This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

Green Symbiotic Cloud Communications: Virtualized Transport Layer and Cognitive Decision Function

H.D. Mustafa*, S.N. Merchant*, U.B. Desai† B.M. Bavejai‡
*Department of Electrical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400055, India.
(mustafahd, merchant)@ee.iitb.ac.in
†Dept. Of Electrical Engineering, IIT Hyderabad, Ordinance Factory Estate, Yeddumailaram, Andhra Pradesh 502205, India
ubdesai@iith.ac.in
‡Dept. Of Electronics and Information Technology, Ministry of Communications and IT, Govt of India, Delhi, 110003, India.
bmbaveja@deity.gov.in

***Abstract*—The evolution of the concept of cloud communications has posed a growing emphasis on virtual and abstract environments for the flow of information, structuring it in similitude to a natural cloud. The Green Symbiotic Cloud Communications (GSCC) paradigm created on this concept facilitates the use of multiple communication mediums concomitantly creating a first of its kind communication cloud. This paper specifically corroborates a virtualized transport layer and network ports and an abstracted Internet protocol scheme in defining the GSCC architecture. We further address the issue of formulating a cognitive decision function based on utility theory, which allows users with GSCC enabled devices to intelligently distribute its bandwidth requirement amongst the available communication mediums. Considering the multiple criteria associated with different networks we formulate an optimization problem to find the solution for this resource allocation problem for single user. We further address the multi-user scenario and formulate and solve the multi-objective optimization problem using goal attainment technique. Results in single and multiple user scenarios, demonstrate that by utilizing multiple mediums as per GSCC paradigm coupled with our proposed decision function improves the functionality of the communication cloud.**

The proposed architecture is dynamic and evolving, embedding greenness by efficiently utilizing the available resources as and when required. The multiple virtual links equate a linearly increasing relationship with the throughput achieved. Experimental results for both real time and static data through the proposed schematic are documented. The augmented paradigm enhances the quality of service, linearly increases throughput and increases the overall security in communications.

***Index terms*— Virtualization, heterogeneous networks, throughput, security, multiple access, socket programming, utility theory, optimization, goal attainment.**

I. Introduction

A cloud is often defined as a visible collection of particles of ice and water suspended in the air, usually at an elevation above the surface. It is generally a dim and obscure area in something otherwise clear and transparent. The clouds appearing in nature even though visible are abstract and virtual, i.e. we are unable to signify a definite boundary of a cloud. The cloud as defined in the field of computing is however very far away from this geographical definition and properties of a natural cloud. Though correlating a cloud with abstraction and virtualization, the existing cloud computing archetypes enfold as backend data or service stations providing bunched or specific services. In building the Green Symbiotic Cloud Communications paradigm, we waver from the existing definition of clouds as outsourced services and define an approach to do justice to the geographical existence of clouds and its emulation in the technological domain. We aim to deviate from the traditional approaches of cloud computing and develop an entirely new way to build, deploy and scale technologies and devices of the future. Paradigms enabling convenient, on demand access to a shared pool of configurable resources that can be rapidly provisioned and released with minimal efforts and interactions constitutes our emblem of a Cloud environment. The GSCC architecture is the first of its kind technology that adopts the idea of cloud communications, wherein abstraction and virtualization currently limited to computing environment, are also embedded in the communications domain an archetype in published literature.

The rapid evolution of the technology used in telecommunication systems, consumer electronics, and specifically mobile devices has been remarkable in the last 20 years. Communication systems handle volumes of data generated by embedded devices, mobile users, enterprises, contextual information, network protocols, location information and such. It is a vast amount of information. For example, a global IP backbone generates over 20 billion records per day, amounting to over 1 Tera Bytes per day! Processing and analyzing this "big data" and presenting insights in a timely fashion will become a reality with advanced analytics to understand the environment, to interpret events, and to act on them. The existing communication systems are just designed as "dumb pipes" to carry information / data from destination to the source. This paper is a positive development that helps unleash the intelligence in communications systems where networks

are no longer labeled as "dumb pipes" but highly strategic and smart cognitive networks. Though there has been development of concepts like MIMO, cognitive radio which target singular link throughput improvements, improved spatial diversity gain, efficient spectrum utilization and increased QoS, the gap between users bandwidth demands and availability is still significant and is expected to increase[1].The next quality of service leap which is fundamentally expected to come from improvements in network topologies, cooperative communication, virtualization and abstraction schemes, the amalgamation of cognitive symbiotic networks and evolving intelligent protocols, all of which is systematically addressed in developing the Green Symbiotic Cloud Communications (GSCC) architecture.

Cloud computing in the past decade has provided a break through in utilizing computing resources efficiently and effectively. According to the traditional definition "Clouds are large pools of virtualized resources which are easy to access, secure and reliable. There are ten characteristics of cloud computing in their sum-up: user friendliness, scalability, resource optimization, pay-per-use, virtualization, Internet centric, variety of resources, learning based adaptation, service SLAs (Service-Level Agreements) and infrastructure SLAs."

The concept of clouds in computing is expanded to communications with evolution of concepts like Cloud Communications[2], which lays the foundation for developing systems of future. It is not difficult to envision that cognitive systems of future will have multiple radio interfaces and will be able to connect to multiple networks simultaneously. The GSCC paradigm as introduced in [3] facilitates the use of multiple communication mediums through virtualization and abstraction yielding a linear increase in communication throughput with minimal power consumption and without minimal addition on infrastructural front.

In recent times, the availability of affordable advanced technological equipments has augmented the exponential surge in network traffic and providers are finding it difficult to support enhanced quality of service and better throughput with the existing infrastructure. It is definitely the need of the hour to support these demands by implementing concepts like virtualization schemes, symbiotic communication, cognitive heterogeneous networks etc. The GSCC paradigm focuses on and evolves around these concepts yielding a comprehensive architecture supporting full duplex cloud communications.

In the existing cloud computing structure the user is oblivious to the back-end processes and is not a participant in the structure of the cloud. The cloud communications concept further extends by integrating the users and other elements of the paradigm as a part of the cloud, facilitating a dynamic and evolving behavior. Specifically in the GSCC paradigm an evolving cloud would mean having servers and clients, equipped with multiple communication mediums and links, permutably connected to each other, communicating seamlessly amongst themselves.

This paper deals with the design of two important aspects of the GSCC architecture: Firstly, the Virtualized Transport Layers and Communication Sockets creating an abstract communication environment and secondly, an intelligent Decision Function that helps in mapping seamless communication in the virtualized and abstracted cloud.

II. Virtualized Transport Layers and Communication Sockets

In this section, we specifically develop the connection capability of devices enabled with the GSCC paradigm to multiple radio access technologies (RATs / CMs) viz, WiFi, LTE, 3G etc. This is achieved by virtualizing the transport layer, followed by virtualizing the communication ports of the devices and finally evolving an abstract and dynamic Internet protocol schematic for the multiple connections. The abstract IP schematic ensures that the multiple connections do not interfere amongst each other while allowing symbiotic, simultaneous and seamless communications.

