two

by Rohit Agrawal

Submission date: 18-Mar-2024 04:38PM (UTC+0530)
Submission ID: 2323598021
File name: Probability11.docx (801.12K)
Word count: 2211
Character count: 14291

Probability-Driven Performance: Enhancing Content Delivery Networks

By

Kalpesh Kumavat: 22BCP024

Harsh Hirani: 22BCP036

Darshan patel:22BCP006

Preet rupapara:22BCP019

Deep shah:22BCP072

Pal shah:22BCP119

Abstract:

Within the domain of digital content distribution, Content distribution Networks (CDNs) are essential for guaranteeing the effective and dependable delivery of multimedia files to end users around the globe. However, standard CDN designs face substantial hurdles from the growing demand for online content, requiring creative ways to improve performance and reduce latency. In this study, we suggest a novel method for CDN optimization that makes use of the Poisson distribution model and probabilistic analysis to inform proactive caching techniques. We employ statistical techniques to precisely forecast future requests and find trends in user behaviour by analysing data collected from a DNS server that logs timestamps of user requests for assets.

By selectively caching assets based on these probabilistic predictions, our technology lowers latency and boosts the effectiveness of content delivery overall. We show the effectiveness of our probability-driven method and its potential to transform CDN performance optimization through thorough experimentation and analysis. Our findings highlight the significance of modifying CDN systems to satisfy the changing demands of the digital ecosystem in addition to furthering the field of content delivery.

Introduction:

It is now crucial to optimize Content distribution Networks' (CDNs') performance in the quickly changing digital content distribution ecosystem. CDNs are the backbone of the exponentially growing online content consumption, guaranteeing the dependable and efficient distribution of multimedia materials to end users around the globe. However, conventional CDN systems frequently find it difficult to sustain ideal performance levels in the face of constantly rising demand and complexity.

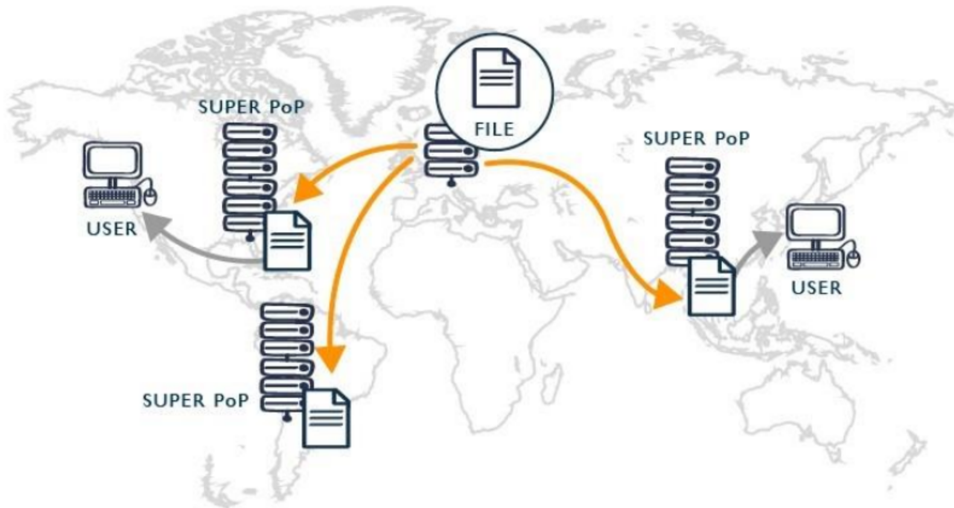
This study explores a unique method for improving CDN performance by using probability-driven tactics. We suggest an anticipatory caching system that makes use of statistical analysis and the Poisson distribution model to anticipate and avert user requests for content assets. Our main goal is to reduce latency and boost overall content delivery efficiency by strategically caching assets at critical network nodes by utilizing the probabilistic nature of user activity patterns.

Utilizing information from one of our DNS servers—which logs the timestamps of user requests for different assets—is the fundamental component of our approach. We want to identify underlying patterns and trends in user behavior through thorough statistical analysis, which will help us predict future requests with a high degree of accuracy. We want to reduce the delay seen in conventional on-demand content delivery methods by proactively storing commonly requested materials based on probabilistic predictions.

