Impacts of Realistic Traffic on the Mapping Scheme in Locator/Identifier Separation Networks

WANG Hongchao, ZHANG Hongke, LUO Hongbin

National Engineering Laboratory for Next Generation Internet Interconnection Devices

School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing, China

e-mail: wanghongchao@126.com

Abstract: There is a growing consensus on the Locator/Identifier separation scheme to address the scalability of the Internet. Its feasibility depends largely upon the performance of the Locator/Identifier mapping system. This paper presents a special study of the characteristics of the realistic campus network traffic and analyses its impacts on the the mapping scheme in Locator/Identifier separation networks. The basis of our study is a 24-hour trace over 13,233 IPv4 addresses, accounting for about 1/3 of the total IPv4 address space of our campus. From the view of the campus network egress, we give out the statistical charateristics of the trace data, for example, the space distribution of the traffic, the average of coresspoding nodes for per local node, the distribution of the packet arrival intervals accoding to their source and destination addresses, respectively. All the analysis results are of great significance to the design and deployment of the the mapping scheme.

Keywords: measurement; traffic analysis; Locator/ID separation; mapping system

1 Introduction

The current Internet is suffering serious challenges in the routing scalability, multi-homing, mobility and security [1][2]. Now, it is known that all the issues are mainly due to the semantics overload of the IP address, presenting for both the identifier (used by transport layer) and the locator (used by the routing system) of the host. Therefore, there is a growing consensus that it is necessary to separate the locator role and the identifier role of IP address in a future Internet [4]-[13].

Base on the consensus, both the Internet Engineering Task Force (IETF) and the Internet Routing Task Force (IRTF) propose a few novel and alternate network architectures to meet the above mentioned challenges. Meanwhile, some evaluations [3][16][17] have been done to evaluate the benefits brought by separating the identifier and the locator. As a common denominator, all of them consider Locator/Identifier separation as a basic component of the future Internet architecture.

In the prior work of our team, we also presented the new network architecture [13]-[15], in which we introduce four identifiers and three mapping processes among them to finish the communication between any two nodes through the new network.

From all the above related work, we can abstract a generic network model that can describe the characteristics of Identifier/Locator separation. Figure 1 shows the basic network model in Identifier/Locator separation networks.

As Figure 1 shown, the whole network model consists of a lot of provider networks and their users, end hosts or customer networks. In the border of the provider network, the mapping router is responsible for the accessing of users and mapping between the host identifier and the locator. Through the mapping resolution system, the mapping routers distributed in all the provider networks finish the mapping processes, from identifier to locator when the packet enters into the provider network and from locator to identifier as the packet leaves the provider network. The mapping resolution system can be viewed as a logic network composed by the mapping routers themselves or the third agents.

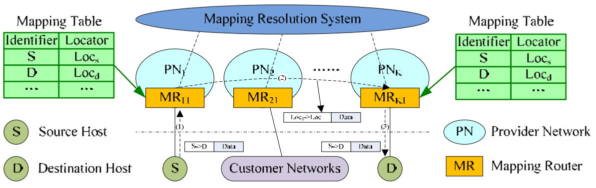

Figure 1. The basic network model in Identifier/Locator separation networks.

In order to introduce less latency when process the packets across between the provider networks and the customer networks or the users, the mapping routers need to cache the mapping items for both the source and destination. Therefore, the mapping routers become the key devices, and their performances directly affect the network performances and the feels of the users.

However, no matter which scheme, as referred to above, is adopted, the quantitative evaluation is needed. Therefore, some related work [16][17] has been done. But the principle of packet arrival and the average of corresponding nodes for per node are not mentioned. They are important to design the mapping scheme supporting host mobility and evaluate its feasibility.

In this paper, on the purpose of evaluating the impacts of realistic traffic on the Identifier/Locator separation scheme, we base our research on a 24-hour traffic trace of our campus. We analyze the characteristics of the traffic, and discuss some considerations of the mapping system.

The remaining parts of the paper are organized as follows. In Section 2, the traffic data set is introduced. Then, Section 3 gives the statistic characteristics of the local addresses. The statistic characteristic of the outside addresses is shown in Section 4. Some impacts of the realistic traffic on the Identifier/Locator separation networks are considered in Section 5. Section 6 concludes the paper.

