JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

3/2015

(Contents Continued on Back Cover)

ISSN 1509-4553 on-line: ISSN 1899-8852

Circulation: 300 copies

Sowa – Druk na życzenie, www.sowadruk.pl, tel. 22 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Over the last three decades, there have been a number of significant technological developments in intelligent network technologies. The most important step in the development of intelligent networks was the agreement of the Regional Bell Operating Companies (RBOCs) and Bellcore to work on the advanced intelligent network (AIN) project.

Today's ultra-scale intelligent networks (modern clusters, cognitive networks, clouds and mobile clouds, federated clusters and modern wireless technologies) enable the aggregation and transmission of huge amount of data generated by geographically-distributed users. In the Big Data era, such data can be incomplete and the whole process of data analytics is dynamic and complex.

The concept of the support the data intensive analytics process by advanced intelligent networks and various scale today's network technologies makes a positive impact on the development of new efficient data, information and network systems. This issue contains ten research papers reporting the recent results on models, solutions, and techniques from such a wide research area, ranging from conceptual and theoretical developments to advanced technologies and innovative applications and tools.

In the first four papers the authors present interesting examples of efficient energy-effective and data resilient approaches in modern intelligent networks. Esmaeilifard and Rahbar developed a novel energy-efficient data transmission model in cellular networks. They considered a realistic scenario in a country area and defined the high capacity energy efficient protocol for the optimization of required output power for network traffic and data transmission between the base station and a mobile user. Although the traffic and number of users can be rather low in such environments, the network access problem and big distance from the base station are the main reasons of high energy consumption in the network and low mobile resource reliability (the batteries can drain very quickly). The presented approach is a simple example of the power management support in wide area cellular networks.

A rapid grow of various public utility services based on the new developments in ICT sector can be observed over the last decades. Intelligent network technologies play a crucial role in supporting the management, monitoring and control of such services. However, the ultra-scale of those networks and a huge volume of different service offered for the user make the service management and control problems very complex and challenging. Therefore, there is a need to make those processes independent from the system and network administrators. In the second paper, Nayaka and Biradar survey the recent network technologies useful in automation of the management and control of water, electricity and gas supply processes. They compare the efficiencies of networks set on the provide a comprehensive network analysis based on communication protocols, topologies, network, hardware, and applications used for automated management of public utility services. As the conclusion from the provided analysis, the authors propose a prototype of a new wireless sensor network topology, where the sensor nodes on two data fusion and aggregation layers are its main components.

The automation of data transmission and traffic in modern networks is an important example of energy optimization method in network management process. Kutin in the third paper addresses the problem of energy-aware data transmission using the wavelets and signal decomposition methods. The author analyzes the signals properties generated by classical modulation methods and spline modulation and defines a novel wavelet coefficients transmission model. The effective reduction of energy used for data transmission and, more general, data migration in heterogeneous large-scale networks such as grids and Infrastructure-as-a-Service (IaaS) layer in cloud systems, remains one of the most important issue in data and tasks allocation processes in such environments. Konovalov and Razumchik try to solve resource allocation problem in a simple 2-servers scenario, where the considered jobs have fixed sizes and the objective is the minimization of mean job sojourn time in the system. The optimal allocation policy in this system is of threshold type with one threshold i.e. if upon arrival of the job the amount of unfinished work at faster server (plus total work in its queue) minus the amount of unfinished work at slower server (plus total work in its queue) exceed the threshold value, job is allocated to slow server. The authors propose a scalable fast iterative non-simulation algorithm for the approximation of the policy parameter (threshold), which has been verified in comprehensive experimental analysis.

Many processes in modern ultra-scale and small-scale networks, as well as the prediction of network reliability and performance, can be defined and estimated by using well-known probabilistic models. In the next four papers, few interesting examples of the application of mathematical models can be found.

Islam *et al.* focus on cognitive radio network with two types of users: primary user (PU) and secondary user (SU). They used 2-dimensional Markov Chain model for traffic analytics and prediction. In the next paper, Krok presents the simulation and prediction of the time series. The author compares the efficiency of two types of Artificial Neural Networks trained by Kalman filter support method and Bayesian-based method. A simple theoretical study of the prediction model and efficiency of both training methods is provided.

Although the OpenStack standard seems to be well explored and exploited in cloud computing, Grzonka analyzed in details the features and performance of the OpenStack cloud in the case of complex data processing and analytics. The prediction of the system performance and cloud resource reliability remain still challenging tasks, mainly because of complex memory management frameworks in OpenStack libraries.

The paper of Grzonka shows indirectly the limitation of today's network systems in the case of massive and Big Data processing. One of the main reasons of low network reliability is, besides the data volume, the data dynamics. Therefore, there is a need of development of simple and scalable data distribution model with the dynamic factor. Krok in the eight paper, analyze Gaussian distribution model for defining the experimental data in complex global optimization tasks. The author used genetic algorithms for estimation of the Gaussian process parameters. Gaussian-based data modeling accuracy was compared with neural networks model learned by Kalman filter method. The results of the performed experiments shows that Gaussian processes can be efficient in simulation and prediction of concrete hysteresis loops in cycling experimental data loading.

The last part of the issue includes two papers with interesting practical example of application large scale networks in e-commerce (Suchacka *et al.*) and historical analysis of analog telephone systems (Patil).

I believe that all papers presented in this issue ought to serve as a reference for students, researchers, and practitioners interested or currently working in the evolving and interdisciplinary areas of energy-aware intelligent networks of various types and scale, where traffic management and data processing and analytics are the crucial issues. I hope that the readers will find there new inspirations for their further research.

I am grateful to all the contributors of this issue. I thank the authors for their time and efforts in the presentation of their recent research results. I also would like to express my sincere thanks to the reviewers, who have helped us to ensure the quality of this publication. My special thanks go to the journal Editors, Managers and Publishers for their great support throughout the entire publication process.

Joanna Kołodziej
Department of Computer Science
Cracow University of Technology, Cracow, Poland
E-mail: jokolodziej@pk.edu.pl
URL: http://www.joannakolodziej.org

Paper

A High Capacity Energy Efficient Approach for Traffic Transmission in Cellular Networks

Maryam Esmaeilifard and Akbar Ghaffarpour Rahbar

Computer Networks Research Lab, Electrical Engineering Technologies Research Center, Sahand University of Technology, Tabriz, Iran

Abstract—The efficiency of cellular networks can be improved in various aspects such as energy consumption, network capacity and interference between neighboring cells. This paper proposes a high capacity energy efficient scheme (HCEE) for data transmission in cellular networks in a country area. In this paper, the authors obtain a new equation to characterize the minimal required output power for traffic transmission between a base station (BS) and a mobile user (MU) based on the MU distance from the BS. Also, the cells boundaries (the boundary of overlapping areas of neighboring cells) by two static and dynamic approaches are specified. This work helps for better frequency allocation to MUs and allows increasing network capacity. In this paper, the analytical modeling in order to formulate the HCEE algorithm and evaluate its performance is used. The performance evaluation results show the simplicity of the HCEE algorithm and its effect on energy consumption decline, network capacity enhancement and the interference reduction.

Keywords—cellular networks, energy efficiency, frequency assignment, interference avoidance, transmission power control.

1. Introduction

In the last decade, the number of subscribers and the traffic in cellular networks has significantly increased. The mobile traffic volume is expected to be almost 26 times in 2015 rather than in 2010. Such a growth in cellular industry has pushed the limits of capacity and energy consumption in mobile networks because mobile equipments are fully working in all days of a year [1].

Information and communication technology (ICT) represents around 30% of total energy consumption in the world. Around 57% of this value is related to MUs, mobile devices and wireless networks. The Global e-Sustainability Initiative (GeSI) research estimates an increase of 72% in ICT energy consumption in 2020. In addition, around 2 to 2.5% of total carbon emissions are related to the ICT industry and it is expected to be nearly doubled by 2020 [2]. Such a value of emission has exceeded the CO_2 output of the entire aviation industry [3]. Moreover, a significant portion of network operator's costs is dependent on energy costs. Therefore, the operators of mobile networks try to provide the capacity and required coverage for users, and to reduce the consumed energy of these networks.

By designing energy aware components in BSs and energy aware network deployment strategies, the idle capacity of BSs can be minimized, and the energy wastage will be reduced [4].

There are many strategies designed for reducing the energy consumption of various elements of mobile networks. A known strategy is the sleeping technique in idle times. The various systems may exploit the idleness at various levels. For example, the Catnap system introduced in [5] saves energy by combining the small gaps between packets by delaying the transmission of them and converting these gaps into significant sleep intervals. In other words, the user's mobile terminal can be switched to the low power consuming states in these sleep intervals. The key design component of the Catnap system is a new scheduler with the goal of increasing the sleep interval for a certain transmission without great effect on the total transmission time. The other method introduced in [6] is based on this reality that the coverage quality of cellular networks is not the same everywhere. In some areas, a signal is strong and in others it is weak. The signal strength has a direct effect on the energy consumption because of two major reasons: increasing the energy consumption and reducing the data rate in unit of time when the signal is weak. Therefore, it is possible to employ the variation of the signal strength for energy saving. To accomplish this, the future signal strength must be predicted based on the location and history. Then, some algorithms use the predicted signal strength to efficiently schedule communications. Therefore, energy can be saved by deferring the communications until a mobile device moves to the area with better signal strength or conversely by prefetching data before the signal decreases.

As discussed in [7], new mobile devices have several radio interfaces (such as Wi-Fi and 3G) for data transmission. Since these interfaces have different features (i.e., the nominal data rate and the achievable data rate are different in these radio interfaces), the consumed energy can be different for transmitting the same amount of data. In addition, the availability of these types of networks is different. The coverage of cellular network is more than Wi-Fi. In general term, transmitting big data units via a Wi-Fi network can achieve more energy efficiency rather than a cellular network, whereas Wi-Fi may be not available every time. Since some applications (such as video capture application on a smartphone that can capture video clips and then automatically upload them to a server) are delay tolerant, therefore it is possible to defer data transmission until a low energy Wi-Fi connection becomes available. Hence, the major point is designing an algorithm for establishing a tradeoff between energy and delay.

The cell zooming technique [8] relies on turning equipments off in idle times. In a cellular network, the cell size is usually defined as the area in which users can receive control signals from a BS. Hence, cell size and capacity are fixed based on the estimation of peak traffic load. Because of users' movements and bursty nature of many applications, network traffic load has temporary fluctuations. If the size and the capacity of each cell are designed based on the maximum traffic load, some of the cells are usually under low traffic load and some others are under heavy traffic load. In this case, the static cell design and deployment for traffic load fluctuations is not optimal. Under cell zooming, the cell size is continuously adjusted according to the traffic load conditions. In this technique, traffic load is concentrated in few cells in order to minimize the number of active cells in the network. In other words, the cells with low traffic load are set to the sleep mode. In this case, neighboring cells are either spread in order to ensure full network coverage or they can cooperate in order to serve the users within the inactive cells.

The switch off scheme is another method for solving the problem of energy aware management in access cellular networks [3]. This method tries to specify the amount of energy that can be saved by reducing the number of active cells in the network when traffic load is low. The reduction of the traffic in a cellular network is due to the combination of two elements: the day-night behavior of users and the daily movement of users between the residential areas and the office regions. Thereupon, the need for capacity is high during the peak time in both areas, but this need is reduced during the period in which those areas are light-populated (e.g., residential areas during day period and office regions during night period). Therefore, some cells can be switched off and radio coverage can be guaranteed by the active cells, with the small increase in the transmission power. Indeed, the main goal in this method is to optimize the saving energy. In other words, it is important to find a suitable number of cells in order to switch them off. Although turning off many cells can lead to significant savings, the traffic pattern shows that this case is possible only for a short period of time. Therefore, turning off a few cells during long period times can cause more energy saving.

Since the current mobile networks are not very energy efficient, in [9] and [10], the mechanisms to improve the efficiency of BSs have been proposed. In [9], a method for dynamic management of BSs has been explored in order to understand the scope for energy saving. The main objective of [10] is to create a suitable radio structure that can save power by reducing the power consumption of various elements of BS such as radio transceivers, power amplifiers, and transmit antennas.

It is difficult for network operators to maintain the capacity growth, utilize bandwidth, decrease delay, and limit the energy consumption at the same time. Therefore, the framework for Green Radio – a wireless architecture in [11] consists of four fundamental efficiency tradeoffs:

– deployment-energy tradeoff to balance the deployment cost, throughput and energy consumption in the whole network,

– spectrum-energy tradeoff in order to balance the accessible rate and the energy consumption of the system,

– bandwidth-power tradeoff to balance the bandwidth utilized and the required power for the transmission,

– delay-power tradeoff to balance the average end-to-end delay and the average consumed power in the transmission.

By using these tradeoffs in different research aspects, such as network planning, resource management and physical layer transmission design, the performance parameters of the network such as power consumption, delay and so on can be achieved simultaneously.

A link adaptive transmission scheme has been proposed in [12] that improves the energy efficiency by adapting both transmission power according to the channel states and the consumed power. Relaying [13] is another way to improve the energy saving in wireless networks. By choosing some nodes as relay nodes, more connections can be established between a source-destination pair. Therefore, data can be delivered through several links. It is clear that one of these links is the shortest path and hence, the required time to transmit a fixed amount of data and thereupon the consumed energy is reduced.

Turning off some elements of data centers such as CPUs, disks and memories when they are in idle state is another effective method for avoiding from unnecessary energy consumption [14]. In addition, cooling of data centers improves the energy efficiency. It is presented in [15] that separating hot air and cold air plays the main role in the cooling energy efficiency. This efficiency is not achievable without designing a HVAC system. It is measured that around 50% of the energy consumption in BSs is related to power amplifiers (PA) [16]. It is presented in [17] that to achieve PA efficiency and energy efficiency in whole network, the BS structure must be changed. In a technique called envelope tracking, the PA power voltage is changed dynamically with ensuring that the output power of transistors remains in a suitable level. This technique achieves high energy efficiency.

It is predicted that mobile applications and services will generate 26 times more traffic load in 2015 than that in 2010. Therefore, they must be able to work based on dynamic users' demands and wireless links. The work in [18] introduces an adaptive approach to reduce energy consumption of multimedia transmissions that it works based on the selecting proper source compression and coding. A method for designing energy efficient applications has been introduced in [19] that works based on the prediction of application activities by learning historical patterns. This approach dynamically limits and adjusts low-layer functions of mobile devices for saving energy.

Another strategy for reducing energy consumption is the control of radio equipment transmission power by focusing on the lovering of transmission power of BSs and MUs. Since interference between cells causes unnecessary energy loss, the work in [20] considers a two layer cellular network in which one frequency channel is used in both layers. The interference between macro cells and femto cells is controlled, and the energy consumption declines by adjusting the transmission power via self-configuration and self-optimization techniques. In self-configuration, a femto cell BS measures the average received pilot powers from neighboring BSs. Then, the femto cell BS adjusts its transmission power based on the strongest pilot power to achieve initial cell coverage. Next, the femto cell BS performs a self-optimization and adjusts the transmission power continuously so that the femto cell coverage does not leak to the outdoor areas while the indoor area coverage of femto cell is provided.

The previous works have some restrictions. For example, the research from [5], [6] and [7] are based on deferring information transmission, thus increasing transmission delay. The paper [6] is applicable only for special applications because all applications do not tend to defer their information transmission. The proposed algorithms in [8] and [14] increase blocking probability of users because it is likely that sufficient bandwidth does not exist when a new user enters to the network and therefore the user is blocked. In [20], femto cells are used for increasing the network capacity, but at the expense of increasing the network equipment costs. Therefore, it is so important to find a method without these restrictions.

The objective of this paper is to design a new method for energy consumption reduction, increasing the capacity, and interference avoidance in cellular networks. The HCEE algorithm is proposed for data transmission in cellular networks in a country area that adjusts the amount of transmission power based on a specific distance from BS in order to reduce the energy consumption and interference probability.

The author's contribution in this paper is to develop a new transmission technique that adjusts the amount of transmission power between a BS and MUs. In addition, the boundaries of neighboring cells overlapping area are specified in order to use all available frequencies for serving those users located outside of the overlapping areas.

2. Network Model

In this paper, a cellular network in a country area is considered, where houses are mostly flat and usually far from each other. This network consists of N cells, each with radius R. Each cell is surrounded by another m overlapping cells. In cellular networks, the used frequencies in each cell must be different from its neighboring cells because of interference. In order to coordinate the assigned frequencies, the cells are clustered. Inside a cluster with C neighboring cells, each cell must have different frequencies from the other cells within the cluster. Let the total number of network frequencies be F. With clustering, each cell can only use the $1/C$ of these frequencies, i.e., the number of assigned frequencies to each cell is $\lfloor F/C \rfloor$.
Figure 1 shows the structure of the network model under study, where the cells are modeled as hexagon. However, in reality, they are overlapping cells in which the interference of frequencies may occur. In this figure, the green (gray) cells show an instance cluster with three cells (the least number of cells for building a cluster), where different frequencies are used in each cell. A BS with h directional antennas is located in the center of each cell. Using directional antenna can lead to reduce transmission power and interference. Therefore, the energy efficiency can be improved [21].

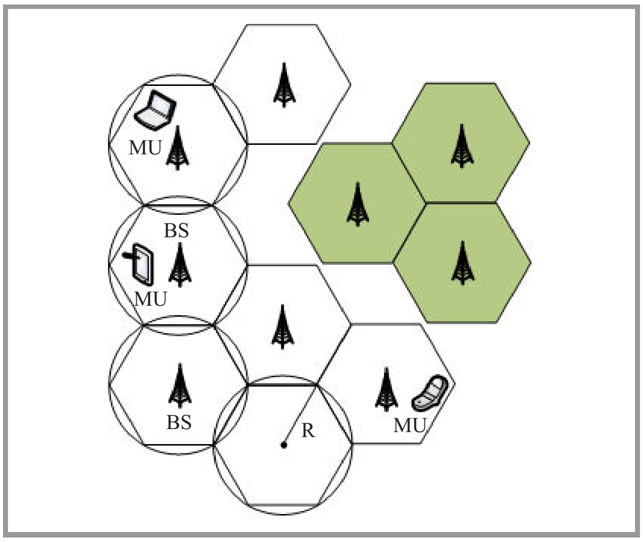


Fig. 1. The network model.

Let the maximum number of active users in each cell be N_a, where each user is equipped with an omnidirectional antenna and located in a random place within the cell.

3. The HCEE Algorithm

The interference phenomenon between neighboring cells is an important issue in a cellular network that not only reduces the network throughput, but also causes the unnecessary energy consumption. The used frequencies in each cell should be different from the other neighboring cells.

The amount of transmission power between the BS and an MU is proportional to the BS range specification and the amount of energy consumption. In other words, whatever transmission power is reduced, the less energy is consumed. Also whereas the transmitted signal cannot leak in the areas far from the location of the MU, the interference phenomenon can be avoided. The minimum amount of power for traffic transmission between the BS and MU_i is given by

$$P_t(\mathrm{dBm}) = S_r(\mathrm{dBm}) + L_p^i(\mathrm{dB})\,, \tag{1}$$

where P_t is the BS transmission power (in dBm) and S_r is the receiver sensitivity of MU_i. Parameter L_P^i is the amount of path loss which is dependent on various factors such as the distance between MU_i and the BS.

Note that in rural or countryside areas, the buildings are low and they are far from each other, therefore the effect of these snags on the transmitted signal power downfall is low. Thus, it is assumed in Eq. (1) that the transmission power downfall occurs only because of path loss.

The amount of path loss is obtained from Eq. (2) known as the "Friis transmission equation":

$$L_p^i(\mathrm{dB}) = 10\log\frac{(4\pi d_i)^2}{G_t G_r \lambda^2}\,, \tag{2}$$

where d_i is the distance between MU_i and BS, G_t is the antenna gain of transmitter and G_r refers to the receiver antenna gain. The antenna gain refers to the directionality of an antenna given by:

$$G = \frac{4\pi A_e}{\lambda^2}\,, \tag{3}$$

where A_e refers to the effective area of an antenna dependent on its shape.

Since P_t in Eq. (1) is the minimum power required between the BS and MU_i, P_t is different for each user. Reducing the amount of P_t could have two advantages:

- The network energy consumption is reduced;
- Since the transmission power for the users located in non-overlapping areas does not leak to other areas, it is possible to use all F available frequencies in a cellular network to service the users. In other words, in the outside of overlapping areas, all cells can equally use all F frequencies because the interference does not occur in these areas, thus enhancing the network capacity.

To achieve the latter advantage, the boundary of overlapping areas of cells must be specified (see Fig. 2). In other words, the goal is to find the value of r (where $r < R$) so that all frequencies can be used for the users located within the circle with radius r. For serving the users located in distances from the BS in between $(R - r)$ and R, different frequencies must be used. In this paper, it is considered that all users exactly located on the boundary of the overlapping area are as those users that located inside the overlapping area in order to avoid the interference.

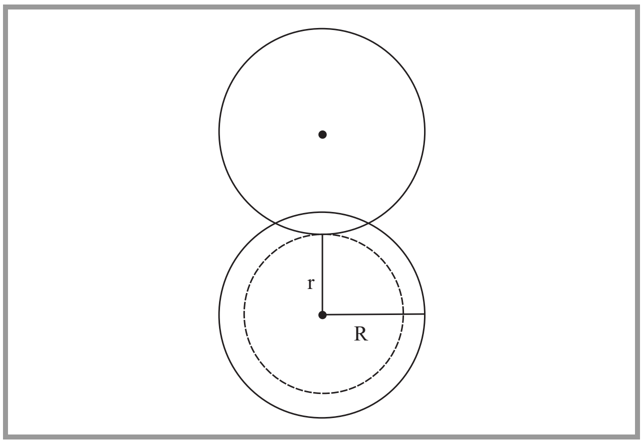


Fig. 2. Specification of radius r.

Using different values of P_t for each user and updating them when a user moves is difficult and also consumes more energy. Therefore, in this paper, the amount of transmission power is equal for all the users allocated within the circle with radius r. This amount of transmission power is obtained from Eqs. (1) and (2), where the value of d_i in Eq. (2) for all users in this area is equal to r. On the other hand, for all users allocated in the overlapping area, the value of d_i is equal to R.

To find r, the two methods as static and dynamic in the following are proposed.

3.1. Static Method

In this method, a constant value for r is obtained so that interference avoidance can be guaranteed. According to Fig. 3:

$$r + 2\alpha = R\,. \tag{4}$$

On the other side, considering the triangle ABC and trigonometric formulas:

$$r + \alpha = R\cos 30^\circ\,. \tag{5}$$

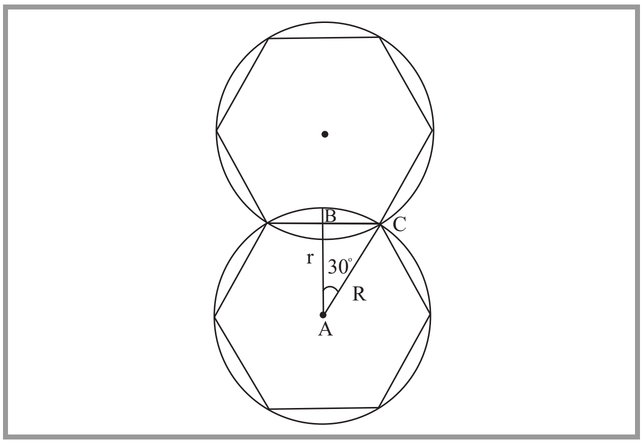


Fig. 3. Static method for finding r.

With solving Eqs. (4) and (5) in an equation system, one can find:

$$r = 0.732R. \quad (6)$$

$$\alpha = 0.134R. \quad (7)$$

3.2. Dynamic Method

In this method, the value of r within cell k is updated based on the received power from neighboring BSs at MUs located within cell k. Indeed, each MU within cell k periodically informs its base station (BS_k) about the amount of received power from neighboring BSs. Then, BS_k can choose the smallest of the received powers as the maximum range of neighboring BS, i.e., the boundary of overlapping area with cell k.

Define an interval with length τ for periodic updates. Let the amount of received power from a neighbor BS at MU_i within cell k at time T in interval τ be $P_{i,T}$. During this interval, MU_i informs its BS_k about the amount of $P_{i,T}$. Note that the amount of received power at MU_i can be variable within the interval because of the MU movement. This is because, for each MU, the average received power must be calculated within interval τ. The BS_k calculates the average value of $\overline{P}_i$ among $P_{i,T}$ values for MU_i in interval τ by:

$$\overline{P}_i = \frac{\sum_{T=1}^{\tau} P_{i,T}}{\tau}. \quad (8)$$

Define U_{NC} to be the set of MUs within cell k that receive power from neighboring BSs during interval τ. Clearly, there could be some MUs that hear nothing from neighboring BSs. Note that the amount of $\overline{P}_i$ for these users is unequal to zero.

Now BS_k chooses the minimum of $\overline{P}_i$ values among all MUs in U_{NC}:

$$P_{\min} = Min\{\overline{P}_i | i \in \{U_{NC}\}\}. \quad (9)$$

Let MU_j has the smallest power, i.e. $P_{\min} = \overline{P}_j$. Then, r is equal to the distance of MU_j from BS_k, i.e.

$$r = \{d_j | \overline{P}_j = P_{\min}\}, \quad (10)$$

where d_j is the distance between MU_j and BS_k.

Since the value of r in this method is periodically updated based on the actual received power from neighboring BSs, the dynamic method may work better than the static method in terms of network throughput.

4. Performance Evaluation

In this section, the performance of HCEE algorithm is evaluated by proprietary program written in C++. All calculations are performed for cell k surrounded by six neighboring cells, where users are randomly located within each cell. Then, the locations of users are randomly changed within each cell at different time snapshots, where each time snapshot is called iteration in the following scenarios. The simulation results show that the percentage of allocated users within the circle with radius r under the static method are between 78.9% and 81.08% with 95% confidence interval, regardless of the number of active users located inside cell k or at radius of cell k (see Fig. 4). In Fig. 4, the horizontal axis shows the number of iterations of the HCEE algorithm during time. The allocated users within the circle with radius of r under the dynamic method are between 77.19–90.46% of total users with 95% confidence interval at various times because of path loss variation that depends on the channel state and weather conditions. Therefore, the value of r and percentage of allocated users within the circle with radius r are different at various times. Figure 5 shows the number of allocated users within the circle with radius r under the dynamic method and Fig. 6 shows the variation of r at various times in both static and dynamic methods for $R = 5000$ m. According to Fig. 6, the value

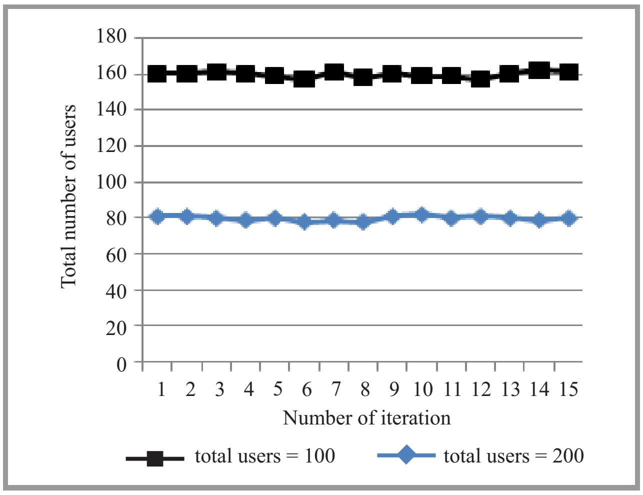


Fig. 4. Number of allocated users within the circle with radius r under the static method.

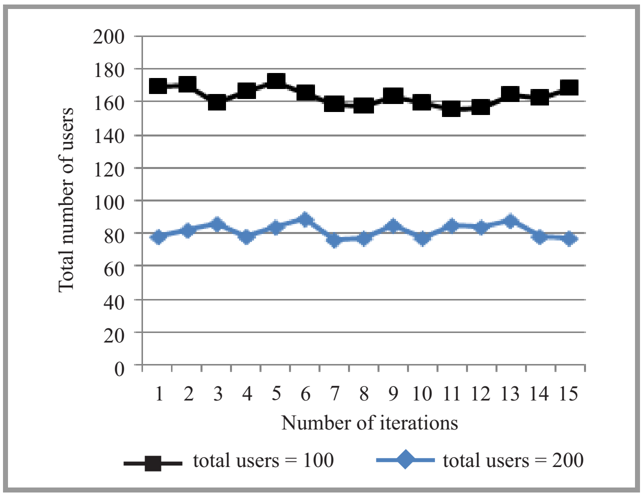


Fig. 5. Number of allocated users within the circle with radius r under the dynamic method.

of r is constant in the static method, but variable under the dynamic method because the value of r in the static method is obtained from Eq. (6) that only depends on R, but in the dynamic method the value of r is obtained based on the received control signal power from neighboring BSs at MUs. Since the amount of received control signal power depends on the weather conditions and path loss, it will be different at various times.

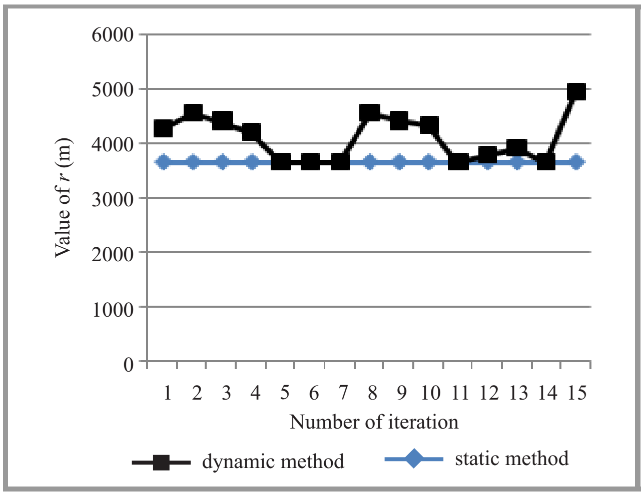


Fig. 6. The value of r in static and dynamic methods.

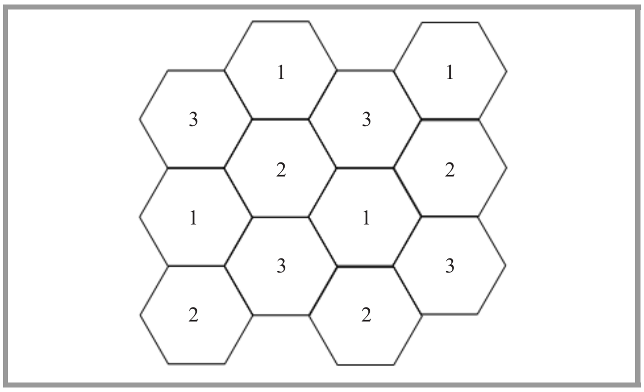


Fig. 7. The network clustering and frequency assignment.

Figure 7 shows the network clustering, so that the same frequencies are assigned to the cells with the same number. According to the Fig. 7 and under conventional clustering, if the cells within ternary clusters is clustered, the number of assigned frequencies to the users of cell k becomes only $F/3$. On the other hand, using HCEE, the number of assigned frequencies to the users within cell k will become F. In other words, the network capacity will be increased three-fold under the static method. However, under the dynamic method, the network capacity will be increased around three-fold up to four-fold.

As stated in Sections 2 and 3, the amount of energy consumption in the network can be reduced by replacing omnidirectional antennas with directional versions and adjusting the amount of transmission power of BSs, i.e., by replacing the conventional method with the HCEE algorithm. Define the level of power saving as:

$$Power\ saving = 1 - \frac{E_x}{E_y}, \tag{11}$$

where E_x is efficient power consumption and E_y is inefficient power consumption. The inefficient power consumption is the consumed power under conventional network structure without using the HCEE algorithm, while the efficient power consumption is defined as the consumed power under using the HCEE algorithm, i.e., using directional antennas and adjusting transmission power using Eq. (1).

Figures 8 and 9 show the effect of replacing omnidirectional versions with directional antennas and adjusting the transmission power, i.e., using the HCEE algorithm, on the energy consumption under dynamic and static methods, respectively, for various radiuses for cell k. This value of power is calculated when maximum number of active users is $N_a = 100$.

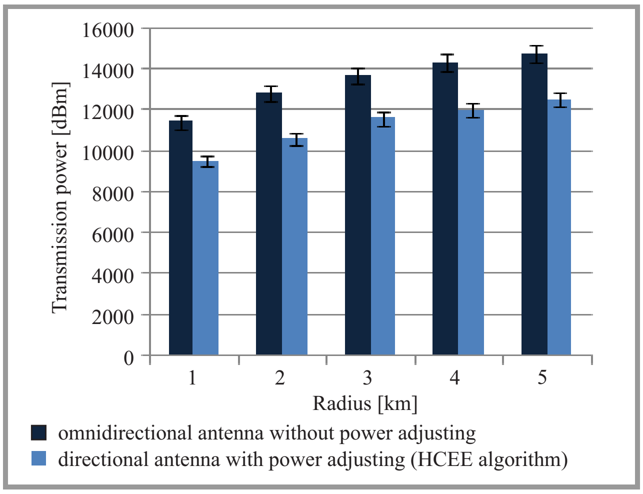


Fig. 8. The effect of using HCEE algorithm on the power consumption under dynamic method.

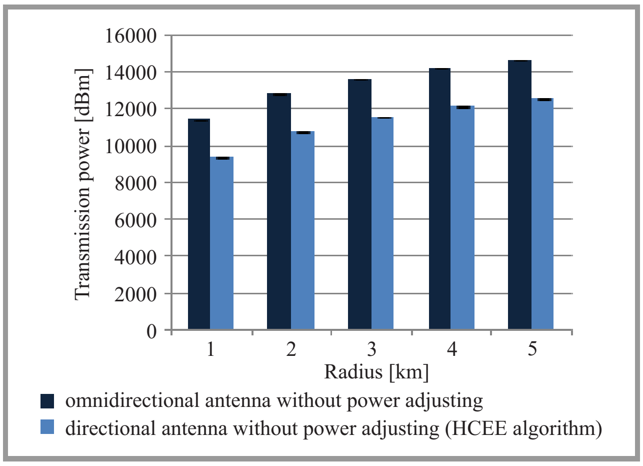


Fig. 9. The effect of using HCEE algorithm on the power consumption under static method.

As it is shown in Figs. 8 and 9, the amount of transmission power goes up with increasing the radius of cell because the transmission power has a direct correlation with the square of value of r that increases with increasing the cell size.

As shown in Figs. 8 and 9 (the inefficient power consumption and efficient power consumption under the HCEE algorithm) and according to Eq. (11), the amount of power saving is around 0.2, i.e., the consumed energy in cell k can be reduced by almost 20% using the HCEE algorithm under both static and dynamic methods rather than the conventional method. Although the amount of transmission power under the dynamic method is different at various iterations because of difference in value of r, but the average of this amount is almost equal to the static method.

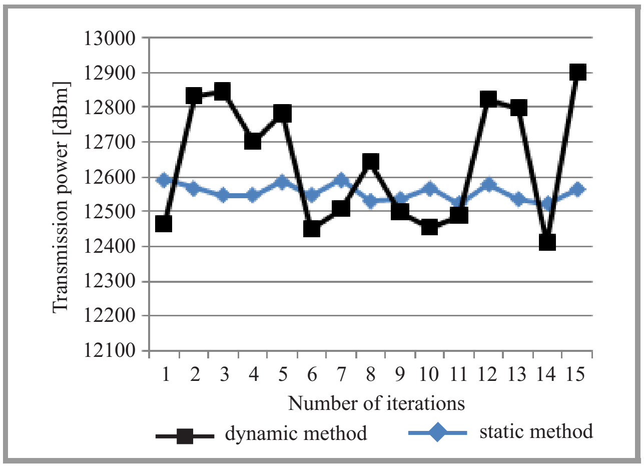


Fig. 10. The variation of transmission power in static and dynamic methods.

In Fig. 10, the variation of transmission power in both static and dynamic methods at various iterations is shown at $R = 5000$ m. As stated earlier, the horizontal axis shows the number of runs of the algorithm.

Figures 11 and 12 show the number of interfered users within cell k at $N_a = 100$ under both dynamic and static methods, respectively. As shown in these diagrams, the interference is reduced by using the HCEE algorithm because in conventional method the transmission power is adjusted to cover the whole area of a cell even for those users allocated near the BS. However in HCEE, for those users located within the circle with radius r, the amount of transmission power is adjusted to only cover the area of this circle. The difference between the number of interfered users in static and dynamic methods is because of the difference between the calculated values for r, i.e., the boundary of cells, in these methods. With the increase in cell size, the number of interfered users is reduced because the same numbers of users, i.e., N_a, are distributed in the larger space and with more distance from each other.

According to the obtained simulation results, the performance of the dynamic method in capacity improvement is better than the static method, but for other performance parameters, such as the amount of transmission power, the dynamic method is better in some cases than the static method and in some cases is worse than the static method. However, when we consider a time interval and calculate the average of the performance metrics, the results are almost equal in the both static and dynamic methods. Furthermore, the performance of HCEE in terms of energy consumption decline, capacity improvement, and interference avoidance is much better than the conventional method.

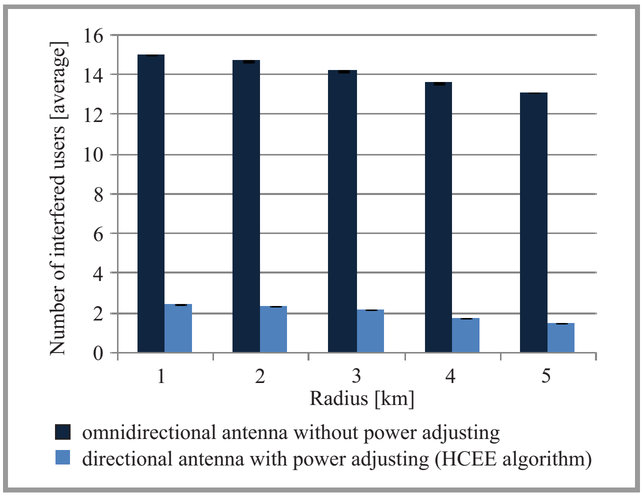


Fig. 11. The number of interfered users under the dynamic method.

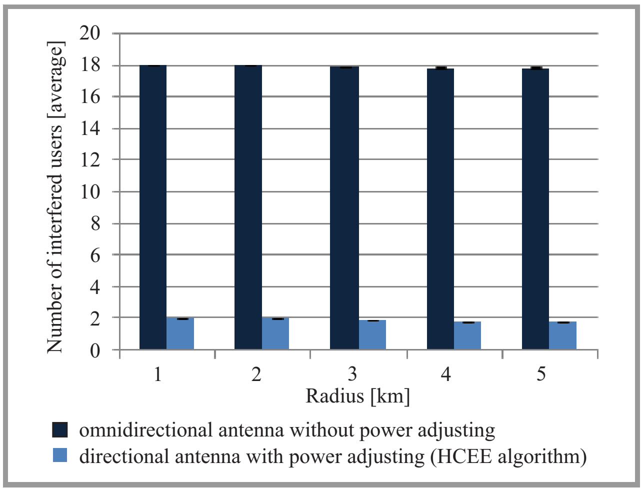


Fig. 12. The number of interfered users under the static method.

5. Conclusion

This paper has proposed a novel efficient scheme for traffic transmission, i.e., the HCEE algorithm that is based on adjusting transmission power and calculating radius r in order to increase the network capacity. An analytic formulation has been provided for the operation of HCEE. Our performance evaluation results show that the HCEE can increase the energy efficiency and capacity of the network compared to the conventional method in which power adjustment and frequency reuse are not addressed inside cells. Furthermore, the HCEE algorithm can decrease the interference probability among neighboring cells.

For future work, the authors consider research covering a city with tall and dense buildings and extend the proposed work for those situations.

References

[1] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges", *IEEE Commun. Surv. & Tutor.*, vol. 13, pp. 524–540, 2011.

[2] GreenTouch [Online]. Available: http://www.greentouch.org

[3] M. A. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks communications", in *Proc. IEEE Int. Conf. Commun. ICC 2009*, Dresden, Germany, 2009, pp. 1–5.

[4] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks", in *Proc. 19th Future Network and Mobile Summit*, Florence, Italy, 2010, pp. 1–8.

[5] F. R. Dogar, P. Steenkiste, and K. Papagiannaki, "Catnap: exploiting high bandwidth wireless interfaces to save energy for mobile devices", in *Proc. 8th Int. Conf. Mob. Syst., Appl., and Serv. MobiSys'10*, San Francisco, CA, USA, 2010, pp. 107–122.

[6] A. Schulman *et al.*, "Bartendr: a practical approach to energy-aware cellular data scheduling", in *Proc. 16th Ann. Int. Conf. Mob. Comput. Netw. MobiCom 2010*, Chicago, IL, USA, 2010, pp. 85–96.

[7] M.-R. Ra *et al.*, "Energy-delay tradeoffs in smartphone applications", *Proc. 8th Int. Conf. Mob. Syst., Appl., and Serv. MobiSys'10*, San Francisco, CA, USA, 2010, pp. 255–270.

[8] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks", *IEEE Commun. Mag.*, vol. 48, pp. 74–79, 2010.

[9] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure", *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, 2011.

[10] C. Han *et al.*, "Green radio: radio techniques to enable energy-efficient wireless networks", *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, 2011.

[11] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks", *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, 2011.

[12] G. Miao, N. Himayat, and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels", *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 545–554, 2010.

[13] G. Y. Li *et al.*, "Energy-efficient wireless communications: tutorial, survey, and open issues", *IEEE Wireless Commun*., vol. 18, no. 6, pp. 28–35, 2011.

[14] J. Liu, F. Zhao, X. Liu, and W. He, "Challenges towards elastic power management in internet data centers", in *Proc. 29th IEEE Int. Conf. Distrib. Comput. Syst. Worksh. ICDCS 2009*, Montreal, Québec, Canada, 2009, pp. 65–72.

[15] M. Stansberry, "The green data cente: energy-efficient computing in the 21st century", 2009 [Online]. Available: http://searchdatacenter.techtarget.com/feature/The-Green-Data-Center-Energy-Efficient-Computing-in-the-21st-Century

[16] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, "A survey of green mobile networks: Opportunities and challenges", *Mobile Netw. and Appl.*, vol. 17, no. 1, pp. 4–20, 2012.

[17] T. Haynes, "Designing energy-smart 3G base stations", *RF Design*, vol. 30, pp. 18–22, 2007.

[18] X. Lu, E. Erkip, Y. Wang, and D. Goodman, "Power efficient multimedia communication over wireless channels", *IEEE J. Selec. Areas in Commun.*, vol. 21, no. 10, pp. 1738–1751, 2003.

