Value-centric Information and Intelligence Sharing Scheme

Shinichi Doi†, Osamu Kamatani, and Ryutaro Kawamura

Abstract

The amount of information in the network has been increasing, and it is dominated by a few heavy users. Unfortunately, because the available resources are not infinite, we must consider resource starvation when we develop a future network architecture. We present value-centric networking (VCN), which is based on *information value*. The main purpose of VCN is to maximize information value in the network, and thus support the entire society. We report our implementation of an architecture for the VCN concept in the form of a proxy server and confirmation that VCN operation forwarded the high-information-value streams passing through the proxy server faster than ordinary first-in first-out (FIFO) operation. In addition, we discuss the controversial issues of VCN: value definition and feasibility.

1. Introduction

The Internet's rapid development is triggering an information deluge. Annualized average traffic growth is estimated to be about 30–40% [1], [2]. If this trend continues, the traffic volume will reach 600 Tbit/s in 2025. Moreover, the traffic imbalance among users will remain unchanged. Traffic forecasts for the Internet given in Cisco's white paper [3] are plotted in **Fig. 1**. It shows that the top 1% users will generate over 1 Tbit/s per month in 2015. That paper contains descriptions such as "the top 10% generate over 60% of the traffic". Previously, users who consumed enormous network resources tended to download and upload illegal data. In the future, ordinary household usage will trigger the traffic imbalance. These forecasts show that this imbalance trend is likely to continue to 2015 at least.

However, the available resources are not infinite. From the environmental and economic aspects, networking cannot use an infinite amount of electricity. Moreover, network technologies cannot keep advancing forever. For example, the growth of optical fiber technology development, which is the basis of current network technology, might slow down at around 100 Tbit/s per fiber. One of the technical difficulties is that the power level of laser light becomes too high for optical fiber.

Although a few heavy users are draining networking resources, there has been little consideration of how to solve this problem in the Future Network. The most straightforward approach is to decrease resource consumption without decreasing the information value interchanged among network users. Furthermore, the network should maximize the information value conveyed using the restricted resources.

One of the future network visions, Content-Centric Networking [4], could decrease the total network traffic, but its use of content caching will result in other resources being consumed in greater amounts. On the other hand, Network Virtualization [5], [6] will require more networking and computing resources. Therefore, we must assume that such resource starvation will be a life-and-death issue in the Future Network.

In this article, we introduce a novel concept for the future network. Called Value-centric Networking (VCN), its main purpose is to maximize the value of information in networks and the value of the intelligence contained in that information and to maximize the contribution of networks to human society.

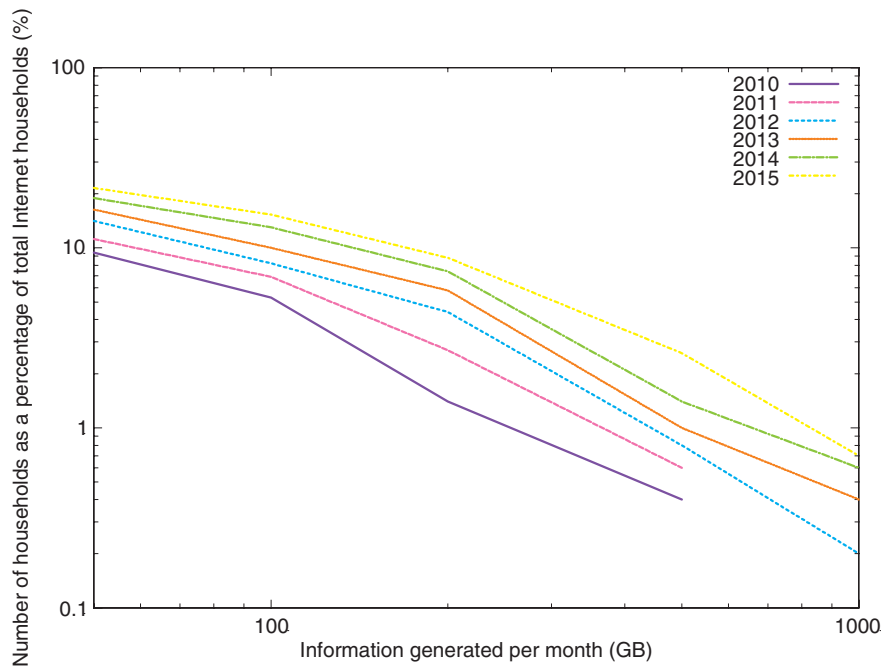† NTT Network Innovation Laboratories
  Yokosuka-shi, 239-0847 Japan