2 Traffic Data Set

Our data set is obtainted from the information center of our university. It manages and maintains all of network resoureces and could collect and trace all of the in/egress traffic data of the campus. Now, our campus totally has 10 CIDR (Classless Inter-Domain Routing) address blocks (added up to 137C of IPv4 address space, 35072 addresses). The campus network has two border egress routers connected to the Internet, one is 1 gigabits/s (Gbits/s) link towards China Telecom, and the other one is also 1 Gbits/s link connecting CERNET, shown as Figure 2. Through the NetFlow [18] measurement facility, we collect the campus in/egress traffic trace, composed of all the information about a flow. From Section 1, we can learn that the traffic trace collected from the campus border router is just fit for the evaluation of the mapping scheme deployed in the mapping routers.

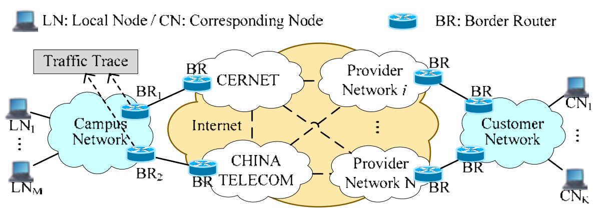

Figure 2. The sketch map of our campus network connecting to the Internet with two physical links.

Because of the similarity of the data among days, we only present and analyze the data on Oct. 12, 2008. Meanwhile, since the ingress packets and the egress ones are mostly symmetrical, we mainly report results for the outgoing traffics. The 24-hour egress traffic traces totally contain 158,844,560 records, related to 13,233 local IPv4 address and 2,129,337 outside ones.

Figure 3 shows the numbers of the unique local and outside IPv4 addresses in the records on the minute basis individually. As shown in the figure, first, the number of the observed addresses over time traces out the principle of the Internet users, fewer addresses observed from midnight to early morning, rapidly increasing addresses in the morning, and then more addresses until people going to sleep again. Second, the number of the unique local addresses observed and the one of outside have the closely similar rate of change, except for the order.

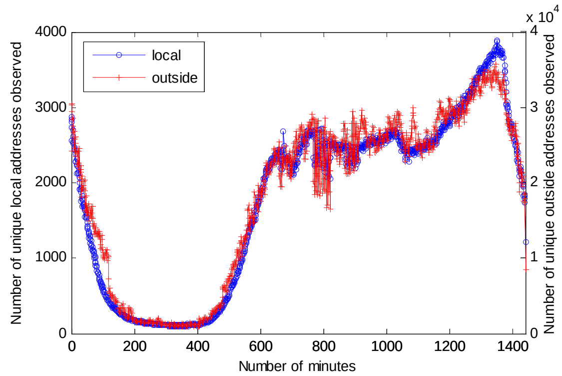

Figure 3. Numbers of records of unique local and outside addresses over time. The x-axis uses 00:00:00, Oct. 12, 2008 as the origin.

3 Characteristics of Local Addresses

In this section, we study the behavior of the local IPv4 addresses. We start from aggregated observations and then delve into finer granularities.

1) Traffic Spatial Distribution

To quantify the characteristics of the local addresses, we first give out the cumulative distribution function (CDF) of the packet number related to the local addresses, illustrated by Figure 4. From the picture, we can learn that most of the local addresses (about 70%~80%) send 10^3~2×10^4 packets individually in the 24 hours. The rate of the local addresses sending less than 10^3 or more than 2×10^4 packets is very low.

Besides that, we further consider the relation between the percentage of the packets and the one of the local addresses. As is illustrated by Figure 5, it is clearly observed that about 20% local addresses send out more than 70% packets, very closed to the Pareto principle [19].

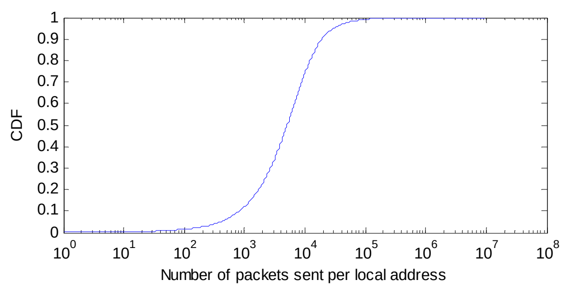

Figure 4. The CDF of the packet number the local addresses send individually. The x-axis uses a log scale.

978-1-935068-06-8 © 2009 SciRes.

