    PathDescription = Publisher acquire shortestPath (Subscriber, Publisher)
# Route acquiring aid the request to the destination to deposit caches on all path
  Begin Request (Interest) (time, DstinationPosition, content, logRecord)
          for i, j,k in path_links(PathDescription);
                Subscriber forward_request via nodes (i, j,k)
                 if node_1_has content (cache j);
                       Publisher(i) = node_1 (j);
                       break
                else
          If cache is not obtain on first check, get content from Publisher
                Subscriber forward_request via (k)
                Publisher(j) = node_2 (k)
Return data
    PathDescription = list (reversed(nodes_shortestPath(Subscriber, Publishers)))
Cache(replicate) content in only on node_x
                  node_x = node_Publisher - 1
                  cached = False
    for i,j,k in path_links(PathDescribtion);
         Subscriber forward_request via nodes (i,j,k)
         if node_1 != content (cache j);
             Cache (j)
             cache = True
Only on_node_path = content(i)
LCD_end_sesson()
```

4.2.3 Probabilistic Cache Strategy

ProbCache [23, 24] strategy handles the cache data in the operation that involves path acquiring, time since interest was sent and the receiver node distance. The content identification also plays important role as the interests from Subscriber to Publisher keeps records in PIT for easy FIB operation of routing and forwarding. In ProbCache, as described and presented in [24, 47, 49], content data (requested) is probabilistically cached according to the amount of space in memory and catalog. The nodes with a better space are used for the data replication in a form of LCE. The main distinction between the LCE and ProbCache is the excessive replication of contents on all traversing nodes in LCE. During ProbCache operation, path indexes are set as the Interest leave the Subscriber, the index score is set at 0. On crossing a hop or node, the index counter is incremented by one to acquire the *Time Since Birth (TSB)*[24]. Probability is thus divided and created using the hop distance between the Publisher and the Subscriber. This has been seen as a motivation for this study to mitigate the redundancy accrued in LCE. ProbCache on the other hand also suffers from lesser content deposition due to its nature of probabilistically getting a caching position in the network. Algorithm analysis (see Algorithm 4.3 [109]) of the ProbCache presents the full theoretical analysis and the extensive description of the ProbCache strategy in ICN.

4.3 Path Redundancy Elimination (*PRE*) Model

Path redundancy has been itemized as one of the leading research gaps in the future ICN deployment [138, 151, 150]. Using the initial design of ICN caching, LCE was proposed and adopted by several designs of projects and ICN approaches. Bearing the initial design proposed by researchers such as [6, 21, 15, 4], LCE form of caching leads the submission that gave rise to the respective approaches and projects of content-centrism. This study aims at mitigating the excesses incurred in the caching form of

Algorithm 4.3 ProbCache Algorithm analysis

```
#Get the parameters in the entire network (SourcePosition, time, DestinationPosition,
content, logRecord)
    Set SourcePosition = Publisher (content)
    PathDescription = Publisher acquire shortestPath (Subscriber, Publisher)
Route acquiring aid the request to the destination to deposit caches probabilistic p path
  Begin Request (Interest) (time, DstinationPosition, content, logRecord)
           for i, j,k in path_links(1, PathDescription(getLenght));
                 i = Path[node - 1]
                 j = Path[node]
                 Subscriber forward_request via nodes (i, j,k)
                  if node_1_has content (cache i);
                        Publisher(i) = node_1 (j);
                        break
            else;
If the content is not hit, keep traversing interest on path_link
      Subscriber forward_request via (k)
      Publisher(i) = node_1 (k);
Get content
     PathDescription = list (reversed(nodes_shortestPath(Subscriber, Publishers)))
      set d = len([j for j in PathDescription if next node_x has cache (j)])
      startPoint = 0.0
      for hop in range (1, len(PathDescription));
           i = PathDescription[hop - 1] #Acquiring the distance from Publisher
           j = PathDescription [hop] #Self cache
          N = SumofCacheSize ([node_1_size[n] for n in the PathDescription [hop - 1]
                                    if n node_1_size])
          if j in node_1_size
           x + =1
          nodes_p_forward_content_in_hops(i,j,k)
          if j != Subscriber and j in node_self
        ProbCache = float (Total_size)/node_travel * node_cache[j]) * (x/d) **d
        if random.random() < ProbCache
              node_x_cacheContent(j)
       ProbCache_end_session()
```

ICN as described in Chapter One.

Mathematical relationship that exhibits the functions and conditions of path redundancy elimination are described in the subsequent subsections of this chapter. Graph theory [105, 106] has been studied extensively through its properties, characteristics and relationships. These properties have thus motivated our design to adapt the hypergraph and extend its qualities to fit in to ICN.

4.3.1 Hypergraph Cache Deployment Strategy

ICN cache deployment strategy can benefit from the characteristics of Hypergraph. In hypergraph, a relationship between the edge nodes and mid-points nodes are obtained. Thus the importance of acquiring the cache path capacity is the first operation for this research. Cache deployment strategy is the form of cache data management, positioning and conditions of data storage. In this study, the deployment will be using a method from graph theory known as Hypergraph [55, 106, 105]. Hypergraph is a standard graph with vertices and edges also known as arcs used for computational decision and network modeling [107].

Let a network be represented as an hypergraph as submitted by [55, 105, 152]:

$$H = (V, E) \tag{4.1}$$

where $V = \{v_1, v_2, v_3, ..., v_n\}$ represents the vertices (Routers) in the network.

Edges $E = \{E_1, E_2, E_3, ..., E_m\}$ be the connectors and relationship in terms of proximity of the routers. Such that there exists a direct relationship between the routers and connectors (E and V) as

$E \subseteq V$ for $i = 1, 2, 3, ..., m$.

A hypergraph H or G as defined by [106, 107] is also denoted as

$(V : E = (e_i)_{i \varepsilon I})$ on a discontinuous set V as $(e_i)_{i \varepsilon I}$ (I is a set of indexes) of smaller set subsets of V called hyperedges.

Sometimes V is denoted by $V(H)$ and E as $E(H)$.

This defines the order of the hypergraph $H = (V : E)$ is the cardinality edged of V;

which results $\mid V \mid = n$; proportionally the resulting edge $E \Rightarrow \mid E \mid = m$

In network terms, a hypergraph cannot be an empty graph due to the presence of edges and nodes termed as routers and connections.

Therefore,

$V \neq \emptyset$,

$E \neq \emptyset$,

However, empty hypergraph are thus represented as

$V = \emptyset$,

$E = \emptyset$, in other fields of study.

Application in ICN

Applying the hypergraph to ICN caching strategy, the edge points are the vertices V, while the connecting links are edges E [106, 153] .

This therefore defines the relationship in the network such that a node

$N = \{n_1, n_2, n_3, ..., n_x\}$

where n_i is the *i-th* vertex and x is the number of the vertices.

Correspondingly, the edges denoted by E represents the elements of the entry as

$E = \{e_1, e_2, e_3, ..., e_y\}$ where e_j is also the *j-th* edge and m the number of edges. This implies that the network representation provides each edge with the pair vertices connected [106, 152].

Applying a popular network topology Abilene in ICN to prove the study proposed hypergraph model, the resulting proposition was considered

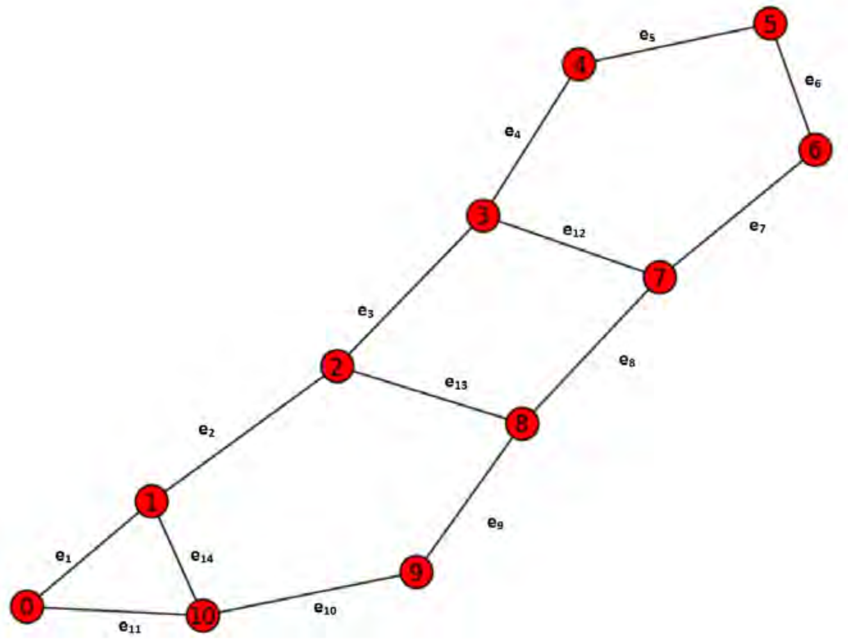

Figure 4.1. Abilene topology

Thus our model definition in accordance with hypergraph theory is defined as:

A network topology T, contains set of N vertices and E edges as

$$T = (N, E) \tag{4.2}$$

such that Abilene upholds the ICN formulation as

$$T^{'} = (N, E) \tag{4.3}$$

where

$$N = \{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

and

$$E = \{e_1, e_2, e_3, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12}, e_{13}, e_{14}\}.$$

By the heterogeneity of the Internet, each edge has a resulting pair interconnection as e_i. This implies that : $e_1 = \{x_0, x_1\}$, $e_2 = \{x_1, x_2\}$, $e_3 = \{x_2, x_3\}$, $e_4 = \{x_3, x_4\}$, $e_5 = \{x_4, x_5\}$, $e_6 = \{x_5, x_6\}$, $e_7 = \{x_6, x_7\}$, $e_8 = \{x_7, x_8\}$, $e_9 = \{x_8, x_9\}$, $e_{10} = \{x_9, x_{10}\}$, $e_{11} = \{x_0, x_{10}\}$, $e_{12} = \{x_3, x_7\}$, $e_{13} = \{x_2, x_8\}$, $e_{14} = \{x_1, x_{10}\}$.

By hypergraph concept and definition, adjacent and disjoint edges are therefore produced.

In this study, x_0 and x_1 are adjacent while x_0 and x_3 are disjoint since there exist no direct linkage of the edge/node. It is thus generalize in the order of the Abilene topology as x_i, x_{i+1} = adjacent; otherwise is a disjoint point. Consequently, a degree relationship of a graph is defined as the total number of edge intersection per edge. For a network node as the case of ICN, it represents the total number of node intersection and inter-connectivity.

Therefore, a degree

$$d(x) = \{number\_of Internode Connection\} \tag{4.4}$$

example $d(x_{10}) = \{3\}$ and $d(x_1) = \{3\}$

while $d(x_0) = \{2\}$, for 2 interconnected node on the edge.

Accordingly, the maximum degree of a network as a graph G is denoted by $\Delta(G)$

according to the graph definition.

In this study, therefore, using the Abilene topology to define and prove our model, the degree of the network is $\Delta(T)$ =3.

Therefore, the general degree of a network which is define as the total cache capacity of the ICN is given as

$$\Delta(T) = \sum_{i=1}^{n} d(x_i) \tag{4.5}$$

In hypergraph, a forwarding arc defines the continuous relationship between a node and a preceding node. Backwards arc defines a link that results to a leaf or edge node.

According to Gallo *et al.*, [55] and [105], the size of the graph is defined as n and m for the standard graph. Size in this study has great significance. As such the size is given as:

$$H = \sum_{E_1 \varepsilon E |E_i|} \tag{4.6}$$

4.3.2 Description of the Proposed *PRE* with Cache Capacity

Suppose there exist a populated ICN *H*, with large number of users spread across several nodes in the network in diverse form as Fricker *et al.,* [149] highlighted; some of the users are hierarchical related while some nodes are inherent. According to the hypergraph description, a node A is a vertice node with the elements connected through the paths or link known as edges

$$\{E_1, E_2, E_3, ..., E_n\}$$

thus

$$H = (E, V)$$

where $V = \{v_1, v_2, v_3, \cdots, v_n\}$ and $E = \{E_1, E_2, E_3, \cdots, E_m\}$ the relationship estimation to construct an ICN scenario with a

Size $H = \sum_{E_1 \varepsilon E|E_i|}$ where $\mid E_i \mid$gives edges.

Assume that there is at least one-to-one between hypergraph and Boolean matrix. Thus n x m matrix

$$A = [a_{ij}] \, \exists a_{ij} \{0, 1\} \tag{4.7}$$

such that there exist a relationship for each row i is related to vertex v_i and each column j is connected to an edge E_j. This assumption therefore generates the hyper-edge in pair $E = (X, Y)$where X is the tail and Y is the head. As described by [55], the relationship can be represented as $T(E)$ and $H(E)$.

To fully describe the relationship of the nodes and edges in ICN, assume the membership of nodes to be in H or T as described in [55], if vertices represents the nodes routers as r_i has an N_i capacity, this implies each node router r_i can hold a chunk. Then N_i can store a minimum time of traffic, such that Capacity of cache (CacheC) in the network is

$$CacheC = \sum_{i=1}^{n} N_i \tag{4.8}$$

This formulates a network H with the elements routers and edges as

$$v_1, v_2, v_3, \cdots, v_n \in N$$

$$\therefore \forall v_i \exists E_i \Rightarrow v \subseteq E$$

if $v = r_i$ represents routers and E represents paths.

Lets assume each r_i has N_i capacity, t represents the minimum time limit a router r_i can hold a cache content based on replacement.

Thus the overall capacity of the entire network path is

$$OverallNetworkCapacity = \Sigma_{i=1}^{n} N_i / tN \tag{4.9}$$

where N is the average Cache size along the path as presented in Algorithm 4.4.

Algorithm 4.4 Cache Capacity

Let S be a Subscriber and P be the Publisher

such that $s = r_i$;

Let d be distance between nodes r_i

where t is the minimum time a chunk can be stored in r_i;

if r_i is not the subscriber router, its $CacheC = N_i$

such that Total Cache Capacity in $H = \sum_{i=1}^{n} N_i$

for each r_i get N_i;

$r_i = 0$, count r, r_i++;

$CacheCP = r_i + r_j + r_k + \cdots, r_x \leq P$

where P = Publisher router

end

where *CacheCP* defines the overall path capacity of the path.

In a similar nature as the TTL in IP addresses, ICN can borrow the concept of *Time of arrival* (the time a process is admitted into the system) in operating system process schedule [62] to record header nodes and time of subscription. According to [24], a Time of Inception (*TSI*) and Time Since Birth (*TSB*) are essential to obtain distances. In operating system, the time of arrival is obtain against *burst time* (run time assigned to the CPU) by calculating the

AverageWaitingTime(AWT)=(TimeofSchedule-TimeofArrival)/NumberofProcesses [62].

Therefore, *TSI* is the time a client sends an interest into the network. A header field is included to the interest to compute the range from *1 to n*. This returns the time a request takes from a client. While *TSB* records the time taken from Publisher to the Client. A header field is inserted on the data to return the total number of hop-distance between the server as Algorithm 4.5 highlights.

Therefore, the request message incorporate the *TSI* to the request by the subscriber and the data includes the *TSI* and *TSB* to obtain the overall hop crossed to gain the message and the total number of hops crossed after serving the data. To generalize this relationship, this study propose the following algorithm.

Suppose an interest I_x is sent into the network, a time TSI is included on the interest I_x to provide the time request interest was created. I_x is recorded in each node it crosses but remain unchanged. However, once the data is obtained, the server records the TSI to actualize the TSI of the interest. Since in ICN, the data is traversed back through the same router, TSB is then marked by the resulting routers on-path of delivering each incrementing its score by one as the data crossed back to the the subscriber node.

To propose the approximate overall number of times a path can cache, we define

$$ApproxTime(z) = \frac{\sum_{i=1}^{c-(x-1)} N_i}{tN} \tag{4.10}$$

where $c = TSI$ and $x = TSB$, $time = t$ and N is the average Cache size along the path.

To determine the probability of where to cache in the divided network position,

we adopt the weight formula in [24] by defining

Algorithm 4.5 *PRE* Algorithm

Let the interest I_x from a subscriber be set at 1;
such that for every node r_i record the *TSI*;
Let *TSI* = α_i;
such that $\forall$ r_i records α_i;
until it arrives a publisher P_x at time $t = 0$;
set a new header message *TSB*
Let TSB = 1 from P_x
such that *TSI* reveals the total r_i crossed to P_x
set a combined header count *TSI*, *TSB*;
as TSB increments by 1 $\forall$ r_j;
New $PathLenght = \sum TSI$
Acquire the proxy position to the server P_x by designating $r_j = 1$ for TSB;
such that $\frac{TSI}{n}$ to gain the centrality positions by assuming the division as $\frac{TSI}{2} = \phi$;
ϕ= mid position routers in the path
Acquire the proxy position of the Subscriber I_x at $TSI = 1$;
such that
proxy to Subscriber $I_x, TSI = 1$
proxy to Publisher $P_x, TSB = 1$
mid point nodes $\frac{TSI}{2} = \phi$
end

$$Cw = \frac{TSB}{TSI} \tag{4.11}$$

where TSI remain unchanged through out a communication between subscriber and publisher.

However, from the Subscriber to Publisher, the TSB increases by 1 each time a router node is crossed.

$$\therefore PRE = \gamma_i \times \left[\frac{\sum_{i=1}^{c-(x-1)} N_i}{tN} \times \frac{c}{x} \right] \tag{4.12}$$

as

$$\gamma_i = \sum TSI_i.I_x.P_x \tag{4.13}$$

After obtaining the approximated time a content is cached, then the proxy relationship of the nodes $r_i, r_j, r_k, \cdots, r_x$ drops the content in immediate nodes and probabilistically cache at mid-points using *PRE* concept.

4.4 *PRE* Model for ICN

This section describes the full operation of the PRE model in ICN. The model designed was treated with mathematical backing using the Hypergraph to create the basis for its implementation. PRE was founded to be needed as the research to come up with the full design description of ICN gained more attention. According to the recent studies by [10, 150, 24, 47], redundancy of content is a huge challenge to actualizing the ICN goal. Initial proposal by Jacobson *et al.,[6]* was focused on gaining an Internet that values the content centrism more than the host-based current Internet. These has therefore yearn the re-modification of the Internet architecture by caching chunk of data at router sides. The sole objective being that information monopoly will be reduced using the LCE deployment approach. LCE however, due to the huge amount of content and the heterogeneous demands posed to the Internet accrued large redundancy. Therefore, our designed PRE model aims to mitigate the path redundancy and through lesser deposition of unused data or content to negate the effects of excessive bandwidth and cache catalog blocks for other data traversing.

Given a case in ICN, a Subscriber (user) node is served as the beginning point of communication. When a Subscriber sends interest into the network, the interest content is registered in a Pending Interest Table (PIT) to score the content availability or otherwise. In the even that the score return a positive value, it means the content has been served and thus the intelligent of the ICN build will serve the content from the Content Store (CS). However, the technique of ICN provide the routing flexibility of checking all traversing routers on-path on an even that the interest has not been serve. This thus, according to the proposed ICN [6], using the LCE results in the acquiring of the content interest from other node routers on the network. In ICN, the content is provided by a Publisher (known as host in the traditional Internet) who publishes into the network the data demanded. The data from the publisher is transferred to the Subscriber using the direct hierarchical track of route where the Interest came from. A challenge is thus recorded as all routers on the path cache the content chunk adopting the LCE in the ICN design. This drop copy on every router known as Leave Copy Everywhere results in the over exhaustion of the cache catalog and router memory spaces for other contents that are heterogeneous. Since ICN depend on next neighbor interactivity to fulfill request, the path thus becomes filled resulting to excess content delays and drop of chunks.

Several studies proposed the need to have the replacement policies of content lodged on the routers in order to reduce the amount of replacing important contents in the router. However, as described in Chapter one and two, Least Recently Used (LRU) replacement policy has been the most adequate content replacement that suit the network dynamism [128, 108, 96, 40]. LRU has also been adequate due to its fair consideration and compromise between two vital variables of complexity and overall performance [149]. The need for the elimination or mitigation of the redundancy is therefore seen when the contents are demanded at high and low interval rates on the Internet. The achievement of the model is therefore driven by the need to place cache contents as proxy to subscriber and reverse proxy to the publisher.

4.4.1 Path Discovery and Selection Model

Assume a simple network model which comprises of n number of nodes/router r such that a proper network is set. A subscriber S posed a request q for a content in an ICN topology; such that a resulting publisher P flood the data requested in to the network. To acquire the requested interest q by the subscriber, the data is thus challenged on the path to deliver q. In ICN, some factors need to be considered in order to serve the data [149]. These include

1. Content placement
2. Traffic behavior and pattern [154]
3. Selection of cache content
4. Flexibility of adapting the traffic [154, 155]
5. Consistency of data in the storage (replacement rate).

These challenges are thus a basis to path discovery in serving the content. Therefore, data delivery need to acquire the best path possible in delivering the data and the network redundancy rate of data deposition is vital.

This is thus represented as

$S \Rightarrow q$

where $r = \{r_1, r_2, r_3, ..., r_n\}$ $\forall r$ in ICN

such that a path is set to be acquired and the overall network capacity is computed

as

$$CacheCP = r_i + r_j + r_k + \cdots, r_x \leq P \tag{4.14}$$

where *CacheCP* define the overall capacity of the node on path.

according to [24] an index is set between the distance from the Subscriber to the Publisher as TimeSinceBirth TSB and TimeSinceInception TSI.

Using our proposed hypergraph concept of proximity of the Subscriber and the Publisher, a delicate computation of location from the Publisher to the Subscriber is thus obtained as hyperedges of the network

as

$I_x, TSI = 1$ from the Subscriber side node to $P_x, TSB = 1$ as described in the *PRE* algorithm.

Once the path is obtained, the request is then traversed and the data is set to be cache in the available nodes on the path and the resulting formula approximate the caching time as thus

we define $ApproxTime(z)$= $\frac{\sum_{i=1}^{c-(x-1)} N_i}{tN}$,

where $c = TSI$ and $x = TSB$, $time = t$ and N is the average Cache size along the path.

4.4.2 Verification of the *PRE* Model

The verification of the PRE model is essential as described in Chapter Three. The verification is to confirm that all the representation of the variables in the PRE model are logically added to the codes. The PRE model was tested against the simulation of different network topology. Codes written in Python on the Ubuntu platform was verified to free the model of all run-time errors and bugs. The path description was added to the lines and verified along heterogeneous paths in the ICN. A clear description and Figure 4.2 presents the snapshot of the verified code and the running on of the proposed PRE model against ProbCache on a DTelekom topology.

Figure 4.2. Simulation run on SocialCCNSim

The codes were followed carefully to link the dependent classes of the tested topologies for the verification and simulation as depicted on Figure 4.3.

