Method of Social Network Service Based Media Contents sharing by Modified URL shortening

Seokhyun Song*, Hyeontaek Oh*, Sangmin Park*, Seung-Hee Kim**, Youngho Jeong** and Junkyun Choi*

* Electrical Engineering Dept., Korea Advanced Institute of Science and Technology, 103-6 Munji-Dong, Yuseong-Gu, Daejeon 305-714, Korea

**Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon, 205-700, Korea

hyungoon@kaist.ac.kr, hyeontaek@kaist.ac.kr, deadpk@kaist.ac.kr, seung@etri.re.kr, yhcheong@etri.re.kr, jkchoi59@kaist.edu

Abstract— **Today, Social Network Services (SNS) has become a part of our life. People can easily publish their own media contents or share ready-made media contents. However, traditional media content sharing is based on the centralized server system. People living in countryside which is far from server farm branch for those services always need to access overseas server to view media contents. To overcome this problem, this paper presents a method for media content sharing on SNS with Modified URL shortening. From HTML5 standard and Web application technologies, we can make web browser work as web server. It has a form of browser extension like Google Chrome Extensions. In addition, friend relationship in SNS means geographical closeness in many cases so it can be used as closer source of media contents than original SNS or media service server. To evaluate this new approach, we simulate the idea with real internet topology and virtual social network structure. From this simulation, we can check the possibility of our suggestion indirectly.**

***Keywords*— Social Network Service, Media Contents Sharing, Web Browser, HTML5, Browser extension, Modified URL shortening**

I. INTRODUCTION

Today Social Network Services (SNS) like Twitter and Facebook have become a part of our life. People can easily produce their own media contents and publish them through their SNS. Moreover, they share lots of ready-made media contents with their SNS friends. On average, in Facebook, more than 250 million photos are uploaded and shared per day [1]. However, traditional media content sharing is based on centralized server system even though the system uses cloud. For those living in countryside which is far from server farm branch for those services always need to access overseas server to view media contents. That is why users who do not live in United States sometimes have inconvenience experience in watching YouTube videos because of buffering or slow loading.

This paper proposes a method for media content sharing with Modified URL shortening to overcome such problems. Since most web browsers cache images or videos, people can use the cache as a new source of contents sharing. From HTML5 standard, Web application and Web browser technologies, we can make web browser working as web server without special native application software. From this change, people can access someone's browser caches and use those media contents. In addition, many SNS friends live nearby themselves; in other words, friend relationship in SNS means geographical closeness in many cases so it can be used as closer source of media contents than original SNS or media service server.

The proposed idea is a simple form of distributed media contents sharing on SNS. Distributed sharing is a general approach in P2P studies. However, constructing the P2P overlay network which makes minimum latency, link/node stress and lookup time is still an open problem. Moreover, free-riding problem is also a big obstacle in real application. Since SNS became popular worldwide, many studies have used social network to solve these problems in P2P are performed [2]-[5]. In the study of Anwar et. al, it shows interest-based routing scheme based on analysis of Orkut, a social networking website which was launched in 2004, being operated by Google, presented better performance than traditional distributed hash table scenario [2]. However, those methods are designed for general file sharing with special native application. Of course, web contents sharing techniques have been studied with P2P concepts. In the study of Wang et. al, it shows non-proxy web caching technique with P2P manner [6]. It provides the services of web caching and searching on the basis of P2P technology but it only consider general P2P with social characteristics. It does not consider SNS.

Our suggestion is a method of sharing media contents on SNS using Web technology especially HTML5. it is still at rough stage, it is a first trial approach of using contents distribution with characteristics of SNS and Web. Because it doesn't use complex Peer-to-Peer (P2P) algorithm, it is free from problems of P2P such as optimization of overlay network, and free-riding. To evaluate this new approach for media contents sharing over social network, we simulate the idea with real internet topology and complex network theory based on virtual social network structure. From this simulation, we can examine the possibility of our suggestion indirectly.

II. SNS BASED MEDIA CONTENTS SHARING SYSTEM

Suggested SNS based media contents sharing system has two parts. First one is a media contents cashing and distribution with Web technologies. The other is a source selection mechanism with Modified URL shortening presentation technique.

A. *Web based media contents sharing*

As already mentioned, traditional media content sharing is based on centralized server system even though the system uses cloud. Not only dedicated contents server but also cloud storage or contents delivery network (CDN) needs pre-constructed server farm. It means that some kinds of traffic concentration always happen near the server farms. P2P is a very good solution for this centralization problem but makes another overhead such as constructing healthy overlay network with complex algorithm.