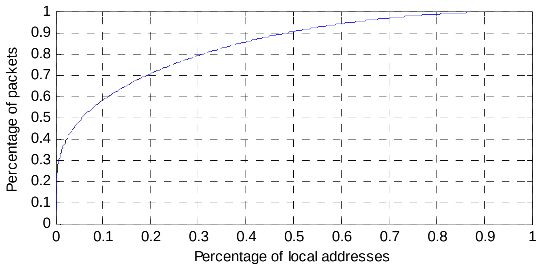

Figure 5. The relation between the percentage of the packets and the one of the local addresses.

2) Packet Arrival Interval Distribution

Packet arrival rate is an important parameter for the selection of the routers, since the routers look up the routing item and forward the packets according to the destination addresses encapsulated in the packet headers. In the Locator/Identifier separation schemes, the routers responsible for the Identifier/Locator mapping still work in the packet granularity. Therefore, we give out the probability distribution function (PDF) of the packet arrival intervals for the local addresses by the Figure 6.

From the figure, we can clearly find that the the PDF of the packet arrival intervals for the local addresses approximately obeys Zipf's law [20], a classic power law function, although it has a heavy tail.

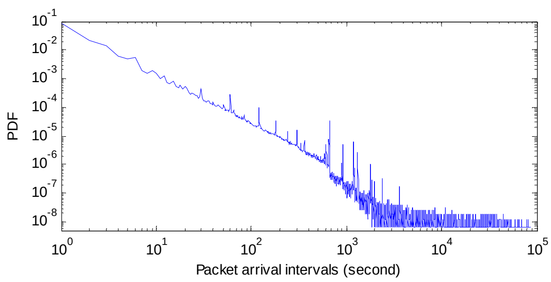

Figure 6. The PDF of the packet arrival intervals for the local addresses. Both the x-axis and y-axis use the log scale.

4 Characteristics of Outside Addresses

In this section, we study the characteristics related to the outside addresses.

1) Traffic Spatial Distribution

While analyzing the local addresses statistically, we also do it as the outside ones. Figure 7 gives out the relation between the percentage of the packets and the one of outside addresses. As can be seen in the figure, only top 1% of outside addresses take up more than 80% of packets. Whereas, more than 90% of the outside addresses only take up less than 10% of packets.

Figure 8 shows the cumulative distribution function (CDF) of the packets number related to the outside addresses individually in the 24 hours. From the figure, we can see that most of the outside addresses (about 96%) appearance less than 10^2 times.

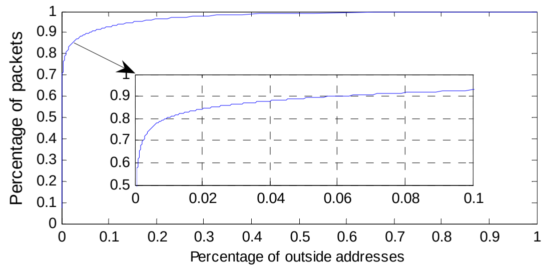

Figure 7. The relation between the percentage of the packets and the one of the outside addresses.

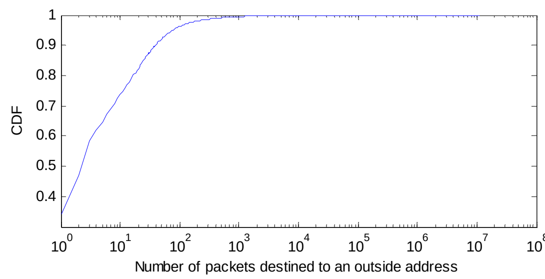

Figure 8. The CDF of the packet number related to the outside addresses. The x-axis uses a log scale.

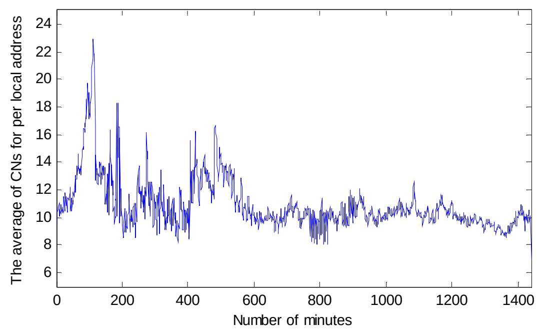

Figure 9. The average of CNs over one local host

2) The Average of outside addresses over one local address

To evaluate the cost of host movement in the Identifier/Locator separation scheme, we first should know how many corresponding nodes (CN) the current host is communicating to. Based on the local and outside addresses observed over the time, we can get the average of CNs for one local host just as shown in Figure 9. From the figure, we can easily learn that the average varies from 7 to 23, with a central value of about 10.