```
#
from cache_manager import CacheManager

class PRE(CacheManager):
    def get_next_node(self, last, path):
        next = last
        while next >= 0 and not self.topology_manager.has_caching_capabilities(path[next]):
            next -= 1
        return next

    def retrieve_from_caches(self, interest, path):
        content_found_caches = False

        for i in range(0, len(path)):
            p = path[i]
            if self.lookup_cache(p, interest):
                # Proxy position
                content_found_caches = True
                break
            else:
                pass


        if content_found_caches:
            next = self.get_next_node(i-1, path)
            if next >= 0:
                self.store_proxS(path[next], interest)
                self.store_ProxP(path[next-2], interest)

        else:

            next = self.get_next_node(i, path)
            if next >= 0:
                self.store_cache(path[next], interest)

        return (content_found_caches, i)
```

Figure 4.3. PRE model in SocialCCNSim

For verification purpose, the selected popular topologies of Abilene, Tiger, GEANT, DTelekom and Tree were subjected to the path redundancy test. The obtained results were tested and compared to the selected Cache management techniques which are Leave Copy Everywhere and ProbCache. The overall simulation was run on several occasion to obtained a good result so as to completely model the entire network as submitted in [125] known as calibration. Figure 4.4 presents the main SocialCCNSim with all tested topology using the VoD content for the verification. This acquire the node and network relationships on all social graph to form the need ISP topology.

For the validation purpose, some parameters were set for the simulation namely: content categorization, replacement policies, cache sizes and topology selection. Content popularity value was run on different multimedia data. Multimedia data are categorized into Web content, Files, User Generated Content (UGC) and Video on Demand

```
        self.users = {}
        for user in self.social_graph.nodes():
            self.users[user] = User(user)

        logging.debug('Start simulation')

        self.sched = sched.scheduler(time.time, time.sleep)

        # Generate Sequence
        #print "generate sequence"
        if sequence_filename == '':
            self.generate_sequence()
        else:
            self.initialize_scheduler_from_file(sequence_filename)

        # Initialize Caches
        self.lock.acquire()

        caching_strategy_upper = self.conf['caching_strategy'].upper()
        cm = getattr(getattr(__import__('cache_management.%s'%caching_strategy_upper),
caching_strategy_upper), caching_strategy_upper)
        self.caches = cm(
                self.conf['cache_policy'],
                cache_size,
                self.social_graph,
                self.topology,
                self.topology_nodes,
                threshold = None
        )
        logging.debug('Loaded caching strategy')
        self.lock.release()