Web based media contents sharing can be a solution for this dilemma. From developing HTML5 standard and various Web App technology, Web browser can perform many functions that are not available in old One. Various technologies standardized in W3C make this possible, and the key technologies are HTML5 and Web API. HTML5 standard defines some leading techniques such as WebSocket, Web Storage and Web Worker. According to the draft document, First WebSocket specification defines an API that enables Web pages to use the Web Socket protocol for two-way communication with a remote host [7]. It means WebSocket supports new way of communication over Web different from current TCP/UDP communication with native application. Second, Web Storage specification defines an API for persistent data storage of key-value pair data in Web client [8]. This can be used as metadata cache for Web browser based contents file sharing. Third, Web Worker specification defines an API that allows Web application authors to spawn background workers running scripts in parallel to their main page [9]. It means we can do thread-like operation over Web. It may reduce the processing overhead of Web based media contents sharing. Lastly, File API is needed in our system. Its specification provides an API for representing file objects in web applications, as well as programmatically selecting them and accessing their data [10].

With these Web Application technologies, we can design system which share cached media contents. Actually, it is not a Web page but a Web application service over Web browser which called Web browser add-on or plug-in. We call it simply browser extension. It can overcome some limits in HTML5 APIs. Of course it is essentially web service, and it can use all the APIs that the browser provides to web pages, from XMLHttpRequest to JSON.

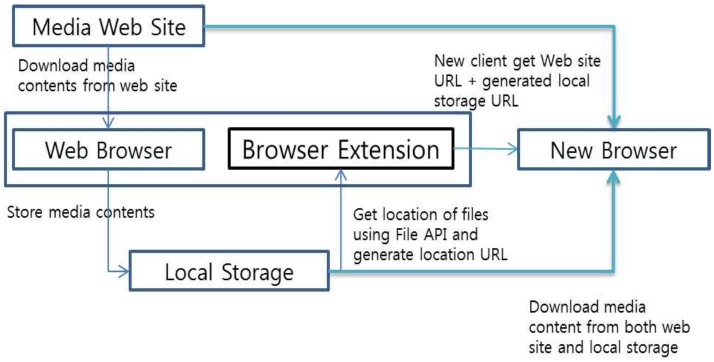

Figure 1. System structure of Web based media contents sharing

This figure shows media contents sharing overview. First, web browser downloads media contents, and save them in temp internet folder or specified folder. That is a basic browser function to show media contents. Browser extension, then, serve this folder and it is indicated with special URL; an array of string.

When the browser which has media contents tries to share the media, then it passes special URL. Actually the browser shares that URL, and the URL is cascaded form of site URL and local storage URL. For this reason, actually new browser can get site URL plus the local storage URL, and it makes new browser can get multiple media file in a two ways.

Therefore, shared data can be transmitted to new browser much faster than original one source transmission.

B. *Source selection with Modified URL shortening*

Traditional URL represents whole information about the web content with readability. However, many SNS limits their message length, so people start using shorten URL to make web content link in SNS. URL -shortening with HTTP-redirect is general technique now. We present Modified URL shortening technique which contains two URLs together. It is a simple application of traditional URL -shortening technique. URL is a standardized technology for representing contents in Internet [11]. Modifying URL can make problem in following URL standard correctly. To overcome this weakness, we just use HTTP redirecting server. If the user has proposed browser extension, it detects the URL which was originated from its peer application. Then, it requests two different original URL, one for original source and the other for peer source, then it uses the URLs for source selection. If the user does not propose browser extension, it just accesses redirecting server and finds original contents URL which is exactly the same with URL shortening.

Modified URL shortening consists of two parts, hashed original contents address and hashed user contents address. It has concatenation form of two shortened URLs. For example, "http://kaist.kr/gf24s/a2g55" is an example of Modified URL shortening. Following figure shows the operation of Modified URL shortening with source selecting.