Fig. 1.   Trend of Internet household traffic per month.

2.   Value-centric Networking

In this section, we describe the concept of VCN and present a simple example of it.

2.1   Main objective

Here, we define the main objective of VCN. First, we define the network components as follows. We denote a piece of information by I_i, where i is the identifier of the information in the network system. Its value is denoted by v_i. The simplest objective of VCN is to maximize the sum of v_i in the network system. We assume that v_i is composed of multidimensional attributes because information has various value aspects.

Furthermore, the network must use some resources, e.g., bandwidth and electricity, when transmitting and receiving information. Let r_i denote resources that I_i uses, and let R denote all of the resources available for use. Like v_i, r_i can consist of multidimensional attributes. We assume that there are n pieces of information in the network system and define the VCN objective by the following formula.

$$\text{Maximize} \sum_{i=1}^{n} v_i I_i$$

$$\text{subject to} \sum_{i=1}^{n} r_i I_i \leq R, \; I_i \in {0,1}.$$

This is a well-known NP-hard problem called the knapsack problem. Although no algorithm is known to be the fastest for knapsack problems, Dynamic Programming (DP) can solve them quickly in practice. A greedy algorithm can give a suboptimal solution to this problem in the practical way.

2.2   Architecture

An overview of the architecture for the VCN concept is shown in **Fig. 2**. There are three domains: users, networks, and services. All layers have resource restrictions. In the user layer, a user's device has resource restrictions such as the device's battery power, central processing unit (CPU) performance, or network bandwidth. With these resources, users get information of value V_U. When users get information, networks must consume resources—mainly bandwidth and electricity. Networks convey information of value V_N, which is distributed without exceeding the resource restrictions. Services also consume resources to provide information. In doing so, services send information of value V_S and consume resources, such as bandwidth, electricity, and computing capacity.

VCN is intended to maximize all three values V_U, V_N, and V_S. However, there may be conflicts where different behaviors maximize different values. For
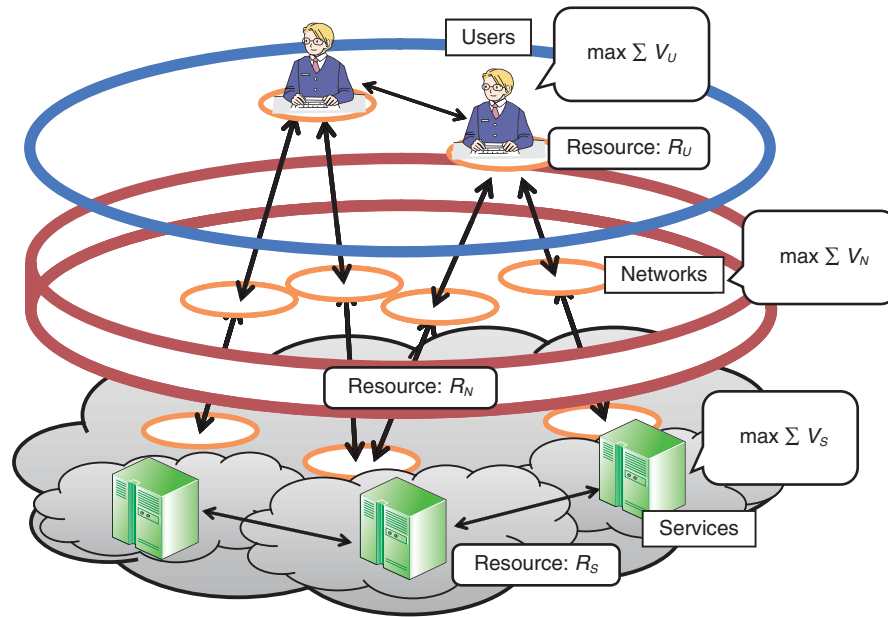
Fig. 2. Overview of VCN architecture. Big ovals represent layers requiring value maximization. Small ovals represent maximizing interfaces.