[19] B. Anand, A. Ananda, M. C. Chan, L. T. Le, and R. K. Balan, "Game action based power management for multiplayer online game", in *Proc. 1st ACM Worksh. Netw., Syst., Appl. Mob. Handhelds MobiHeld 2009*, Barcelona, Spain, 2009, pp. 55–60.

[20] H. Claussen, L. T. Ho, and L. G. Samuel, "Self-optimization of coverage for femtocell deployments", in *Proc. Wirel. Telecommun. Symp. WTS 2008*, Pomona, CA, USA, 2008, pp. 278–285.

[21] M. V. Rali, M. Song, and S. Shetty, "Virtual Wired Transmission scheme using Directional antennas to improve Energy Efficiency in Wireless Mobile Ad-hoc Networks", in *Proc. IEEE Milit. Commun. Conf. MILCOM 2009*, Boston, MA, USA, 2009.

Maryam Esmaeilifard received the B.Sc. degree in Information Technology from Azarbaijan Shahid Madani University, Tabriz, Iran, in 2011 and the M.Sc. degree in Information Technology from Sahand University of Technology, Sahand New Town, Tabriz, Iran, in 2013. Her research interests include wireless networks, green networks, network security, energy efficient security algorithms.

E-mail: m.esmaeili_69@yahoo.com
Computer Networks Research Lab
Electrical Engineering Technologies Research Center
Sahand University of Technology
Tabriz, Iran

Akbar Ghaffarpour Rahbar received the B.Sc. and M.Sc. degrees in Computer Hardware and Computer Architecture from the Iran University of Science and Technology, Tehran, Iran, in 1992 and 1995, respectively, and the Ph.D. degree in Computer Science from the University of Ottawa, Ottawa, Canada, in 2006. He is currently a Professor with the Electrical Engineering Department, Sahand University of Technology, Sahand New Town, Tabriz, Iran. He is the director of the Computer Networks Research Laboratory, Sahand University, Iran. Dr. Rahbar is a senior member of the IEEE, and member of the Editorial Board of the Wiley Transactions on Emerging Telecommunications Technologies Journal and the Journal of Convergence Information Technology. He is editor-in-chief of Journal of Nonlinear Systems in Electrical Engineering. His current research interests include optical networks, optical packet switching, scheduling, PON, IPTV, VANET, network modeling, analysis and performance evaluation.

E-mail: ghaffarpour@sut.ac.ir
Computer Networks Research Lab
Electrical Engineering Technologies Research Center
Sahand University of Technology
Tabriz, Iran

Paper

A Survey on Wireless Network Applications in Automated Public Utilities Control and Management

Raja Jitendra Nayaka[1] and Rajashekhar C. Biradar[1,2]

[1] *Department of Electronics and Communications, REVA Institute of Technology and Management, Bangalore, Karnataka, India*
[2] *School of Electronics and Communication Engineering (UG), REVA University, Bangalore, India*

Abstract—Public utilities such as water, electricity and gas are essential services that play a vital role in economic and social development. Automation of public utility services addresses the huge losses faced by the utility services today, due to non-accounting of distributed utility resources. Automation improves government revenues. The different type of architectures are proposed and designed for automated metering, control and management of public utilities like water, gas, and electricity for effective management and control of resources. The various network topologies, hardware and software architectures to automation and management of public utilities are proposed by researchers. In this paper, the different technologies proposed by various researchers across the globe are surveyed and list of issues and challenges for automated meter reading and control of public utilities is identified.

Keywords—Automated Meter Reading (AMR), Wireless Sensor Networks (WSN), Wireless communication, Short Message Services (SMS).

1. Introduction

At present, most of the houses across globe have the traditional electromechanical or digital energy, water or gas meters. These public utilities have individually managed by connected service departments. Today the billing system and control and management of public utilities are not fully automated. Presently person from the electricity or water or gas board goes to every building and takes the meter reading manually. These meter readings are used for electricity or water or gas bill calculation. It requires a large staff for reading the meters, control and manage public utilities, and eventually sending the bills to customer. A new technology named Automatic Meter Reading System (AMR) is discussed. AMR is a sophisticated system, which allows companies to collect the reading without visiting the site. As number of meter grows the manual collection of data is becomes cumbersome task and time consuming. Sometime task become infeasible if the data terminals are unreachable. Therefore, a wireless based data collection mechanism is needed. The mentioned task can be achieved by using wireless communication network. AMR is a system and process used to remotely collect electrical, water or gas meter data without the physical presence of meter readers at the user premises. With such automation, system it is possible to read multiple meters remotely at any time. AMR is also known as smart meters and associated network is called smart grid. It provides cost effective solution to meter reading service. AMR use a real time wireless communication network to connect meters with a central system.

Public utilities are essential services that play a vital role in economic and social development. Quality utilities are a prerequisite for effective poverty eradication. Increased competition in the utilities sectors in recent years has entailed changes in regulatory frameworks and ownership structures of enterprizes, in addition to business diversification and enhancing efficiency of delivery and reviewing tariffs and other sources of income collection remotely.

The authors contribution in this paper is the extensive survey of proposed electricity, water and gas meters based automated public utility control across the globe. The comparative analysis of related works with parameters which includes communication protocols, topologies, network, hardware, and applications used for automated management of public utilities is made.

In Section 2 the taxonomy of automated utility management and comparative study of related works is provided. Section 3 provides a survey of integrated solutions for utility management. Section 4 presents the issues, challenges in automated management of public utilities in present scenario, Section 5 presents future research directions and overview of proposed methodology. The paper is concluded in Section 6.

2. Taxonomy of Automated Utility Management

Recent research has developed several techniques that deal with various types of automated control and management of public utilities control at different networks. To assist in understanding the assumptions, the authors focus design and development of these techniques. In this section the taxonomy of different automation technologies used in traditional systems and survey various research on automated

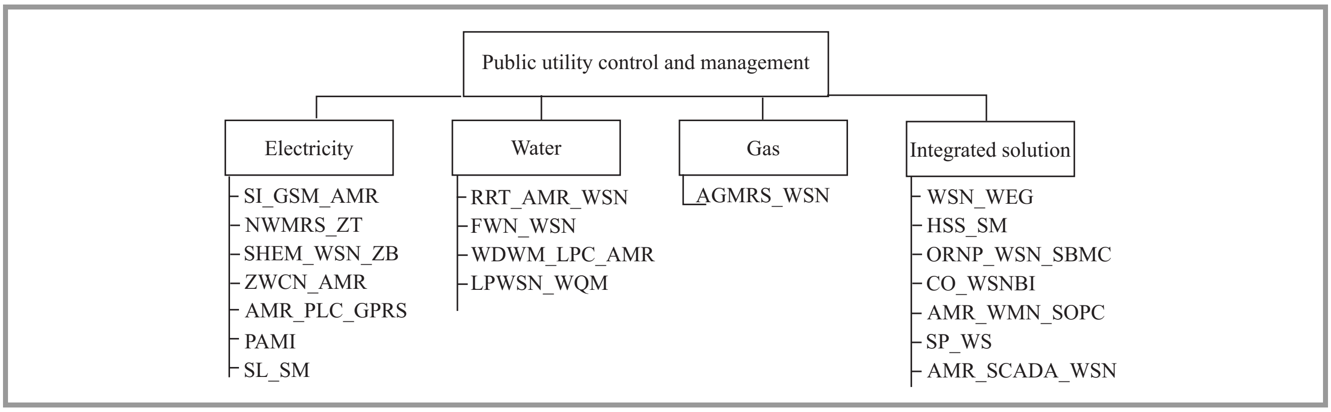


Fig. 1. Taxonomy of public utility management and control.

public utility management shown in Fig. 1 is provided. Public utility management and control mechanisms are categorized based on type of utility and applied technologies. The basis of utility classification includes: (a) electricity, (b) water, (c) gas and (d) integrated solution. Public utility management and control mechanisms can be further classified based on the technologies used.

Examples of automated electricity based utility management are:

- SI_GSM_AMR – smart and intelligent GSM based AMR system,
- NWMRS_ZT – networked wireless meter reading system based on ZigBee technology,
- SHEM_WSN_ZB – smart home energy management system using IEEE 802.15.4 and ZigBee,
- ZWCN_AMR – ZigBee wireless communication network design as an answer for last-mile problem in consumer AMR systems,
- AMR_PLC_GPRS – implementation of AMR system using PLC and GPRS,
- EEAMI – an energy efficient advanced metering infrastructure,
- PAMI – the prepaid advanced metering infrastructure,
- WDTMS_RE – the application of wireless data transmit module in monitoring system under remote environment,
- WSN_SM – wireless sensor network based smart meter.

Examples of automated water based utility management are:

- RRT_AMR_WSN – remote real time AMR system based on WSN,
- FWM_WSN – a flexible architectural framework for water management based on wireless sensor network,
- WDWM_LPC_AMR – wireless digital water meter with low power consumption for AMR,
- LPWSN_WQM – low power wireless sensor system for water quality monitoring.

Example of automated gas based utility management is:

- AGMRS_WSN – automatic gas meter reading system based on WSN,

Examples of automated Integrated solution based utility management are:

- AMR_SA_WSN – AMR and SCADA application for wireless sensor network,
- WSN_WEG – WSN in water, electricity and gas industry,
- AMR_SCADA_WSN – AMR and SCADA application for WSN,
- HSS_SM – hybrid spread spectrum based wireless smart meter,
- ORNP_WSN_SBMC – optimal relay node placement in WSN for smart buildings metering and control,
- CO_WSNBI – cost optimization of wireless-enabled metering infrastructures,
- AMR_WMN_SOPC – AMR system based on wireless mesh networks and SOPC technology,
- SP_WS_WGMS – self powered wireless sensors for water gas meter system.

Automated reading technologies are classified based on the parameters such as bandwidth, delay, reliability, etc. The reason behind classifying automated reading mechanisms based on technology is that the topological analysis gives better way of representing the utility technology to provide clarity and effectiveness. Further classification is based upon various services that every topology based technology supports. The services supported by the topology based on utility can be seen at broader way [1].

2.1. Electricity as Public Utility

A. Jain *et al*. in 2012 presented the development of Smart and Intelligent GSM based Automatic Meter Reading System, which is has capabilities of remote monitoring and controlling of energy meter [2]. AMR continuously monitors the energy meter and sends data on request of service provider through SMS. The data received from an energy meter is stored in database server located at electricity board station through SMS gateway. It is further processed by energy provider. Energy provider sends electricity bill either by e-mail, SMS or by post [3]. This system allows to the customers to pay bill online by card or by money transfer. This system helps electricity companies to take action against lenient customers who have outstanding dues. Otherwise, companies can disconnect the power and reconnect it after deposition of dues. This system also gives the power cut information and tempering alert.

L. Cao *et al*. in 2008 carried research on networked wireless AMR system based on wireless sensor networks and ZigBee technology. The paper presented the mechanism for solving the problems existed in the present meter reading system. The hardware structure of system employs WSNs consists of measure meters, sensor nodes, data collectors, server and wireless communication network [4], [5]. The mesh network topology was adopted in this design. For a short distance transmission, the data sink collects data from the meter sensors using the ZigBee [6]. For a long distance transmission system uses TCP/IP protocol from the data sink to the server. A modified routing protocol is used based on LEACH is adopted in this system.

D.-M. Han *et al*. in 2010 proposed smart home energy management system using IEEE 802.15.4 and ZigBee. This mechanism introduces smart home interfaces and device definitions to allow interoperability among ZigBee devices produced by various manufacturers of electrical equipment, meters and smart energy enabling products. It introduces home energy control systems design to provide intelligent services for users and tested in real test bed [7].

In the research of K. Marcinek in 2011 design and implementation of ZigBee wireless communication network design as an answer for last-mile problem in consumer AMR systems [8] are proposed. The main stress was given to compatibility and integration with existing AMR networks along with installation complexity. It maintains low bit error rate (BER). Real life implementation of ZigBee based energy meter reading in one of the modern estates in Warsaw was presented in this work. In this mechanism ZigBee wireless communication network design for last-mile problem in consumer AMR systems has been proposed. Author achieved results of 98% of delivered packet rate. It also said this solution provides both transparent and buffered communication independent to attached device protocol. It is proposed that, usage dedicated GSM modem with ZigBee integration board with into existing AMR systems. This solution reduces both costs and installation time. PC software allows the system owner remotely monitoring network along with modem parameters management for flexibility. Future work proposed by author concern further system functionality improvements. Usage of module with on-board application processor and external memory for implementing local buffered data read out and gaining independence from the radio link quality fluctuations or temporary inaction. Tests with 868 MHz modules are also proposed. It is indicated that the only disadvantage is in Industrial, Scientific and Medical (ISM) band has its 10% duty cycle limitation. On heavy network load, this regulation decreases the real RF data rate from 24 Kb/s to 2400 b/s, which is over 100 times slower in comparison with 250 Kb/s on 2.4 GHz band.

J. Yang *et al*. in 2011 presented an AMR system using PLC and GPRS communication [9]. Authors have studied that electricity meters installed in every household is connected to a collector through RS-485 interface. The communication between data sink and collectors is done using Power Line Carrier (PLC). The data sink is connected to master station via a GPRS accessing to Internet. It is indicated that proposed solution can be used for water and gas meters. Authors proposed an energy efficient advanced metering infrastructure (EEAMI) for meter data collection and energy management. An energy efficient Advanced Metering Infrastructure (AMI) is an AMR infrastructure with bidirectional meters. These meters are called smart meters they are connected to the gateway through power lines and gateway that communicates to the central station. The central station communicates through GSM.

Kishore *et al*. in 2012 proposed the prepaid advanced metering infrastructure (PAMI). It combines with 3G network technology [10]. This research claims that proposed technology will make the processing fast and reduce the theft of electricity. It will make people more conscious and will save electricity. Moreover, people in this mechanism can recharge their smart cards with the desired amount even at the end of the month. Future work includes using supercapacitor instead of using fixed battery inside the electronic meter and develop the system in ASIC chip. In addition, real time clock can be interfaced with the electronic meter so that when credit finishes at night or at holidays, the meter will not close the valve at that time, rather continue with negative billing and finally close the valve at working hours.

Z. Ailing *et al*. in 2004 proposed an AMR system using the application of wireless data transmit module in monitoring system under remote environment, which is low cost, high performance solution [11].

2.2. Smart Grid in Electricity Distribution

A smart grid is an electricity distribution network that combines a bidirectional power flow network integrated with a bidirectional information flow in such a manner as to facilitate various optimization and control features.

Table 1
Automation of electricity utility management

No.	Research	Protocol	Topology	Network	Hardware	Application
1	SI_GSM_AMR	SMS, GSM	—	GSM	—	Electricity
2	NWMRS_ZT	LEACH, TCP/IP	Mesh	ZigBee, IP	S344B0X	Electricity
3	SHEM_WSN_ZB	Disjoint multipath routing protocol (DMPR)	—	ZigBee, HAN	CC2420, CC2430, 8051 MCU	Consumer electronics device
4	ZWCN_AMR	ZigBee	—	WSN	—	Electricity
5	AMR_PLC_GPRS	SMS, GPRS, TCP/IP	—	PLC, GPRS	—	Electricity
6	PAMI	GSM	—	3G	GPRS modem	Electricity
7	WSN_SM	Reliable block transport (RBT)	Cluster tree	—	LPC1763 ARM, CC1120	Electricity

A key enabler for the smart grid is a distribution utility is the final touch point with residential and industrial consumers. Apart from delivering quality power to the consumer premises, it has to manage consumer expectations, environmental implications, and billing to generate revenue for the entire stakeholders in the power ecosystem. It acts as the perfect input for revenue management, energy accounting, and billing for the utility. One of the most compelling benefits of the smart grid is the promise of delivering demand management or load control. Utilities will save energy, lower costs and defer additional transmission and generation expenses with the ability to shape load and curtail load to mitigate grid events. Additionally, consumers will be able to conserve energy use to benefit from time of use or time based rate structures. Various studies have shown that these actions can give 15 to 20% savings.

Next generation power grid, uses two-way flows of electricity and information to create a widely distributed automated energy delivery network.

The smart grid in electricity distribution was discussed by X. Fang *et al.* in 2011 [12]. The three major systems were explored, namely the smart infrastructure, the smart management and the smart protection. Authors also proposed possible future directions in each system. For the smart management system, the authors presented various management objectives, such as improving energy efficiency, profiling demand, maximizing utility, reducing cost and controlling emission. The authors also explored various management methods to achieve these objectives. For the smart protection system, they discussed failure protection mechanisms, which will improve the reliability of the smart grid and explored the security and privacy issues [7], [13].

Comparison in Table 1 summarizes automation of electricity utility management in terms of protocol, topology, network, hardware chips or modules and applications used for automated management of electricity. As per survey of related works, its observed that automated electricity meter reading uses GSM, GPRS, 3G, PLC and ZigBee based networks for a communication. SMS based prepaid electricity billing systems are also used [1], [14].

2.3. Water as Public Utility

Water meter should be introduced across the country to help tackle water shortages. Meters provide water use information that will help the department to monitor the effectiveness of water resource plans and their progress in meeting environmental flows and water allocation objectives. Metering water use encourages more efficient management practices, allowing a better usage of water used and improve water use efficiency. There are a number of advantages of using Supervisory Control and Data Acquisition (SCADA) in water distribution [15].

L. Cao *et al.* in 2008 presented a remote real time AMR system based on wireless sensor networks [4], where AMR sensors were implemented on water supply system. Presented mechanism employs distributed structure based on WSN, which consists of measure meters, sensor nodes, data collectors, and server. For short-range communication, RF and ZigBee are used to collect data, and CDMA cellular network is used to collect data from the server. The water meter data are received at the server through LAN using TCP/IP protocol.

M. Fayed *et al.* in 2011 proposed a flexible architectural framework for water management based on WSN and highlighted the need for water management [5]. The authors found that, WSN could play a very important role in helping to reduce water wastage, increase water efficiency and utility. Building such a WSN presents many challenges, which are different from those of other applications. In this mechanism, authors discussed and proposed an architectural framework to address those challenges. The proposed framework allows for substitution of the mechanics of transport for information, thereby increasing fault tolerance, and resiliency during natural or manmade disasters.

Y.-W. Lee *et al.* in 2008 proposed wireless digital water meter with low power consumption for AMR [16]. Authors used magnetic hole sensors to compute water consumption. The readings are transferred via ZigBee to a gateway. It is suggested that dual 3-volts batteries having 3 Ah, would last 8 years by analyzing the real power consumption.

Table 2
Automation of water utility management

No.	Research	Protocol	Topology	Network	Hardware	Application
1	RRT_AMR_WSN	ZigBee, TCP/IP	—	WSN, CDMA	ARM MPU S344B0X	Water well
2	FWN_WSN	Relay agent routing and scheduling	—	WSN, Multi-tiered hybrid network	—	Water management
3	WDWM_LPC_AMR	ZigBee	—	ZigBee	MSP430 MCU, magnetic Hall sensor, HLL	Water meter
4	LPWSSN_WQM	ZigBee	—	WSN, ZigBee	MPC82G516A, nRF24L01, PIC12F629	Water quality management

The researchers W.-Y. Chung *et al*. in 2011 proposed low power wireless sensor system for water quality monitoring. In this mechanism the MPC82G516A 8-bit microcontroller, and a nRF24L01 2.4 GHz wireless transceiver module, together with a PIC12F629 8-bit micro-controller are used to design a basic wireless node. These water quality parameters acquired from the sensor node are transmitted to the repeat node via 2.4 GHz wireless signal. The repeat node receives data and transmits to the main node [17], and then to PC by the using of RS232 interface. In addition, a wireless signal path from the sensor node to repeat node uses a single direction relaying method, thereby making the sensor node and repeat node to be in sleep mode when idle. In sleep mode, authors claim all nodes consume only 27 μA at 3 V.

2.4. SCADA Systems

Water SCADA is the term usually used to describe the computerized central control system used in many drinking water utilities. SCADA system replaced the legacy control schemes, which utilized electromechanical process control. The components were fitted with Programmable Logic Controllers (PLCs), the individual wires were replaced by Ethernet cables and the control panel was replaced by Human Machine Interface (HMI) software operating on a PC. The dedicated communications channels for remote facilities were replaced by Internet cable connections and wireless link.

The Internet SCADA system was implemented by Z. Feng *et al*. in 2008. A computer screen replaced the large mechanical control panel with its dozens of dials, levers and mechanical control. The data logger is connected to the electronic readout on pressurized systems and pressure transducer on canal systems. Basic data for water users is to collect water flow. For water users, the minimum interval for collecting data would be once weekly, although most users find that collecting at more frequent intervals.

Recent advances in communications technology and WSN made new trends to emerge in agriculture sector [13]. One such new trend is using WSNs in monitoring water level in the farm area for precision agriculture. Few algorithms offer a maximum opportunity of delivery of water level information packets or signals to base station.

2.5. SCADA Security

SCADA systems or distributed control systems are widely utilized in industries plant or infrastructure like electric or water or gas production and distribution systems. In the other words, in case of disruption or destruction of these critical services, catastrophic events might be occurred. Research has revealed that there is a lack of security in SCADA systems. Although they have historically been isolated from other computers like enterprise networks, they have been interconnecting with enterprise network or Internet by spreading with TCP/IP as a carrier protocol. This has been led to emerge new vulnerabilities while proprietary OS have been incapable of performing emerging security mechanism. Control system communication protocol like Modbus, Profibus and DNP are still used widely in control system network, even though some more secure protocol or version are developed. Existing vulnerable protocols will continue to be used in the future by reason of economy and backward compatibility. The lack of strict security policy including weak protection of user credentials cause, information leak through insecure service configuration, services running with unnecessary privileges as well as unauthorized physical access to devices. It must also ensure that only authorized parties have access to system, services and sensitive information about system structure and elements. Comprehensive strategy for cyber attacks against the nation's critical infrastructure requires understanding the nature of the threats. Therefore, it is necessary to create depth defense and proactive solutions in terms of improving the security of SCADA control systems.

Table 2 summarizes automation of water utility management in terms of protocol, topology, network, hardware and applications used for integrated automated management of water metering and control.

2.6. Gas as Public Utility

The wireless remote gas meter is a new type intelligent measurement equipment with an added function of remote

Table 3
Automation of gas utility management

No.	Research	Protocol	Topology	Network	Hardware	Application
1	AGMRS WSN	ZigBee	—	WSN, GPRS	—	Gas

control as compared to ordinary gas meter. Automation enables the gas companies not only to monitor and control the meters remotely. It also accurately counts the gap between total gas supply and gas consumption, so as to digitize and modernized management of the gas. All the diaphragm gas meters have their relative wireless remote meter product model with direct-reading technology and built-in antenna.

In view of the status and shortcomings of existing Gas meter reading systems, Y. Jie Yang *et al*. in 2012 propose an automatic gas meter reading system based on the WSN and GPRS technologies [18]. The proposed mechanism is consisting of three level network. The first and the second level use the different frequencies during communication. It decreases the interface of each other during the data translation and increase the smart gas meter lifetime. The proposed system has the advantages of easy construct, flexible layout and lower power consumption. Table 3 summarizes automation of gas utility management in terms of technology, protocol, topology, network, hardware and applications used for integrated automated management of gas metering and utility.

2.7. Communication Technologies for Utility Management

A communication technologies used in each type of utility is shown in Fig. 2. For electricity utility management, there are three types of controls such as single phase, three phase and smart grid. The communication technologies used for electricity utility management under single phase and three

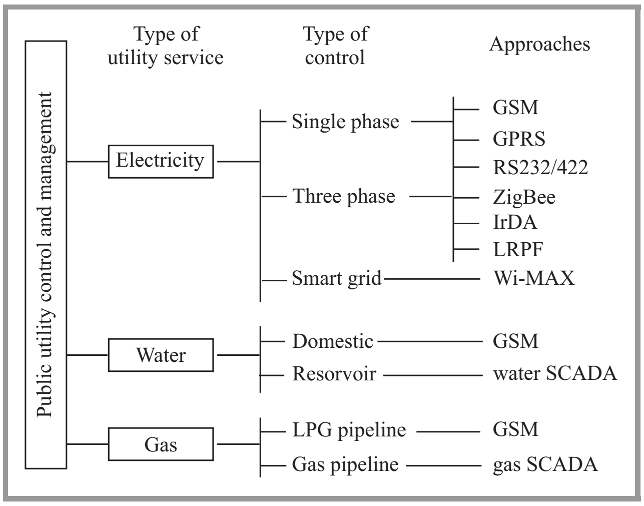


Fig. 2. Technologies used for automated public utility management.

are GSM, GPRS, RS232/422, ZigBee, IrDA and LPRF. Smart grid management uses Wi-MAX.

For water utility management, there are two types of controls such as domestic and reservoir control. For gas utility management, there are two types of controls such as LPG pipeline and gas control. The communication technologies used are GSM and gas SCADA respectively.

AMR uses various technologies and protocols for meter data collection such as PLC, Wi-Fi, ZigBee, Bluetooth, LPRF and GSM network. The taxonomy is as shown in Fig. 2. The smart meter can be used to implement the features of billing, credit management, communication, remote connection or disconnection, revenue protection, power quality measurement, load and loss control, load forecast, common user console with power management. The new designs will also support advanced tamper detection and as well support geographical information by GPS. The smart meter will have features like remote unit with display, touch screen to set different load profile options and control. AMR system helps the customer and energy service provider to access the accurate and updated data from the energy meter. AMR system can send energy consumption in hourly, monthly or on request. This data is sent to central system for billing and troubleshooting. These data are stored into the database server for processing and recording.

2.8. Comparison of Communication Technologies

There are various advantages and drawbacks in communication technologies used for utility management. Table 4 summarizes pros and cons of various technologies used in this area.

A WSN consists of densely distributed sensor nodes in a geographical area for collecting and processing data. Further transmitting the data to nearby base stations for processing. Short-range radio communication is used for transmission of data between the sensors. Transmission coverage becomes very important parameter, it measures how effectively monitored data by sensor network. The WSNs coverage issues are shown in Table 5.

Table 5
Frequency and range for AMR communication

Band [MHz]	Indoors [m]	Outdooors (with obstacles) [m]	Outdoors (with in line-sight)
433	< 100	< 500	1–2 km
800	< 50	< 300	0.8–1.5 km
900	< 30	< 200	< 800 m

Table 4
Technologies used for AMR communication

Wi-MAX (IEEE 802.16)	It does not require deployment of a costly wired infrastructure	Early stage of deployment, uncertain whether the technology will meet its range targets
Narrow band	Field-proven in Europe	Expensive deployment and not suited for particular application
Cellular service	High coverage area and potentially low costs	Fast development of new technology and their is danger of being tied to one service provider some packet-switched services not very reliable
Satellite services	Universally available, regardless of concrete location	It has high cost of maintenance, low reliability during bad weather conditions
SMS/GPRS	Low costs and highly reliable	Low bandwidth and thus only support of a few applications such as simple emergency alerts
ZigBee (IEEE 802.15.4)	Low power requirements low implementation cost with good scalability. It is particularly designed for use in industrial and home automation or security applications	It is suitable for low power requirements, it requires low implementation cost. It has good scalability. They are particularly designed for use in industrial and home automation or security applications
Bluetooth (IEEE 802.15.1)	It is more mature than ZigBee and many products already available. It permits higher data rates than ZigBee	Most meters do not have Bluetooth implementation. It has limited maximum number of devices in a network. There are issues with security vulnerabilities
Wi-Fi (IEEE 802.11)	Deployment is easy and cost is falling	Only useful within the customer site. It requires additional security layers

Cluster tree is the best topology for a large public utility control and management network. Ring and star topology are unsuitable due to their sheer scale of the entire network. Cluster tree topology allows network to divide into sub parts and they are easily manageable. A cluster tree network provides enough room for future expansion. It overcomes the limitation of star network topology, which has a limitation of hub connection points and the broadcast traffic induced limitation of a bus network topology. All nodes in cluster tree have access to their immediate network neighbors. The cluster tree network makes it possible for multiple network nodes to be connected with the central hub or data fusion.

2.9. Battery and Network Life

The reliability of a WSN is related to the availability of a communication path between two wireless devices. The sensor nodes in WSN operates on battery power. The power source for wireless sensors has mainly been disposable as primary batteries. Since networks consists thousands of sensor nodes, challenge is with using disposable batteries and maintaining sensors in service locations. Changing the batteries in the field is therefore a cumbersome task [16], [19]. Randomly distributed sensor networks makes difficulty to changing batteries. Making recharging almost impossible during operations. This problem has forced WSN and node developers to make changes in the basic architecture to minimize the energy consumption. Nodes make the network and overall system application more energy efficient.

Recently the IEEE 802.15.4 standard was developed for low data-rate application, which needed to last for longer duration by consuming relatively less energy. ZigBee 802.15.4 technology is one of a number of promising technologies for wireless communication due to its low cost and complexity, and overall energy efficiency.

Studies by Polastre *et al*. (2005), Chiasserini (2002), Kumar *et al*. (2005), mentioned that lifetime of wireless sensor networks before their installation is an important concern. Study shows, still there are some precautions to be taken by which a sensor network system can be made to run for longer time [17], [20], [21]. The problem of battery energy consumption in sensor networks depends on node architecture, network structure and routing algorithm to support energy saving in the network. Stand-alone measures such as selecting a low-power microcontroller with embedded transceiver will important factor to achieve energy saving over the entire network. Energy efficient WSN design objectives needs to look at different aspects, such as application code, network configuration, routing algorithms, etc.

Numerous types of batteries are available including alkaline, carbon zinc, zinc air, and lithium based batteries. Traditional energy harvesting such as solar, piezo, and thermal, share a common limitation of being reliant on ambient sources generally beyond their control. Hence, these solutions are not suitable for WSN. The majority of researches use a definition suitable for the context of their work [10].

The novel AMR devices generate high current pulses at periodic intervals with little current between signal transmissions. Lithium thionyl chloride batteries are generally preferred to power AMR devices due to their inherent long life and high-energy density. Among of all the available lithium battery chemistries, bobbin-type Li-$SOCl_2$ cells offer the advantages of higher energy density and voltage, excellent temperature characteristics, low self-discharge rates and good safety. Many of these components have been operating for over 15 years without a battery change. Study shows that these batteries can last up to 20 years. Reliability is another major advantage, as these batteries can operate in severe environmental conditions from –40 to 85°C.

3. Integrated Solutions for Utility Management

Apart from the wide variety of individual electricity or water or gas or automated public utility management solutions discussed in previous sections, there are many integrated automated public utility management services, which fit into more than one category. In this section all such solutions proposed by research community across globe in this domain have been briefly described.

The author Aghaei in 2011 proposed WSN in water, electricity and gas industry in which sensors compose main components [22]. Sensors are deployed and are connected with each other in environment dynamically. In this mechanism, it is presented a model for processes which are related to user of water, electricity and gas meter reading. Distribution of bills, sending notice, cutting, and reconnection of flow by using WSN were tested in Iran. The researcher showed that the proposed model leads to a great deal of costs savings.

Garlapati *et al.* in 2012 proposed a hybrid spread spectrum based smart meter network design that reduces the overhead, latency and power consumption in data transfer when compared to the 3G cellular technologies [23].

ZhiliZhou *et al.* in 2013 proposed an optimal relay node placement in WSN for smart buildings metering and control. This mechanism examines WSN communication infrastructure for smart grid implementation in building. Proposed a scheme for the deployment in buildings, in which sensors are massively placed to meter electricity consumption and collect illumination, thermal, pressure information and multiple base stations are connected with the communication network for power grid distribution network [24]. The paper exploits the software tool to simulate building environment and to test optimal deployed WSN. Integer programming approaches for both deterministic and robust cases are considered.

Das *et al.* in 2012 proposed cost optimization of wireless-enabled AMI, which measures, collects and analyzes information by communicating with metering devices either on request or on a schedule. The AMI consists of a collection of Neighborhood Area Networks (NANs), which include smart, wireless-enabled mesh-connected meters or sensor nodes. Each NAN is controlled by a gateway or Access Point (AP). These devices in turn are usually mesh-connected using wireless or wire-line backhaul links to a servers. This contribution develops an elegant graph-theoretic approach for optimizing the cost of an AMI by maximizing the ratio of the number of sensors nodes in a NAN to that of gateways or APs. A Matlab program has been implemented to automate their approach, which can deal with random and complex NAN topologies.

Cao *et al.* proposed in 2009 remote wireless AMR system based on wireless mesh networks and SOPC technology [14]. The system consists of measure meters, wireless sensor nodes, data collector, management centre and wireless communication networks. The data is transmitted from the sensor nodes to the data collector using ZigBee. The system uses Ethernet to transmit data from the data collector to the management centre. The data collector acts as gateway, it is adopted wireless mesh network topology structures. Management center is based on FPGA chip. In ZigBee sensor node design, Atmel MEGA128 microcontroller is used. Wireless chip TI CC2420 is used for communication unit. The systems presented in this mechanism have many significant excellences, such as networked, wireless, moveable and lower power consuming. The proposed system have abroad application foreground in the real application field to remote measure and manage of electric power, water supply, gas supply and heat supply.

Di Zenobio *et al.* in 2012 carried research on self powered wireless sensors. Described mechanism is a new solution for a wireless self-powered sensors network, which allows the energy harvesting from the action of a turbine wheel rotating in the path of a fluid stream environment [25].

New devices family was introduced to find application in water or gas smart metering systems. Reinhardt *et al.* in 2011 and Cao *et al.* in 2008 presented low-power hardware mechanism and incorporates it in a reprogrammable microcontroller, which allows developers easily deploying new algorithms. This IEEE 802.15.4-compliant radio transceiver makes its integration with existing sensor [3], [25], [26].

Francisco *et al.* in 2013 proposed automated meter reading and SCADA application for WSN. The authors found that currently, there are many technologies available to automate public utilities services [13]. AMR and SCADA are the main functions these technologies must support. In their work, authors propose a low cost network with a similar architecture to a static ad-hoc sensor network based on low power and unlicensed band radio. Topological parameters for this network are analyzed to obtain optimal performance and to derive a pseudo-range criterion to create an application-specific spanning tree for polling optimization purposes. In application layer services, authors analytically studied different polling schemes.

Table 6 summarizes automation of utility management in terms of various technologies, protocol, topology, network, hardware and applications used for integrated automated management of utilities.

This research study shows that, there is scope to integrate multiple utility sensors using common platform. The proposed hardware are SoC based on FPGA using ZigBee tech-

Table 6
Automation of integrated utility management

No.	Research area	Protocol	Topology	Network	Hardware	Application
1	WSN_WEG	Semiautomatic WASN, SMS	—	GPS, WSN	—	Water, gas, electricity
2	HSS_SM	HSS, multiuser detection	—	HSS based AMI network	PIC MCU	Electricity
3	ORNP_WSN_SBMC	DRP_IS, RNP_IS	—	WSN, ZigBee	—	Electricity, thermal, pressure
4	CO_WSNBI	—	—	WSN	—	Electricity
5	AMR_WMN_SOPC	ZigBee, TCP/IP, CSMA/CD	Mesh network	ZigBee, IP	FPGA, EP2C35F67272C8, MCU, MEGA128, CC2420	Water, gas, electricity
6	SP_WS_WGMS	M_Bus	—	UMTS, LTE, 3G	MSP430L092 MCU	Water, gas
7	AMR_SCADA_WSN	Spanning tree	Static ad-hoc sensor network	WSN	—	Water, SCADA

nology. The Hybrid Spread Spectrum (HSS), ad-hoc, mesh and NAN topologies are found to be suitable for utility management and water SCADA.

In fact, the actual selection of utility technology depends on several factors such as geographic coverage of the communication architecture, the locations of substations, cost of communication architecture, and a remote control center with network management types. As a result, electric utilities should evaluate their unique communication requirements and the capabilities oftechnologies comprehensively in order to determine the best solution for automation applications [20], [27].

4. Issues and Challenges in Automated Control and Management of Public Utilities

Global metering service industry is a heterogeneous one with multiple communication protocols and interfaces. Another issue is the difficulties in integration of different make of meters at the field level. Evolution of the electricity has historically taken place with proprietary protocols and interfaces to provide internally stored values in formats is unique to the manufacturer. With the change in requirements of the utilities, additional parameters and features have been added resulting in different versions of meters even from the same manufacturer. The users and service providers with these multiple versions of meters are burdened with multiple data formats on proprietary protocols. The utilities service providers have to buy and maintain separate Application Program Interface (API) software from each meter manufacturer in order to make use of the data from different versions. In addition, third party hand held readers and remote metering systems have to be updated for every new meter type or version. The proprietary protocols results in dependence on the vendors of meters as the APIs are needed for integration of metering information with the IT infrastructure. This resulted in focus on the development of open integrated utility metering protocol and interface.

The interoperability is the capability of the data collection system to exchange data with meters of different makes. This necessitates the presentation of the meter data in predefined common formats and interface. This calls for the meters which results in compact, low cost and efficient programming effort for applications using IT infrastructure. The evolution of enhanced capabilities of microprocessor based meters and the benefits led to development of open protocols independent of make. With the availability of open protocols, many options and features become available to the purchaser or software developer who may want to take advantage of them to optimize their operations or to maximize their commercial benefit.

Battery life is a critical issue for communicating AMR meters. The communicating AMR meters rely on batteries. There is a need to obtain the longest possible period of performance battery.

5. Future Research Directions

Despite the extensive research in AMR technologies, there are still several open research issues, i.e., efficient resource and route management mechanisms, inter domain network management, that need to be developed for automation applications. Possible the research scope are:

1. network topology design and development for public utilities control and management,
2. design and development of single integrated electric, water and gas meters for efficient control and management,

3. protocol design and development using analytical and probabilistic models of spectrum management,

4. design and development of utilities database management, analysis and control using software,

5. security issue analysis, tamper proof and protection,

6. development of WSN test-bed for testing public utilities management and control protocols,

7. verification of designed meters, network topologies, communication protocols and database management using developed test-bed for improvement of energy efficiency, latency, throughput, network lifetime using spectrum management and reducing control overhead for automated public utilities management using WSNs.

5.1. Overview of Proposed Methodology

In this paper the various research for automated management of utilities have been explored. The authors believe that cluster-based routing with new protocol are good candidates that can benefit integrated automation of utility services.
The utility industry is aggressively growing towards automation by cutting operating cost and increased efficiency. To meet next generation utility management network, a new architecture for integrated automation of metering, control and management of public utilities for effective management and control of resources using wireless sensor network is proposed. For conducted research an electrical, water and gas utilities are considered. This three types sensors can be integrated into single node called as EWGSN (i.e. electrical, water and gas sensor node). The EWGSN does preliminary collection of data from three utility sensors and stores into single node. Hence, EWGSN node is considered as level one data aggregation and fusion (L1DA),

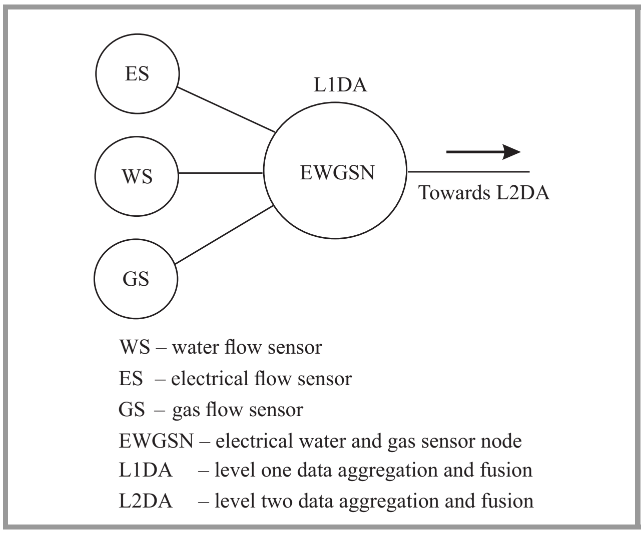


Fig. 3. Integrated sensors based AMR.

which is PAN coordinator with FFD functionality as shown in Fig. 3. The L1DA function is preliminary packetizing of sensor data into standard frame and transmits to level two data aggregation and fusion center (L2DAC). Transmit frame is used to transmit data readings from sensors at EWGSN node to data fusion centre. EWGSN Control-T consists of special requests to data aggregation centre at L2DA data fusion centre. Field check error is performed on payload using CRC 32 algorithm. Receive frame at EWGSN is used to transmit control commands data from L2DA data fusion centre to EWGSN. This frame is used for configuration to be done in EWGSN nodes remotely from L2DA. EWGSN ControlR consists of special requests from data aggregation centre at L1DA to sensors.
The proposed system are energy efficient, sensor node communicates with fusion centre on event based and nodes are always in listening mode. The single packet carries three type of payload. Since, sensor nodes are placed in utility houses rechargeable battery can be used.
The WSN based integrated control and management system mainly performs the following functions: data acquisitions, data communication, information and data presentation, monitoring and control.

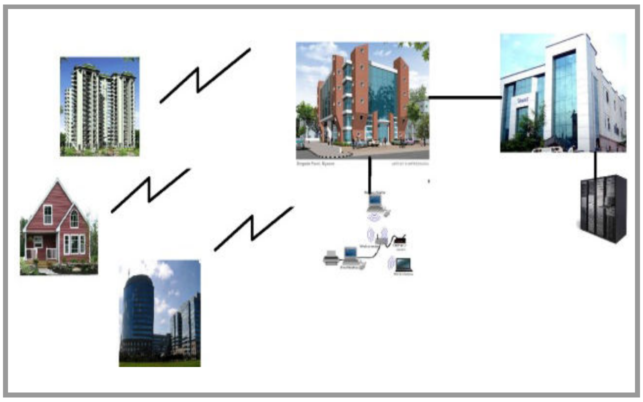


Fig. 4. Integrated AMR.

Integrated WSN based metering consist of users, i.e., individual home, apartments, offices and industries, which ware connected to nodal office as shown in Fig. 4. Nodal office is a area wise a office it communicates to users RTU in particular area through exchange of information. It is also connected to central metering and data control office. All nodal offices are connected to central office through Internet cloud using IoT as shown in Fig. 5. The central office maintains database and control part of SCADA system. The main aim of deploying the WSNs based automated utility services is to make the real time decision which has been proved to be very challenging due to the highly utility service resource constrained and communicating capacities. Huge volume of data generated by WSN based AMR sensor nodes, motivates the research community to explore novel data mining techniques, and dealing with extracting knowledge from large continuous arriving data from WSN based integrated AMR sensor node.