The scope of this section of the paper targets connections to different RATs / Communication Mediums (CM) at the same instance. However connections to multiple CMs of the same type is not covered here. To quantify the advantage of the proposed schematic, consider a scenario wherein a user is running several applications that have bandwidth requirement of 5 Mb/s, but the available networks, WiFi and LTE, have available bandwidth as 3 Mb/s and 3.5 Mb/s respectively. Individually, these networks will not be able to fulfill user's requirement. However, it is easy to see that if the networks decide to cooperate, they will collaboratively be able to satisfy user's requirement. Once the user is connected to both the mediums WiFi and LTE then the decision function as described in the subsequent section helps in deciding the symbiotic cognitive usage of both these mediums. By using multiple mediums symbiotically, the total capacity achieved would be the linear sum of the individual capacity per link. This would increase the throughput of the device (reduce the overall delay) and intelligently utilize the availability of resources resulting in a greener communication paradigm. Furthermore in the proposed architecture no additional hardware or infrastructure needs to be involved thus making it extremely cost effective and easy to implement.

Existing communication schematics that use multiple mediums concomitantly, are mainly software based and suffer major drawbacks. A software-based approach is adapted in [4], which virtualizes a single wireless adapter as multiple entities. An augmented system based approach proposed in [5] scans for the availability of WLAN networks for the users and subsequently offloads the cellular data through these access points. Coordinated Multipoint (COMP) [6], a concept of spatial reuse, adapted in cellular networks to increase throughput suffers major implementation issues. Generalized heterogeneous networks mainly employ the use of low powered nodes like femtocells and pico-cells, which are in turn powered by macro base stations. Another approach as mentioned in [4] deals with the handover taking place between two different links during abrupt connection termination. An implementation onto the android kernel was also proposed in [7].

In this section we develop the run time process of the GSCC system model including creating virtual transport layers, session creation and running multiple sessions in parallel.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

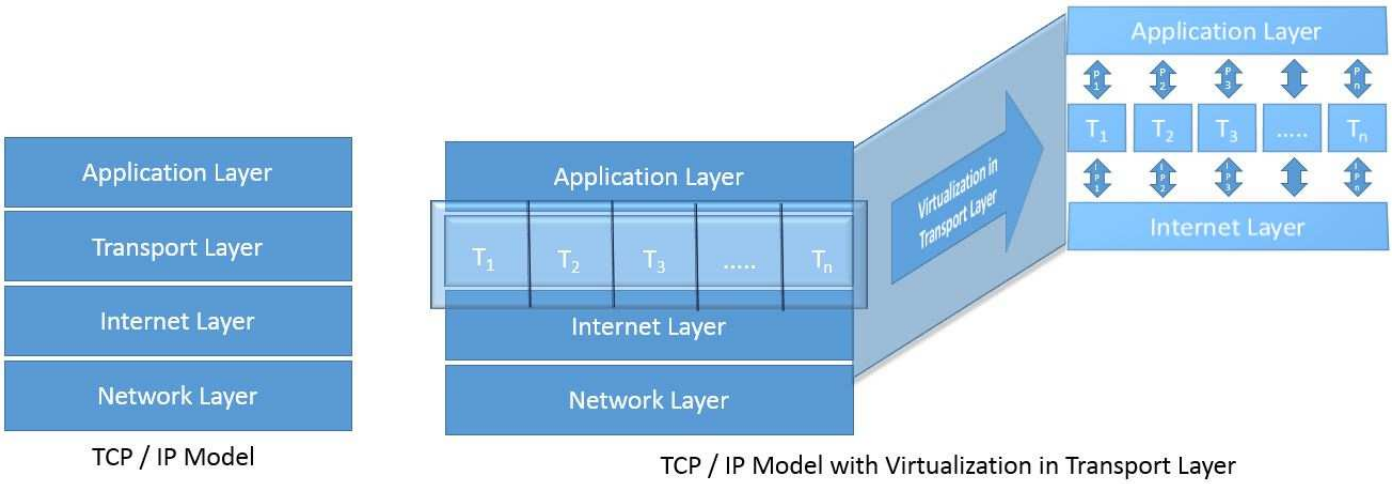

Figure 1. TCP/IP reference model and its modifications: Transport layer is virtualized to multiple sessions. Each session T_1, T_2, ... T_n is communicating to application layer and transport layer through a unique set of [PortNumber, IPAddress]

Simulation and experimental results conducted in static scenario with Ethernet, Wi-Fi access points, 3G access points and compared with the proposed GSCC based paradigm show promising results.

A. System Model: Virtualized Transport Layers

The communication device system model initiates with implementing a virtualized transport layer schematic.

1) Session generation: Consider a generic communication scenario where a duplex communication scenario is to be established between a client to a remote server. We define a communication medium (CM) as any specific available RAT. For instance if a user has three WiFi access points and two LTE communication SIMs then they have access to two distinct CMs and five distinct Communication Links (CL). As per the current schematic enlisted in this section the user can simultaneously connect to any one of the WiFi points and one of the LTE sims. Here a Communication Link (CL) is considered as a connection established from the server to the client using a specific Communication Medium (CM). The transport layer initially implements a decision function as proposed in the subsequent section, which predetermines channel and utility parameters of the network. The cognitive decision function will also identify and select the strongest CM, if multiple CMs of the same kind (2 WiFi's / 2 LTE) are available.

The decision function on the basis of the requirements of the client and server selects the optimum number of CMs needed for completing the communication. The transport layer is then split into multiple parallel virtual layers. The number of these virtual layers are dynamic and changes with the output of the decision function based on the unique CMs available. The information packets are then split among the different links with different sessions each randomly over a predetermined distribution based on inputs from the decision function.

As shown in Figure 1 the virtual transport block on the server side has a logical many-to-many mapping with the virtual transport block on the client side. For instance consider the client side being connected to 3 different unique CMs each with a 0.33 ratio of transmittance for uplink and downlink as decided by the decision function. Similarly the server side is connected to 4 unique CMs each with a 0.25 ratio of information transmittance for uplink and downlink. The GSCC paradigm now establishes a many-to-many link structure, where the 3 links of the client side are mapped to the 4 links of the server for full duplex communications. The distribution of the packets transmitted over a such a communication cloud, increases the complexity of reassembling the received packets at both the server and client side in the correct sequence. To counter this challenge a buffer is introduced on both the server and client sides to store the unsorted data and reassemble the streamlined data segments at the application layer.

The total number of links possible is a function of the available network adapters at both client and server side. The maximum number of logical links possible are $N_s \times N_c$ where N_s and N_c are the number of connections on server and client side respectively. This is due the possibility of a many-to-many logical mapping.

However, under the assumptions of a one-to-one mapping, the maximum possible throughput is the linear sum of throughputs of all individual links and maximum power consumption during the transfer of data is the combined power consumption of all links. Figure 2 describes the entire process of the duplex communication scenario.

