Transactions on
Networks and
Communications

TABLE OF CONTENTS

Dr K. Ty Bae
Department of Radiology
University of Pittsburgh
United States

Dr Jiang Hsieh
Illinois Institute of Technology
University of Wisconsin-Madison
United States

Dr David Bulger
Department of Statistics
MACQUARIE University
Australia

Dr YanXia Lin
School of Mathematics and Applied Statistics
University of Wollongong
Australia

Dr Marek Reformat
Department of Electrical and Computer Engineering
University of Alberta
Canada

Dr Wilson Wang
Department of Mechanical Engineering
Lake head University
Canada

Dr Joel Ratsaby
Department of Electrical Engineering and Electronics
Ariel University
Israel

Dr Naoyuki Kubota
Department of Mechanical EngineeringTokyo
Metropolitan University
Japan

Dr Kazuo Iwama
Department of Electrical Engineering
Koyoto University
Japan

Dr Stefanka Chukova
School of Mathematics and Statistics
Victoria University of Wellington
New Zealand

Dr Ning Xiong
Department of Intelligent Future Technologies
Malardalen University
Sweden

Dr Khosrow Moshirvaziri
Department of Information systems
California State University Long Beach
United States

Dr Kechen Zhang
Department of Biomedical Engineering
Johns Hopkins University
United States

Dr. Jun Xu
Sun Yat-Sen University , Guangzhou
China

Dr Dinie Florancio
Multimedia Interaction and Collaboration Group
Microsoft
United States

Dr Jay Stokes
Department of Security and Privacy, Microsoft
United States

Dr Tom Burr
Computer, Computational, and Statistical Sciences Division
Los Alamos National Laboratory
United States

Dr Philip S. Yu
Department of Computer Science
University of Illinois at Chicago
United States

Dr David B. Leake
Department of Computer Science
Indiana University
United States

Dr Hengda Cheng
Department of Computer Science
Utah State University
United States

Dr. Steve Sai Ho Ling
Department of Biomedical Engineering
University of Technology Sydney
Australia

Dr. Igor I. Baskin
Lomonosov Moscow State University,
Moscow
Russian Federation

Dr. Konstantinos Blekas
Department of Computer Science & Engineering,
University of Ioannina
Greece

Dr. Valentina Dagiene
Vilnius University
Lithuania

Dr. Francisco Javier Falcone Lanas
Department of Electrical Engineering,
Universidad Publica de Navarra, UPNA
Spain

Dr. Feng Lin
School of Computer Engineering
Nanyang Technological University
Singapore

Dr. Remo Pareschi
Department of Bioscience and Territory
University of Molise
Italy

Dr. Hans-Jörg Schulz
Department of Computer Science
University of Rostock
Germany

Dr. Alexandre Varnek
University of Strasbourg
France

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

TNC **TRANSACTIONS ON NETWORKS AND COMMUNICATIONS**

Implementation and Verification of High Data Availability on Database

Yen-Jen Chen , Han Tsai
Ming Chi University of Technology, Taiwan, R.O.C.
yjchen@mail.mcut.edu.tw; alsontsai@gmail.com

ABSTRACT

This study provides a low-cost and high-availability database management system architecture for general Small/Medium Enterprises (SMEs) to backup database data access. To prove that the proposed architecture can support the high availability of the database, and can effectively avoid data loss in memory caused by *failovers*, this study applies the main test method of powering off the virtual machine and verified three cases on two commonly used databases MySQL and PostgreSQL: Case 1 proves that this study combines the database native disaster recovery mechanism to effectively achieve high availability of the database. Case 2 proves that it effectively controls the WAL (Write Ahead Log) of the PostgreSQL database and Redo log mechanism of the MySQL database, so that data correctness is maintained during failovers. Case 3 proves that it can analyze and control the timing of the database in writing data in the cache memory to the hard disk. This study also designed a failover process to avoid data loss during failovers due to no enough time to write the data in the cache memory back to the hard disk; and finally to realize the high-availability of the database management system architecture in a practical way.

<u>*Keywords*</u>: Database, DBMS, High Availability, Failover, DRBD

1   Introduction

1.1   Motivation

Information technology infrastructure is an indispensable part of the information environment of modern SMEs. The database (DB) for storing enterprise operational data is the most important part of the information environment because it relates to the lifeblood of a company. In the course of business operation, in order to avoid business interruption, enterprises adopt a backup mechanism so that the network and system services in the information environment will be uninterrupted. Database service is no exception. However, during the operation of the backup mechanism of the database, since the database software has a mechanism for storing data in the memory to enhance the operational efficiency of the system; this mechanism will also lead to data loss due to the switching of the database server while the backup mechanism is in operation.

Therefore, the subject of this study is data loss caused by the operation of database backup mechanism, and we discuss about high data availability mechanism on database management systems.

1.2   Objective

In the past, the high availability mechanism of the database was divided into server parts and database segments, and the two segments need to be separately constructed and maintained. Considering the data consistency of the database, it is necessary to modify the internal source code or use the database communication protocol. In order to achieve this effect, in addition to a certain understanding of the source code, it is necessary to do a certain degree of secondary development, and the complexity is relatively improved. Nowadays, the database of the high availability mechanism is not commonly. One architecture can only achieve a database high availability mechanism for a database, and it needs to pay a high amount of software authorization. Therefore, the architecture flexibility and scalability of the enterprise at this part are in limited This study uses the Clustering and Virtualization technologies of open source licensed Linux to achieve a system platform of high availability, scalability, flexibility, and efficiency, in addition to support commercial software operations for business operations. Clustering is mean that at least two homogeneous devices share a service, and the devices can support each other or share the work. Clustering technology can make the platform have high availability and scalability; On the other hand, Virtualization refers to virtualizing a physical machine into multiple virtual machines, so that hardware resources can be fully utilized to achieve so-called efficiency. Since virtual machines can be transferred between different physical machines, managers can flexibly schedule them. Thus achieving the so-called flexibility. According to DB-Engines, the first global database ranking list [1] in the second half of 2018 was released. The top 4 were found to be 1. "Oracle", 2. "MySQL", 3. "Microsoft SQL Server", and 4. "PostgreSQL". This study uses two high market share databases: PostgreSQL and MySQL to discuss the data high availability mechanism of the database system.

1.3   Organization

This paper is organized as follows. In the second section, the research is explained on the high availability mechanism of the database in recent years. In section 3, the system architecture and research method of proposed scheme is presented. Section 4 shows the analysis of the experimental results and three cases of verifying that the high availability architecture is indeed implemented. Finally, section 5 gives the conclusions of this study.

2   Background

In recent years, among studies on the high availability mechanism of database, C. Jaiswal et al. [2] proposed a mechanism that regardless of whether the backup mechanism is activated, the database system will complete the write operation and write the data to the disk. That mechanism is called Always Ahead Processing (AAP). With this mechanism, high availability on database management system is achieved; RH de Souza et al. [3] proposed an architectural model for assessing the reliability, availability, and maintainability of decentralized databases. The model is based on an approach similar to service quality, and provides database services based on the Service Level Agreement (SLA), to increase the availability of the database.