example, when there are two users, A and B, V_U is represented by A's max value (ΣV_{UA}) plus B's max value (ΣV_{UB}). However, if the network resources are insufficient to satisfy the requests of both users, then the information requested by one of the user's, either A or B, is not forwarded; this behavior maximizes V_N but not V_U.

2.3 Maximization methods

In this subsection, we describe how to get the optimal combination of information transferred. As previously mentioned, we assume that v is one of the information metrics (e.g., the popularity of the information) and is a cardinal number and that r is one of the resources, e.g., bandwidth. One of the simplest approaches is to use a greedy algorithm. The greedy algorithm for knapsack problems selects the information combinations sorted by v_i/r_i. We show one of the simplest examples in **Fig. 3**. This example shows the user getting some information, which is shown in the information table, from a service via the network. The network has restricted resources represented by $R_N = 20$. Examples of this restriction include bandwidth. First, we use the information set $\{I_1, I_2, I_3, I_4, I_5\}$. With this information set, the combination that maximizes the utilization of resources is $\{I_1, I_2, I_3\}$. On the other hand, the combination that maximizes the network value is $\{I_4, I_5\}$. These results are obvi-

ously different.

The greedy algorithm sorts the information set by v_i/r_i and returns its answer $\{I_4, I_5\}$ as a solution. The procedure for obtaining this solution is similar to quality-of-service (QoS) classification: QoS uses priority, and we can use v/r priority here for the greedy algorithm. However, it is not always true that the greedy algorithm can get the optimal solution.

If we add information I_6 to the above information set, the greedy algorithm selects only $\{I_6\}$ because its v_i/r_i is higher than those of others; this is not an optimal solution. As previously mentioned, DP is well known as one method for solving knapsack problems. DP selects $\{I_4, I_5\}$ as the optimal solution. However, its calculation cost is high because it uses a matrix table having n rows and R columns. If the number of information bits is n, then the order of the greedy algorithm is $O(n)$, while the order of DP is $O(n \times \frac{R_N}{g})$, where g is resource granularity. For VCN, this cost might be excessive. R_N, which may be bandwidth, might be a big value, such as 100 Gbit/s or 1 Tbit/s. In addition, the amount of information ranges from several bytes to several gigabytes. This indicates that if we use DP to get the optimal solution, g must be small and the DP table must be extremely large.
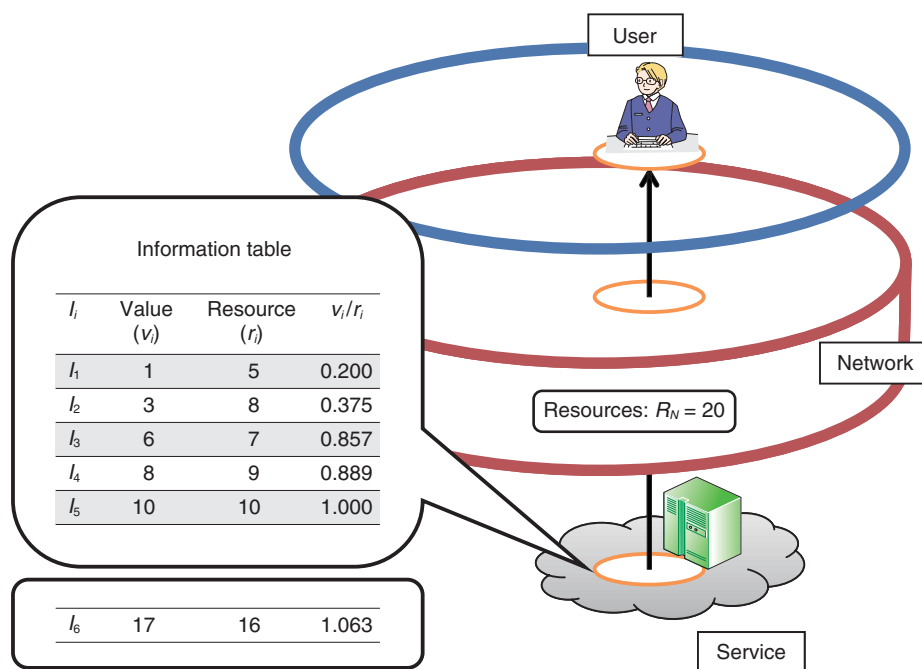
Fig. 3. One of the simplest examples of VCN. A user is receiving several information streams from a service, and network resources are restricted to $R_N = 20$.