2) Structure of the virtual TCP/IP stack: With reference to the requirement, the virtual transport layer performs the functionality of transferring the data from server to client or vice-versa. The virtual transport layer is divided in to 2 segments:

- Decision function
- Multiple transport layers

3) Decision function: The Decision function has a crucial role in the segmentation and reassembling of data. Assume a scenario where a file needs to be sent from the server to the client. The server divides the file into n different segments and is transmitted simultaneously through the multiple links

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

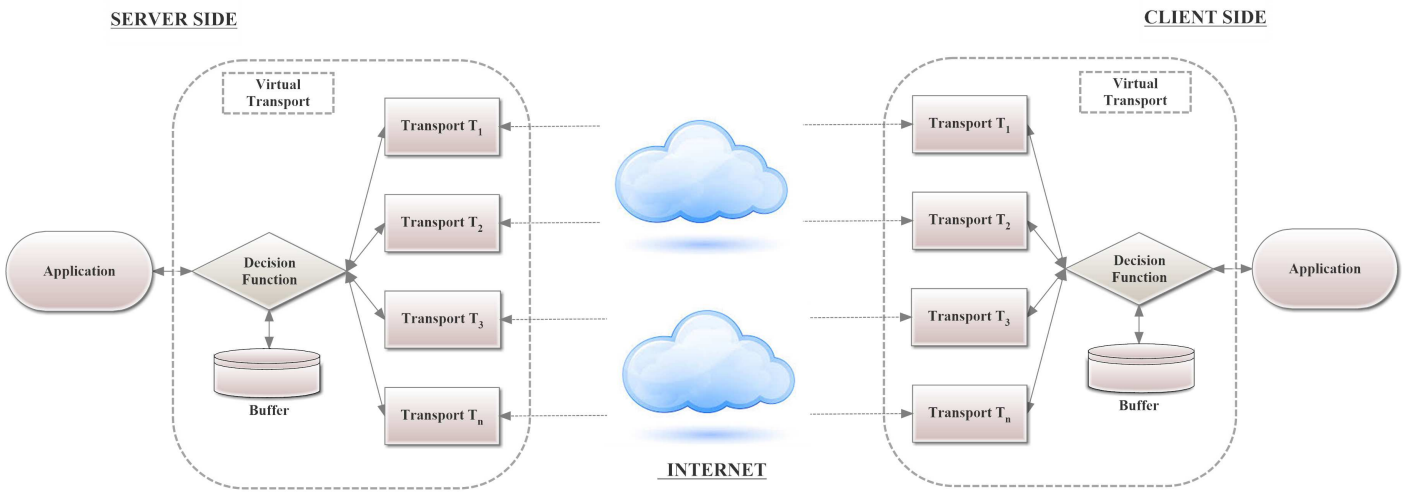

Figure 2. Process layout

created. Depending on the results of the cognitive decision function it will split the file in order to optimize the total time taken and the energy consumed. The decision function also adheres to the requirements as set by the user, in which preferential setting for any of the parameters could be done and weight-age to the parameters is so allocated in the decision making process. Optimization of energy and throughputs is done depending on the multiple logical connection scenarios and user preference.

We define the communication clouds under the following classification:

- Homogeneous clouds: When the links are homogeneous, the energy consumed for transferring the data is constant Therefore there is no energy factor involved during the splitting.
- Heterogeneous clouds: These links are different as they use different last hops such as LTE, 3G, WIFI and Ethernet. For instance, if LTE consumes the least power but has lower throughput and if the user requires the data at higher speed the decision function acts accordingly by considering only maximizing throughput compromising on the energy consumption. And if the user wants the system to run in power saving mode then the decision function disables the higher power consuming CMs, and uses only lower ones in ascending order of throughput requirement.

The decision function is dependent on :

1) Capacity of the link $\{T_1, T_2, T_3, ...T_n\}$
2) Power consumption of the link $\{P_1, P_2, P_3, ...P_n\}$
3) Cost of the link $\{C_1, C_2, C_3, ...C_n\}$

Assume there are n network adapters for a machine and T_i , P_i and C_i denotes the capacity, power consumption and the cost of the link for transferring data for the i^{th} link respectively.

4) Multiple Transport Layers: Data after division will be sent through different transport layers which corresponds to the different network adapters available. The modified TCP/IP stack will appear as shown in Figure 1. T_1, T_2 ... T_n are the original transport layers. These layers can be used for creating multiple sessions for transferring data in parallel using the socket API. Thus the virtual transport layers present in the stack are independent on the n different network mediums present e.g. WiFi, 3G/LTE, Ethernet etc. from the application viewpoint.

The Decision function further governs the generation of sessions. Primary and secondary links are created according to the priority order of the connections, depending on outputs of the decision making process. The client initiates a request and the server acknowledges this to initiate a handshake and generate sessions. For a session to be created, client creates a socket which will be used as the unique identity vector C_i containing $\{\text{clientip}_i, \text{clientport}_i, \text{serverip}, \text{serverport}\}$ where clientip_i and clientport_i are the IP address and the port number alloted to the i^{th} network interface. Similarly we have another vector S on the server which contains $\{\text{clientip}_i, \text{clientport}_i, \text{serverip}, \text{serverport}\}$. This unique vector is used by the sockets to differentiate between the virtualized sockets. When the client needs to download/upload data this vector will be used in address headers to identify the client. The data will be sent along with the packet numbers in order to assemble at the other end.

The problem occurs when data is sent in multiple sessions as the data needs to be assembled correctly at the receiving end. Thus after dividing the data in to packets, they will have a new local virtual packet number which will be embedded in the data field of the packet frame and a session's packet number that will be included in the packet number field of the packet frame present in header of the packet. Session packet numbers, which are in sequence, are present in the header of the packet as the firewalls / NATs will start discarding the packets if continuity in the packet number is not preserved.

At the server end the parameters are same as that of the client except for an additional parameter status that is used to record the current status of the link. During the time a client initiates a request and the server acknowledges it, a handshake is initiated where a list containing the priority order of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

links along with the throughputs, cost and power consumption of the links is transferred from the client to the server. This information is further dynamically updated at regular time intervals.

For simplicity let us assume that the file is split uniformly into n segments. Since all n segments are sent simultaneously, the i^th segment of the data will travel on the i^th link. Packet reassembling is done at the application layer. However we will require an additional field $datasequence$ in the packet frame which will store the segment number. The connections after they are established and the dynamic duplex cloud communication scenario of the GSCC paradigm with the proposed virtualization schematic of the transport layers, is shown in Figure 3.

5) Creating multiple virtual sessions: In the existing state of the art communications, a server stores the session vector for defining the session. If multiple sessions are created from the same client but from a different network adapter (different IP addresses) the server will not be able to recognize the link if its from same client or a different one. Creating multiple sessions from the same client with n different network adapters with the GSCC paradigm helps in increasing the throughput. Since each network adapter has a specific IP address, the client has a pool of n IP addresses. Therefore while initiating multiple connections to the server we have to make certain changes to the default TCP protocol in order for the server to recognize that the same client is creating those multiple sessions.

We define each connection from the client to the server as a link. We classify the links into two different types:

1) Primary link
2) Secondary link

The primary link firstly initiates a handshake with the server. This link is chosen by the decision function according to the priority order and inputs if any from the users preference. This link is used for sending information regarding all the other available links from client which can be used for establishing the connection. Apart from sending this initial modified handshake information it will act as a normal link for sending the data packets. Other links that are created serve as secondary links and are used for sending only the packet data.

A buffer is then created for allowing seamless duplex communications. The main functionality of the buffer is storing the data till the layer receives a successful acknowledgment from the receiver for the earlier transmission. If the data is not received, retransmission of the data is done directly from buffer saving on overheads and reducing delay. At the receiving end it is also used for sorting the received data before sending it to the application layer i.e., assuming if the server is sending / receiving data through multiple links there will be delay of some packets or some packets might get missed. Thus until the ordered sequence is received, data will be stored in the buffer which will be later on assembled and sent to the application layer.