This research has implications that go beyond performance optimization; it emphasizes the need to modify CDN infrastructures in order to satisfy the changing needs of the digital world. Through adoption of a probabilistic paradigm and matching caching tactics to user behavior dynamics, we see a world in which CDNs function with never-before-seen responsiveness and efficiency, meeting the varied demands of a worldwide user base.

We explore the technique of our probability-driven approach in more detail in the next sections of this paper, providing empirical proof of its effectiveness through thorough analysis and testing. We also address possible directions for future study and development in the field of content delivery optimization, as well as the consequences of our results. By doing this project, we hope to further the existing conversation about improving CDN performance and open the door to a more robust and flexible digital infrastructure.

Real world problem:



One of the most pressing issues facing content delivery networks (CDNs) nowadays is how to effectively distribute multimedia material that is in high demand to consumers while reducing latency and optimizing bandwidth use. In particular, unanticipated surges in user traffic are a common issue for CDNs, which can lead to higher server stress and possible performance deterioration. The dynamic and geographically scattered nature of internet material consumption exacerbates this problem.

As an example, let us contemplate a well-known streaming service that encounters an abrupt spike in viewer demand during prime watching hours or during a much-awaited event, like a live sports game or the premiere of a new episode. Conventional CDN designs could find it difficult to handle the surge of requests in these situations, which would cause buffering problems, sluggish loading times, and ultimately a bad user experience.

The geographical dispersion of consumers further compounds the intricacy of the issue. A CDN system must be able to dynamically adapt to these various needs while effectively routing material to reduce latency, as users in different locations may have different access behaviours and preferences.

Innovative solutions that make use of sophisticated caching techniques, predictive analytics, and clever routing algorithms are needed to solve this real-world issue. Using probabilistic models to predict user behaviour and store content assets strategically, content delivery networks (CDNs) may effectively minimize the effects of sudden spikes in traffic and maximize resource use. Furthermore, the responsiveness and scalability of CDN infrastructures may be further improved by utilizing edge computing technologies and dynamic load balancing methods, guaranteeing smooth content delivery across a variety of user demographics and geographic locations.

Study Area

The interaction of network optimization, probability theory, and content delivery systems in computer science is the main topic of this research study. In particular, it belongs to the more general categories of:

1. DNSs, or content delivery networks:
    - content delivery networks are looking at ways to improve the effectiveness and performance of CDNs, which are essential parts of the internet infrastructure that are in charge of providing web content to users all over the world.
    - tackling issues with CDN architecture scalability, bandwidth optimization, and latency reduction.
2. Optimization Driven by Probability:
    - investigating the use of statistical models, such as the Poisson distribution, and probability theory in CDN caching strategy optimization.
    - looking at the ways proactive caching choices might be informed by probabilistic user behaviour predictions to increase the effectiveness of content delivery

Solutions:

Certainly! Achieving the solution of enhancing Content Delivery Networks (CDNs) through probability-driven caching involves several key methods. Here's a structured approach:

1. Data Collection and Analysis:

- Collect data from the DNS server that records timestamps of user requests for assets.
- Analyse the collected data to identify patterns and trends in user behaviour. This may involve statistical techniques such as frequency analysis, time-series analysis, and pattern recognition.
- Determine the frequency of asset requests, the popularity of assets, and any temporal patterns in user access.

Sample data :

IP	URL	Time
69.57.172.25	/wp-admin/admin-ajax.php?action=async_litespeed&nonce=ed6552142e&litespeed_type=imgoptm	3/18/24, 5:38 AM
69.57.172.25	/wp-cron.php?doing_wp_cron=1710739717.7233130931854248046875	3/18/24, 5:28 AM
49.36.212.15	/wp-content/uploads/2023/08/Blood-over-bright-Haven-3.png	3/18/24, 5:24 AM
66.249.79.36	/book-review-summary-if-we-were-villains/	3/18/24, 5:28 AM
69.57.172.25	/wp-admin/admin-ajax.php?action=async_litespeed&nonce=ed6552142e&litespeed_type=imgoptm	3/18/24, 5:28 AM
66.249.79.36	/book-review-summary-if-we-were-villains//1000	3/18/24, 5:28 AM
66.249.79.35	/wp-content/litespeed/css/226cacb7689d1848e47e7b9ada68aad0.css?ver=0548f	3/18/24, 5:28 AM
66.249.79.35	/book-review-summary-if-we-were-villains/1000	3/18/24, 5:28 AM
40.77.167.8	/book-review-summary-caraval/?s=	3/18/24, 5:18 AM
40.77.167.71	/sitemap.xml	3/18/24, 5:06 AM
207.46.13.230	/robots.txt	3/18/24, 5:06 AM
69.57.172.25	/wp-admin/admin-ajax.php?action=async_litespeed&nonce=ed6552142e&litespeed_type=imgoptm	3/18/24, 5:07 AM