        self.initialize_catalog()
```

Figure 4.4. Topology Manager in SocialCCNSim

(VoD) [149, 156]. Each of the content differing in their respective popularity value through Zipf distribution [157]. The Distribution of the popularity of the content are on different range as Cisco Virtual Index [3] submitted that about 96% of traffic are content based [149]. LRU [158, 157, 159] was the optimal replacement policy used in the verification due to the submission by researcher of its complexity and a true demonstration of the heterogeneous demands on content on the Internet.

4.4.3 Simulation and Validation of the PRE

PRE model was designed and simulated (using SocialCCNSim [136]) to verify the performance among LCE, ProbCache being the leading studies in the cache manageability of ICN. The simulation was ran on four distinct network topologies to enable PRE compare the redundancy of unsolicited contents in an heterogeneous network. The contents were selected for the verification based on the submission that the Zipf value of VoD ranges at about 1.2 according to [149, 159, 160]. The simulation results shows that the amount of caching operation was reduced using the PRE concept of redundancy mitigation using the open social traffic generator [110].

In Abilene topology using the LRU form of replacement, it was observed that the cache redundant content in LCE was higher than the PRE model due to the operation caching content at all crossed nodes. This logically proved that the redundancy experience in LCE and ProbCache were higher that PRE in the Abilene topology as presented in Figure 4.5. Caching operations define the numbers of data replication as caches on all traversed nodes in the network. The higher the caches, the more likely the excessive redundant data would be lodged on the paths and devices. This thus, causes the aforementioned problems of content and path redundancy.

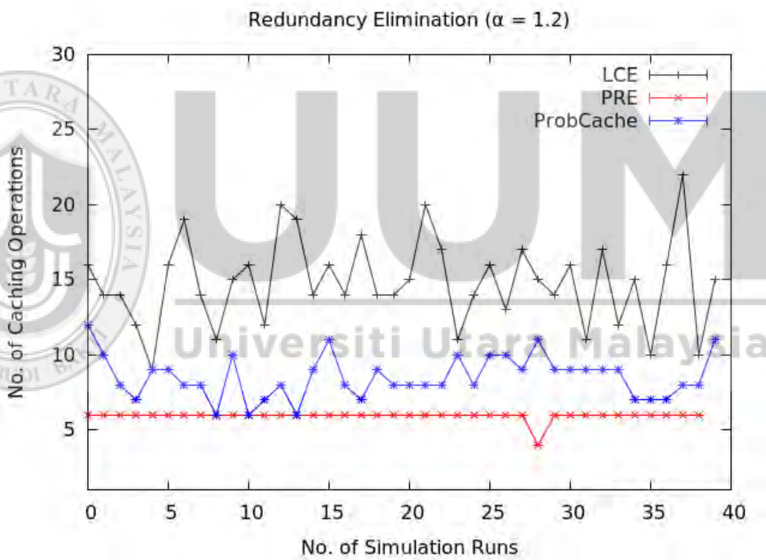

Figure 4.5. PRE, LCE and ProbCache on Abilene

From 4.5, it was observed that the *PRE* concept achieved lesser redundant content on each run of the simulation. ProbCache was also seen to be achieving lesser redundancy of content as against the LCE. However, at some instances of the simulation, the ProbCache was achieving the same numbers of content caching. This was a result of high demand of heterogeneous content and the replacement policy of LRU. ProbCache as described in Chapter two probabilistically cache contents on the nodes depending on the available memory on the nodes.

On the Tree network topology, the simulation results shows higher values of content deposition due to the nature of the topology having more cluster and nodes than the Abilene topology. Consequently, it was also deduced that the PRE model performed better with lesser contents to cache in the entire network. Figure 4.6 presents the results from the Tree topology.

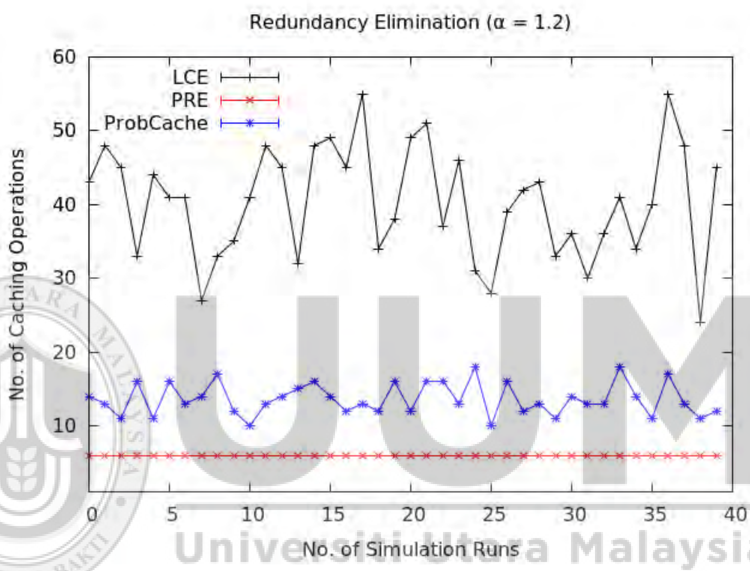

Figure 4.6. PRE, LCE and ProbCache on Tree

Additionally, the result of the PRE was also tested on GEANT and DTelekom as shown on Figure 4.7 and 4.8 respectively.

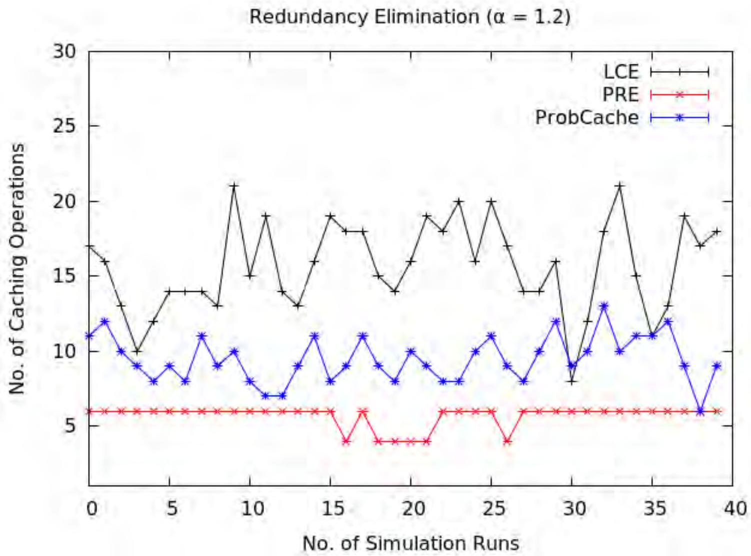

Figure 4.7. PRE, LCE and ProbCache on GEANT

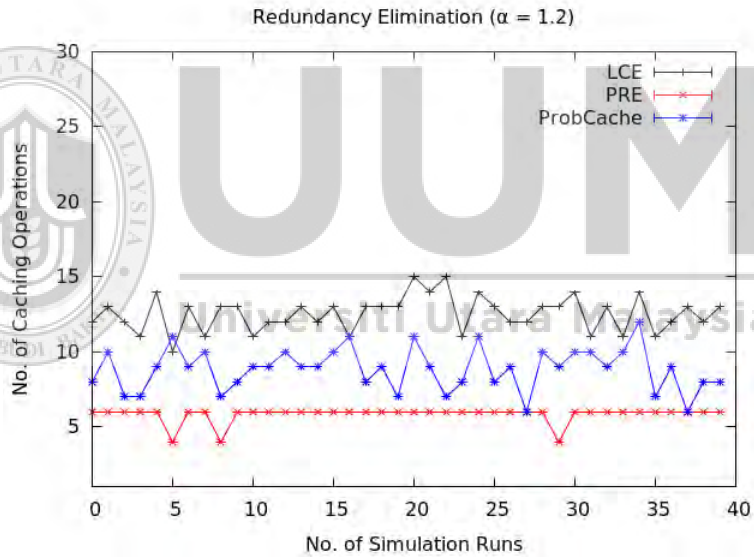

Figure 4.8. PRE, LCE and ProbCache on DTelekom

4.5 Summary

This chapter introduced and described the PRE model by proposing the relationship of content caching on the ICN using four different topologies. In the model, hypergraph concept was extended using the hyperedges to represent a proxy node to a subscriber at the interest end while a proxy to the publisher at the data end. The resulting model PRE was used to test the efficiency of the redundancy elimination of less popular contents. In the verification, the programs were added to the simulator and verified to treat all possible errors.

A better content count was recorded for the PRE model. Hence from the results obtained from simulation against LCE, ProbCache and PRE, the proposed proved the lesser redundancy as the PRE model. It is thus conclusively reached that the overall content redundancy in the network was reduced to about a lower ratio as compared to the LCE and ProbCache. This chapter therefore achieved the first objective of mitigating the path redundancy suffered by the LCE and ProbCache management.

CHAPTER FIVE
CONTENT REDUNDANCY ELIMINATION

5.1 Introduction

Determining the position of network caches has been considered as an important variable for effective cache manageability on the Internet [161, 162, 163]. Several studies have taken keen points in actualizing the positions of cache on the Internet to enhance overall network efficiency and better bandwidth saving [164]. ICN has thus been essential in its design of effective content delivery to the Subscriber. Recent studies by [165, 166, 164] and [50] have hence submitted the importance of content caching on specific location with more emphasis on caching at proxy positions. Cache positioning also referred to as cache deployment strategy [50, 23, 53] could affect the cache-hit ratio, content redundancy, content diversity, network stretch, bandwidth usability and lessen the overall burden of mistakenly replacing popular contents when caches are full.

This chapter presents a proposed cache deployment strategy called *ProXcache* (pronounce as Prowess cache) to reduce content redundancy, cache-hit rate and increase content diversity in the popular ICN cache management schemes. The *ProXcache* shall use the *PRE* model proposed in Chapter Four for assigning the proxy-positions to the Publisher and Subscriber respectively. The outline of the chapter is as thus presented: Section 5.2 highlights the theoretical background in guiding the design of the strategy alongside the cost of coordinating cache, categories of cache-able contents, time of content caching, cache sizes and catalog and the evaluated metrics in the study. Section 5.3 presents the ProXcache deployment strategy including the design and case scenario. The test analysis including the simulation setup are discussed in Section 5.4 and the summary of the chapter is presented in Section 5.5.

5.2 Theoretical Analysis

For clearer perspective into our proposed ProXcache, it is important to note the variables that set the need for cache replication and deployments in proxy position in an ICN. Among the issues to be considered are:

5.2.1 Cache-able Contents

In ICN, due to its cache-driven architecture, various algorithms and strategies have been designed to choose based on the set of rules on the contents to cache. These contents span through varieties of popular contents, content defined names, topological content information among others. For the purpose of achieving the set out objectives in this study, the research shall be caching contents of four categories based on their popularity defined through Zipf distribution [167, 168] as web contents, file sharing, user generated content (UGC) and video on demand (VoD) [149, 142, 139].

5.2.2 Cost of Coordination

Studies in ICN caching by [149, 147, 146, 24] have proven that it is essential to have large cache sizes due to the heterogeneous nature of the Internet. Therefore, there is always need to have a form of replication and replacement flexibility in an event that the caches are full. Cost of coordinating the caches of replacement and replicating contents is vital to free assigned memory through the conditions of recency and frequency in use (popular among are the Least Recently Used and the Most Frequently Used). Fricker *et al.*, [149] studies proved that the flexibility and intelligent information is of paramount importance to solve the cache coordination issue

In MAGIC by Ren et al., [54], the coordination was proposed to handle the content with the highest number of hits. Thus categorizing the content as popular and coordinating its sharing and request. However, in ProbCache by [24], the use of the topological information was most essential as contents are carefully distributed on the paths with lesser congestion and request flooding. Therefore, our concept of ProXcache coordination shall be driven by the proxy-positions of the Publisher and Subscribers end.

5.2.3 Time of Content Phase

In ICN study it is also important to predict the amount of time a content has or need to move from one cache-able device to another which was also an interest in the study by [169]. These time of content residency is usually set by the cache algorithm or cache administrator in some cache deployment strategies (example in Adaptive cache) [50, 161]. In active and adaptive cache deployments, cache contents are manipulated based on some criteria and time duration. This is thus similar to the *quantum time* (q) [62] set in round-robin scheduling algorithm in operating system and grid networks. In network studies, this time is essential and highly important as the contents are updated, the cached contents requires the use of the phase time to re-cache from the original content provider known as the Publisher. This study therefore used the replacement policies along side the elapsing time to re-cache the contents unto the cache-enabled devices.

5.2.4 Cache Strategies

This defines the order of orchestrating the manageability of contents in the network [42, 23, 46]. Some contents are temporarily deployed onto the network while some contents assume permanent deployment according to the algorithm binding the contents. In this study, particularly in this Chapter, the proposed ProXcache would be subjected to ProbCache and LCE caching schemes to investigate cache-hit ratio,diversity, redundancy and bandwidth utilization measures.

5.2.5 Cache Size

Cache size provides the overall memory and spaces available in the nodes of the network for content chunk residency. This node spaces known as cache sizes provide the storage platform temporarily for all contents in the network. It is usually temporal due to the limited sizes and spaces available in caching devices in practical deployments. Caches are sizable with lower capacity as compared to primary and secondary storage. Several researches expressed the cache sizes as multiples of powers. Chai *et al.*, [75, 43], Saino [128] and Cesar [146] expressed the cache sizes as powers (example 2 x 10^2, 10^5). A recent study by Bilal and Shin [42] and Fricker *et al.*, [149] represented the cache sizes as 100, 200, 500, 1000, 10000 each bearing the capacity of elements. While few studies expressed the sizes in percentages.

For the purpose of evaluation, this study shall also adopt the use of 100, 200, 500 and 1000 (representing 1GB to 10GB for clearer presentation) as the capacity of elements in all the chosen topologies of evaluation and taking chunk sizes as 10MB as evaluated by [42, 149].

5.2.6 Cache Hit

Cache Hit ratio returns the amount in average of available content hit (found) when requests are sent [135]. Several studies in database and networking domain have emphasized on the performance of the network through better cache-hit ratio. To acquire a cache-hit rate for a particular strategy, a full note of misses are used for nodes and interests. From the results of the proposed strategies, LCE earned higher cache hit as against the probabilistic ProbCache and ProXcache. However, the study agreed on the good rates acquired in LCE due to the enormous deposition of contents on all traversing nodes in the network but suffers high content and traffic redundancy.

Cache-Hit is therefore define by [135, 146] as

$$Cache\_Hit = \frac{\sum_{i=1}^{|N|} hit_i}{\sum_{n=1}^{|N|}(hit_i + Miss_i)} \tag{5.1}$$

5.2.7 Diversity

In ProXcache, part of the objectives set is to mitigate high number of replicas that might result in content redundancy or overall network overload due to traffic and high coordination rate. With lower diversity tending inversely away from the value 1.0, implies that huge amount of replicas of the content flood several nodes and path. Diversity is thus defined in two folds; namely: Higher diversity implicates the number of heterogeneity which shall benefit the ICN. Consequently, lower diversity implies less heterogeneous accommodation of interests and data. Diversity is defined by [146] as:

$$Diversity = \frac{len(\cup_{n=1}^{|N|} C_n)}{\sum_{n=1}^{|N|} |C_n|} \tag{5.2}$$

where $\cup_{n=1}^{|N|} C_n$ defines all contents Cache node n and the $|C_n|$ defines each content at each node.

Conclusively, higher diversity (towards 1.0 or 100) using the LRU replacement is termed better in performance and redundancy elimination measurement.

5.2.8 Cache Catalog

The catalog size defines the overall number of contents in a network per environment. Literature by leading researchers have also used several notions of presenting catalog sizes. Studies by Chai *et al.*,[75] used 10^3, Rossi *et al.*,[145, 144] considered a YouTube-like catalog, which includes 10^8 objects with an average of 10MB per chunk. Bilal and Shin-Gak [42] used a catalog size as 10^4. Bernardini studies [147] considered a catalog of 10^6 elements, while Mau *et al.*,[170] used catalog value as 10^4.

Our strategy and simulation analysis shall be setting the catalog size to 10^6 according to the flexibility in SocialCCNSim to accommodate and mimic a real ICN network traffic and setting.

5.2.9 Traffic

Traffic in the network has been investigated as a keen variable for the performance and evaluation of a network. To obtain the nature of traffic and the entire relationship of network and the data flooding nodes, hops in traversing messages and information dissemination, traffic characteristics is of the essence [110] and [171]. Traffic is thus investigated through the total number of hops crossed during a request and submission actions. The attributes also include the distances covered by the central requests (Interest) to the data acquiring node (Publisher). Exponential growth of requests posed to the network can further be represented using the following notation for better understanding of our description.

Suppose a network G has $|N|$ nodes and $|L|$ links, there exist an Interest chunk from a Subscriber I_i and a resulting data D_i traversing through node N_i. Traffic is the defined as all forwarded data and chunks through the link $|L_i|$. The nature of traffic hence, gives a better representation of computing the magnitude of clusters and traffic using the *stretch* metrics. Stretch is defined as the amount in ratio of traversing path from Client (Subscriber) to Publisher (Server). While a proportional traffic metric known as hop-reduction ratio is the degree of the nodes in the network that interest from

Subscribers have not crossed.

Stretch is thus given as

$$Stretch = \frac{\sum_{i=1}^{|R|} hops\_hits_i}{\sum_{i=1}^{|R|} total\_hops_i} \tag{5.3}$$

$$Reductionratio = 1 - Stretch \tag{5.4}$$

where $|R|$ =Total number of requests issued by User

and $hops\_hits_i$ =Total number of hops for the request I_i from Source to a node where cache-hit occurs for the i_{th} request. $total\_hops_i$ =Total number of hops between the Subscriber to the Publisher.

5.2.10 Content Categorization with Popularity Models

Content categorization known as popularity function model is essential. This defines the order of content requests. Content popularity is achieved through the measurements of frequency on demand for a particular class of content. Previous studies of the Internet has shown that users were more interested in web files. However, according to the Cisco VNI [3, 1, 30], the distribution of content popularity has changed.

Power law [168, 172] are the most common form of achieving and computing the content popularity. Studies such as [77, 173] submitted that the Zipf distribution is the ideal solution for content categorization on the Internet. This therefore forms the basis of this study to use the Zipf model in the simulation popularity parameter setting.

In the characteristics the alpha α presents the slope of the distribution and q describes the placement. This is why the Zipf is referred to as scale-free approach. Several usage of Zipf was tested in similar network paradigm, example in P2P network by Ripeanu *et al.*, [173] with large sets of content with different popularity. Current studies in ICN and current Internet have categorized the ranges of content popularity to distinct scores. According to Fricker *et al.*, and Bernardini [149, 146], the popularity of the contents range between 0.6 to 2.5; such that the popularity values x is $2.5 \geq x \leq 0.6$. In similar example, Pirate Bay is modeled at $\alpha = 0.75$, Daily motion $\alpha = 0.88$. Additionally, large samples of Video on Demand (VoD) in China has α ranging between 0.65 to 1.0 according to Fricker *et al.*, [149].

