Reducing the Download Time in Stochastic P2P Content Delivery Networks by Improving Peer Selection  
by  
Nicholas J. Hays  
Dissertation Advisor Dr. Gregory Simco

Degree of Doctor of Philosophy in Computer Science

2017  
College of Computing and Engineering   
Nova Southeastern University

What research problem was the reviewed dissertation trying to address?

The research problem that the reviewed dissertation is trying to address is that Peer-to-Peer (P2P) networks take up enormous portions of the open internet's bandwidth (Zuo & Iamnitchi, 2016). The P2P networks consuming so much bandwidth on the open internet is a problem because of the congestion put on Internet Service Provider’s (ISP’s) networks (He, et al., 2016). Because open networks get saturated with high bandwidth and inefficient P2P traffic, other applications can see reduced performance (Ijaz, Saleem, & Welzl, 2013). The congestion in turn leads to poor user experience through low throughput and intermittent inefficiency (Brienza, et al., 2016). There have been attempts to solve the bandwidth issues through biased and random based switching techniques (Hsiao, et al., 2011; Pacifici, et al., 2016). The problem with biased based switching is the requirement of client overhead because connection data of peers is needed to poll, store, and analyze (Wilkins & Simco, 2013). Biased based switching cannot switch peers if the selected nodes performance degrades (Wilkins & Simco, 2013). The problem with random switching is it only performs switching after the peers have been selected and allows for a potential problem of downloading from a faulty peer before a better one can be chosen (Chiu & Eun, 2008). The research problem of the reviewed dissertation proposes to improve P2P networks by addressing how to improve performance of peers before a connection is made and while a file transfer is in progress.

What was the research goal of the reviewed dissertation?

The research goal of the reviewed dissertation was to reduce the average download time for individual clients and application loads on networks. The hypothesis was that a random peer selection strategy could be improved by replacing the worst performing peers with advanced knowledge given to the clients. The goal was proposed by the reviewed dissertation using advanced knowledge for peer selection, monitoring the performance of the server peer after the connection has been made, and only replacing the worst performing peers. Previous research from (Lehrfeld & Simco, 2010) as well as (Wilkins & Simco, 2013) showed that limiting the amount of time spent with poor performing peers reduces the average download time for the client and that the reviewed dissertation’s goal was achieved based on improved metrics compared to the metrics from the previous research simulations.

What evidence (literature, data, etc) was presented to support the reviewed dissertation research and achieve the research goal?

A myriad of research literature was presented as evidence for the dissertation research and the research goal. The literature attempts to cover some of the accomplishments made to the problem of reducing download time in stochastic P2P content delivery networks through five peer selection strategies. The first being from Chiu & Eun, 2008 proposing that the heterogeneity of nodes needs to be considered in P2P download strategies. The second being the literature on random peer selection that has the advantage of having low overhead, being robust, and easy to implement (Sherman, Neih, & Stein, 2009). Third is literature about advance knowledge peer selection. The advanced peer selection has high overhead and cost but better download times when compared to random selection (Xie, et al., 2008). Fourth is ISP based selection, the problem being that cross ISP connections can be expensive and causes bottlenecks (Varvello & Steiner, 2011). The dissertation proposes that the research goal will need to reduce cross ISP traffic without the cooperation from the ISP’s since that information can not be obtained because of privacy and security concerns. (Ijaz, et al., 2013; Varvello & Steiner, 2011). Last is proximity strategies that use the location of peers to make informed decisions using concepts such as Round-Trip Time (RTT), jitter, and number of hops (Fiorese, et al., 2013). The five strategies presented help support the reviewed dissertation and achieve the research goal of reducing download time with less network traffic.

What research methodology used and who claim this methodology is valid?

The research methodology used was a simulation method. A P2P simulation based on the prior works of (Chui & Eun, 2008; Lehrfeld & Simco, 2010; Wilkins & Simco, 2013) was set up. Server peers with IP addresses and service capacity metrics were populated from the (OOKLA, 2017) and (MaxMind, 2017) databases. To validate the simulation environment a baseline was created by replicating findings from the previous research of Lehrfeld & Simco, 2010 and Wilkins & Simco, 2013 using the same data populated from the (OOKLA, 2017) and (MaxMind, 2017) databases. Once the previous research's metrics were replicated in the simulation environment, the new algorithm experiments could be performed. The test environment comprised of the C++ programming language written in a Microsoft Visual Studio Integrated Development Environment (IDE). The experiments were run on an AsusG751J Graphics Processing Unit (GPU) with an Intel i7-470HQ processor. The system comprised of 32GB of RAM on the Windows 10 Operating’s System and the random number generator for each experiment had the same seed for each experiment. The simulation P2P environment used Auto Regressive One (AR-1) to simulate the stochastic nature of the server peer’s service capacity, the same way as in the previous works of (Chiu & Eun, 2008; Lehrfeld & Simco, 2010; Wilkins & Simco, 2013). The reviewed dissertation does admit to not accurately reflect real world P2P networks because of three reasons. The first being the amount of peers remained constant in each experiment where in real world P2P networks the amount of peers are constantly increasing and decreasing (Zuo & Iamnitchi,2016). The second being the file sizes accounted for were medium to large sizes where a good portion of P2P traffic are small size files such as bursts of audio (Gummadi, et al, 2003). Lastly the size of the simulation network maxed out at 700 peers where real world P2P networks have shown to be several hundred thousand peers (Zuo & Iamnitchi, 2016). Since the simulation environment from the research method was able to replicate data from the previous research of (Chiu & Eun, 2008; Lehrfeld & Simco, 2010; Wilkins & Simco, 2013) it claims its validity as a reputable simulation environment when comparing to those previous works.