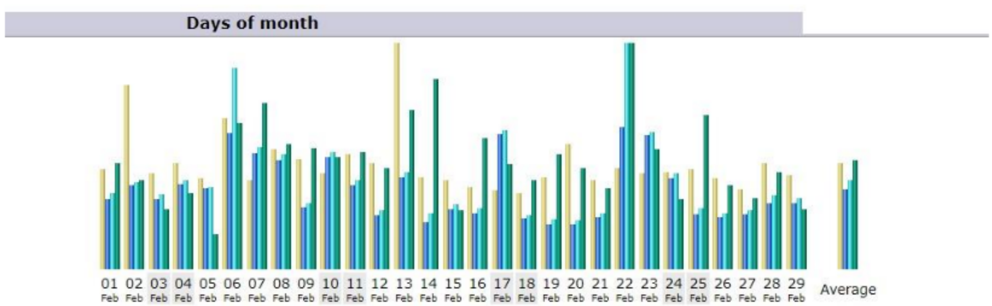
Analytics CHARTS ON DATA:

Summary

Reported period	Month Feb 2024				
First visit	01 Feb 2024 - 00:16				
Last visit	29 Feb 2024 - 23:46				
	Unique visitors	Number of visits	Pages	Hits	Bandwidth
Viewed traffic *	**1,864**	**2,901** (1.55 visits/visitor)	**13,526** (4.66 Pages/Visit)	**15,140** (5.21 Hits/Visit)	**391.95 MB** (138.35 KB/Visit)
Not viewed traffic *			**16,176**	**19,019**	**563.21 MB**

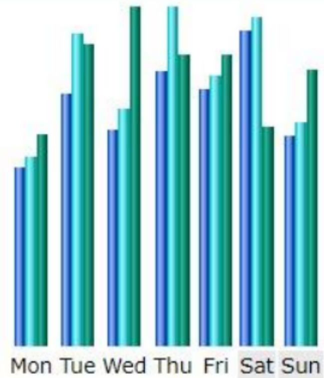
* Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.

(total data on month:March 2024)



(day wise data)

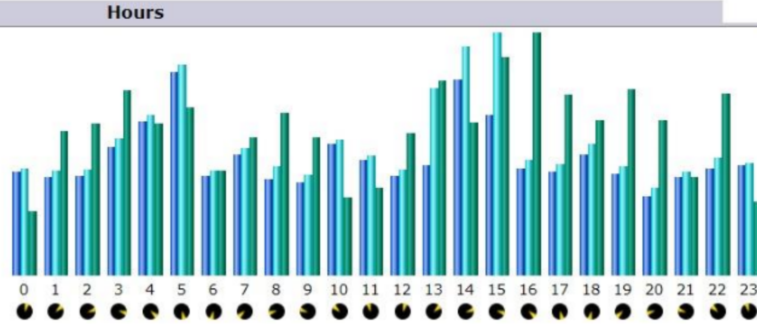
Days of week



Day	Pages	Hits	Bandwidth
Mon	339	359	10.36 MB
Tue	480	594	14.75 MB
Wed	413	451	16.58 MB
Thu	524	647	14.26 MB
Fri	490	517	14.25 MB
Sat	601	626	10.75 MB
Sun	400	426	13.47 MB

(week wise data)

Hours



(Hour wise data)