The architecture design in this study has a high availability mechanism, so there will be two DBMS servers, and the servers are connected to the storage via the ISCSI (Internet Small Computer System Interface) [4][5] protocol. All data is uniformly stored in the storage unit to maintain data consistency. However, in order to back up the data, the data is synchronized to another ISCSI storage by means of DRBD (Distributed Replicated Block Device) [6] distributed mirror. Backup between the two DBMS systems is achieved

through the connection management software Keepalived [7] for the failover function [8][9]. Relevant information such as technologies, systems, and open source software used are as follows:

2.1 DRBD

DRBD (Distributed Replicated Block Device) is based on a distributed storage system on the Linux platform. It is also a technology for instantly synchronizing the data content of the inter-host block storage device through the network. It can be regarded as a kind of network RAID1, that can use storage devices on two independent server hosts through the network as RAID1. Management can be done via the two hosts, and the two sides of RAID1 data are placed respectively on the storage device of these two hosts. Therefore, even if the storage device of one of the hosts is damaged, it will not affect client end usage. After the damaged host is repaired, the data can be recovered by synchronizing the data.

2.2 ISCSI

Internet SCSI (Internet Small Computer System Interface) is a standard developed by the IETF. The principle is to use the network to transport SCSI commands for data transmission. Block-level data transmission is used. The storage device and the host can be divided into two parts. The ISCSI Target, that is the storage device side, is used to store data on the ISCSI disk array. The ISCSI Initiator, that is the client side which uses the Target. The storage space provided by Target can only be used by installing the Initiator.

2.3 Keepalived

A high availability solution is usually used to avoid single-node failures. The mechanism is generally divided into two servers, a primary server and the secondary (i.e. backup) server. To maintain the connection, the primary server sends specific messages to the backup server. When the backup server cannot receive the message, the backup server will take over the virtual IP and continue to provide services to ensure high availability.

2.4 Failover

Whenever a node in a cluster is unavailable, the current service providing node's resources is switched to an available node. A procedure that moves a service from an active node in a cluster to a passive node. It is also a backup. If one or more of the cluster nodes fail, the other nodes will start to take over and continue to provide the service through a procedure called "Failover". There will also be at least one monitoring node responsible for monitoring to confirm that the service is working properly. If the service is not working properly, the user's connection to the service will be switched to another node.

2.5 WAL for PostgreSQL

WAL (Write ahead log) [10] is a technique used in relational database systems to provide Atomicity and Durability, where both are in database ACID properties. The database system software has a mechanism for storing data in the memory to improve the performance of the system. If the database server is unexpectedly shut down, the data in the memory will lost. Then, when the database service is interrupted by power, when the database is restored, it is restored to the hard disk according to the recorded content of the WAL, so that the data remains consistent.

2.6 Redo log for MySQL

MySQL's Redo log [11] and PostgreSQL's WAL have the same effect. If the database server is shut down unexpectedly, the data in the memory is lost. Then, when the database service is interrupted, when the database is restored, the contents of the Redo log are restored to the hard disk to keep the data consistent.

3 Research Methods

In this research, we use the database, MySQL and PostgreSQL, to verify the proposed scheme. The database management systems (DBMS) and databases are stored separately on servers and storage units; the server utilizes ISCSI to connect to storage; for the system backup mechanism, two DBMS servers utilizes high data availability mechanisms to toggle one on active and the other on standby. For the data backup mechanisms, the two storage units are separated as primary and secondary storage with both systems using DRBD mirroring mechanism to achieve data synchronization; in other words, whenever DBMS writes data into primary storage, DRBD synchronizes the data into secondary storage. The further details for the proposed design is shown in Fig. 1.
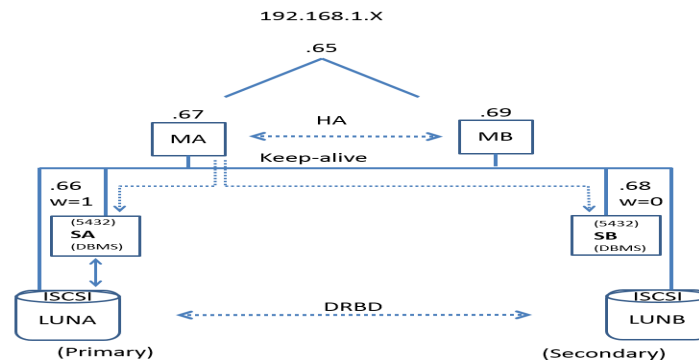

Figure. :1 Database System and Network Structure

MA and MB are master servers while SA and SB are salve servers; the slave is the DBMS server responsible accessing and computing of database. One of the masters (i.e. MA and MB) is active while the other is on standby; the active master always dispatches the database service request to a highly weighted (w=1) slave, and thus the low weighted (w=0) slave is always on standby. The default weight value of SA and SB are 1 and 0 respectively, and when database services of the SA cannot be provided normally, the master will change the weight of SB from 0 to 1 and dispatches the database service request to SB to ensure no data is lost during the failover process. This study was implemented in the PostgreSQL and MySQL database system using its disaster recovery mechanism WAL and Redo log to prevent data loss during failovers. It records all write activities of PostgreSQL and MySQL by recording writing to ensure that the time of WAL and Redo log doesn't fall behind the database. The WAL and Redo log and database are placed in storage together. Since the database cache memory cannot be shut off, DBMS writes data in the memory into database when shutting down normally, aside from the periodic writing of data in the memory into the database. This is used as a reference to write a control program in the masters responsible for activating and shutting off DBMSs in order so that when a failover occurs, the control program will link the database to a new highly weighted slave, say, SB after shutting off DBMS in the other slave, say, SA. If DBMS cannot be shut off, SA will be forced to shut down for flushing the database cache

memory. After shutdown is confirmed, the program activates SB's DBMS to connect to the database and perform the unwritten log in WAL and Redo log to avoid the loss of data in memory having not yet to be written into the database.

The storage DRBD used in this paper adopts the single master mode, which refers to any resource at any specific time. There is only one master node in this two-node cluster. Therefore, the DRBD has two states: Primary and Secondary. As shown in Figure 2, Users' requests will generally be handled by the Primary. The Secondary node is in the Standby state and is responsible as a real-time backup for the Primary. When a hardware failure occurs in the Primary, there is an immediate backup of the data in the Secondary that can be restored.

DRBD Status

Figure. 2: DRBD status display

In order to achieve High Availability (HA), the cluster monitoring software Keepalive in this study detects the database's udp port. When the SA cannot provide services, the master will trigger the control of weight values through the utility "notify_down" as shown in Figure 3.

HA Service instance   Failover set up

Figure. 3: Keepalived Failover set up

Figure 3 shows the Keepalived profile with control settings of connections and weights. Set the SB weight value from 0 to 1, then the connections are directed to SB, which continues to provide the database service for users. If the SA database returns to normal, with "notify_up" trigger to set the SB weight value from 1 to 0, then connections will then be directed back.