3) Packet Arrival Interval Distribution

Since the mapping routers in the Identifier/Locator separation networks impossibly have all the mapping items, they need to cache the most of the mapping items currently being used; otherwise it would have to spend more cost to query them. Thus, the study on the arrival interval distribution of packets related to the outside addresses is very valuable for the mapping systems.

Figure 10 shows the PDF of the packet arrival intervals for the outside addresses. From the figure, we can see that the PDF of the packet arrival intervals for the outside addresses also approximately obeys the Zipf's law [20], as is the one of the local addresses.

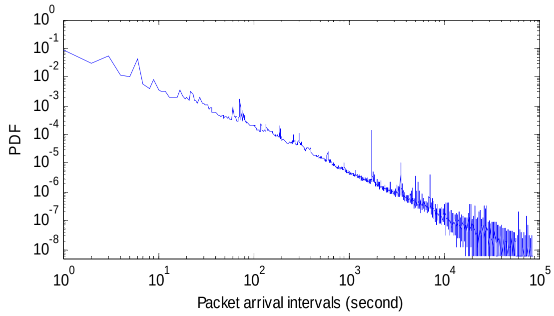

Figure 10. The PDF of the packet arrival intervals for the outside addresses. Both the x-axis and the y-axis use the log scale.

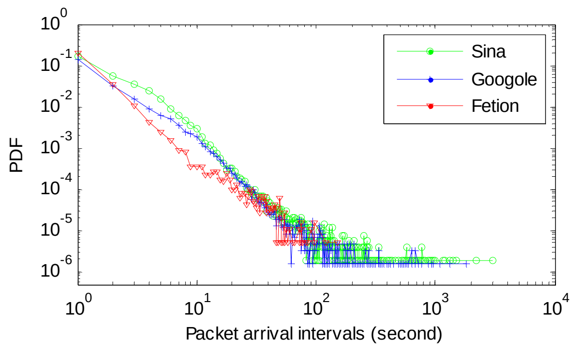

Figure 11. The PDF of packet addrival intervals about three classic addresses. Both the x-axis and the y-axis use the log scale.

Next, we study whether the Zipf's law is suitable for the individual outside address. To answer this question, we select three classic addresses, corresponding to sina, google and fetion individually, and plot their PDF of packet arrival intervals in Figure 11. From the figure, we can learn that the PDF of the individual outside address still approximately obeys the Zipf's law [20], although they have slightly different origins and slops.

5 Discussions

Based on the above analysis of the realistic traffic trace, we will consider the impacts of the realistic traffic on the design and deployment of the mapping scheme in Identifier/Locator separation networks in the flowing aspects:

1) Management Scale of per Mapping Router

The management scale of one mapping router is a very practical metric on the deployment. It cannot be too small or too large. If too small, first it would bring the flexibility of management; second it would introduce more cost of mapping items update due to host mobility. Otherwise, the mapping router cannot bear the overload caused by much more users.

The selection of the management scale of one mapping router can be determined according to its performance. Moreover, the performance of the mapping router exactly is closely related to the characteristics of the traffic and the parameters setting by the mapping scheme.

Therefore, the analysis results of the realistic traffic provide some important references for the design of the mapping scheme. For example, we can make a model to evaluate the feasibility of the mapping scheme according to the change principle of users, the pcaket arrival rate, the average of CNs for one user, etc. Due to the space limitation, the specific model will not be introduced in this paper.

2) Cache Time of Mapping Items

Given the number of users, the number of cache item in the mapping router depends largely on the cache time of mapping items. It also becomes the most important factor that affects the deployment of the mapping scheme.

Relative to the items of outside addresses, the cache size for the local addresses is very small. Take our campus for an example, even the mapping router caches all the mapping items of the local users all the time, the total number of cache items only is up to 13,233. We will not give more descriptions about the local addresses in the paper. Conversely, the cache size for the outside addresses is very large. Figure 12 shows the number of cache items for the outside addresses when the cache time is 3 minutes, 30 minutes and 300 minutes, respectively.