6) Data transfer on virtualized communication links:

a) Initialization: The decision block initializes by detecting the number of network adapters present on the communication device. Decision function calculates the priority of the links using the parameters such as channel conditions, throughput of the link, user defined or model based power consumption and cost of the link.

b) Prioritizing the links: The decision block then creates a table containing the possible available links to the server, which are sorted according to their priority. These are stored according to the priority in an array using IPs and port numbers.

c) Creation of the communication Cloud: Highest priority link will be created first and considered as primary link. While handshaking it will discover whether the server is also GSCC enabled. If server is GSCC compliant, the remaining connection links will be created according to the priority and are considered as secondary links. Once these connections are established they are executed in parallel using the socket API.

d) Multiple sessions handling: The Server has a table of IP address, Port number, Primary session ID and their Status (connected/disconnected). The Server needs to know all the IP addresses linked up with the client to recognize the connections from same client. Once the primary link is created it transfers the priority table to the server during handshaking. The Primary Session ID is used as unique key for identifying the client. Once the secondary connections are established the status field turns into $connected$ in the table. During uplink the server receives data from all the connections and using the table, it will assemble the packets in the buffer and send to the application layer. During downlink, server stores some data in the buffer and sends the data to all the links, which are having connection status as $connected$ in the table for that particular client.

e) Closing the Cloud: Any link can send an ack to close the session. Closing the session will clear all the IP and Port address attached to the client primary session ID.

7) Connection errors: Connection errors are due to connection breakage. It is detected when data cannot be transferred through that link. We consider the following connection error scenarios:

- The primary link is broken
  When the primary link is broken the client can no longer update the server table. In order to remove such an error, the server chooses the next connected link from the priority table as the primary link. Same is the case at the client side, it will consider the next connected connection as primary link from the priority table. The connection status of the erroneous link is then changed to $disconnected$ and the decision block stops sending data through that link.
- Secondary link is broken
  If the secondary link is broken the connection status of the link is changed to $disconnected$ and the decision block stops sending data through that link.
- Broken link is re-created
  Client keeps on trying to establish the broken link at regular predefined time intervals and if it succeeds in establishing a broken primary link both client and server changes it to $connected$ status. The reconnected link can again be classified as primary or kept as secondary

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

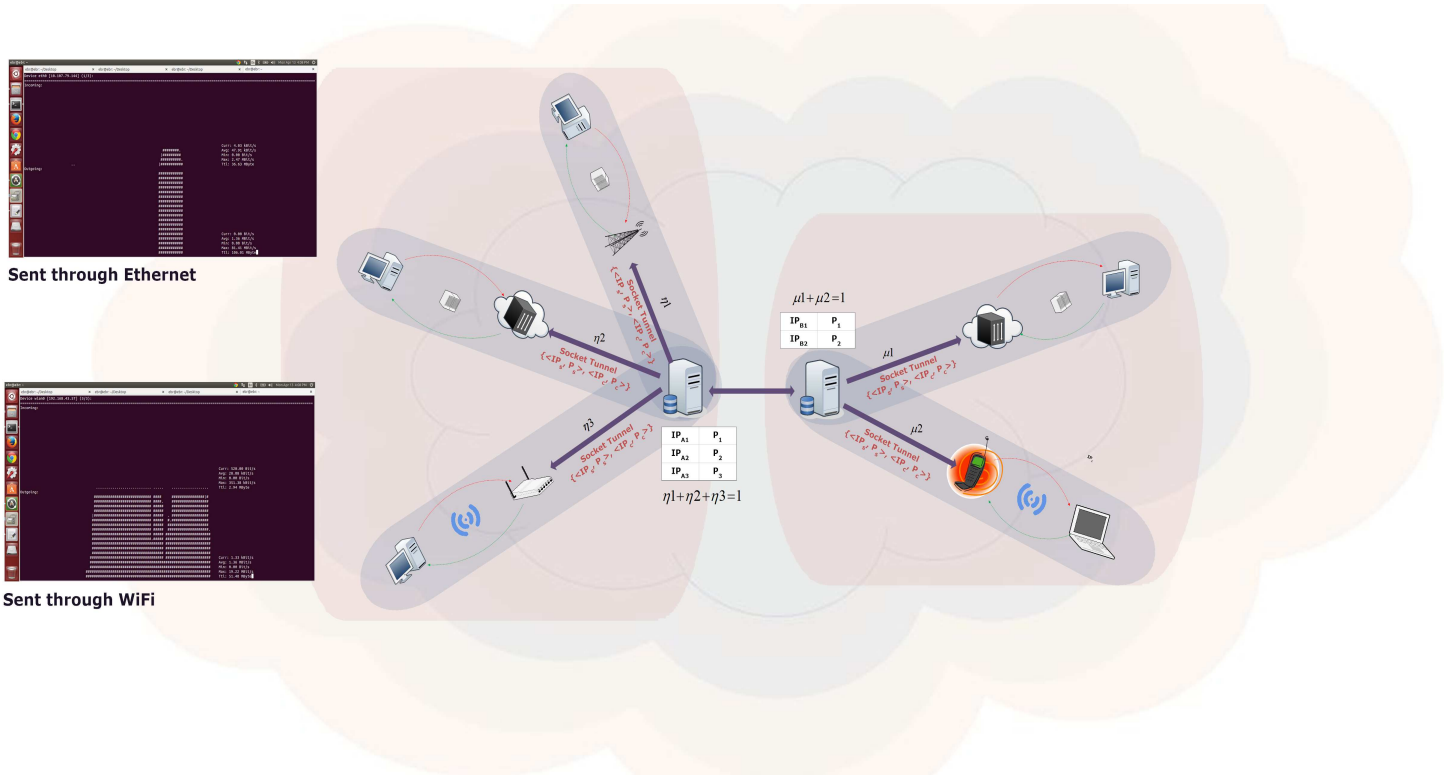

Figure 3. Server is connected to multiple clients via socket tunnel through different communication links. As more clients enters the system, cloud starts evolving. n_1, n_2, n_3 and u_1 u_2 indicates the fraction transferred through different links decided by the decision function. Each socket is a unique combination of port number and IP address through which servers and clients are sending or receiving data. Data transferring through multiple links simultaneously is shown in the two figures on the left.

based on its evaluation by the decision function. The data transmission is restored on the re-connected link and if required the existing primary link is changed to secondary. If a secondary link establishes re-connection, the status of the link will be changed from disconnected to *connected* and the data transmission will be restored through the link. The decision function will not be evaluated again in case of secondary link, to reduce on the system overheads.

The client and server sides algorithms are enlisted as follows:

III. Cognitive Decision Function

It is not difficult to foresee that communication devices of next generation will have multiple radio interfaces and will be able to connect to multiple networks simultaneously; however, not much effort is directed towards using multiple CMs simultaneously. The Virtualized Transport Layer architecture of the GSCC paradigm as described in the previous section facilitates the use of multiple communication mediums through virtualization and abstraction yielding a linear increase in communication throughput with minimal power consumption, without minimal addition on infrastructural front. The GSCC paradigm stresses the need for a decision function that cognitively allows the use of these varied communication mediums simultaneously. To quantify this problem, consider the same scenario again wherein a user is running several

Algorithm 1: Client side process

Data: Initializes port numbers and IP addresses of the client
Input the desired location;
while *Connection is not established* **do**
  check for host;
  if *Host is unreachable* **then**
    initiate random backoff;
    attempt for connection;
  else
    Create an end communication socket;
    Connect to host using request and wait for acknowledgement
Allocate Memory to the buffer and keep application on standby;
while *Check for connectivity;*
do
  initiate data transfer;
  check the status;
  if *connection terminated* **then**
    update the server and wait until connection re-established
  else
    check transfer status;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