In this study, the architecture integrates multiple open source software and database services (PostgreSQL and MySQL). Usually every project here uses memory (cache). To avoid data loss caused by Failover when the data is in memory, the ISCIS cache has been closed. By changing parameters in the ISCSI Target profile, set the IOMode to wt to change the data writing mode as shown in Figure 4.

```
Target iqn.2010-07.com.wishdb:data.iscsi-target
Lun 0 Path=/drbdaa/iscsi/sqdata12.img,IOMode=wt
Alias Lun 0

root@cephadmin:~# cat  /etc/iet/ietd.conf | more
```

Figure. 4: Data Access Mode Settings in an ISCSI Profile

IOMode is set to wt so that data will not pass through the memory, but will be directly written to the hard disk. Some software including the operating system can not close the cache, such as PostgreSQL and MySQL. PostgreSQL and MySQL does not read and write directly to the hard disk. Instead, it puts the data in the hard disk into the Shared Buffers (cache), and uses the LRU (Least Recently Use) algorithm to make the data access complete in memory to improve the performance. Data will then be flushed to the hard disk at regular intervals.

4   Experiment and Verification

The purpose of this study is to ensure the high reliability and the status of the data transferred to the database. There will be three cases in which a failover will be caused by powering off the virtual machines, and the consistency of the data will be compared. The content and principles of each case will be explained below:

Case 1 : Failover data consistency

Figures 5 verify the consistency of the data before and after failover. A dataset is transmitted continuously every second, and the serial number and time recorded in the data is used to verify whether the data is consistent. During the data transmission, the physical machine forces the virtual machine to power off via the command "virsh destroy" to cause a Failover. Data in the user end and the database end are compared for consistency.

| (a) User end | (PostgreSQL) | (b) Database end |

| (b) User end | (MySQL) | (b) Database end |

Figure. :5 (a), (b) Comparison of user-end and database-end's data after a failover

It can be confirmed from Figure 5 that even if an unexpected accident (power outage) occurs, the data can still remain consistent. The key to maintaining consistency is the WAL and Redo log, which records in advance the dataset transmitted by the user into the database. Even if the data is lost in the memory, data can still be recovered through the records in the WAL and Redo log. In order to verify the function of

the WAL and Redo log, this study uses the file system monitoring software Inotifywait [12] to monitor the directory where the WAL and Redo log are located as shown in Figure 6.

(PostgreSQL)

(MySQL)

Figure. :6 Inotifywait WAL and Redo log record screen

It can be confirmed from Figure 6 that when the user transmits a dataset, a modification record will be recorded in the WAL and Redo log file. Therefore, although the data has not been stored in the hard disk, when a Failover occurs, master will connect the database to SB. SB can also recover data through WAL and Redo log.

Case 2: Failover data recovery verification

In Case 1, during continuous data transfer, the virtual machine is powered off to cause a Failover and the data consistency is verified, so as to confirm that a data loss can be recovered by WAL and Redo log. However, in Case 2, only PostgreSQL can be tested because MySQL writes memory data back to its database with an interval not bigger than 1 second. Before transferring the data, copy WAL to the temporary folder first. The WAL saved in this folder has not recorded the data transferred yet. Then 10 datasets are transferred to the database. At this time, because there is a record in WAL about the 10 datasets just transferred, even after the failover, the data will remain consistent. Therefore, after the data is transferred, copy the WAL in the temporary folder to overwrite the WAL in PostgreSQL. Then the virtual machine is powered off to cause a Failover. Afterwards, verify again the state of the data on the user end and database end as shown in Figure 7.
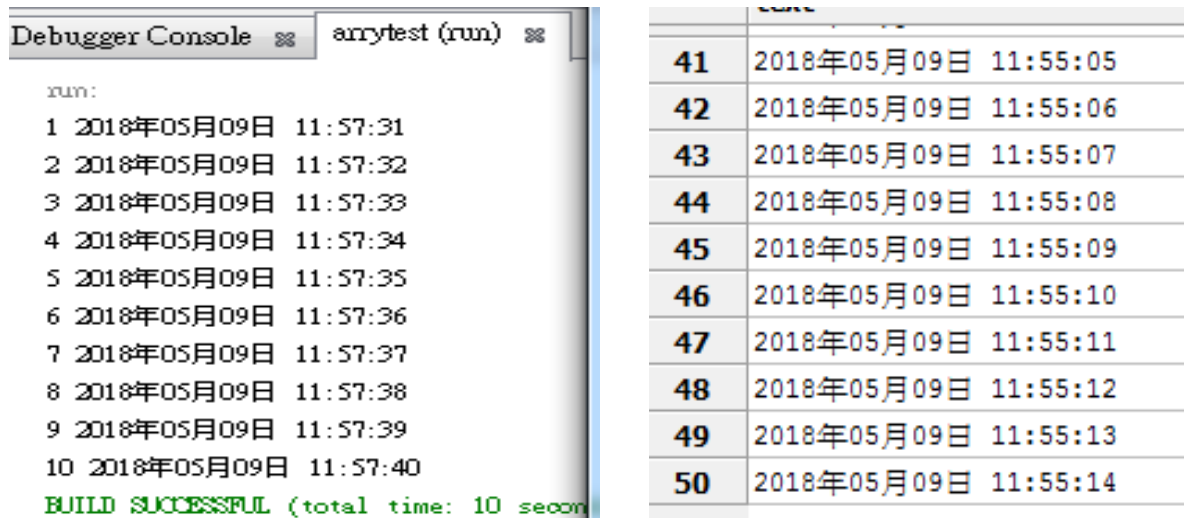
(a) User end   (b) Database end

Figure. :7 (a), (b) Comparison of user-end and database-end data after a Failover

It can be seen from Figure 7 that after the failover, the data in the database-end is inconsistent with the user-end. The reason is that when the data is transferred to the database, the system will store the frequently used data in the memory for access without directly storing it into the database, and copy the WAL that has not recorded the incoming data to overwrite the WAL in PostgreSQL. At this time, if Failover occurs, since the data is not stored into the database, nor has WAL recorded the incoming data, a data loss will be caused. After verifying a Failover by replacing WAL, PostgreSQL uses WAL to recover lost data in memory.

Case 3 : Checkpoint function data is written to the hard disk for verification

Checkpoint [13] is one of PostgreSQL's and MySQL's checkpoints for writing data back to the hard disk. It will write the data in the memory back to the hard disk at intervals, based on the time set by the user. The PostgreSQL default is every five minutes, while the MySQL default is every second. Since the MySQL setting is forced to one second, MySQL is hard to be verified in Case2. After the data is written back to the hard disk, the checkpoint record will be written to the WAL and Redo log. After the database service is restarted, the latest checkpoint will be loaded, and the write operations after this checkpoint will be re-executed to ensure that the data will not be lost. While the data before this point has been written back from the memory to the hard disk by Checkpoint, therefore no action is required.