Locales (Top 25) - Full list				
Locales		Pages	Hits	Bandwidth
United States	us	10,397	11,430	140.71 MB
Unknown	ip	1,210	1,487	110.82 MB
Great Britain	gb	384	409	18.57 MB
Israel	il	358	358	5.28 MB
India	in	182	259	25.81 MB
Russian Federation	ru	147	160	4.71 MB
China	cn	139	169	8.43 MB
Canada	ca	72	89	8.85 MB
Germany	de	70	75	3.76 MB
Ukraine	ua	45	45	703.85 KB
Sweden	se	38	41	4.03 MB
France	fr	33	40	2.18 MB
Netherlands	nl	26	34	1.61 MB
Poland	pl	23	23	122.51 KB
Australia	au	21	35	6.28 MB
Seychelles	sc	20	21	239.89 KB
Hungary	hu	19	19	285.36 KB
Bulgaria	bg	16	17	232.27 KB
Indonesia	id	16	20	1.90 MB
South Africa	za	15	21	2.18 MB
Estonia	ee	14	14	192.21 KB
Lithuania	lt	14	16	284.82 KB
Brazil	br	13	15	2.69 MB
Philippines	ph	12	15	2.01 MB
Pakistan	pk	10	19	1.22 MB

(Location wise data)

2. Probability Modelling:

- Utilize probability theory and statistical distributions to model user behaviour.
- Employ the Poisson distribution, which is commonly used to model the arrival of events over time, to predict the likelihood of future requests for specific assets.
- Calculate probabilities of asset requests based on historical data and user behaviour patterns.

3. Cache Management:

- Develop algorithms for proactive caching based on probabilistic predictions.
- Determine which assets to cache and where to cache them within the CDN infrastructure to minimize latency and maximize hit rate.
- Consider factors such as asset popularity, access frequency, and geographic distribution of users.

4. Cache Replacement Policies:

- Implement cache replacement policies that prioritize assets based on their probability of future requests.
- Consider policies such as Least Recently Used (LRU), Least Frequently Used (LFU), or variants that take into account probabilistic predictions.

Create a table of top assets that will be in demands based on given data and accordingly cache assets in given time frame and refresh it after a certain amount of time whichever time duration works for you

Key point to create a table are:

Consider :

Terms	Abbreviations
Hourly Hits	HH
Daily Hits	DH
Weekly Hits	WH
Monthly Hits	MH

Rank forgiven time = 2*HH + 1.5*DH + WH + MH

By assigning every assets with this formula you get a number let's call this number prob number of an asset
then accordingly prob number create a table with highest prob number having high rank

Sort it

And as your cache storage add top assets in caching

Here's sample :

Pages-URL (Top 25) - Full list - Entry - Exit

484 different pages-url	Viewed	Average size	Entry	Exit
/wp-admin/admin-ajax.php	4,517	38 Bytes	45	174
/wp-cron.php	3,950		132	72
/	923	15.06 KB	550	449
/wp-login.php	601	1.75 KB	467	472
/feed/	533	19.31 KB	430	308
/questions-i-ponder-bloghop/	242	8.87 KB	69	165
/wp-json/oembed/1.0/embed	224	1.13 KB	40	86
/comments/feed/	194	2.05 KB	145	149
/contact/	160	11.45 KB	94	83
/summary-captive-prince/	134	20.47 KB	119	104
/aksara-bhagavad-gita-book-review/	100	19.11 KB	89	87
/feed/atom/	80	64.21 KB	66	68
/jade-city-summary/	68	19.39 KB	63	54
/dawn-gods-united-reich-by-d-alexander-book-review/	56	4.61 KB	13	48
/sitemap_index.xml	56	351 Bytes	31	32
/book-review-malice/	52	18.48 KB	41	35
/pestilence-book-review/	45	19.24 KB	39	32
/wp-json/wp/v2/posts/883/autosaves	43	1.39 KB		
/wp-json/wp/v2/tags	38	321 Bytes		
/category/personal-blog/	38	14.91 KB	15	18
/wp-json/wp/v2/categories	34	483 Bytes		
/book-review-summary-the-ninth-rain/	32	18.91 KB	12	5
/the-battle-of-vathapi-nandis-charge-book-review/	30	17.55 KB	11	11
/blood-over-bright-haven-book-review/	27	19.85 KB	20	16
/wp-admin/post.php	25	118.76 KB		

Analysis:

To analyse the data of request URLs with time for a statistical model, you can follow these steps:

1. Data Preparation:

- Organize your data into a structured format, with each row representing a request URL and its associated timestamp.
- Ensure that timestamps are in a consistent format and that they capture the time accurately (e.g., date and time with milliseconds if available).