Algorithm 2: Server side process

```
Data: Initializes port numbers and IP addresses of the
      server
;
while Listen for incoming connections do
    if Connection using GSCC protocol then
        Accept initialization headers ;
        Connect to GSCC client;
        Create database containing IP addresses and
        Port numbers;
        Create end point communication segments;
        initialize buffer and segment the data;
    else
        Create an end communication socket;
        Connect to host using request and wait for
        acknowledgement
Select decision criteria;
Commence data transfer;
Check and update the status of the transfer;
Terminate connection when data transfer complete;
```

applications that have bandwidth requirement of 5 mb/s, but the available networks, Wifi and LTE, have available bandwidth as 3 mb/s and 3.50 mb/s respectively. Individually, these networks will not be able to fulfill user's requirement. However, it is easy to see that if the networks decide to cooperate, they will collaboratively be able to satisfy user's requirement. Furthermore, the question arises is what would be the best division of the required 5 mb/s and how can we dynamically adapt the paradigm to the changing conditions of the communication cloud? We aim to address this problem in this paper by developing a decision function suited to the GSCC paradigm.

Substantial work has been done to solve the issue of best network selection in heterogeneous networks scenario. In [8] presented an overview of different mathematical models and approaches to select best network amongst those available. They discussed the different tools used to solve this problem like utility theory, MADM, combinatorial optimization, game theory, etc. We target the problem of developing a decision function that allocates the bandwidth requirement of a user among the multiple Communication Links using utility theory. We assume that the communication devices are GSCC enabled and has multiple radio interfaces and are able connect to and communicate using multiple CMs as described earlier. We also investigate the multi-user version of this scenario, wherein multiple users are competing for available bandwidth of CMs.

The development of the decision function is organized as follows where we introduce the adapted system model in next subsection, followed by the structuring of the proposed optimization problem for single and multiple user scenario for the GSCC Paradigm.

A. System Model

We adapt the system model used by Kosmides et. al. [9] as shown in figure III-A in which, each CM informs the control plane about the bandwidth that it can provide to the users, referred as $b_i,\ i \in \{1, 2, ..., N\}$ for i^{th} network. Every user also informs the control plane about its utility functions for various criteria and its bandwidth requirement. Let bandwidth requirement for i^{th} user be $M_i,\ i \in \{1, 2, ..., k\}$ and utility functions be $f_{ij},\ i \in \{1, 2, ..., k\},\ j \in \{1, 2, ..., r\}$ where r are the number of criteria based on which a decision is to be made. The control plane solves the optimization problem and informs the users about their respective allocation vectors.

B. Optimization problem

1) Single user: In the single user scenario we consider that a user has multiple CMs to which it can simultaneously connect and is the only user operating on all the CMs it is connected to. Consider a set of criteria c, an allocation vector α and the bandwidth required by the user M_0 as:

- $c = (c_1, c_2, ..., c_r)$ - criteria set - A set of network criteria based on which a decision is to be made
- $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n),\ \alpha_i \in [0, 1]\ \forall\ i \in \{1, ..., N\}$ - allocation vector
- M_0 - bandwidth requirement of the user

Thus, bandwidth allocated to i^{th} network is $\alpha_i M_0$.
The optimization problem for the single user becomes -

$$\max\ U \tag{1}$$

$$\text{subject to } \sum_{i}^{n} \alpha_i = 1 \tag{2}$$

$$\alpha_i M_0 \leq b_i\ \forall\ i \in \{1, 2, ..., n\} \tag{3}$$

where $U = \prod_{i=1}^{r} [U_i]^{w_i}$ is the total utility obtained from all the criteria(w_i is the relative weight assigned to i^{th} criterion, $\sum_{i=1}^{r} w_i = 1$). This way of aggregating utility functions to evaluate multi-criteria utility is suggested by Nguyen-Vuong et. al [10].

Intuitively, by this formulation of optimization problem, we're interested in an allocation which will maximize user's utility. Constraint (2) suggest that the bandwidth requirement of the user is fulfilled. Constraint (3) suggests that the bandwidth allocated to a network can't exceed the bandwidth that the network is able to provide.

2) Multi-user: In case where k users are connected to any specific CM, the optimization problem becomes -

$$\max_{\alpha \in [0,1]}\ U_i\ \forall\ i \in \{1, ..., k\}$$

$$\sum_{j=1}^{n} \alpha_{ij} = 1\ \forall\ i \in \{1, ..., k\}$$

$$\sum_{i=1}^{k} \alpha_{ij} \leq b_j\ \forall\ j \in \{1, ..., N\}$$

where $U_i = \prod_{j=1}^{r} U_{ij}$ is the total utility for i^{th} user. The optimizer for the above problem gives an allocation matrix where i^{th} row correspond to allocation for i^{th} user.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

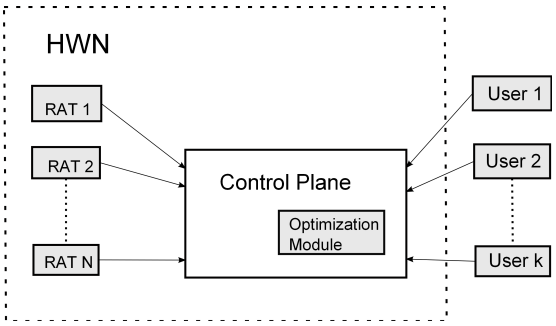

Figure 4. System Model

3) Utility Theory: In microeconomics, utility refers to amount of satisfaction obtained by consumption of a good or service [10]. Utility function maps from a value of a good/service to the utility obtained by it. Depending on users preferences, same value of a good/service may give different utility to different users. For single criterion decision making problems, it is fairly straightforward to directly use the utility and make a decision. For multi-criteria decision problems, the utility of different parameters can be combined together by some mathematical operation, also incorporating the relative preferences of the different parameters for a user, and a decision can be made.

One might argue why to use utility functions at all. One could instead formulate an optimization problem to minimize power or formulate a multi-objective problem to minimize power and cost. However, solution for such a formulation may not give user the same utility as a formulation of above form. So an optimization problem maximizing utility is more intuitive.

Nguyen-Vuong et. al. studied the different utility functions for single criterion and aggregate utility function forms in the context of wireless network selection and came up with conditions suitable for an ideal utility function. They proposed that a utility function of the following form that satisfies all the discussed conditions [10]:

$$u(x) = \begin{cases} 0 & x < x_\alpha \\ \frac{\left(\frac{x-x_\alpha}{x_m-x_\alpha}\right)\zeta}{1+\left(\frac{x-x_\alpha}{x_m-x_\alpha}\right)\zeta} & x_\alpha \leq x \leq x_m \\ 1 - \frac{\left(\frac{x_\beta-x}{x_\beta-x_m}\right)\gamma}{1+\left(\frac{x_\beta-x}{x_\beta-x_m}\right)\gamma} & x_m < x \leq x_\beta \\ 1 & x > x_\beta \end{cases}$$

where

$$\gamma = \frac{\zeta(x_\beta - x_m)}{x_m - x_\alpha}$$

$$\text{and } \zeta \geq \max\{\frac{2(x_m - x_\alpha)}{x_\beta - x_m}, 2\}$$

ζ and γ are the tuned steepness parameters.