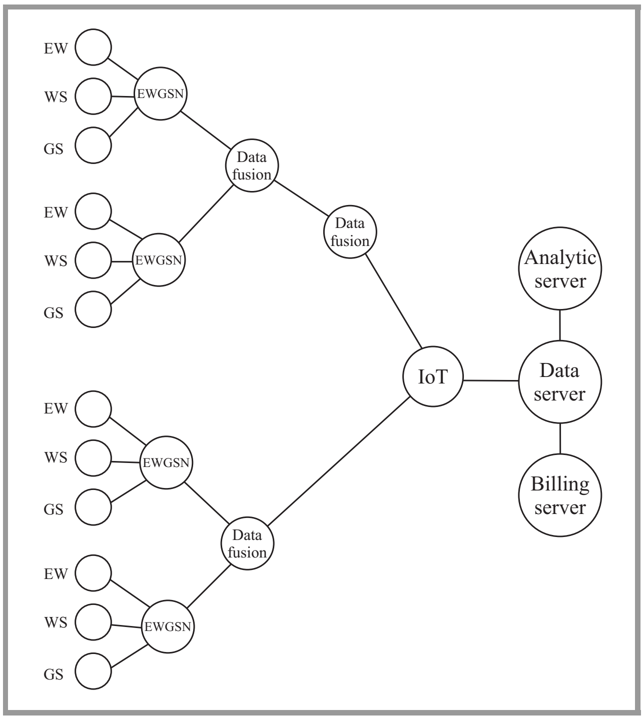


Fig. 5. Integrated sensors based AMR management.

Proposed system has following benefits:

- control units can integrate wide range of data in single frame,
- it provides on board mathematical and graphical information,
- it has ability to measure and store the historical information,
- it is easily expandable,
- to handle multiple daily data transmissions, it uses Tadiran PulsesPlus hybrid lithium battery.

6. Conclusion

Currently, there are many technologies available for automated management and control of public utilities services. There are various types of design, protocol, communication technologies and interfaces. In this paper, authors studied different methodologies proposed by various researches for automated reading and control. It is observed that WSN based solutions are cost effective for automation of integrated utility services. The conducted comparative analysis shows that automation of integrated solution with single window management of utilities appears to be highly challenging. There is further research scope for building standard frame formats, interfaces, network topology, database management, billing and network security requirement for integrated public utility management and control.

References

[1] M. Kohvakka, M. Hannikainen, and T. D. Hamalainen, "Wireless sensor network implementation for industrial linear position metering", in *Proc. IEEE 8th Euromicro Conf. Digit. Sys. Design*, Porto, Portugal, 2005, pp. 267–273.

[2] A. Jain, D. Kumar, and J. Kedia, "Smart and intelligent GSM based automatic meter reading system", *Int. J. Engin. Res. Technology* (*IJERT*), vol. 1, no. 3, 2012.

[3] A. Reinhardt *et al.*, "SmartMeter.KOM: A low-cost wireless sensor for distributed power metering", in *Proc. IEEE 36th Conf. Local Comp. Netw. LCN 2011*, Bonn, Germany, 2011, pp. 1032–1039.

[4] L. Cao, J. Tian, and Y. Liu, "Remote real time automatic meter reading system based on wireless sensor networks", in *Proc. 3rd Int. Conf. Innov. Comput. Inform. Control. ICICIC '08*, Dalian, China, 2008.

[5] A. Shah *et al.*, "Aqua-net: A flexible architectural framework for water management based on wireless sensor network", in *Proc. IEEE 24th Canadian Conf. Elec. Comp. Engin. CCECE 2011*, Ontario, Canada, 2011, pp. 481–484.

[6] A. Weigel *et al.*, "Wireless sensor networks for smart metering", in *Proc. IEEE Energy Conf. ENERGYCON 2014*, Dubrovnik, Croatia, 2014, pp. 722–729.

[7] D. Man Han and J. Hyun Lim, "Smart home energy management system using IEEE 802.15.4 and ZigBee", *IEEE Trans. Consum. Electron.*, vol. 56 , no. 3, pp. 1403–1410, 2010.

[8] K. Marcinek, "Design and implementation of ZigBee solution for last-mile problem in automatic meter reading", *Przegląd Elektrotechniczny* (*Electrical Review*), vol. 87, no. 12a, 2011.

[9] W. J. Yang "Implementation of automatic meter reading system using PLC and GPRS", *J. of Inform. Comput. Sci.*, vol. 8, pp. 4343–4350, 2011.

[10] K. Peshwani and S. Choudhary, "Energy efficient advanced metering infrastructure using 3G network", *Int. J. Power Syst. Oper. Energy Manag.*, vol. 2, no. 1, 2012.

[11] Z. Ailing and Q. Runhe, "The application of wireless data transmit module in monitoring system under remote environment", *Microcomp. Inform.*, vol. 10, pp. 61–64, 2004.

[12] X. Fang *et al.*, "Smart grid the new and improved power grid: A survey", *IEEE Commun. Surveys and Tutorials*, vol. 14, no. 4, pp. 944–980, 2011.

[13] F. Javier *et al.*, "Automated meter reading and SCADA application for wireless sensor network", in *Proc. 2nd Int. Conf. ADHOC-NOW 2003*, Montreal, Canada, 2003. LNCS, vol. 2865, pp. 223–234, Springer, 2003.

[14] L. Cao, W. Jiang, and Z. Zhang, "Automatic meter reading system based on wireless mesh networks and SOPC technology", in *Proc. 2nd Int. Conf. Intell. Netw. Intell. Sys. ICINIS '09*, Tianjin, China, 2009, pp. 142–145.

[15] N. Cochrane, "Auditor-General slams Victorian smart meters", IT News for Australian business, 12 Nov. 2009 [Online] Available: http://www.itnews.com.au/News/160398,auditorgeneral-slams-victorian-smart-meters.aspx

[16] Y.-W. Lee, E. Seongbae, and O. Seung-Hyueb, "Wireless digital water meter with low power consumption for automatic meter reading", in *Proc. Int. Conf. Converg. Hybrid Inform. Technol. ICHIT'08*, Daejeon, Korea, 2008, pp. 639–645.

[17] J. Polastre *et al.*, "Telos: enabling ultra-low power wireless research", in *Proc. 4th Int. Symp. Inform. Proces. Sensor Netw.*, Los Angeles, CA, USA, 2005, pp. 364–369.

[18] Y. Jie Yang *et al.* "Automatic gas meter reading system based on wireless sensor network", Advanced Materials Research, Jan. 2012, pp. 424–425 [Online]. Avaiable: www.scientific.net/AMR

[19] Z. Feng and S. Xian-Feng, "Remote well monitoring and control system based on wireless data transmission technology", *Process Automation Instrumentation*, no. 12, pp. 64-66, 2008.

[20] C.-F. Chiasserini, I. Chlamtac, P. Monti, and A. Nucci, "Energy efficient design of wireless ad-hoc networks", in NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications, E. Gregori *et al.*, Eds. *LNCS*, vol. 2345, pp. 376–386, Springer, 2002.

[21] C. Hsun Huang, T. Tsun Hsien, and G.-J. Jong, "Indoor power meter combined wireless sensor network for smart grid application", in *Proc. 8th Int. Con. Informa. Sci. Digit. Content Technol. ICIDT 2012*, Jeju Island, South Korea, 2012, pp. 336–339.

[22] B. Aghaei, "Using wireless sensor network in water, electricity and gas industry", in *Proc. IEEE 3rd Int. Conf. Elec. Comp. Technol. ICECT 2011*, Kanyakumari, India, 2011, vol. 2, pp. 14–17, 2011.

[23] S. Garlapati *et al.*, "Performance evaluation of hybrid spread spectrum based wireless smart meter sensor network with multi-user detection techniques", in *Proc. IEEE Int. Con. Wirel. Inform. Technol. Syst. ICWITS 2012*, Maui, Hawaii, USA, 2012.

[24] Z. Zhou *et al.*, "Optimal relay node placement in wireless sensor network for smart buildings metering and control", in *Proc. 15th IEEE Int. Conf. Commun. Technol. ICCT 2013*, Guilin, China, 2013, pp. 456–461.

[25] S. Di Zenobio *et al.*, "A self powered wireless sensor for water/gas metering systems", in *Proc. IEEE Int. Conf. Commun. ICC 2012*, Ottawa, Ontario, Canada, 2012, pp. 5772–5776.

[26] L. Cao, W. Jiang, and Z. Zhang, "Networked wireless meter reading system based on ZigBee technology", in *Proc. 20th Chinese Control and Decision Conf. CCDC 2008*, Yantai, China, 2008, pp. 3455–3460.

[27] S. Kumar, A. Arora, and T. Lai, "On the lifetime analysis of always on wireless sensor network applications", in *Proc. IEEE Int. Conf Mob Ad-Hoc Sensor Sys MASS 2005*, Washington, USA, 2005, pp. 188–190.

[28] N. Mahalakshmi, P. Krishnaiah, and E. Elavarasi, "Design of an intelligent SMS based remote metering system for AC power distribution to HT and EHT consumers", *Int. J. Comp. & Electron. Res.* (*IJCER*), vol. 2, no. 3, pp. 901–911, 2012.

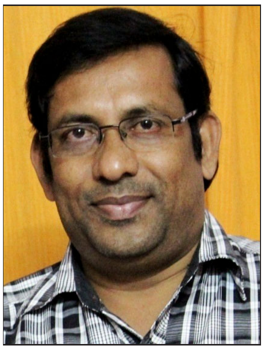
[29] N. Nhat-Quang *et al.*, "Improving the performance of mobile data collecting systems for electricity meter reading using Wireless Sensor Network", in *Proc. Int. Conf. Adv. Technol. Commun. ATC 2012*, Hanoi, Vietnam, 2012, pp. 241–246.

[30] P. Koponen *et al.*, "Interfaces of consumption metering infrastructures with the energy consumers. Review of Standards", VTT Tiedotteita – Research Notes 2542, Technical Research Centre of Finland (VTT), 2010.

[31] K. Tatsiopoulos, "A smart ZigBee based wireless sensor meter system", in *Proc. 16th Int. Conf. Sys. Sig. Image Proces. IWSSIP 2009*, Chalkida, Greece, 2009, pp. 1–4.

[32] L. Tao and X. Jianzheng, "Application of GPRS technology in monitoring system of distribution transformer", *Electrical Measurement & Instrumentation*, Jun. 2004.

[33] W.-Y. Chung, C.-L. Chen, and J.-bin Chen, "Design and implementation of low power wireless sensor system for water quality monitoring", in *Proc. 5th Int. Conf. Bioinform. Biomed. Engin. iCBBE 2011*, Wuhan, China, 2011, pp. 1–4.



Raja Jitendra Nayaka is working as senior engineer at Core R&D, ITI, Bangalore, India, Ministry of Communication, Gov. of India. He has over 18 years experience in design and development of telecom products. He obtained his M.Sc. in VLSI design and embedded systems from REVA Institute of Technology and Management, Bangalore, India. Presently he is pursuing his Ph.D. from VTU Belgaum, India. Mr. Nayaka has an experience in cryptography, network security, switching, transmission, Internet, SDH and optical communication. He has published in national and international journals and conferences. His field of interest is telecommunication, wireless sensor networks, FPGA based embedded system and IP core designs.

E-mail: rjnayaka@gmail.com
Department of Electronics and Communications
REVA Institute of Technology and Management
Bangalore, Karnataka, India

Rajashekhar C. Biradar is working as Professor and head in the Department of Information Science and Engineering, REVA Institute of Technology and Management, Bangalore, India. He obtained his Ph.D. from VTU Belgaum, India. He has many publications in reputed national/international journals and conferences. His research interests include multicast routing in mobile ad hoc networks, wireless Internet, group communication in MANETs, software agent technology, network security, multimedia communication, VLSI design and FPGA.

E-mail: raj.biradar@revainstitution.org
School of Electronics
and Communication Engineering (UG)
REVA University
Department of Information Science and Engineering
REVA Institute of Technology and Management
Bangalore, Karnataka, India

Paper

Comparison of Wavelet Decomposition Coefficients Transmission Systems Using Splines and Classical Types of Modulation

Artem I. Kutin

National Aviation University, Kiev, Ukraine

Abstract—Wavelets are becoming increasingly used as a tool for the analysis of non-stationary data. To transmit the coefficients resulting from the signal decomposition traditionally their combination into a single data packet is used, without including unequal energy contribution of each factor and impact of the decomposition level. This paper analyzes (at different transmission speeds) the signals properties produced by classical modulation methods and spline modulation for wavelet coefficients transmission proposed by the author. For all signal types the additive Gaussian noise is used as a noise disturbance.

Keywords—multi-speed channels, spline-modulation, spline Savitzky-Golay filter, wavelet decomposition.

1. Introduction

Filter banks and wavelet decomposition, are widely used for the analysis of non-stationary one-dimensional and two-dimensional data in many areas of research, such as processing video, audio, seismic, cardiology and many other signals [1]–[4].

Continuous expansion of the multiscale data representation and their applications is an important factor that determines the development of transmission methods. Easy hardware and software implementation are the fundamental conditions for most data transmission systems. To one of feasible solution for the wavelet coefficients transmission this article is dedicated.

Classical structure of three level algorithm of wavelet decomposition proposed by Mallat is shown in Fig. 1 [5].

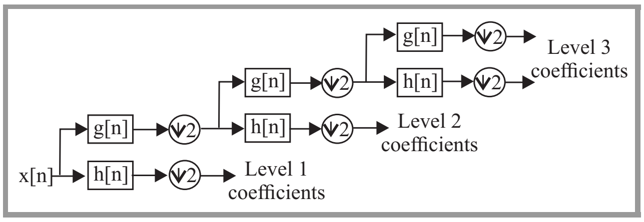


Fig. 1. Block-diagram of three-level wavelet decomposition.

At each decomposition level of approximating coefficients, detailing bits twicely reduce their quantity at the expense of decimation data, which is clearly seen in Fig. 2 [6]. On the last decomposition's level the number of approximating and detailing coefficients become equal. If the signal does not contain low frequency components, its approximation coefficients during the expansion are close to zero and don't have to be transmitted [4]. The simultaneous transfer of approximating and detailing coefficients is planned to be investigated later and in this article will not be considered.

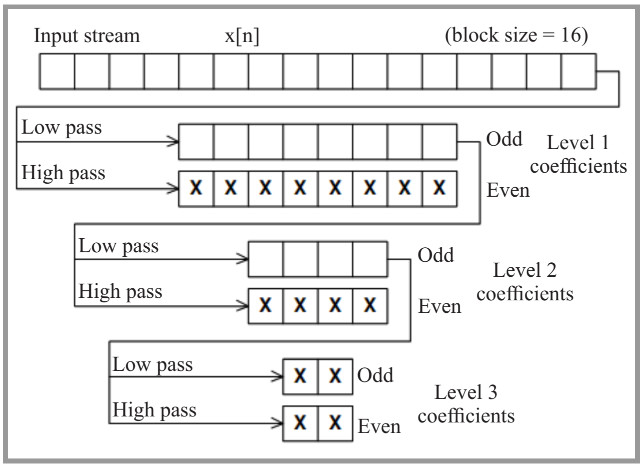


Fig. 2. Sequential decimation in the discrete wavelet decomposition.

While considering the hierarchy of wavelet decomposition, then the highest level of decomposition is based on the coefficients that correspond to low frequency signal components.

In existing data transmission systems, the expansion coefficients obtained for direct transmission over the communication channel are encoded using Huffman or arithmetic compression algorithm and transmitted using the serial two- or multiposition modulation techniques, divided into packets using channel coding or not. The use of coding aligns the error's probability of all the coefficients, however, in many applied problems it is possible, from the perspective of reducing energy contribution, to assert with confidence that the coefficients significance proportionally decreases from the base of the pyramid to its top.

The situation is similar for data transmission systems of telemetry and remote control. The bits significance of binary words that correspond to the parameters of object's motion, i.e. speed, direction, for example, transmitted from the drone, is also not the same, decreases in proportion to the reduction in the weight category of the word. Errors in transmission LSBs will have less effect than MSBs.

Therefore, if the data transmission system due to its-characteristics cannot guarantee the transmission's reliability of the coefficients at all decomposition's levels during changes of interference, e.g. without the use of channel coding. It must take into account their importance mentioned-above.

The transmission through the communication channel of all coefficients can be done sequentially: first all the coefficients of one level of decomposition, then the next and so on and in parallel (simultaneous transmission of sequences of coefficients at all levels of decomposition).

Unlike parallel transmission, serial transmission the whole tree of coefficients requires an increase of the data rate in proportion to the number and capacity of coefficients, which is not always possible. Also, in this case is more difficult to provide a inverse relationship mechanism between the probability of occurrence of errors in the transmission coefficient and its significance in the wavelet decomposition.

In the process of wavelet decomposition of the generated coefficients for the same period of time, the ratio of the scale factor changes twice between adjacent levels. Respectively, it is possible to consider the data received on the output of the decomposition circuit (algorithm), as several multi-speed channels.

2. Analysis of Recent Research and the Problem Statement

It has to be noted that the transfer of the wavelet coefficients as variable-speed flow, until this time was not considered. Possible variant of implementation can be the integration of individual channels in the OFDM system (using classic types of modulation in each) or wavelet packet modulation [7]–[11]. But the applied mathematical apparatus requires complex hardware and DSP processor. To transfer data from sensors using inexpensive systems their application would not be justified.

Let us consider possible ways of transmission of the variable-speed data stream, that is existing, widely used modulation types: Multilevel Phase Shift Keying (MPSK), Multilevel Frequency Shift Keying (MFSK), M-ary Quadrature Amplitude Shift Keying (MQASK) and compare them with the spline modulation proposed by the author in [12] with certain modifications:

- interpolation directly exposed fragments of sine waves (previously fragments of sine waves were interpolated using cubic Hermite splines),
- the value of the signal amplitude across different channels (previously scope of each of the channels was the same),
- the number of channels is 8 (compared with 2 channels discussed earlier [12]).

Note, that the analysis is carried out at baseband, or rather its complex envelope, not a band-pass signal that can be considered like equivalent to [13], although it requires less time spent on modeling.

3. The Wavelet Coefficients Transfer's Characteristics of the Classical Types of Modulation and Spline-modulation

3.1. Transfer by the Spline Signal

Block diagram of a multi-speed digital data transmission system, implemented by using cubic Hermit splines (as Nyquist pulse) is shown in Fig. 3. It uses 8 channels in it is eight and equal to the number of levels of the wavelet expansion coefficients.

Binary sequence x1-8, corresponding to the coefficients of detail wavelet decomposition from the generator coefficients of decomposition arrive at 8 interpolators, consisting of devices adding zeros (represented with arrows up) and filters with finite impulse response. The impulse response filter is a cubic spline samples.

Basic functions of local cubic Hermit spline $B(t)$ is a smooth function with continuous first derivative and allows interpolated value of a specific function $f(t)$ (in this case binary sequences in any one of the outputs of the expansion coefficients generator). The function $f(t)$ takes in time one of the two possible values 0 or 1.

The general formulas of the spline equations and four fragments of which it is composed are of the form:

$$B(t)=f(t_1)\cdot X_0(t)+f(t_2)\cdot X_1(t)+f'(t_1)\cdot X_2(t)+f'(t_2)\cdot X_3(t),$$

$$X_0(t)=\frac{2\cdot t^3-3\cdot t^2\cdot(t_1+t_2)+6\cdot t_1\cdot t_2\cdot t-t_2^2\cdot(3\cdot t_1-t_2)}{(t_1^2-2\cdot t_1\cdot t_2+t_2^2)\cdot(t_2-t_1)},$$

$$X_1(t)=\frac{(t-t_1)^2\cdot(2\cdot t+t_1-3\cdot t_2)}{(t_1-t_2)\cdot(t_1^2-2\cdot t_1\cdot t_2+t_2^2)},$$

$$X_2(t)=\frac{(t-t_1)\cdot(t^2-2\cdot t_2\cdot t+t_2^2)}{(t_1-t_2)^2},$$

$$X_3(t)=\frac{(t-t_1)^2\cdot(t-t_2)}{(t_1-t_2)^2},$$

where t – discrete times in the interval $[t_1,t_2]$; t_1,t_2 – time value of binary values occurrence at the output of the expansion coefficient generator (interpolation nodes); $f(t_1),f(t_2)$ – binary values at the output of the decomposition generator (the function values of the nodal points);

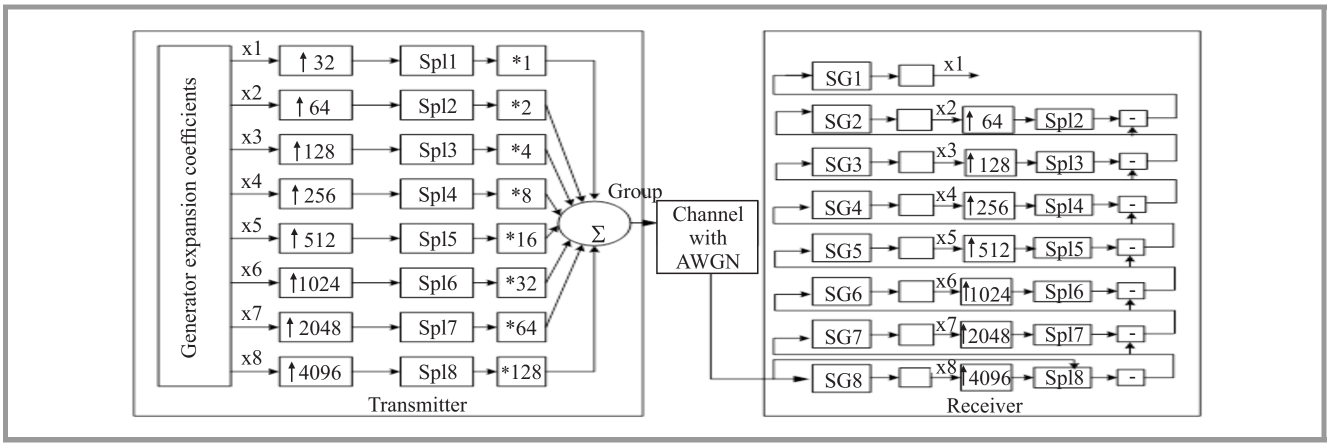


Fig. 3. Block diagram of an 8 multi-speed spline system of transmition of digital data.

$f'(t_1), f'(t_2)$ – derivatives of the function $f(t)$ in time moment t_1 and t_2 (defined as the difference between the current value of the function and the value of in the previous time of binary values occurrence at the output of the expansion coefficients generator). Figure 4 shows the value of the function $f(t)$ for the case when the signals at the output of the expansion coefficients generator at times $t_1 = 0$ s and $t_2 = 1$ s are equal respectively $f(0) = 1$ and $f(1) = 0$. The derivatives in these times are equal respectively $f'(0) = 0$ (assuming equal to zero the output state of the expansion coefficients generator until moment t_1), and $f'(1) = -1$.

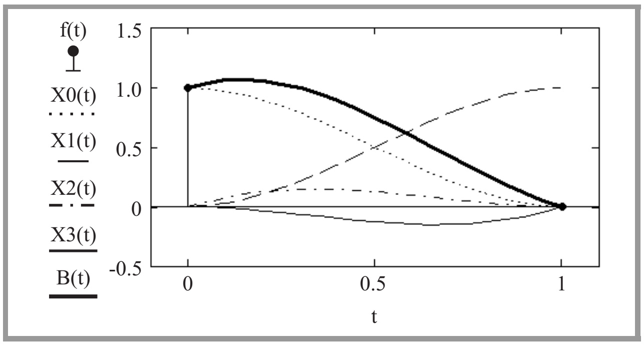


Fig. 4. The interpolation process of the binary sequence fragment using a spline.

The number of counts added to the interpolation process, is proportional to the interpolation factor as a multiple of two and is equal to $2^5, 2^6, 2^7, 2^8, 2^9, 2^{10}, 2^{11}, 2^{12}$ for channels respectively from the first to the eighth.

Interpolators equalize the number of samples in the signal of each channel for the same period and generate a corresponding channel's spectrum. The resulting smoothed binary signals (Fig. 4) are multiplied by the appropriate scale factors, shown as numbers with multiplication sign ahead, and fed to an adder at whose output a group signal appears (group). It is equal to the algebraic sum of the channels signals. The described elements are transmitted (transmitter).

The rate of binary sequences arrival at each level of decomposition increases with decreasing channel number.

The spectrums of video signals corresponding to the coefficients of the expansion, are shown in Fig. 6 above. The spectra of these signals after the spline interpolation, for each of the channels are shown in Fig. 6 below.

Detailing of the baseband signal into the transmission channels from the first to the eighth 1024, 512, 256, 128, 64, 32, 16 and 8 bits respectively, and its energy spectrum is shown in Fig. 7a from above, and below is shown the corresponding energy spectrum.

Figure 7a shows that the energy of the baseband signal due to the uneven capacity of each separate channels (due to scaling factors), focuses in the band the least speed channel. However, during transmission, the bandwidth should be limited to the first zero of the energy spectrum lobe highest speed channel, to store information in all channels. Group signal, after passing through the channel (channel with AWGN) in which it is added to the Gaussian noise arrives to the receiver (receiver). At the receiver group signal is in the process of decomposition to the individual channel signals, starting with the least-speed channel (Channel 8). Next, the signal passes through the spline filter Savitzky-Golay eighth channel (SG8) (filter Savitzky-Golay with the basis spline). This filter generates an estimate of the transmitted signal form of the eighth channel given by:

$$\widehat{S8} = (\mathrm{P8}^T \cdot \mathrm{P8})^{-1} \cdot \mathrm{P8}^T \cdot G\_N,$$

where $\widehat{S8}$ – evaluation form transmitted in the eighth channel signal, P8 – planning matrix composed of four sample's fragments of the spline, and G_N – vector samples of mix group signal and noise (the values that come with the block channel with AWGN).

Based on the obtained evaluation of the transmitted signal's form, a threshold device is used (hard decisions) determining the evaluation of binary values (zero or unity) transmitted in the eighth channel, which is input to the eighth channel's spline interpolator, identical to that used in the transmitter.

The restored copy of the eighth channel signal is subtracted from the signal G_N. The difference is input to the next spline filter Savitzky-Golay (SG7), where the procedure is

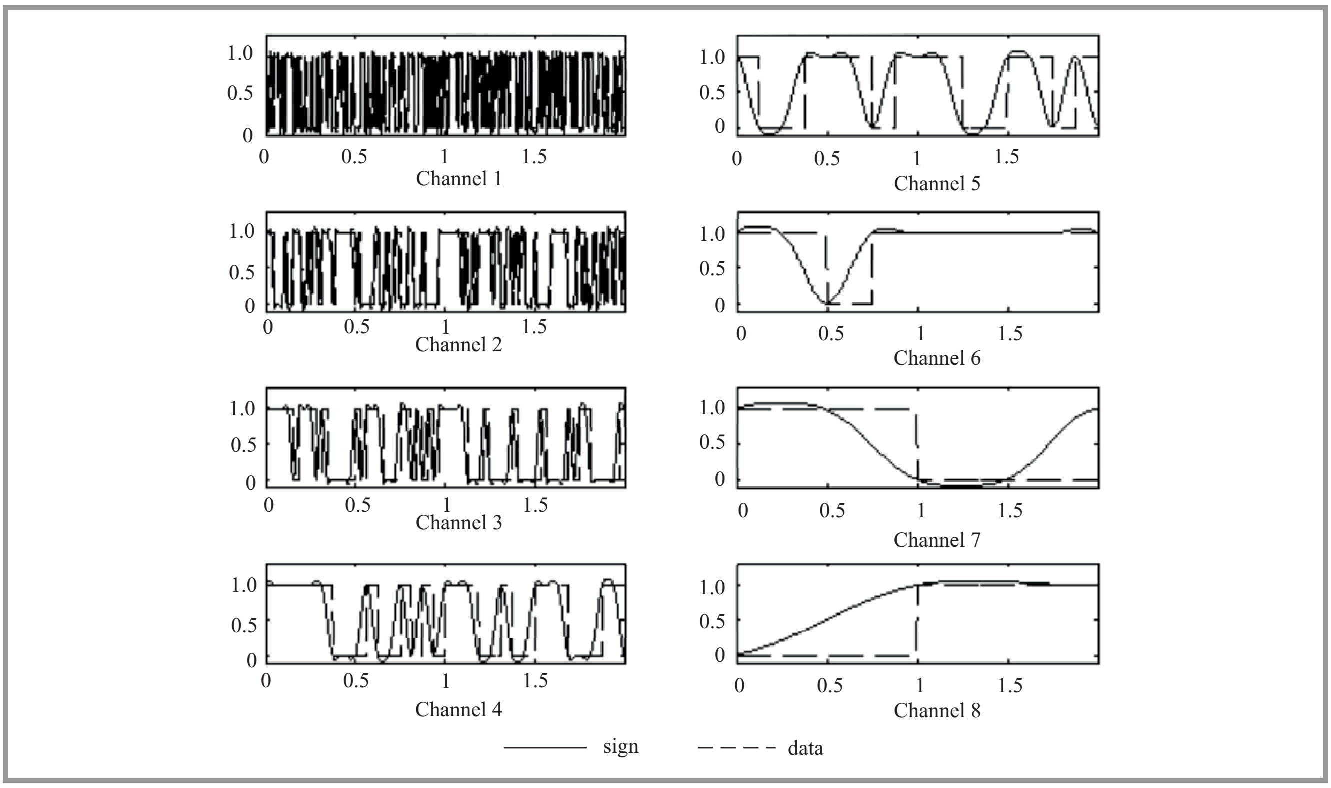


Fig. 5. Timing diagrams of signals at the output of the expansion coefficients generator (data) and spline interpolators (sign), for each of the 8 channels (Ch1–Ch8).

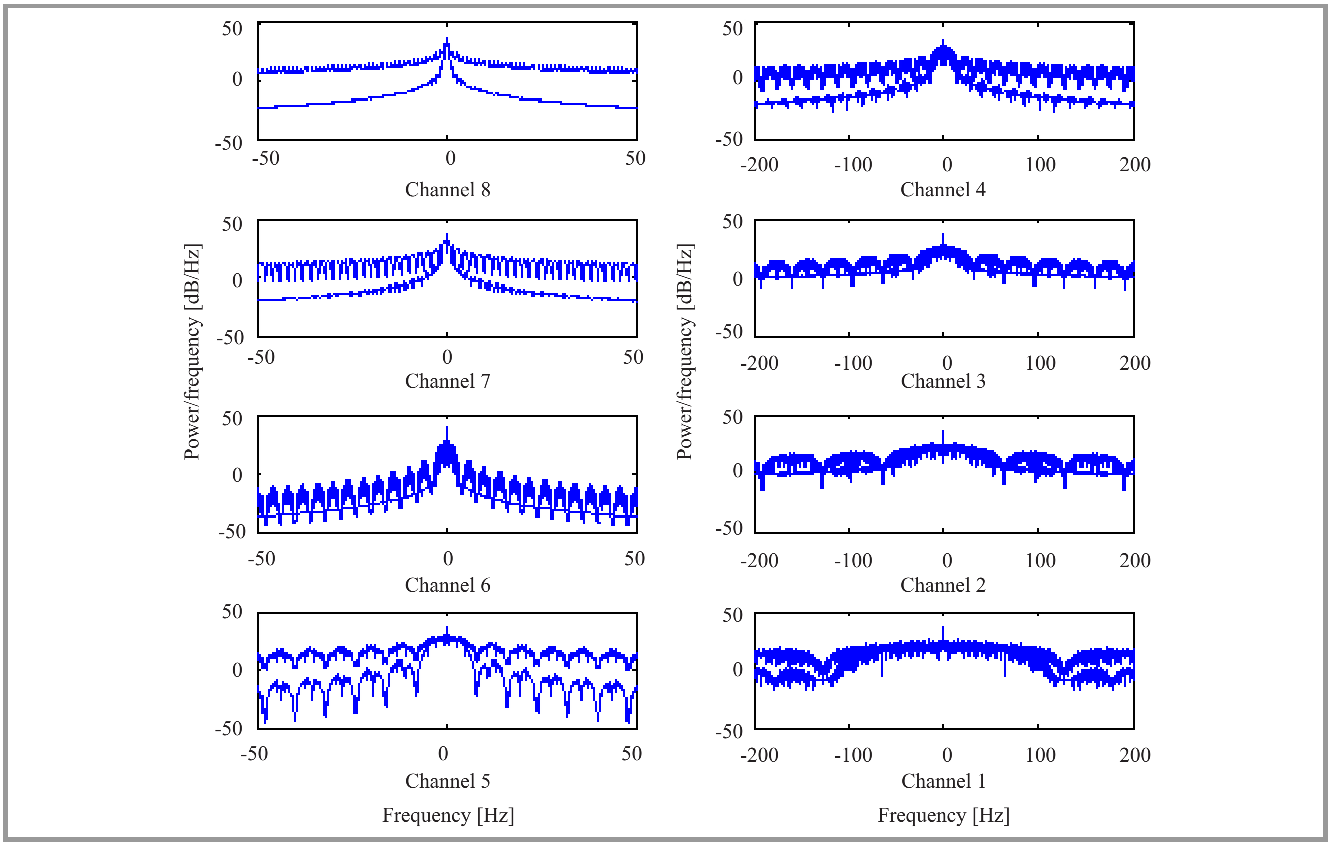


Fig. 6. The energy spectrum of binary sequences (for all the channels is pictured above) and spline signals of each channel (for all channels is pictured below).

repeated. The only difference is in the planning matrix, which is composed of four fragment of spline seventh channel and the signal difference *G_N*. *G_N* replaced and restored copies of the 8 signal channel. Thus, binary values are defined in all 8 channels of the system.

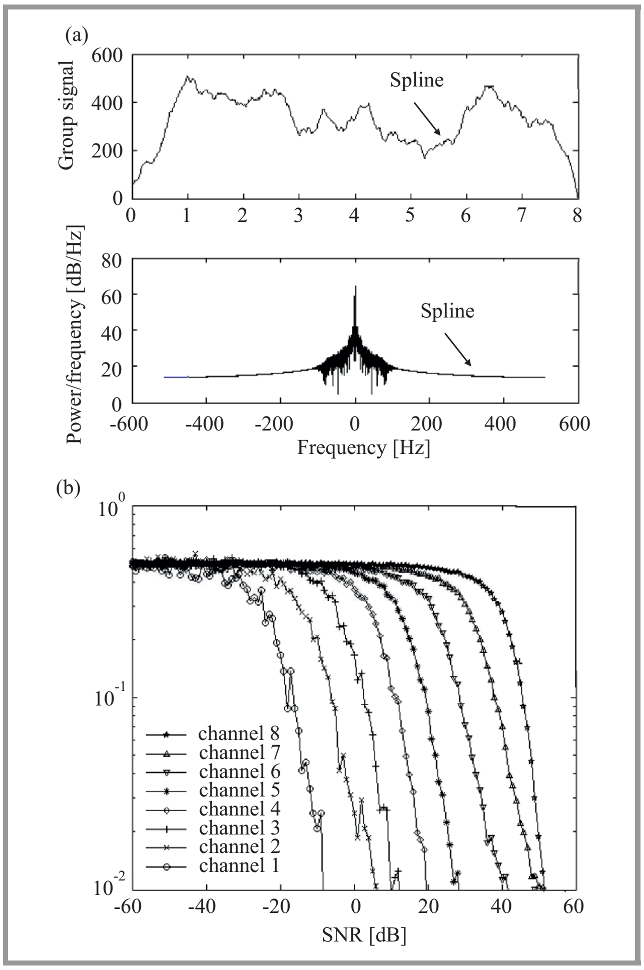


Fig. 7. Spline signal and its characteristics: (a) the baseband signal (top) and its energy spectrum (bottom), (b) the dependence of the probability error on the SNR for the data transmission system using spline modulation.

On the basis of ten measurements is estimated averaged error probability to each of the eight channels according on the signal to noise ratio (SNR), range from –60 to 60 dB, depicted in Fig. 7b. It shows that the probability of error in each channel are not the same and increases with speed. It is the predictable reaction to a different power level and bandwidth of each channel. The presence of local peaks of errors similar to multiple channels is due to an error in lower-speed channel that was not correctly restored and compensated in a certain period of time, resulting in incorrect identification bit (bits) in it. This feature may be reduced by identifying higher average level of the signal at the next channel's input (in this embodiment, this feature still needs more research), or application code protection from errors.

3.2. Transfer of MFSK

Consider the transmission of the same data using MFSK modulation with minimum shift. To do this, each of the binary sequences of wavelet decomposition coefficients was seen as a change in the individual bits of the binary bit character with 256-character alphabet. To narrow signal's spectrum, the most high-speed channel corresponds to a smaller value for the discharge character, as shown in Fig. 8.

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
x8	x7	x6	x5	x4	x3	x2	x1

Fig. 8. Distribution of the expansion coefficients according to categories of characters.

Because the signal changes of x8 are 128 times slower than x1, its value stored (interpolated one and the same value) for transmittion of 128 bits of x1 channel. Similarly, for channel x7 (but for transfer 64 bits of x1 channel) and so on.

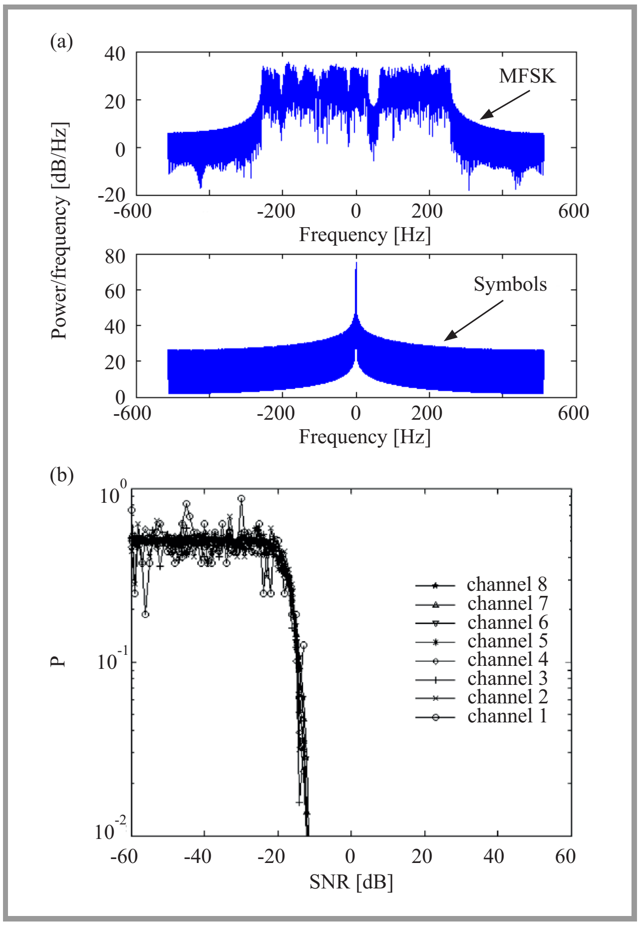


Fig. 9. Characteristics MFSK: (a) the energy spectrum of MFSK (top) and a sequence of characters (bottom), (b) dependence of the error probability on signal to noise ratio for a data transmission system with help of minimum shift MFSK.

The above-described feature of the wavelet coefficients representation as the data will be preserved when MPSK and MQASK are used.

Figure 9a depicts the energy spectra of MFSK and signal as a sequence of characters having the same sampling rate (1024 Hz) and the number of transmitted bits in the channels on the splines schema. As can be seen the width of the MFSK signal's spectrum is n times greater (where n is the number of the alphabet's characters) than the spectrum's width of the symbols sequence, which in turn is determined by the change's rate of bits in the symbol of the most high-speed channel.

The dependence of the error probability on the signal to noise ratio is shown in Fig. 9b. In spite of some bursts in the most high-speed channel, the main form of dependence is almost the same for all channels. Note characteristic for frequency modulated signals resistance to Gaussian noise.

3.3. Transfer of MPSK

Consider the transmission characteristics compared to similar signal MPSK modulation. The spectral width of the signal by the zeros of the first lobe in the range of 256 Hz corresponds to double value of the transfer rate of the fastest channel, 128 bit/s (Fig. 10a).

The dependence of the error probability on the signal to noise ratio between the channels is characterized by uneven curve, as in the spline-system, with decreasing reliability at increasing bit rate (Fig. 10b).

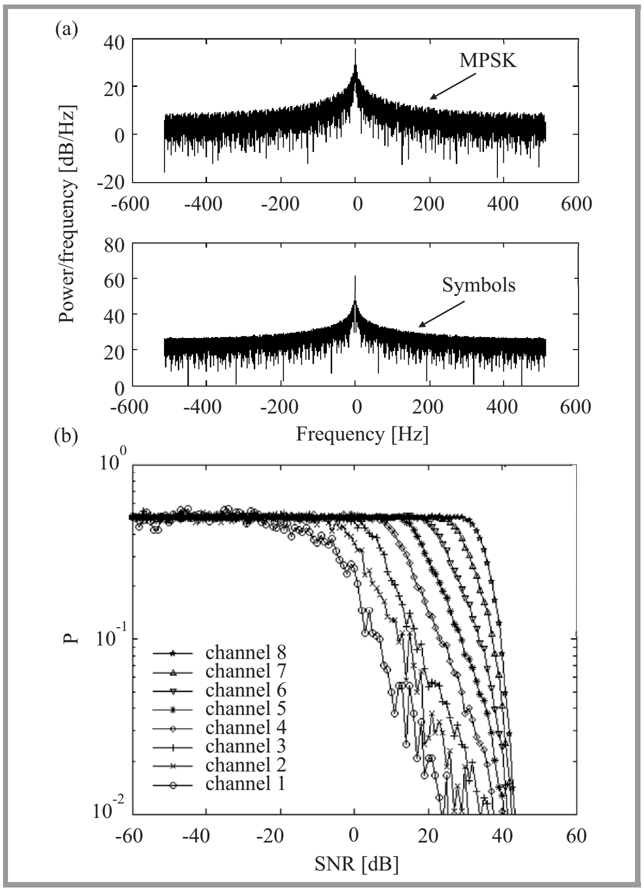


Fig. 10. Characteristics of MPSK: (a) the energy spectrum of MPSK (top) and a sequence of characters (bottom), (b) the dependence of the probability of error on the signal to noise ratio for the data transmission system with help of MPSK minimum shift.

3.4. Transfer of MQASK

As for the signal modulation MQASK, its spectrum has the same features as that of MPSK (Fig. 11a), but the dependence of an error probability is specific. Channel pairs 1-5, 2-6, 3-7, 4-8 show almost the same dependence on the signal to noise ratio (Fig. 11b).