This is a recovery mechanism that PostgreSQL and MySQL does to avoid data loss. In order to verify that the data is actually written to the hard disk after the checkpoint, the verification steps is as below. The software Inotifywait monitors the folder where the WAL and Redo log are located to observe its status. Before the data is transferred, copy the WAL and Redo log to the temporary folder, and then transfer 10 datasets to the database. Wait for five minutes until Checkpoint is executed. After that, the execution record is written into the WAL and Redo log as shown in Figure 8.

(PostgreSQL)

(MySQL)

Figure. :8 Inotifywait Checkpoint record screen

For PostgreSQL, it can be observed from Figure 8 that there is a modification record in the WAL after five minutes. This modification is to record the data in the memory being written back to the WAL after the

Checkpoint execution is completed. In order to verify that the data in the memory has been written back to the hard disk by Checkpoint, again, use the WAL which has not yet recorded the transferred data to overwrite the WAL in PostgreSQL. Power off the virtual machine to cause a Failover. Afterwards, the master will direct the database connection to SB to reconfirm the status of the data at user-end and the database-end as shown in Figure 9.

For MySQL, it can be observed that there are twenty modified records in the Redo log, ten of which are incoming data, and the other ten are completed by Checkpoint. The data in the memory is written back to the hard disk and recorded to the Redo log. In order to verify that the data in the memory has been written back to the hard disk by Checkpoint. Use the Redo log which has not yet recorded the transferred data to overwrite the Redo log in MySQL. Power off the virtual machine to cause a Failover. Afterwards, the master will direct the database connection to SB to reconfirm the status of the data at user-end and the database-end as shown in Figure 9
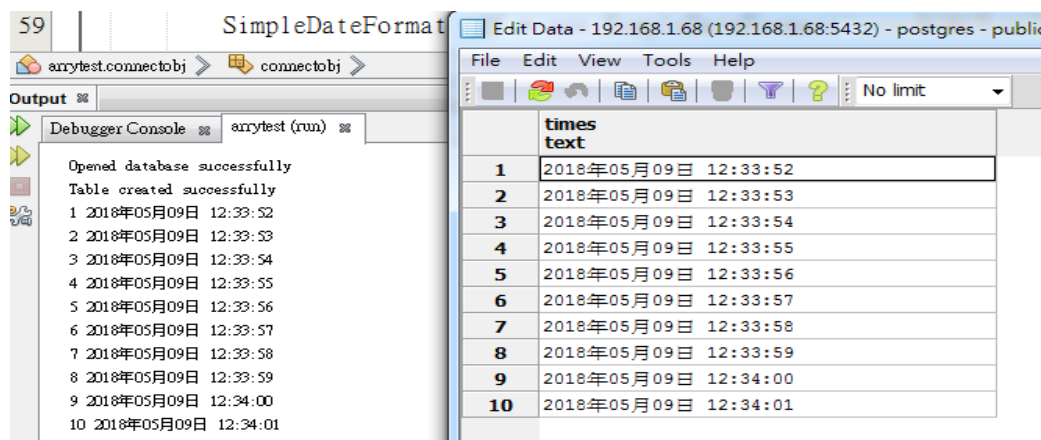
(a) User-end   (PostgreSQL)   (b) Database-end
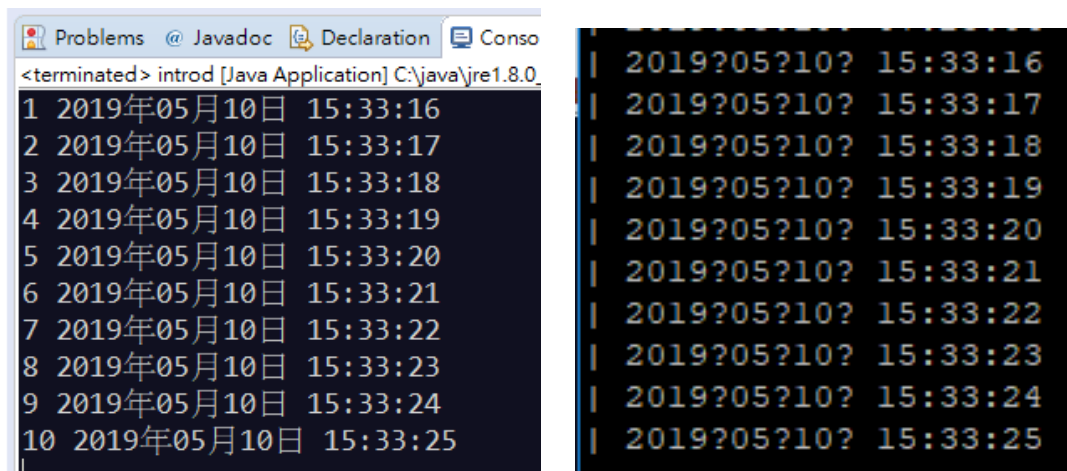
(a) User-end   (MySQL)   (b) Database-end

Figure. :9 (a),(b) Comparison of user-end and database-end data after a failover

From Figure 9 it can be observed that the data on the user-end and on the database-end is consistent. PostgreSQL compared with Case 2, there is 5 minutes more waiting time for Checkpoint to complete and

to record the time in the WAL. No matter is PostgreSQL or MySQL. The result data of this experiment is consistent.

Data still exists in memory and is not yet saved to the hard drive. Also, use the WAL before the data is entered to overwrite the WAL in PostgreSQL; use the Redo log before the data is entered to overwrite the Redo log in MySQL. Therefore, WAL and Redo log did not record the data entry. The data was lost in the memory and WAL and Redo log did not record the written data. The data could not be recovered via the WAL and Redo log. However, after the Failover, it was observed that the data still existed. This can prove that Checkpoint writes data from the memory into the hard disk, and SB can display the ten datasets that the user just entered after the failover.

5   Conclusions

This study proposes a high-availability database management system architecture. For the universality of this research, two databases, MySQL and PostgreSQL, are used to increase the chances of combining with various systems. In order to achieve the capabilities of data backup and system backup at the same time, there are three system operating mechanisms proposed. First, the architecture combines the WAL and Redo log mechanism of the database to implement a database disaster recovery mechanism. Next, this study can control the WAL and Redo log mechanism of the database, so that the correctness of the data is maintained even during a database failover. Finally, this study can analyze and control the timing when the database system writes cached data to the hard disk, and to verify with three experimental cases that the operation of the above three mentioned mechanisms are feasible. Based on the above three mechanisms, this study designs a failover process to avoid data loss in the cache memory due to a failover, so as to achieve a high availability database management system architecture.