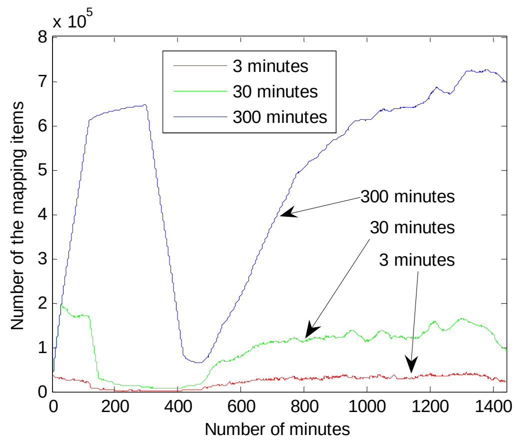

Figure 12. The number of cache items for the outside addresses over the time.

From Figure 12, we can see that when the cache time is set to 3 minutes, the maximum cache items could be below 10^5 (actually 42,679). However, the number of cache items increases significantly when the cache time is set to be longer. When the cache time is 300 minutes, the peak of cache items can reach to more than 7×10^5. This number means that the mapping router must assign a very large space of cache memory

978-1-935068-06-8 © 2009 SciRes.

and spend much processing time managing and maintaining the cache items. Maybe, the cost would not be afforded for most of the current used routers according to their performances.

However, the increasing number of the cache items means the low cache misses. If the item exists in the cache, then the mapping router can directly use it, without sending a request to the mapping resolver for querying the mapping item. Figure 13 shows the percentages of the misses per minute when the cache time is 3 minutes, 30 minutes and 300 minutes, respectively. From the figure, we can see that the percentage of cache misses is reduced with the increasing cache time, but the difference is slight, since both the three curves keep a steady value below 4%.

Anyhow, the cache time is an important parameter for the deployment of the mapping scheme. We should balance the cost of maintaining the cache items and the one due to cache misses to get its optimal value.

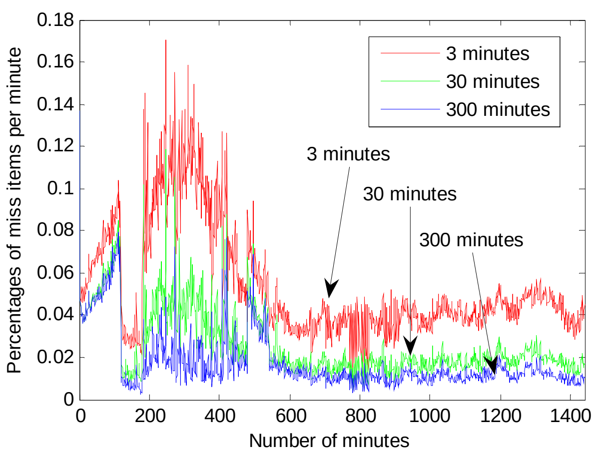

Figure 13. The percentage of cache misses for the outside addresses per minute.

3) Cost of Node Movement

In the most of mapping schems, there are few discussions about the cost when the node movement occurs. However, the node movement can bring a lot of problems, one host usually has more than one correspongding node after all illustrated by Figure 9, not as one in the traditional mobile telecom networks.

Handover lantency is an important metrics to evaluate the mobility scheme. Identifier/Locator separation brings an advantage that the connection established based on the identifiers between communicating parties can be keep alive. However, it is still a problem how to make the mapping routers serving for the CNs to sense the updated mapping items of the moble host immediately.

In addition, the network overhead due to host movement is another important metrics. An effective scheme should make the network overhead as little as possible.

Studying the characteristics of the realistic traffic is helpful to design and optimize the mapping scheme to support the host mobility.

4) Cache Privilege of Top Addresses

In Section 4 and Section 5, we can see that most of the network traffic is made by very few addresses. If they were given the privilege of mapping cache, the most of the network traffic can be guaranteed. The specification description will be introduced in the future paper.

6 Conclusions

In this paper, we analyse a 24-hour traffic trace of our campus, and give out some statistic characteristics of the traffic. We have focused on the space distribution of network traffic, the CDF of the packet number related to one address, the PDF of packet arrival intervals according to the local and outside addresses, respectively, and the average of CNs for one local host. Our study leads to some interesting observations, which provide important references for the design and the deployment of the mapping scheme in Identifier/Locator separaion networks. For instacnce, most of the network traffic is due to very few addresses; the PDF of packet arrival intervals approximately obeys Zipf's law no matter according to the whole address space or the single address; the average of CNs for one host is about 10, etc.