3. Evaluation

In this section, we describe our preliminary results for evaluating VCN, which was using our prototype proxy server.

3.1 Overview

In order to evaluate the VCN concept and conduct a practical experiment, we implemented a VCN proxy server (VPS) in a system. With this system, we assumed that all connections are controlled by VPS at the network entrance. We assumed that VPS has information about the network's available resources. VPS controls the hypertext transfer protocol (HTTP) connections by assessing their values and resource consumption. Below, we describe VPS and the experimental results.

3.2 Implementation of VCN proxy server (VPS)

We implemented VPS as a Java program. VPS maximizes the value of information passing through it. We used a greedy algorithm to maximize information value because DP's calculation cost is too high. VPS behaves as follows.

When a user sends an HTTP request to VPS, VPS accepts the connection and creates a proxy connec-

tion instance. The proxy connection instance includes two connections: Client Side and Server Side. The user's HTTP request is read by Client Side. If it contains a host entry, Server Side tries to connect to the server identified in the host entry. If the connection is successful, Server Side writes the user's HTTP request to the server and reads the server's HTTP response. However, that read operation is restricted by the bandwidth controller. If the read size exceeds the restriction set by the bandwidth controller, Server Side suspends the read operation. Under the restriction, Server Side forwards the HTTP response to Client Side, and Client Side sends the HTTP response to the user. Server Side and Client Side have trash queues for storing packets forwarded from the server. A connection manager analyzes packets in the trash queue and controls the bandwidth according to the connection's value (v) and the resource (r) using each connection's bandwidth controller. Value v is defined in a value table, which stores tuples of a regular expression for the universal resource locator (URL) and value v. Each connection is evaluated against the value table by using its content's URL. Resource r is defined by its content length, and the connection manager sorts the connections by v/r. The top P_v% of all connections at a given moment are allocated R_v%
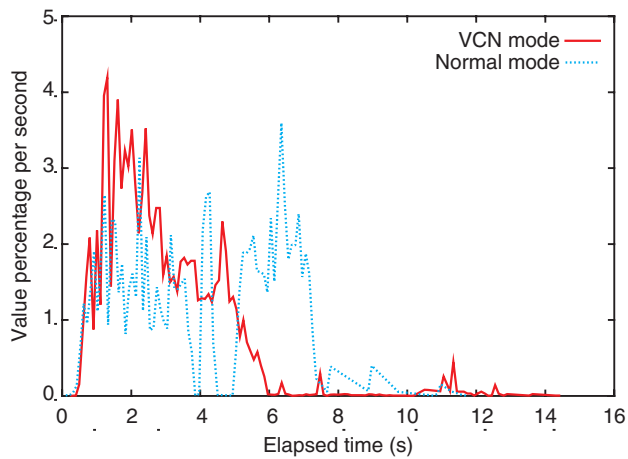
Fig. 4.   Experimental results for value percentage per second.

Fig. 5.   Experimental results for cumulative value.

of the bandwidth. P_v and R_v can be set independently to any value. For example, setting $P_v = 50$ and $R_v = 50$ yields normal proxy server usage.