Therefore, this research shall be adopting the categorization of the range of contents popularity into four sets. Each set distinguishing the other in its value; namely: File Systems $\alpha = 0.75$, Web contents $\alpha = 0.80$, User Generated Content (UGC) $\alpha = 0.88$ and the Video on Demand (VoD) $\alpha = 1.2$. The Probability Mass Function (*pmf*) is given as

$$pmf(x,N,q,\alpha) = \frac{(q+x)^{-\alpha}}{\sum_{k=1}^{N}(q+k)^{-\alpha}} \tag{5.5}$$

where N = Number of elements, k = rank and q define the Zipf plateau and α is the exponential.

The summary is therefore depicted on Figure 5.1 from the studies of popular researchers in ICN.

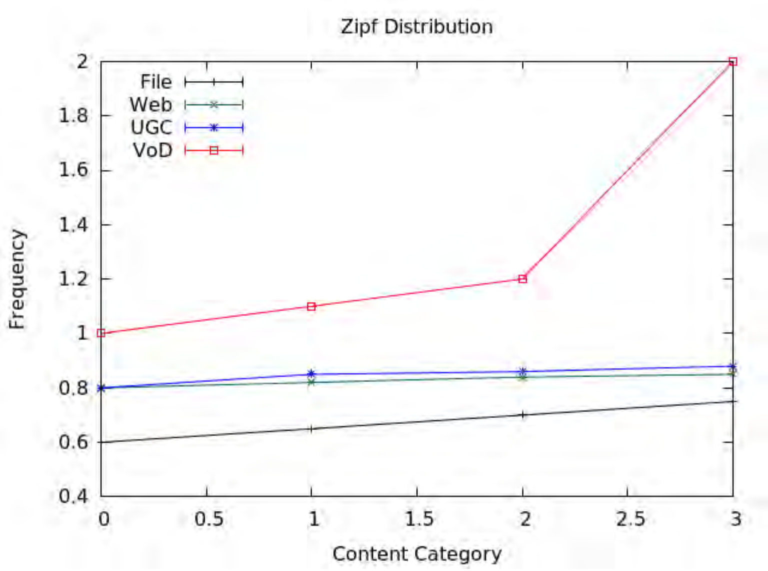

Figure 5.1. Zipf Content Categorization

In reference to related studies [149, 42, 174, 144, 47], it is worthy to mention some studies that explored the Zipf parameter in ICN cache managements. Studies conducted by Psaras *et al.*, used the media driven content with the parameter being $\alpha = 0.8$. The study by Saino *et al.*, explored the ranges from 0.6 to 1.1. Similarly, the α parameter was taken between 0.65 to 2.5 in Rossi [144] and [175]. Wang *et al.*, [94] in an autonomous ICN setting used the values between 0.7-0.96 while Ren *et al.*, [54] took the values between 0.7-1.1.

It is thus conclusively deduced that there has not been fixed parameters for certain categories except as taken in several studies [149]. However, according to the ICN baseline and [149], VoD has higher traffic generation due to users tending their interests in media-driven information ahead of web content, file and UGC [33]. Additionally, alpha value has tremendous effect on the overall results through the better demand for diversified content in the heterogeneous network.

5.3 ProXcache Strategy (Prowess cache)

ProXcache (*prowess cache*), proposed strategy for ICN. The proposed ProXcache is mainly focused on the position and centrality of cache routers and the flexibility of caching at proxy positions after a hit has occurred. This aims at mitigating the overall flooding of less referred or less needed deposited data on the paths of network data traversing. ProXcache demonstrate the need to reduce content and path redundancy through the approach of hypergraph proposed in Chapter Four of this study. In ICN LCE strategy (the initial submitted paradigm structure), users (Subscribers) are posed at a position in the network. The Subscribers send their Interests through the node directly neighboring their position. Once the node receives the Interest, the Interest are check in the Pending Interest Table (PIT) to confirm its identification and relate the request to the already saved contents through request history. Once the data is found in the PIT to be positive, the data is looked-up in the Content Store (CS) and forwarded to the Subscriber through the same route of request.

However, if the request is not found, the FIB is charged with the operation of forwarding the request to the available node to serve the Interest. The need for ProXcache becomes paramount as the default ICN caching strategy results in excessive data deposition on all traversed and crossed nodes. This as a result creates the overloaded traffic and redundancy to the network. Given the dynamic structure of the Internet and acknowledging the predicted future traffic (VNI [3, 1, 30]) redundancy of data needs to be palliated to benefit users sending heterogeneous requests at about unpredicted time intervals.

To this end, the ProXcache shall be exploring the path redundancy elimination model in Chapter Four and carefully caching contents on Subscriber and Publisher proxy positions; thereby probabilistically caching at a central position [75, 43] on the network to mitigate content redundancy.

5.3.1 ProXcache Design

The proposed ProXcache incorporates a general traffic prediction to model a true network. The model took into cognizance the nature of Interests demanded and the time of approximating a replacement policy. Replacement policy defines the order in which a content is replaced (pre-empted) out of a cache-able device due to time allotted, algorithm definition and the exhaustion of memory and catalog sizes. In ProXcache, every node has a distinct identification through the point and direction of requests for a particular data/Interest. The functionality shall test and predict the overall network node cache-able routers on the path of requests. This will aid in approximating the overall capacity of the network and return the values of neighbor node distances to create the proxy-position. During the ProXcache operation, probabilistic mid-node caching operations shall be added to increase the chances of data subscription coming into the network but logically far from either the Publisher or Subscriber proxy. This is the basis of carefully presenting the ProXcache with the capital letter 'X' denoting the central probabilistic position whenever a request is sent.

5.3.2 ProXcache Case Scenario

Given a network G, there exist several nodes for content requests and data serving. To aid the proposed concept in comparison, ProXcache shall be compared along LCE to verify its contribution through path redundancy mitigation and compare the cache-hit metric against the popular ProbCache concept [24].

Suppose a network with ten (10) nodes named R1 to R10, assuming the content class is a VoD according to [149] as described in Section 5.2.10. Subscriber S1, and S2 request heterogeneous content at an unpredicted time intervals (See Figure 5.2).

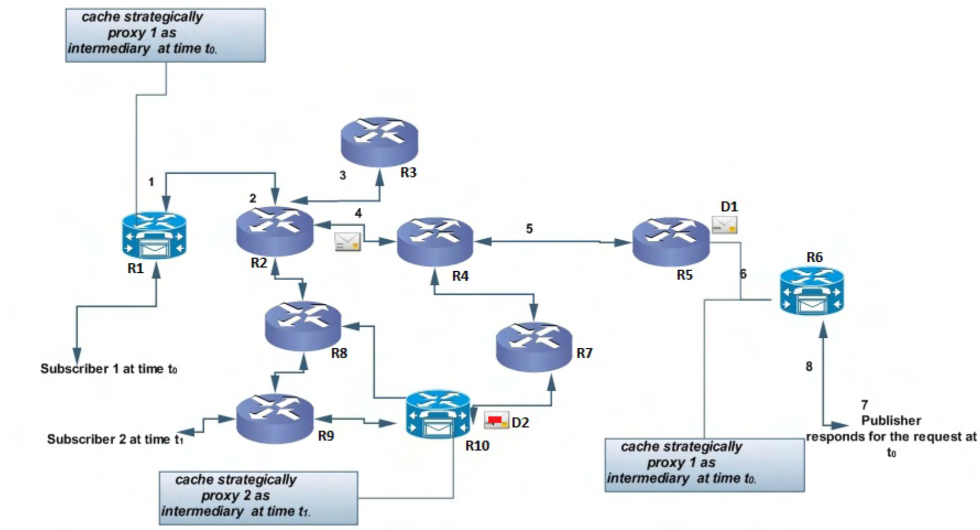

Figure 5.2. Network Case 1

Subscriber 1 at time t= 0 sends an Interest of data D1 which traversed through R1, R2, R3, R4, R5-R6. The record is checked and found in R6, using the *PRE* model the algorithm 4.5, the path is computed and the approximated cache capacity is stored. Path description and traffic distance is returned as described by Psaras *et al.*, in [24] as *TSB*. R6 (Blue with a data) then returns according to the hyper edge properties of a graph with its degree of other attached nodes. Such that it is assumed as an edge by all path nodes R1-R5. The concept thus produce a cache-proxy operation for R5 such that it clears paths for subsequent requests that might require the path and also predicatively relief R5 of caching further contents. R6 serving the request of S1 can logically have longer time to keep data D1.

Consequently, serving data D1 to R1 requires the ProXcache concept not to cache on any traversing node except a mid point between R5-R1; which is then followed by the Subscriber proxy position before finally dropping the content to Subscriber S1. Objecting the normal practice in ICN caching using LCE, all traversing and crossed nodes are expected to cache D1. Additionally, if Subscriber S2 sends a request at time t = 1 through R9 and the data D2 was to be served by R10, then using the *PRE* and ProXcache, similar operation shall be required to satisfy the request of S2. However, a keen point to note here will be the reconfiguration of the ProXcache operation which will use the Path-capacity approximation to make a neighboring node a proxy to a Publisher and Subscriber. For every operation, depending on the availability of memory spaces, the mid points are recomputed as the does the proxy positions. This thus reduces the overall path and content redundancy suffered by the LCE strategy and decreasing the rate of cache eviction due to memory utilization burden and further improves the overall content diversity in ICN.

This flexibility will then improve the effective use of memory, save bandwidth and mitigate the excessive caching operations in ICN as compared to the LCE. Furthermore, since ProbCache initial objective was to reduce redundancy suffered by LCE, it in-turn scored lesser cache-hit ratio. Therefore, the ProXcache aims at improving the cache-hit ratio reduced in ProbCache by dynamically creating the proxies for the Publisher and the Subscriber.

5.3.3 ProXcache Simulation Case

For effective evaluation, ProXcache shall be using the CCNSim platform as the underlying simulation with overall case simulations on SocialCCNSim [136]. For the caching operation, LCE was designed in SocialCCNSim which shall be one of the strategy to compare with. For the replacement operations, LRU was adopted as the optimal replacement algorithm in most ICN cache studies. LRU was agreed as the replacement policy due to its better trade-off and consideration ahead of other policies in terms of complexity and overall performance. According to studies by [149, 54, 144, 175, 40] and [24], Zipf distribution was used in the traffic and content popularity selection. In this test case, File sharing, Web and Video on Demand (VoD) was used to investigate the performance of ProXcache along other strategies on the

Abilene topology.

Table 5.1 presents the parameters used in SocialCCNSim for the evaluation of the proposed strategy.

Table 5.1
Simulation Description Value

Parameter	Value/Description
Runs	10
Chunk Size	10MB
Cache Size	100, 200, 500, 1000
Catalog Size	10^6
Alpha value	1.2
Topology	Abilene, GEANT, Tree, DTelekom
Simulator	SocialCCNSim
Traffic Source	SONETOR (Facebook [176, 113])

5.4 Simulation Setup and *ProXcache* Test

In this study, a proposed Cache Deployment Strategy was presented. The proposed strategy aims at curtailing the slated challenges faced by the current design of ICN through mitigating content and path redundancy. In order to investigate the performance of the ProXcache, a Social graph [177] comprising of several users was subjected to the usual operation of content request as Interest and serving by holders of the data was simulated. Users in the social network was mapped to exhibit social relationship as the demand for contents and network relationship that exist between nodes. For the benefit of the simulation, the large set of users were 4,039 with 88,234 friends across the network known as edges [113, 177]. For the purpose of the simulation, parameters set were the number of runs, cache sizes were carefully selected as 1000.

The Catalog Size and topologies tested against were as depicted on Table 5.1.

For comparison purpose, ProXcache was tested against the LCE and ProbCache to investigate the effect of content redundancy, cache-hit ratio and the diversity on Abilene. Inline with the huge demands on media data, the Zipf popularity used was set at 1.2 (VoD) according to Fricker *et al.*, [149] categorization. The Figure Social Network Graph presents the pictorial demonstration of what the Facebook [177] social network looks like. Each node having social and Network relationship through content popularity demand and the network link inter-connectivity.

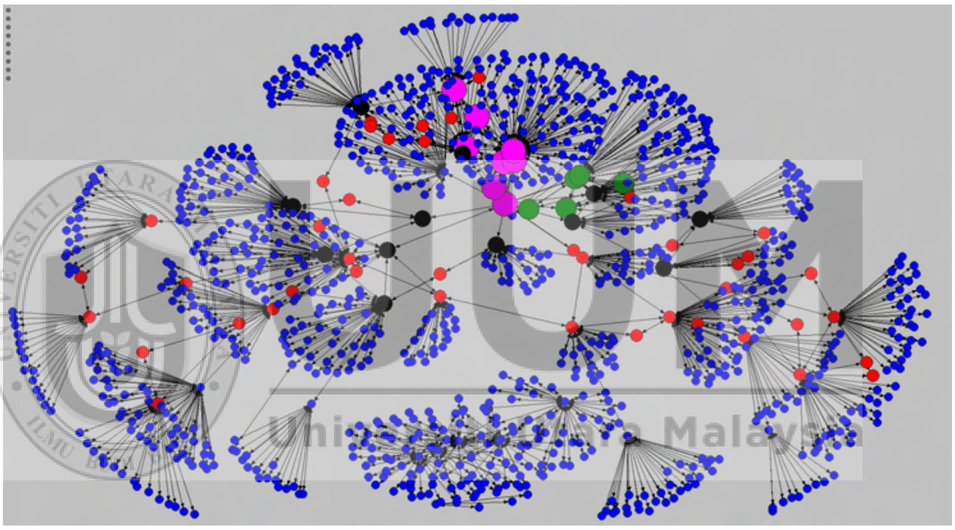

Figure 5.3. Social Network Graph [178]

5.4.1 Simulation Result for Content Redundancy

As presented in Subsection 5.2, redundancy of content is the excessive caching on traversing nodes which result into less or no referral to the content. This as a result affect the performance of the network and result into unnecessary delay of content coordination and high number of eviction to accommodate new heterogeneous content in the network.

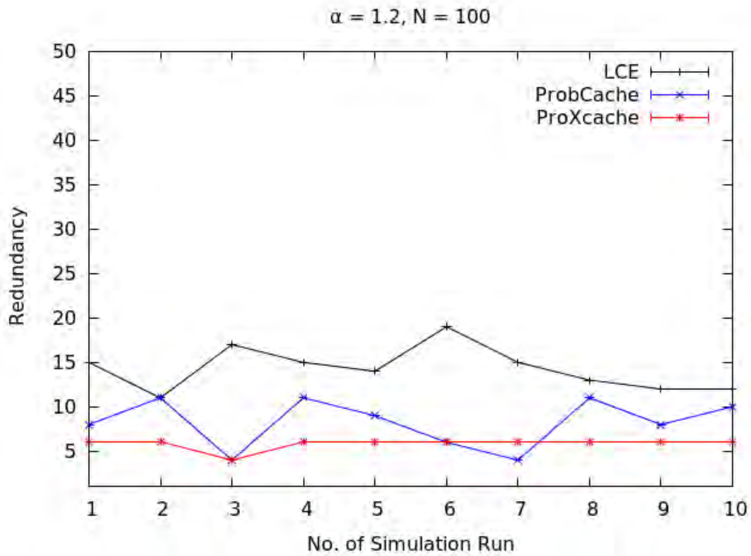

Figure 5.4. Content Redundancy on Abilene *N=100*

The result presents the relationship between the tested LCE, ProbCache and the ProXcache in terms of content redundancy using the cache size *N=100,* $\alpha = 1.2$. The result presents the overall content deposition on the nodes; as described in Chapter 2 and Chapter 4, the LCE was observed to have higher content deposition due to its design of caching on all nodes crossed. ProXcache was observed to clear the high traffic incurred by LCE. However, the ProbCache performed better than LCE and also had a value competitively lower than ProXcache at 7 run. The result therefore showed that the ProXcache reduced the redundancy at about 55%.

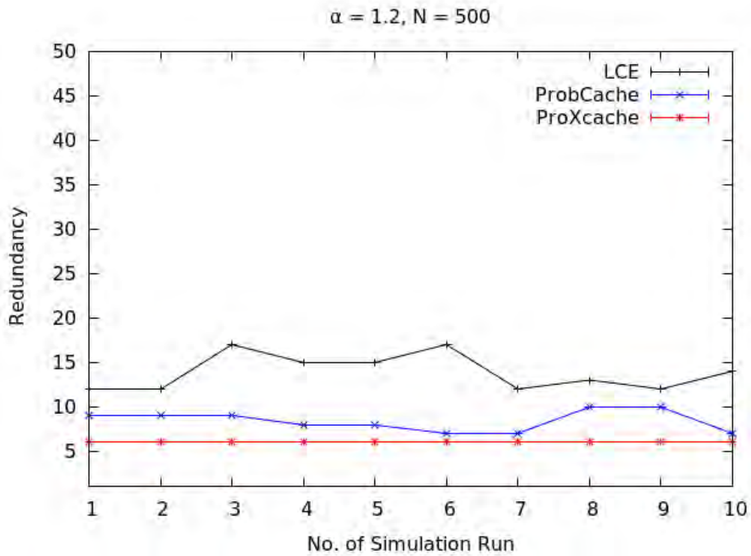

Figure 5.5. Content Redundancy on Abilene *N=500*

Increasing the cache size to 500, a high impact was displayed with high redundancy due to the increase space for contents to flow and choose more cache-able nodes. However, the ProXcache still recorded a better redundancy score compared to LCE and ProbCache.

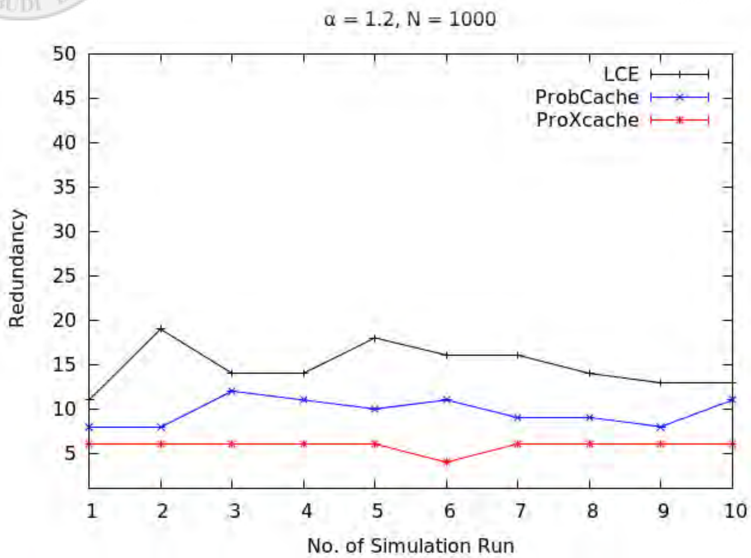

Figure 5.6. Content Redundancy on Abilene *N=1000*

5.4.2 Simulation Result on Diversity

Diversity defines the unique caching of contents on selective nodes. It provides the score ratio that is tending toward 1.0 for distinct contents and less than 1.0 for similar contents and replica.

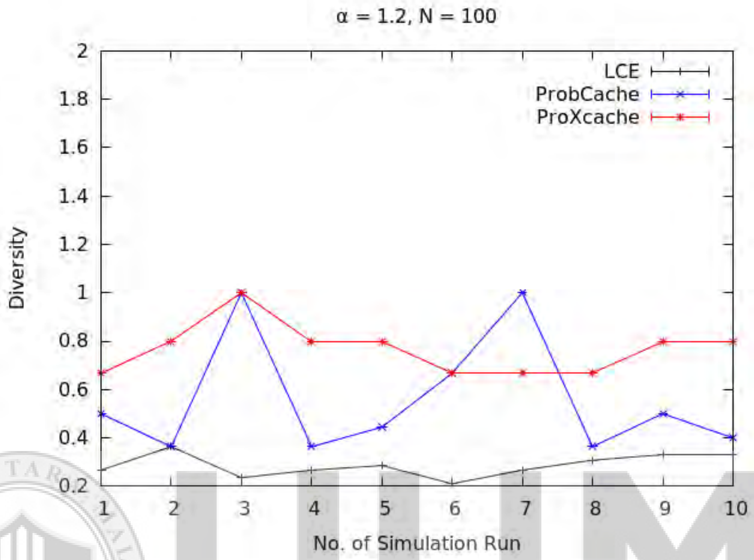

Figure 5.7. Diversity on Abilene *N=100*

From the result of the simulation, the LCE had lower diversity. This implies that the replica in LCE was far higher than that of the ProXcache. This reveals that as the cache size was low, the diversity on LCE was weaker while in ProXcache the value hit the distinct uniqueness at run 3 while in ProbCache the diversity of 100% was observed at run. This implies the path was a bit clearer that ProXcache for distinct caching content.

As the cache size increase from 100 to 500, the diversity of the concepts were unfavorable. However, the ProXcache still recorded a better diversity ratio. Most of the diversity ratio was mostly 0.8 in ProXcache and the LCE going as low was 0.3 while ProbCache was lowest only at the diversity ratio 0.4

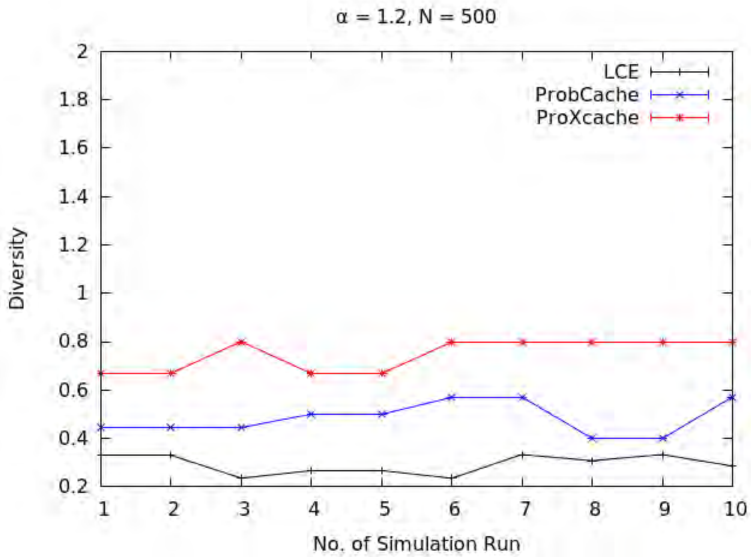

Figure 5.8. Diversity on Abilene *N=500*

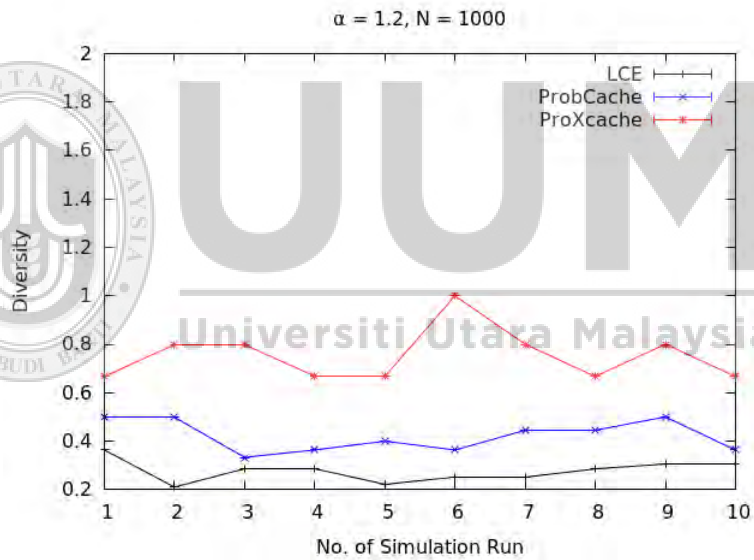

Figure 5.9. Diversity on Abilene *N=1000*

Increasing the cache size had negative effects on LCE as the number of replica contents increased thereby earning lower diversity while ProXcache had 100% at run 6 due to the large flexibility that lead to more options to cache distinct contents.

5.4.3 Simulation Results on Cache-Hit ratio

For the benefit of comparison, the results of cache hit is shown on Figure 5.10. When the cache size increases, the overall performance of the network seen to be better except for the ProbCache due to its probabilistic nature of caching on a free probability position node. Figure 5.10 (a) presents the result obtain on the Abilene using the $\alpha = 0.75$. The runs depict the cache size starting from 10 to 1000 as 1 to 10 respectively. Each value on the x-axis represents a multiple of 10 as the cache size. The cache size at position 1 (1GB), the cache hit ratio was about 0.0476 on LCE against the 0.0332 in ProbCache and 0.0385 for ProXcache. However, increasing the sizes of caches provides better hit rates.

Consequently the Figure 5.10 (b) presents the result obtain on the Abilene using the $\alpha = 0.80$ of the Web contents. The graph shows close relationship between the ProbCache and the ProXcache. However, the result of the ProXcache is seen to be slightly competing with the ProbCache as the cache size increases. The improvement due to the few heterogeneous deposition of contents, the ProXcache was able to grant better hits as compared to the LCE that almost cache similar contents at every request.

Figure 5.10 (c) depicts the results of the simulation ran on the UGC content category. It was not far away from the previously tested content categories as the Abilene topology was used also. As the cache size increased, it was also noted that the ProXcache improved due to its design of caching at proxy positions and probabilistically caching on the mid parts. For the VoD on Figure 5.10 (d) content, the amount of caches was determined in relation to the fast replacement of contents in LCE and the exhaustion of paths due to the caching operations performed on each node.

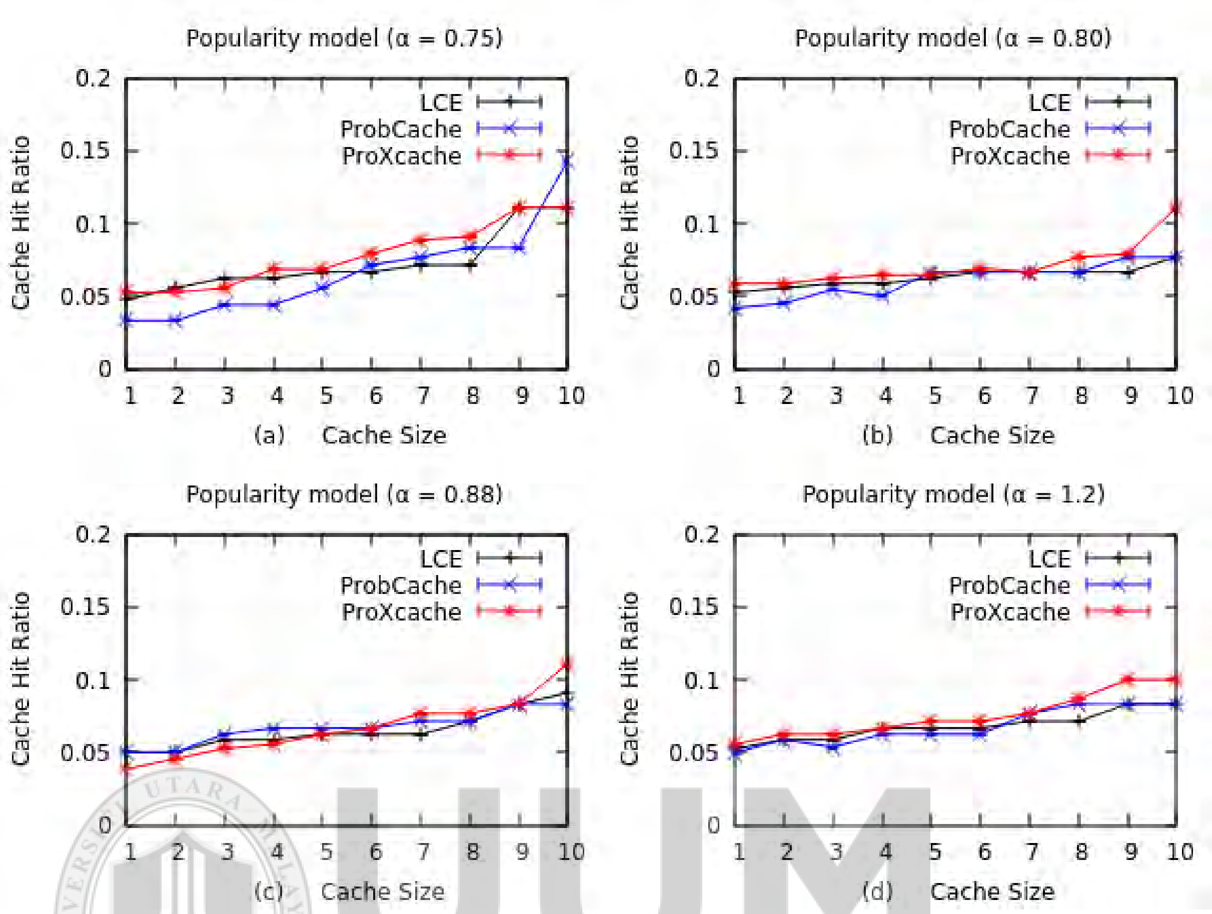

Figure 5.10. Cache Hit Simulation Results on Abilene

This has yielded the lower cache hit ratio as compared to LCE and ProXcache. The result provided that the nature of VoD demand on Abilene topology is complex as compared to file sharing with the alpha parameter 0.75 to be compared with in Chapter 6.

5.4.4 Results Summary on GEANT and Abilene

5.4.4.1 Content Redundancy

Content redundancy is an important metric to evaluate our ICN heterogeneous data dissemination. It was deduced that LCE suffers huge deposition of replicas that affect the throughput and easy content delivery to Subscribers of different content category.

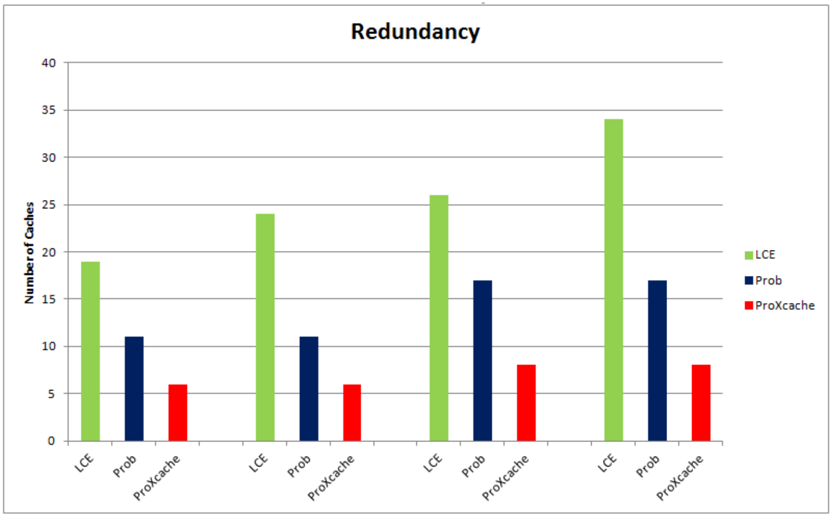

Figure 5.11. Content Redundancy Summary on Abilene