In addition, we discuss some impacts of the realistic traffic characteristics on the design and the deployment the mapping scheme in Identifier/Locator separation networks. These considerations include the management scale of per mapping Router, the cache time of mapping items, the cost of node movement, the cache privilege of top addresses. However, it is not specific that how to use the analysis results to model and evaluate the mapping scheme, which is our future research work.

Acknowledgements

This work is partially supported by National Basic Research Program of China (“973 program”) under contract No. 2007CB307101 and No. 2007CB307106. We would like also to thank the BJTU Information Center, for the dataset they help to make available.

References

[1] G. Huston. “Growth of the BGP Table-1994 to Present,” BGP Routing Table Analysis Reports. http://bgp.potaroo.net.

[2] D. Meyer, L. Zhang, and K. Fall, “Report from the IAB workshop on routing and addressing,” IETF RFC 4984, September 2007.

[3] B. Quoitin, L. Iannone, C. de Launois, and O. Bonaventure, “Evaluating the benefits of the locator/identifier separation,” in Proceedings of ACM SIGCOMM 2007 Workshops, Kyoto, Japan, August 2007, pp. 35–40.

[4] R. Moskowitz and P. Nikander, “Host identity protocol (HIP) architecture,” IETF RFC 4423, May 2006.

[5] E. Nordmark and M. Bagnulo, “Shim6: level 3 multihoming shim protocol for IPv6,” IETF draft-ietf-shim6-proto-12, February 2009.

[6] M. O'Dell, "GSE-an alternate addressing architecture for IPv6," IETF draft-ietf-ipngwg-gseaddr-00, February1997.

[7] D. Farinacci, V. Fuller, D. Oran, *et al*., "Locator/ID separation protocol (LISP)," IETF draft-farinacci-lisp-12, March 2008.

[8] B. Zhang, D. Massey, D. Pei, L. Wang, *et al*., "A secure and scalable Internet routing architecture (SIRA)," University of Arizona, Technical Report TR06-01, April 2006.

[9] R. Whittle. "Ivip (Internet Vastly Improved Plumbing) Architecture," IETF draft-whittle-ivip-arch-02, August 2008.

[10] E. Lear, "NERD: A Not-so-novel EID to RLOC Database," IETF draft-lear-lisp-nerd-04, April 2008.

[11] D. Jen, M. Meisel, D. Massey, L. Wang, B. Zhang, and L. Zhang, "APT: A Practical Transit Mapping Service," IETF draft-en-apt-01, November 2007.

[12] Xiaohu Xu, Dayong Guo, "Hierarchical Routing Architecture (HRA)," in Next Generation Internet Networks, 2008.

[13] H. K. Zhang, W. Su, "Fundamental Research on the Architecture of New Network--Universal Network and Pervasive Services," Acta Electronica Sinica, Vol. 35, No. 4, pp. 593-598, 2007.

[14] P. Dong, Y. J. Qin, H. K. Zhang, "Research on Universal Network Supporting Pervasive Services," Acta Electronica Sinica, Vol. 35, No. 4, pp. 599-606, 2007.

[15] D. Yang, H. C. Zhou, H. K. Zhang, "Research on Pervasive Services Based on Universal Network," Acta Electronica Sinica, Vol. 35, No. 4, pp. 607-613, 2007.

[16] L. Iannone and O. Bonaventure, "On the cost of caching locator/id mappings." in the 3rd Annual CoNEXT Conference, December 2007.

[17] H. Zhang, M. Chen, and Y. Zhu, "Evaluating the Performance on ID/Loc Mapping," in IEEE GLOBECOM, New Orleans, LA, USA, 2008.

[18] Cisco. "Introduction to Cisco IOS NetFlow – A Technical Overview," Available Online at: http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6555/ps6601/prod_white_paper0900aecd80406232.html.

[19] Wikipedia, "Pareto principle," Available online at: http://en.wikipedia.org/wiki/Pareto principle.

[20] Wikipedia, "Zipf's law," Available online at: http://en.wikipedia.org/wiki/Zipf%27s_law.

978-1-935068-06-8 © 2009 SciRes.