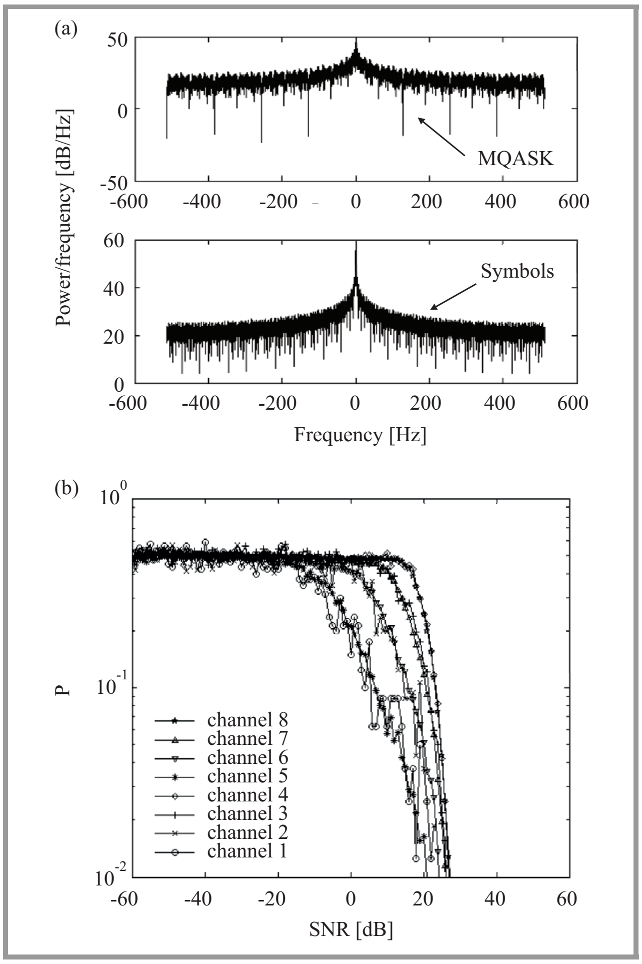


Fig. 11. Features: (a) power spectrum MQASK (top) and a sequence of characters (bottom), (b) the dependence of the error probability on the signal to noise ratio for the data transmission system with help of MQASK.

4. Conclusions

All the considered modulation schemes allow to arrange the values of the decomposition's coefficients, but only in the spline and MPSK modulation during the transmission of the coefficients remains the dependence of their validity and reliability. In this case the spectral width of the first lobe corresponds to the speed of transmission in most high-speed channel. For MQASK modulation it is observed the irregular dependence of the error probability between the channels, and the spectral width equal to the width of the MPSK spectrum.

The MFSK modulation transmits all coefficients with the same error probability, and requires significantly greater bandwidth than all the other methods of modulation.

In many cases, when transmitting signals from the direct source of information to the receiver in the data collection systems it is rational to simplify the scheme of the transmitter and complexity of the receiver. The use of classical modulation methods for transmission of wavelet coefficients requires additional use of the quadrature modulator, which typically represents a separate functional unit that complicates and increases the cost of the transmitter. Group spline signal can be generated by means of pulse-width modulator, which is common, even in low-cost types of controllers.

Together with built in it analog to digital converter, and a fast algorithm for multiresolution wavelet decomposition, it forms an element of data collection, as suggested in [14] for transmission on two-wire or coaxial cable.

References

[1] M. Akay, "Wavelet applications in medicine", *IEEE Spectrum*, vol. 34, no. 5, pp. 50–56, 1997.

[2] G. Evangelista "Comb and multiplexed wavelet transforms and their applications to signal processing". *IEEE Trans. Sig. Proces.*, vol. 42, no. 2, pp. 292–303, 1994.

[3] F. Bömers, "Wavelets in real-time digital audio processing: analysis and sample implementations", M.Sc. thesis, Universitat Mannheim, 2000 [Online]. Available: http://www.boemers.com/personal/thesis.pdf

[4] D. V. Beilekchi, "Model' ustroistva peredachi tsifrovogo rechevogo signala po kanalu svyazi", *Metody i Ustroistva Peredachi i Obrabotki Informatsii*", no. 1, pp. 83–89, 2007 (in Russian) [Online]. Available: http://rts-md.com/docs/archives/MIU_07_1/chapter%204.pdf

[5] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. Orlando, FL, USA: Academic Press, 1992.

[6] T. Edwards, "Discrete wavelet transforms: theory and implementation", *Stanford University*, CA, USA, 1992.

[7] A. Jamin and P. Mähönen, "Wavelet packet modulation for wireless communications", *J. Wirel. Commun. & Mob. Comput.*, vol. 5, no. 2, pp. 123–137, 2005.

[8] J. Monera-Llorca and W.-S Lu, "An improved wavelet-packet-division multiple access system", in *Proc. IEEE Pacific Rim Conf. on Comp. and Sig. PACRIM 1999*, Victoria, BC, Canada, 1999, pp. 495–498.

[9] W. Z. Zhong and Q. Guo, "Performance evaluation of wavelet packet modulation over mobile satellite channel", *Inform. Technol. J.*, vol. 8, no. 3, pp. 310–317, 2009.

[10] M. C. Gill, "Coded-waveform design for high speed data transfer over high frequency radio channels", Ph.D. thesis, School of Electronic Engineering, University of South Australia, 1998.

[11] N. Nikolov and Z. Nikolov, "A communication system with wavelet packet division multiplexing in an environment of White Gaussian Noise and narrow-band interference," *Cybernet. & Inform. Technol.*, vol. 5, no. 1, pp. 100–114, 2005.

[12] A. I. Kutin, "Mnogokanal'naya sistema peredachi dannykh v splainovykh bazisakh (Multi-channel data transmission system in the spline bases)", *Vesnik Natsional'nogo universiteta "L'vovs'kaya politekhnika", Radioelektronika i Telekommunikatsii (Radio electronics and Telecommunications)*, no. 796, pp. 75–82, 2014 (in Ukrainian).

[13] B. Sklar, *Digital Communications: Fundamentals and Applications*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[14] R. Stojanovic and S. Knezevic, "Optimization and implementation of the wavelet based algorithms for embedded biomedical signal processing", *Comp. Sci. & Inform. Sys.*, vol. 10, no. 1, pp. 503–523, 2013.

Artem I. Kutin received his M.Sc. in 2009 from the direction of radio engineering and television systems Institute of Air Navigation of National Aviation University (NAU), Kiev, Ukraine. He is currently a lecturer in electronics Krivoy Rog college NAU. He is doing research in the use of splines for data transmission in wired and cordless communication systems.

E-mail: art-kutin@yandex.ru
Radio Engineering Faculty
Institute of Air Navigation
National Aviation University
Kosmonavta Komarova 1
03058 Kiev, Ukraine

Paper

Iterative Algorithm for Threshold Calculation in the Problem of Routing Fixed Size Jobs to Two Parallel Servers

Mikhail Konovalov and Rostislav Razumchik

Institute of Informatics Problems, Federal Research Center "Computer Science and Control", Russian Academy of Sciences, Moscow, Russia

Abstract—At present, solutions of many practical problems require significant computational resources and systems (grids, clouds, clusters etc.), which provide appropriate means are constantly evolving. The capability of the systems to fulfil quality of service requirements pose new challenges for the developers. One of the well-known approaches to increase system performance is the use of optimal scheduling (dispatching) policies. In this paper the special case of the general problem of finding optimal allocation policy in the heterogeneous *n*-server system processing fixed size jobs is considered. There are two servers working independently at constant but different speeds. Each of them has a dedicated queue (of infinite capacity) in front of it. Jobs of equal size arrive at the system. Inter-arrival times are i.i.d. random variables with general distribution with finite mean. Each job upon arrival must be immediately dispatched to one of the two queues wherefrom it will be served in FCFS manner (no pre-emption). The objective is the minimization of mean job sojourn time in the system. It is known that under this objective the optimal policy is of threshold type. The authors propose scalable fast iterative non-simulation algorithm for approximate calculation of the policy parameter (threshold). Numerical results are given.

Keywords—continuous MDP, discretization, job allocation, optimal policy, threshold.

1. Introduction

For high-performance processing systems, consisting of several servers working independently and in parallel one of the fundamental problems is the problem of optimal allocation (or routing) of arriving jobs. Allocation happens at instants of each job arrival and means that job is assigned to one of the servers where it will be served. This decision cannot be undone later. It is assumed that each server has a dedicated queue of infinite capacity where jobs assigned to this server can wait for service.

The optimal allocation (or optimal policy) is the one which provides optimal value of the value function. As the example of simple (but sometimes difficult to compute) value function one can imagine mean sojourn time in the system, tail of the sojourn time distribution. The optimal policy typically depends on value function, service discipline (FIFO, LIFO, PS, etc.) and on the amount of information about the state of the system, which is available at decision instants. One can identify are three main approaches for finding optimal policy for the type of problems described above. The first approach is to choose, based on preliminary qualitative system analysis, the most "promising" policy and then to check the "degree" of its optimality. According to the second approach one chooses the parametrized policy (for example, SITA policy in [1], [2]), then finds the value function under this policy and estimates the values of the policy parameters which provide optimal value of the value function. The third relied on ideas from Markov decision processes and is used in many jobs and resource allocation problems (see for example [3]–[9]). In the majority of the problems the system state space is very complex (for example, due to the need to track elapsed/remaining service times, allow infinite storage capacities, etc.). Thus the class of considered policies is usually reduced to static policies which allow sometimes decomposition of the system and its study in component-wise manner. The are also policies which allow look-ahead actions and still tractable solution (see, for example, [10]).

The problem of finding optimal allocation policy in a heterogeneous two-server system processing fixed size jobs, which is the subject of this paper, has already been considered before and the apparently latest results appear in [11]. In [11] the flow of jobs is Poisson and jobs are served in FCFS manner from queues. The objective is minimization of mean sojourn time in the system. Authors show that this problem is related to the well-known slow-server-problem ([11], [12]). From this observation they derive the following result: optimal allocation policy is of threshold type with one threshold i.e. if upon arrival of the job the amount of unfinished work at faster server (plus total work in its queue) minus the amount of unfinished work at slower server (plus total work in its queue) exceed the threshold value, job is allocated to slow server. The simplicity of the problem formulation and the known (but nonconstructive) answer makes even more sticking the fact that its analytic solution is not known: one can determine the threshold value only using numerical methods. In [11] authors provide one of such methods based on Markov decision processes and Monte-Carlo simulation and also provide several heuristic policies which show near optimal results for the wide range of initial system's values.

The use of Monte-Carlo simulation for the threshold value estimation in the considered problem is greatly complicated by the fact that the curvature of the value function in the neighborhood of its minimum is very low and thus it requires very long simulation time in order to achieve high accuracy.

In this paper a new method for estimation of the threshold value of the optimal policy is provided, which does not rely on any simulation results and is based only on probabilistic arguments and properties of threshold policy. In this respect from one point of view it is free from disadvantages inherent to simulation methods (like those in [11]) and from the other point of view it serves as a case study of efficient handling of Markov decision process problem with continuous state space by discretization.

The paper is organized as follows. In the Section 2 the description of the system is given and the question under study is formulated. Section 3 is devoted to detailed description of the solution method and in Section 4 some numerical results are presented. In conclusion, obtained results are briefly discussed.

2. Description of the System and Problem Formulation

Consider heterogeneous dispatching system with two parallel servers processing fixed size jobs. Jobs inter-arrival times are i.i.d. random variables with known distribution function $F(x)$ with finite mean. Servers are working independently and at constant rates: service rate of one server equals 1 and of the other equals $\nu > 1$. Henceforth, the server working at rate 1 will be referred as server I and to the server working at rate ν as server II. Clearly, time it takes server I and server II to complete one job equals 1 and ν^{-1} respectively. Each server has its own queue (of infinite capacity) and arriving job must be immediately upon arrival assigned (or routed) to one of the queues wherefrom it will be served. For the sake of brevity in what follows authors will refer to the decision to route a job to the queue in front of server I or server II by saying that action 1 or 2 was chosen. No jockeying between queues is allowed. Each server serves jobs only from its own queue on a first-come-first-served basis. Pre-emption is not allowed. The objective is to find the sequence of actions that minimizes mean job sojourn time in the system[1]. It is known that such sequence of actions is fully described by threshold-type policy (see details in, for example [11]). The most interesting is the non-simulation estimation of the value of the policy parameter, i.e. threshold value.

Let us denote by x current workload at server I which equals the number of jobs in the queue in front of server I plus the remaining service of the job in server I. Current workload at server II is denoted by y. Following queueing theory terminology x and y can be understood as virtual waiting times. The evolution of the system in time is fully described by changing values of the pair (x,y) with state space $S = \{(x,y),\ x \in [0,\infty),\ y \in [0,\infty)\}$.

Assume that upon arrival of a job the system is in the state $s = (x,y) \in S$. At this time instant the job must be routed to one of the two queues. If the job is routed to queue in front of the server I (i.e. action 1 is chosen), then at time instant of the next job arrival system's state will be s' equal to

$$s' = ((x+1-\tau)^+, (y-\nu\tau)^+),$$

where τ is the time until next job arrival and $a^+ = \max(0,a)$. The set of states to which transitions from state $s = (x,y)$ can occur is $A_1(s) = \{(x',y'),\ x' = (x+1-t)^+, y' = (y-\nu t)^+,\ t \geq 0\}$. The probability distribution that governs these transitions is denoted by $\mathbf{P}_1(s'|s)$, $s' \in A_1(s)$, $s \in S$. Note that given the distribution $F(x)$ of inter-arrival times, the distribution $\mathbf{P}_1(s'|s)$ can be calculated in straightforward manner.

In case the job is routed to server II (i.e. action 2 is chosen), then at time instant of the next job arrival the state of the system will be s' equal to

$$s' = ((x-\tau)^+, (y+1-\nu\tau)^+).$$

When action 2 is chosen the set of states to which transitions from state $s = (x,y)$ can occur is $A_2(s) = \{(x',y'),\ x' = (x-t)^+,\ y' = (y+1-\nu t)^+,\ t \geq 0\}$. Probability distribution that governs such transitions is denoted by $\mathbf{P}_2(s'|s)$, $s' \in A_2(s)$, $s \in S$. It can be calculated just like $\mathbf{P}_1(s'|s)$.

For fixed s both sets $A_1(s)$ and $A_2(s)$ are one-dimensional. Specifically, each of them is the composition of two line segments: one segment is part of the line with slope ν going through point (x,y) between point (x,y) and intersection of the line with one of the coordinate axes (segment AB in Fig. 1) and the other segment is part of the line from the intersection to point $(0,0)$ (segment OA in Fig. 1).

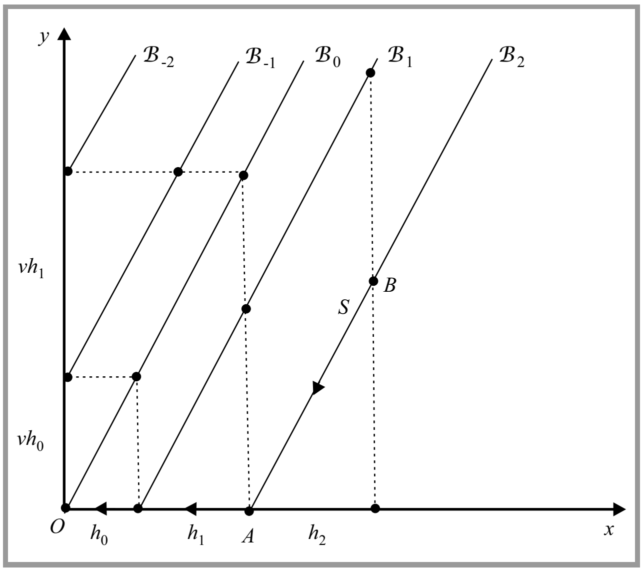


Fig. 1. Discretization of the state space.

[1] Sojourn time for a given job starts from the instant when it arrives at a queue and stops when its service is completed. It is assumed that the decision process does not incur any delay.

Let τ_n, $n \geq 1$, be the arrival instant of the n-th job. Denote by s_n the state of the system at time τ_n but *before* any action is chosen, i.e. before decision where to route the arrived job is made. The authors assume that threshold-type policy is implemented in the system. This implies that at time τ_n one must choose action 1 if

$$s_n \in S^{\xi} = \{(x,y) : \frac{y}{\nu} - x > \xi\} \tag{1}$$

and one must choose action 2 if

$$s_n \in \overline{S}^{\xi} = S \backslash S^{\xi} = \{(x,y) : \frac{y}{\nu} - x \leq \xi\}. \tag{2}$$

Here $\xi \geq 0$ is the parameter of the policy (i.e. threshold value). For detailed discussion of this threshold-type policy one can refer to [11]. Given initial system's state, say $s_1 = (0,0)$, the sequence of s_n, $n \geq 1$, constitutes the Markov chain with transition probabilities

$$\mathbf{P}_{\xi}(s'|s) = \begin{cases} \mathbf{P}_1(s'|s), & \text{if} \quad s \in S^{\xi}, \\ \mathbf{P}_2(s'|s), & \text{if} \quad s \in \overline{S}^{\xi}. \end{cases}$$

Denote the stationary distribution of this Markov chain by π_{ξ}. Given sufficient condition for the stability of the system $(\lambda/(1+\nu)) < 1$, where $\lambda = (\int xF(x))^{-1}$, is satisfied, the stationary distribution exists.

With each state of the Markov chain s_n, $n \geq 1$, one can associate a "reward" $g_n(s_n)$ equal to the sojourn time of the n-th job in the system. If $s_n = (x,y)$ then, due to the fact that jobs are served from queues on FCFS basis, we have

$$g_n(s_n) = \begin{cases} x+1, & \text{if} \quad s_n \in S^{\xi}, \\ \frac{y+1}{\nu}, & \text{if} \quad s_n \in \overline{S}^{\xi}. \end{cases}$$

The limiting expected reward or limiting expected sojourn time T_{ξ} in the system can be defined as

$$T_{\xi} = \int_S g(u)\pi_{\xi}(du).$$

As mentioned above, the most interested is the non-simulation estimation of the value ξ, which minimizes the value of T_{ξ}. Despite the fact that the value function depends only one parameter, the analytical solution of the optimization problem is not known. To author's knowledge there are no analytical results concerning the exact expression and properties of T_{ξ} such as monotonicity, concavity, unimodality, differentiability, which makes impossible the application of standard optimization methods. One of the main solution approaches is the use of simulation in conjunction with ideas of Markov decision processes. This was done in [11], where authors have thoroughly studied the behavior of T_{ξ} experimentally and proposed method for the estimation of the threshold value ξ. But the problem of minimization of T_{ξ} basing only on system's initial parameters ($F(x)$ and ν) without the use of simulation remains open and in the next section fast iterative algorithm is provided, which allows one to find solution with prescribed accuracy.

3. Iterative Algorithm

The idea of the iterative algorithm for computation of the approximate value of the optimal threshold is based on the following observation. Assume the system is in state $\hat{s} = (\hat{x},\hat{y})$ such that $\frac{\hat{y}}{\nu} - \hat{x} = \xi_{opt}$, where ξ_{opt} is the optimal (still unknown) threshold value. Then the threshold policy introduced in the previous section tells us that action 2 must be chosen. But in fact it is irrelevant, which action one chooses when system is in the state $\hat{s}$. Otherwise the current threshold value is not the true optimal value, because we have to prefer one action to another (and thus the threshold value must be shifted and the value of the value function will be improved[2]). As we don't know the optimal threshold value we fix (almost) arbitrary value $\xi \geq 0$ and assume the system is in the state $\hat{s} = (\hat{x},\hat{y})$ such that $\frac{\hat{y}}{\nu} - \hat{x} = \xi$. In state $\hat{s}$ two actions can be chosen. Denote by $\sigma_{\xi}^{(1)}$ the policy that chooses action 1 and then follows Eqs. (1)–(2) rule. The policy that at first chooses action 2 and then also follows Eqs. (1)–(2) rule denote by $\sigma_{\xi}^{(2)}$. Let us compare $\sigma_{\xi}^{(1)}$ and $\sigma_{\xi}^{(2)}$. Consider the difference

$$\Delta_{\xi} = g^{(1)} - g^{(2)} + \sum_{n=1}^{\infty} \left(\int_S g(u)\pi_n^{(1)}(du) - \int_S g(u)\pi_n^{(2)}(du) \right), \tag{3}$$

where $g^{(1)}$ and $g^{(2)}$ are rewards for the first action when system is in state $\hat{s}$, and $\pi_n^{(i)}$ is stationary distribution at n-th step of Markov chain (corresponding to fixed value of ξ) given that first action was i, $i = 1,2$. From definition of strategies $\sigma_{\xi}^{(i)}$ it follows that

$$g^{(1)} = \hat{x} + 1, \quad g^{(2)} = \frac{\hat{y}+1}{\nu}. \tag{4}$$

Thus $g^{(1)} - g^{(2)} \neq 0$ and the other terms in (3) are non-zero because the distibutions $\pi_n^{(1)}$ and $\pi_n^{(2)}$ are different for any n. But due to the fact that $\lim_{n\to\infty} \pi_n^{(1)} = \lim_{n\to\infty} \pi_n^{(2)} = \pi_{\xi}$ at exponential rate, the sum in Eq. (3) converges. If $\Delta_{\xi} = 0$ then the value of ξ is the value of the optimal threshold. If $\Delta_{\xi} \neq 0$ then the value of ξ must be increased or decreased depending on the sign of Δ_{ξ}.

The implementation of this idea heavily depends on the opportunity to compute distributions $\pi_n^{(i)}$. The obvious approach is to approximate Markov chain $\{s_n, n \geq 1\}$ by a finite-state Markov chain with transition probability matrix P and use the relation $\pi_n = \pi_{n-1}\mathrm{P}$. If one partitions state space by equal rectangles then the cost to compute with such an approach becomes too high. As experiments show the curvature of function T_{ξ} in the neighborhood of its minimum is very low and thus, in order to obtain suitable results, one has to use very high level of discretization. Eventually matrix P becomes too big (storage requirements become too high) making impossible to use relation $\pi_n = \pi_{n-1}\mathrm{P}$. In the next subsection a new discretization method based on non-uniform grid spacing, which does

[2]Here is implicitly assumed that T_{ξ} is a continuous function of ξ.

not require the calculation of matrix P and gives accurate results is proposed.

3.1. Discretization of the State Space and Construction of Approximating Finite-State Markov Chain

Before constructing finite-state Markov chain $\{\hat{s}_n, n \geq 1\}$, which approximates Markov chain $\{s_n, n \geq 1\}$ discretization of state space S must be performed.
In order to do this some notation need to be introduced. Denote by h_i be the sequence of numbers

$$h_i = h_0(1+\alpha)^i, \; i = 0, 1, \ldots, \tag{5}$$

where $h_0 > 0$ and $\alpha > 0$ are arbitrary small numbers and introduce the following sets:

$$\mathcal{B}_i = \{(x,y) : y = \nu(x - a_i)\}, \; i = 0, \pm 1, \pm 2, \ldots, L,$$

where L is arbitrary big whole number and

$$a_i = \begin{cases} 0, & \text{if } \; i = 0, \\ a_{i-1} + h_{i-1}, & \text{if } \; i > 0, \\ a_{i+1} - h_{-i-1}, & \text{if } \; i < 0. \end{cases}$$

The sets $\mathcal{B}_i$ are straight lines with slope ν shifted along the x-axis. Denote also by $\mathcal{C}_j^+$ and $\mathcal{C}_j^-$ the following sets:

$$\begin{aligned} \mathcal{C}_j^+ &= \{(x,y) : x = a_j\}, \; j = 1, 2, \ldots, L, \\ \mathcal{C}_j^- &= \{(x,y) : y = \nu a_j\}, \; j = 1, 2, \ldots, L. \end{aligned}$$

Define the set of points $\tilde{S}^{\alpha,L}$ as union of the following sets

$$\tilde{S}^{\alpha,L} = \{\tilde{S}_{00}\} \cup \{\tilde{S}_{ij}, \; i = 0, \pm 1, \pm 2, \ldots, L; j = 1, 2, \ldots, L\},$$

where $\tilde{S}_{00} = (0,0)$, $\tilde{S}_{ij} = \mathcal{B}_i \cap \mathcal{C}_j^+$ if $i \geq 0$ and $\tilde{S}_{ij} = \mathcal{B}_i \cap \mathcal{C}_j^-$ if $i < 0$. The set of points $\tilde{S}^{\alpha,L}$ consists of $(L+1)^2$ points and represents the grid, which covers the rectangle area of the first quadrant of the xy-plane. One vertex of the rectangle coincides with $(0,0)$ and sides along the x-axis and y-axis equal H and νH respectively, where

$$H = \sum_{i=0}^{L-1} h_i = \frac{h_0(1+\alpha)^L - h_0}{\alpha}. \tag{6}$$

The points in the set $\tilde{S}^{\alpha,L}$ are distributed non-uniformly (Fig. 1). As one moves towards the origin and line $y = \nu x$ (set $\mathcal{B}_0$) the concentration increases. As one move in the opposite direction the concentration goes down.
The set of points $\tilde{S}^{\alpha,L}$ is used to construct state space of approximating finite-state Markov chain $\{\hat{s}_n, n \geq 1\}$. The following argumentation follows from the description of the sets $A_1(s)$ and $A_2(s)$ given in the previous section.
Assume that *after* decision on the n-th step Markov chain $\{s_n, n \geq 1\}$ was in state $s = (x,y) \in \mathcal{B}_i$ (see Fig. 1). Then until the arrival instant of the next job system's state, i.e. values of pair (x,y) will "belong" to the line indicated with the arrow in Fig. 1. The start point of the route is s and finish point is $A = (0,0)$ which means that system is empty. By the arrival of $(n+1)$-th customer the system may be at any point *only* on this route. As the continuous state space of Markov chain $\{s_n, n \geq 1\}$ is discretized, then this route must consist of finite number of points.

Remark 1. The way in which the grid $\tilde{S}^{\alpha,L}$ was constructed tells that the length of any segment (either vertical or horizontal, or slanted) of arbitrary route equals h_i. It can easily be seen that time it takes system to pass a segment also equals h_i.
Now one needs to define the set of possible routes. In order to do this the following sets are defined:

$$\mathcal{A}_0 = \mathcal{B}_0 \cap \tilde{S}^{\alpha,L}, \tag{7}$$

$$\mathcal{A}_i = \left(\mathcal{B}_i \cap \tilde{S}^{\alpha,L}\right) \cup \{\tilde{S}_{i-1,i-1}, \ldots, \tilde{S}_{1,1}, \tilde{S}_{0,0}\}, \; i > 0, \tag{8}$$

$$\mathcal{A}_i = \left(\mathcal{B}_i \cap \tilde{S}^{\alpha,L}\right) \cup \{\tilde{S}_{i+1,-i-1}, \ldots, \tilde{S}_{-1,1}, \tilde{S}_{0,0}\}, \; i < 0. \tag{9}$$

In Eqs. (8)–(9) each set in parentheses contains points from the set $\tilde{S}^{\alpha,L}$ which belong to slanted segment $\mathcal{B}_i$. The set in braces contains points of the line connecting origin O and intersection of $\mathcal{B}_i$ with one of the coordinate axes. The routes $\mathcal{A}_i$ can be also represented in a different way:

$$\begin{aligned} \mathcal{A}_0 &= \{\tilde{S}_{0,L}, \tilde{S}_{0,L-1}, \ldots, \tilde{S}_{0,0}\}, \\ \mathcal{A}_i &= \{\tilde{S}_{i,L}, \tilde{S}_{i,L-1}, \ldots, \tilde{S}_{i,i}, \tilde{S}_{i-1,i-1}, \ldots, \tilde{S}_{1,1}, \tilde{S}_{0,0}\}, \; i > 0, \\ \mathcal{A}_i &= \{\tilde{S}_{i,L}, \tilde{S}_{i,L-1}, \ldots, \tilde{S}_{i,-i}, \tilde{S}_{i+1,-i-1}, \ldots, \tilde{S}_{-1,1}, \tilde{S}_{0,0}\}, \; i < 0. \end{aligned}$$

Any discretized route, just like OAB depicted in Fig. 1, is the subset of $\mathcal{A}_i$. The elements of $\mathcal{A}_i$ can be enumerated in a natural way, starting from point $\tilde{S}_{0,0}$. For $i \geq 0$ it holds that

$$\begin{aligned} &S_{i,0} = \tilde{S}_{0,0}, S_{i,1} = \tilde{S}_{1,1}, \ldots, S_{i,i-1} = \tilde{S}_{i-1,i-1}, \\ &S_{i,i} = \tilde{S}_{i,i}, \ldots S_{i,L-1} = \tilde{S}_{i,L-1}, S_{i,L} = \tilde{S}_{i,L}, \end{aligned}$$

and for $i < 0$

$$\begin{aligned} &S_{i,0} = \tilde{S}_{0,0}, S_{i,1} = \tilde{S}_{-1,1}, \ldots, S_{i,-i-1} = \tilde{S}_{i+1,-i-1}, \\ &S_{i,-i} = \tilde{S}_{i,-i}, \ldots S_{i,L-1} = \tilde{S}_{i,L-1}, S_{i,L} = \tilde{S}_{i,L}. \end{aligned}$$

Thus for any i the route $\mathcal{A}_i$ can be represented as $\mathcal{A}_i = \{S_{i,0}, S_{i,1}, \ldots, S_{i,L}\}$. As the state space $S^{\alpha,L}$ of the approximating finite-state Markov chain $\{\hat{s}_n, n \geq 1\}$ we will take the union of possible routes, i.e. $S^{\alpha,L} = \cup_{i=-L}^{L} \mathcal{A}_i$. The size of the set $S^{\alpha,L}$ is $(L+1)(2L+1)$, which is greater than the size of the set $\tilde{S}^{\alpha,L}$. This is due to the fact that some points of the grid $\tilde{S}^{\alpha,L}$ belong to different routes $\mathcal{A}_i$ at the same time. Such points are those which lie on coordinate axes (excluding extreme points). For example the route $\mathcal{A}_i$ includes point $S_{i,0} = \tilde{S}_{0,0}$ corresponding to empty state of the system. Such duplication may seem unnatural but, as will be shown further, it greatly simplifies the calculation of transition probabilities.
Now let us dwell on description of transitions of approximating Markov chain $\{\hat{s}_n, n \geq 1\}$. Let at the time of the n-th job arrival the system be in the state $\hat{s}_n = S_{ij} \in S^{\alpha,L}$ and assume that after a decision the system entered state

$S_{kl} \in S^{\alpha,L}$[3]. The state S_{ij} (point on the grid) to which transition from state S_{ij} occurs is completely defined by threshold policy and there is one-to-one correspondence between indexes k and i (l and j, as well). After transition to state S_{kl} system evolves deterministically until the next arrival. At next arrival instant system finds itself in the new state $\hat{s}_{n+1}$, which coincides with one of the points S_{kl}, $S_{k,l-1}$, $\ldots$, S_{k0} of the grid. From description of the set $S^{\alpha,L}$ and *Remark 1* it follows that the transition probabilities $S_{kl} \to S_{km}$, $m = 0,1,\ldots,l$, depend only on index l and do not depend on index k. Let us denote these probabilities by q_{lm}, i.e. $q_{lm} = \mathbf{P}\{S_{kl} \to S_{km}\}$. Clearly $q_{00} = 1$. Let $l = 1$. From system standpoint it means that there is unfinished work in the system equal to h_0. Due to the fact that the state space at the instant of the next job arrival have been discretized, there are only two options: either unfinished work in the system will be the same (say, with probability q_{11}), or the system will be empty (with probability $1 - q_{11} = q_{10}$). The value of q_{11} may be taken equal to probability that inter-arrival time does not exceed $0.5h_0$, i.e. $q_{11} = F(0.5h_0)$[4]. By the same argument the following expression for arbitrary value of $l = 1,\ldots,L$ is obtained:

$$q_{lm} = F(H_{l-m+1}) - F(H_{l-m}),\ m = 0,\ldots,l,$$

where

$$H_m = \begin{cases} 0, & \text{if } m=0, \\ h_{l-1}+h_{l-2}+\ldots+h_{l-m+1}+0.5h_{m-1}, & \text{if } m=1,\ldots,l, \\ 1, & \text{if } m=l+1. \end{cases}$$

3.2. Description of the Iterative Procedure

In order to be able to compute transition probabilities of approximating Markov chain $\{\hat{s}_n, n \ge 1\}$ one has to know how to jump from bevel coordinates given by indexes of elements S_{ij} to rectangular coordinates $(x,y) \in S$ and back. This transform follows directly from the way the sets $S^{\alpha,L}$ was constructed. Let x and y be rectangular coordinates of point $S_{ij} \in S^{\alpha,L}$. If $i = 0$, then clearly $x = y = 0$. For $i > 0$ it holds that

$$x = h_0 \frac{(1+\alpha)^j - 1}{\alpha}, \tag{10}$$

$$y = \max\big(0, \nu x - \nu h_0 \frac{(1+\alpha)^i - 1}{\alpha}\big), \tag{11}$$

and for $i < 0$

$$x = \max\big(0, \frac{y}{\nu} - h_0 \frac{(1+\alpha)^{-i} - 1}{\alpha}\big), \tag{12}$$

$$y = \nu h_0 \frac{(1+\alpha)^j - 1}{\alpha}. \tag{13}$$

The inverse transform is not unique. This is because there are different ways in which one can choose point $S_{ij} \in S^{\alpha,L}$, which approximates point $(x,y) \in S$. For example, one can use the following rule:

$$i = \max\big(-L, \min(L, i')\big), \tag{14}$$

$$j = \max\big(-L, \min(L, j')\big), \tag{15}$$

where

$$i' = \operatorname{sign}\Big(x - \frac{y}{\nu}\Big) \left\lfloor \frac{\ln\big(1 + \frac{\alpha}{h_0}\big|x - \frac{y}{\nu}\big|\big)}{\ln(1+\alpha)} \right\rfloor,$$

$$j' = \begin{cases} \left\lfloor \frac{\ln\left(1+\frac{\alpha x}{h_0}\right)}{\ln(1+\alpha)} \right\rfloor, & \text{if } y < \nu x, \\ \left\lfloor \frac{\ln\left(1+\frac{\alpha y}{h_0 \nu}\right)}{\ln(1+\alpha)} \right\rfloor, & \text{if } y \ge \nu x, \end{cases}$$

where $\operatorname{sign}(a)$ denotes signum function and $\lfloor a \rfloor$ denotes integer part of a.

Assume Markov chain $\{s_n, n \ge 1\}$ generated by threshold policy ξ is in state $s = (x,y)$ at the time of n-th arrival. Then after a decision it will move to state

$$(\tilde{x},\tilde{y}) = \begin{cases} (x+1,y), & \text{if } \frac{y}{\nu} - x > \xi, \\ (x,y+1), & \text{if } \frac{y}{\nu} - x \le \xi. \end{cases} \tag{16}$$

Let the approximating Markov chain $\{\hat{s}_n, n \ge 1\}$ be in state $\hat{s} = (\hat{x},\hat{y})$ such that $\frac{\hat{y}}{\nu} - \hat{x} = \xi$. Consider again policies $\sigma_\xi^{(1)}$ and $\sigma_\xi^{(2)}$ introduced at the beginning of Section 3 and denote by $\hat{\pi}_n^{(1)}$ and $\hat{\pi}_n^{(2)}$ respectively stationary distribution over the state space $S^{\alpha,L}$ under these policies. The discrete version of the difference Δ_ξ, introduced in Eq. (3), is given by

$$\Delta_\xi^{\alpha,L} = g^{(1)} - g^{(2)} + \sum_{n=1}^{\infty} \sum_i \sum_j g(i,j)\big(\hat{\pi}_n^{(1)}(i,j) - \hat{\pi}_n^{(2)}(i,j)\big),$$

where $g^{(1)}$ and $g^{(2)}$ are computed from Eq. (4), $g(i,j) = g(S_{ij})$, and $\hat{\pi}_n^{(i)}(i,j)$ are the values of the distributions $\hat{\pi}_n^{(1)}$ and $\hat{\pi}_n^{(2)}$ at point S_{ij}.

The step-by-step procedure for the update of the value $\Delta_\xi^{\alpha,L}$ is given below in Algorithm 1. It also shows how stationary distributions $\hat{\pi}_n^{(i)}$ can be calculated on the fly.

The x_{ij}, y_{ij} are rectangular coordinates of point S_{ij}, calculated from Eqs. (10)–(13), I_{xy}, J_{xy} are indexes of inverse transform calculated from Eqs. (14)–(15) and $\tilde{x} = \tilde{x}(x,y)$, $\tilde{y} = \tilde{y}(x,y)$ are given by Eq. (16).

Remark 2. Algorithm 1 is only the basic version which can be modified in order to improve its efficiency. For example, one can shift the area of the grid $\tilde{S}^{\alpha,L}$ where the most points are concentrated from the neighborhood of $(0,0)$ (which is the case in Algorithm 1) to the neighborhood of the more frequent states of the system. Such states can be determined using simulation.

Remark 3. Proposed algorithm allows one to check whether the chosen value of threshold ξ is the optimal value. Algorithm 1 does not contain the description of the exact

[3] Note that the authors are working under assumption that transitions $S_{ij} \to S_{kl}$ do not incur any delay.

[4] This value is taken by an agreement. The is no other reasoning behind this choice except for common sense.

Algorithm 1: Algorithm for computation of steady state probabilities and approximating value of Δ_ξ

Step 1
Initialize $\Delta_0 = g^{(1)} - g^{(2)}$;
if $i = i_{\hat{x},\hat{y}}$ and $j = j_{\hat{x},\hat{y}}$ **then**
  $\hat{\pi}_0^{(1)}(i,j) = 1$, $\hat{\pi}_0^{(2)}(i,j) = 0$;
else
  $\hat{\pi}_0^{(1)}(i,j) = 0$, $\hat{\pi}_0^{(2)}(i,j) = 1$;
end if
Step 2
$x = \hat{x} + 1$, $y = \hat{y}$;
$k = I_{xy}$, $l = J_{xy}$;
$\hat{\pi}_1^{(1)} = 0$;
for $m = 0$ to l **do**
  $\hat{\pi}_1^{(1)}(k,m) = \hat{\pi}_1^{(1)}(k,m) + q_{lm}\hat{\pi}_0^{(1)}(k,l) = 0$; // Compute initial state probabilities after action 1
end for
$x = \hat{x}$, $y = \hat{y} + 1$;
$k = I_{xy}$, $l = J_{xy}$;
$\hat{\pi}_1^{(2)} = 0$;
for $m = 0$ to l **do**
  $\hat{\pi}_1^{(2)}(k,m) = \hat{\pi}_1^{(2)}(k,m) + q_{lm}\hat{\pi}_0^{(2)}(k,l) = 0$; // Compute initial state probabilities after action 2
end for
$\Delta\hat{\pi}_1 = \hat{\pi}_1^{(1)}(k,m) - \hat{\pi}_1^{(2)}(k,m)$; // component-wise difference
$\Delta_1 = \Delta_0 + \sum_{n=1}^{\infty}\sum_{i=-L}^{L}\sum_{j=0}^{L} g(i,j)\Delta\hat{\pi}_1(i,j)$;
$n = 1$;
Step 3
$n = n + 1$;
for $i = -L$ to L **do**
  for $j = 0$ to L **do**
    $x = x_{ij}$, $y = y_{ij}$; // rectangular coordinates of point S_{ij}
    $k = i_{\tilde{x}\tilde{y}}$, $l = j_{\tilde{x}\tilde{y}}$; // index values after making decision
    $\Delta\hat{\pi}_n = 0$;
    for $m = 0$ to l **do**
      $\Delta\hat{\pi}_n = \Delta\hat{\pi}_n + q_{lm}\hat{\pi}_{n-1}^{(1)}(k,l) - q_{lm}\hat{\pi}_{n-1}^{(2)}(k,l)$
    end for
  end for
end for
$\Delta_n = \Delta_{n-1} + \sum_{n=1}^{\infty}\sum_{i=-L}^{L}\sum_{j=0}^{L} g(i,j)\Delta\hat{\pi}_n(i,j)$;
if $|\Delta_n - \Delta_{n-1}| < \varepsilon$ **then** // ε – parameter of the algorithm
  goto *Step 3*;
else
  $\Delta_\xi^{\alpha,L} = \Delta_n$.
end if

procedure for the calculation of the threshold because it can be performed in different ways. For example, one can choose (using qualitative analysis of the system behavior) interval which contains the (unknown) value of threshold ξ. For example, in current setting this interval is $(0, \nu^{-1})$. Then use bisection method can be applied.

4. Numerical Example

Let us give simple comparison of results, which were obtained from proposed algorithm with results obtained from Monte-Carlo simulation.

Let the service rate of server II be equal to $\nu = 2$. The threshold value of the optimal policy for two types of inter-arrival distributions is then computed: exponential with parameter $\lambda = 2.4$ and Pareto with scale $b = 0.21$ and shape $a = (1 - \lambda b)^{-1} \approx 2.016$. Both these distributions have equal mean inter-arrival times but their variances differ significantly. For exponential distribution the variance is $\lambda^{-2} \approx 0.417$ and for Pareto it is $ab^2/[(a-1)^2(a-2)]) \approx \approx 25.43$.

In order to construct the grid $\tilde{S}^{\alpha,L}$ let us fix the minimum and maximum grid spacing by letting $h_0 = 0.005$ and $h_{L-1} = 0.025$. Let the total length of the approximating area along the x-axis be $H = 10$. Given the value of h_0, h_{L-1} and Hm, other parameters of the grid can be calculated from Eqs. (5) and (6). That is

$$\alpha = \frac{h_{L-1} - h_0}{H - h_{L-1}} \approx 0.0001, \quad L = \left\lfloor \frac{\ln\left(1 + \frac{\alpha H}{h_0}\right)}{\ln(1+\alpha)} \right\rfloor \approx 1600.$$

The total number of states after discretization is

$$(L+1)(2L+1) \approx 5.2 \times 10^6.$$

Having applied iterative algorithm described in Section 3 we obtained that for exponential inter-arrival times the optimal threshold ξ_{opt} lies in the interval $(0.166, 0.167)$ and for Pareto inter-arrival times the interval is $(0.150, 0.151)$. In order to understand how accurate these results are, let us have a look at the value of value function T_ξ (estimated from Monte-Carlo simulation) for the threshold values ξ, which are in the neighborhood of the obtained intervals. The results are given in Table 1.

Table 1
(Approximate) values of the value function T_ξ in the neighborhood of the intervals, containing optimal threshold

Exponential inter-arrival times		Pareto inter-arrival times	
Threshold ξ	Mean T	Threshold ξ	Mean T
0.160	1.25459	0.144	0.93638
0.162	1.25458	0.146	0.93637
0.164	1.25456	0.148	0.93636
0.166	1.25454	0.150	0.93636
0.168	1.25454	0.152	0.93636
0.172	1.25454	0.154	0.93637
0.174	1.25455	0.156	0.93638

One can see from Table 1 that the proposed algorithm gives good results up to (and including) the third digit after the decimal point. In order to check the value of the fourth digit one has to be able to estimate value function T_ξ from Monte-Carlo simulation up to the sixth digit after the decimal point. Such estimation is far from being simple because estimation of T_ξ up to fifth digit already takes several hours on standard PC. Meanwhile the proposed algorithm finds the interval (up to third digit after decimal point), which contains optimal threshold value usually in 5–10 minutes.