The result on Figure 5.11 shows the obvious reason of high caching operations suffered by LCE. ProbCache was observed to minimize the huge data deposition using the probability of available cache spaces for its caching operation. However, ProXcache performed better by limiting the replica contents at every instance of interest flooding in the network.

5.4.4.2 Diversity Summary

For comparison and to enable performance check, GEANT and Abilene topologies were used to test the performance of the proposed concept in content diversity for the cache sizes 100, 200, 500 and 1000. The results from Figure 5.12 shows that the diversity of the ProXcache is better than the LCE and ProbCache. The result also reflects the better diversity in ProXcache to the nature of degree interconnection of nodes in the GEANT topology. This made it flexible for the PRE model use of the ProXcache.

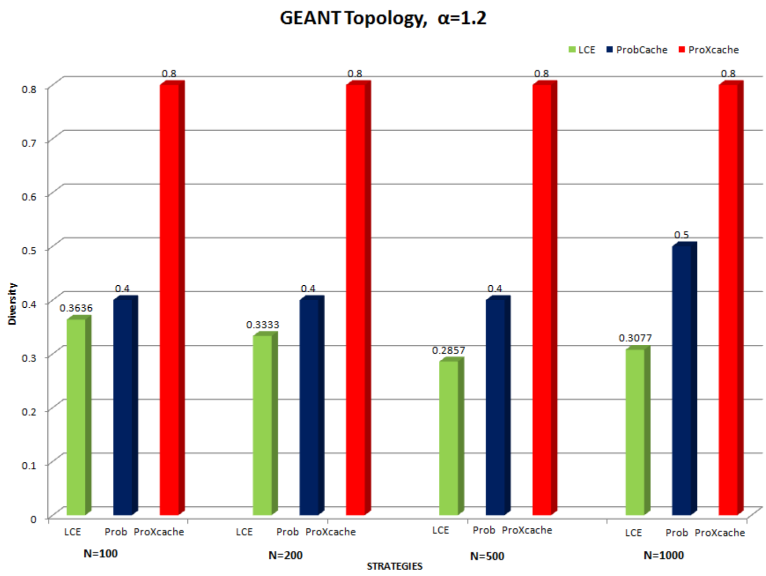

Figure 5.12. GEANT topology summary on Diversity

Diversity of the contents implies that content redundancy has also been minimized in our strategy using the 1.2 alpha parameter and testing against variable cache sizes (N=100, 200, 500 and 1000) as depicted on Figure 5.12.

5.4.4.3 Cache-Hit Rate Summary

Increasing the simulation runs and presenting the summary shows that the LCE has better cache-hit similar to the analysis performed on the Abilene topology using the File sharing. However, our strategy outperforms the ProbCache strategy in all cases. The sizes were constantly increased to accommodate more heterogeneous content. This proves that the amount of hits in ProXcache is better compared to the ProbCache.

The results summary presents the values as the topology changes, the structure has an obvious effect on the number of hits. This is attributed to the simple hierarchical structure of the Abilene network topology. The result shows that part of the research objective was to improve the performance of ProbCache. This proves that if proxy-positions are carefully added to the ProbCache strategy, the amount of bandwidth wastage and excesses could be reduced thereby improving the overall throughput.

5.5 Summary

The Chapter presented the ProXcache deployment strategy. The strategy adopted the LRU replacement policy and was designed and simulated in SocialCCNSim. GEANT, and Abilene network topologies were used to test the efficiency of the proposed concept against the popular LCE and ProbCache. The results obtained proved that ProXcache performed better than the LCE and ProbCache in terms of content redundancy elimination. Content redundancy was reduced thereby implying the amount of caches on the path was reduced to accommodate the free caching of other interest request. The results from the simulation also shows that the Diversity of contents was improved as LCE seem to cache higher number of replica. ProbCache was observed to have a good diversity when cache sizes increase.

ProXcache can thus be seen as an alternative to ICN practice in the future Internet of a deployed network case with large media-driven data of heterogeneous categories.

CHAPTER SIX
SIMULATION ANALYSIS AND EVALUATION

6.1 Introduction

The simulations of path and content redundancy as presented in Chapter Four and Five have yielded positive results. These have necessitated the interest of the study to try more extensive simulations with varying parameters to re-establish the findings in this study. This chapter shall be taking note and presenting the results obtained on different parameters and topologies. Section 6.2 presents the simulation and evaluation of content and path redundancy for file documents, web contents, user generated content and the video on demand. Section 6.3 details the information of the simulation phase to investigate the ProXcache cache-hit ratio. The analysis of diversity on different content categorization was covered in Section 6.4. The effects of diversity was investigated on Abilene, GEANT and DTelekom network topologies for different sampling.

To carefully evaluate our strategies, the simulation was selectively started on ICN-SIM [129, 130] as depicted in Figure 6.1 to verify the management and pattern of ICN communication. The simulation was carefully set to match the specification of this study. The only limitation was the lack of flexibility to reconfigure the needed caching parameters. This thus prompted the research to extend the exploration of ICN in SocialCCNSim [136, 127] owing its background to the CCNSim [126].

Figure 6.1, presents the typical ICN topology and the Subscribers, Publisher relationship in this study. The flexibility grant the relationship between nodes and the stations as Interests and data flood the communication to establish the social relationship through contents request and the network relationship through link and connectivity as presented by [111].

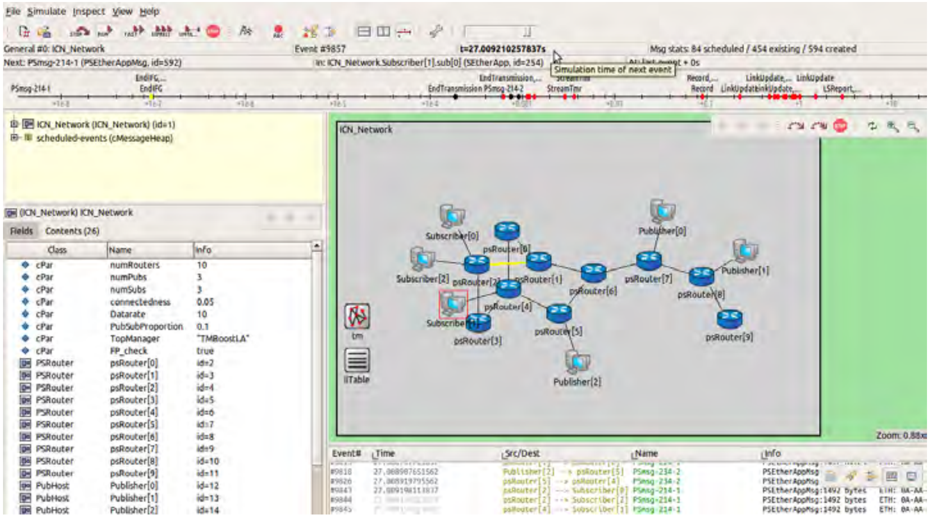

Figure 6.1. Information-Centric Network Simulator

6.2 Simulation and Evaluation of Content and Path Redundancy

Path redundancy as we have itemized was the major parameter of the study in regards to evaluation. As achieved in Chapter Four, the analysis was done using the VoD data category. This as such, is inline with the common data transmitted and highly shared on the Internet. This was also accordingly chosen based on the Cisco Virtual Network Index 2015 [1] that states that the source of traffic shall be media-driven data in the future and most of the traffic shall be subjected through smart devices.

In this evaluation, the study ran the strategies in different data category; viz: File sharing $\alpha = 0.75$, Web data $\alpha = 0.80$ and Video on Demand (VoD) $\alpha = 1.2$. The results of the simulation shows that the amount of data traversed are highly demanded on the VoD and the redundancy experienced in VoD was enormous in terms of path exhaustion.

6.2.1 Simulation for Path Redundancy

The path and content redundancy was evaluated and compared with other strategies. The simulation was subjected to distinct contents types for better evaluation. The results of the sets on different topologies varied due to the structure and the node network relationship that exist in the network. The result was evaluated on Abilene for the Figure 6.2 (File sharing $\alpha = 0.75$) subjecting the tests to varying cache sizes from 1GB to 10GB (1-10). It was observed that as the cache size increases, there exist a direct proportionality in the amount of data deposited as caches on the nodes. This resulted in the excessive cache deposition of LCE. However, this study enhanced the overall performance by proposing the ProXcache strategy.

The overall results as presented in Table 6.1 depicts the rate in percentage the amount of redundancy incurred on the three tested strategies on $\alpha = 0.75, 0.80, 1.2$ respectively. With the cache size increasing, LCE almost exhausted the path of request. This was part of the reason why LCE returns lesser cache hit rates on heterogeneous content request.

Table 6.1
Content-Path Redundancy on File sharing with $\alpha = 0.75$

ProXcache	**ProbCache**	**LCE**
0.341	0.311	0.490
0.207	0.311	0.658
0.282	0.276	0.658
0.335	0.301	0.860
0.260	0.417	0.860
0.293	0.487	0.860
0.335	0.641	0.99
0.412	0.694	1.0
0.444	0.781	1.0
0.444	0.968	1.0

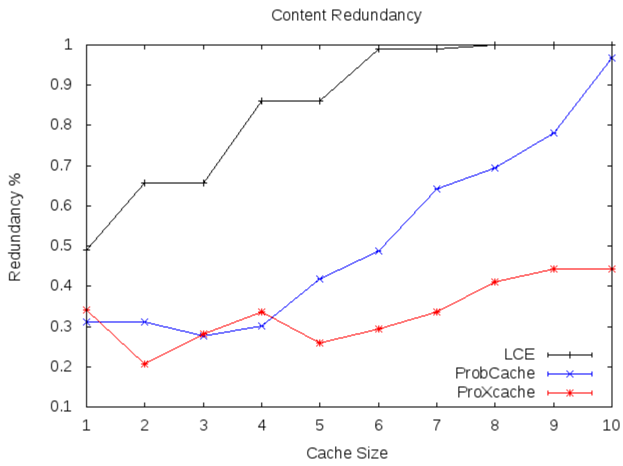

Figure 6.2. File sharing with $\alpha = 0.75$

Consequently, the results were also evaluated for the Web Content request. The File sharing $\alpha = 0.75$ results proved that the demand for the Web content recorded more popularity than the Files ($\alpha = 0.75$). The results presented on Table 6.2 and Figure 6.3 provides the data in rates of percentages of the effects of caching rates that resulted into path redundancy after high number of caching operation. ProXcache minimized the total number of caches in the overall simulation.

Table 6.2

Content-Path Redundancy on Web Content with $\alpha = 0.80$

ProXcache	**ProbCache**	**LCE**
0.300	0.281	0.784
0.278	0.515	0.784
0.522	0.546	0.833
0.313	0.660	0.860
0.294	0.593	0.860
0.335	0.485	0.926
0.300	0.593	0.881
0.451	0.445	1.0
0.512	0.615	1.0
0.571	0.809	1.0

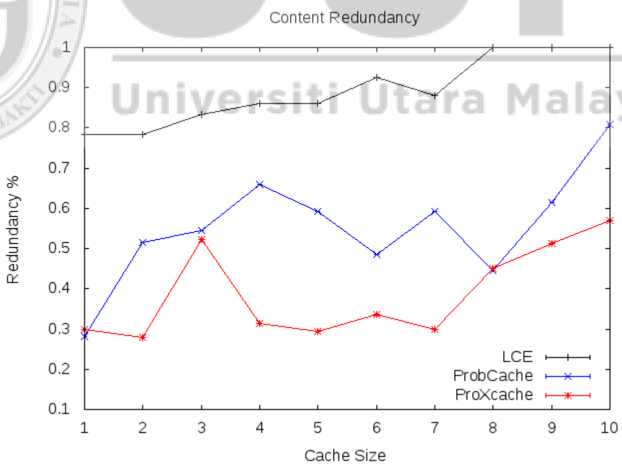

Figure 6.3. Web Content with $\alpha = 0.80$

VoD was also evaluated on the Abilene topology to test the effects of the caching everywhere of the LCE and the ProbCache against ProXcache with varying Cache Sizes. The results shows the high connectivity and caching of VoD to be higher than the caching operation in File and Web data. The redundancy of data on LCE was high as the Cache Size increase to 5GB. This is attributed to the large demand of media-driven data predicted on VNI 2015 [179, 1] and [2].

The amount of redundancy thus affects the number of content hit as the replacement policy is frequently called upon and high replacement of contents are observed in LCE and ProbCache as against the ProXcache. Table 6.3 presents the results of the evaluation see Figure 6.4.

Table 6.3
Content-Path Redundancy on Video on Demand with $\alpha = 1.2$

ProXcache	ProbCache	LCE
0.573	0.748	0.994
0.371	0.880	0.990
0.504	0.537	1.00
0.572	0.682	1.00
0.445	0.75	1.00
0.500	0.681	1.00
0.532	1.00	1.00
0.612	1.00	1.00
0.769	1.00	1.00
0.769	0.908	1.00

6.3 Simulation and Evaluation on Cache-Hit Rate

For the purpose of comparison and evaluation, the studied strategies were subjected to various run varying the parameters as shown on Path redundancy. For the cache-hit rates, File, Web, and VoD was evaluated on Abilene, GEANT and DTelekom.

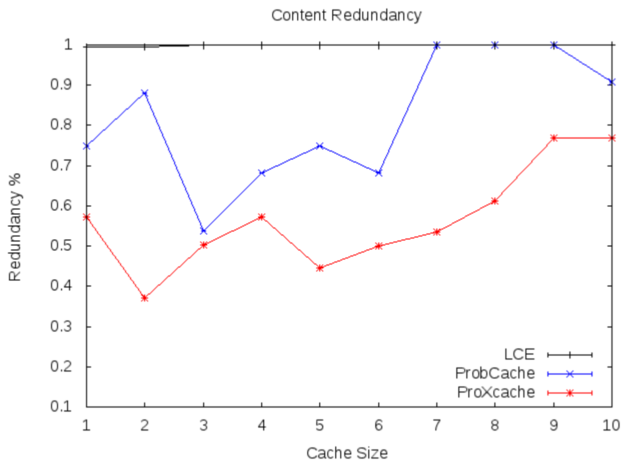

Figure 6.4. Video on Demand with $\alpha = 1.2$

6.3.1 Cache-Hit Evaluation on File Sharing with $\alpha = 0.75$

The following results presents the summary of evaluation carried out on File sharing content category.

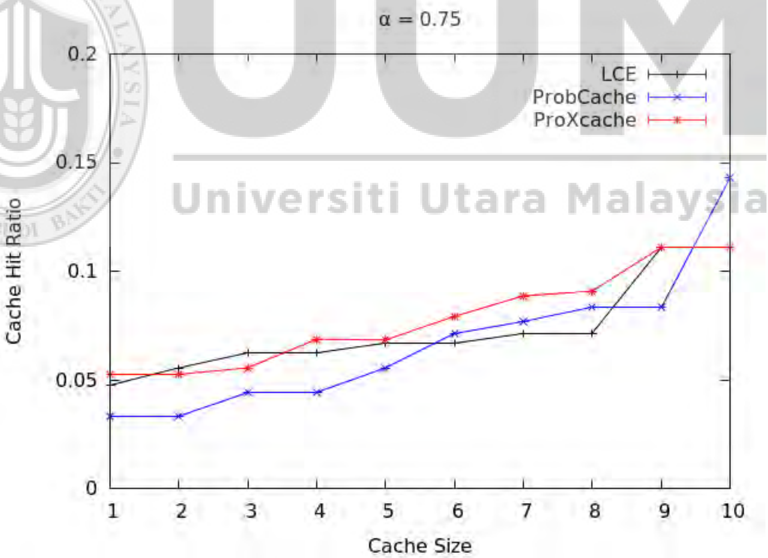

Figure 6.5. Cache-hit Case 1in Abilene topology

When cache sizes ranged from 1-1000 (1GB to 10GB), the performance proves that with the added concept of path redundancy elimination proposed in Chapter Four, cache-hit rate in ProXcache was improved. However, LCE due to its large set of caching contents on path recorded higher cache-hit rates. This is therefore an advantage in a network with lesser heterogeneous contents. The result presented on Figure 6.5 was simulated on the Abilene topology. The structure and hierarchical nature of the topology had effects on the rate of cache hit. Content category is the ordinary File ($\alpha = 0.75$) for the first part of the evaluation and comparison in cache-hit.

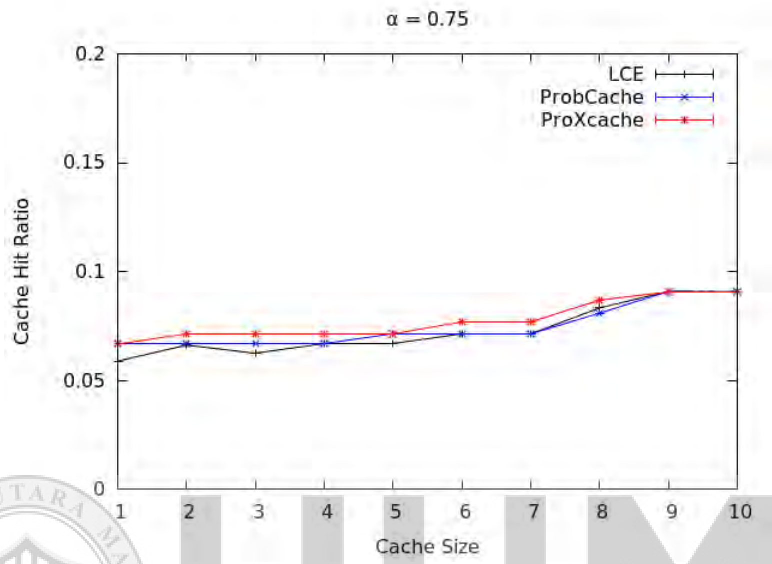

Figure 6.6. Cache-hit Case 2 in DTelekom topology

With cache sizes of 1-1000, the cache-hit ratio of the proposed ProXcache performed almost at a competitive rate to the LCE. The simulation was subject to a different network topology DTelekom. The cluster based DTelekom was used to investigate the potential of the proposed strategy as benefited both the LCE and ProbCache. It was observed that the ProbCache had the same parameter on cache size 1000 as depicted on Figure 6.6. A slight variation was observed in the events of the simulation on GEANT topology as shown in Figure 6.7.

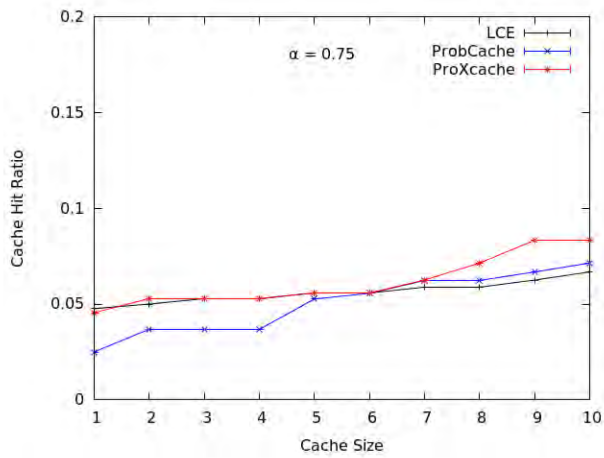

Figure 6.7. Cache-hit Case 3 in GEANT topology

6.3.2 Cache-Hit Evaluation on Web Content with $\alpha = 0.80$

The evaluation was also performed on different topology to juxtapose the performance of ProXcache against the selected caching strategies of LCE and ProbCache. The cases are on Abilene, GEANT and DTelekom.

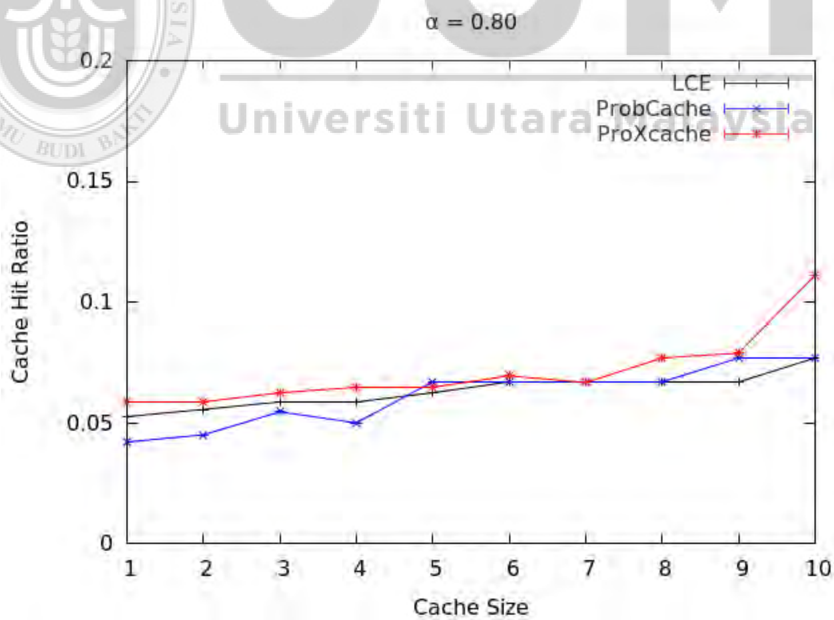

Figure 6.8. Cache-hit Case 4 in Abilene topology

Figure 6.8 presents the results obtained using the Abilene topology. The result shows that ProXcache outperform the LCE and ProbCache to deliver heterogeneous contents to Subscribers. As the Cache Size increases, so is the rate of hits in the Cache. At the lower memory, the ProbCache recorded lower cache-hit rates due to the Probabilistic nature of the design to available spaces. However, as the cache size increases the rate of content replacement was reduced and thus yielded in improve results against the LCE for heterogeneous requests.

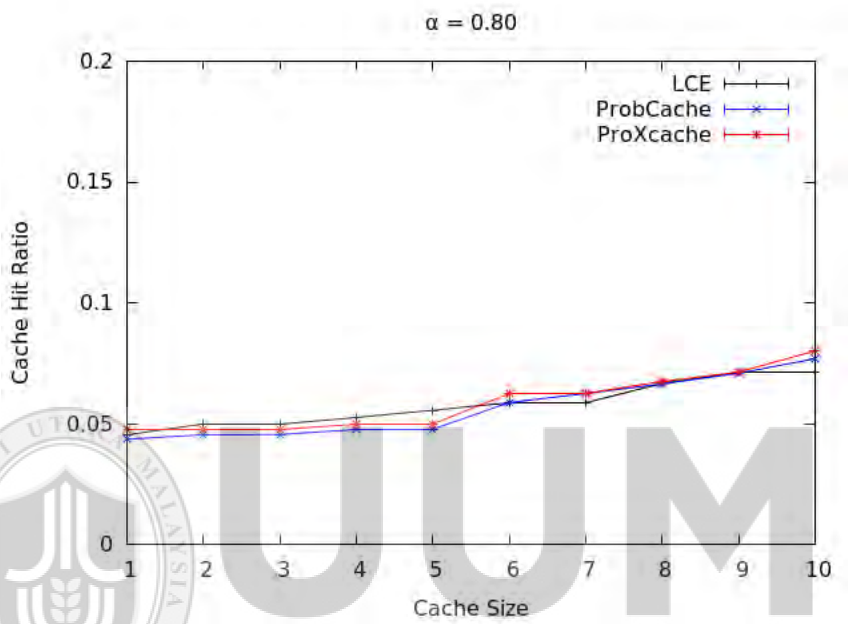

Figure 6.9. Cache-hit Case 5 in GEANT topology

The case expression and comparison on GEANT was subjected to review the effects of the hit rates on a more diversified network topology as shown on Figure 6.9. The close relationships experienced on GEANT is as a result of higher degree on node interactions. GEANT topology structure is a double representation of the Abilene in terms of coverage.

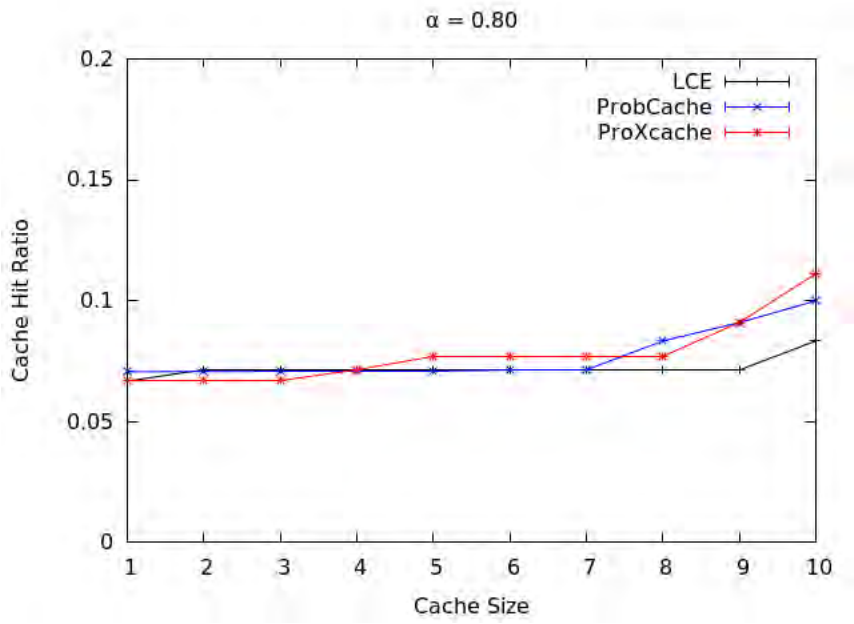

Figure 6.10. Cache-hit Case 6 in DTelekom topology

Flexibility of the DTelekom also showed little relationship of the enormous clusters in the topology (see Figure 6.10). ProXcache on lesser cache size performed lower. However, as the cache sizes increases the degree of cache hit improved making ProXcache outperforming other strategies. This implies that as the size is increasing, caching unique contents are achieved. This also improved the diversity and reduced the amount of replacement.

6.3.3 Cache-Hit Evaluation on Video on Demand with $\alpha = 1.2$

VoD contents have more request as submitted by [1, 30].

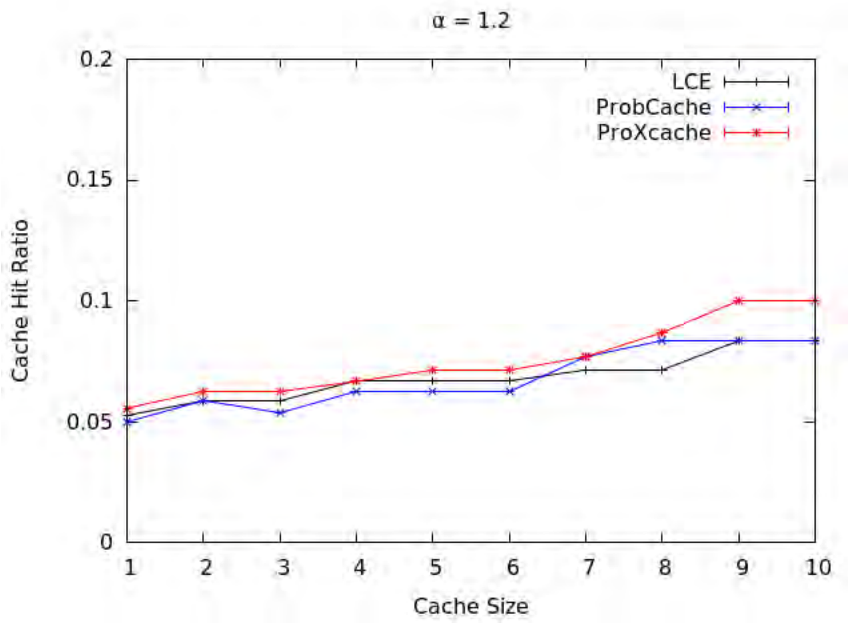

Figure 6.11. Cache-hit Case 7 in Abilene topology

It was also noticed by the results as presented on Figure 6.11, the lower the cache sizes, the hit rates are lesser. ProXcache thus improved the hit against the VoD as discovered and observed on File $\alpha = 0.75$ and Web contents $\alpha = 0.80$.

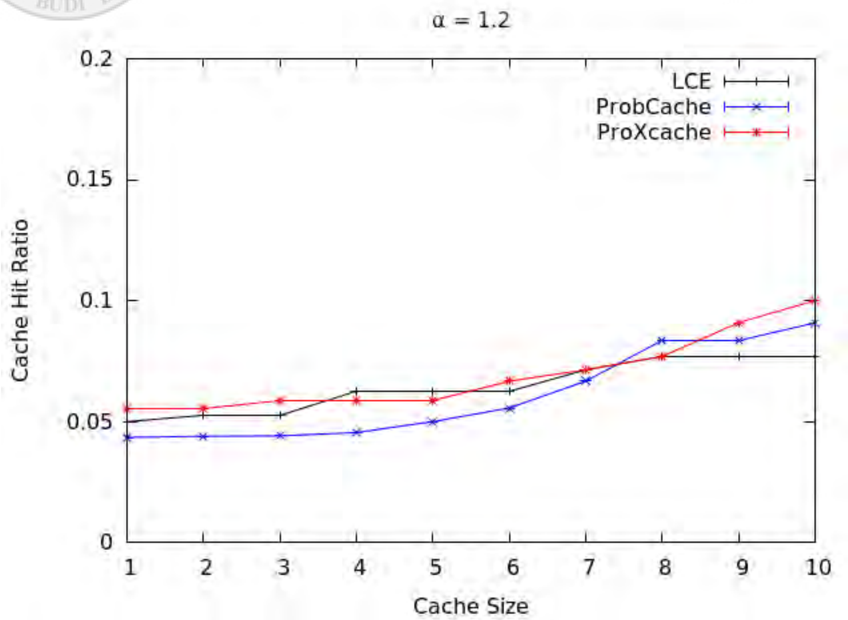

Figure 6.12. Cache-hit Case 8 in GEANT topology

For the evaluation and comparison purpose, GEANT mostly returned more complex relationship as it also yielded better performance by adding the proxy-positions in our strategy as presented on Figure 6.12 and 6.13 for the DTelekom topology respectively.

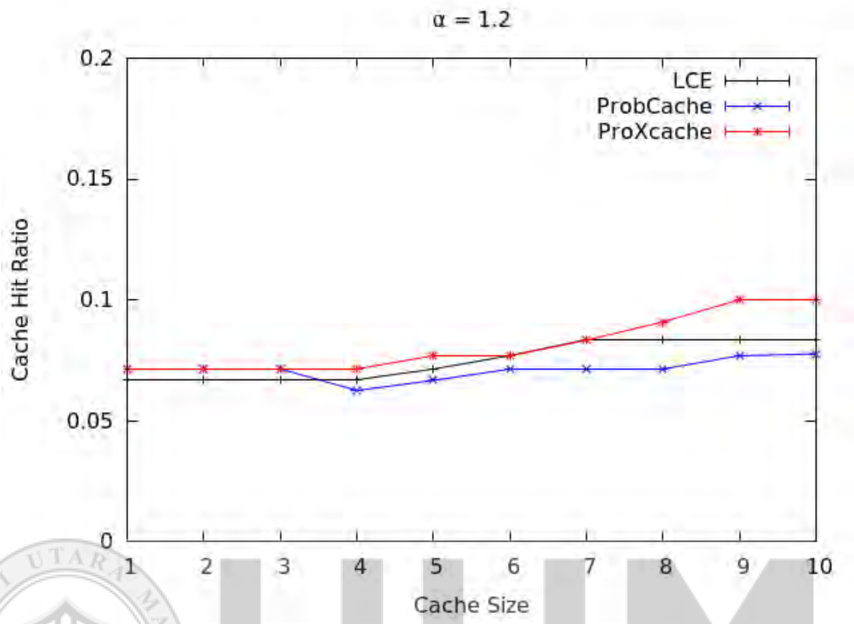

Figure 6.13. Cache-hit Case 9 in DTelekom topology

6.4 Simulation and Evaluation on Diversity with $\alpha = 1.2$

Diversity defines the degree of caching distinct and unique contents on the network path [144, 146]. The lower and higher diversity implies the number of similar contents while a higher diversity score returns better number of unique or heterogeneous contents on caching devices in ICN.

6.4.1 Diversity on Abilene Topology

The more the diversity value ratio tends towards one the better the performance and caching of the contents.

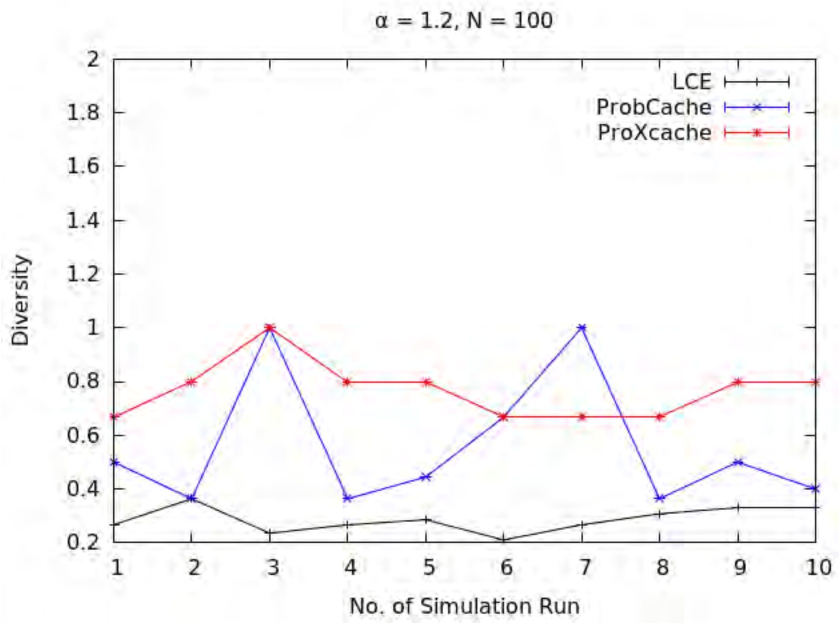

Figure 6.14. Diversity on Abilene topology

This implies the number of caching unique contents. On the Abilene topology, the reason of evaluation and comparison, the experiments were tested on cache sizes 100 (1GB) where ProXcache recorded more caching of unique and diversified contents as presented in Figure 6.14. However, the ProbCache outperformed the LCE and ProXcache on some trials since it was randomly subjected to the simulation according to the traffic by [112].

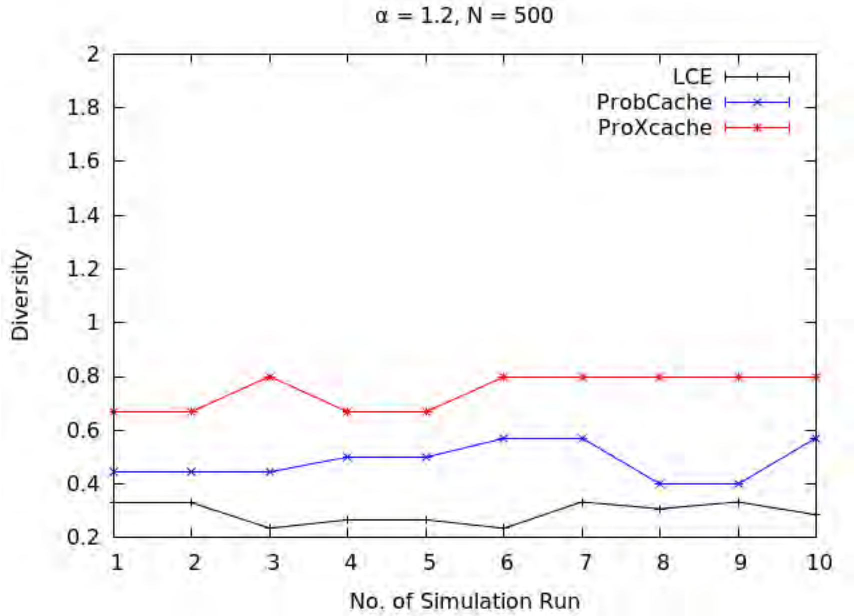

Figure 6.15. Diversity Case 6 in Abilene topology

On the increase of cache sizes from 100 to 500, the ProXcache recorded better values of diversity as compared to LCE and ProbCache. LCE due to its caching operation on all traversed nodes, recorded lesser diversity. The performance was enhanced to about 80% of the contents as diversifies unique contents were cached as depicted on Figure 6.15.

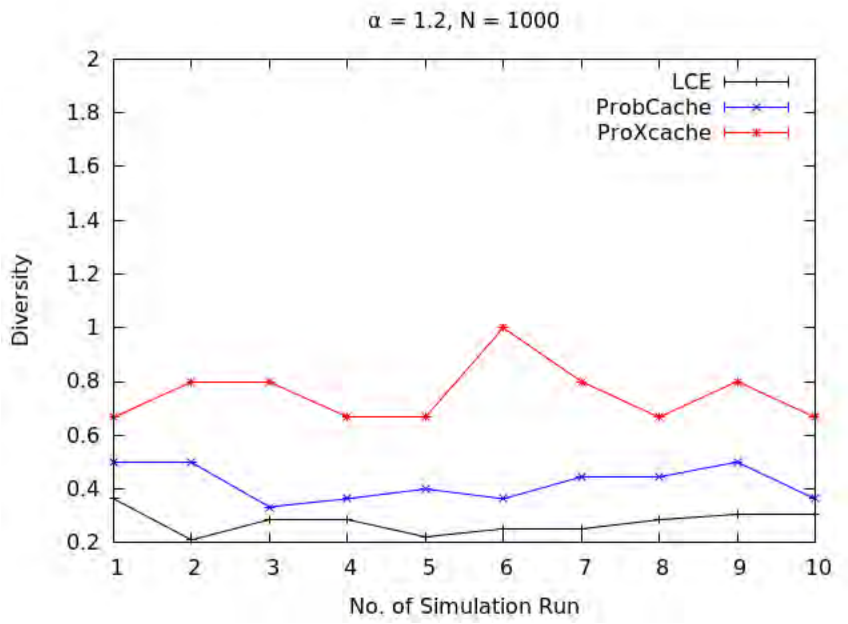

Figure 6.16. Diversity Case 7 in Abilene topology

Increasing the cache size to 1000 = 10GB, yielded better diversity as LCE and ProbCache performed lesser than the proposed ProXcache on almost all the runs and trials on Abilene topology. Figure 6.16 presents the graphical representation of the description.

6.4.2 Diversity on GEANT Topology

Diversity on GEANT provide better results as compared to Abilene.

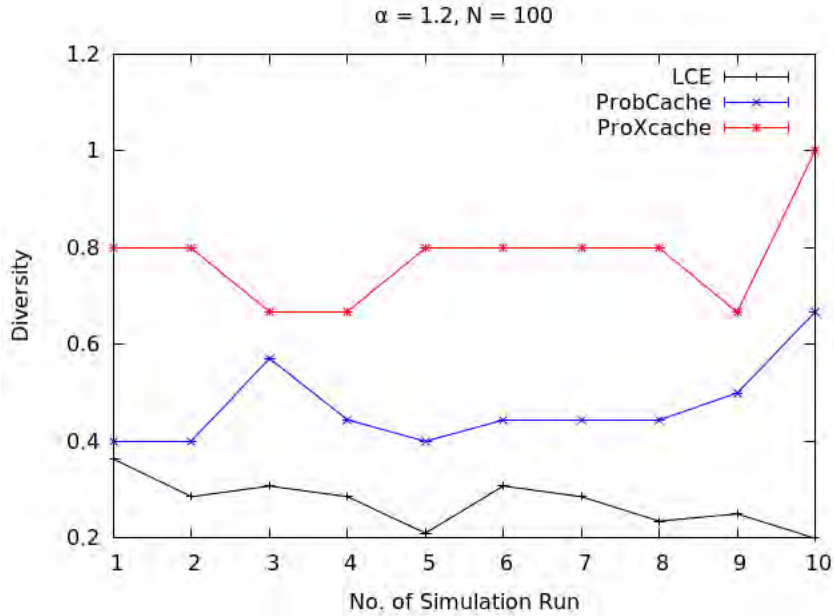

Figure 6.17. Diversity Case 1in GEANT topology

The evaluation was done on popular VoD. The trials proved the high diversity of contents cached on the network. The result of GEANT is presented for cache sizes 500 and 1000 on Figure 6.18 and 6.19 respectively.

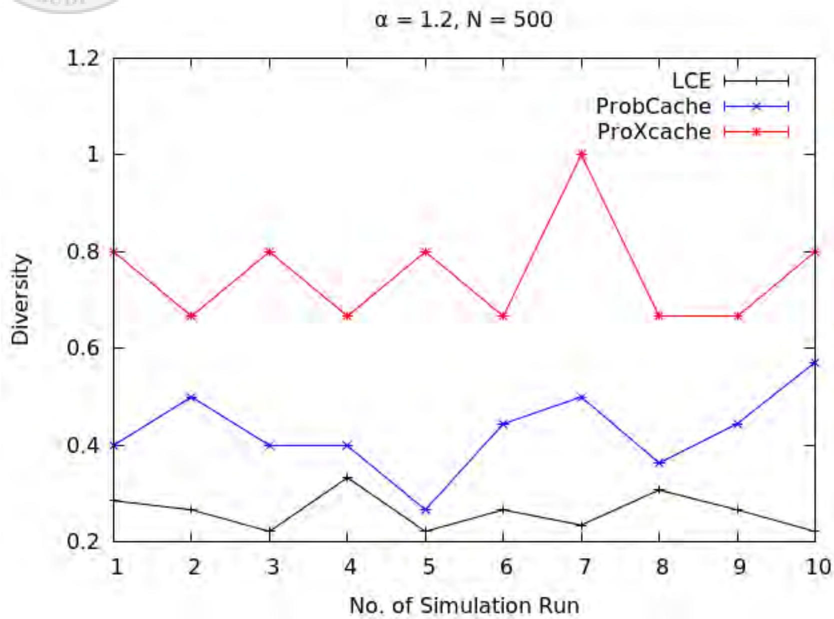

Figure 6.18. Diversity Case 2 in GEANT topology

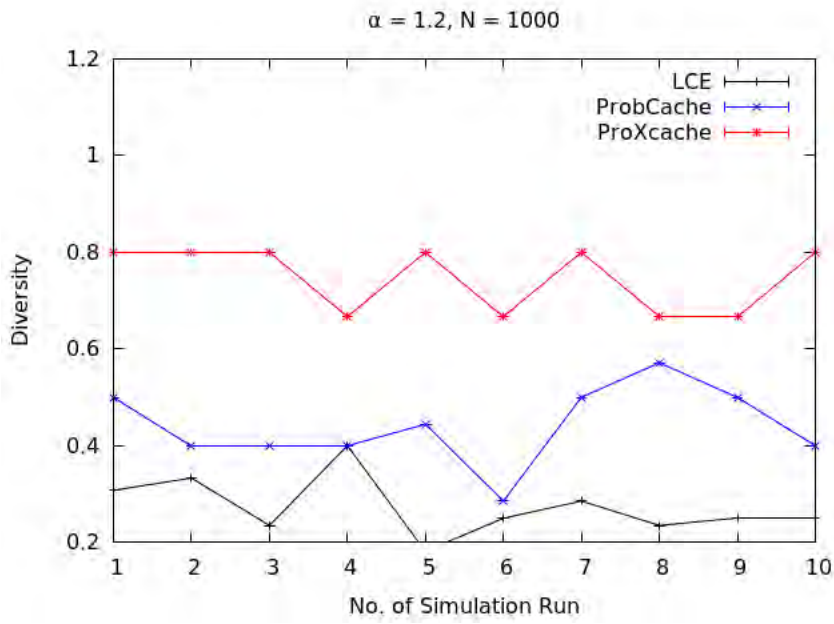

Figure 6.19. Diversity Case 3 in GEANT topology

6.4.3 Diversity on DTelekom Topology

Attempts and trials were also verified on the DTelekom network topology varying the cache sizes from 100- 1000 as presented on the following Figures 6.20 and Figure 6.21.