REFERENCES

[1]   DB-Engines. (2019). The first global database ranking list in the second half of the 2018, https://buzzorange.com/techorange/2018/08/02/2018-july-database-ranking/

[2]   C. Jaiswal and V. Kumar, "DbHAaaS: Database High Availability as a Service," 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Bangkok, 2015, pp. 725-732.doi: 10.1109/SITIS.2015.25

[3]   R. H. de Souza, P. A. Flores, M. A. R. Dantas and F. Siqueira, "Architectural recovering model for Distributed Databases: A reliability, availability and serviceability approach," 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, 2016, pp. 575-580.doi: 10.1109/ISCC.2016.7543799

[4]   Adnan, A. A. Ilham and S. Usman, "Performance analysis of extract, transform, load (ETL) in apache Hadoop atop NAS storage using ISCSI," 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Kuta Bali, 2017, pp. 1-5.doi: 10.1109/CAIPT.2017.8320716

[5]   A. Elghazi, M. Berrezzouq and Z. Abdelali, "New version of iSCSI protocol to secure Cloud data storage," 2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech),   Marrakech,   2016,   pp.   141-145. doi: 10.1109/CloudTech.2016.7847690

[6]   M. Riasetiawan, A. Ashari and I. Endrayanto, "Distributed Replicated Block Device (DRDB) implementation on cluster storage data migration," 2015 International Conference on Data and Software Engineering (ICoDSE), Yogyakarta, 2015, pp. 93-97.doi: 10.1109/ICODSE.2015.7436978

[7]   Rahul k.(2018). How to Setup IP Failover with KeepAlived on Ubuntu & Debian. Retrieved from https://tecadmin.net/setup-ip-failover-on-ubuntu-with-keepalive/

[8]   Yen-Jen Chen & An-Liang Lo. Design and implementation of construction system with high-availability application service environment, IJCEE 2015 Vol.7(6): 357-369 ISSN: 1793-8163 DOI: 10.17706/IJCEE.2015.7.6.357-369

[9]   R. F. Gibadullin, I. S. Vershinin and R. S. Minyazev, "Realization of replication mechanism in PostgreSQL DBMS," 2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM),   St.   Petersburg,   2017,   pp.   1-6. doi: 10.1109/ICIEAM.2017.8076380

[10]   S. Ryu, K. Lee and H. Han, "In-memory write-ahead logging for mobile smart devices with NVRAM," in IEEE Transactions on Consumer Electronics, vol. 61, no. 1, pp. 39-46, February 2015.doi: 10.1109/TCE.2015.7064109

[11]   Peter Frühwirt , Peter Kieseberg, Sebastian Schrittwieser, Markus Huber, and Edgar Weippl, "InnoDB Database Forensics: Reconstructing Data Manipulation Queries from Redo Logs," in Seventh International Conference on Availability, Reliability and Security, 2012, doi:10.1109/ARES.2012.50

[12]   Rohan McGovern.(2018). Home · rvoicilas/inotify-tools Wiki · GitHub. Retrieved from https://github.com/rvoicilas/inotify-tools/wiki

[13]   S. i. Sou and Y. b. Lin, "Modeling mobility database failure restoration using checkpoint schemes," in IEEE Transactions on Wireless Communications, vol. 6, no. 1, pp. 313-319, Jan. 2007.doi: 10.1109/TWC.2007.05200

The Consumers' perceptive of Internet-of-Things: Exploring Users' Acceptance and Purchase Behavior

Hui-Hsin Huang

Associate Professor, Department of Business Administration, Aletheia University, Taiwan
hoyasophia@gmail.com

ABSTRACT

This paper stands on customer's view of point to discuss the usage on Internet of Things(IoT). A purchase behavior model of IoT is constructed based on ChaCha model to explore the impact factors that influence customer's adoption of IoT. The characteristics of IoT itself and customer individual characteristics are also included. Then a survey data will be conducted to estimate the parameters of the proposed model. These empirical data are also used to make model calibration. Finally, the application of IoT in marketing management will be present in the conclusion.

Keywords: Internet-of-Things(IoT), purchase behavior, ChaCha model

1   Introduction

The Internet of Things (IoT) promises a new technological paradigm(Luet al.,2018), when everyday objects can be augmented with RFID tags and intelligent sensors to become ICCs, real-time data flow can be automatically produced. This trend can improve marketing research into new broad areas (Ngand Wakenshaw, 2017).

Previously physical product consumption would occur in consumers' private spaces, an era of IoT would user in a substantial amount of quantitative data on consumption and experience (Ngand Wakenshaw, 2017). This would create a dimension of visibility where none had previously existed, particularly with information on consumption quantity and **depletion** for particular contexts, interactions between things and between people and things, and information about the environment (Parry et al., 2016).

IoT offers products with various functions and target scope. In the business literature, these characteristics have a significant impact on consumer purchase intention and good-practice principles in product design. Therefore, the purpose of this paper is to explore the IoT adoption or purchase intention of customers and the relationships of its impact factors. This research will construct a model to use these factors to predict the probability of customers' IoT purchase intention (adoption).

2   IoT and Customers' Purchase behavior

There are other researches (Bao et al., 2014; Gao and Bai, 2014; Chang et al.2014;Rau et al., 2015) find that the customers 'experience is the mediator variable which can mediate the effect between IoT characteristics and purchase intention.

Customers' experiments of IoT can be divided into functional experience and emotional experience. The marketing activities such IoT promotion, product design and management can focus on these experiences to facilitate customers' purchase intention. Thus, this paper combines the characteristics of IoT itself and customer individual characteristics to formulate a model to predict the purchase intention that customer's adoption of IoT.

The paper is constructed into four parts. Firstly, we make a literature review to demonstrate the purchase intention model of customer's adoption of IoT. In this part, the ChaCha model (Wang,2014) and its literature is introduced. The previous researches of impact factors on customer's adoption of IoT which include characteristics of IoT and customer individual characteristics are also reviewed. Secondly, the empirical data is used to make parameters estimation and model calibration. In this part, the method of data collection and analysis are present. The theory background of questionnaires is introduced. Thirdly, the results of parameters estimation and model calibration are shown. Finally, the conclusion is made.

3   The theory background of Model

In this research, we based on the concept of ChaCha model (Wang,2014) which combinesthe motivated media processing and selective exposure behavior to discuss the TV viewing experiment and channel choice behavior. In the literature review, we can find that the impact factors of customers' IoT adoption or purchase can be attributed to their motivation. ChaCha model includes the audience's motivation value and motivation strength to forecast viewing behavior. Thus, we use the framework of ChaCha model to portray the customer's IoT purchase behavior.

Based on Wang (2014), we propose that the customers' purchase intention of IoT(p) is a random variable which is following the Wald distribution(inverse Gaussian) with its probability density function(pdf),

$$f(p) = \left(\frac{\sigma}{2\pi p^3}\right)^{\frac{1}{2}} exp\left[-\frac{\sigma(p-\delta)^2}{2\delta^2 p}\right] \tag{1}$$

4   Motivation

Acording to Chang et al.(2014) , Lu et al.(2018), Rau et al.(2015), Kim and Kim, (2016), we can divide the motivation of customers' purchase intention of IoT into the characteristics of IoT itself and customer individual characteristics.