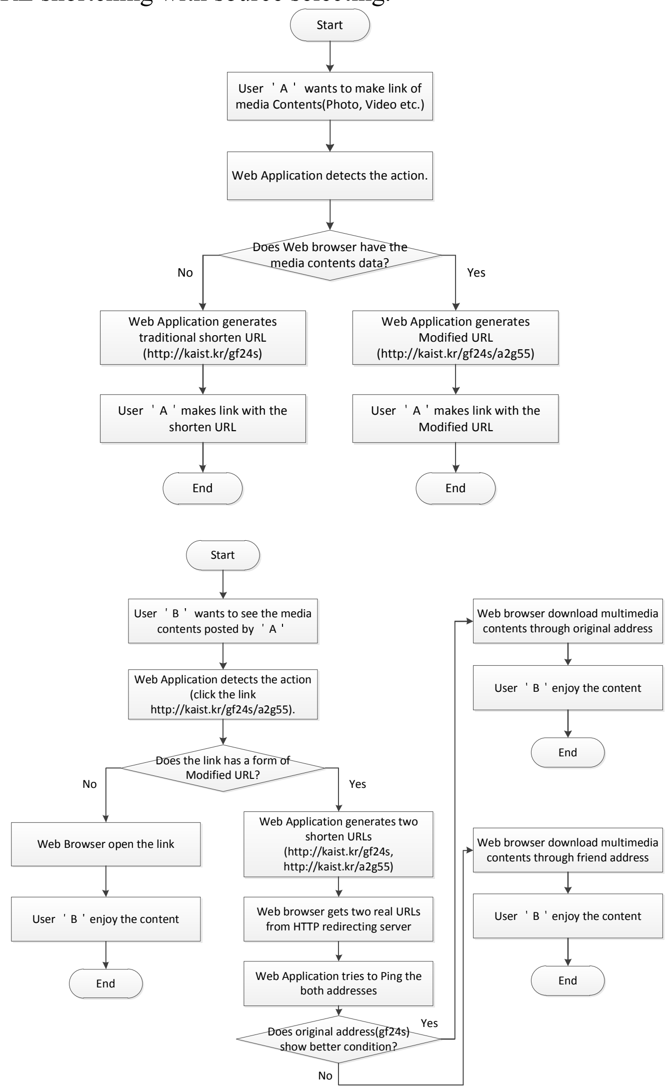

Figure 2. Operation of Source election with Modified URL shortening

HTTP redirecting server does not consider the last part of URL "a2g55" because "http://kaist.kr/gf24s" is enough for URL shortening. Therefore, it redirects the user to some original contents URL whatever address is attached below the last slash. If the user who has special browser extension for Modified URL shortening clicks the link, the browser extension detects the URL and makes two different shortened URL, "http://kaist.kr/gf24s" and "http://kaist.kr/a2g55". Then, it tries to ping both addresses. From the result of ping, browser selects the source to download media contents such as photos, videos and flash games.

III. SIMULATION & EVALUATION

1960s, Stanley Milgram shows 'six degrees of separation' in social network with passing letter experiment [12]. This phenomenon came from characteristics of social network. Social network is a complex network which has scale-free characteristics. A scale-free network is a network whose degree distribution follows a power law. It is sufficiently clustered and has short path length between nodes. Peer relationship in SNS reflects these social network features [13]. Therefore, it can be represented as scale-free network.

SNS in real world has millions and billions of users and there is various attitudes about privacy. From this reason, it is hard to experiment proposed idea as real application. However, simulation is a good alternative for testing possibility of idea. To evaluate this new approach for media contents sharing over social network, we simulate the idea with real internet topology and complex network theory based virtual social network structure. From this simulation, we can check the possibility of our suggestion indirectly.

A. Simulation environment

There are daily internet AS(Autonomous System)-level topology maps using following data sources [14][15].

- BGP routing tables + updates: Route Views, RIPE-RIS, Abilene, CERNET BGP View
- Route Servers: Packet Clearing House, UCR, traceroute.org, Route Server Wiki
- Looking Glasses: traceroute.org, NANOG, Looking Glass Wiki

From this data, we can get a real internet topology map and use it as basic internet topology structure in our simulation. The topology map has over 40,000 nodes and over 144,000 links.

For the SNS, we make virtual social network topology map through complex network theory. There are various models to build scale-free network. The most popular model is ER(Erdos and Renyi) Random graph model [16]. However, it is not suitable for describing social network because it has phase transition with seed probability. The characteristic makes unstable fragmentation or single giant component in the map. Therefore we use BA(Barabási–Albert) preferential attachment model in our simulation. The study of Barabási and Albert suggests an algorithm for generating random scale-free networks using a preferential attachment mechanism [17][18]. Its characteristic is quite suitable for describing social network. SNS user relationship is an example of social network so we use BA model for generating virtual SNS friend map.