They also proposed that a suitable aggregation for multi-criteria utility function is formulated as

$$U(x) = \prod_{i=1}^{n} [u_i(x_i)]^{w_i} \quad (4)$$

where n is the number of criteria, w_i is the weight vector for criterion i($\sum_{i=1}^{n} w_i = 1$) and $u_i(x_i)$ is the elementary utility of criterion i that follows the utility form mentioned above.

4) Multi-objective optimization: A basic multi-objective optimization problem is mathematically described as

$$min[f_1(x), f_2(x), ..., f_n(x)]$$
$$x \in S$$

where n > 1 and S represents set of feasible points.

The concept of optimality does not directly apply in the context of multi-objective optimizationand hence we adapt the concept of pareto-optimality [11]. A feasible point $x*$ is said to be pareto-optimal if for no $x \in S$, all the objective functions improve over $x*$. In our proposed scenario, pareto-optimality is as follows:

- Weak pareto-optimality - $\nexists x \in S$ such that $f_i(x) < f_i(x^*)\ \forall\ i \in \{1, 2, ..., n\}$
- Strong pareto-optimality - $f(x^*) \leq f(x)\ \forall\ x \in S$ and $\forall\ i \in 1\{1, ..., n\}$ with strict inequality for atleast one i.

The image of all pareto-optimal points under $F = [f_1(x), f_2(x), ..., f_n(x)]$ is called pareto-curve or pareto-front. The points on pareto-front are also called non-inferior or non-dominated points.

In principle we're interested not in pareto-front but a particular optimizer for the problem. Hence, there is a need of decision function that provides subjective performance preferences, to choose the best solution among the set of pareto-points. A basic categorization is made of the techniques for solving multi-objective problems based on the instant at which decision is required to provide preference information:

- Prior to the search (a-priori approaches)
- During the search (interactive approaches)
- After the search (a-posteriori approaches)

Of the several techniques available to solve these problems, we used goal attainment for our scenario because the quantities needed to characterize this technique have a simple intuitive interpretation in our scenario. Goal attainment is an a-priori approach in which the decision making preferences are available before the search begins. Mathematically, for the above

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

problem, the goal attainment gives the following optimization problem:

$$\begin{aligned} & min\ \alpha \\ \text{subject to } & f_i(x) - \alpha w_i \leq z_i^{ref}\ \forall\ i \in \{1, 2, ..., k\} \\ & \sum_{i=1}^{k} |w_i| = 1 \\ & x \in S \end{aligned}$$

It has been shown [11] that an optimizer for the above problem gives a pareto-optimal solution.

To characterize our multi-objective optimization problem using goal attainment technique, we need to define the goal and the weight vectors as $z^{ref} \in \mathbb{R}^k$, $w \in \mathbb{R}^k$ respectively.

- **w** - w reflects the relative amount by which under- or over-attainment of the desired goals is allowed. It gives an indication of the priority order of the objective functions. In our case, as all the users are of equal priority, $w = [1/k, ..., 1/k]^T$, where k = number of users, k > 1.
- $\mathbf{z}^{ref}$ - z^{ref} is the goal vector that we want the objective functions to achieve. In our case, the goal vector is the maximum value of the utility of a user when it is not competing with any other user i.e. when the full bandwidths of all the networks are available to it.

IV. Experimentation and Results

The experimental setup comprises of a regular personal computers communicating with ISP servers. We implement the proposed algorithm of virtualizing the transport layers and API sockets using GSCC designed programmable ethernet, wifi and 3G shields as the communication interfaces to the client. Each communication link established by these interfaces are connected through heterogeneous networks independent of each other but converge to the host server, creating a cloud of communications. During the initial experimental setup, we implement the modified proposed file transfer protocol (FTP) using our algorithm on both the server and client making them GSCC enabled.

A. Case I: Generic Internet Browsing and File Downloads

The task performed is to connect to the remote server and retrieve specific files of interest. The size of each of files is designed to be large ($\approx 500MB$). A comparative experiment was performed to find the throughput of the system and its performance using normal FTP on each individual link to the server as to when the task is performed in the GSCC paradigm using all the links at a time.

In our scenario, Ethernet posses the highest throughput and least power consumption, whereas 3G has the lowest transmission rate and highest transmission power consumption. When the requirement is to save energy, the decision function will establish connection using only Ethernet, subject to the Ethernet being able to satisfy the minimum required throughput. If high throughput is demanded, the system will utilize all the available connected links for the transmission of data and decision function will split the data among various links in-order to maximize the throughput. In case where throughput is not being satisfied by one particular link and energy conservation is also required additional links will be added to the primary link on basis of their priority.

Table I
Experimental parameters for transmitting a 500MB file through the GSCC paradigm.

Experimental results	Ethernet	WIFI	3G	GSCC
Maximum throughput (Mbps)	9.2	4.6	2.4	15
Average throughput (Mbps)	8.5	4.25	2.1	14.1
Power for transmission (dBm (mW))	9 (8.8)	15 (32)	24 (251)	25 (292)
Time taken(Sec)	57	96	181	39

The decision function is dependent on four main parameters namely channel conditions, throughput, cost of link and power consumed. In this experiment we try to maximize throughput, neglecting the other parameters as it demonstrates the most common practically experienced scenario. Hence we use all the three links simultaneously to transfer the data.

The experiment is intended to show the improvement in throughput though neglecting the energy saving in this scenario. We have captured the parameters of 500 MB file transfer through different links which are shown in Table I. The experiment is averaged over 150 different trials in different operating scenarios and data content. We observe that with a properly functioning cognitive decision function we are able to obtain a near complete linear summation of the throughputs of each link without addition of any new infrastructure.

B. Cognitive Decision Function enabled Duplex Communications in GSCC

For simulation, we consider a heterogeneous network scenario consisting of Wifi (IEEE 802.11g), GSM / LTE and WiMax (IEEE 802.16 $-$ 2004 version) as the available RATs. The criteria set for the allocation problem is **c** $=$ {Power, Cost}.

To calculate power at the allocated bandwidth, we adapt the power model of [12] (for WiMax, Wifi and GSM):

$$P_{Wifi} = 4.652 + 0.024f\ W \tag{5}$$

$$P_{GSM} = 0.024 + 11.9f\ W \tag{6}$$

$$P_{WiMax} = 16 + 0.174f\ W \tag{7}$$

$$P_{Total} = P_{Wifi} + P_{GSM} + P_{WiMax} \tag{8}$$

where f is allocated bandwidth in Mb/s. This model is only for the transmission power when adaptive modulation scheme is used, and the datagram is of size 1280 bytes.

To calculate cost at the allocated bandwidth, we assumed a linear model of cost vs bandwidth as shown in figure 5 with different slopes for different CMs. Specifically, for simulations, we used the cost/bandwidth of {WiMax, Wifi, GSM} as {2,1,3}.