2. Exploratory Data Analysis (EDA):

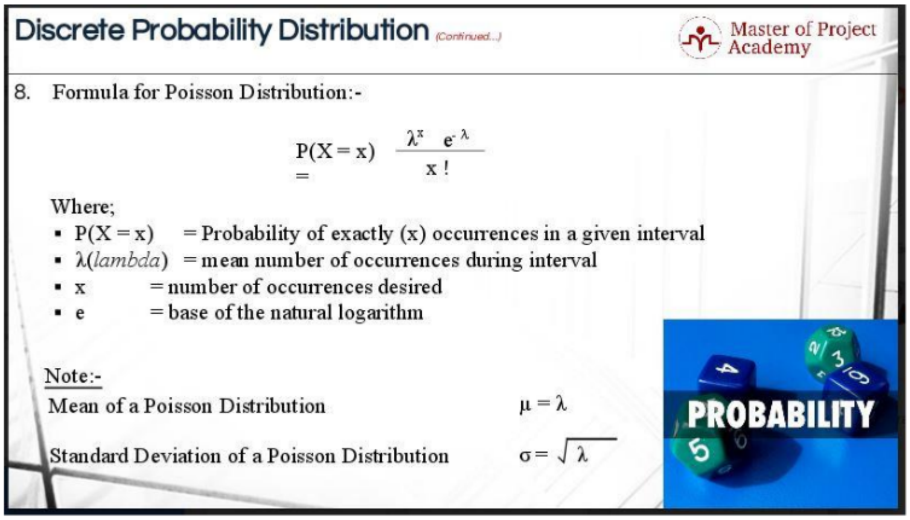
- Conduct exploratory data analysis to understand the distribution of your data.
- Check for missing values, outliers, and inconsistencies.
- Visualize the distribution of timestamps (e.g., histogram, time series plot) to understand patterns and trends.

3. Statistical Modelling:

- Choose an appropriate statistical model based on the nature of your data and the questions you want to answer.
- For time series data, consider models like ARIMA (Autoregressive Integrated Moving Average), SARIMA (Seasonal ARIMA), or Prophet.

Formula:

- N,the number of trials in indefinitely large,n →infinity
- P,the constant probability of success for each trials in indefinity small p→0
- Np=lambda is finite



Solving CDN with poison distribution:

1. Using a CDN with a Poisson distribution may look like this:

   Traffic Modeling: To simulate the traffic that your CDN will receive, you may utilize a Poisson distribution. This entails calculating the typical pace at which requests (such those for web pages, downloaded files, etc.) reach your CDN servers.

2. Configure your CDN: Using your Poisson distribution model as a guide, configure your CDN settings to manage the anticipated load. To effectively manage the incoming requests, this may entail allocating bandwidth, optimizing server locations, and putting up suitable caching mechanisms.

3. Load Balancing: To disperse incoming traffic among several servers or edge locations, CDN providers frequently include load balancing features. You may optimize load balancing algorithms to guarantee that traffic is dispersed efficiently, minimizing delay and optimizing throughput, by comprehending the Poisson distribution of incoming requests.

4. Capacity Planning: You may also plan the capacity of your CDN infrastructure by using the Poisson distribution. You may scale your CDN resources (such as server capacity and bandwidth) to manage peak loads while lowering expenditures during periods of low demand by examining past traffic patterns and forecasting future demand.

5. Performance Optimization: You may enhance the efficiency of your CDN infrastructure by comprehending the features of the Poisson distribution. To enhance response speeds and lessen the strain on origin servers, you may, for instance, modify cache expiration regulations, put prefetching techniques into place, or fine-tune caching parameters.

Conclusions:

The research highlights the transformative potential of probability-driven caching in improving Content Delivery Networks (CDNs) and enhancing user experiences across digital platforms. It demonstrates a significant improvement in performance metrics, reducing latency and optimizing content delivery efficiency compared to traditional methods. The Poisson distribution model effectively pre-empts user requests, minimizing content delivery time and enhancing responsiveness.

Optimized resource utilization, based on predictive analytics, increases cache hit rates, reduces load on origin servers, and enhances bandwidth efficiency. This not only saves costs but also enhances the scalability and resilience of CDN infrastructures. The predictive power of probability models allows proactive caching decisions to adapt dynamically to changing network conditions and user behaviours, ensuring optimal performance across diverse geographical regions and content types.

The findings highlight the pivotal role of probability-driven caching in delivering seamless, high-quality user experiences in a digital world. As the digital landscape evolves, the adoption of advanced caching strategies will be crucial for CDN operators to meet the growing demands of online content consumption.