5   Customer individual characteristics

We use α as the customer individual characteristics such as demographic variables orperceived usefulness and perceived ease of use. Thus, α is shown in equation (1) with δ =α/τ^{-1}=ατandσ isthe shape of the Wald distribution (Luce, 1986).

$$\alpha = \beta_0 + \beta_1\alpha'_1 + \beta_2\alpha'_2 + \beta_3\alpha'_3 + ... + \beta_n\alpha'_n \tag{2}$$

in which $\beta_1...\beta_n$ is the coefficient of *n* individual characteristics.

T is motivational strength, τ*(h)* = exp*[m(h)]* , IoT motivational values, which can range from negative (aversive motivational) to positive (appetitive motivational), are transformed by exponential transformation(Wang, 2014).Motivation strength will decay when time passing.

6   Characteristics of IoT

We propose γ is the characteristics of IoT which can attract customers to adopt are a linear equation as

$$m(h) = m(h-1) + a_1\gamma_1 + a_2\gamma_2 + a_3\gamma_3 + \ldots + a_k\gamma_k \qquad (3)$$

in which $m(h)$ as the motivational value of IoT, which is the motivational value used to predict the hth purchase after viewing all the previous $h-1$ occasions. There are k characteristics of IoT.

7 Empirical data

The empirical survey data are used to estimate the parameters of present model and make model calibration. We use survey ethod to collect the empirical data. The duration of survey is from 1 September to 31 October in 2019. There are 2416 sample sizes.

The questionnaire design is based on the previous theory background. For the IoT factors, based on Chang et al.(2014) and Lu et al.(2018), there are six characteristics of IoT which can influence the customers' purchase intention. They are (1) IoT Convenience is "the degree to which consumers save time and effort in the process of planning, purchasing, and using a product".(2) IoT Security is "damage avoidance when it comes to any vulnerable and valuable assets".(3) IoT Telepresence is "the subjective feelings of customers about the extent to which media represent the physical and social environment".(4) IoT Intelligence is" intricate and accurate recognition functions, correct thinking and judgment capabilities". (5) IoT Connectivity is "the degree to which things are interconnected".(6) IoT Interactivity is "the customers' feeling that occurs when information communication is bidirectional and response is timely". These items are measured by Likert five scales.

For customers' characteristics, the most used is technology acceptance model(Kim and Kim, 2016) in which perceived usefulness and perceived ease of use(Bao et al., 2014; Gao and Bai, 2014)to predict IoT adoption. Lu et al.(2018) also mention that the customers' experience is the mediator variable which can mediate the effect between IoT characteristics and purchase intention. And fun and enjoyment are obtaind in IOT function. Thus, we include perceived usefulness, perceived ease of use and customers' experience (fun and enjoyment) as three individual factors to influence the customers' IoT purchase behavior. These items are also measured by Likert five scales.

8 The Parameters estimation

We divide the 2416empirical data randomly into two parts. One(1408 sample size) is for parameters estimation, another is for model calibration. For the questionnaires design, some parameters in model such as ni ndividual characteristics are n=4(there are four impact factors of individual characteristics in questionnaires design), k=6there are four impact factors of IoT characteristics in questionnaires design) and h=4(There are four duration of observing customer purchase occasions which are from 1September to 15September, from 16 September to September 30, from 1 October to 15October, from 16October toOctober). The results of parameters estimations are shown in table 1.

Table 1 The results of parameters estimations

β_0	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	σ
2.417	0.591	0.355	0.677	0.412	3.424	0.268	0.359	0.821	0.664	0.565	0.2

The root-mean-square deviation(RMSD)(Busch et al., 2014) is used as model calibrationto calculate the real data and the data which is stimulated from using the results of parametersestimations. We denote $r_d$is the real data, s_c is the simulation data. To use the 2416 total sample sizes, we divide the whole samples

into two parts. Thus, d=1408 and c=1408. We calculate the average distance between r_d and s_c. The result is shown as

$$\frac{\sum_{d=1,c=1}^{1408} \sqrt{(r_d - s_c)^2}}{1408} = 0.673.$$

The result of RMSD is 67.3% which is smaller than 70%. It means the distance between real data and simulation data is not far. Therefore, the goodness of fit of the proposed model is acceptable.

9   Conclusion and Discussion

This paper proposes a probability model to predict customers purchase intention of IoT. To merge the motivation concept into model construction, the author bases onthe business literature of IoT to include the characteristics of both individual and IoT to make the proposed model more useful. For marketing, managers can find which factors will have great impact on the customer's purchase behavior or what marketing strategies can be controlled to contribute customers to IoT adoption. This proposed model combines the IoT theory background such as technology acceptance model and probability calculation to make the theory more concert and more operational.

In the future, researchers can try to use other individual factors such as customers' demographic data to forecast the IoT adoption behavior. Beside the survey method, other research method such as experiment method also can be conduct to collect customers' psychological variables and find more customers' inside factors to approach IoT use.

For the company side, the traits of IoT such as different functions can be included into the model to provide the reference for IoT product design and make model more application.

THE REFERENCE

[1]   Bao, H., Yee-Loong Chong, A., Ooi, K.-B., Lin, B.(2014). Are Chinese consumers ready toadopt mobile smart home? An empirical analysis.International Journal of Mobile Communications (Int. J. Mob.Commun.), 12(5), pp.496-511.

[2]   Busch,P., Lahti,P. and Werner,R. F.(2014). Colloquium: Quantum root-mean-square error and measurement uncertainty relations.Reviews of Modern Physics, 86 , pp. 1261 -1283.

[3]   Chang, Y., Dong, X., Sun, W.(2014). Influence of characteristics of the internet of things onconsumer purchase intention. Social Behavior and Personality: An International Journal, 42(2), pp. 321-330.

[4]   Gao, L., Bai, X.(2014). A unified perspective on the factors influencing consumer acceptanceof internet of things technology.Asia Pacific Journal of Marketing and Logistics (Asia Pacific J. Market. Log.)26, pp.211-231.

[5]   Kim, S. and Kim, S. (2016). A multi-criteria approach toward discovering killer IoT applicationin Korea. Technological Forecasting and Social Change(Technol. Forecast.Soc. Chang.), 102, pp.143-155.

[6]   Lu,Y.,Papagiannidis, S. andAlamanos, E.(2018). Internet of Things: A systematic review of the business literature from theuser and organisational perspectives.Technological Forecasting & Social Change ,136, pp.285-297.

[7]   Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization.New York, NY: Oxford University Press.

[8]   Ng, C.L. I. and Wakenshaw, Y.L. S. (2017).The Internet-of-Things: Review and research directions. International Journal of Research in Marketing 34 ,3-21.

[9]   Parry, G., Brax, S., Maull, R., & Ng, I. (2016). OperationalisingIoT for reversesupply: The development of use-visibility measures. Supply Chain Management: AnInternational Journal, 21(2), 228–244.

[10]   Rau, P.-L.P., Huang, E., Mao, M., Gao, Q., Feng, C., Zhang, Y.(2015). Exploring interactivestyle and user experience design for social web of things of Chinese users: a case studyin Beijing.International Journal of Human-Computer Studies, 80(8), pp.24-35.