Based on internet topology map, we build general user map with SNS friendship relationship. Since Internet topology map is generated in AS-level, we assume that an average 20 general users are attached in each AS. Frankly, an AS can handle thousands of users. However, it means that we makes virtual map which have over billion users. It needs lots of computing power. In our test, Python simulator needs over 7 GB memories for one time simulation although we use only 20 users for each AS. Virtual social network map has over 4 million nodes and over 80 million links.

B. Simulation result and Evaluation

Although it deals with large data set, the simulation is quite simple. In our simulation, we aim to check how the geographical closeness affects the efficiency of proposed idea. Firstly, we build virtual SNS topology map and randomly map SNS topology on Internet topology. Then, we operate SIR(Susceptible, Infectious and Recovered) process with randomly selected seed node. SIR model is simple and popular epidemic model with only here compartments and fixed population [19]. The spread of media contents within people is very similar to the spread of the disease. Therefore, we borrow the idea of epidemics to simulate media contents usage in SNS. From SIR process, we can get a range of media spreading. It represents how many users show the content which is shared by seed user. As the final step, we calculate an average hop count between seed and infected users. If the average hop count is not quite bigger than conventional way, we can accept the idea as a possible solution.

In this simulation process, we adopt geographical closeness as parameter "p". If the parameter is 0.1, it means 10% of SNS friends of someone who are located in same AS. Two users in the same AS have maximum geographical closeness. For example, two school mates in the same dormitory belong to one AS. Since the data transmission between them is not affected by outer internet topology or traffic condition, it assures the best transmission performance. This is reflected as a zero-hop between them in our simulation. In addition, we fix the probability of becoming infected as 36.68%. The value is obtained by survey on SNS users. We aggregate the response about the rate of open multimedia link on SNS stream and use the average as a constant parameter from about 100 users. Another fixed parameter is the link parameter in BA model. In BA model, it needs the number of links for preferential attachment process. We fix the parameter as 13 based on statistical report about Facebook and Twitter. From the statistics page in Facebook, it says that an average Facebook user has 130 friends [1]. However, twitter shows a little different aspect. From the Barracuda Lab's annual report on 2009, 20% of Twitter users are not following anyone and 51% follow less than 5 people and the number of Twitter users following more than 10 people is about 40%. The report defines a true Twitter user as someone who has three main attributes: it has at least 10 followers, follows at least 10 people, and has tweeted at least 10 times [20]. Based on these two reports, we set 13 as average number of attached links in our virtual social network.

Following table shows the simulation result. We take several testing and calculate the average of them for each "p". "p" represents geographical closeness parameter, "H" represents the average distance, a hop count, between seed and infected nodes, "σ" represents a standard deviation of distance between seed and infected nodes, "N_I" represents the number of infected nodes and "T" represents the average simulation time (seconds) for a trial.

TABLE 1. SIMULATION RESULT

p	The average of simulation result			
	H	σ	N_I	T
0.0	3.866	0.169	294219	2876
0.1	3.599	0.843	294554	3006
0.2	3.316	0.421	294339	3244
0.3	3.280	0.517	294408	3143
0.4	2.958	0.358	294131	3280
0.5	2.555	0.799	294246	3359
0.6	2.385	1.023	294446	3512
0.7	2.191	1.579	294098	3535
0.8	2.099	1.497	294143	3590
0.9	1.470	1.716	294449	3503
1.0	0	0	294320	2850

H: Avg. hop count, σ: std. dev.
N_I: #of Infected nodes, T: Avg. simulation time

The average hop between every pair node in given internet topology map is about 3.74. The maximum hop count is 13. It means the traditional centralized media contents sharing needs at least 3 hop even though there is proper distribution of server farm. In our simulation, the average hop distance is 3.866. It shows that longer hop distance can be needed in spreading patter that is different from every pair case. Naturally, close geographical closeness can make smaller hop count in proposed method. The table shows the efficiency is proportion to geographical closeness. It's not affected by the number of infected node. Actually, the simulation time and the number of infected time are affected by probability of becoming infected which is fixed in our simulation. Moreover, the result shows that proposed manner makes smaller hop count if "p" is over 0.4; the average hop count is 2.958.

In our original method, the host of media contents sharing is continuously changed by the user. It means the browser extension replaces the friend address with its own address when the contents are re-shared. Then, the distance between friend host and new content consumer can be more directly affected by geo-graphical closeness between SNS friendship. However, we simply calculate the distance between seed node and infected node. We assume that replacing friend address process is skipped and Modified URL shortening is just spread through SNS. This makes some problem in terms of appropriateness of simulation but it is enough to check the possibility of the suggested idea. We can find out that the geographical closeness within SNS friend can lead to a meaningful result in this simulation. Since we cannot know the actual dynamics in real SNS, the best way of evaluating the idea is practical experiment with real SNS environment.