4.   Evaluation results

We evaluated the effect of VCN as follows. A user browsed a web page containing many pictures via VPS. The storage size was used as the metric for resources and content. Most web browsers use several HTTP connections in rendering a web page that has many pictures. Accordingly, VPS maximized the value of information passing through itself. That is, the user received the more valuable pictures rapidly.

VPS and the Apache server providing web content were instantiated on the same server computer. The web page requested includes 64 pictures (in JPEG format) whose values ranged from low to high at random. Because the browser made one connection for each JPEG picture, this web page can emulate 64 users each getting a JPEG picture. The server computer and the user's terminal (a laptop personal computer) were linked by Ethernet via the same network switch. The total size of all 64 JPEG images was 44 Mbytes, and the bandwidth limitation of VPS was set to 24 Mbit/s, i.e., the performance limit of VPS.

We ran two trials: VCN mode and normal mode. In VCN mode, the VPS settings were $P_v = 30$ and $R_v = 90$, which means that the top of 30% of connections were able to use 90% of the bandwidth. In normal mode, VPS was configured with $P_v = 50$ and $R_v = 50$.

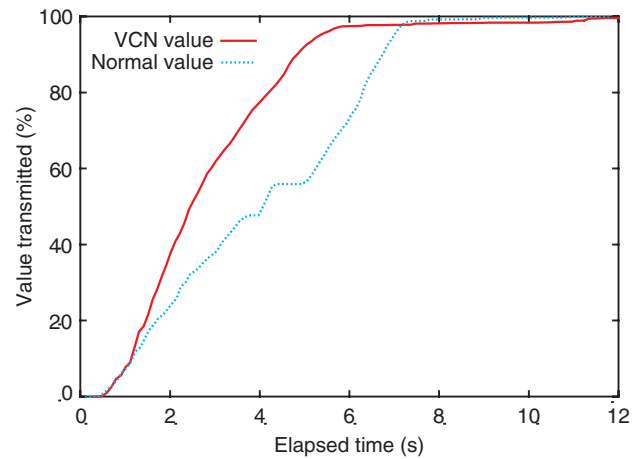The value per byte of information passing through

VPS is shown in **Fig. 4**, where value percentage means the percentage of the value transmitted (i.e., 100% means that all of the value is transmitted). We confirmed that high-value information was forwarded more rapidly in VCN mode than in normal mode. The cumulative value of information passing through VPS is shown in **Fig. 5**. Although both modes completed page loading at the same time, VCN mode forwarded high-value information more rapidly. This means that the user could get more valuable information before less valuable information. These results show that VCN enables users who request valuable information to get it faster than other users.

5.   Discussion

In this section, we discuss two remaining issues with VCN.

5.1   Value definition
In this article, we used the term *information value* many times. However, we avoided the fundamental question of "What is the information metric?" Defining the information metric is a difficult and controversial problem. We think that there is no universal solution that will suit everyone. Our assumption is that the definition will be grounded in two fundamental principles: objectivity and cardinal number. From this standpoint, some examples of information metrics are given below.
Objectivity: The information metric should be based on engineering factors and thus not be subject to the feelings or opinions of individuals. We assume that

there are two aspects: objectivity and subjectivity. Using measures associated with subjectivity prevents comparisons because they are deeply related to individual feelings. For example, subjective metrics include the beauty of movies and photos, the appeal of music, and human interest in information. By contrast, the resolution or bitrate of movies, photos, or music and the popularity of content based on the number of requests for it are objective metrics.

Cardinal number: The information value should be a quantity, i.e., a numerical value. When characterizing information by a number, we need to consider two types of number: cardinal numbers and ordinal numbers. A cardinal number expresses a numerical quantity (count) while an ordinal number indicates a ranking. An example of a cardinal value is information popularity. Popularity can be expressed by the number of people who like it, and this is a cardinal number.