5. Conclusion

As it is mentioned in many research papers quite a few problems which one may encounter in practice (for example, building schedulers in distributed processing systems) can be formulated in terms of flows, servers, queues. The considered problem is only the special case of far more general model which may encompass many details of real-life systems and the need for appropriate solution methods seems to be high. At present the most popular "attack" method for such problems is the use of heuristics and their validation using simulation. Even for the considered special case the non-simulation solution is far from being simple (and exact solution is not known at all). Analytic solution methods for arbitrary $n > 2$ number of servers have not yet been developed and the structure of optimal policy is not known. It must not necessarily be of threshold type. Although if one decides that threshold policy should be used in the n-server system, then the proposed algorithm can be scaled in a straightforward manner, but the obtained results may not be optimal. Here one of the appealing ways to check the quality of the solution is again the comparison with simulation. Our experiments show that Monte-Carlo simulation in combination with adaptive algorithms for partially observed Markov chains is the most suitable approach for this purpose.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research (grant 15-07-03406).

References

[1] M. E. Crovella, M. Harchol-Balter, and C. D. Murta, "Task assignment in a distributed system: Improving performance by unbalancing load", in *Proc. of Joint Int. Conf. on Measur. and Model. of Comp. Syst. SIGMETRICS'98/PERFORMANCE '98*, Madison, Wisconsin, USA, 1998, pp. 268–269.

[2] M. Harchol-Balter, M. E. Crovella, and C. D. Murta, "On choosing a task assignment policy for a distributed server system", *J. of Parallel and Distrib. Comput.*, vol. 59, pp. 204–228, 1999.

[3] S. Bhulai, "On the value function of the M/Cox(r)/1 queue", *J. of Appl. Probab.*, vol. 43, no. 2, pp. 363–376, 2006.

[4] E. Hyytiä, J. Virtamo, S. Aalto, and A. Penttinen, "M/M/1-PS queue and size-aware task assignment", *Perform. Eval.*, vol. 68, no. 11, pp. 1136–1148, 2011.

[5] E. Hyytiä, A. Penttinen, and S. Aalto, "Size- and state-aware dispatching problem with queue-specific job sizes", *Eur. J. of Oper. Res.*, vol. 217, no. 2, pp. 357–370, 2012.

[6] E. Hyytiä, S. Aalto, and A. Penttinen, "Minimizing slowdown in heterogeneous size-aware dis-patching systems", in *Proc. of the 12th ACM SIGMETRICS/Performance Conf. Measur. Model. Comp. Syst.*, London, UK, 2012, *Eval. Rev.*, vol. 40, pp. 29–40, 2012.

[7] J. Kołodziej and F. Xhafa, "Enhancing the genetic-based scheduling in computational grids by a structured hierarchical population", *Future Gener. Comp. Syst.*, vol. 27, no. 8. pp. 1035–1046, 2011.

[8] K. R. Krishnan and T. J. Ott, "State-dependent routing for telephone traffic: Theory and results", in *Proc. 25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986, vol. 25, pp. 2124–2128.

[9] M. G. Konovalov, "About one task of overload control", *Informatics and its Applications*, vol. 7, no. 4, pp. 34–43, 2013 (in Russian).

[10] E. Hyytiä, "Lookahead actions in dispatching to parallel queues", *Perform. Eval.*, vol. 70, no. 10, pp. 859–872, 2013.

[11] E. Hyytiä, "Optimal routing of fixed size jobs to two parallel servers", *INFOR: Inform. Syst. and Oper. Res.*, vol. 51, no. 4, pp. 215–224, 2013.

[12] R. L. Larsen, "Control of multiple exponential servers with application to computer systems", Ph.D. thesis, University of Maryland at College Park, College Park, MD, USA, 1981.

[13] K. R. Krishnan, "Markov decision algorithms for dynamic routing", *IEEE Commun. Mag.*, vol. 28, no. 10, pp. 66–69, 1990.

[14] J. Leeuwaarden, S. Aalto, and J. Virtamo, "Load balancing in cellular networks using first policy iteration", Tech. Rep., Networking Laboratory, Helsinki University of Technology, 2001.

[15] K. R. Krishnan, "Joining the right queue: a state-dependent decision rule", *IEEE Trans. on Autom. Contr.*, vol. 35, no. 1, pp. 104–108, 1990.

[16] S. A. E. Sassen, H. C. Tijms, and R. D. Nobel, "A heuristic rule for routing customers to parallel servers", *Statistica Neerlandica*, vol. 51, no. 1, pp. 107–121, 1997.

Mikhail Konovalov, Ph.D., is the head of the sector at Information Technologies Department at Institute of Informatics Problems FRC CSC RAS. His interests include adaptive control of random sequences, modeling and simulation of complex systems.

E-mail: mkonovalov@ipiran.ru
Institute of Informatics Problems
Federal Research Center "Computer Science and Control"
Russian Academy of Sciences
Vavilova st 44-2
119333 Moscow, Russia

Rostislav Razumchik received his Ph.D. degree in Physics and Mathematics in 2011. Since then, he has worked as a senior research fellow at Institute of Informatics Problems FRC CSC RAS. His current research activities focus on stochastic processes and queuing theory.

E-mail: rrazumchik@ipiran.ru
Institute of Informatics Problems
Federal Research Center "Computer Science and Control"
Russian Academy of Sciences
Vavilova st 44-2
119333 Moscow, Russia

Paper

Performance Evaluation of Cognitive Radio Network Based on 2-D Markov Chain

Md. Imdadul Islam, Md. Fazlay Rabbi, Risala Tasin Khan, and Jesmin Akhter

Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh

Abstract—The objective of cognitive radio network is to enhance the wireless network spectrum utilization. In such a network, two types of users are enlisted, namely primary user (PU) and secondary user (SU). The PU can access any channel in case of its availability, but SU users have lower priority and can access a channel only when it is unused by PUs. The performance of such a network solely depends on two traffic parameters: probability of false alarm and probability of misdetection. In this paper the performance of such a network is analyzed based on two dimensional Markov chain including those parameters. The main contribution of this paper is to evaluate blocking probability and PU and SU throughput using the state transition chain instead of existing statistical analysis.

Keywords—cognitive radio networks, misdetection, probability of false alarm, throughput and traffic of limited channel.

1. Introduction

In a conventional network during low offered traffic period, few traffic channels are idle. Hence network experiences a waste of channel utilization. A cognitive radio network (CRN) overcomes the situation, since the unutilized channels are used by unlicensed users called secondary users (SUs) in case of their availability. The main job of a secondary user is to sense the presence of a primary user (PU), which is done by managing the traffic channel. The detail statistical analysis of CRN spectrum sensing technique is presented in [1]–[4]. To enhance the success rate of sensing a PU by SU, several detected signals of SUs are further analyzed at a host level. Two important traffic parameters under CRN are probability of false alarm and misdetection. Recent literature deals with a statistical model of spectrum sensing based on absence of PU and presence of PU known as hypothesis H_0 and H_1 respectively. If there is no PU on a physical channel and the received signal strength of SU is over a signal to noise ratio (SNR) threshold, then the phenomenon is called false alarm (FA). Again, under presence of a PU on physical channel, if the received SNR is lower than the threshold value, then the phenomenon is called misdetection. Both the parameters govern the performance of a CRN, which are explained in [5]–[8] with a detailed statistical model. Sometimes the false alarm is categorized based on coverage range of sensing shown in [9]. In this case the SU has to confirm whether the PU is in paging mode inside the sensing zone or in transmitting mode outside the sensing region but this concept is beyond the analysis of the paper. To combat the false alarm and misdetection under fading wireless channel, the space diversity and different combining schemes are incorporated with the sensing unit of SU shown in [10]. The traffic analysis of cognitive radio under Internet, specially for Voice over IP (VoIP) is discussed in [10]–[11]. When a PU requests for a physical channel, the SU on that channel is removed and the phenomenon is called forced termination (FT). In this case the SU has to move to another available traffic channel. In this paper this idea is modeled by two dimensional state transition chains, which is then solved to evaluate the CRN traffic parameters. The detailed statistical analysis of Markov chain and its solution is presented in [13]–[15] but authors emphasis on two dimensional Markov chain with some modification to cope with the CRN traffic. Finally the CRN traffic performance based on blocking probability and throughput (carried traffic) of both PU and SU varying other traffic parameters is analyzed.

The paper is organized as follows. Section 2 provides the traffic model of CRN based on two dimensional Markov chain and the traffic parameters pertinent to the CRN performance are derived solving the state transition chain. Section 3 depicts the results based on analysis of Section 2 and finally Section 4 concludes the entire analysis.

2. System Model

Let us first consider a Markov chain of unlimited channel and user case, which is represented by Kendall's notation as $M/M/\infty$ shown in Fig. 1. The λ and μ are the call arrival and termination rate respectively as discussed in [16]–[18]. The chain states P_0, P_1, P_2 etc. are the probability of arrival

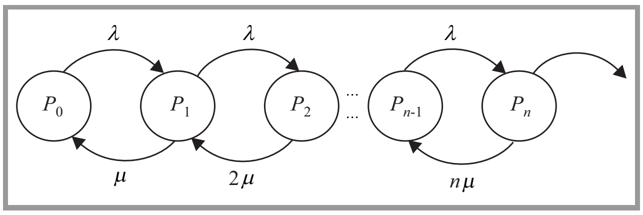


Fig. 1. State transition chain of unlimited trunk network.

of no call, one call, two calls and so on. Basic assumption in Markov chain is that probability of reaching the next state depends on present state but not on previous states. Any new call arrival rate λ will make transition of a probability state P_i to P_{i+1} similarly any call termination rate $i\mu$ will make transition from P_i to P_{i-1}.

Applying cut equation between nodes in Fig. 1 the generalized probability state x known as Poisson's probability density function (PDF) is:

$$P_x = \frac{A^x}{x!}e^{-A} . \tag{1}$$

Poisson's distribution is valid for infinite number of trunks or channels but in real conditions number of channels are limited therefore traffic should be analyzed for limited channel case. A limited trunk traffic model is shown in Fig. 2, where number of users N is infinite, users offer an average arrival rate of λ and average holding time of $t_h = 1/\mu$. If the number of trunk is n then any arrival beyond the state P_n will not get the service and the call will be lost shown in Fig. 3 using Markov chain.

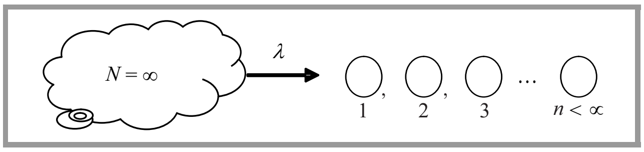


Fig. 2. Limited server or channel network.

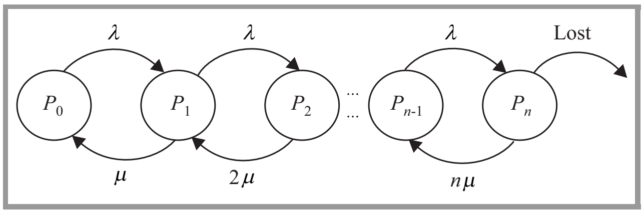


Fig. 3. State transition of limited trunk network.

Again applying cut equations, the probability state x in generalized form from Fig. 3 becomes:

$$P_x = \frac{\frac{A^x}{x!}}{\sum_{i=0}^{n}\frac{A_i}{i!}} , \tag{2}$$

which is known as Erlang's PDF.

In teletraffic engineering probability of occupancy of all channels is called call blocking probability:

$$B_n = P_n = \frac{\frac{A^n}{n!}}{\sum_{i=0}^{n}\frac{A_i}{i!}} . \tag{3}$$

Let us now focus on 2-D Markov chain having two types of Poisson's offered traffic A_1 and A_2 [19]. The arrival and termination rates are: λ_1, λ_2, μ_1, μ_2 and $n = \infty$ servers or channels shown in Fig. 4. It is convenient to arrange states x_1 and x_2 along X and Y direction, where any probability state $P_{x_1x_2}$ indicates probability of x_1 and x_2 calls of type 1 and type 2 traffic occupancy, respectively. Because the system is reversible it is convenient to apply cut equation between nodes.

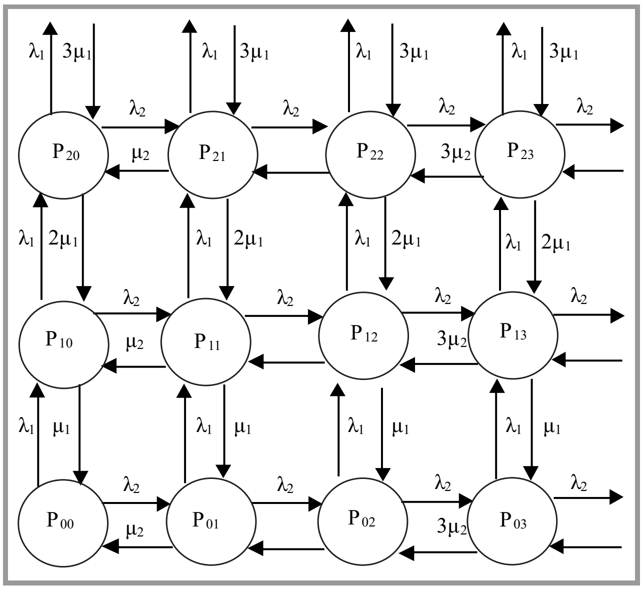


Fig. 4. 2-D Markov chain of Poisson traffic.

Considering x_2-th column in Fig. 4 and applying cut equation between first and second node we get

$$\lambda_1 P_{0x_2} = P_{1x_2}\mu_1 \Rightarrow P_{1x_2} = \frac{\lambda_1}{\mu_1}P_{0x_2} = A_1 P_{0x_2} .$$

Similarly between second and third node:

$$\lambda_1 P_{1x_2} = P_{2x_2}2\mu_1 \Rightarrow P_{2x_2} = \frac{\lambda_1}{2\mu_2}P_{1x_2} = \frac{A_1^2}{2!}P_{0x_2} .$$

In generalized form

$$P_{x_1x_2} = \frac{A_1^{x_1}}{x_1!}P_{0x_2} . \tag{4}$$

Again considering x_1-th row in Fig. 4:

$$P_{x_1x_2} = \frac{A_2^{x_2}}{x_2!}P_{x_10} . \tag{5}$$

Putting $x_2 = 0$ in Eq. (4)

$$P_{x_10} = \frac{A_1^{x_1}}{x_1!}P_{00} . \tag{6}$$

From Eqs. (5) and (6)

$$P_{x_1x_2} = \frac{A_2^{x_2}}{x_2!}\frac{A_1^{x_1}}{x_1!}P_{00} . \tag{7}$$

Considering entire sample space

$$\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}P_{x_1x_2} = 1 \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}\frac{A_2^{x_2}}{x_2!}\frac{A_1^{x_1}}{x_1!} = 1 \Rightarrow$$

$$\Rightarrow \leftarrow -- \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}\frac{A_2^{x_2}}{x_2!}P_{00} = 1 \Rightarrow e^{A_1}e^{A_2}P_{00} = 1 . \tag{8}$$

From Eqs. (7) and (8) the probability state (x_1, x_2) in normalized form becomes

$$P_{x_1 x_2} = \frac{A_1^{x_1}}{x_1!}\frac{A_2^{x_2}}{x_2!}e^{-(A_1+A_2)}. \tag{9}$$

When multidimensional Poisson traffic is applied to a limited number of servers ($n < \infty$) fully available to all traffic components (complete sharing), loss probabilities are calculated using multidimensional Erlang's loss formula. Using the concept of Eqs. (4) to (9) the sample space for n channels case becomes

$$\sum_{x_1=0}^{n}\frac{A_1^{x_1}}{x_1!}\sum_{x_2=0}^{n-x_1}\frac{A_2^{x_2}}{x_2!}P_{00} = 1 \therefore P_{00} = \frac{1}{\sum\limits_{x_1=0}^{n}\frac{A_1^{x_1}}{x_1!}\sum\limits_{x_2=0}^{n-x_1}\frac{A_2^{x_2}}{x_2!}}. \tag{10}$$

The probability state in normalized form is:

$$P_{x_1 x_2} = \frac{\frac{A_1^{x_1}}{x_1!}\frac{A_2^{x_2}}{x_2!}}{\sum\limits_{x_1=0}^{n}\frac{A_1^{x_1}}{x_1!}\sum\limits_{x_2=0}^{n-x_1}\frac{A_2^{x_2}}{x_2!}}. \tag{11}$$

Blocking probability is the sum of all state $P_{x_1 x_2}$ where $x_1 + x_2 = n$,

$$B_n = \sum_{x_1=0}^{n} P_{x_1} P_{n-x_1}. \tag{12}$$

Figure 5 shows the probability states of a limited channel system ($n = 3$), where sum of the complete occupied states is the blocking probability expressed as $B = P_{30} + P_{21} + P_{12} + P_{03}$.

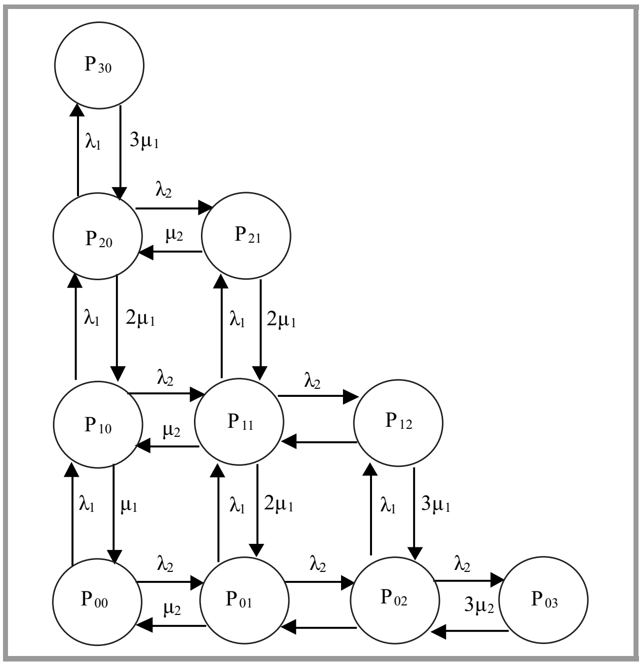


Fig. 5. 2-D Markov chain of Erlang's traffic ($n = 3$).

In cognitive radio network the Markov chain from Fig. 5 has to be modified like it is presented in Fig. 6 (the only $n = 3$ channels are considered). Since any arrival of PU under a complete occupied state, a SU has to be terminated

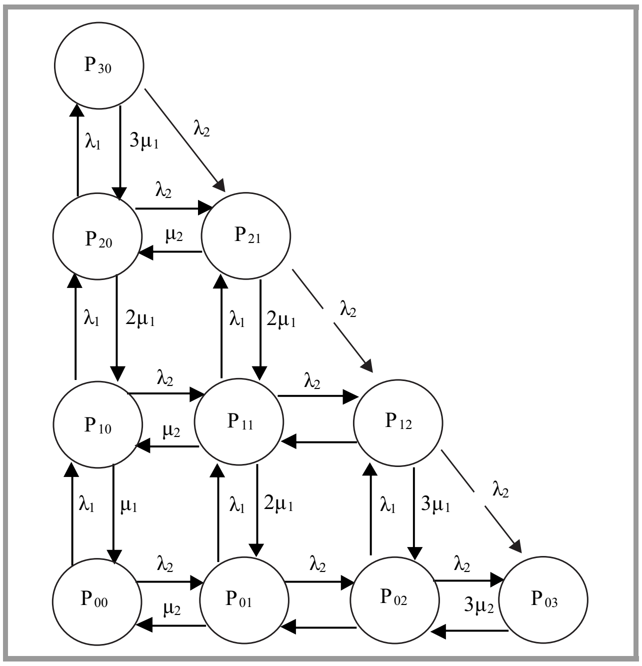


Fig. 6. 2-D Markov chain of Erlang's traffic ($n=3$) for performance evaluation of cognitive radio network under unlimited user.

forcibly as discussed in [20], [21]. The annotation is as follows: $\lambda_1 = \lambda_2 =$ arrival rate of SU and $\lambda_2 = \lambda_p =$ arrival rate of PU, $\mu_1 = \mu_s =$ termination rate of SU and $\mu_2 = \mu_p =$ termination rate of PU. The offered PU traffic and SU are $A_p = \lambda_p/\mu_p$ and $A_s = \lambda_s/\mu_s$ respectively. For example a CRN of n channel, the complete occupied states will be $(x_1, x_2) = (i, n-i)$; $i = 0, 1, 2, 3, \ldots, n$, where x_1 is the number of SU and x_2 is the number of PU. Any arrival of PU at state $(i, n-i)$ will change the state to $(i-1, n-i+1)$ forcefully.

Let us introduce the traffic parameters of CRN like: P_{md} – probability of misdetection, P_f – probability of false alarm, $A_{s\_original}$ – offered traffic of SU, A_p – offered traffic for primary user. Channel will experience the traffic from SU side as

$$A_s = A_{s\_original}\cdot(1-P_{md})P(\mathrm{H}_1) + A_{s\_original}\cdot(1-P_f)P(\mathrm{H}_0),$$

where $P(\mathrm{H}_1)$ and $P(\mathrm{H}_0)$ are the probability of presence and absence of an PU on a traffic channel. The probabilities are widely used in test symbols statistical analysis of detected energy, and are known as two hypothesis model.

Solving the Markov chain from Fig. 6 for n channels case, the entire sample space is

$$S(n) = A + B + C + D, \tag{13}$$

where

$$A = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1-i}\frac{A_p^i}{i!}\frac{A_s^j}{j!},$$

$$B = \sum_{x=1}^{n-1}\frac{A_p\left\{\frac{A_p^{n-x-1}}{(n-x-1)!}\frac{A_s^x}{x!}+\frac{A_p^{n-x-1}}{(n-x-1)!}\frac{A_s^{x+1}}{(x+1)!}\right\}+A_s\frac{A_p^{n-x}}{(n-x)!}\frac{A_s^{x-1}}{(x-1)!}}{n+A_p},$$

$$C = \frac{A_p \left\{ \frac{A_p^0}{0!} \cdot \frac{A_s^{n-1}}{(n-1)!} \right\}}{n + A_p} \quad \text{and} \quad D = \frac{A_p \left\{ \frac{A^{n-1}}{(n-1)!} \cdot \frac{A_s^0}{0!} \right\} + A_p \frac{A_p^{n-x}}{(n-x)!} \cdot \frac{A_s^1}{1!}}{n}.$$

The blocking probability of SU is:

$$B_{SU}(n) = \frac{B + C + C}{S(n)}, \tag{14}$$

and the blocking probability of PU:

$$B_{PU}(n) = \frac{D}{S(n)}. \tag{15}$$

The throughput of PU could be expressed as

$$X_{p\_bar}(n) := \big(1 - B_{pu}(n)\big) \cdot A_p \tag{16}$$

and the throughput of SU

$$X_{p\_bar}(n) := \big(1 - B_{su}(n)\big) \cdot A_{s\_original}. \tag{17}$$

The next section provides the profile of throughput and blocking probability to evaluate the CRN performance.

3. Results

Figure 7 shows that the variation of throughput against number of channels taking. The offered traffic of PU is $A_p = 15$ Erl, offered traffic of SU $A_{s\_original} = 10$ Erl, the probability of misdetection $P_{md} = 0.12$, the probability of false alarm $P_f = 0.04$ and the probability of hypothesis $P(H_1)$ and $P(H_0)$ are both 0.5. The picture shows that the throughput gradually increases with the number of channels of both the PU and SU. The throughput of PU is found much larger than the SU because of its priority of getting a channel.

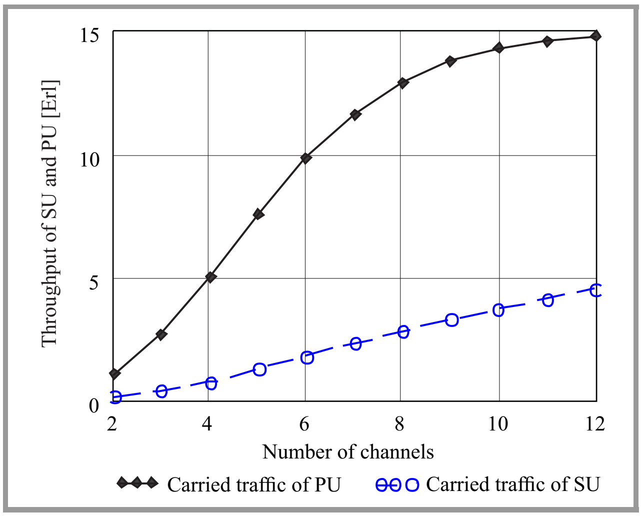


Fig. 7. Throughput of SU and PU with the variation of channel.

The variation of blocking probability (PU and SU) against the number of channel is shown in Fig. 8. The blocking probability of PU is much lower than SU at the same time the rate of decrement of blocking probability with increase in the channel number is more rapid for PU for the same reason. From the profile of blocking probability one shall notice that $B_{PU} \approx 0$ for $n > 12$ but $B_{SU} > 0.35$, for the same channel condition. The blocking probability is more sensitive to P_{md} than P_f. This is also visualized in Fig. 8.

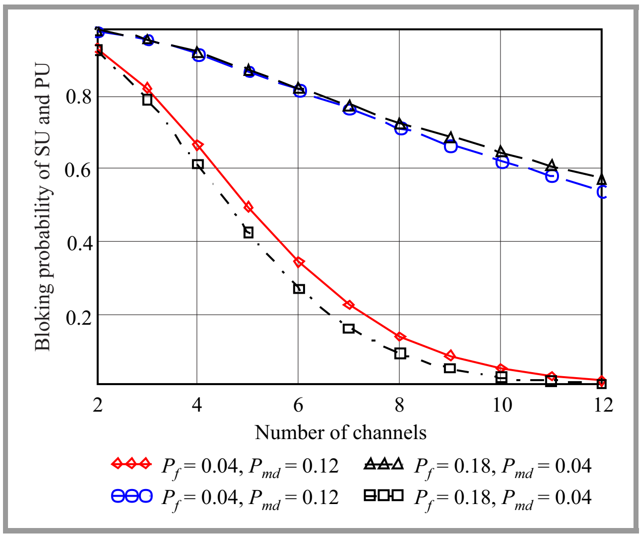


Fig. 8. Blocking probabilities of SU and PU with the variation of channel.

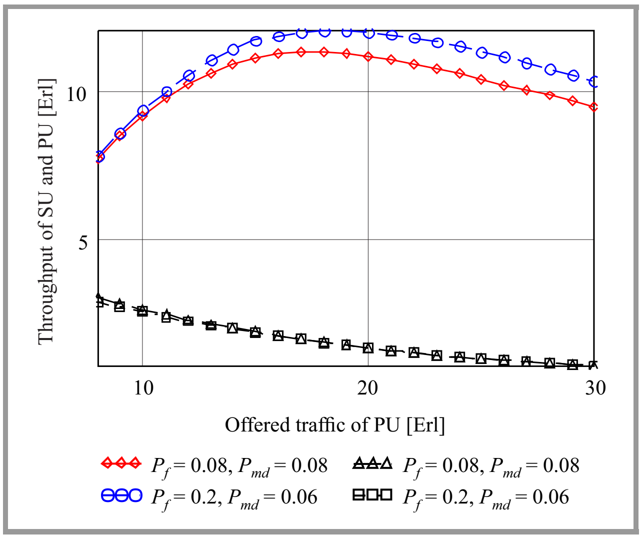


Fig. 9. Throughput of SU and PU with the variation of offered traffic.

Figure 9 shows that the throughput variation versus offered traffic of PU, under the conditions: $n = 8$, offered traffic of PU $A_p = 8 \ldots 30$ Erl, offered traffic of SU $A_{s\_original} = 6$ Erl, probability of misdetection $P_{md} = 0.08$ and probability of false alarm $P_f = 0.08$. The graph shows that the PU throughput increases within offered traffic of PU and attains at a maximum value then starts to decrease. At low offered traffic, most of the channels remains idle, again at higher offered traffic the blocking probability increases, hence throughput lowers. Therefore, there must be an optimum offered traffic with maximum throughput. There is a tradeoff between idle channel and blocking probability visualized in Fig. 9. The impact of P_{md} is more prominent than P_f the throughput always decreases with increase in the offered traffic of PU because of its forced termination.

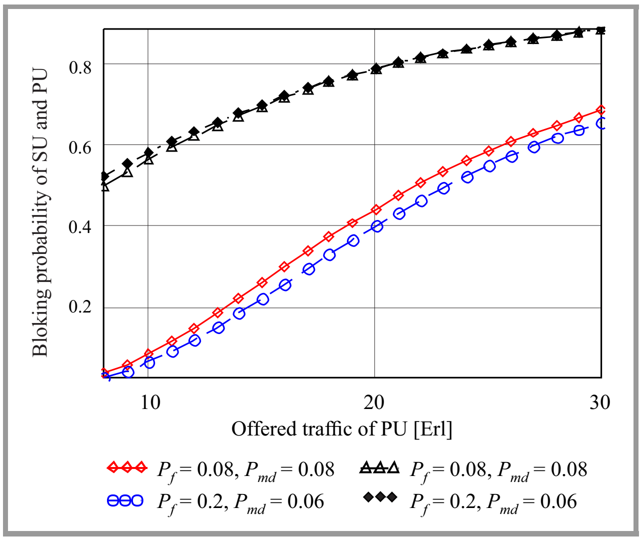


Fig. 10. Blocking probability of SU and PU with the variation of offered traffic.

Finally Fig. 10 shows that the variation of blocking probability of both PU and SU against the offered traffic of PU under following conditions: the number of channels $n = 8$, offered SU traffic $A_{s\_original} = 6$ Erl, probability of misdetection $P_{md} = 0.08$ and probability of false alarm $P_f = 0.08$. The graph shows that the blocking probability of both type of user increases with the PU offered traffic. The blocking probability is much higher for SU than PU like before, but the rate of B_{PU} increment is more prominent than B_{SU} since offered PU traffic affects the PU users directly and to SU indirectly.

4. Conclusion

In this paper the performance of a cognitive radio network using M/M/n/k traffic model based on 2-D Markov Chain is analyzed. The results shows the performance of the network varying different traffic parameters, provide expected results. Still the scope to use M/G/1/K traffic of packet switch network of finite queue has to be researched. In this case two dimensional traffic models using state transition chain will not be possible because of general PDF service time. The authors can apply tabular form of 2-D traffic model of M/G/1/K to pave the way for evaluating packet loss PU and SU probability. Finally the impact of fading channel on false alarm and misdetection can also be included on the traffic model to observe the performance under small scale fading environment. One of the major components of 5G mobile communications is the concept of cognitive radio under each node-B to increase the carried such a network traffic. The proposed traffic model will be helpful for developing the next generation wireless networks.

References

[1] S. Cheng, *Foundation of Cognitive Radio Systems*. Rijeka, Croatia: InTech, 2012.

[2] Y. Xiao and F. Hu, *Cognitive Radio Networks*. CRC Press, Taylor & Francis Group, 2009.

[3] J. Ma, G. Y. Li, and B. H. Juang, “Signal processing in cognitive radio”, in *Proc. of the IEEE*, vol. 97, no. 5, pp. 805–823, 2009.

[4] S. Wang, Y. Wang, J. P. Coon, and A. Doufexi, “Energy-efficient spectrum sensing and access for cognitive radio networks”, *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 906–912, 2012.

[5] K. Thakur and N. K. Dewangan, “Cognitive radios and their role in efficient allocation of the spectrum”, *Mediter. J. Social Sci.*, vol. 3, no. 15, pp. 152–164, 2012.

[6] D. Cabric, S. Mishra, and R. W. Brodersen, “Implementation issues in spectrum sensing for cognitive radios”, in *Proc. 38th Asilomar Conf. Sig., Syst. and Comp. ASILOMAR 2004*, Pacific Grove, CA, USA, 2004, pp. 772–776.

[7] A. Ghasemi and E. S. Sousa, “Spectrum sensing in cognitive radio networks: the cooperation-processing trade-off”, *Wirel. Commun. and Mobile Comp.*, Special Issue on Cognitive Radio, Software-Defined Radio, and Adaptive Wireless Systems, vol. 7, no. 9, pp. 1049–1060, 2007.

[8] R. Xie, F. R. Yu, and H. Ji, “Dynamic resource allocation for heterogeneous services in cognitive radio networks with imperfect channel sensing”, *IEEE Trans. Veh. Technol.*, vol. 61, no. 2, pp. 770–780, 2012.

[9] W. Han, J. Li, Q. Liu, and L. Zhao, “Spatial false alarms in cognitive radio”, *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 518–520, 2011.

[10] R. T. Khan, T. N. Shabnam, Md. Imdadul Islam, and M. R. Amin, “Enhancement of performance of cognitive radio network with incorporation of MRC scheme at secondary receiver”, *IACSIT Int. J. Engine. Technol.*, vol. 4, no. 4, pp. 495–499, 2012.

[11] Y. Y. Mihov and B. P. Tsankov, “Cognitive system with VoIP secondary users over VoIP primary users”, in *Proc. 3rd Int. Conf. Adv. Cogn. Technol. Appl. COGNITIVE 2011*, Rome, Italy, 2011, pp. 30–35.

[12] Y. Y. Mihov, “Cross-layer QoS provisioning in cognitive radio networks”, *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 678–681, 2012.

[13] Md. Imdadul Islam, J. K. Das, and S. Hossain, “Modeling of mixed traffic for mobile cellular network”, *J. Telecommun. Inform. Technol.*, no. 1, pp. 83–89, 2007.

[14] Md. S. Hossain and M. Imdadul Islam, “A proposed 2-D queuing model of PCT-I traffic”, in *Proc. 6th Int. Conf. Comp. Informa. Technol. ICCIT 2003*, Dhaka, Bangladesh, 2003, pp. 114–118.

[15] Y. Fang, “Thinning scheme for call admission control in wireless networks”, *IEEE Transa. Comp.*, vol. 52, no. 5, pp. 685–687, 2003.

[16] G. D. Morley and W. D. Grover, “Strategies to Maximize Carried Traffic in Dual-Mode Cellular Systems”, *IEEE Trans. Veh. Technol.*, vol. 49, no. 2, pp. 357–366, 2000.

[17] B. Li, L. Li, B. Li, K. M. Sivalingam, and X.-R. Cao, “Call admission control for voice/data integrated cellular networks: performance analysis and comparative study”, *IEEE J. Selec. Areas Commun.*, vol. 22, no. 4, pp. 706–718, 2004.

[18] B. Jabbari and W. F. Fuhrmann, “Teletraffic modeling and analysis of flexible hierarchical cellular network with speed sensitive handoff strategy”, *IEEE J. Selec. Areas Commun.*, vol. 15, no. 8, pp. 1539–48, 1997.

[19] F. Pavlidou, “Two-dimensional traffic models for cellular systems”, *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1505–1511, 1994.

[20] Md. Imdadul Islam and J. K. Das, "A mathematical model of traffic performance of mobile cellular network", *J. Elec. Comp. Sci.*, vol. 8, pp. 1–9, 2007.

[21] M. B. Bhuiyan *et al.*, "Performance evaluation of cognitive radio network under limited user traffic", *Jahangirnagar Univer. J. Elec. Comp. Sci.*, vol. 15, pp. 23–28, 2014.

Md. Imdadul Islam has completed his B.Sc. and M.Sc. in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 1993 and 1998, respectively. He has completed his Ph.D. degree from the Department of Computer Science and Engineering, Jahangirnagar University (JU), Dhaka, Bangladesh in the field of network traffic engineering in 2010. He is now working as a Professor at the JU Department of Computer Science and Engineering. Previously, he worked as an Assistant Engineer in Sheba Telecom, from 1994 to 1996. Mr. Islam has a good field experience in installation of radio base stations and switching centers for WLL. His research field is network traffic, wireless communications, wavelet transform, OFDMA, WCDMA, adaptive filter theory, ANFIS and array antenna systems. He has more than 150 research papers in national and international journals and conference proceedings.

E-mail: imdad@juniv.edu
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

Md. Fazlay Rabbi has completed his B.Sc. in Electrical and Electronics Engineering from United International University, Dhaka in 2011 and since 2013 he is doing his M.Sc. in Computer Science at Jahangirnagar University. He worked as a system support engineer in United International University, Bangladesh from 2012 to 2014. He also worked as an instructor in UIU CISCO Networking Academy from 2011 to 2014. He is now working as an ICT officer with an international NGO (Action Contre La Faim, France, ACF International, Bangladesh Mission). His research field is network traffic, network protocols, Virtual Private Network, wireless communications, VoIP, design network system and deployment.

E-mail: rabbi_mit34@yahoo.com
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

Risala Tasin Khan has completed her B.Sc. (Hons) in Computer Science and Engineering from Jahangirnagar University (JU), Savar, Dhaka in 2006 and M.Sc. in Computer Science and Engineering from JU. She is now doing Ph.D. at the JU Department of Computer Science and Engineering. She worked as a lecturer in the Department of Computer Science and Engineering, Daffodil International University, Dhaka. She is now working as an Assistant Professor at Institute of Information Technology, JU. Her research field is network traffic and network security.

E-mail: risala@juniv.edu
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

Jesmin Akhter received her B.Sc. degree in Computer Science and Engineering from Jahangirnagar University (JU), Savar, Dhaka, Bangladesh in 2004 and M.Sc. degree in Computer Science and Engineering from JU in 2012. Since 2008 she is an Assistant Professor at the Institute of Information Technology in JU. Her research areas are on network traffic, complexity and algorithms and software engineering. Now she is pursuing Ph.D. at the Department of Computer Science and Engineering, JU, in the field of 4G wireless networks.

E-mail: togorcse@gmail.com
Department of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

Paper

Multi-layered Bayesian Neural Networks for Simulation and Prediction Stress-Strain Time Series

Agnieszka Krok

Tadeusz Kościuszko Cracow University of Technology, Cracow, Poland

Abstract—The aim of the paper is to investigate the differences as far as the numerical accuracy is concerned between feed-forward layered Artificial Neural Networks (ANN) learned by means of Kalman filtering (KF) and ANN learned by means of the evidence procedure for Bayesian technique. The stress-strain experimental time series for concrete hysteresis loops obtained by the experiment of cyclic loading is presented as considered example.

Keywords—Bayesian Neural Networks, Kalman filtering.

1. Introduction

Kalman filtering and Bayesian learning methods are based on the same assumption of modeling neural networks as the combination of random variables. In both cases, ANNs are layered, feed forward, learned by supervised method with a teacher. Learning set and testing set are considered. Learning process is based on methods known from probability theory and statistical analysis: Kalman filtering and Bayes theorem. The aim of the paper is to make a comparison of the two approaches from the same family techniques.

As far as the network architecture, the most common multilayer perception ANN were considered. Two hidden layers were used for their ability to model nonlinear functions, according to the universal approximation theorem [1]. For a comparison purpose the same architecture of ANNs were considered. It results in the same number of ANN weights to be found. The Mean Squared Error (MSE) for learning and testing set was considered as the measure of learning efficiency. In addition, the qualitative criteria was examined. The shape of modeled time series is calculated by ANNs according to the experimental data. Possibilities of easy designing the network shape (number of neurons in each layer), the number of parameters that control the process of selection model and the time for implementing both methods were also verified.

2. Motivation and Related Background

Bayesian Neural Networks (BNN) are constructed as layered, feed forward networks learned by supervised methods that involves Bayes theorem [2]. The following four steps are considered:

1. make predictions including error bars for new input data;
2. estimate the weight parameters and their uncertainties.
3. estimate the weight decay parameters and their uncertainties;
4. repeat steps 2–3 with different initial conditions and different network architectures. Select the architecture and w-minimum with highest evidence. Optionally select a committee to reflect the uncertainty on this level [3].

The BBN were recently used for the case problems including a regression, a classification, and an inverse problem. The Internet traffic classification [4], modeling protein family [5], concrete quality estimation problem [6], assessment of lean manufacturing effect on business performance [7], medicine diagnoses finding [8], forecasting performances over the weekly sales of a Finance Magazine [9], image skin segmentation [10], classification of file system activities [11], analyzing weather data [12], classifying segmented outdoor images [13], were analyzed.

The traditional approach to the hysteresis modeling assumes using differential equation models that involves the parameters that are specific to the modeled material: Jiles–Atherton model [14], Ylinen's Model [15], Takács model [16], Prandtl–Ishlinskii model [17]. In most cases, the models are in the form of piece-wise functions different for the particular branches of the hysteresis [15], [18].

Also soft methods was considered: neural networks in the form of multi-layer perceptions, learned by the back-propagation algorithm for supervised training [19], [20], or the Levenberg-Marquardt algorithm [21]–[23] .

3. Kalman Filtering as the ANN Learning Technique

The KF as a method was adapted to ANN nonlinear models [24], as the learning technique and developed excessively using selected nodes learning and as far as pruning the ANN is concerned [25].

The basic KF learning method – Node Decoupled Extended Kalman Filter (NDEKF) consists of process equation and measurement equation. After modifications they may be adopted to learn standard Multi-layered ANN [26]:

$$\mathbf{w}^i_{k+1} = \mathbf{w}^i_k + \omega^i_k\,, \tag{1}$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{w}_k, \mathbf{x}_k) + \nu_k\,, \tag{2}$$

where: k – discrete pseudo-time parameter, i – the number of neuron in ANN- w^i_{k+1} for $i = 1, 2, \ldots, W$ – state vector corresponding to the set of synaptic weights and biases, $\mathbf{h}$ – non-linear vector-function of input-output relation, $\mathbf{x}/\mathbf{y}$ – input/output vectors, ω^i_k, ν_k – Gaussian process and measurement noises with mean and covariance matrices defined by:

$$\mathbf{E}(\nu_k) = \mathbf{E}(\omega^i_k) = 0\,, \tag{3}$$

$$\mathbf{E}(\omega^i_k * \omega^{i^T}_l) = \mathbf{Q}^i_k \delta_{k,l}\,, \tag{4}$$

$$\mathbf{E}(\nu_k \nu^T_l) = \mathbf{R}_k \delta_{k,l}\,, \tag{5}$$

where: $\delta_{k,l} = 1$ for $k = l$, $\delta_{k,l} = 0$ for k not equal l.