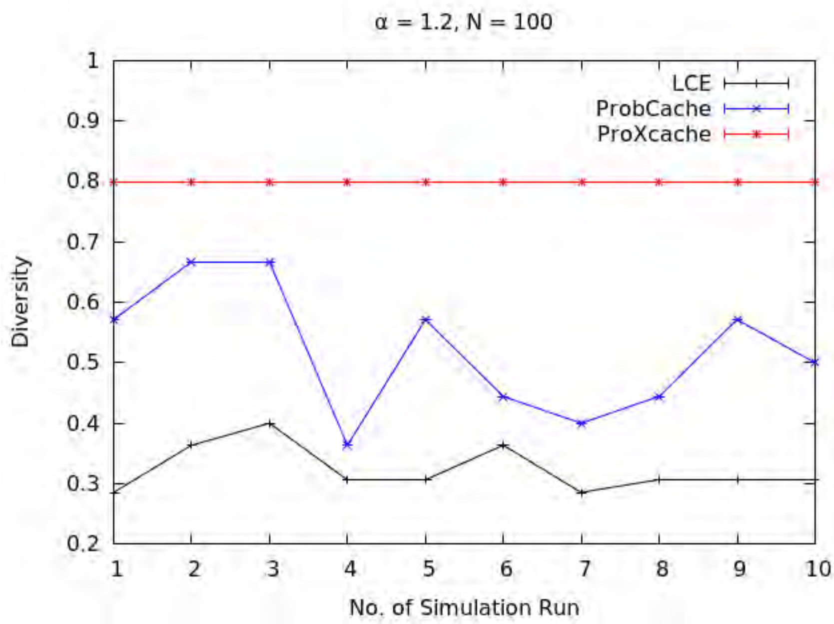

Figure 6.20. Diversity Case 4 on DTelekom topology

DTelekom performance was almost on an equal percentage on all the series of trial for ProXcache. This is as a result of higher node degrees between the nodes on the topology. Figure 6.20 and Figure 6.21 presents the results of increasing the cache size to 500 and 1000 respectively.

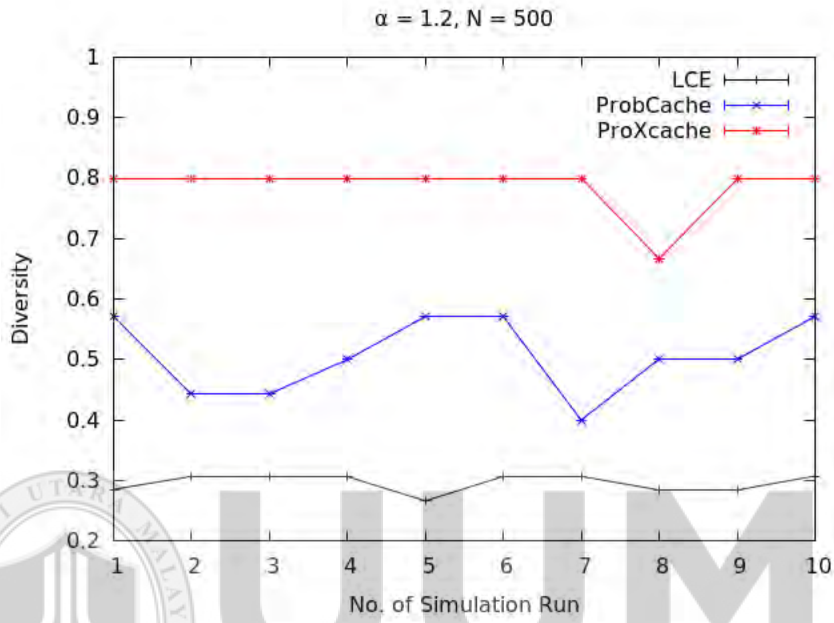

Figure 6.21. Diversity Case 5 on DTelekom topology

Results of case 6.22 provides the flexible percentage of N =1000. ProXcache outperformed the LCE and ProbCache in all the trials. The evaluation and comparisons showed that on all cache sizes, ProXcache performed better than the LCE and ProbCache.

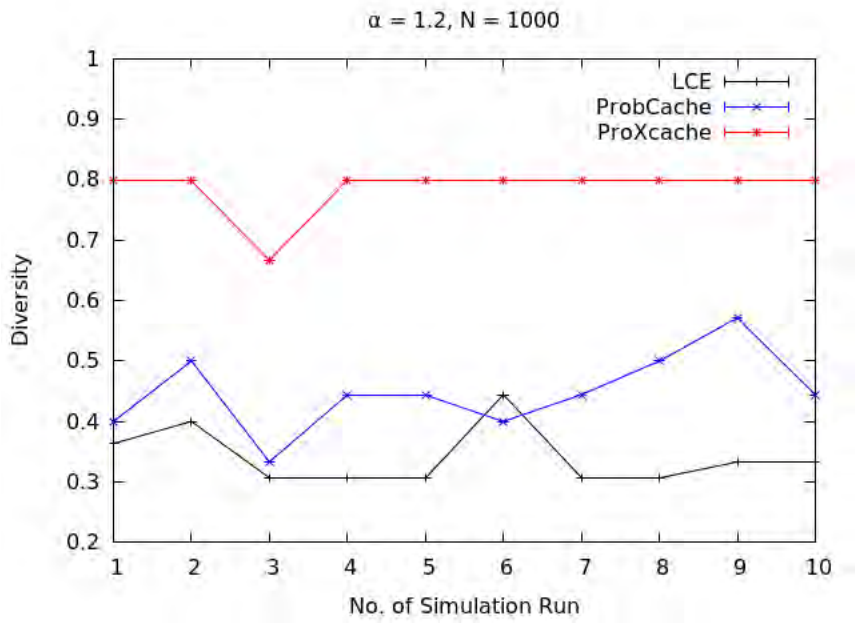

Figure 6.22. Diversity Case 6 on DTelekom topology

6.5 Summary

The chapter introduced the amount of parameters used in the simulation analysis. The study covered the simulation of the popular contents. From the File $\alpha = 0.75$, Web files $\alpha = 0.80$, and the VoD $\alpha = 1.2$ contents subjected to various trials and increasing cache sizes between 100-1000 (1GB-10GB). The results proved the comparisons and evaluation of the cache-hit, path and content redundancy. The path and content redundancy of the contents were tested against ranging cache sizes from 1GB to 10GB on various topologies. The evaluation shows that ProXcache achieved better rates of mitigating path and content redundancy. The evaluation proved that as the value of cache sizes increases for the tested case VoD, the path were easily congested due to its popularity and higher Zipf value. It is thus conclusive to submit that our strategy recorded better performance in mimicking an ICN heterogeneous content flooded scenario. Our strategy recorded a less significant result as it was seen on alpha 0.8 for web content. It was also noticed that DTelekom topology favors ProXcache performance. Therefore, for a network scenario or structure that depicts DTelekom architecture, our strategy could be applied. Diversity performance was also better for the ProXcache on the GEANT and DTelekom topologies as the numbers of diversity ratio was more stable as against Abilene.

The chapter also evaluated the results of cache-hit rates in different topologies. Ranging from the Abilene, GEANT and DTelekom. The overall performance according to the results in the study, shows that better cache-hit rates are achieved by ProXcache in heterogeneous content-driven network. The results of the content diversity was evaluated at the end of the chapter. ProXcache cached more diversified contents due to its lessen content deposition on the paths. This thus provides path and memory spaces of caching new contents. The results proved that ProXcache also had overall better performance on content diversity thereby providing lesser bandwidth consumption through less eviction operation in case of a cache full devices.

CHAPTER SEVEN
CONCLUSION AND FUTURE WORKS

7.1 Introduction

For the past decades of Information dissemination. Internet has engrave all the competitive forms of delivery and productivity by huge margin. This is as a result of the Internet diversified structure that enables it to handle all forms of data. Internet structure have never seize from the alteration of its primitive design through proposals and forms.

However, major architectural techniques and theoretical updates have been fruitful without dissociating its original semantics. Initial Internet design was model to share little memory demanding resources when compared to the needs of its users at present. The demands of the current Internet usage have boosted the usability to the staggering traffic and memory growth in the capacity of the data traversed via user connectivity and device-to-device interactivity. These along the predicted traffic gave rise to the leading research of Information-Centric Networking (ICN). The ICN paradigm which yearns to bring information much closer to the users by dissociating partly the address of the host in lieu of the content names. ICN has been agreed as a future paradigm of man-in-middle (caching) form of communication. Its deployment is predicted to be a leading hoping ground for the Internet of Things (IoT) and Internet of Everything (IoE).

Consequently, ICN seems practically un-deployable without adequate caching schemes. Several leading studies have investigated the need to cache data in all traversing nodes or devices as submitted by [6, 15, 4]. A popular strategy in ICN initial design was the Leave Copy Everywhere (LCE) caching management. Researchers have therefore agreed with its initial configuration; vise-a-vis the LCE suffers from lots of deposition of contents on all traversing nodes leading to high request of network resources. This thus led to device and memory challenges of path redundancy, content redundancy, excessive data replacement (preemption through deletion) as cache sizes are limited in space for heterogeneous and multicached driven networks. LCE have been identified to suffer from the aforementioned challenges.

This study evaluated the research problems in caching techniques of ICN and thereby proposed a novel caching form called ProXcache to mitigate the excessive path and content redundancy in ICN. The remaining part of the chapter is organized as thus: Section 7.2 summarized the research overall, Section 7.3 describes the network perspectives and the research contribution. Section 7.4 presents the research limitation while Section 7.5 highlights the future direction of the study.

7.2 Summary of the Research

Generally, the Internet description presents several section and operation of content dissemination through various protocols driven by the host dependency. As described in Chapter one, the study extensively described the general promising potentials of subscribing to the Internet of content known as Information-Centric Network (ICN). Chapter one introduced the need to vend into ICN by identifying some of its potentials of achieving the IoT, IoE and other advances envisioned in today's Internet. The general architecture and research issues of the study was described. Motivation , Research Questions, objectives and scopes were covered to juxtapose the results of the overall research.

The study went into the related relevant literature in Chapter Two. A full description of the research efforts made by several researchers in different capacity was explored.

In description, careful reporting of the concerned topics were generally itemized and categorized. Cache categorization to on-path and off-path shows the full direction of this research. It explored the related approaches by introducing the content deployments strategies and how the improvements are required to better the understanding of the study.

Empirical simulation comparison of other related studies were presented at the end of the chapter. Chapter three expressed the overall research methodology adopted for the study using the techniques of DRM by [121, 125] and the extension of other research methods. Chapter four of the study, described the full actualization of the first objective of the study. A full detail description of the forms of achieving the path redundancy elimination using Hypergraph concept. The chapter also presented the details of the verification of the proposed *PRE* model. In conclusion of the chapter, a comparison of the model was presented and the results showed the improvements in terms of mitigating the enormous amount of caches previously suffered by the LCE.

Chapter Five of the study covered the ProXcache deployment and development. The chapter presented the extensive description of all the model information, topologies and relationship of the study through comparison of results with other studies. The results obtained proved the implementation of the ProXcache in performance to be better than the compared LCE and ProbCache is some instances of the metrics. The results also submitted the significant reduction of path and content redundancy in a network. Chapter Six presented the extended results comparison using more parameters to further substantiate the performance of metrics chosen in chapter four and five.

7.3 Research Contribution

The general research contribution of this study was to develop a new cache deployment strategy that will curtail and minimize the amount of content deposited on traversing nodes of the Internet. The strategy was developed and named ProXcache. Its development included the proposed *PER* model and the exploration of the hypergraph concept of the graph theory. The specific contribution are as thus classified into the following:

1. The novel contribution through the use the hypergraph model in the ICN cache manageability by introducing the path description and positioning address of proxy caching.
2. A new description of the mathematical relationship using hypergraph in ICN for the formulation of node relationships
3. The study formulates a new form of cache deployment strategy that achieved the following:
   a) Mitigates the high path redundancy in LCE
   b) Reduced the overall content deposition through ProXcache
   c) Improved the cache-hit rates of the ICN heterogeneous demand of data for LCE and ProbCache
   d) An improved form of caching that increases the number of caching distinct contents by reducing the amount of popular content deletion in ICN
4. An extension of the SocialCCNSim simulator by incorporating the ProXcache design for future strategy comparison

A summary of the entire research contribution can be themed along the following representation as shown on Figure 7.1. The first stage is the PRE model development through mathematical relationship formulation to acquire distance and node relation-

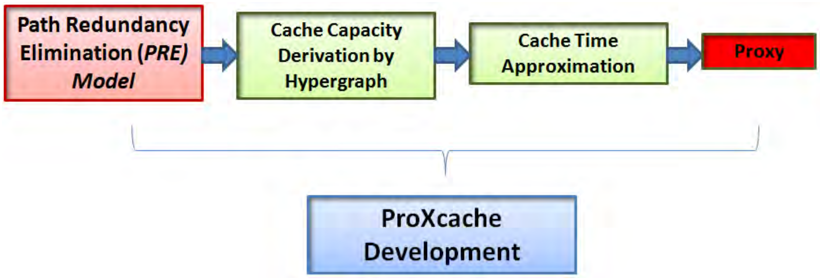

Figure 7.1. Summary of Research Contribution

ships. Cache Capacity computation and approximate cache time development was successful using the hypergraph properties of hyper arcs and edges. Unique inclusion of proxy at the user and publisher ends concludes the ProXcache development.

7.4 Limitation of the Study

The research has indeed explored and extended the study through varying scenario by mimicking the set-up of the real network using the SocialCCNSim and Social media traffic generator (SONETOR). However, the research was only subjected to the on-path caching scheme. The research is thus limited to some chosen parameters due to time constraints which among was the catalog size of 10^6. Since ICN research is still in its early states in terms of the real world implementation; it would go a long way if the proposed strategy could be subjected to a testbed to evaluates the performance of the strategy. With the tough task of finding ICN simulators that supports adding more optional flexibility, the simulation was fully done on SocialCCNSim.

With the diversify nature of the ad hoc network, the study is limited to a discrete controlled network model using the specifications expressed on the parameter table. Therefore, the researchers have not tested the performance on mobility driven network settings such as mobile ad hoc networks (MANET) and vehicular ad hoc networks (VANETs).

7.5 Future Research Direction

The proposed strategy was fully deployed through simulation to a fixed network. The strategy could be reproduced on different specification to test the following as a future direction:

1. To test the implementation of the strategy on different scenario setting of other future network paradigm (e.g. 5G network, Grid network and cloud infrastructure)
2. Testing the strategy against other network evaluation metrics such as network stretch, bandwidth utilization and delay.
3. Extend and test the study against other research proposed cache deployments strategies such as Active caching, Adaptive caching and transparent caching.
4. To measure the degree of performance against the cache management schemes such as Most Popular Caching scheme, MaX-Gain (MAGIC) caching as the strategies are gaining popularity in the field of ICN research.
5. Extend the concept of ICN deployment in a more diverse network such as MANET.
6. Expose the test and implementation of ICN deployment on caching in a disaster scenarios challenged network as submitted in the ICN IETF-Draft [35].

REFERENCES

[1] Cisco, “Cisco Visual Networking Index:Global Mobile Data Traffic Forecast Update, 2014-2019,” Cisco, Tech. Rep., 2015. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[2] J. F. Gantz, D. Reinsel, C. Chute, W. Schlichting, J. McArthur, S. Minton, I. Xheneti, A. Toncheva, and A. Manfrediz, “The expanding digital universe: A forecast of worldwide information growth through 2010.” An IDC White Paper- Sponsored by EMC, Technical Reoprt, March 2007.

[3] Cisco, “Cisco Visual Networking Index:Global Mobile Data Traffic Forecast Update, 2013-2018,” Cisco, Tech. Rep., 2014.

[4] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, “A survey of information-centric networking,” *Communications Magazine, IEEE*, vol. 50, no. 7, pp. 26–36, July 2012.

[5] E. Bardram, “Activity-based computing: Support for mobility and collaboration in ubiquitous computing,” *Personal Ubiquitous Comput.*, vol. 9, no. 5, pp. 312–322, Sep. 2005. [Online]. Available: http://dx.doi.org/10.1007/s00779-004-0335-2

[6] V. Jacobson, D. K. Smetters, J. D. Thornton, M. Plass, N. Briggs, and R. Braynard, “Networking named content,” in *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT ’09. ACM, 2009, pp. 1–12.

[7] G. Xylomenos, C. Ververidis, V. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. Katsaros, and G. Polyzos, “A survey of information-centric networking research,” *Communications Surveys Tutorials, IEEE*, no. 99, pp. 1–26, 2013.

[8] ICNRG-IRTF. (2015) Information-centric networking research group (ICNRG). [Online]. Available: https://irtf.org/icnrg

[9] G. Xylomenos, C. Ververidis, V. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. Katsaros, and G. Polyzos, “A survey of information-centric networking research,” *Communications Surveys Tutorials, IEEE*, vol. 16, no. 2, pp. 1024–1049, Second 2014.

[10] G. Pavlou, “Keynote 2: Information-centric networking: Overview, current state and key challenges,” in *Computers and Communications (ISCC), 2011 IEEE Symposium on*, June 2011, pp. 1–1.

[11] Y. Zhu and A. Nakao, “A deployable and scalable information-centric network architecture,” in *Communications (ICC), 2013 IEEE International Conference on*, June 2013, pp. 3753–3758.

[12] S. Shailendra, B. Panigrahi, H. Rath, and A. Simha, "A novel overlay architecture for information centric networking," in *Communications (NCC), 2015 Twenty First National Conference on*, Feb 2015, pp. 1–6.

[13] K. Katsaros, G. Xylomenos, and G. Polyzos, "A hybrid overlay multicast and caching scheme for information-centric networking," in *INFOCOM IEEE Conference on Computer Communications Workshops , 2010*, March 2010, pp. 1–6.

[14] V. Jacobson, D. K. Smetters, J. D. Thornton, M. Plass, N. Briggs, and R. Braynard, "Networking named content," *Commun. ACM*, vol. 55, no. 1, pp. 117–124, 2012.

[15] B. Ahlgren, M. D'Ambrosio, M. Marchisio, I. Marsh, C. Dannewitz, B. Ohlman, K. Pentikousis, O. Strandberg, R. Rembarz, and V. Vercellone, "Design considerations for a network of information," in *Proceedings of the 2008 ACM CoNEXT Conference*, ser. CoNEXT '08. New York, NY, USA: ACM, 2008, pp. 66:1–66:6. [Online]. Available: http://doi.acm.org/10.1145/1544012.1544078

[16] X. Vasilakos, V. A. Siris, G. C. Polyzos, and M. Pomonis, "Proactive selective neighbor caching for enhancing mobility support in information-centric networks," in *Proceedings of the Second Edition of the ICN Workshop on Information-centric Networking*, ser. ICN '12. New York, NY, USA: ACM, 2012, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/2342488.2342502

[17] C. Dannewitz. (2011) 4WARD and SAIL NetInf Project. http://www.netinf.org/4ward-netinf/. [Online]. Available: http://www.netinf.org/4ward-netinf/

[18] C. Dannewitz, D. Kutscher, B. Ohlman, S. Farrell, B. Ahlgren, and H. Karl, "Network of information (netinf) - an information-centric networking architecture," *Comput. Commun.*, vol. 36, no. 7, pp. 721–735, Apr. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.comcom.2013.01.009

[19] H. R. Sadjadpour, "A new design for information centric networks," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, March 2014, pp. 1–6.