[11]   Wang, Z.(2014). Bridging media processing and selective exposure: a dynamic motivational model of media choices and choice response time.Communication Research, 41(8), pp.1064-1087.

TNC **TRANSACTIONS ON NETWORKS AND COMMUNICATIONS**

Regularities in the Social Network's User Distribution by the Number of Mutual Contacts

[1,2]Ibragim E Suleimenov, [2]Akhat S Bakirov
[1] Almaty University of Power Engineering and Telecommunications, Almaty, Kazakhstan;
[2]National Academy of Engineering of the Republic of Kazakhstan, Almaty, Kazakhstan;
esenych@yandex.ru; axatmr@mail.ru

ABSTRACT

This paper describes a study in which the pattern of distribution of the number of friends among users of the Vkontakte social network among residents of eight large cities of the CIS (Moscow, St. Petersburg, Almaty, Novosibirsk, Tashkent, Kiev, Yekaterinburg, Pavlodar) was studied. Experimental data show that this distribution is of a similar nature for all selected cities. A semi-empirical model was built, on the basis of which an explicit form of theoretical dependence was obtained, describing the nature of the distribution of the number of mutual contacts ("friends") of users of social online networks. It is shown that this theoretical dependence agrees with satisfactory accuracy with experimental data for a sufficiently large sample of cities. It is established that the Dunbar number, which is included in the dependencies considered as a control parameter, is a characteristic of the communication environment of each specific city and correlates with the population of the city.

Keywords: Online social networks, Dunbar number, City communication environment.

1   Introduction

The study of the communication space formed by the rapid development of telecommunication technologies is of considerable interest in several aspects.

Thus, a significant part of campaigns to promote products on the market is currently conducted in online social networks [1-5], and such networks are already used not as an auxiliary means of campaigning, but as the main one.

Social online networks connecting a large number of people are an ideal platform for conducting various informational influences of a mass nature, including in the framework of information wars [6-11].

At the same time, modern telecommunication technologies make it possible to receive education online, which creates a number of advantages (access to educational courses of universities around the world, saving travel time, convenience for people with advanced requirements) [12-16]. Massive open online courses (MOOCs) are beginning to gain great popularity [17-21].

The consistent use of social online networks for various purposes requires their adequate theoretical description on a quantitative level.

In this paper, we analyze the distribution of users of social online networks by the number of friends (mutual contacts) and propose a model that describes this distribution.

2 Research method

The data used in the work was collected on the basis of direct collection of information regarding the number of friends of users of social online networks VKontakte.

A random sample of users living in the following cities was created: Almaty, Moscow, Novosibirsk, Kiev, Pavlodar, Yekaterinburg, St. Petersburg, Tashkent. The number of friends each user has was counted. Based on these data, the dependencies presented in Fig. 1 - 8 (dots). The abscissa axis shows the numbers n falling at the center of the intervals on which the axis is divided. The partition scale is selected uneven. The ordinates show the ratio of the total number of users, whose number of friends falls within this interval, to the length of this interval. Such a construction approximately corresponds to the density of distribution of users by the number of friends.

It can be seen that the curves obtained for various cities in which the Runet is actively exploited have a similar character.

In general, the distribution of the number of users by the number of friends can be found on the basis of the following equation:

$$\frac{dn_j}{dt} = n_{j-1}\sum_{i=0} w_{j-1,i}n_i - n_j \sum_{i=0} w_{j,i}n_i - \frac{n_j}{\tau} \tag{1}$$

nj – the number of users whose friends are j, $w_{j,i}$ — the frequency of the formation of a new relationship between people with j and i friends, τ - network user lifetime (actual time of using the resource).

Assuming that the functions under consideration with a change in the number j change slowly, we can proceed to the continuous form of writing equation (1). Replacing the discrete numbers j and i with continuous variables x and y, we can write

$$\frac{d}{dt}n(x) = n(x - \Delta x)\int w(x - \Delta x, y)n(y)dy - n(x)\int w(x,y)n(y)dy - \frac{n(x)}{\tau} \tag{2}$$

Applying the Taylor series expansion up to quadratic terms, we have

$$\frac{d}{dt}\frac{n(x)}{\Delta x} = -n(x)\int \frac{\partial}{\partial x}w(x,y)n(y)dy - \frac{dn}{dx}\int w(x,y)n(y)dy - \frac{1}{\tau}\frac{n(x)}{\Delta x} \tag{3}$$

Whence we get that for the stationary case

$$\frac{df}{dx} = f(x)K(x) \tag{4}$$

where

$$K(x) = \frac{\left[\int \frac{\partial}{\partial x}W(x,y)f(y)dy - \frac{1}{\tau}\right]}{\int W(x,y)f(y)dy} \tag{5}$$

It is essential that the functions included in the factor for the function f (x) in expression (4) are determined by the entire profile of the distribution under consideration.

Expression (4) allows us to interpret the establishment of an equilibrium distribution of users by the number of friends through a stream directed towards their increase. This, in particular, allows us to use heuristic considerations in finding the function K (x).

These considerations are as follows. The literature currently uses the Dunbar number [22-25], which represents the maximum number of constant social contacts (ties) that one person can support. Initially, this number was obtained by studying the influence of the size and development of the neocortex on the size of a flock of monkeys [22-23], but now it is increasingly used to describe the communication space [26-29].

It can be assumed that the Dunbar number describes the potential causing the flow in a positive direction along the x axis in the model under consideration. If the number of friends of this user is less than the Dunbar number, then we can expect that the derivative in formula (4) will have a positive sign (and vice versa).

The simplest record satisfying this requirement has the form

$$\frac{df}{dx} = (a - x)f^\alpha \tag{6}$$

where the factor α takes into account the fractal nature of the communication space.

This nature of the communication space follows from visual considerations. The formation of a relationship between two persons is facilitated if they have a common friend, i.e. there is a certain chain linking them. The inclusion of such chains leads to a power-law dependence in formula (6) and the fractal nature of the communication space as a whole.

Equation (6) is of the first order and is easily integrated; we have

$$-(1 - \alpha)f^{1-\alpha} = (a - x)^2 + C \tag{7}$$

where C is the integration constant.

From here

$$f = \frac{A}{((a-x)^2+C)^\beta} \tag{8}$$

where the normalization factor A is entered in the record,

$$\beta = \frac{1}{\alpha-1}; \quad \alpha = 1 + \frac{1}{\beta} \tag{9}$$

A comparison of the experimental results with theoretical calculations based on the proposed model is presented in Fig. 1-8.

The theoretical curves (solid lines) in these figures were obtained based on the model described above. It is seen that there is good agreement between the experimental and theoretical data.

The control parameters defining the theoretical curve were determined on the basis of experimental data using the least squares method. The obtained numerical values, using which theoretical curves are plotted, shown in Fig. 1 to 8 are summarized in Table 1.