Total cost is given by -

$$c_{Total} = c_{WiMax} + c_{Wifi} + c_{GSM} \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

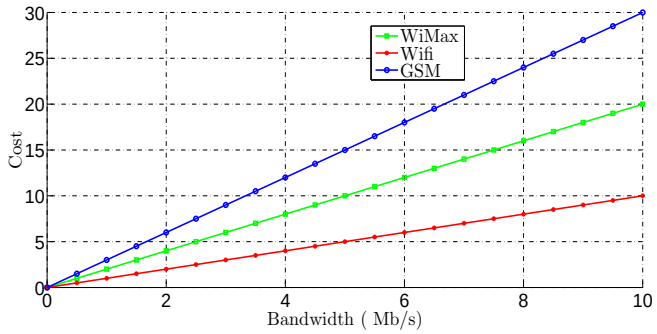

Figure 5. Cost vs bandwidth

Table II
PARAMETERS CHARACTERIZING UTILITY FUNCTION FOR USER 1

Parameters	Cost utility	Power utility
x_α	0	0
x_β	5	18
x_m	2	10
ζ	2	2

Table III
PARAMETERS CHARACTERIZING UTILITY FUNCTION FOR USER 2

Parameters	Cost utility	Power utility
x_α	0	0
x_β	5	18
x_m	3	8
ζ	2	2

1) Single user case:

- CMs, **N** = {WiMax, Wifi, GSM}
- Available bandwidths of networks, **b** = {500, 500, 500} Kb/s ({0.488, 0.488, 0.488} Mb/s)
- Relative weights, **w** - It refers to the relative importance of different parameters in the multi-criteria problem (not to be confused with **w** of goal attainment). In their work by Song et. al. [13], they demonstrate the use of Analytic Hierarchy Process (AHP) to determine the relative weights of different criterion. However, WLOG and for the sake of simplicity, we did not use AHP and instead choose to assign equal priority to power and cost $\implies$ **w** $= \{1/2, 1/2\}$

Utility function is characterized by specifying $(x_\alpha, x_\beta, x_m, \zeta)$. The parameters for utility function are described in the table II:

The parameters are chosen considering the entire range of power consumption and total cost for range of bandwidth requirement from 300 - 1500 Kb/s. The formulation of an appropriate utility function in different practical conditions is still open ended and is discussed more in a later section.

Thus the optimization problem for single user is as follows

$$\max_{\alpha} u_p u_c$$
$$\sum_{i=1}^{3} \alpha_i = 1$$
$$\alpha_i M_0 \leq b_i \; \forall \; i \in \{1, 2, 3\}$$

Figure 6 shows allocation to different networks versus the bandwidth requirement of user, keeping the available bandwidths of the CMs as fixed (b = [500 500 500] Kb/s).

We observe that at lower bandwidth requirement of the user, GSM is preferred as its power consumption is significantly low. As the bandwidth requirement increases, the power benefits for using GSM decreases and the cost benefits of Wifi increases. As such, user's requirement is split between GSM and Wifi. At $\approx 400Kb/s$, Wifi is both cost-efficient and power efficient than other CMs and hence the user uses only Wifi as its CM and does not use any bandwidth from other CMs. Above 500 Kb/s, as Wifi alone cannot fulfill the user's demand, the function uses full available bandwidth of Wifi and allocates the remainder to GSM. Above 600 Kb/s, though WiMax is less power-efficient than GSM, it has less cost/bandwidth and hence some allocation goes to WiMax. This demonstrates the compromise between the utilities of cost and power and the requirement of a cognitive decision function that constantly evaluates the networks and conditions available to the users.

2) Multi-user case: For multi-user scenario, we consider a simple case of 2 users. The first user is more sensitive to cost than power. This means that the utility changes sharply with change in cost; however the change in utility for the corresponding change in power is less. The seconds user is modeled as more power sensitive, with the context of sensitivity taken as explained earlier. To capture the complex interaction between the users, we assume that user 1 and user 2 have bandwidth requirements of 400 Kb/s and 450 Kb/s respectively. The parameters characterizing the utility functions of user 2 is shown in Table III.

The optimization problem is structured as:

$$\max_{\alpha} [U_1, U_2]$$
$$\text{subject to} \sum_{i=1}^{3} \alpha_{1i} = 1 \, , \sum_{i=1}^{3} \alpha_{2i} = 1$$
$$\alpha_{1i} M_1 + \alpha_{2i} M_2 \leq b_i \; \forall \; i \in \{1, 2, 3\}$$

where $U_1 = u_{1p}.u_{1c}$ and $U_2 = u_{2p}.u_{2c}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

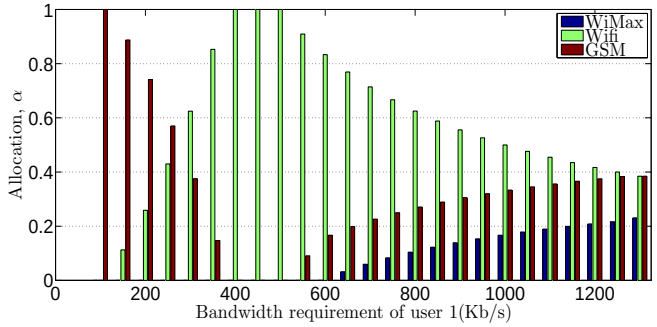

Figure 6. Allocation vs Bandwidth requirement of user 1.

For the following plots, x represents the available bandwidth of all the networks. For example, at $x = 400Kb/s$, the available bandwidths are $[400400400]Kb/s$.

In the following results in figure 7 to figure 8, we vary the available bandwidths with the networks keeping the requirements of both users as constant and observe the allocations of different users.

In the results, we can clearly see the compromise between the users. Initially, when b = [300 300 300], user 1, who is more cost sensitive, should have allocated most to Wifi. However, due to conflict of utility with user 2, it has to compromise by allocating more to WiMax (which is the next best in terms of cost). We also see that user 2, who is more power sensitive, does not allocate anything to WiMax in the entire range of bandwidth requirement from 300-1000 Kb/s. This is because WiMax is least power-efficient in this bandwidth range, and also due to the fact that GSM is not so preferable to user 1 as compared to user 2 (due to higher cost/bandwidth). Also, the total bandwidth requirement of both the users is 850 Kb/s. So when the available bandwidth with the individual CMs exceeds 850 Kb/s (i.e. all the networks individually can support both the users), both the users use only Wifi, which is both power efficient and cost efficient in that bandwidth range.

Figure 9 Shows the total allocation to different CMs and their available bandwidths:

Initially, full bandwidth of GSM and Wifi is used, but not of WiMax as the utility it brings is less in that bandwidth range. We can see that the full available bandwidth of Wifi is used at all points. This is intuitively expected as Wifi is power and cost efficient. We also observe that WiMax and GSM are allocated lesser bandwidth as the available bandwidths of the CMs increase because users are allocating more and more to Wifi, and the benefits of using GSM and WiMax are decreasing.

An added advantage of the proposed GSCC paradigm is receiving added security when a cloud of communications is formed. Consider a scenario where a connection has to be established between two entities for duplex communication. Traditional secured communication protocol establishes a single link for data transfer with only encrypted security. An attacker knowing the encryption may attempt to penetrate into the network to capture the data transmitted and received in the connection. The penetration has more chances of success as we are transmitting over a single link.

In the proposed GSCC paradigm the multiple communication links that are established creates a cloud of communications. The data is traversing over multiple links simultaneously. Now the data transfer at this point will be proceeded such that the data will be distributed randomly and uniformly over the links established as decided by the decision function. Now if the attacker attempts to penetrate into the network, they will have to hack into all the links simultaneously at the same time to obtain the full data. Thus the proposed GSCC paradigm establishes a secure cloud which becomes very difficult for intruders to penetrate.

V. Conclusion

The GSCC paradigm adhering to the concept of cloud communications benefits from incorporating our proposed algorithms and virtualization schematics by gaining a linear addition in the throughput for the user with minimally increased power consumption and added security. The results obtained, clearly corroborate an improvement in the overall throughput of the system. We have also notice an overall increase in the power consumption during the communication period in the proposed protocol. This overall increase in power is compensated by the reduction in time required for overall transmission implying that our energy required for the overall task is constant. This constant energy output has been augmented by the increased throughput without any additional infrastructural burden or cost to the communication architecture.