[20] S. DiBenedetto, C. Papadopoulos, and D. Massey, "Routing policies in named data networking," in *Proceedings of the ACM SIGCOMM Workshop on Information-centric Networking*, ser. ICN '11. New York, NY, USA: ACM, 2011, pp. 38–43. [Online]. Available: http://doi.acm.org/10.1145/2018584.2018595

[21] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, "A data-oriented (and beyond) network architecture," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 181–192, Aug. 2007. [Online]. Available: http://doi.acm.org/10.1145/1282427.1282402

[22] M. Bari, S. Chowdhury, R. Ahmed, R. Boutaba, and B. Mathieu, "A survey of naming and routing in information-centric networks," *Communications Magazine, IEEE*, vol. 50, no. 12, pp. 44–53, December 2012.

[23] N. Laoutaris, S. Syntila, and I. Stavrakakis, "Meta algorithms for hierarchical web caches," in *Performance, Computing, and Communications, 2004 IEEE International Conference on*, 2004, pp. 445–452.

[24] I. Psaras, W. K. Chai, and G. Pavlou, "Probabilistic in-network caching for information-centric networks," in *Proceedings of the Second Edition of the ICN Workshop on Information-centric Networking*, ser. ICN '12. New York, NY, USA: ACM, 2012, pp. 55–60. [Online]. Available: http://doi.acm.org/10.1145/2342488.2342501

[25] C. Fang, R. Yu, T. Huang, J. Liu, and J. Liu, "A survey of green information-centric networking: Research issues and challenges," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

[26] Q. Nguyen, M. Arifuzzaman, T. Miyamoto, and S. Takuro, "An optimal information centric networking model for the future green network," in *Autonomous Decentralized Systems (ISADS), 2015 IEEE Twelfth International Symposium on*, March 2015, pp. 272–277.

[27] Z. Aihua, "Study of ubiquitous learning environment based on ubiquitous computing," in *Ubi-media Computing (U-Media), 2010 3rd IEEE International Conference on*, July 2010, pp. 136–138.

[28] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in *Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing*, ser. MCC '12. New York, NY, USA: ACM, 2012, pp. 13–16. [Online]. Available: http://doi.acm.org/10.1145/2342509.2342513

[29] A. Banerjee, A. Mukherjee, H. S. Paul, and S. Dey, "Offloading work to mobile devices: An availability-aware data partitioning approach," in *Proceedings of the First International Workshop on Middleware for Cloud-enabled Sensing*, ser. MCS '13. New York, NY, USA: ACM, 2013, pp. 4:1–4:6. [Online]. Available: http://doi.acm.org/10.1145/2541603.2541605

[30] Click, "Internet time line," ClickTale, Tech. Rep., February 2015. [Online]. Available: http://blog.clicktale.com/2010/11/17/a-brief-history-of-web-analytics/

[31] R. Gallen, A. Cord, N. Hautiere, E. Dumont, and D. Aubert, "Nighttime visibility analysis and estimation method in the presence of dense fog," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–11, 2014.

[32] Intel, "What happens in an internet minute," Intel Corporation, Tech. Rep., February 2014. [Online]. Available: http://www.intel.com/content/www/us/en/communications/internet-minute-infographic.html#

[33] Livestats. (2015, June) Internet Live Stats. [Online]. Available: http://www.internetlivestats.com/

[34] K. Pentikousis, B. Ohlman, D. Corujo, G. Boggia, G. Tyson, E. B. Davies, D. Gellert, P. Mahadevan, S. Spirou, A. Molinaro, and S. Eum, "ICN Baseline Scenarios and Evaluation Methodology," Internet Engineering Task Force, Internet-Draft draft-pentikousis-icn-scenarios-04, Jan. 2014, work in Progress. [Online]. Available: https://tools.ietf.org/html/draft-pentikousis-icn-scenarios-04

[35] J. Seedorf, A. Tagami, N. Blefari-Melazzi, M. Arumaithurai, and K. Ramakrishnan, "Using ICN in disaster scenarios," Internet Engineering Task Force, Internet-Draft draft-seedorf-icn-disaster-05, Dec. 2015, work in Progress. [Online]. Available: https://tools.ietf.org/html/draft-seedorf-icn-disaster-05

[36] V. Dimitrov and V. Koptchev, "PSIRP Project – Publish-Subscribe Internet Routing Paradigm: New Ideas for Future Internet," in *Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies*, ser. CompSysTech '10. New York, NY, USA: ACM, 2010, pp. 167–171. [Online]. Available: http://doi.acm.org/10.1145/1839379.1839409

[37] A. Dominguez, O. Novo, W. Wong, and T. Valladares, "Publish/subscribe communication mechanisms over psirp," in *Next Generation Web Services Practices (NWeSP), 2011 7th International Conference on*, Oct 2011, pp. 268–273.

[38] C. Dannewitz, M. Herlich, and H. Karl, "Opennetinf: prototyping an information-centric network architecture," in *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*, Oct 2012, pp. 1061–1069.

[39] A. Araldo, M. Mangili, F. Martignon, and D. Rossi, "Cost-aware caching: Optimizing cache provisioning and object placement in icn," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, Dec 2014, pp. 1108–1113.

[40] C. Bernardini, T. Silverston, and O. Festor, "MPC: Popularity-based Caching Strategy for Content Centric Networks," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3619–3623.

[41] G. Bianchi, A. Detti, A. Caponi, and N. Blefari Melazzi, "Check before storing: What is the performance price of content integrity verification in lru caching?" *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 3, pp. 59–67, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2500098.2500106

[42] M. Bilal and S.-G. Kang, "Time aware least recent used (tlru) cache management policy in icn," in *Advanced Communication Technology (ICACT), 2014 16th International Conference on*, Feb 2014, pp. 528–532.

[43] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache "less for more" in information-centric networks," in *Proceedings of the 11th International IFIP TC 6 Conference on Networking - Volume Part I*, ser. IFIP'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 27–40. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-30045-5_3

[44] M. Draxler and H. Karl, "Efficiency of on-path and off-path caching strategies in information centric networks," in *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*, Nov 2012, pp. 581–587.

[45] K. Katsaros, G. Xylomenos, and G. Polyzos, "Multicache: An incrementally deployable overlay architecture for information-centric networking," in *INFOCOM IEEE Conference on Computer Communications Workshops , 2010*, March 2010, pp. 1–5.

[46] Y. Li, H. Xie, Y. Wen, and Z.-L. Zhang, "Coordinating in-network caching in content-centric networks: Model and analysis," Center for Innovation, Huawei Central Research Institute, Tech. Rep., 2012. [Online]. Available: http://www.ntu.edu.sg/home/ygwen/Paper/LXW-ICDCS-13.pdf

[47] I. Psaras, W. Chai, and G. Pavlou, "In-network cache management and resource allocation for information-centric networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[48] Y. Li, H. Xie, Y. Wen, and Z.-L. Zhang, "Coordinating in-network caching in content-centric networks: Model and analysis," Center for Innovation, Huawei Central Research Institute, Tech. Rep., 2012. [Online]. Available: http://www.ntu.edu.sg/home/ygwen/Paper/LXW-ICDCS-13.pdf

[49] C. Wei-Koong, H. Diliang, P. Ioannis, and P. George, "Cache less for more in information-centric networks (extended version)," *Computer Communications*, vol. 36, pp. 758–770, 2013.

[50] S. V. Nagaraj, *Web Caching and Its Application*, S. Science, Ed. Kluwer Academic Publishers New York, Boston, Dordrecht, London, 2004, no. 1-4020-8050-6. [Online]. Available: http://ebooks.kluweronline.com

[51] H. Zhai, H. Jiang, Y. Sun, J. Li, J. Liu, G. Fang, and E. Dutkiewicz, "Lhcds: A novel deployment strategy of proxy caches for p2p traffic in isp networks," in *Computer and Information Technology (CIT), 2011 IEEE 11th International Conference on*, Aug 2011, pp. 224–230.

[52] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking: a survey," *Computer Networks*, vol. 57, no. 16, pp. 3128–3141, 2013.

[53] A. Luotonen and K. Altis, "World-wide web proxies," *Computer Networks and ISDN systems*, vol. 27, no. 2, pp. 147–154, 1994.

[54] J. Ren, W. Qi, C. Westphal, J. Wang, K. Lu, S. Liu, and S. Wang, "Magic: A distributed max-gain in-network caching strategy in information-centric networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, April 2014, pp. 470–475.

[55] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, "Directed hypergraphs and applications," *Discrete Appl. Math.*, vol. 42, no. 2-3, pp. 177–201, Apr. 1993. [Online]. Available: http://dx.doi.org/10.1016/0166-218X(93)90045-P

[56] B. M. Leiner, V. G. Cerf, D. D. Clark, R. E. Kahn, L. Kleinrock, D. C. Lynch, J. Postel, L. G. Roberts, and S. Wolff, "A brief history of the internet," *SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 5, pp. 22–31, Oct. 2009. [Online]. Available: http://doi.acm.org/10.1145/1629607.1629613

[57] B. Aboba, *The Online User's Encylopedia: Bulletin Boards and Beyond.* Reading, MA:: Addison-Wesley, 1993.

[58] K. Pentikousis, P. Chemouil, K. Nichols, and G. Pavlou, "Information-centric networking: part ii [guest editorial]," *Communications Magazine, IEEE*, vol. 50, no. 12, pp. 15–17, December 2012.

[59] D. Petcu, "Multi-cloud: Expectations and current approaches," in *Proceedings of the 2013 International Workshop on Multi-cloud Applications and Federated Clouds*, ser. MultiCloud '13. New York, NY, USA: ACM, 2013, pp. 1–6. [Online]. Available: http://doi.acm.org/10.1145/2462326.2462328

[60] A. Schmidt, "Ubiquitous computing: Are we there yet?" *Computer*, vol. 43, no. 2, pp. 95–97, Feb 2010.

[61] W. Liang, B. Suzan, and K. Jussi, "Optimal chunking and partial caching in information-centric networks," *Computer Communications*, 2015.

[62] A. Silberschatz, P. B. Galvin, and G. Gagne, *Operating System Concepts*, 8th ed. Wiley Publishing, 2008.

[63] *Enhancing cache robustness for content-centric networking*, March 2012.

[64] A. Bengt, K. Holger, K. Dirk, O. B., O. Sara, and S. Ignacio, Eds., *Information-Centric Networking*, ser. Dagstuhl Seminar Proceedings, no. 10492. Dagstuhl, Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2011. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2011/2943

[65] L. Muscariello, G. Carofiglio, and M. Gallo, "Bandwidth and storage sharing performance in information centric networking," in *Proceedings of the ACM SIGCOMM Workshop on Information-centric Networking*, ser. ICN '11. New York, NY, USA: ACM, 2011, pp. 26–31. [Online]. Available: http://doi.acm.org/10.1145/2018584.2018593

[66] K. Su, F. Bronzino, K. K. Ramakrishnan, and D. Raychaudhuri, "MFTP: A Clean-Slate Transport Protocol for the Information Centric Mobilityfirst Network," in *Proceedings of the 2Nd International Conference on Information-Centric Networking*, ser. ICN '15. New York, NY, USA: ACM, 2015, pp. 127–136. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2810156.2810169

[67] G. Tyson, N. Sastry, R. Cuevas, I. Rimac, and A. Mauthe, "A survey of mobility in information-centric networks," *Commun. ACM*, vol. 56, no. 12, pp. 90–98, Dec. 2013. [Online]. Available: http://doi.acm.org/10.1145/2500501

[68] Z. Hao, W. Yonggang, X. Haiyong, and Y. Nenghai, "DHT Platforms," in *Distributed Hash Table*, ser. SpringerBriefs in Computer Science. Springer New York, 2013, pp. 23–38.

[69] J. Kubiatowicz, D. Bindel, Y. Chen, S. Czerwinski, P. Eaton, D. Geels, R. Gummadi, S. Rhea, H. Weatherspoon, W. Weimer, C. Wells, and B. Zhao, "Oceanstore: An architecture for global-scale persistent storage," *SIGPLAN Not.*, vol. 35, no. 11, pp. 190–201, November 2000. [Online]. Available: http://doi.acm.org/10.1145/356989.357007

[70] A. Karila and P.-P. Team. (2008) PSRIP Project. http://www.psirp.org/home.html. [Online]. Available: http://www.psirp.org/home.html

[71] C. Tsilopoulos and G. Xylomenos, "Supporting diverse traffic types in information centric networks," in *Proceedings of the ACM SIGCOMM Workshop on Information-centric Networking*, ser. ICN '11. New York, NY, USA: ACM, 2011, pp. 13–18. [Online]. Available: http://doi.acm.org/10.1145/2018584.2018588

[72] H. Nicanfar, P. TalebiFard, C. Zhu, and V. Leung, "Efficient security solution for information-centric networking," in *Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing*, Aug 2013, pp. 1290–1295.

[73] X. Zhang, K. Chang, H. Xiong, Y. Wen, G. Shi, and G. Wang, "Towards name-based trust and security for content-centric network," in *Network Protocols (ICNP), 2011 19th IEEE International Conference on*, Oct 2011, pp. 1–6.

[74] X. Chen, Q. Fan, and H. Yin, "Caching in information-centric networking: From a content delivery path perspective," in *Innovations in Information Technology (IIT), 2013 9th International Conference on*, March 2013, pp. 48–53.

[75] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache less for more in information-centric networks (extended version)," *Computer Communications*, vol. 36, no. 7, pp. 758 – 770, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S014036641300025X

[76] A. Dabirmoghaddam, M. M. Barijough, and J. Garcia-Luna-Aceves, "Understanding optimal caching and opportunistic caching at "the edge" of information-centric networks," in *Proceedings of the 1st International Conference on Information-centric Networking*, ser. ICN '14. New York, NY, USA: ACM, 2014, pp. 47–56. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2660129.2660143

[77] G. R. Dario Rossi, "Caching performance of content centric networks under multi-path routing (and more)," Telecom ParisTech, Paris, France, Tech. Rep., 2011.

[78] S. Podlipnig and L. Böszörmenyi, "A survey of web cache replacement strategies," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 374–398, Dec. 2003. [Online]. Available: http://doi.acm.org/10.1145/954339.954341

[79] D. Christian, D. Kutscher, B. Ohlman, S. Farrell, B. Ahlgren, and H. Karl, "Network of information: An information centric network architecture," *Computer Communications*, no. 36, pp. 721–735, 2013.

[80] A. Detti, M. Pomposini, N. Blefari-Melazzi, and S. Salsano, "Supporting the web with an information centric network that routes by name," *Comput. Netw.*, vol. 56, no. 17, pp. 3705–3722, Nov. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.comnet.2012.08.006

[81] S. Eum, K. Nakauchi, Y. Shoji, N. Nishinaga, and M. Murata, "CATT: Cache aware target identification for icn," *Communications Magazine, IEEE*, vol. 50, no. 12, pp. 60–67, December 2012.

[82] S. Eum, K. Nakauchi, Y. Shoji, M. Murata, and N. Nishinaga, "Staleness verification of caching data in icn," in *ICT Convergence (ICTC), 2013 International Conference on*, Oct 2013, pp. 565–570.

[83] M. Skjegstad, F. Johnsen, T. Bloebaum, and T. Maseng, "Information-centric networking in the tactical domain," *Communications Magazine, IEEE*, vol. 51, no. 10, pp. 34–41, October 2013.

[84] T. Biswas, A. Chakraborti, R. Ravindran, X. Zhang, and G. Wang, "Contextualized information-centric home network," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, ser. SIGCOMM '13. New York, NY, USA: ACM, 2013, pp. 461–462. [Online]. Available: http://doi.acm.org/10.1145/2486001.2491691

[85] M. J. Rajahalme, K. V. Sarela, and J. Riihijarvi, "On name-based inter-domain routing," *Computer Networks*, vol. 55, pp. 975–986, 2011.

[86] K. Katsaros, N. Fotiou, X. Vasilakos, C. Ververidis, C. Tsilopoulos, G. Xylomenos, and G. Polyzos, "On Inter-Domain Name Resolution for Information-Centric Networks," in *NETWORKING 2012*, ser. Lecture Notes in Computer Science, R. Bestak, L. Kencl, L. Li, J. Widmer, and H. Yin, Eds. Springer Berlin Heidelberg, 2012, vol. 7289, pp. 13–26. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-30045-5_2

[87] H. Liu, X. De Foy, and D. Zhang, "A multi-level dht routing framework with aggregation," in *Proceedings of the Second Edition of the ICN Workshop on Information-centric Networking*, ser. ICN '12. New York, NY, USA: ACM, 2012, pp. 43–48. [Online]. Available: http://doi.acm.org/10.1145/2342488.2342498

[88] P. Flegkas, V. Sourlas, G. Parisis, and D. Trossen, "Storage replication in information-centric networking," in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, Jan 2013, pp. 850–855.

[89] Z. Ming, M. Xu, and D. Wang, "Age-based cooperative caching in information-centric networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*, March 2012, pp. 268–273.

[90] A. A. A. Mohammed and K. Okamura, "Distributed ga for popularity based partial cache management in icn," in *Proceedings of The Ninth International Conference on Future Internet Technologies*, ser. CFI '14. New York, NY, USA: ACM, 2014, pp. 18:1–18:2. [Online]. Available: http://doi.acm.org/10.1145/2619287.2619305

[91] J. W. Sen Wang, Jun Bi, "Collaborative caching based on hash -routing for information-centric networking," in *SIGCOMM'13*. Hong Kong, China: ACM, August 2013, pp. 535–536.