The NDEKF algorithm assumes splitting state vector into groups. The single group was assigned to single neuron (nodes $i = 1, 2, \ldots, N$). Similar to all teacher based learning techniques the change of $\mathbf{w}^i$ is made during the presentation of each k-th learning pattern:

$$\mathbf{K}^i_k = \mathbf{P}_k \mathbf{H}_k \Big[\sum_{j=1}^{g} (\mathbf{H}^j)^T_k \mathbf{P}^j_k \mathbf{H}^j_k + \mathbf{R}_k \Big]^{-1}\,, \tag{6}$$

$$\mathbf{w}^i_{k+1} = t\mathbf{w}^i_k + \mathbf{K}^i_k \xi_k\,, \tag{7}$$

$$\mathbf{P}^i_{k+1} = (\mathbf{I} - \mathbf{K}^i_k (\mathbf{H}^i_k)^T) \mathbf{P}^i_k + \mathbf{Q}^i_k\,, \tag{8}$$

where: $\mathbf{K}^i_k$ – Kalman gain matrix, $\mathbf{P}^i_k$ – approximate error covariance matrix, g – the number of ANN nodes (neurons), $\xi_k = \mathbf{y}_k - \widehat{\mathbf{y}}_k$ – error vector, with the target vector $\mathbf{y}_k$ for the k-th presentation of a training pattern, $\widehat{\mathbf{y}}_k$ – output vector given by ANN.

$\mathbf{H}$ is the matrix of current linearization of Eq. (2)

$$\mathbf{H}^i_k = \frac{\partial \mathbf{h}}{\partial \mathbf{w}^i}\,. \tag{9}$$

The considered parameters for the Gaussian noise adopted are e.g. in the form:

$$\mathbf{Q}^i_k = \alpha_1 \cdot \mathrm{e}^{\frac{s-1}{\beta_1}} \cdot I\,, \tag{10}$$

$$\mathbf{R}_k = \alpha_2 \cdot \mathrm{e}^{\frac{s-1}{\beta_2}} \cdot I\,, \tag{11}$$

where: I – identity matrix which dimension depends on the state vector dimension in ANN, s – the number of learning epoch, and α_1, α_2, β_1, β_2 are real numbers.

4. Bayesian Neural Networks

The ANN is formulated as [27]:

$$t = y(\mathbf{x}; \mathbf{w}) + \varepsilon\,, \tag{12}$$

where y is the non-linear vector function of input-output relation, ε – noise incorporated to the model, $\mathbf{w}$ – vector of ANN weights interpreted as the random variables, t is the target output variable interpreted as a random variable. Next, the

$$p(\mathbf{w}) \tag{13}$$

is the prior broad probability distribution of the $\mathbf{w}$, and representing little knowledge about values of $\mathbf{w}$:

$$p(\mathbf{w}|D) = \frac{p(D|\mathbf{w})p(\mathbf{w})}{p(D)} \tag{14}$$

is the posterior probability distribution of the $\mathbf{w}$. It representing knowledge about values of $\mathbf{w}$ after data set D is presented to the network, $p(D|\mathbf{w})$ is the data set likelihood.

$$p(t|\mathbf{x}^*, D) = \int p(t|\mathbf{x}^*, \mathbf{w}) p(\mathbf{w}|D) \mathrm{d}\mathbf{w} \tag{15}$$

is the predicted distribution of the ANN output y for the particular input vector $\mathbf{x}^*$;

$$E(t|\mathbf{x}^*, D) = \int t\, p(t|\mathbf{x}^*, \mathbf{w}) p(\mathbf{w}|D) \mathrm{d}\mathbf{w} \tag{16}$$

is the point prediction of the ANN output t for the particular input vector $\mathbf{x}^*$. The requirement for small values of $\mathbf{w}$ suggests a Gaussian prior distribution the the ANN weights

$$p(\mathbf{w}) = \frac{1}{Z_W(\alpha)} \cdot \mathrm{e}^{\frac{-\alpha||\mathbf{w}||^2}{2}}\,, \tag{17}$$

where α represents the inverse variance of the distribution of $\mathbf{w}$ and

$$\alpha = \frac{1}{D^2(\mathbf{w})}\,. \tag{18}$$

$Z_W(\alpha)$ is the normalization constant $Z_W(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{W}{2}}$ where W is the number of ANN weights.

It is assumed the target data is given by the Gaussian distribution with zero mean and the constant inverse variance β, so the data set likelihood $p(D|\mathbf{w})$ is

$$p(D|\mathbf{w}) = \frac{1}{Z_D(\beta)} \cdot \mathrm{e}^{\frac{-\beta \sum_{n=1}^{N} \|t(\mathbf{x}^n;\mathbf{w}) - t^n\|^2}{2}}\,, \tag{19}$$

where β represents the inverse variance of the ε distribution defined as:

$$\beta = \frac{1}{D^2(\varepsilon)}\,. \tag{20}$$

The $Z_D(\alpha)$ is the normalization constant given by $Z_D(\alpha) = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}}$, where N is the number of data point in D. Then assuming α, β are random variables with their own probability distributions:

$$p(t|\mathbf{x}^*, D) = \iint p(t|\mathbf{x}^*, \mathbf{w}, \beta) p(\mathbf{w}|\alpha, \beta, D) p(\alpha, \beta|D) \mathrm{d}\alpha \mathrm{d}\beta , \tag{21}$$

$$p(t|\mathbf{x}^*, \mathbf{w}, \beta) = N\big(t|t(\mathbf{x}^*, \mathbf{w}), \beta^{-1}\big) , \tag{22}$$

$$\ln p(\mathbf{w}|\alpha, \beta, D) = p(D|\mathbf{w}) . \tag{23}$$

5. Evidence Procedure for Bayesian Neural Networks

The evidence procedure was used as an iterative algorithm for determining optimal weights and hyper parameters during Bayesian learning of the ANN [28].
Presented method is based on the approximating the hyper parameters posterior distribution with its value at the most probable (MP) values

$$p(\mathbf{w}|D) \sim \int p(y|\mathbf{w}, \beta_{MP}) p(w|\alpha_{MP}, \beta_{MP}, D) \mathrm{d}w . \tag{24}$$

To find the MP values of α and β one have to find the maximum of:

$$p(\alpha, \beta|D) = \frac{p(D|\alpha, \beta) p(\alpha, \beta)}{p(D)} . \tag{25}$$

In the further calculation $p(\alpha, \beta)$ is assumed to be uniform and ignored. Maximizing $p(D|\alpha, \beta)$ equals finding the maximum of:

$$p(D|\alpha, \beta) = \int p(D|\mathbf{w}, \beta) p(\mathbf{w}|\alpha) \mathrm{d}\mathbf{w} \tag{26}$$

$$p(D|\alpha, \beta) = \frac{1}{Z_D(\beta)} \frac{1}{Z_W(\alpha)} \int \mathrm{e}^{-S(\mathbf{w})} \mathrm{d}\mathbf{w} , \tag{27}$$

where:

$$S(\mathbf{w}) = \frac{\beta}{2} \sum_{i=1}^{N} (y(\mathbf{x}^n; \mathbf{w}) - t^n)^2 + \frac{\alpha}{2} \sum_{i=1}^{W} \mathbf{w}_i^2 = \beta E_D + \alpha E_W \tag{28}$$

is the misfit function. The t^n and $y(\mathbf{x}^n; \mathbf{w})$ are the target and computed output values for n-th pattern scaled to the interval $0 \ldots 1$, $\mathbf{w} = w_i, \ldots, w_W$ is the vector of ANN weights. By computing the logarithm of the Eq. (28) and the partial derivative with respect to α one can obtain:

$$\alpha = \frac{W - \sum\limits_{i=1}^{W} \frac{\alpha}{\lambda_i + \alpha}}{2E_W(w_{MP})} = \frac{\gamma}{2E_W(w_{MP})} , \tag{29}$$

where $\mathbf{w} = \mathbf{w}_{MP}$, λ_i is the i-th eigenvalue of the Hessian matrix H:

$$H = \nabla\nabla E_D , \tag{30}$$

$$\gamma = \sum_{i=1}^{W} \frac{\lambda_i}{\lambda_i + \alpha} . \tag{31}$$

This implicit solution is used for the iterative procedure: after setting initial values of α that is used to find $\mathbf{w}_{MP}$ and $S_W(\mathbf{w}_{MP})$ the α is re-estimated according to [6]:

$$\alpha = \frac{\gamma}{2E_D(\mathbf{w}_{MP})} , \tag{32}$$

where $\mathbf{w} = \mathbf{w}_{MP}$.
By computing the logarithm of the objective function and the partial derivative with respect to β one can obtain:

$$\beta = \frac{N - \gamma}{2E_D(\mathbf{w}_{MP})} . \tag{33}$$

The procedure scheme can be written in following steps:

1. Choose the initial values of hyper parameters α and β, initial ANN weights drown from prior distribution given by α,

2. Train the ANN with Scaled Conjugate Gradients Algorithm (SCGA) [1], to minimize negative log probability of weight posterior probability misfit function $S(\mathbf{w})$, where $N = L$ is the number of learning patterns to find $\mathbf{w}_{MP}$,

3. Hyper parameters re-estimate:

$$\alpha(new) = \frac{\gamma}{2E_W(\mathbf{w}_{MP})} , \tag{34}$$

and

$$\beta(new) = \frac{N - \gamma}{2E_D(\mathbf{w}_{MP})} , \tag{35}$$

4. Update the log evidence

$$p(D|\alpha(new), \beta(new), \gamma) , \tag{36}$$

5. Repeat steps 2–4 until convergence.

Number of training cycles is the steps number during SCGA performance, number of inner loops is the number of updating $\alpha \Rightarrow \alpha(new)$, $\beta \Rightarrow \beta(new)$, number of outer loops is the number of repeating the $\mathbf{w}$ re-estimation.

6. Experimental Results for Simulation and Prediction of Steel Hysteresis Loops

6.1. Experimental Data

Many time series for simulation and prediction stress-strain relation was considered for steel and concrete. In this paper, the one specific numerical result would be presented. The main tendency and numerical accuracy during modeling the rest of the data was similar. All the tested examples may be found in [29].
Presented data set was the result of uniaxial low cyclic tension-compression test for stainless steel AISI 316L [24], see Fig 1.

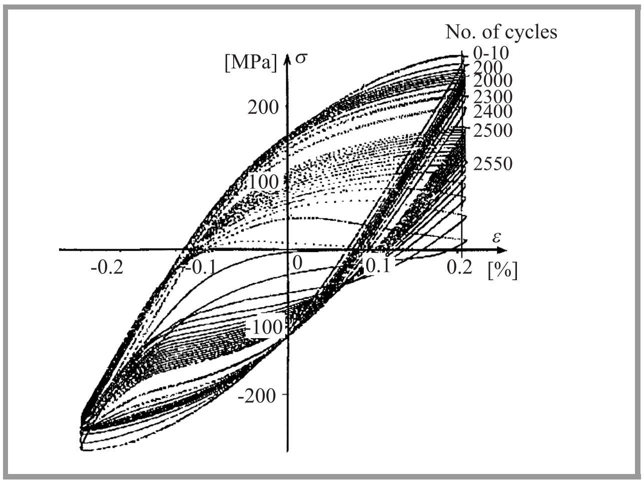


Fig. 1. Experimental data on σ-ε plane.

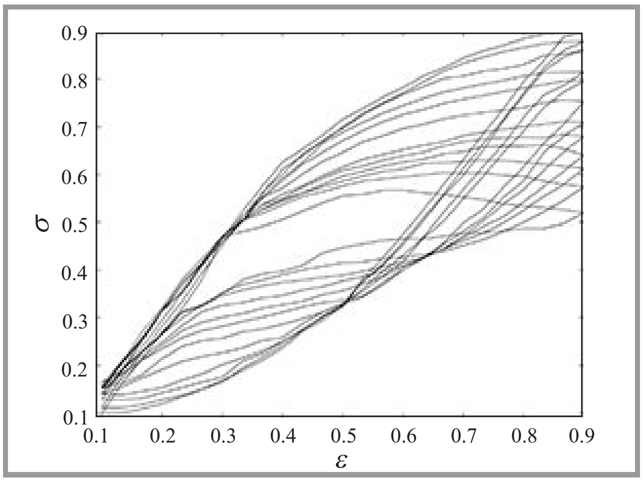


Fig. 2. Data on σ-ε plane for ANN learning and testing.

The aim of the conducted neural analysis was to simulate first part of experiment and to predict the phase before the material damage. Presented time series is based on the inner processes inside the material and is self-dependent because each next state of the material depends on all the previous states during experiment. During learning stage, time series simulation was performed, and whole testing stage time series prediction was made.

The twelve representative loops were selected for the neural computation. The loops were selected to the constant maximal stain value inclination. The first and second loop selected for neural analysis were taken from range of 0–2000 experimental loops, three next loops from range of 2000–24000 experimental loops, the remaining 7 loops from range, were the changes in stress and stein values were the largest. Each loop was discretized on 49 $(\sigma(k), \varepsilon(k))$ points for:

$$\varepsilon(k) = 0.2 - (k-1) \cdot \Delta_1 \varepsilon \,, \tag{37}$$

for $k = 1, 2, \ldots, 25$ and

$$\varepsilon(k) = -0.26 + (k-25) \cdot \Delta_2 \varepsilon \,, \tag{38}$$

for $k = 26, \ldots, 49$ with

$$\begin{aligned} \Delta_1 \varepsilon(k) &= 0.2/25 = 0.008\% \,, \\ \Delta_2 \varepsilon(k) &= 0.26/25 = 0.0104\% \,. \end{aligned} \tag{39}$$

Adopted discretization results in $P = 12 \cdot 49 = 588$ data points for learning and testing. Given data ware scaled to the interval $0.1 \ldots 0.9$ for the ANN processing, see Fig. 2.

The first nine loops containing $L = 949$ patterns for the $k = 1, \ldots, 441$ were used for the learning and $T = 588 - 441 = 147 = 3 \cdot 49$ patterns form final loops for $k = 442, \ldots, 588$ for testing.

The input vector $\mathbf{x}$ consists of scaled marker of current pattern $k/587$, scaled marker of current pattern number inside each loop separately $\mathrm{mod}(k, 49)/49$ and the previous σ value given by ANN, marked $\sigma_{ANN}(k-1)$ [29]:

$$\mathbf{x}(k) = \left[\sigma_{ANN}(k-1), k/587, \quad \mathrm{mod}\,(k, 49)/49\right]. \tag{40}$$

The output vector for k-th input takes the form $\sigma(k)$.

7. Comparison of Neural Networks Learned by Bayesian Evidence Procedure Accuracy with KF Models

The basic KF model is simple, that makes is easy to implement. However, the model does not have many parameters to exploit. One may split ANN differently (layers not nodes) [26], or use different parameters for the Gaussian noise, α_1, α_2, β_1, β_2 in Eqs. (10)–(11). Also different noise models instead of Eqs. (10)–(11) may be adopted. First possibility enlarges excessively the model dimensions given by Eqs. (1)–(5) makes the model very time consuming. This second option does not influence the computational results much, see [29]. The possibility of automation of setting the network structure without the stopping learning process is very valuable. One may start from the large network and switch off the some of the network nodes during learning (pruning). The author's model development proved that ANN learned by KF may be successfully designed by pruning, and the approximate error covariance P, matrix may be used to more accurate learning, see Eq. (8) [29], [30]. KF learning technique was stated to be very promising tool as far as time series simulation and prediction [29], [31]–[34].

In comparison to KF, Bayesian learning technique is much more complicated to implement but have many more free parameters to change to adjust the model. It allows the better flexibility, but incorporates the problem of searching the parameter space for a suboptimal solution. For example the changing of characteristics of hiper parameter distribution significantly influence the model, see Eqs. (19)–(36). During pruning process the model given by Eqs. (24)–(36) have to be reformulated and there is a need to compute prior for sparsely connected ANN [35]. The method also depends on the SCGA performance that should be implemented and used correctly, whereas the in KF no additional high-level tool is needed.

The simulation for Bayesian learning process was made using modified Netlab Software [1]. Simulation for KF was made for software developed fully by author in Matlab environment. The same ANN architecture was considered, and the same networks input vectors were used, see Eq. (40).

For Bayesian learning the initial prior hyper parameter $\alpha = 0.01$, initial noise hyper parameter $\beta = 50$, number of training cycles in inner loop 500, number of inner loops 3, number of outer loops 3 was found as the suboptimal solution for the given data set. The results of ANN simulation and prediction are presented in Figs. 3 and 4.

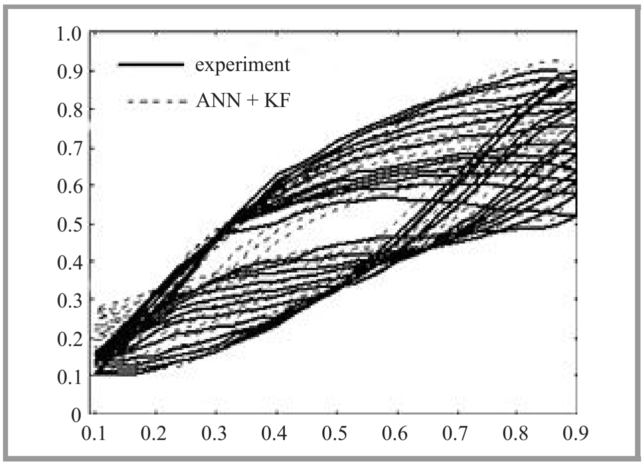


Fig. 3. Experimental vs. simulated by KF neural network hysteresis loops, 9 first loops for learning and remaining for testing on σ-ε plane.

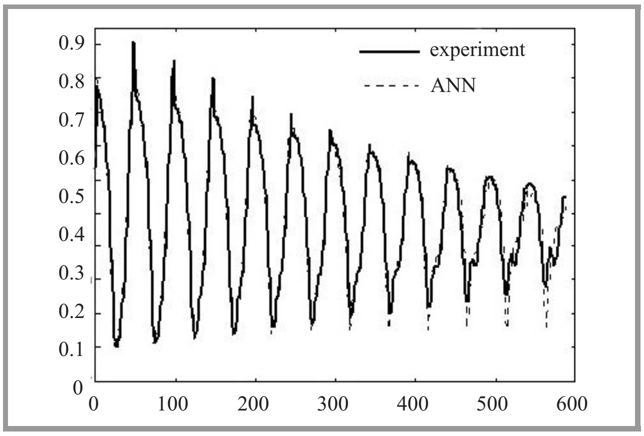


Fig. 4. Experimental vs. simulated by KF neural network hysteresis loops, 9 first loops for learning and remaining for testing on σ – no. of pattern plane.

For KF learning method with 1000 epochs and $\alpha_1 = 0.001$, $\alpha_2 = 7$, $\beta_1 = 50$, $\beta_2 = 50$ was adopted. The results of ANN simulation and prediction are presented in Figs. 5 and 6.

The presented method of ANNs learning enables simulation of the hysteresis loops with a very high accuracy using ANN of small number of parameters (first hidden layer 6 nodes, second hidden layer 6 nodes). ANN predicts the behavior of the considered material during the final step of loading and unloading properly.

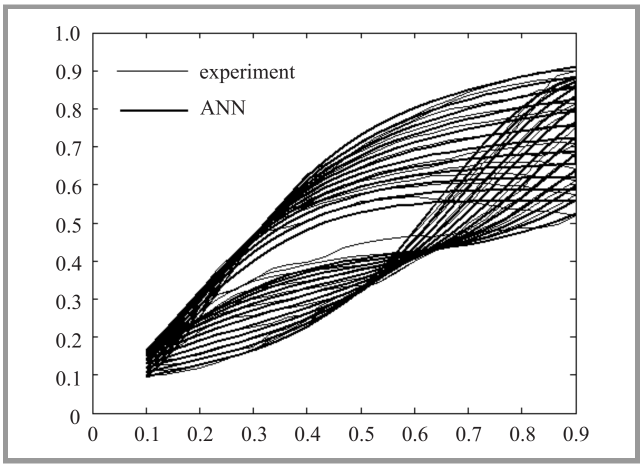


Fig. 5. Experimental vs. simulated by Bayesian neural network hysteresis loops, 9 first loops for learning and remaining for testing on σ-ε plane.

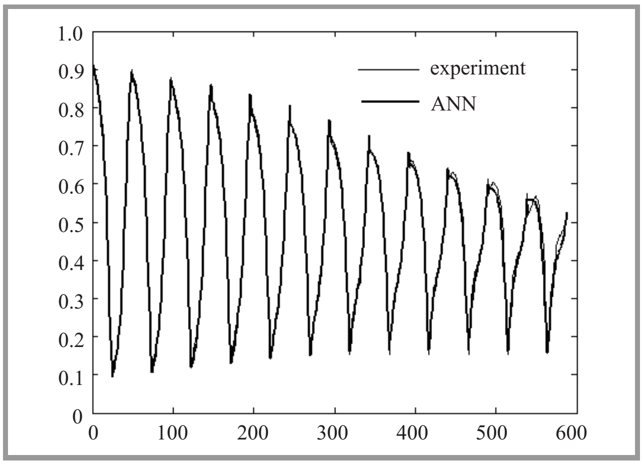


Fig. 6. Experimental vs. simulated by Bayesian neural network hysteresis loops, 9 first loops for learning and remaining for testing on σ – no. of pattern plane.

Comparison with the results obtained by modeling with damage mechanics was made.

In [36] two theoretical models for stress-strain relation for the considered material were proposed. They were based on uniaxial nonlinear elasto-plastic Ylinen model [37]. The relationship between considered quantities had to be separated for two phases, two damage parameters was needed to obtain the proposed models. The models took into consideration the discrete process of opening and close of the cracks, see Fig. 7 for model A, and continuous process of opening and close of the cracks, see Fig. 8 for model B.

Proposed model A is inconsistent with the experiment as far as continuity of first order derivative of $\partial\sigma/\partial\varepsilon$ is considered. Both models are incorrect concerning negative values of stress strain close to their minimal values.

Simulating the stress-strain relation, using damage mechanics equations, require the values of material constants to be properly chosen for the theoretical model and calibration

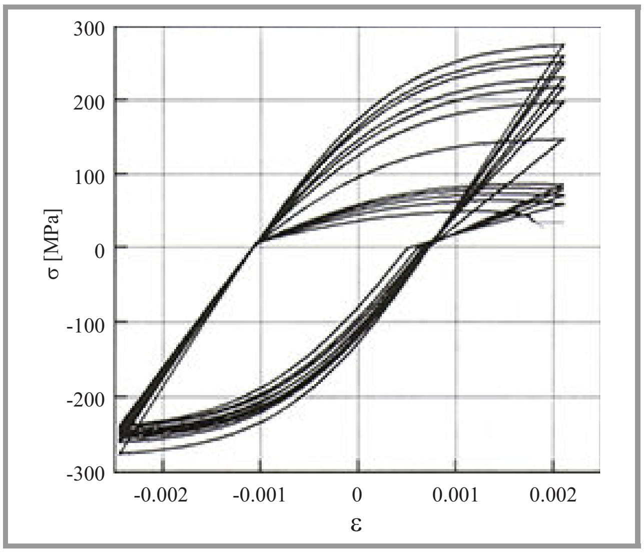


Fig. 7. Hysteresis loops for model A, σ-ε plane.

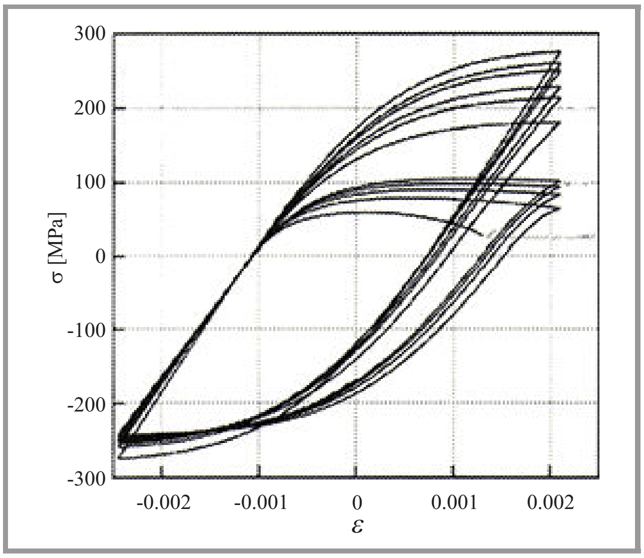


Fig. 8. Hysteresis loops for model B, σ-ε plane.

as far as model parameters are considered. During ANN modeling none of this component is necessary. The model is based only on the experimental data and simple markers of the experiment phase. Information about mechanical effects is independent of arbitrarily chosen mechanical model and imposed model parameters.

KF and Bayesian ANN modeling incorporates no prior knowledge about mechanical model.

8. Conclusions

The evidence procedure for the Bayesian Neural Networks enables hysteresis loops simulation with a very high accuracy as far as results quality is concerned. The model fits the data better than model based on KF and it is superior to known mechanical models. Presented Bayesian model has three basic parameters to set: Bayesian learning initial prior hyper parameter α, initial noise hyper parameter β, and number of training cycles in inner loop k.

Presented KF model has five parameters. The coefficients for Gaussian noise incorporated into the model (four values), and number of epochs of learning.

The influence of the Bayes model parameters is meaningful. It enables the model to adjust to the data better, but makes searching for optimal parameter set more demanding. For some set of parameters model calibrated to the data is incorrect or significantly worse.

The influence of the KF model parameters is hardly a significant. Different parameter setting leads to slightly longer teaching. The most significant parameter is the number of epochs of learning.

For both models, there is also the need for searching initial weight space. For different initial weight sets the results of simulation and prediction differs slightly.

The automatic setting of neural network shape, i.e. number of neurons in each hidden layer, by pruning procedure during learning process is much easier to implement in KF model. To adjust both models to different kind of data the distributions incorporated into models may be changed. However, any change in Bayesian theoretical model has more severe consequences into computational process, because all Eqs. (12)–(36) have to be changed.

Acknowledgements

Author would like to acknowledge support from the Polish Ministry of Science and Higher Education Grant “Application of Bayesian neural networks in experimental mechanics of structures and materials”.

References

[1] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice Hall, 1999.

[2] T. Bayes, “An Essay towards solving a problem in the doctrine of chances”, *Philosoph. Trans. of the Royal Society of London*, vol. 53, pp. 370–418, 1763.

[3] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.

[4] T. Auld, A. W. Moore, and S. F. Gull, “Bayesian neural networks for Internet traffic classification”, *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 223–239, 2007.

[5] D. MacKay, “Bayesian neural networks and density networks”, *Nuclear Instrum. and Methods in Phys. Res. Section A*, vol. 354, no. 1, pp. 73–80, 1995.

[6] J. Lampinen and A. Vehtari,“ Bayesian approach for neural networks and case studies”, *Neural Netw.*, vol. 14, no. 3, pp. 257–274, 2001.

[7] G. Büyüközkan, G. Kayakutlu, and I. S. Karakadılar, “Assessment of lean manufacturing effect on business performance using Bayesian Belief Networks”, *Expert Syst. with Applications*, vol. 42, no. 19, pp. 6539–6551, 2015.

[8] I. Kononenko, “Bayesian neural networks”, *Biolog. Cybernetics*, vol. 61, pp. 361–370, 1989.

[9] O. Kocadağli and B. Aşıkgil, “Nonlinear time series forecasting with Bayesian neural networks”, *Expert Syst. with Applications*, vol. 41, no. 15, pp. 6596–6610, 2014.

[10] A. A. Zaidan *et al.*, “Image skin segmentation based on multi-agent learning Bayesian and neural network”, *Engin. Appl. of Artif. Intell.*, vol. 32, pp. 136–150, 2014.

[11] M. N. A. Khan, "Performance analysis of Bayesian networks and neural networks in classification of file system activities", *Comp.& Secur.*, vol. 31, no. 4, pp. 391–401, 2012.

[12] H. S. Hippert and J. W. Taylor, "An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting", *Neural Netw.*, vol. 23, no. 3, pp. 386–395, 2010.

[13] Ch. K. I. Williams and F. Vivarellia, "Comparing Bayesian neural network algorithms for classifying segmented outdoor images", *Neural Netw.*, vol. 14, no. 4–5, pp. 427–437, 2001.

[14] M. Hamimid, S. M. Mimoune, and M. Feliachi, "Minor hysteresis loops model based on exponential parameters scaling of the modified Jiles–Atherton model", *Physica B: Condensed Matter*, vol. 407, no. 13, pp. 2438–2441, 2012.

[15] A. Ganczarski and L. Barwacz, "Concept of continuous damage deactivation in modelling of low cycle fatigue", *in Proc. Int. Conf. on Fracture ICF11 2005*, Turin, Italy, 2005.

[16] K. Chwastek, "Modelling hysteresis loops in thick steel sheet with the dynamic Takács model", *Physica B: Condensed Matter*, vol. 407, no. 17, pp. 3632–3634, 2012.

[17] M. Al Janaideh, "A time-dependent stop operator for modeling a class of singular hysteresis loops in a piezoceramic actuator", *Physica B: Condensed Matter*, vol. 413, pp. 100–104, 2013.

[18] A. P. S. Baghel, A. Gupta, K. Chwastek, and S. V. Kulkarni, "Comprehensive modelling of dynamic hysteresis loops in the rolling and transverse directions for transformer laminations", *Physica B: Condensed Matter*, vol. 462, pp. 86–92, 2015.

[19] I. Kucuk, "Prediction of hysteresis loop in magnetic cores using neural network and genetic algorithm", *J. Magnetism and Magnetic Materials*, vol. 305, no. 2, pp. 423–427, 2006.

[20] R. Dong, Y. Tan, H. Chen, and Y. Xie, "A neural networks based model for rate-dependent hysteresis for piezoceramic actuators", *Sensors and Actuators A: Physical*, vol. 143, no. 2, pp. 370–376, 2008.

[21] A. Nouicer, E. Nouicer, and F. Mouloudc, "A neural network for incorporating the thermal effect on the magnetic hysteresis of the 3F3 material using the Jiles–Atherton model", *J. Magnetism and Magnetic Materials*, vol. 373, pp. 240–243, 2015.

[22] V. Wolfs and P. Willems, "Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural networks", *Environ. Model. & Softw.*, vol. 55, pp. 107–119, 2014.

[23] X. Zhang, Y. Tan, and M. Su, "Modeling of hysteresis in piezoelectric actuators using neural networks", *Mechan. Syst. and Sig. Process.*, vol. 23, no. 8, pp. 2699–2711, 2009.

[24] R. E. Kalman, "A new approach to linear filtering and prediction problems", *Trans. ASME – J. of Basic Engin.*, vol. 82, series D, pp. 35-45, 1960.

[25] L. Prechelt, "Adaptive parameter pruning in neural networks", ICSI Tech. Rep. TR-95-009, Berkeley, CA, USA, 1995.

[26] S. Haykin, Ed., *Kalman Filtering and Neural Networks*. New York: Wiley, 2001.

[27] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[28] R. M. Neal, "Bayesian Learning for Neural Networks", *Lecture Notes in Statistics*, vol. 118. Springer, 1996.

[29] A. Krok, "Analysis of mechanics of structures and material problems applying artificial neural networks learnt by means of Kalman ftering", Ph.D. thesis, Institute of Computer Mhetods in Civil Engineering, Cracov University of Techology, 2007 (in Polish).

[30] A. Krok, "The development of Kalman filter learning technique for artificial neural networks", *J. Telecom. Inform. Technol.*, no. 4, pp. 16–21, 2013.

[31] A. Krok and Z. Waszczyszyn, "Kalman filtering for neural prediction of response spectra from mining tremors", in *Proc. Int. Conf. Artif. Intell. AI-METH 2004*, Gliwice, Poland, 2004.

[32] A. Krok and Z. Waszczyszyn, "Simulation of building loops for a superconductor using neural networks with Kalman filtering", *Comp. Assisted Mechanics and Engin. Sci.*, vol. 13, pp. 575–582, 2006.

[33] A. Krok and Z. Waszczyszyn, "Kalman filtering for neural prediction of response spectra from mining tremors", *Computers and Structures*, vol. 85, no. 15–16, pp. 1257–1263, 2007.

[34] A. Krok, "Simulation of concrete hysteresis loops using gaussian processes calibrated by genetic algorithm", in *Proc. Int. Symp. IPM on Inverse Problems in Mechanics of Struc. Mater.*, Rzeszów-Łańcut, Poland, 2009.

[35] H. H. Thodberg, "Ace of bayes: application of neural networks with pruning", Tech. Rep. 1132E, The Danish Meat Research Institute, 1993.

[36] A. Ganczarski and L. Barwacz, "Low cycle fatigue based on unilateral damage ewolution", *Int. J. Damage Mech.*, vol. 16, no. 2, pp. 159–177, 2007.

[37] A. Ylinen, "A method of determining the buckling stress and the required cross-sectional area for centrally loaded straight columns in the elastic and inelastic range", *Int. Assoc. for Bridges and Struct. Engin.*, vol. 16, pp. 529–550, 1956.

[38] I. T. Nabney, *Netlab: Algorithms for Pattren Recognition*. Springer, 2002.

[39] H. H. Thodberg, "Review of Bayesian neural networks with application to near infrared spectroscopy", *IEEE Trans. Neural Netw.*, vol. 7, no. 1, pp. 56–72, 1996.

[40] Y. S. Othmana *et al.*, "Frequency based hysteresis compensation for piezoelectric tube scanner using Artificial Neural Networks", *Procedia Engin.*, vol. 41, pp. 757–763, 2012.

Agnieszka Krok received her M.Sc. in the field of stochastic processes at the Jagiellonian University, Cracow, Poland and Ph.D. degree in the field of neural networks at Tadeusz Kosciuszko Cracow University of Technology, Poland, in 2003 and 2007, respectively. From 2009 she is an Assistant Professor at Faculty of Physics, Mathematics and Computer Science, Tadeusz Kościuszko Cracow University of Technology. Her main scientific and didactic interests are focused mainly on artificial intelligence: artificial neural networks, genetic algorithms, and additionally on parallel processing and cryptography.

E-mail: agneskrok@gmail.com

Faculty of Physics, Mathematics
and Computer Science
Tadeusz Kościuszko Cracow University of Technology
Warszawska st 24
31-155 Cracow, Poland

Paper

The Analysis of OpenStack Cloud Computing Platform: Features and Performance

Daniel Grzonka

Institute of Computer Science, Tadeusz Kościuszko Cracow University of Technology, Cracow, Poland

Abstract—Over the decades the rapid development of broadly defined computer technologies, both software and hardware is observed. Unfortunately, software solutions are regularly behind in comparison to the hardware. On the other hand, the modern systems are characterized by a high demand for computing resources and the need for customization for the end users. As a result, the traditional way of system construction is too expensive, inflexible and it doesn't have high resources utilization. Present article focuses on the problem of effective use of available physical and virtual resources based on the OpenStack cloud computing platform. A number of conducted experiments allowed to evaluate computing resources utility and to analyze performance depending on the allocated resources. Additionally, the paper includes structural and functional analysis of the OpenStack cloud platform.

Keywords—cloud computing, high performance computing, OpenStack, parallel environments, resource utilization analysis, virtualization.

1. Introduction

The rapid development of computer hardware caused a new issue for software developers. With the increase of computing power, the problem with efficient use of physical resources occurred. In many cases, an increase in the available resources, especially in the number of computational units, can have opposite effects and cause significant decrease in performance. Therefore, despite the fact that the development of computer hardware and programming languages has enabled the creation of a much more complex systems, the development of software techniques doesn't keep up with the technological development. This phenomenon is called software crisis [1], [2]. Over the years there has been proposed a number of solutions, i.e., new paradigms, programming languages, hardware architectures or software solutions, which are more or less, eliminate that problem. One of the latest is known as cloud computing based on abstraction of resources and ability to user's needs adoption, but the challenge is still actual [3], [4]. Therefore, in the present work, attempts are made to analyze both resource consumption and performance basing on OpenStack cloud computing platform.

The research from this paper is an extended version of preliminary results published in [1] and presented at the 29th European Conference on Modeling and Simulation (ECMS 2015) in Albena, Bulgaria.

The paper is organized as follows. In Section 2 the idea of resources abstraction is presented. Section 3 describes objectives, assumptions, and architecture of cloud computing. The structural analysis of OpenStack cloud computing platform is presented in Section 4. Finally, the Section 5 describes conducted experiments and evaluates them. The paper is summarized and concluded in Section 6.

2. Idea of Resources Abstraction

In the 1970s, when mainframe computers gain great popularity, the significant waste of computing resources caused by the single application execution at a time was observed. Due to the their high cost, to remedy this problem, the new idea of virtualization was proposed. Initially, the virtualization was defined as parallel work environments on a single computer and was considered as a method for logically dividing mainframes. This division allowed running multiple applications simultaneously [5].

In the second half of the 1990s, when personal computers obtain a sufficiently high computing power, the use of virtualization becomes a popular solution. The continuous development of idea allowed to formulate general definition of virtualization as technique for simulating the software and the hardware upon which other software runs [6]. In accordance to this definition, the term includes widely understood abstraction of various software and hardware areas like operating systems, storage, memory, networks, servers, CPUs and other. The key benefits are reduction of energy, computations, heat generation and maintenance costs, better resources utilization, space-saving for computers, the ability to adapt to the end user needs, safety increase and high flexibility in system design [1], [6]–[8].

Virtualization is based on the coexistence of host and guest machines. The host machine is the actual machine on which the virtualization takes place, and the guest machines are the virtual machines operating through the host. The tool supervising the virtual machines is hypervisor (or virtual machine monitor). When the hypervisor replaces host system we are dealing with native (or bare-metal) hypervisor. If the hypervisor operates under the operating system, we are dealing with hosted hypervisor. The most popular hypervisors are KVM, VirtualBox, VMware ESX, Xen, Oracle VM, Microsoft Hyper-V [1], [7], [9].

The virtualization method, due to the way of implementation, can be divided on two approaches, namely paravirtualization and full virtualization [10]. Full virtualization might be based on both hosted and native hypervisor that provide virtual hardware for virtual guest operating system. Guest operating system has the impression that it operates on a real, physical hardware. In fact, the virtualized OS instructions, which would interfere with the activities of other virtualized environments or host operating system, are captured and executed by virtualization layer. Performance in this solution is, unfortunately, less than in paravirtualization, but in return it does not impose any restrictions or special requirements for guest machine [6], [10].

Paravirtualization bases on native hypervisor, that operates directly on hardware and provides it to virtual machines and guest operating systems. This solution requires modifying the guest kernel to translate nonvirtualizable instructions to hypercalls that communicate directly with the native hypervisor. The mechanism provides very good performance that is very close to the system native operation. Very good performance for the virtual machines is huge paravirtualization advantage. The disadvantage is the need to modify the systems kernel or the use of dedicated systems [6], [10].

3. Cloud Computing Technology

Linking the virtualization with distributed systems and with the concept of services resulted in the new idea called cloud computing. Cloud computing is defined as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [11]–[13].

Cloud computing paradigm is based on flexible idea of services providing, where the virtual machines are the base for delivering any cloud services. A service can be defined by a number of virtualized functionalities provided by servers including hardware, software, storage, computing capacities and infrastructure with server rooms. Services are usually priced on a pay-per-use business model or subscription fee, and the access is realized via Internet with thin-client, e.g., mobile device, tablet or computer with Web browser. Additionally, this approach allows for effective resource utilization, flexible developing of complex systems, customization of services, elastic system management, on-demand resource provision and, as a result, reduction of energy consumption and maintain costs [9], [12]. The services are referred as:

- Infrastructure as a Service (IaaS),
- Platform as a Service (PaaS),
- Software as a Service (SaaS).

IaaS is a model that relies on providing client IT infrastructure, i.e., hardware like machines, network, storage, software and support. In this model the client have the remote access to infrastructure. This level is responsible for managing and virtualization of the resources.

PaaS provides ready, often tailored to the needs of the user, application set. This model does not require to purchasing hardware or software installation – all the necessary programs are located on servers provider. The customer on his side has access to the interface via Internet and a thin client.

In SaaS customer receives a needed functionality and he doesn't have to worry about the infrastructure and the work environment. This model provides access to specific, functional tools. Programs are running on the provided server. The customer is not obliged to purchase software licenses – only is paying for every use, and access to them is obtained upon request [4], [13], [14].

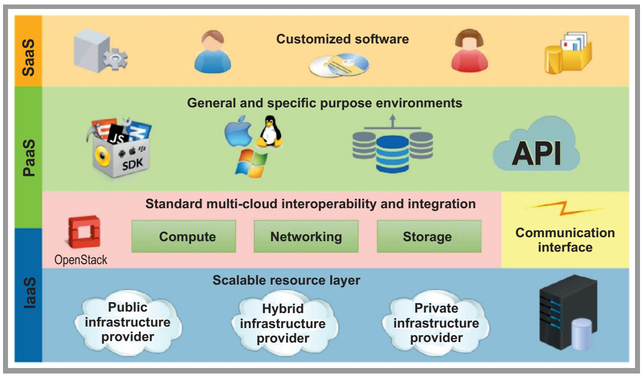


Fig. 1. Cloud computing fundamental models.

Figure 1 presents general model of cloud computing consisting of three levels: IaaS, PaaS and SaaS. In terms of internal policies, the following part can be distinguished [4]:

- Private cloud, which is part of the organization or autonomous provider of services;
- Public cloud, which is offer by external, publicly accessible provider (e.g., Amazon, Google, Microsoft);
- Hybrid cloud, which is a combination of private and public clouds. A part of the applications and infrastructure is placed in the private cloud, and the rest in the public cloud.

Benefits of cloud computing include [4], [15]:

- universal access – the user can use the service from anywhere in the world, using devices with Internet access. On the other hand, cloud computing may cause excessive network traffic;
- saving money – user does not need to maintain the entire infrastructure and is exempt from the costs of purchase and repair servers, maintenance of rooms and charges for electricity;

- performance and scalability – at any time the user may obtain the required hardware resources;
- reliability – thanks to the distribution of resources, and the systems flexibility in case of failure, the virtual environment is recreated on the new infrastructure;
- popularity – services in cloud computing are becoming more widely available, even for small businesses;
- security – use of cloud computing reduces the risk of errors related to the human factor. Simultaneously, the virtualization technology greatly increases the operations security level in cloud computing environments.

4. The OpenStack Project

OpenStack is a free and open source platform under the terms of the Apache license that possesses a set of tools for the creation and management of private, public and hybrid cloud computing. The software is developed for a control of wide range of processing, storage and networking resources throughout a data centre. It can be treated as an Infrastructure as a Service model strongly connected with Platform as a Service model (see Fig. 1). OpenStack manage the IT infrastructure, provide communication interface, virtualizes resources and prepare environment. It provides a modular architecture that gives the flexibility in the clouds design, including integration with existing systems and third-party technologies, e.g., Amazon EC2, GoGrid, Rackspace [1], [16].

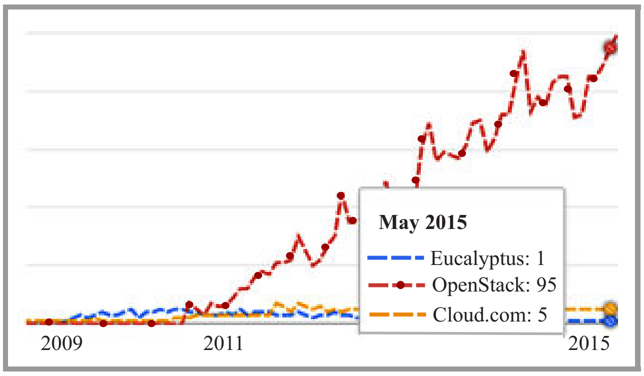


Fig. 2. Interest rate of most popular cloud platforms.