Virtualizing only the transport layer using socket programming has some limitations as it can result in increased overheads. Those limitations will be addressed in our future endeavors, where complete virtualization of network cards is targeted. Furthermore the current paradigm restricts the usage to unique RATs / mediums. Multiple RATs of the same type can be utilized by embedding virtualization at all levels. For instance, the network cards available are capable of transferring data at very high rates, but due to bottleneck somewhere in the path, they don't utilize their capacity. If a traditional wifi card has a capacity of around 54 Mbps, it might not be fully utilized and generally under very good scenarios we might obtain speeds around 4-5 Mbps. Hence about 50

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2644727, IEEE Access

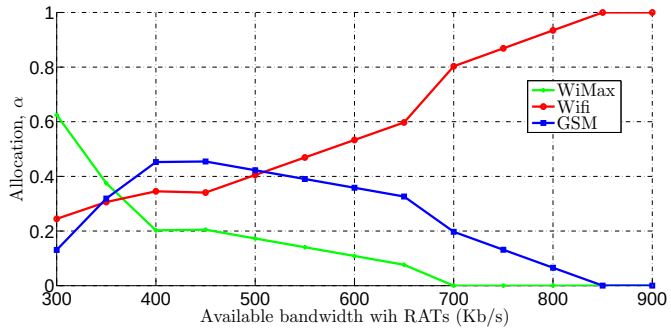

Figure 7. Allocation for User 1 vs Available bandwidths of networks.

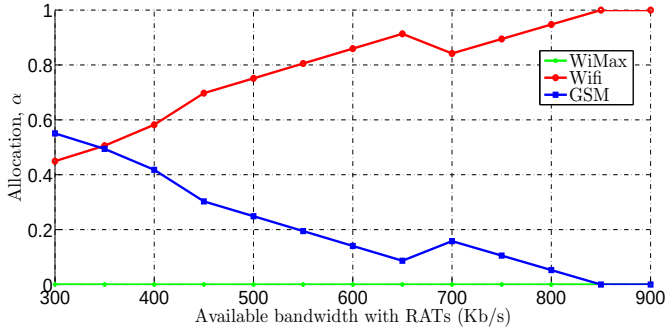

Figure 8. Allocation for User 2 vs Available bandwidths of networks.

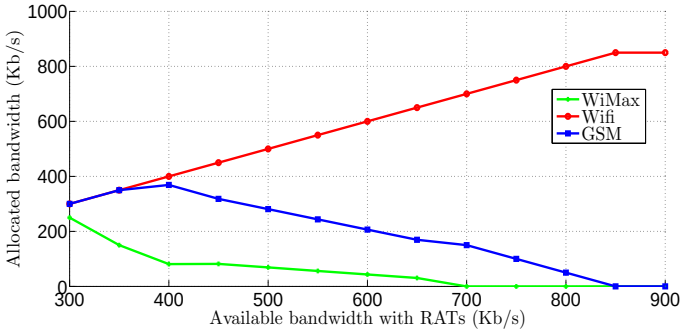

Figure 9. Bandwidth allocated to different CMs vs Available bandwidths of CMs.

Mbps capacity of that WiFi card is currently underutilized. With virtualization in the network card, we can connect to multiple WiFi's, where we are authorized to, with the same card at the same time. This would result in full utilization of our resources thus adhering to greener emblem of GSCC paradigm.

The paper further develops a decision function to efficiently utilize the benefits of sharing the bandwidth requirement of users in the GSCC paradigm. However, this formulation is valid for the scenario when the network parameters determining the allocation are static i.e. there is complete information about them before making the decision. However, parameters like Bit error rate(BER) are probabilistic in nature and to incorporate such scenarios a stochastic multi-criteria decision problem is aimed to be developed in our future endeavors providing robustness to the proposed model.

Furthermore in the initial characterization of the decision function the utility function forms are known but very little is done about their characterization. There is no straightforward approach to determine the parameters for utility functions in a practical scenario and needs to be explored further.

It is envisioned that when the above improvements be incorporated, the decision function will approach the real-world scenario and will give a better practical result and can be applied for standardizing with the GSCC paradigm.

The proposed algorithms may also be extended on real-time applications such as voice telephony, conferencing and circuit switched networks which is currently being explored in our future endeavors.

REFERENCES

[1] A. Tolli, P. Hakalin, and H. Holma. Performance evaluation of common radio resource management (crrm). In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 5, pages 3429–3433 vol.5, 2002.

[2] HD Mustafa, BM Baveja, S Vijayan, SN Merchant, and UB Desai. Replicating the geographical cloud: Provisioning omnipresence, omniscience and omnipotence. *Future Generation Computer Systems*, 2015.

[3] H.D. Mustafa, S. Vijayan, S. Katta, S. Sagar, S.N. Merchant, and U.B. Desai. Green symbiotic heterogeneous networks. In *Consumer Communications and Networking Conference (CCNC), 2014 IEEE 11th*, pages 95–100, Jan 2014.

[4] Christoph Paasch, Gregory Detal, Fabien Duchene, Costin Raiciu, and Olivier Bonaventure. Exploring mobile/wifi handover with multipath tcp. In *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, pages 31–36. ACM, 2012.

[5] Bo Han, Pan Hui, and Aravind Srinivasan. Mobile data offloading in metropolitan area networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 14(4):28–30, 2011.

[6] G. Nigam, P. Minero, and M. Haenggi. Coordinated multipoint joint transmission in heterogeneous networks. *Communications, IEEE Transactions on*, 62(11):4134–4146, Nov 2014.

[7] Kok-Kiong Yap, Te-Yuan Huang, Masayoshi Kobayashi, Yiannis Yiakoumis, Nick McKeown, Sachin Katti, and Guru Parulkar. Making use of all the networks around us: a case study in android. In *Proceedings of the 2012 ACM SIGCOMM workshop on Cellular networks: operations, challenges, and future design*, pages 19–24. ACM, 2012.

[8] L. Wang and G. S. G. S. Kuo. Mathematical modeling for network selection in heterogeneous wireless networks x2014; a tutorial. *IEEE Communications Surveys Tutorials*, 15(1):271–292, First 2013.

[9] P. Kosmides, A. Rouskas, and M. Anagnostou. Network selection in heterogeneous wireless environments. In *Telecommunications (ICT), 2011 18th International Conference on*, pages 250–255, May 2011.

[10] Q. T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, pages 144–151, April 2008.

[11] Y. L. Chen and C. C. Liu. Multiobjective var planning using the goal-attainment method. *IEE Proceedings - Generation, Transmission and Distribution*, 141(3):227–232, May 1994.

[12] ”Karina Gomez, Dejene Boru, Roberto Riggio, Tinku Rasheed, Daniele Miorandi, and Fabrizio Granelli”. ”measurement-based modelling of power consumption at wireless access network gateways ”. *”Computer Networks ”*, ”56”(”10”):”2506 – 2521”, ”2012”. ”Green communication networks”.

[13] Qingyang Song and A. Jamalipour. Network selection in an integrated wireless lan and umts environment using mathematical modeling and computing techniques. *IEEE Wireless Communications*, 12(3):42–48, June 2005.