[92] M. Xie, I. Widjaja, and H. Wang, "Enhancing cache robustness for content-centric networking," in *INFOCOM, 2012 Proceedings IEEE*, March 2012, pp. 2426–2434. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6195632

[93] W. Sen, B. Jun, and W. Jianping, "Collaborative caching based on hash-routing for information-centric networking," in *ACM SIGCOMM 2013 Conference on SIGCOMM, Hong Kong, China*, ser. SIGCOMM '13. ACM, 2013, pp. 535–536.

[94] W. J. Min, Z. Jun, and B. Brahim, "Intra-as cooperative caching for content-centric networks," in *Proceedings of the 3rd ACM SIGCOMM Workshop on Information-centric Networking*, ser. ICN '13. New York, NY, USA: ACM, 2013, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/2491224.2491234

[95] Y. Wang, Z. Li, G. Tyson, S. Uhlig, and G. Xie, "Optimal cache allocation for content-centric networking," *IEEE Intl. Conference on Network Protocols*, 2013. [Online]. Available: http://www.internet-science.eu/sites/internet-science.eu/files/biblio/Optimalchacheallocation.pdf

[96] V. Sourlas and L. Tassiulas, "Effective cache management and performance limits in information-centric networks," in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, Jan 2013, pp. 955–960.

[97] S. Wang, J. Bi, and J. Wu, "Collaborative caching based on hash-routing for information-centric networking," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 535–536, Aug. 2013. [Online]. Available: http://doi.acm.org/10.1145/2534169.2491727

[98] L. Qin, H. Li, K. Pan, F. Chen, W. Chen, and C. Yu, "Exploring cache coding scheme for information-centric networking," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, Dec 2014, pp. 557–562.

[99] H. Jeon, B. Lee, and H. Song, "On-path caching in information-centric networking," in *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, Jan 2013, pp. 264–267. [Online]. Available: http://ieeexplore.ieee.org.eserv.uum.edu.my/stamp/stamp.jsp?tp=&arnumber=6488185

[100] M. Draxler and H. Karl, “Efficiency of on-path and off-path caching strategies in information centric networks,” in *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*, Nov 2012, pp. 581–587.

[101] I. Abdullahi, S. Arif, and S. Hassan, “Survey on caching approaches in information centric networking,” *Journal of Network and Computer Applications*, vol. 56, pp. 48 – 59, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804515001381

[102] X. Tang and S. Chanson, “Coordinated en-route web caching,” *Computers, IEEE Transactions on*, vol. 51, no. 6, pp. 595–607, Jun 2002.

[103] H. Shen and S. Xu, “Coordinated en-route web caching in multiserver networks,” *Computers, IEEE Transactions on*, vol. 58, no. 5, pp. 605–619, May 2009.

[104] M. Korupolu and M. Dahlin, “Coordinated placement and replacement for large-scale distributed caches,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 14, no. 6, pp. 1317–1329, Nov 2002.

[105] A. Bretto, *Hypergraph Theory: An Introduction*. Springer Publishing Company, Incorporated, 2013.

[106] V. I. Voloshin, “Introduction to graph and hypergraph theory.” 2009.

[107] M. Junghans, “Visualization of hyperedges in fixed graph layouts,” Ph.D. dissertation, Thesis, Brandenburg University of Technology, Cottbus, 2008.

[108] V. Sourlas, L. Gkatzikis, P. Flegkas, and L. Tassiulas, “Distributed cache management in information-centric networks,” *Network and Service Management, IEEE Transactions on*, vol. 10, no. 3, pp. 286–299, September 2013.

[109] B. Cesar. (2014, March) Social ccn sim is a ccn simulator. [Online]. Available: https://github.com/mesarpe/socialccnsim.git

[110] C. Bernardini, T. Silverston, and O. Festor, “SONETOR: A social network traffic generator,” in *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3734–3739.

[111] J. Leskovec and R. Sosič, “SNAP: A general purpose network analysis and graph mining library in C++,” http://snap.stanford.edu/snap, Jun. 2014.

[112] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” http://snap.stanford.edu/data, Jun. 2014.

[113] J. Leskovec and J. J. Mcauley, “Learning to discover social circles in ego networks,” in *Advances in neural information processing systems*, 2012, pp. 539–547.

[114] J. Leskovec and R. Sosič, “Snap.py: SNAP for Python, a general purpose network analysis and graph mining tool in Python,” http://snap.stanford.edu/snappy, Jun. 2014.

[115] S. Wang, J. Bi, J. Wu, and A. Vasilakos, “Cphr: In-network caching for information-centric networking with partitioning and hash-routing,” *Networking, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[116] B. Wang, W. Yu, X.-H. Sun, and X. Wang, “Dacache: Memory divergence-aware gpu cache management,” in *Proceedings of the 29th ACM on International Conference on Supercomputing*, ser. ICS ’15. New York, NY, USA: ACM, 2015, pp. 89–98. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2751205.2751239

[117] V. Sourlas, P. Flegkas, L. Gkatzikis, and L. Tassiulas, “Autonomic cache management in information-centric networks,” in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, April 2012, pp. 121–129.

[118] L. Galluccio, G. Morabito, and S. Palazzo, “Caching in information-centric satellite networks,” in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 3306–3310.

[119] K. Suksomboon, S. Tarnoi, Y. Ji, M. Koibuchi, K. Fukuda, S. Abe, N. Motonori, M. Aoki, S. Urushidani, and S. Yamada, “Popcache: Cache more or less based on content popularity for information-centric networking,” in *Local Computer Networks (LCN), 2013 IEEE 38th Conference on*, Oct 2013, pp. 236–243.

[120] S. Saha, A. Lukyanenko, and A. Yla-Jaaski, “Cooperative caching through routing control in information-centric networks,” in *INFOCOM, 2013 Proceedings IEEE*, April 2013, pp. 100–104.

[121] L. T. M. Blessing and A. Chakrabarti, *DRM, a Design Research Methodology*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[122] A. Habbal, “Tcp sintok: Transmission control protocol with delay-based loss detection and contention avoidance mechanisms for mobile ad hoc networks,” Ph.D. dissertation, School of Computing, Universiti Utara Malaysia, 2014.

[123] G. Rossini and D. Rossi, “Coupling caching and forwarding: Benefits, analysis, and implementation,” in *Proceedings of the 1st International Conference on Information-centric Networking*, ser. ICN ’14. New York, NY, USA: ACM, 2014, pp. 127–136. [Online]. Available: http://doi.acm.org/10.1145/2660129.2660153

[124] C. Westphal, B. Mathieu, and O. Amin, “Using ccn for discovery of missing physical items,” in *Proceedings of the 2Nd International Conference on Information-Centric Networking*, ser. ICN ’15. New York, NY, USA: ACM, 2015, pp. 213–214. [Online]. Available: http://doi.acm.org/10.1145/2810156.2812613

[125] M. Guizani, A. Rayes, B. Khan, and A. Al-Fuqaha, *Network modeling and simulation: a practical perspective*. John Wiley & Sons, 2010.

[126] R. Chiocchetti, D. Rossi, and G. Rossini, “ccnsim: An highly scalable ccn simulator,” in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2309–2314.

[127] M. Tortelli, D. Rossi, G. Boggia, and L. A. Grieco, "Ccn simulators: Analysis and cross-comparison," in *Proceedings of the 1st International Conference on Information-centric Networking*, ser. ICN '14. New York, NY, USA: ACM, 2014, pp. 197–198. [Online]. Available: http://doi.acm.org/10.1145/2660129.2660133

[128] L. Saino, I. Psaras, and G. Pavlou, "Icarus: a caching simulator for information centric networking (icn)," in *Proceedings of the 7th International ICST Conference on Simulation Tools and Techniques*, ser. SIMUTOOLS '14. ICST, Brussels, Belgium, Belgium: ICST, 2014.

[129] N. Vastardis, A. Bontozoglou, K. Yang, and M. Reed, "Simulation tools enabling research on information-centric networks," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5833–5838.

[130] N. Vastardis, "The ICN Simulator," July 2014. [Online]. Available: http://privatewww.essex.ac.uk/~nvasta/ICNSim.htm

[131] A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM: NDN simulator for NS-3," NDN, Technical Report NDN-0005, October 2012. [Online]. Available: http://named-data.net/techreports.html

[132] S. Mastorakis, A. Afanasyev, I. Moiseenko, and L. Zhang, "ndnSIM 2.0: A new version of the NDN simulator for NS-3," NDN, Technical Report NDN-0028, January 2015.

[133] M. Hassan and R. Jain, *High Performance TCP/IP Networking*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2003.

[134] A. L. Van Den Wollenberg, "Redundancy analysis an alternative for canonical correlation analysis," *Psychometrika*, vol. 42, no. 2, pp. 207–219, 1977.

[135] L. R. Beaumont, "Calculating web cache hit ratios," *Accessed on*, vol. 2, no. 3, p. 2015, 2000.

[136] C. Bernardini. SocialCCNSim. [Online]. Available: https://github.com/mesarpe/socialccnsim

[137] W. Wang, Y. Sun, Y. Guo, D. Kaafar, J. Jin, J. Li, and Z. Li, "Crcache: Exploiting the correlation between content popularity and network topology information for icn caching," in *Communications (ICC), 2014 IEEE International Conference on*, June 2014, pp. 3191–3196.

[138] S. K. Fayazbakhsh, Y. Lin, A. Tootoonchian, A. Ghodsi, T. Koponen, B. Maggs, K. Ng, V. Sekar, and S. Shenker, "Less pain, most of the gain: Incrementally deployable icn," *SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 147–158, Aug. 2013. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2534169.2486023

[139] L. Saino, I. Psaras, and G. Pavlou, "Hash-routing schemes for information centric networking," in *Proceedings of the 3rd ACM SIGCOMM workshop on Information-centric networking*. ACM, 2013, pp. 27–32.

[140] W. Li, S. Oteafy, and H. Hassanein, "Dynamic adaptive streaming over popularity-driven caching in information-centric networks," in *Communications (ICC), 2015 IEEE International Conference on*, June 2015, pp. 5747–5752.

[141] R. Li and H. Asaeda, "A community-oriented route coordination using information centric networking approach," in *Local Computer Networks (LCN), 2013 IEEE 38th Conference on*, Oct 2013, pp. 793–800.

[142] K. Cho, M. Lee, K. Park, T. T. Kwon, Y. Choi, and S. Pack, "Wave: Popularity-based and collaborative in-network caching for content-oriented networks," in *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*. IEEE, 2012, pp. 316–321.

[143] Dante, "Geant: The pan-european and educational network," 2015. [Online]. Available: https://www.restena.lu/restena/Pdfs/geant3-nov-09-Poster_HighRes.pdf

[144] D. Rossi and G. Rossini, "Caching performance of content centric networks under multi-path routing (and more)," *Relatório técnico, Telecom ParisTech*, 2011.

[145] D. Rossi, G. Rossini *et al.*, "On sizing ccn content stores by exploiting topological information."

[146] C. Bernardini, "Strategies de cache basees sur la popularite pour Content Centric Networking," Ph.D. dissertation, Universite de Lorraine, France, 2015.

[147] C. Bernardini, T. Silverston, and O. Festor, "MPC: Popularity-based caching strategy for Content Centric Networks," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3619–3623.

[148] A. Xu, X. Tan, and Y. Tian, "Design and evaluation of a utility-based caching mechanism for information-centric networks," in *Communications (ICC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 5535–5540.

[149] C. Fricker, P. Robert, J. Roberts, and N. Sbihi, "Impact of traffic mix on caching performance in a content-centric network," in *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*. IEEE, 2012, pp. 310–315.

[150] D. Perino, M. Varvello, and K. P. N. Puttaswamy, "ICN-RE: Redundancy elimination for information-centric networking," in *Proceedings of the Second Edition of the ICN Workshop on Information-centric Networking*, ser. ICN '12. New York, NY, USA: ACM, 2012, pp. 91–96. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2342488.2342508

[151] A. Anand, C. Muthukrishnan, A. Akella, and R. Ramjee, "Redundancy in network traffic: Findings and implications," *SIGMETRICS Perform. Eval. Rev.*, vol. 37, no. 1, pp. 37–48, Jun. 2009. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2492101.1555355

[152] M. Xiao, "Finding minimum 3-way cuts in hypergraphs," *Information Processing Letters*, vol. 110, no. 14, pp. 554–558, 2010.

[153] R. Klimmek, F. Wagner *et al.*, "A simple hypergraph min cut algorithm," 1996.

[154] A. Mahanti, C. Williamson, and D. Eager, "Traffic analysis of a web proxy caching hierarchy," *Network, IEEE*, vol. 14, no. 3, pp. 16–23, 2000.

[155] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 15–28.

[156] Y. Carlinet, B. Kauffmann, P. Olivier, and A. Simonian, "Trace-based analysis for caching multimedia services," Orange labs technical report, Tech. Rep., 2011.

[157] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: Evidence and implications," in *INFOCOM'99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1. IEEE, 1999, pp. 126–134.

[158] H. Che, Y. Tung, and Z. Wang, "Hierarchical web caching systems: Modeling, design and experimental results," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 7, pp. 1305–1314, 2002.

[159] H. Yao, C. Fang, C. Qiu, C. Zhao, and Y. Liu, "A novel energy efficiency algorithm in green mobile networks with cache," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 139, 2015.

[160] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng, "Understanding user behavior in large-scale video-on-demand systems," in *ACM SIGOPS Operating Systems Review*, vol. 40, no. 4. ACM, 2006, pp. 333–344.

[161] G. Barish and K. Obraczke, "World wide web caching: Trends and techniques," *IEEE Communications magazine*, vol. 38, no. 5, pp. 178–184, 2000.

[162] J. Erman, A. Gerber, M. T. Hajiaghayi, D. Pei, and O. Spatscheck, "Network-aware forward caching," in *Proceedings of the 18th international conference on World wide web*. ACM, 2009, pp. 291–300.

[163] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: exploiting content caching and delivery techniques for 5g systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 131–139, February 2014.

[164] M. Amade, C. Campolo, J. Quevedo, and D. CorujoT, "Information-centric networking for the internet of things: Challenges and opportunities," 2015.

[165] P. Agyapong and M. Sirbu, "Economic incentives in information- centric networking: implications for protocol design and public policy," *Communications Magazine, IEEE*, vol. 50, no. 12, pp. 18–26, December 2012.

[166] C. Bernardini, T. Silverston, and O. Festor, "Using social network information into icn," 2013.

[167] M. E. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[168] X. Gabaix, "Zipf's law for cities: an explanation," *Quarterly journal of Economics*, pp. 739–767, 1999.

[169] I. Psaras, R. G. Clegg, R. Landa, W. K. Chai, and G. Pavlou, "Modelling and evaluation of ccn-caching trees," in *NETWORKING 2011*. Springer, 2011, pp. 78–91.

[170] D. O. Mau, M. Chen, T. Taleb, X. Wang, and V. Leung, "Fgpc: Fine-grained popularity-based caching design for content centric networking," *Procedure ACM MSWIM*, pp. 295–302, 2014.

[171] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security*. Springer, 2011, pp. 338–340.

[172] L. A. Adamic and B. Huberman, "Zipfs law and the internet," *Glottometrics*, vol. 3, no. 1, pp. 143–150, 2002.

[173] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design," *arXiv preprint cs/0209028*, 2002.

[174] G. Rossini and D. Rossi, "Coupling caching and forwarding: Benefits, analysis, and implementation," in *Proceedings of the 1st International Conference on Information-centric Networking*, ser. ICN '14. New York, NY, USA: ACM, 2014, pp. 127–136. [Online]. Available: http://doi.acm.org.eserv.uum.edu.my/10.1145/2660129.2660153

[175] ——, "A dive into the caching performance of content centric networking," in *Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2012 IEEE 17th International Workshop on*. IEEE, 2012, pp. 105–109.

[176] I. Cantador, P. Brusilovsky, and T. Kuflik, "Second workshop on information heterogeneity and fusion in recommender systems (hetrec2011)." in *RecSys*, 2011, pp. 387–388.

[177] J. Leskovec, "Social circles: Facebook," December 2015, stanford University. [Online]. Available: https://snap.stanford.edu/data/egonets-Facebook.html

[178] S. Penson. (2014, October) Penguin: What happens next? 10 data-led predictions. Search Engine Watch. Social Network. [Online]. Available: http://searchenginewatch.com/sew/how-to/2375404/penguin-what-happens-next-10-data-led-predictions

[179] D. Chang, H.-G. Yeh, and P. Wang, "Dynamic power allocation via wavefront multiplexing through multiple base stations," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 3789–3794.