5.2 Feasibility

In achieving the VCN concept, there are two main difficulties. One is the technical problems posed by VCN's need for deep packet inspection and by solutions to the knapsack problem. The calculation costs might be high enough to cause network services to suffer delay. Actually, the network performance of VPS is only 24 Mbit/s, despite its fully asynchronous operations. We believe that higher-grade computers and more sophisticated programs will run VCN at wire speed; nevertheless, electricity consumption will still be problematic.

The remaining difficult issue is "Who decides the information metric and how?" Because numerical information values are essential in VCN, the metric should be decided carefully. For example, if only the network provider determines the metric, the result may not be acceptable to users and service providers. For this issue, we think that it is important to have transparency in the valuation process. For example, users and services might request the network to provide their own values, and networks would run the value-maximizing process on the basis of those values.

6. Conclusion

Today's network allows access to rich information resources, but it is unaware of the value of the information. One of the main objectives of VCN is to make a value-aware network and maximize the value of information conveyed in the network system given the limited resources available. We assume that in VCN there are three domains requiring value maximization: users, networks, and services. Each has its own resource restrictions and information metrics. In order to evaluate VCN, we implemented it in the form of proxy server, VPS. An experiment conducted in a practical environment showed that VPS could maximize information value. This experiment used an existing browser, web server, and typical web page. Thus, it showed the practicality of VCN. Finally, we discussed two remaining issues: value definition and feasibility. We think that they should be discussed in far more detail.

References

[1] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet Inter-domain Traffic," SIGCOMM'10, 3.1, New Delhi, India, August, 2010.

[2] "Cisco Visual Networking Index: Forecast and Methodology," 2010–2015. CISCO White Papers, 2011.

[3] "Entering the Zettabyte Era," In CISCO White Papers, 2011.

[4] V. Jacobson, D. K. Smetters, J. D. Thrnton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking Named Content," CoNEXT'09, Rome, Italy, 2009.

[5] A. Bavier, N. Feamster, M. Huang, L. Peterson, and J. Rexford, "VINI Veritas: Realistic and Controlled Network Experimentation," SIGCOMM06, New York, USA, 2006.
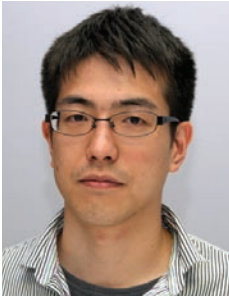
Shinichi Doi

Research Engineer, Media Innovation Laboratory, Network Innovation Laboratories.

He received the M.E. and D.Eng. degrees from Nara Institute of Science and Technology in 2007 and 2010, respectively. He joined NTT Network Innovation Laboratories in 2010 and studied future networking technologies. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Information Processing Society of Japan.

Osamu Kamatani

Senior Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E., M.E., and Ph.D. degrees in electrical engineering from the University of Tokyo in 1988, 1990, and 1993, respectively. He joined NTT Transmission Systems Laboratories in 1993 and worked on high-speed optical time division multiplexed transmission systems, optical networking architectures, and high-speed interactive data transfer applications. He moved to NTT Service Integration Laboratories in 2004 and worked on the NGN architecture and transport control protocols and was also involved in related standardization activities in ITU-T and ETSI TISPAN. He moved to the Media Innovation Laboratory in 2008. His current interest is in terabit-class high-end computing systems based on multicore processors, and a future network architecture that can offer end-to-end terabit scalable data processing.

Ryutaro Kawamura

Director, Vice President, Photonic Transport Network Laboratory, NTT Network Innovation Laboratories.

He received the B.S. and M.S. degrees in precision engineering and the Ph.D. degree in electronics and information engineering from Hokkaido University in 1987, 1989, and 1996, respectively. He joined NTT Transport Systems Laboratories in 1989. From 1998 to 1999, he was a visiting researcher in Columbia University, USA. From 2008 to 2010, he was a visiting research expert in the National Institute of Information and Communications Technology. Since 2003, he has been a member of the Board of Directors of the OSGi Alliance, and since 2005, he has been a Vice President for the Asia/Pacific region. He has been engaged in research on network reliability techniques, network control and management, high-speed computer networks, active networks, network middleware, software component technology, and a future network. He is a member of IEEE and IEICE.