IV. CONCLUSIONS & FURTHER WORKS

This paper proposes a method of distributed media contents sharing on SNS. Although it is rough, it is a first trial approach of using contents distribution with characteristics of SNS. Because it doesn't use complex Peer-to-Peer (P2P) algorithm, it is free from problems of P2P such as optimization of overlay network, and free-riding. Actually, this is a work-in-progress paper. To evaluate this new approach for media contents sharing over social network, we simulate the idea with real internet topology and complex network theory based on virtual social network structure. From this simulation, we can check the possibility of our suggestion indirectly. In the further studies, we expect that real web browser based contents sharing App. can be developed after finalizing HTML5 standard in W3C.

ACKNOWLEDGMENT

This work was supported by the ETRI R&D Program of KCC(Korea Communications Commission), Korea [11921-03001, "Development of Beyond Smart TV Technology"].

REFERENCES

[1] (2011) Facebook official website. [Online]. Available: http://www.facebook.com/press/info.php?statistics

[2] Z. Anwar, W. Yurcik, V. Pandey, A. Shankar, I. Gupta, and R. H. Campbell, "Leveraging Social-Network Infrastructure to Improve Peer-to-Peer Overlay Performance: Results from Orkut," ACM Computing Research Repository, p. 9, 2005.

[3] B. C. Le Roy, A. Sanabria, and S. Calvert, "PEER TO PEER SOCIAL NETWORKING," US Patent App. 12/139,342. Google Patents, 13-Jun-2008.

[4] J. Altmann, "A P2P file sharing network topology formation algorithm based on social network information," INFOCOM Workshops 2009, IEEE, pp. 1-6, Apr. 2009.

[5] K. C.-J. Lin and L. Golubchik, "SocioNet: A Social-Based Multimedia Access System for Unstructured P2P Networks," IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 7, pp. 1027-1041, Jul. 2010.

[6] X. Wang, Ng, B. Ooi, K. Tan, and A. Zhou, "Buddy Web: A P2P-based Collaborative Web Caching System," NETWORKING 2002 Workshop on Web Engineering and Peer-to-Peer Computing, 2002.

[7] The WebSocket API , I. Hickson, Editor, W3C Working Draft (work in progress), 29 September 2011, http://www.w3.org/TR/2011/WD-websockets-20110929/. Latest version available at http://www.w3.org/TR/websockets/.

[8] Web Storage , I. Hickson, Editor, W3C Working Draft (work in progress), 1 September 2011, http://www.w3.org/TR/2011/WD-webstorage-20110901/. Latest version available at http://www.w3.org/TR/webstorage/.

[9] Web Workers , I. Hickson, Editor, W3C Working Draft (work in progress), 1 September 2011, http://www.w3.org/TR/2011/WD-workers-20110901/. Latest version available at http://www.w3.org/TR/workers/.

[10] File API , A. Ranganathan, J. Sicking, Editors, W3C Working Draft (work in progress), 26 October 2010, http://www.w3.org/TR/2010/WD-FileAPI-20101026/. Latest version available at http://www.w3.org/TR/FileAPI/.

[11] Uniform Resource Locators (URL) Specification, IETF Std. RFC1738, 1994.

[12] M. Newman, "Models of the small world," Journal of Statistical Physics, vol. 101, pp. 819-841, 2000.

[13] M. E. J. Newman, D. J. Watts, and S. H. Strogatz, "Random graph models of social networks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99 Suppl 1, no. March, pp. 2566-72, Feb. 2002.

[14] (2011) Internet Topology Collection – Internet Research Lab in UCLA website. [Online]. Available: http://irl.cs.ucla.edu/topology/

[15] Beichuan Zhang, Raymond Liu, Daniel Massey, Lixia Zhang, "Collecting the Internet AS-level Topology," ACM SIGCOMM Computer Communication Review (CCR), special issue on Internet Vital Statistics, January, 2005

[16] Erdős, Paul, A. Rényi, "On the evolution of random graphs". Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5: 17-61, 1960

[17] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, p. 509, 1999.

[18] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[19] W. O. Kermack and a G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 115, no. 772. pp. 700-721, 01-Aug-1927.

[20] Paul Judge, "Barracuda Labs Annual Report 2009," Barracuda Networks Inc., Technical Report, 20