The OpenStack Cloud Computing Platform is probably the most popular open source software for creating and managing private, public, and hybrid clouds. Referring to the Google Trends Explore [17], [18], the OpenStack is the most popular platform among three the most known cloud platforms, i.e., Eucalyptus, OpenStack and Cloud.com. Figure 2 presents interest rate basing on the number of searches have been done for a particular term in relation to the total number of searches done on Google over time (a higher ratio means higher interest).

OpenStack was started in 2010 as a common project of Rackspace (a large US hosting company) and NASA (the US space agency). Currently, it is managed by the OpenStack Foundation, a non-profit corporate entity established in 2012. At the moment, the foundation has over 18,000 members from over 150 countries around the world. The purpose of the foundation is to promote the OpenStack software and its community, which develop project. The Foundation is composed of three committees: Board of Directors, Technical Committee and User Committee [1], [16].

The software is built modular and consists of many services working together through the open APIs. The newest version of the OpenStack platform – Kilo (2015.1.0) released on 30 April 2015 – may use the following modules (some of which are optional): Nova, Glance, Swift, Horizon, Keystone, Neutron, Cinder, Heat, Ceilometer, Trove, Sahara, Ironic, Zaqar, Manila, Designate, Barbican [1], [16]. The most important and fundamental are the first three components – they are implementing major features [19]:

- Compute (Nova service) – module for arranging, managing and providing virtual machines. It is "designed to provide power massively scalable, on demand, self service access to compute resources" [16];
- Object Store (Swift service) – module for creation and managing object storage system;
- Image Service (Glance service) – module which provide a service for uploading and discovering data assets. It retrieves and process data about virtual machine images.

The installation can be carried out in several ways, e.g. [1], [16], [19], [20]:

- directly from GitHub repository,
- Ubuntu Juju (based on installation Metal as a service layer),
- Vagrant OpenStack Provider (popular tool for creation and configuration of virtual environments),
- DevStack (tool for the installation of the platform from source with the set of initial configuration).

Details of the installation process are described in a previous publication [1].

The platform is open source and can be modified and adapted as needed. Services can be managed from Horizon dashboard (single Web-based on user interface) or by custom developed software. The access and communication is possible after authentication based on security certificates, which are the responsibility of Keystone Identity Service [1], [16].

5. Experimental Results

In this section the results of conducted experiments are presented. The study include:

- in-depth analysis of resources (CPU and memory) utility (both at full load and without) and consumption,
- the performance measurement of tasks execution depending on selected OpenStack set of virtualized resources,
- evaluation of effectiveness of the applied technologies.

In this research, 5 sets of virtual resources named flavors were considered. These were, namely: tiny, small, medium, large and extra large sets presented in the Table 1. Default environment configuration allowed to create up to 10 instances and assigning among others 20 VCPUs, 50 GB of memory and 1 TB of storage.

Table 1
Available sets of the virtual resources

Flavor	VCPUs	Disk [GB]	RAM [MB]
m1.tiny	1	1	512
m1.small	1	20	2048
m1.medium	2	40	4096
m1.large	4	80	8192
m1.xlarge	8	160	16384

OpenStack provides two virtual operating systems images by default [16]:

- CirrOS (9.3 MB) – a minimal Linux distribution that was designed for use as a test image on clouds,
- Trusty Tahr (244.4 MB) – distribution of Ubuntu 14.04.

According to OpenStack minimum requirements for proof-of-concept environment with several instances of CirrOS the following resources are needed [16]:

- Controller node: 1 processor, 2 GB memory, and 5 GB storage,
- Network node: 1 processor, 512 MB memory, and 5 GB storage,
- Compute node: 1 processor, 2 GB memory, and 10 GB storage.

The initial part of the research was based on CirrOS Linux distribution. The experimental server (specification is presented in Table 2) allowed to create 2 virtual compute nodes with maximum 3 instance per node [1].

Firstly, the impact of the existence of running instances in environment was measured. Analyzed parameters were CPU and RAM usage. For this purpose the 3 non-load instances had been deploying simultaneously every time.

Table 2
Specification of Pinokio experimental server

Processor	Intel Core i7-4710HQ CPU @ 2.50 GHz
Cores	4
Threads	8
Architecture	x86 64 bit
Virtualization	Intel VT-x
Memory	16 GB SODIMM DDR3 Synchronous 1600 MHz (0.6 ns)
Disk	256 GB ADATA SP600
Operating system	Ubuntu Server 14.10 (x64)

The results are presented in Table 3. With increasing number of running instances, the growing CPU usage was observed. Each started instance irreclaimable get and use about 2% of computing power (about 5–7% per 3 running instances). The operating memory usage was constant [1].

Table 3
Resources consumption caused by running instances

Instances	CPU usage [%]	RAM usage [%]
0	5	97.1
3	12	97.1
6	18	97.1

Next, resources usage was analyzed in the full load scenario. In this study, the matrix multiplication was performed as the CPU intensive task. It is one of the most popular benchmark tasks [21]. The size of used matrices was 1024×1024. The results of measurements are presented in Table 4. It should be noted, that instances number 1, 3, and 5 were automatically assigned to first compute node. As it can be seen, the full use of available resources was not reached. This is due to the amount of virtual computing machines limited by server hardware specification. But the obtained results lead to the conclusion that the OpenStack platform allows to fully use available computing resources [1].

Table 4
Resources consumption caused by running instances

Instances	CPU usage [%]	RAM usage [%]
0	5	97.1
1	19	97.2
2	32	97.3
3	41	97.3
4	51	97.4
5	55	97.5
6	56	97.5

The extended part of the research was based on Ubuntu 14.04 Trusty Tahr Linux distribution. The experiments were performed on external cloud provider servers equipped with AMD Opteron 6274 @ 2.2 GHz processors. The aim of the study was to analyze computing performance depending on selected OpenStack flavor (set of available virtual resources). As in previous experiment, for this purpose the matrix multiplication with matrices of 2048×2048 size were used. Multiplication was parallelized using OpenMP API with the static scheduling clause. The results of the measurements are presented in Table 5. It should be noted that only parallelizable part of the task (multiplication) was measured. The results seem to be astonishing, because the top performance for sequential scenario (one thread) was received for medium flavor. In this case, the characteristics of the environment should be take into account, namely its distribution and virtualization. The experiment shows that the increase of available virtual resources may reduce the performance. It can be caused by communication between distributed resources and as well as by the observed switching between physical resources (threads and fragmented memory).

Table 5
Time of sequential (one thread) and parallel matrix multiplication for each flavor

Flavor	Time [s]				
m1.tiny	231.94	232.80	233.90	234.54	234.90
m1.small	54.79	55.02	55.97	56.39	57.42
m1.medium	48.91	35.74	37.01	37.83	38.92
m1.large	51.93	37.07	19.95	22.04	23.43
Number of threads:	1	2	4	8	16

Table 5 also presents scenarios where the software takes full advantage of available resources – to those achieved the best time results. This study shows how important is the selection of appropriate virtual resources to the task, and how much impact on the performance has distribution and virtualization of resources.

Table 6
Speedup of parallel matrix multiplication depending on the selected flavor

Flavor	Speedup [s]				
m1.tiny	1	0.996	0.992	0.989	0.987
m1.small	1	0.996	0.979	0.972	0.954
m1.medium	1	1.368	1.322	1.293	1.257
m1.large	1	1.401	2.603	2.356	2.216
Ideal speedup	1	2	4	8	16
Number of threads:	1	2	4	8	16

The last Table 6 presents achieved speedup in relation to set of virtual resources. The results were determined on the basis of the Amdahl's law that is a model for the expected speedup and the relationship between parallelized application [22]. The highest speedups were gained in the cases where the number of VCPUs is equal the number of threads. However, comparing the results in Table 5, the speedup often does not result in faster task execution.

6. Summary and Conclusions

Presented article is an extension of previous studies on OpenStack cloud computing platform analysis [1]. The main topic of the paper is a problem with effective use of available physical and virtual resources based on the OpenStack cloud computing platform. Conducted research has shown that OpenStack well manages and utilizes resources, but the platform takes significant part of resources. The process of creating virtual machines, consumes very large amounts of memory. Even in the idle time, the virtual servers are constantly maintained and incessantly keep resources. In this experiment, the cost of platform implementation was greater than the benefits.
The second part of the research was based on the extended infrastructure. The performance analysis of tasks execution depending on selected OpenStack flavor has shown that increase of allocated resources may reduce the performance – especially in the case of distributed environments. Selection of appropriate resources is a very important and non-trivial task. An interesting issue in the future would be a development of a trade-off model as a solution of this problem.

References

[1] D. Grzonka, M. Szczygieł, A. Bernasiewicz, A. Wilczyński, and M. Liszka, "Short analysis of implementation and resource utilization for the OpenStack cloud computing platform", in *Proc. 29th Eur. Conf. Modell. Simul. ECMS 2015*, Albena (Varna), Bulgaria, 2015, pp. 608–614.

[2] E. W. Dijkstra, "The humble programmer", *Commun. ACM*, vol. 15, no. 10, 1972.

[3] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", *Future Gener. Comp. Syst.*, vol. 25, no. 6, pp. 599–616, 2009.

[4] M. Armbrust *et al.*, "A view of cloud computing", *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[5] C. D. Graziano, "A performance analysis of Xen and KVM hypervisors for hosting the Xen Worlds Project", Graduate Theses and Dissertations, Iowa State University, paper 12215, 2011.

[6] K. Scarfone, M. Souppaya, and P. Hoffman, "Guide to Security for Full Virtualization Technologies", Recommendations of the National Institute of Standards and Technology, Special Publication 800-125, USA, 2011.

[7] M. Serafin, *Wirtualizacja w praktyce* (*Virtualization in practice*). Gliwice, Poland: Wydawnictwo Helion, 2012 (in Polish).

[8] G. J. Popek and R. P. Goldberg, "Formal requirements for virtualizable third generation architectures", *Commun. ACM*, vol. 17, no. 7, pp. 412–421, 1974.

[9] M. Tajvidi, R. Ranjan, J. Kołodziej, and L. Wang, "Fuzzy cloud service selection framework", in *Proc. IEEE 3rd Int. Conf. Cloud Networking CloudNet 2014*, Luxembourg, 2014, pp. 443–448.

[10] J. Sahoo, S. Mohapatra, and R. Lath, "Virtualization: A survey on concepts, taxonomy and associated security issues", in *Proc. 2nd Int. Conf. Comp. Netw. Technol. ICCNT 2010*, Bangkok, Thailand, 2010, pp. 222–226.

[11] P. Mell and T. Grance, "The NIST Definition of Cloud Computing", Recommendations of the National Institute of Standards and Technology, Special Publication 800-145, USA, 2011.

[12] Y. Mhedheb, F. Jrad, J. Tao, J. Zhao, J. Kołodziej, and A. Streit, "Load and Thermal-Aware VM Scheduling on the Cloud", in *Algorithms and Architectures for Parallel Processing*, *LNCS*, vol. 8285, pp. 101–114. Springer, 2013.

[13] M. Marks and E. Niewiadomska-Szynkiewicz, "Hybrid CPU/GPU platform for high performance computing", in *Proc. 28th Eur. Conf. Model. Simul. ECMS 2014*, Brescia, Italy, 2014, pp. 508–514.

[14] K. Smelcerz, "Recent developments in mobile cloud scheduling: State-of-the-art, challenges and perspectives", *J. Telecommun. Inform. Technol.*, no. 4, pp. 51–57, 2013.

[15] F. Gens, "Defining Cloud Services and 'Cloud Computing", IDE, 2008 [Online]. Available: http://blogs.idc.com/ie/?p=190

[16] OpenStack Website and Documentation [Online]. Available: http://www.openstack.org/

[17] Google Trends [Online]. Available: http://www.google.com/trends/

[18] O. Sefraoui, M. Aissaoui and M. Eleuldj, "OpenStack: Toward an Open-Source Solution for Cloud Computing", *Int. J. Comp. Appl.*, vol. 55, no. 3, pp. 38–42, 2012.

[19] K. Pepple, *Deploying Openstack*, 2nd ed. O'Reilly Media, USA, 2011.

[20] K. Jackson and C. Bunch, *OpenStack Cloud Computing Cookbook*, 2nd ed. Packt Publishing, UK, 2013.

[21] J. M. Cecilia, J. M. García, and M. Ujaldón, "The GPU on the Matrix-Matrix Multiply: Performance Study and Contributions", in *Parallel Computing: From Multicores and GPU's to Petascale*, B. Chapman *et al.*, Eds. *Advances in Parallel Computing*, vol. 19, pp. 331–340, 2010.

[22] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities", in *Proc. Spring Joint Comp. Conf. AFIPS '67 (Spring)*, Atlantic City, New Jersey, USA, 1967, pp. 483–485.

Daniel Grzonka received his B.Sc. and M.Sc. degrees with distinctions in Computer Science at Cracow University of Technology, Poland, in 2012 and 2013, respectively. Currently, he is Research and Teaching Assistant at Cracow University of Technology and Ph.D. student at Jagiellonian University in cooperation with Polish Academy of Sciences. He is also a member of Polish Information Processing Society and IPC member of several international conferences. The main topics of his research are grid and cloud computing, multi-agent systems and high-performance computing.

E-mail: grzonka.daniel@gmail.com
Institute of Computer Science
Faculty of Physics, Mathematics and Computer Science
Tadeusz Kościuszko Cracow University of Technology
Warszawska st 24
31-155 Cracow, Poland

Paper

The Use of Genetic Algorithms for Searching Parameter Space in Gaussian Process Modeling

Agnieszka Krok

Tadeusz Kościuszko Cracow University of Technology, Cracow, Poland

Abstract— The aim of the paper is to present the possibilities of modeling the experimental data by Gaussian processes. Genetic algorithms are used for finding the Gaussian process parameters. Comparison of data modeling accuracy is made according to neural networks learned by Kalman filtering. Concrete hysteresis loops obtained by the experiment of cyclic loading are considered as the real data time series.

Keywords—Gaussian processes, genetic algorithms.

1. Modeling Time Series

Modeling processes that are time-dependent is made based on many techniques. Methods for time series analyzes may be: correlation analyzes, autoregressive or moving average model, trend estimation and decomposition of time series, principal component analysis, Fast Fourier Transform, continuous wavelet transform. Also many tools are used for time series modeling, e.g., general state space models, unobserved components models, and machine learning methods such as artificial neural networks (ANN), support vector machines, Gaussian processes (GPs) [1], [2]. The main concern of the paper is Gaussian processes modeling. Genetic algorithms (GAs) are well known tools for searching the space of sub-optimal solutions, for example for supporting neural networks learning processes [3]. The investigation was made for the possibility of transferring the techniques known in the field of neural networks for accelerating the search of Gaussian models parameters.

2. Motivation and Related Background

Nowadays Gaussian processes are used for modeling different kind of data and wide variety of time series. In [4] GPs were used for modeling super resolution images, confirming the ability to deal with data read out from the image. In [5] complicated problem of probabilistic prediction of Alzheimer's disease from multimodal image was solved. GPs were also used for modeling time dependent processes inside materials, i.e. wax precipitation model in crude oil systems [6].

Also time depended signals such as speech [7], wind energy systems [8], and facial expressions [8], economical time series, was successfully modeled using GPs.

Modeling stress-strain hysteresis loops involves the representation of changes in time the material properties during tension-compression test. The data for analysis were discrete points taken from the graphical representation of stress-strain relation considered as the time series for artificial unit of time strictly related to the consecutive experiment stages.

The traditional approach to the hysteresis modeling assumes using differential equation models that involve the parameters that are specific to the modeled material as: Jiles–Atherton [9], Ylinen [10], Takács model [11], Prandtl–Ishlinskii model [12]. In most cases, the models are in the form of piecewise functions different for the particular branches of the hysteresis [10], [13].

In addition, the soft methods were considered: neural networks in the form of multi-layer perceptions, learned by the back propagation algorithm for supervised training [14], [15], or the Levenberg-Marquardt algorithm [16]–[18]. For considered experiments, successful modeling using supervised artificial neural networks learned by Kalman filter was already made [19].

MacKay in [20] suggested that Gaussian processes might be a replacement tool for supervised neural networks.

During numerical experiments, the influence of parameters of GP was examined. It was stated that the parameters of GP models are much more significant for the proper time series modeling then the parameters of ANN. The number of neurons, the initial values of ANN, values of the parameters that govern the learning process does not influence the numerical results much. The improper parameters of GP lead to incorrect modeling. Using GA is the well-known technique for supporting the process of ANN learning process [2], [21]–[24].

The aim of this survey was to confirm or subvert the thesis of the possibility of using GP instead of ANN for modeling hysteresis loops of stress-strain relation for concrete specimens and to find methodologies for effective selection of GP parameters. The tool, selected for this purpose was GA, and each individual in the population represents a possible solution of the Gaussian model, similarly to [18], but in this paper scatter, crossover operator was used.

3. Gaussian Process Model

Lets consider the stochastic process Y, generated by the set of fixed basis functions with random weights [3]:

$$Y(x) = \sum_{j=1}^{M} W_j \phi_j(x)\,, \tag{1}$$

where x is the input vector indexing random variables [3]. If weights vector has normal distribution with zero mean, and particular standard deviation $W_j \sim N(0,\Sigma)$ then $E_W[Y(x)] = 0$ and $E_W[Y(x)Y'(x)] = \phi^T(x)\Sigma\phi(x)$. The training data set consist of pairs (x_i,t_i), where t_i, $i = 1, 2, \ldots, N$ is the sample from the random variable $T(x_i)$.
To make the prediction in the new input x_* it is necessary to compute conditional distribution $p(T(x_*)|T(x_1),\ldots,T(x_N))$. Let C denote covariance matrix of the training data, $t = [T(x_1), \ldots, T(x_N)]$, k denote the covariances vector between the training data $T(x_1),\ldots,T(x_N)$ and $T(x_*)$, V denote the prior variance of $T(x_*)$ that is $Cov(T(x_*),T(x_*))$. Then [3]:

$$E(T(x_*)|T(x_1), \ldots, T(x_N)) = k^T C^{-1} t\,, \tag{2}$$

$$D^2(T(x_*)|T(x_1), \ldots, T(x_N)) = V - k^T C^{-1} k\,. \tag{3}$$

Two covariance functions were considered:

- Squared exponential:

$$C(x^i,x^j) = v_0 \exp\Big(\sum_{l=1}^{d} a_l (x_l^i - x_l^j)^2\Big) + b\,, \tag{4}$$

where $x^i,x^j \in R^d$, $x^i = [x_1^i, \ldots, x_d^i]$. Then target covariance is given by:

$$C(x^i,x^j) + \sigma_v^2 \delta_{i,j}\,, \tag{5}$$

where

$$\sigma_v^2 \tag{6}$$

is the variance for the $p(T(x_*)|T(x_1), \ldots, T(x_N))$ $\delta_{i,j} = 1$ for $i = j$, $\delta_{i,j} = 0$ for i not equal j. The parameters of the model ale considered in the log space:

$$\theta = (\ln(v_0),\ln(b),\ln(a_1), \ldots, \ln(a_d),\ln(\sigma_v^2),\ln(v))\,. \tag{7}$$

- Rational quadratic:

$$C(x^i,x^j) = v_0\Big(1 + \sum_{l=1}^{d} a_l (x_l^i - x_l^j)^2\Big)^{-\tau} + b\,, \tag{8}$$

$$\theta = \big(\ln(v_0), \ln(b), \ln(a_1), \ldots, \ln(a_d), \ln(\sigma_v^2)\big)\,. \tag{9}$$

Gaussian process structure may be viewed in the ANN form:

$$\theta = \big(\ln(v_0), \ln(b), \mathbf{w}, \ln(v)\big)\,, \tag{10}$$

where b is the network bias, σ_v^2 is the noise incorporated into the network, $\mathbf{w}$ is the vector of weights, v_0 is the scaling parameter, x^i,x^j are input vectors [25]. That is why the effectiveness of GP models was compared to ANN models.

3.1. Learning Hyper Parameters

The non-linear optimizer is used to find the maximum likelihood values of the parameters θ_i. It is done by equaling to zero partial derivatives of log likelihood and using one of the standard optimization algorithms. The Scaled Conjugate Gradient optimization (SCG) was used with the standard Matlab implementation options [26]. The analysis showed that the key role in the efficiency of the procedure plays the number of SCG steps marked with k parameter.

The starting parameter values for the algorithm are randomly chosen from the $N(m,\sigma)$ distribution, where m and σ are the hyper parameters for this model. Additional noise term is added to the noise σ_v^2 in Eqs. (5) and (6) to make sure that noise variance never collapse to zero.

3.2. Making Predictions

During this stage the parameters for the predicted Gaussian distribution are computed according to the Eqs. (11) and (12) [3]:

$$E(T(x_*)|T(x_1), \ldots, T(x_N)) = k^T C^{-1} t\,, \tag{11}$$

$$D^2(T(x_*)|T(x_1), \ldots, T(x_N)) = V - C^T C^{-1} k\,. \tag{12}$$

3.3. Parameters of the Presented Procedure

The starting parameter values for the algorithm m and $sigma$ and the SCG algorithm steps k number have to be chosen before each algorithm run. In the paper, the research for this parameters is made. Software for Flexible Bayesian Modeling and Markov Chain Sampling implementation [27] with own author's modification was used to implement the theoretical model. The weight number in each Gaussian model corresponds fully to the feed forward ANNs of the architecture that were considered in [19]. Each ANN weight has its equivalent in GP model.
To summarize the data effectiveness simulation set using one single value, MSE error was introduced:

$$MSEV = \frac{1}{V}\sum_{l=1}^{V}(y_l - \overline{y}_l)^2\,, \tag{13}$$

where: $V = L$, T is the number of learning and testing patterns, respectively; y_l – the target $\overline{y}_l$ is computed output mean value for l-th pattern scaled to the interval $[0\ldots1]$, see Eq. (1).

3.3.1. Calibrating the Parameters of the Numerical Models

As far as squared exponential covariance function is concerned the parameters of models are:

$$\theta = (\ln(v_0),\ln(b),\ln(a_1), \ldots, \ln(a_d),\ln(\sigma_v^2),\ln(v))\,. \tag{14}$$

For the rational quadratic covariance:

$$\theta = (\ln(v_0),\ln(b),\ln(a_1), \ldots, \ln(a_d),\ln(\sigma_v^2))\,. \tag{15}$$

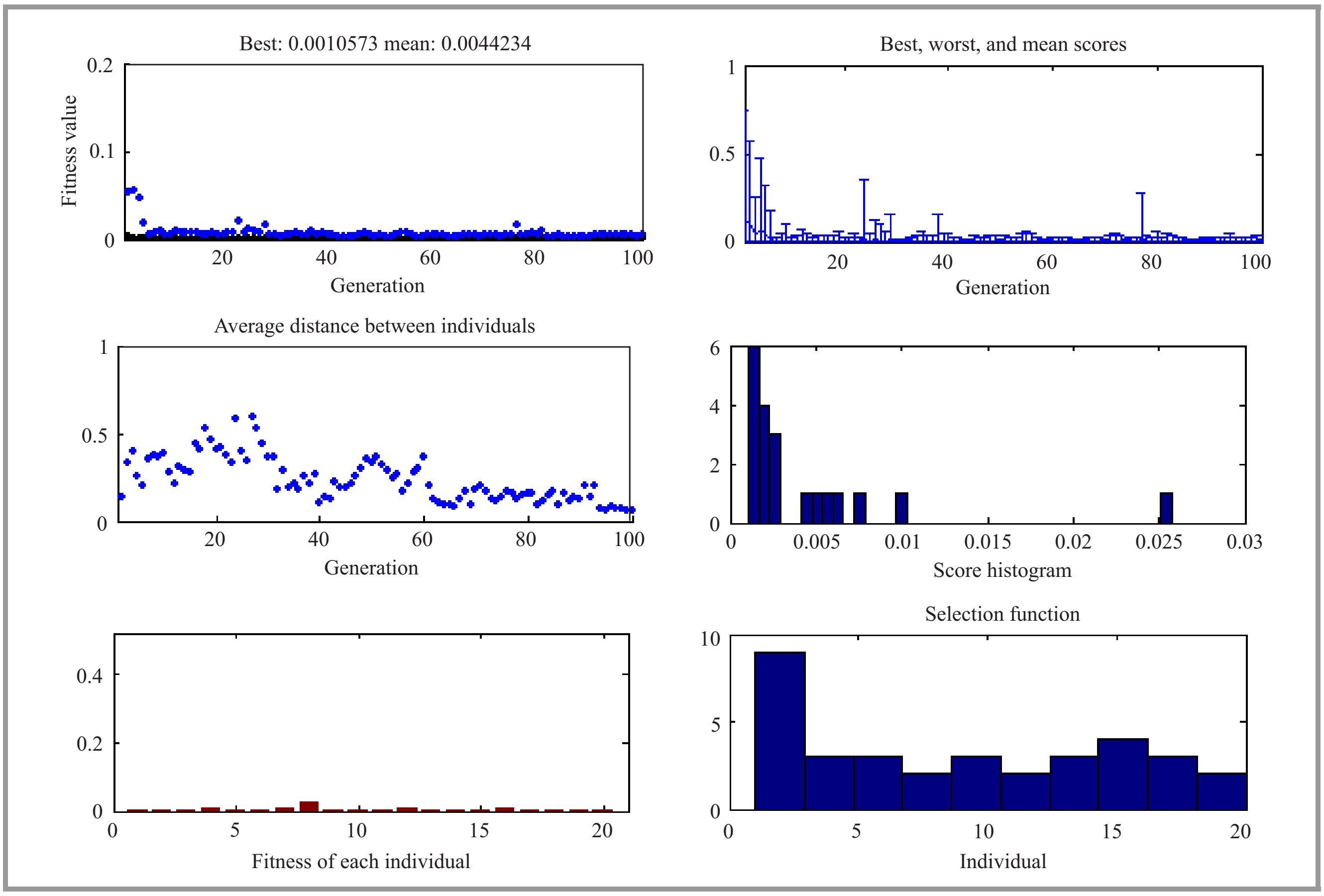


Fig. 1. Sample results of genetic algorithm proceeding.

To initialize the process of finding the optimal values of the parameters mean and variance of their prior Gaussian densities have to be set:

$$\theta(i) \sim N(m,\sigma)\,. \tag{16}$$

3.3.2. The Genetic Algorithm for the GP Models Parameter Finding

There was stated that there is the significant influence of the parameters on the value of *MSE* errors as far as parameters (m,σ) and the length of SCG process k is concerned. This was the main reason for applying the additional procedure of calibrating both models. To reach that GA was chosen, and the fitness function to minimize was set in the form:

$$MSE(\overline{k},m,\sigma)\,, \tag{17}$$

where $\overline{k}$ is k scaled to the selected range by the linear scaling parameter ω: $\overline{k}=\omega k$.

In this research, the Matlab Genetic Algorithm Tool was used [28]. Population type, which specifies the type of the input to the fitness function was the vector $(\overline{k},m,\sigma)\in R^3$. Population size, which specifies how many individuals there are in each generation, was assumed 20 individuals. The uniform creation function creates the initial population from the given interval of initial range. The scaling function, which converts raw fitness scores returned by the fitness function to values in a range that is suitable for the selection function was used. Next, the rank scaling function was applied. Rank scales the raw scores and is based on the rank of each individual, rather than its score. The rank of an individual is its position in the sorted scores. The rank of the fittest individual is 1, the next fittest is 2 and so on. Rank fitness scaling removes the effect of the spread of the raw scores [29].

The selection function chooses parents for the next generation based on their scaled values from the fitness scaling function.

Stochastic uniform selection function was then applied. It lays out a line in which each parent corresponds to a section of the line of length proportional to its expectation. The algorithm moves along the line in equal size steps, one step for each parent. At each step, the algorithm allocates a parent from the section it lands on. The first step is a uniform random number less than the step size. Reproduction options determine how the genetic algorithm creates children at each new generation. Elite count specifies the number of individuals that are guaranteed to survive to the next generation. Set elite count to be a positive integer less than or equal to population size. Here this number was set to 2. Then crossover fraction specifies the fraction of the next generation, other than elite individuals, that are produced by crossover. The remaining individuals, other than elite individuals, in the next generation are produced by mutation. Set crossover fraction is a fraction between 0 and 1, and was set 0.8.

Mutation functions make small random changes in the individuals in the population, which provide genetic diversity and enable the GA to search a broader space. In presented research a Gaussian mutation functions was used. It adds a random number to each vector entry of an individual. This random number is taken from a Gaussian distribution centered on zero. The variance of this distribution can be controlled with two parameters. The scale parameter determines the variance at the first generation, and the shrink parameter controls how variance shrinks as generations go by. If the shrink is 0, the variance is constant. If the shrink is 1, the variance lowers to 0 linearly as the last generation is reached. Scale and shrink parameters was set at 1.

Crossover combines two individuals, or parents, to form a new individual, or child, for the next generation. Scattered crossover was used. It creates a random binary vector. It then selects the genes where the vector is a 1 from the first parent, and the genes where the vector is a 0 from the second parent, and combines the genes to form the child [29].

Stopping criteria determine what causes the algorithm to terminate:

- generations parameter specifies the maximum number of iterations the genetic algorithm performs (the value 100 was set),
- time limit specifies the maximum time (in seconds) the genetic algorithm runs before stopping (inf. was set),
- fitness limit – if the best fitness value is less than or equal to the value of fitness limit, the algorithm stops (inf. was set),
- stall generations – if there is no improvement in the best fitness value for the number of generations specified by stall generations, the algorithm stops (50 was set,
- stall time limit – if there is no improvement in the best fitness value for an interval of time in seconds specified by stall time limit, the algorithm stops (10^5 was set).

Best fitness plots the best function value in each generation vs. iteration number. Score diversity plots a histogram of the scores at each generation. Best individual plots the vector entries of the individual with the best fitness function value in each generation. Scores plots the scores of the individuals at each generation. Distance plots the average distance between individuals at each generation. Range plots the minimum, maximum, and mean fitness function values in each generation, see Fig. 1.

4. Simulation

Tests were made according to previously, investigated data sets [19]. Many experimental data sets coming from different kind of loading-unloading concrete and steel specimens were considered. These time series reflects the behavior of the material over time.

In this Section the time series coming from 12 concrete cylindric samples 3×6 inches size, that were compressed according to the following cyclic loading plan are presented [26]:

- monotonic increasing of the load to the maximal value,
- decreasing of the load to the 0,
- monotonic increasing of the load to the maximal value, and the stress-strain relation in time was consider as the modeled time series.

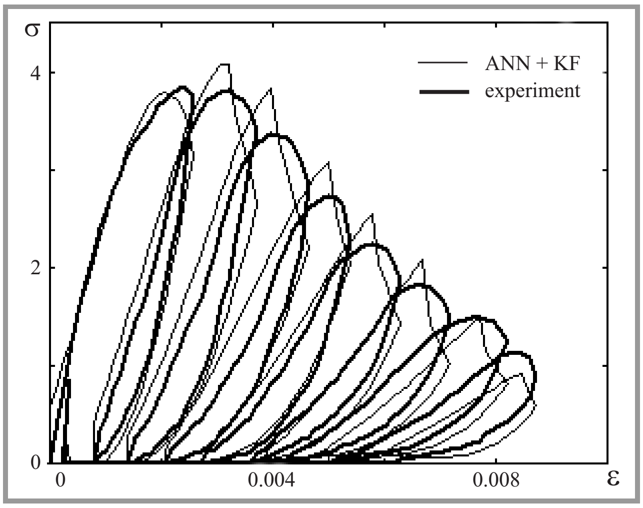


Fig. 2. Simulation and prediction of hysteresis loops from ANN models, learned by Kalman filtering (vertical axis contains values of stress, horizontal axis shows values of strain).

Data for calibrating and testing the models were discrete points from stress–strain σ-ε relation, see Fig. 2. As a preprocessing scaling to the internal range $[0.1 \ldots 0.9]$ was applied. This operation was done to correspond to the learning and testing for ANN, considered in [19] for the same experiment. Given data sets were divided into calibrating and testing set, corresponding to the learning set and testing set in [19]. The used testing set consists of points from 3 last of 8 hysteresis loops. For properly calibrated models it results in simulating behavior of the material in the first experiment part and prediction of the material behavior in the final part of the experiment basis on the material behavior in the first part of the experiment. The number of weight in each Gaussian model corresponds feed forward ANN of the architecture, which were considered in [19]. Each ANN weight has its equivalent in Gaussian process model. The output Gaussian process models were the stress σ value, predicted in the current step of computation.

The calibrating set for the experiment consist of first six hysteresis loops, which gave $L = 273$ data points. The testing set were selected as the following three loops, what resulted in $T = 132$ data points. From among many input vectors the most effective

$$\mathbf{x}(j) = [\sigma(j-1), j/(273+132), marker_2],$$

was found in [19], where i is the number of current pattern, $j = 1, \ldots, 405$. The *marker* is the parameter numbering patterns for network learning and testing inside each loop separately independently from another loops. Parameters $marker_1$ and $marker_2$ were adopted. Inside i-th hysteresis loop the following values of these parameters were the most numerically effective:

$$marker_{1,i} = \Big[1/M_i, 2/M_i, \ldots, M_i/M_i, (M_i-1)/M_i, (M_i-2)/M_i, \ldots, (M_i - N_i)/M_i\Big] \tag{18}$$

where M_i is the number of experimental points for which the material is loaded, N_i is the number of experimental points which material is unloaded inside i-th hysteresis loop. Parameter $marker_{2,i}$ is based on $marker_{1,i}$ scaled to the interval $[0.1 \ldots 0.9]$.

There were the sets of parameters when both models are simulating and predicting the data set correctly. The combination of algorithm parameters was for example $m = 0$, $\sigma = 1$, $k = 10$.

Results of simulation given data set is presented for both considered covariance models initialized by the same set of parameters $m = 5, \sigma = 1, k = 40$. The numerical accuracy of both models differs, see Fig. 3. The predicted mean values of the experimental data for the rational quadratic covariance function are shown on the Fig. 4.

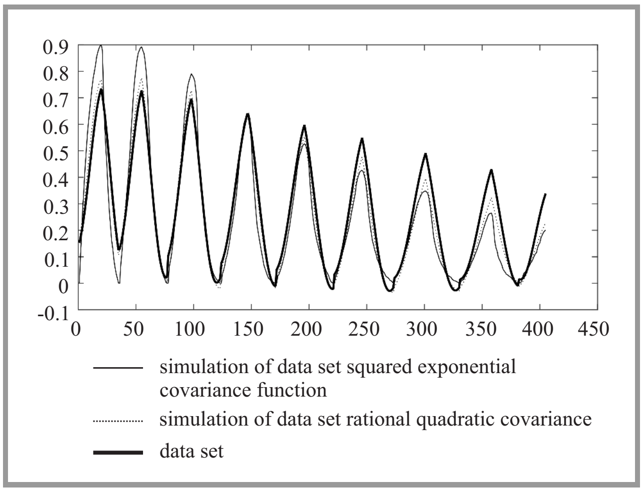


Fig. 3. Simulation and prediction of hysteresis loops for GP models (vertical axis presents values of stress, horizontal axis is number of data point j).

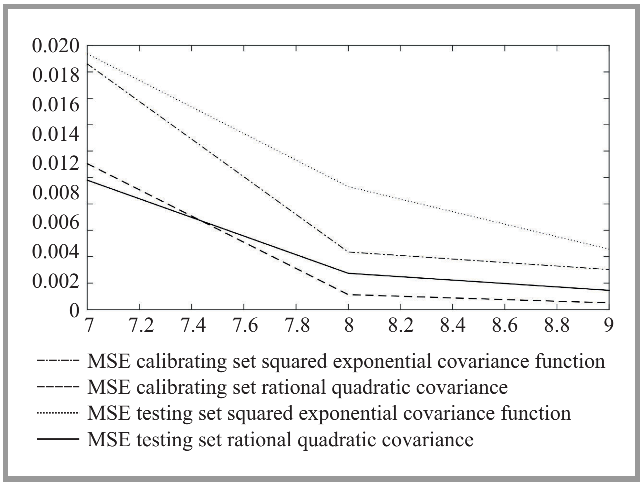


Fig. 4. MSE errors for calibration and prediction of hysteresis loops for GP models.

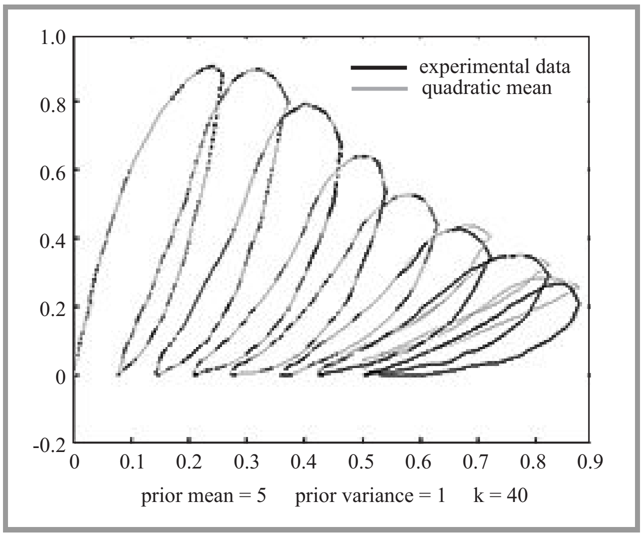


Fig. 5. Simulation and prediction of hysteresis loops for the rational quadratic covariance function GP models (vertical axis shows values of stress, horizontal axis presents values of strain).

5. Results Discussion

The analysis of the presented results suggest that the k parameter setting plays the key role in the presented numerical method. For the chosen values of k both covariance function can be used to simulate presented data correctly. Setting the SCG algorithm steps maximal number too small or two large can make one of the covariance function model ineffective. Simulation made shows that the values of k in the range $[10 \ldots 45]$. For the squared exponential covariance function much longer SCG operating phase is necessary.

For the rational quadratic covariance function $k = 10$ is enough to receive correct results, see Figs. 4 and 5. For the squared exponential covariance model processing SCG algorithm for too less steps result in averaging the obtained results. Neither learning nor testing set is simulated correctly – obtained model do not distinguish the process taking place in time as far experiment is taking place. Then,

setting k larger gives optimal behavior of the model. Processing SCG algorithm for too much steps result in averaging the obtained results again. The rational quadratic covariance function seems to be less sensitive to the k parameter setting. The effect of averaging results is taking place for considered range of $k > 70$ steps.
The comparison with earlier results when ANN learned by Kalman Filtering demonstrated superiority of Gaussian processes, as far as the quality of modeling is concerned, see Fig. 2. The model of first 7 hysteresis loops is more accurate. Last loop is model with less precision but the tendency still may be found.

6. Final Remarks

Gaussian processes were found as a very accurate tool for simulation and prediction of concrete hysteresis loops. The use of genetic algorithm as a method for automatic setting parameters of GP occurred to shorten the parameters setting process much. The simulation and prediction of the stress-strain relation is much precise then made by neural networks models.

References

[1] J. D. Hamilton *Time Series Analysis*. Princeton University Press, 1994.

[2] S. O. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2008.

[3] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[4] Y. Kwon, K. Kim, K. J. Tompkin, J. H. Kim, and C. Theobalt, "Efficient learning of image super-resolution and compression artifact removal with semi-local Gaussian processes", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. PP, no. 99, 2014.

[5] J. M. Young, "Probabilistic prediction of Alzheimer's disease from multimodal image data with Gaussian processes", Ph.D. thesis, University College London Press, 2015.

[6] A. K. Manshada and H. Rostamib, "Prediction of Wax precipitation in crude oil systems using Gaussian processes", *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 37, no. 1, pp. 84–91, 2015.

[7] J. Nielsen and J. Larsen, "Perception-based personalization of hearing aids using Gaussian processes and active learning", *IEEE/ACM Trans. Speech, and Lang. Process.*, vol. 23, no. 1, pp. 162–173, 2015.

[8] J. Diwale, S. Sanjay, L. Ioannis, and J. Colin, "Optimization of an airborne wind energy system using constrained Gaussian processes with transient measurements", in *Proc. Indian Control Conf. ICC 2015*, Chennai (Madras), India, 2015.

[9] M. Hamimid, S. M. Mimoune, and M. Feliachi, "Minor hysteresis loops model based on exponential parameters scaling of the modified Jiles–Atherton model", *Physica B: Condensed Matter*, vol. 407, no. 13, pp. 2438–2441, 2012.

[10] A. Ganczarski and L. Barwacz, "Low cycle fatigue based on unilateral damage evolution", *Int. J. of Damage Mechanics*, vol. 16, no. 2, pp. 159–177, 2007.

[11] K. Chwastek, "Modelling hysteresis loops in thick steel sheet with the dynamic Takács model", *Physica B: Condensed Matter*, vol. 407, no. 17, pp. 3632–3634, 2012.

[12] M. Al Janaideh, "A time-dependent stop operator for modeling a class of singular hysteresis loops in a piezoceramic actuator", *Physica B: Condensed Matter*, vol. 413, pp. 100–104, 2013.

[13] A. P. S. Baghel, A. Gupta, K. Chwastek, and S. V. Kulkarni, "Comprehensive modelling of dynamic hysteresis loops in the rolling and transverse directions for transformer laminations", *Physica B: Condensed Matter*, vol. 462, pp. 86–92, 2015.

[14] I. Kucuk, "Prediction of hysteresis loop in magnetic cores using neural network and genetic algorithm", *J. Magnetism and Magnetic Materials*, vol. 305, no. 2, pp. 423–427, 2006.

[15] R. Dong, Y. Tan, H. Chen, and Y. Xie, "A neural networks based model for rate-dependent hysteresis for piezoceramic actuators", *Sensors and Actuators A: Physical*, vol. 143, no. 2, pp. 370–376, 2008.

[16] A. Nouicer, E. Nouicer, and F. Mouloudc, "A neural network for incorporating the thermal effect on the magnetic hysteresis of the 3F3 material using the Jiles–Atherton model", *J. Magnetism and Magnetic Materials*, vol. 373, pp. 240–243, 2015.

[17] V. Wolfs and P. Willems, "Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural networks", *Environ. Model. & Softw.*, vol. 55, pp. 107–119, 2014.

[18] X. Zhang, Y. Tan, and M. Su, "Modeling of hysteresis in piezoelectric actuators using neural networks", *Mechan. Syst. and Sig. Process.*, vol. 23, no. 8, pp. 2699–2711, 2009.