**What were the findings of the reviewed dissertation research?**

The results of the reviewed dissertation research claim to have developed a historic based peer selection strategy for P2P content delivery networks that reduced download time by making informed decisions of selected peers thus reducing network throughput. The research dissertations proposed findings claimed to be achieved through modifying the random based strategy to favor advanced knowledge of peer selection, monitoring the connections, logging the events, and replacing weaker peers with a preemptive choke algorithm based off of Lehrfeld & Simco, 2010. By constantly eliminating weaker peers the findings proposed to have been able to reduce the clients average download duration in the simulated P2P environment. The proposed historic based peer selection strategy improved download time, network latency, and reduced the amount of cross ISP traffic when compared to both the Wilkins & Simco, 2013 smart peer replacement strategy and the Lehrfeld & Simco, 2010 smart peer replacement strategy with choke. Another finding was that incorporating the Lehrfeld & Simco, 2010 choke algorithm to the proposed historic based strategy did not significantly change the results of the proposed historic based strategy. The findings show that the proposed historic based strategy with and without choke perform better in the simulated environment then the previous Wilkins & Simco, 2013 smart peer replacement strategy, the Lehrfeld & Simco, 2010 smart peer replacement strategy with choke, and the random periodic switching strategy presented by Chiu & Eun, 2008.

## **References**

Brienza, S., Cebeci, E., Masoumzadeh, S., Hlavacs, H., Özkasap, Ö., & Anastasi, G. (2016). A survey on energy efficiency in peer-to-peer systems: file distribution, content streaming, and epidemics. *Proceedings of the* *ACM Computing Surveys,* pp. 36-48.

Chiu, M., & Eun, Y. (2008). Minimizing File Download Time in Stochastic Peer-to-Peer Networks. *Proceedings of the* *IEEE/ACM Transactions on Networking*, p. 16.

Fiorese, A., Simoes, P., & Boavida, F. (2013). Approach for service search and peer selection in peer-to-peer service overlays. *Proceedings of the* *2013 International Conference on Information Networking*, pp. 303-308.

Gummadi, P., Dunn, J., Saroiu, S., Gribble, D., Levy, M., & Zahorjan, J. (2003). Measurement, modeling, and analysis of a peer-to-peer file-sharing workload. *Proceedings of Operating Systems Review*, pp. 314-329. doi:10.1145/1165389.945475

He, Q., Dong, Q., Zhao, B., Wang, Y., & Qiang, B. (2016). peer-to-peer traffic optimization based on congestion distance and DHT. *Journal of Internet Services and Information Security*, pp. 53-69.

Hsiao, H., Hsu, H., & Miao, B. (2011). Adaptive and efficient peer selection in peer-to-peer Streaming Networks. *2011 IEEE 17th* *International Conference on Parallel and Distributed Systems*, pp. 753-758.

Ijaz, H., Saleem, S., & Welzl, M. (2013). Fewest common hops: an improved peer selection approach for peer-to-peer applications. *2013* *21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*, pp. 449-453.

Lehrfeld, M., & Simco, G. (2010). Choke-based switching algorithm in stochastic peer-to-peer networks to reduce file download duration. *Proceedings of the IEEE Southeast Conference* *2010*, pp.127-130.

MaxMind. (2017). GeoLite ISP Database. Retrieved from https://www.maxmind.com/en/geoip2-isp-database#features

OOKLA. (2017). Speedtest Intelligence from Ookla. *Internet performance database,* Retrieved from http://www.ookla.com/speedtest-intelligence

Sherman, A., Nieh, J., & Sten, C. (2009). FairTorrent bringing fairness to peer-to-peer systems. *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pp. 133-144. doi: 10.1145/1658939.1658955

Varvello, M., & Steiner, M. (2011). Traffic localization for DHT-based BitTorrentnetworks. *International Conference on Research in Networking*, pp. 40-53.

Wilkins, R., & Simco, G. (2013). Download Time Reduction Using Recent Performance- Biased Peer Replacement In Stochastic P2P Content Delivery Networks. *2013 International Conference on Selected Topics in Mobile and Wireless Networking*, pp. 86-91.

Xie, H., Yang, R., Krishnamurthy, A., Liu, G., & Silberschatz, A. (2008). P4P:Provider Portal for Applications. Proceedings of *ACM SIGCOMM 2008 conference on Data Communication*, pp. 351-362.

Zuo, X., & Iamnitchi, A. (2016). A Survey of Socially Aware Peer-to-Peer Systems. *ACM Computing Surveys*, p. 49.