References:

Configanator: A Data-driven Approach to Improving CDN Performance. Usama Naseer Brown University Theophilus A. Benson Brown University

Popularity prediction in content delivery networks
https://ieeexplore.ieee.org/abstract/document/7343641

RevOPT: An LSTM-based Efficient Caching Strategy for CDN
https://ieeexplore.ieee.org/abstract/document/9685051

LiveNet: a low-latency video transport network for large-scale live streaming
https://dl.acm.org/doi/abs/10.1145/3544216.3544236

Predictive CDN Selection for Video Delivery Based on LSTM Network Performance Forecasts and Cost-Effective Trade-Offs

Evaluating and Improving a Content Delivery Network (CDN) Workflow using Stochastic Modelling

An Overview of Cloud Based Content Delivery Networks: Research Dimensions and State-of-the-Art
https://www.researchgate.net/publication/286221526_An_Overview_of_Cloud_Based_Content_Delivery_Networks_Research_Dimensions_and_State-of-the-Art

An Analysis of Internet Content Delivery Systems Stefan Saroiu, Krishna P. Gummadi, Richard J. Dunn, Steven D. Gribble, and Henry M. Levy Department of Computer Science & Engineering University of Washington

https://citeseerx.ist.psu.edu/document?doi=131c76a896805508583b6d8bc046d0f8673ed5f8&repid=rep1&type=pdf

Challenges and Opportunities in Content Distribution Networks: A Case Study

https://www.diva-portal.org/smash/get/diva2:1010458/FULLTEXT01.pdf

Orchestrating Massively Distributed CDNs

https://conferences.sigcomm.org/co-next/2012/eproceedings/conext/p133.pdf

Structured Ranking Learning using Cumulative Distribution Networks Jim C. Huang Probabilistic and Statistical Inference Group University of Toronto Toronto, ON, Canada M5S 3G4 jim@psi.toronto.edu Brendan J. Frey Probabilistic and Statistical Inference Group University of Toronto Toronto, ON, Canada M5S 3G4

https://proceedings.neurips.cc/paper_files/paper/2008/file/03c6b06952c750899bb03d998e631860-Paper.pdf

Darwin: Flexible Learning-based CDN Caching Jiayi Chen1 , Nihal Sharma1 , Tarannum Khan1 , Shu Liu2 , Brian Chang1 , Aditya Akella1 , Sanjay Shakkottai1 , Ramesh K.

Sitaraman3 1The University of Texas at Austin, 2UC Berkeley, 3UMass Amherst & Akamai Tech

https://dl.acm.org/doi/pdf/10.1145/3603269.3604863

SIMULATION AND OPTIMIZATION OF CONTENT DELIVERY NETWORKS CONSIDERING USER PROFILES AND PREFERENCES OF INTERNET SERVICE PROVIDERS Peter Hillmann Tobias Uhlig Gabi Dreo Rodosek Oliver Rose Department of Computer Science Universitat der Bundeswehr M ¨ unchen ¨ Werner-Heisenberg-Weg 39 Neubiberg, 85577, GERMANY

https://arxiv.org/pdf/2005.13896.pdf

AdaptSize: Orchestrating the Hot Object Memory Cache in a Content Delivery Network Daniel S. Berger, University of Kaiserslautern; Ramesh K. Sitaraman, University of Massachusetts Amherst and Akamai Technologies; Mor Harchol-Balter, Carnegie Mellon University

https://www.usenix.org/system/files/conference/nsdi17/nsdi17-berger.pdf

SDFVAE: Static and Dynamic Factorized VAE for Anomaly Detection of Multivariate CDN KPIs
https://www.researchgate.net/profile/Tao-Lin-58/publication/352111679_SDFVAE_Static_and_Dynamic_Factorized_VAE_for_Anomaly_Detection_of_Multivariate_CDN_KPIs/links/62a29c8d55273755ebe1a1c2/SDFVAE-Static-and-Dynamic-Factorized-VAE-for-Anomaly-Detection-of-Multivariate-CDN-KPIs.pdf

two

ORIGINALITY REPORT

3%	2%	1%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.oocities.org Internet Source	1%
2	fastercapital.com Internet Source	1%
3	Submitted to Yew Chung International School of Beijing Student Paper	1%

Exclude quotes	On	Exclude matches	< 15 words
Exclude bibliography	On		