It can be seen that analogues of the Dunbar number obtained on the basis of the analysis of the presented dependencies fit into the range (100-230) indicated in the literature [22-25]. It is also seen that the analogue of the Dunbar number for Moscow is noticeably higher than for Almaty and Novosibirsk, which could be expected from general considerations.

This means that the analogue of the Dunbar number included in the dependencies used as a parameter can be considered as a characteristic of the communication space for each specific city.

The most significant result is that the fractal dimension of the communication space is the same for all eight cities, and the value is 2/3 with high accuracy. It can be assumed that this value reflects some fundamental features of the formation of the communication space.

Table 1. Empirically determined constant formulas (8) for various settlements and the population of cities.

City	A	C	a	β	Population, thousand people
Almaty	820	5500	95	0,654	1806,6
Moscow	1290	43000	300	0,657	12506,5
Novosibirsk	1120	14000	95	0,651	1612,8
Kiev	1100	20000	130	0,656	2934,5
Pavlodar	800	3000	110	0.648	334,9
Yekaterinburg	800	9800	120	0,.649	1468,8
St. Petersburg	716	9350	205	0.65	5351,9
Tashkent	200	1500	140	0.654	2424,1

It should be noted that the obtained values of the Dunbar number (considered as parameters of the theoretical dependence) correlate with the population of the city (Fig. 9).

This figure shows that the Dunbar number is not an absolute indicator, but significantly depends on the characteristics of the communication environment, mainly the population of the city.

This circumstance cannot be surprising - the larger the city, the greater the number of communications a particular person is involved. He, on average, is better informed, more sociable, etc. Of course, these factors are not the only ones, but we can speak with all certainty about the influence of the nature of the communication environment on the value of the Dunbar number.

3 Conclusion

In conclusion, we note that based on heuristic considerations and literature data on the nature of the Dunbar number, we can propose a simple semi-empirical model that adequately describes the experimental distributions of users of social online networks by the number of friends.

This model is based on the assumption of the fractal nature of the communication space. This assumption seems to be justified also because the fractal dimension obtained for various cities in which the Runet is actively used remains constant with high accuracy

REFERENCES

[1].   Chen W., Wang C., Wang Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM. P. 1029-1038.

[2].   Doyle S. (2007). The role of social networks in marketing. Journal of Database Marketing & Customer Strategy Management. T. 15. №. 1. P. 60-64.

[3].   Stelzner M. (2014). 2014 Social Media Marketing Industry Report. Social media examiner. P. 1-52.

[4].   Chaffey D. (2016). Global social media research summary 2016. Smart Insights: Social Media Marketing.

[5].   Ilyashenko S.B. (2016). Social networks internet as a tool to improve e-commerce. In Proceedings of the International Scientific and Practical Conference. P. 105-107.

[6].   Hunter E., Pernik P. (2015). The challenges of hybrid warfare. International Centre for Defence and Security, 2015.

[7].   Klausen J. (2015). Tweeting the Jihad: Social media networks of Western foreign fighters in Syria and Iraq. Studies in Conflict & Terrorism. T. 38. №. 1. P. 1-22.

[8].   Kamalipour Y.R., Friedrichsen M. (2017). Introduction: Digital Transformation in a Global World. Digital Transformation in Journalism and News Media. Springer, Cham. C. 1-4.

[9].   Sergeev I.V. (2015). Social networks on the Internet as a means of implementing operations of information and psychological warfare. International Scientific Research Journal. №. 9(40) Part 4. P. 101-104.

[10].   Chernova, T.A., Slepovronskaya, K.Yu. (2015). Historical memory in the information war. How are social networks used in the ideological confrontation between Ukraine and Russia. Philosophical Sciences. №. 5. P. 16-23.

[11].   Zheltukhina M.R., Pavlov P.V. (2016). Social network "Facebook" in the XXI century: from the communication tool to the information war tool. Philology. Theory and practice. №. 7-3. P. 89-93.

[12].   De-Marcos L. et al (2014). An empirical study comparing gamification and social networking on e-learning. Computers & Education. T. 75. P. 82-91.

[13].   De-Marcos L. et al (2016). Social network analysis of a gamified e-learning course: Small-world phenomenon and network metrics as predictors of academic performance. Computers in Human Behavior. T. 60. P. 312-321.

[14].   De Meo P. et al (2017). Combining trust and skills evaluation to form e-Learning classes in online social networks. Information Sciences. T. 405. P. 107-122.

[15].   Bukayeva A.A., Magzumova A.T. (2015). The use of social networks in the educational process. Innovations in Science. №42 p. 120-126.

[16]. Kalmykova S.V., Krasnoshchekov A.V. (2017). Approaches to e-learning organization through the social networks. Modern education: content, technology, quality. T. 1. P. 195-198.

[17]. Makhmutova M.V., Makhmutov G.R. (2015). Models and platforms for the implementation of mass open online courses. Modern information technologies and IT education. T. 1. №. 11. P. 486-496.

[18]. Kurzaeva L.V., Grigoriev A.D. (2015). Mass open online courses: essence, specific characteristics. New information technologies in education. P. 250-253.

[19]. Bremer C., Weiß D. (2013). How to analyze participation in a (C) MOOC?. Univ.-Bibliothek Frankfurt am Main. P. 992-1002.

[20]. Khalil M., Brunner H., Ebner M. (2015). Evaluation grid for xMOOCs. International Journal of Emerging Technologies in Learning (iJET). T. 10. №. 4. P. 280-314.

[21]. Colas J.F., Sloep P.B., Garreta-Domingo M. (2016). The effect of multilingual facilitation on active participation in MOOCs. The International Review of Research in Open and Distributed Learning. T. 17. №. 4. P. 4-11.

[22]. Dunbar R.I.M. (1993). Coevolution of neocortical size, group size and language in humans. Behavioral and brain sciences. T. 16. №. 4. P. 681-694.

[23]. Dunbar R. (1998). Grooming, gossip, and the evolution of language. Harvard University Press.

[24]. Dunbar R.I.M. (2009). The social brain hypothesis and its implications for social evolution. Annals of human biology. T. 36. №. 5. P. 562-572.

[25]. Shultz S., Dunbar R.I.M. (2012). The social brain hypothesis: an evolutionary perspective on the neurobiology of social behavior. I Know What You Are Thinking: Brain Imaging and Mental Privacy. P. 13-28.

[26]. Gilyazova A.A. (2014). On the estimation of the number of Dunbar and its influence on the interaction of users in the social network Facebook. In XII All-Russia Conference on Control Problems ARCCP-2014. P. 6292-6300.

[27]. Chesnokov V.O. (2017). The use of the algorithm for isolating communities in the information confrontation in social networks. Cybersecurity issues. №. 1 (19). P. 37-44.

[28]. Katasev A.S., Kirpichnikov A.P., Ramazanova R.I. (2015). Method of analyzing the security of social network accounts from malicious content. Bulletin of Kazan Technological University. T. 18. №. 18. P.195-198.

[29]. Rzayeva M.E. (2015). Evaluation of the effectiveness of special projects in Russian online media through the determination of the price of contact. The media sphere: problems and points of development.