[19] A. Krok, "Analiza wybranych zagadnień mechaniki konstrukcji i materiałów za pomocą SSN i filtrów Kalmana (Analysis of mechanics of structures and material problems applying artifcial neural networks learnt by means of Kalman filtering)", Ph.D. thesis, Tadeusz Kościuszko Cracow University of Technology, 2007 (in Polish).

[20] D. J. C. MacKay, "Gaussian processes – a replacement for supervised neural networks?", Lecture notes for a tutorial at Neural Information Processing Systems (NIPS) 1997, Cambridge University, 1997.

[21] D. Whitley, T. Starkweather, and C. Bogart, "Genetic algorithms and neural networks: optimizing connections and connectivity", *Parallel Comput.*, vol. 14, no. 3, pp. 347–361, 1990.

[22] D. Pham and D. Karaboga, *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks*. Springer, 2011.

[23] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[24] Ms. Dharmistha and D. Vishwakarma, "Genetic algorithm based weights optimization of artificial neural network", *Int. J. Adv. Res. Elec., Electron. and Instrumen. Engin.*, vol. 1, no. 3, 2012.

[25] R. M. Neal, "Regression and classification using Gaussian process priors", in *Bayesian Statistics 6*, J. M. Bernardo *et al.*, Eds. Oxford University Press, 1998, pp. 475–501.

[26] B. P. Sinha, K. H. Gerstle, and L. G. Tulin, "Stress-strain relations for concrete under cyclic loading", *J. of the American Concrete Institute*, no. 61-12, 1964.

[27] R. Neal, Software for Flexible Bayesian Modeling and Markov Chain Sampling [Online]. Available: http://http://www.cs.toronto.edu/~radford/fbm.software.html

[28] Optimization Toolbox User's Guide – MathWorks, The MathWorks Inc. [Online]. Available: http://uk.mathworks.com/help/pdf-doc/optim/optim_tb.pdf

[29] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*. Springer, 2008.

[30] Y. S. Othmana *et al.*, "Frequency based hysteresis compensation for piezoelectric tube scanner using Artificial Neural Networks", *Procedia Engin.*, vol. 41, pp. 757–763, 2012.

[31] F. Pernkopf and D. Bouchaffra, "Genetic-Based EM Algorithm for Learning Gaussian Mixture Models", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 8, pp. 1344–1348, 2005.

Agnieszka Krok – for biography, see this issue, p. 51.

Paper

A k-Nearest Neighbors Method for Classifying User Sessions in E-Commerce Scenario

Grażyna Suchacka[1], Magdalena Skolimowska-Kulig[1], and Aneta Potempa[2]

[1] *Institute of Mathematics and Informatics, Opole University, Opole, Poland*
[2] *GEA Technika Cieplna Sp. z o.o., Opole, Poland*

Abstract—This paper addresses the problem of classification of user sessions in an online store into two classes: buying sessions (during which a purchase confirmation occurs) and browsing sessions. As interactions connected with a purchase confirmation are typically completed at the end of user sessions, some information describing active sessions may be observed and used to assess the probability of making a purchase. The authors formulate the problem of predicting buying sessions in a Web store as a supervised classification problem where there are two target classes, connected with the fact of finalizing a purchase transaction in session or not, and a feature vector containing some variables describing user sessions. The presented approach uses the *k*-Nearest Neighbors (*k*-NN) classification. Based on historical data obtained from online bookstore log files a *k*-NN classifier was built and its efficiency was verified for different neighborhood sizes. A 11-NN classifier was the most effective both in terms of buying session predictions and overall predictions, achieving sensitivity of 87.5% and accuracy of 99.85%.

Keywords—data mining, e-commerce, k-Nearest Neighbors, k-NN, log file analysis, online store, R-project, supervised classification, Web mining, Web store, Web traffic, Web usage mining.

1. Introduction

Electronic commerce has gained tremendous popularity in recent years, especially in the area of Business-to-Consumer (B2C) trade, completed through Web stores. A still increasing number of online customers and their heterogeneous needs motivate online retailers to search for ways of predicting e-customer behavior and providing them with customized service. Especially valuable would be the ability to identify potential or future buyers as it could make it possible to enforce different personalized service strategies, e.g., by offering special discounts to encourage undecided visitors to buy or by presenting them with complementary products to increase online sales.

The problem of analyzing Web user behavior and anticipating their needs has been intensively explored. Various approaches have been proposed in the literature to discover hidden patterns in user navigation paths and online purchase transactions [1]–[4], as well as to identify users with high purchasing intentions and predict online sales [5]–[9]. This study fits into this research area by addressing the problem of predicting online purchases. Authors propose an approach to classify user sessions in online store into buying sessions and browsing sessions based on some session features using a k-Nearest Neighbor (k-NN) technique.

The rest of the paper is organized as follows. Section 2 provides a brief background on k-NN classification. Section 3 discusses related work on application of k-NN classifiers in e-commerce environment. Section 4 presents a methodology underlying the presented approach and Section 5 discusses the performance evaluation of the approach using data from a real online store. The final Section 6 provides some concluding remarks and directions for future work.

2. K-Nearest Neighbors Classifier

A k-Nearest Neighbors algorithm is a supervised learning technique often used in pattern recognition for classification, although it can also be used for estimation and prediction [10]. A k-NN classifier is memory-based (instance-based) and requires no model to be fitted, it is also conceptually simple. No explicit training procedure is required for the set of observations apart from collecting vectors of features with labels of classes they belong to. All intensive computations are performed at classification, which involves two steps for each test observation: finding k nearest neighbors among the observations in a training set and performing a majority voting among the retrieved k neighbors to assign the most frequent class label among them [11].

Let $\{(x_i, y_i),\ i = 1, 2, \dots, n\}$ be the training set of observations, where x_i is a vector of features and y_i is a class label of x_i. We assume that every x_i is in some multidimensional feature space with a metric ρ and $y_i \in \{1, 2, \dots, l\}$, where l is the number of considered classes, $i = 1, 2, \dots, n$. The task is to assign an unlabeled vector x to a proper target class from $\{1, 2, \dots, l\}$. The simplest version of k-NN algorithm is the nearest neighbor rule (1-NN) which assigns x to the class of its closest neighbor. It means that if x_j, for some $j \in \{1,2,\dots,n\}$, is the nearest to x in the sense of distance ρ, i.e.:

$$x_j = \arg \min_{\{x_i, 1 \le i \le n\}} \rho(x, x_i), \tag{1}$$

then the rule labels x the number y_i. The k-NN method for $k > 1$ is a natural extension of the foregoing rule. It classifies x by assigning it the label which is most frequently represented among the k nearest training points x_i, where k is a user-defined constant. Thus, a decision is made by examining class labels of the k nearest neighbors and taking a majority vote.

The 1-NN rule usually classifies with an error rate greater than the minimum possible one – the Bayes rate. However, when the number of observations n tends to infinity, the error rate is not worse than twice the Bayes rate. For more details and discussion on the k-NN classification see e.g. [12] and [13].

The most common similarity measure (distance measure) between observations is the Euclidean metric. The Euclidean distance between two J-dimensional vectors a and b is expressed by the formula:

$$d_{a,b} = \sqrt{\sum_{j=1}^{J} (a_j - b_j)^2}\,. \tag{2}$$

Additional metrics which can be used as distance measures are, e.g., standardized Euclidean, weighted Euclidean, Mahalanobis, Minkovsky, and Chebyshev distances, or in the case of a discrete type of data – a Hamming distance.

Huge advantages of the k-NN algorithm, especially important in practical applications, are its great scalability, linear computational complexity, robustness to data sparseness and skewed target class distribution, and interpretable target class predictions [11]. The k-NN method has been successfully applied in a large number of classification problems, like handwriting detection, gene expression, EKG patterns or satellite image scene detection. It has also been applied to some problems related to WWW and e-commerce, which are briefly reviewed in the Section 3.

3. Related Work

The most popular application area of the k-NN method in the e-commerce environment has been recommender systems. Generally, product recommendation methods in online stores may be divided into content-based recommendation and collaborative filtering (CF) recommendation. Content-based recommendation methods are based on the similarity of products (item profiles are built). On the other hand, CF recommender systems try to assess the utility of items for a given customer based on the items connected with other, similar users (user profiles are built) [14].

A traditional CF method is similarity-based, i.e., products to be recommended to a particular user are selected based on product preferences of a group of their nearest neighbors (with preferences similar to those of the target user) [15]. A comprehensive survey of CF techniques, divided into memory-based, model-based, and hybrid CF algorithms have been discussed in [16].

A key task in CF recommender systems is computing the all-to-all similarity between customers in order to form the neighborhood for a particular customer, i.e., a group of the most similar customers in terms of preferences, tastes, and purchase patterns. The proximity between two e-customers in conventional CF methods has been the most frequently measured using the Pearson correlation coefficient [15], [17]–[19] or the cosine metric [11], [17]. Many extensions to the standard correlation-based and cosine-based techniques have been proposed, including default voting, inverse user frequency, case amplification, and weighted-majority prediction [14].

The most popular method for neighborhood formation is a center-based scheme, that forms a neighborhood for a particular customer by simply selecting the k nearest other customers. Other methods may be applied as well, e.g., in the case of very sparse data sets an aggregate neighborhood scheme may be applied, which allows the nearest neighbors to affect the process of a neighborhood formation for a particular customer [17].

Two main techniques are generally used to determine the neighborhood size [18]. The first one is a *correlation-thresholding* technique, in which the nearest neighbors are customers with absolute correlates greater than a given threshold. The second technique is a *best-k-neighbors* technique, in which the best k correlates are selected for the neighbors. In practice, the value of k is often determined experimentally.

Many improvements have been proposed for CF recommendation in e-commerce. Some studies have used the fact that a low-dimensional space is less sparse than the corresponding high-dimensional space and thus, they have applied dimensionality reduction techniques, e.g., Singular Value Decomposition (SVD) to generate a low-dimensional feature space in which the neighborhood has then been formed [17]. In [18] a CF-based recommendation methodology based on Web usage mining and product taxonomy was proposed to address the sparseness and scalability problems of CF recommender systems. The product taxonomy was used to improve the performance of searching for nearest neighbors through dimensionality reduction of the rating database. Jiang *et al.* [19] proposed a clustering-based k-NN approach which combines the k-NN and the iterative clustering algorithm. The iterative clustering approach allows them to solve the data sparseness problem by fully exploiting the voting information first. Then, a cluster-based k-NN is applied to improve the performance of CF.

Other application areas of the k-NN method in the e-commerce environment besides recommender systems have been based on the Web contents analysis. In [20] a method for e-commerce website trust assessment based on the analysis of textual contents and page layout was proposed. Two approaches were applied to construct a feature vector for each analyzed document: in a *baseline* approach all words extracted from the document were used and in a *EC-word* approach the extracted words were mapped into the meaningful groups of e-commerce terminology. Then, three different methods were applied to classify the text into a proper trust level class: k-NN, Naïve Bayes, and Support Vector Machine (SVM) based on Sequential Minimal Optimization (SMO) training. The k-NN method was used to find k nearest neighbors of the analyzed document in the training set of documents using the Euclidean distance as a similarity measure. The categories of the nearest neighbors were used to determine the resultant trust level class. Results of k-NN classification were good, although a bit worse than SVM results.

In [11] the k-NN method was used in a hierarchical approach to the problem of large-scale text-based item categorization on an e-commerce site (a major online auction site). Categorization of items in very large datasets was formulated as a supervised classification problem where the categories are the target classes and words composing some textual description of the items are the features. The classification problem was decomposed into a coarse level task and a fine level task. The coarse level classifier was responsible for classifying items into so-called latent groups. A simple and scalable k-NN classifier for the reduced feature space was applied at this level. The similarity between items was measured using the cosine metric and the item was assigned to the class with the majority votes of its k nearest neighbors. On the other hand, the fine level classifier was responsible for assigning items to the right class (category) in a given latent group. To this end, an SVM classifier was applied for all latent groups for the original latent group feature space.

To the best of authors knowledge, the problem of k-NN classification of e-customer sessions in terms of a purchase confirmation has not been investigated so far. This work is a continuation of authors previous studies on the application of various data mining techniques (e.g., association rules [8] and SVM [9]) to predict online purchases.

4. Research Methodology

4.1. Analysis of the Website Contents and Distinguishing Session States

A typical online store is implemented as a website hosted on a Web server on the Internet. The website consists of many pages, each of which is related to some function (e.g., reading general information about the store, searching for a product, adding the product to a shopping cart, etc.). At any time multiple users may interact with the site by opening different pages and performing various functions, i.e., there may be many active user sessions on the server.

The presented research was based on data obtained from an online bookstore (the name of the store is not given in the paper due to a non-disclosure agreement). The analyzed website contained not only typical Web store pages but also related pages with multimedia entertainment contents, like movies, quizzes, games, etc.

The website contents was analyzed in detail and as a result each page was assigned to one of the following 15 session states: *Home* – the home page of the bookstore, *Information* – pages containing general information about the store, *Entertainment* – pages with entertainment contents, *Shipping* – pages with information on shipping cost and terms, visited before starting the checkout process, *Shipping_checkout* – pages with shipping information visited during the checkout process, *Browse* – browsing interactions, *Search* – interactions connected with searching for products according to some keywords, *Details* – pages with product information, *Add* – adding a product to the shopping cart, *Register_success* – successful user registration, i.e., creating a new user account, *Register_try* – pages connected with the registration process, other than the successful user registration, *Login_success* – successful user logging into the site, *Log_off* – user logout, *Checkout_try* – pages connected with the checkout process, other than a purchase confirmation, *Checkout_success* – purchase confirmation operation, i.e., a successful finalization of a purchase transaction.

4.2. Source Data

A single user session in an online store is represented on the Web server as a sequence of HTTP requests sent to the server by a client (i.e., an Internet browser). All HTTP requests coming to the server are registered in a server access log. In the case of a popular NCSA combined log format the following data is written in a log file for each request: client IP address, identifier and username, for authentication, date and time stamp, HTTP method, version of HTTP protocol, Uniform Resource Identifier (URI) of the requested server resource, HTTP status code, the volume of data transferred for the HTTP request in response, a referrer that linked the user to the store website, and a user agent string with some information on the client browser.

In this research the data recorded from 1 to 30 April 2014 was used. All user session data set contained 39,000 observations. This set was divided into two subsets: a training set and a test set. The training set was created by drawing 26,000 observations (among which 146 observations contained a purchase confirmation operation). The remaining 13,000 observations (including 72 observations with a purchase confirmation operation) were in the test set.

4.3. Reconstruction and Description of User Sessions

Using a dedicated C++ program data was read from log files, merged, pre-processed and cleaned. Based on IP addresses and user agent strings of HTTP requests, request streams for individual users were distinguished. Assuming that intervals between each two subsequent requests of the same user in session do not exceed a 30-minute threshold [21]–[24] user sessions were then identified.

Each user session was described with 23 session features. The first group of features includes 15 elements connected with visits to the session states occurred in session. A variable $V_{Checkout\_success}$ is a boolean variable (equal to one if a purchase transaction was successfully finalized in session and zero otherwise). Other variables: V_{Home}, $V_{Information}$, $V_{Entertainment}$, $V_{Shipping}$, $V_{Shipping_checkout}$,

V_{Browse}, V_{Search}, $V_{Details}$, V_{Add}, $V_{Register\_success}$, $V_{Register\_try}$, $V_{Login\_success}$, V_{Log_off}, $V_{Checkout\_try}$, are connected with numbers of visits to the corresponding session states occurred in session.
The second group of features includes the following 6 session characteristics:

- $V_{Requests}$, the number of HTTP requests in session,
- $V_{Transfer}$, the volume of data downloaded in session (in kilobytes),
- V_{Pages}, the number of pages visited in session,
- $V_{Duration}$, the session duration (in seconds),
- $V_{Time\_per\_page}$, the mean time per page (in seconds),
- V_{Source}, the source of the visit (a reference from natural or paid search engine results, an e-mail newsletter, or a social media site, internal reference from a page with entertainment content, or other source).

Two last session features, V_{Is_bot} and V_{Is_admin}, are boolean variables indicating whether the session was performed by a Web bot or the website administrator, respectively.

4.4. Problem Formulation

The research goal is to predict user sessions that ended with purchases, so the sessions were classified depending on if a purchase transaction was finalized in session or not. That is why two session classes (browsing sessions and buying sessions) are considered, differing in the value of $V_{Checkout\_success}$ variable, which is a class label $y_i \in \{0,1\}$, $i = 1,2,\ldots,n$, in fact. A vector of features, x_i, contains 22 predictor variables corresponding to the remaining session features.
K-NN classifiers for different k values were built based on the training set and their performance was evaluated based on the test set. This research was realized using a free software for statistical computing, R-project [25].

5. Results and Discussion

First, each of 22 predictor variables for all observations in the training set and the test set was standardized to have a mean equal to zero and a variance equal to one. After the standardization k-NN classifiers for $k = 1,2,\ldots,40$ were built based on the training set. In all cases the proximity between user sessions was measured by using the Euclidean metric.
Then, the classifiers performance was evaluated using the test set. A classification result for each observation may be a buying session, which is considered to be positive, or a browsing session, considered to be negative. Depending on the actual value of the variable $V_{Checkout\_success}$ (1 or 0) for each observation, the session classification may be true or false. Results of the classification may be expressed with the numbers of true and false positives and negatives by comparing predicted classifications vs. actual classifications:

- true positives (TP) is the number of correctly classified buying sessions,
- true negatives (TN) is the number of correctly classified browsing sessions,
- false positives (FP) is the number of browsing sessions which were incorrectly classified as buying sessions,
- false negatives (FN) is the number of buying sessions incorrectly classified as browsing sessions.

As one can see in Figs. 1 and 2, the classification results differ depending on the number of nearest neighbors taken into account when determining a class label. The highest number of correctly classified buying sessions, i.e., true positives was 63 and this result was achieved for k ranging from 11 to 15 (Fig. 1). Thus, the classifier with 11–15 nearest neighbors was the most effective in predicting online purchases. On the other hand, the highest number of correctly classified browsing sessions, i.e., true negatives was 1292 for k equal to 5 (Fig. 2).

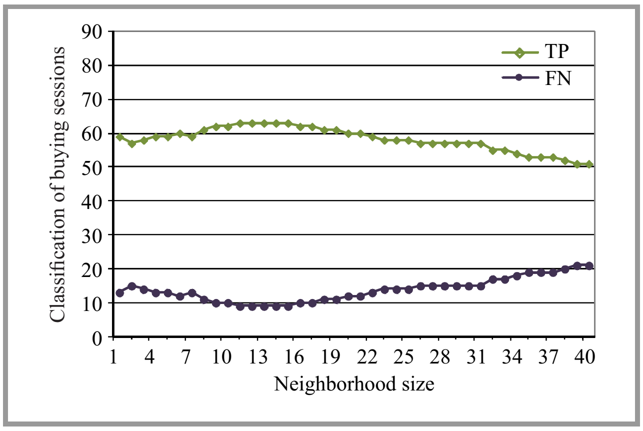


Fig. 1. Results of k-NN classification for buying sessions depending on the neighborhood size.

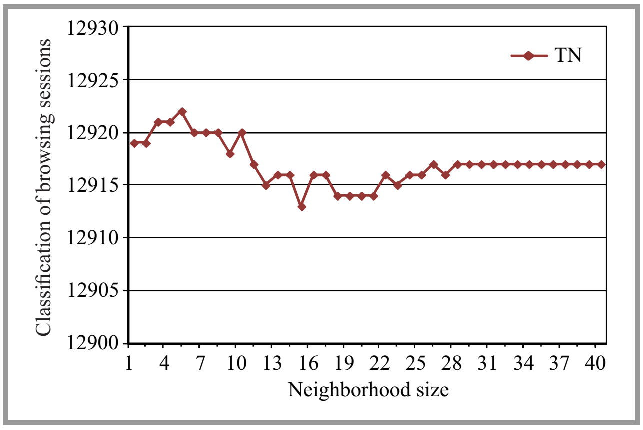


Fig. 2. Results of correct k-NN classification for browsing sessions depending on the neighborhood size.

The quality of the classifiers may be assessed with measures of predictive accuracy, error rate, and sensitivity. The accuracy is defined as the percentage of all correct classifications:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{3}$$

Similarly, the error rate is the percentage of all incorrect classifications:

$$Error\ rate = \frac{FP + FN}{TP + TN + FP + FN}. \tag{4}$$

A sensitivity measure is the percentage of correct classifications of buying sessions, so it is an estimate of the probability of predicting a buying session:

$$Sensitivity = \frac{TP}{TP + FN}. \tag{5}$$

This measure is of the most importance from a retailer's point of view, because the ability to identify potential buyers in an online store makes it possible to apply different incentives to purchase, e.g., by offering special discounts to make an undecided visitor buy.

The performance rates for k-NN classifiers are illustrated in Fig. 3. All the classifiers have a very high predictive accuracy exceeding 99.74% and a correspondingly low error rate below 0.25%. The sensitivity of the classifiers is more differentiated, however, and ranges from 70.83% for k equal to 40 to 87.50% for k ranging from 11 to 15. One can observe that in general, as k increases the classifier sensitivity increases as well until the neighborhood size of 15 and then continues to drop.

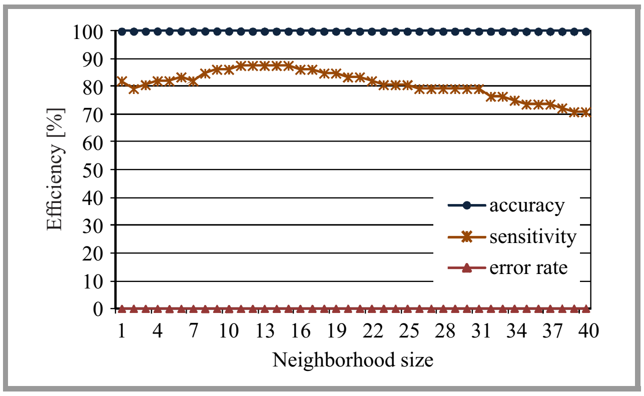


Fig. 3. The efficiency of k-NN classification in terms of all user sessions and buying sessions depending on the neighborhood size.

To sum up, the classifier taking into account the 11 nearest neighbors was the most effective in predicting buying sessions and in terms of overall predictions, achieving sensitivity of 87.5% and accuracy of 99.85%.

6. Conclusions

In this paper the problem of supervised classification of user sessions in a Web store in terms of purchase transaction realization was investigated. Using data from a real online retailer, the authors described each user session with a 22-element feature vector and performed a k-NN classification of sessions into two classes: buying sessions and browsing sessions. Evaluation of the efficiency of k-NN classifiers for different neighborhood sizes showed that a classifier based on the 11 nearest neighbors was the most effective, achieving the overall predictive accuracy of 99.85%, while being capable of correctly classifying 87.5% buying sessions.

For future work, the authors would like to continue research on k-NN classification of e-customer sessions by examining various similarity measures and voting schema among the k nearest neighbors, as well as to apply various dimensionality reduction techniques to the session feature space. It would also be worth verifying the efficiency of the proposed approach for other e-commerce data sets.

References

[1] Q. Duan, J. Li, and Y. Wang, "The application of fuzzy association rule mining in e-commerce information system mining", *Adv. Engin. Forum*, vol. 6–7, pp. 631–635, 2012.

[2] G. Kuang and Y. Li, "Using fuzzy association rules to design e-commerce personalized recommendation system", *TELKOMNIKA Indonesian J. Elec. Engin.*, vol. 12, no. 2, pp. 1519–1527, 2014.

[3] Y.-S. Lee and S.-J. Yen, "Mining web transaction patterns in an electronic commerce environment", in *Advances in Web and Network Technologies, and Information Management – Proc. APWeb/WAIM'07 International Workshops*, Huang Shan, China, 2007, *LNCS*, vol. 4537, pp. 74–85. Springer, 2007.

[4] N. D. Thuan, N. G. Toan, and N. L. V. Tuan, "An approach mining cyclic association rules in e-commerce", in *Proc. 15th Int. Conf. Network-Based Inform. Syst. NBiS 2012*, Melbourne, Australia, 2012, pp. 408–411.

[5] W. Hop, "Web-shop order prediction using machine learning", Masters Thesis, Erasmus University Rotterdam, 2013.

[6] M. Mohammadnezhad and M. Mahdavi, "Providing a model for predicting tour sale in mobile e-tourism recommender systems", *IJITCS*, vol. 2, no. 1, pp. 1–8, 2012.

[7] N. Poggi *et al.*, "Web customer modeling for automated session prioritization on high traffic sites", in *Proc. User Modeling'07*, Corfu, Greece, 2007, LNCS, vol. 4511, pp. 450–454. Springer, 2007.

[8] G. Suchacka and G. Chodak, "Practical aspects of log file analysis for e-commerce", in *Proc. Computer Networks'13*, Lwówek Śląski-Brunów, Poland, 2013, CCIS, vol. 370, pp. 562–572. Springer, 2013.

[9] G. Suchacka, M. Skolimowska-Kulig, and A. Potempa, "Classification of e-customer sessions based on Support Vector Machine", in *Proc. 29th Eur. Conf. Model. Simul. ECMS'15*, Albena, Bulgaria, 2015, pp. 594–600.

[10] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining. Wiley, 2005.

[11] D. Shen, J.-D. Ruvini, and B. Sarwar, "Large-scale item categorization for e-commerce", in *Proc. 21st ACM Int. Conf. Inform. Knowl. Manag. CIKM'12*, Maui, HI, USA, 2012, pp. 595–604.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley, 2000.

[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2009.

[14] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.

[15] J. Cho, K. Kwon, and Y. Park, "Collaborative filtering using dual information sources", *IEEE Intel. Syst.*, vol. 22, no. 3, pp. 30–38, 2007.

[16] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques", *Adv. Artif. Intel.*, vol. 2009, Article no. 4, 2009.

[17] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce", in *Proc. 2nd ACM Conf. Elec. Commerce EC'00*, Minneapolis, MN, USA, 2000, pp. 158–167.

[18] Y. H. Cho and J. K. Kim, "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce", *Expert Syst. Appl.*, vol. 26, no. 2, pp. 233–246, 2004.

[19] X.-M. Jiang, W.-G. Song, and W.-G. Feng, "Optimizing collaborative filtering by interpolating the individual and group behaviors", in *Proc. 8th Asia-Pacific Web Conf. Frontiers of WWW Res. Develop. APWeb'06*, Harbin, China, 2006, LNCS, vol. 3841, pp. 568–578. Springer, 2006.

[20] B. Soiraya, A. Mingkhwan, and C. Haruechaiyasak, "E-commerce web site trust assessment based on text analysis", *Int. J. Business Inform.*, vol. 3, no. 1, pp. 86–114, 2008.

[21] M. Adnan, M. Nagi, K. Kianmehr, R. Tahboub, M. Ridley, and J. Rokne, "Promoting where, when and what? An analysis of Web logs by integrating data mining and social network techniques to guide ecommerce business promotions", *Soc. Netw. Anal. Min.*, vol. 1, no. 3, pp. 173–185, 2011.

[22] L. D. Catledge and J. E. Pitkow, "Characterizing browsing strategies in the World-Wide Web", in *Proc. 3rd Int. World-Wide Web Conf. Technol., Tools Appl.*, 1995, pp. 1065–1073.

[23] Z. Chen, A. W.-C. Fu, and F. C.-H. Tong, "Optimal algorithms for finding user access sessions from very large Web logs", *World Wide Web*, vol. 6, no. 3, pp. 259–279, 2004.

[24] D. Stevanovic, N. Vlajic, and A. An, "Unsupervised clustering of Web sessions to detect malicious and non-malicious website users", *Procedia Comput. Sci.*, vol. 5, pp. 123–131, 2011.

[25] The R project for statistical computing [Online]. Available: http://www.r-project.org

---

Grażyna Suchacka received the M.Sc. degrees in Computer Science and in Management from Wrocław University of Technology, Poland. She received her Ph.D. degree in Computer Science from Wrocław University of Technology. Now she is an assistant professor in the Institute of Mathematics and Informatics at Opole University, Poland. Her research interests include Web mining, Web analytics, and Quality of Web Service with special regard to electronic commerce.

E-mail: gsuchacka@math.uni.opole.pl
Institute of Mathematics and Informatics
Opole University
Oleska st 48
45-052 Opole, Poland

Magdalena Skolimowska-Kulig received her Ph.D. degree in Mathematics from University of Wrocław, Poland. Now she works as an assistant professor in the Institute of Mathematics and Informatics at Opole University, Poland. Her research interests include theory of probability and mathematical statistics.

E-mail: mskolimow@math.uni.opole.pl
Institute of Mathematics and Informatics
Opole University
Oleska st 48
45-052 Opole, Poland

Aneta Potempa received the B.S. degrees in Mathematics from Opole University, Poland and in Economy, accounting and audit from WSB Schools of Banking in Opole, Poland. She received the M.Sc. degree in Financial Mathematics from Opole University in 2014. She works in a controlling department in GEA Technika Cieplna, the company specializing in production of industrial heat exchangers.

E-mail: aneta.potempa@gea.com
GEA Technika Cieplna Sp. z o.o.
Kobaltowa st 2
45-641 Opole, Poland

Review

Historical Perspectives of Development of Antique Analog Telephone Systems

Vinayak L. Patil

Trinity College of Engineering and Research, University of Pune, Pune, India

Abstract—Long distance voice communication has been always of great interest to human beings. His untiring efforts and intuition from many years together was responsible for making it to happen to a such advanced stage today. This paper describes the development time line of antique telephone systems, which starts from the year 1854 and begins with the very early effort of Antonio Meucci and Alexander Graham Bell and ends up to the telephone systems just before digitization of entire telecommunication systems. The progress of development of entire antique telephone systems is highlighted in this paper. The coverage is limited to only analog voice communication in a narrow band related to human voice.

Keywords—antique telephones, common battery systems, crossbar switches, PSTN, voice band communication, voice communication, strowger switches.

1. Initial Claims and Inventions

Since centuries, telecommunications have been of great interest to the human beings. One of the dignified personality in the field of telecommunication was Antonio Meucci [1]–[7] (born in 1808) who worked relentlessly for communication to distant person throughout his life and invented telephone in 1849. Although he was nowhere near to claim his invention, he will be remembered for his noble work he did at that time.

In 1854, Meucci migrated to Clifton area of Staten Island near New York [8]. He set up communications link in his Staten Island home that connected the basement with the first floor, and later, when his wife began to suffer from crippling arthritis, he created a permanent link between his lab and his wife's second floor bedroom. Meanwhile, he perfected his instrument to be able to demonstrate for financial backup. In 1860 he demonstrated it publicly, which appeared in New York's Italian language newspaper. His life's saving was exhausted in perusing his invention. Meucci was unable to raise sufficient funds to pay his way through the patent application process. However, he was able file the caveat in 1871 and renewed in 1872 and 1873 but could not renew it after that. The constructional details of Meuicci's telephone instrument are shown in Fig. 1.

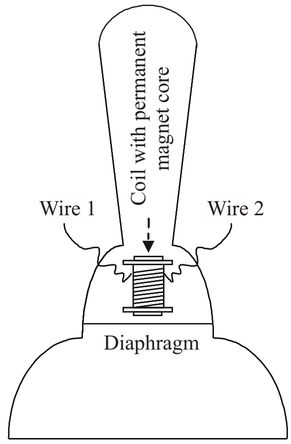


Fig. 1. The details of Meucci's telephone.

Many contestants were in fray for the invention related to telephones. In 1854, Bourseul [9] publishes description regarding make and break telephone transmitter and receiver to put forward an idea of how to transmit speech electrically. Due to this idea, many of the scientific community consider him as one of the inventors of telephone [10]. Boursuel used term "make and break" telephone in his work. In 1850, Philip Reis [11]–[13] began work on telephone. Working towards the Bourseul's idea, he was successful in his attempts in 1860. His work appeared in New York Times editorial on March 22, 1876. His instrument consisted of a horn to collect the sound that was allows to strike the diaphragm, which was in loose contact with point contact. The loose contact represented a variable resistance responsible for modulating signaling current. While receiver was consisted of needle that was made to vibrate by signaling current, while these vibrations were acoustically amplified by mounting vibrating needle on sound box

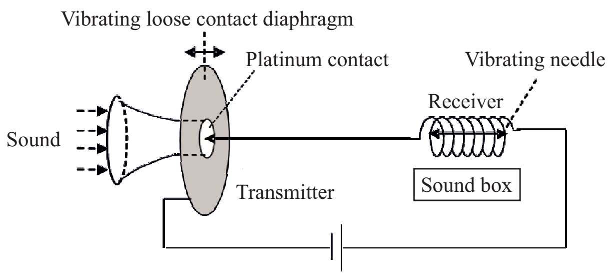


Fig. 2. Reis's principle of sound transmission.

itself. The adjustments related to this needle were critical for device operation. However, with these arrangements, he was able to transmit tones and some vowels the Reis instrument shown in Fig. 2 was better suitable for music than human voice. The era of work of Reis was nearly coinciding with an era that of Meucii.

1.1. Bells's First Telephone Design

On June 2, 1975 Bell along with his assistant Thomas A. Watson working in Williams machine shop in Boston discovered that sound notes can be electrically transited trough wires from one room to other using electromagnet. The Bell's first telephone also called as Bell's Gallows Telephone [12], [14]–[17] consists of parchment membrane, reed relay armature, core and a coil (Fig. 3a). The voice/sound is responsible making membrane to vibrate and the movements are then picked up by the reed relay armature. As this reed relay armature is moved through the magnetic field of relays coils, it induces fluctuating electric current in the coil. The variation in the current is then transmitted to receiving relay for further reproduction.

This first telephone design never worked, possibly because the coil did not use permanent magnet to produce sufficient current and possibly the membrane could have made up of iron disk instead of parchment membrane. This gallows type telephone is shown in Fig. 3b.

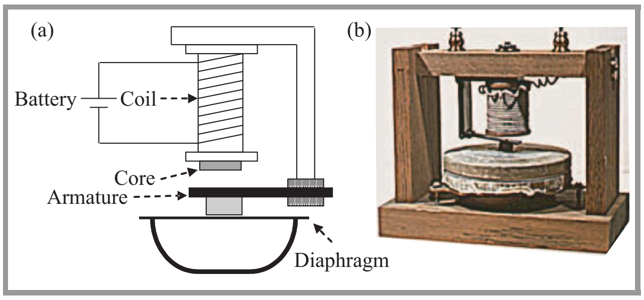


Fig. 3. Bell's first gallows telephone.

As per the records of Bell's note book pages, his invention related to telephone was ready on January 20, 1876 and on February 14, 1876, he filed a patent application regarding "Improvement in telegraphy" [19], [20], while just few hours later the caveat entitled "Transmitting vocal sounds telegraphically" was filed by attorney of Elisha Gray [21], [22] professor at Oberlin College. Transmitting end of Elisha's instrument (Tone telegraph) operated on similar way as that of Reis. It used vibrating steel rod that interrupted a current in the circuit. While at receiving end, it used a electromagnet with steel reed near the magnetic pole (Fig. 4).

On March 7, 1876, as per United States patent law and its clause, issued a patent to Alexander Graham Bell (No. 174465) declaring him, the legal inventor of telephone. It was not also free from controversies those still exists till today. The controversy was triggered by application filed on the same day by Gray [19].

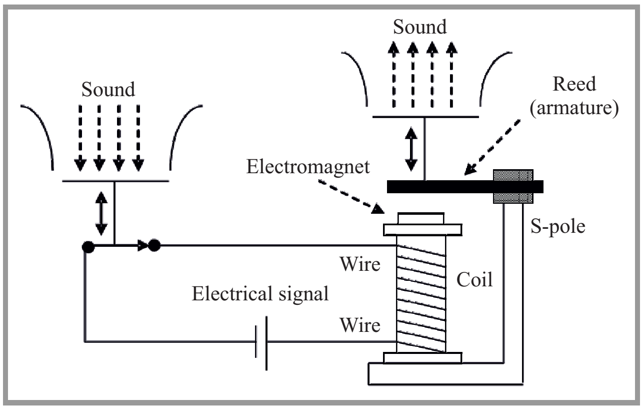


Fig. 4. Elisha Gray's work on sound transmission.

The inventions made by various researchers were shadowed by legal status awarded by United States patent office to Bell and all the issues related to "Invention of telephone" were then buried for silence.

2. Producing Electrical Voice Signals

Earlier mechanisms were based on coils constructed using permanent magnets. One pole of magnet is brought near to other pole through armature and is nearly in contact with each other. Sound vibrations sensed using a diaphragm are transmitted to armature changing the gap between two poles of permanent magnet. These changes of flux are then responsible for change of current in coil. This mechanism of sensing is shown in Fig. 5.

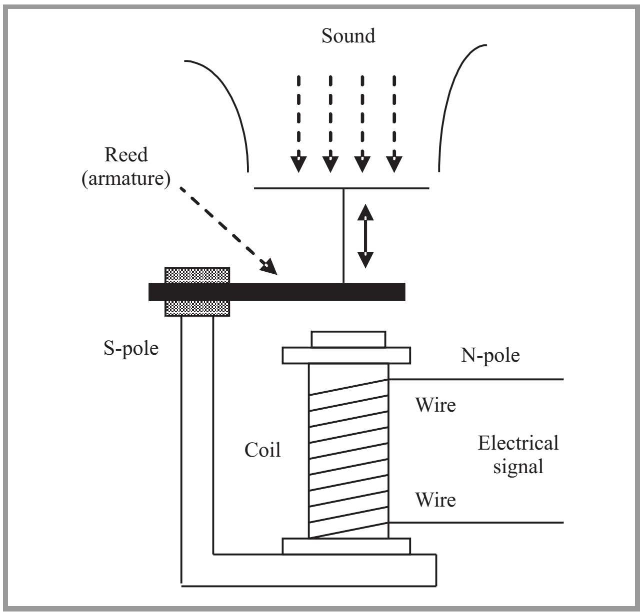


Fig. 5. Conversion of vibrations to electrical signals.

At this point it is to note that inward diaphragm stroke produces current in one direction while outward diaphragm stroke produces current in opposite direction. This opposing current has an effect to degradation of magnetic properties of a permanent magnet used and hence magnitude reverse current should never be great enough to nullify the reverse magnetism of core. Degradation of magnetic property leads to reduction of armature gap. Figure 5 presents typical magneto transmitter and receiver that can be used for voice transmission.

3. Telegraph and Telephone Experiment by Alexander Graham Bell

Bell used reed relays in his experimental setup [23], [24] to show that vibrations generated by tuning fork are picked up and converted to electrical pulses by reed relay at one end. These electrical pulses are sent over a pair of wires. At other end, yet another reed relay converts these electrical pulses back to vibrations. Figure 6 shows the representative experimental setup used by Bell experiment. He has used magneto design using soft iron instead of hard magnetized steel. The coil used in Bell's experiment is shown in Fig. 6.

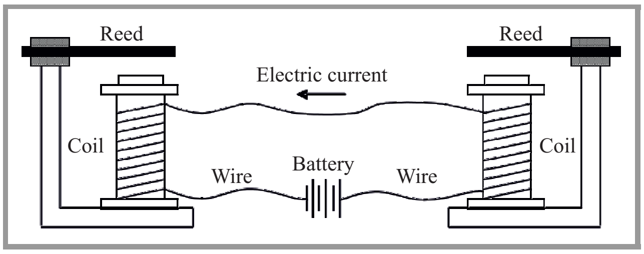


Fig. 6. Experimental setup used by Bell.

With this experiment, Bell was able to show that vibrations from sounds can be electrically transmitted from one end and reproduced at other end, can serve our purpose of communication to distant person. The dc current in the wire loop gets modulated due vibrations and variations in the current are sensed by a relay at other end. This experiment was the basis for further developments in telephony.

4. Developments of Microphones

To incorporate transmission of human voice required the sensor that can sense vibrations related to human voice for transmission over a device.

In 1827, Sir Charles Wheastone made an instrument to amplify the weak signals and called it as microphone and from that time the phrase microphone existed [25]. Later on, Emile Berliner, invented first microphone in the year 1876 for converting a voice signal to electrical signals [26]. Emile Berliner filed a caveat on April 14, 1877, and then he advanced an application for United States patent on June 4, 1877 and to this effect he was awarded with a patent of "Contact telephone (US Patent No. 222652, December 16, 1879)". As per his invention, sound striking point of contact can cause pressure difference that is responsible for producing a change in current related to sound. Its construction is shown in Fig. 7.

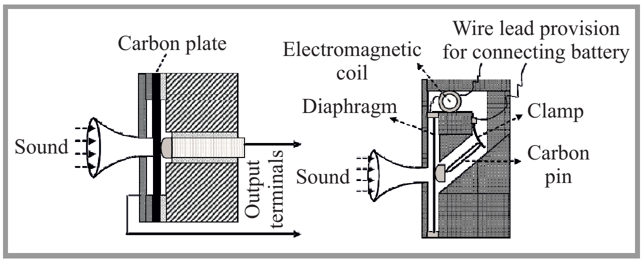


Fig. 7. Emile Berliner's microphones.

Later on Berliner filed a patent for improvement of microphone [27] on August 11, 1979. This invention consists of dispensing with the clamping device for fixing the carbon pin in position, maintaining a contact with diaphragm under gravity (Fig. 7).

He also invented the transformer that prevents signals from weakening. These inventions were responsible for transforming Reis's and Meucci's telephone instruments to practical one. Both Emile Berliner and Thomas Alva Edison filed the patents for carbon microphone. The Berliner filed his patents in the years 1877 and 1879 while Thomas Alva Edison filed patent in the year March 1878. However, in 1892, Thomas Alva Edison won the legal battle on the pretext that "Edison preceded Berliner in the transmission of speech" [28]. Carbon microphone [29], [30] invented by Edison can convert human voice to electrical signal to be put on dc current loop responsible for long distance voice communication.

The carbon-button microphone invented by Edison generated better signals. The mechanism consisted of carbon button [19] formed out of carbon powder collected from kerosene lamp and chimneys, which was pressed into a compact button form. A change of resistance of this button occurred due to sound pressure and hence changes of current which is representative of voice. The Edison's carbon transmitter is shown in Fig. 8.

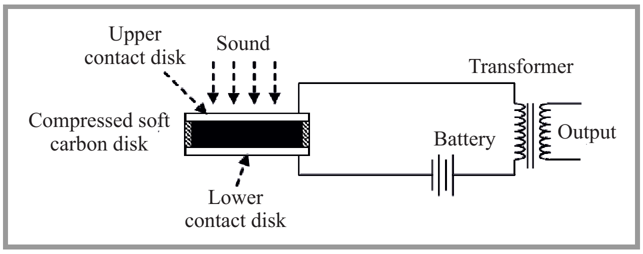


Fig. 8. Edison's carbon microphone.

One of the early models of carbon microphone was invented by David Edwards Hughes [31]–[33] in 1878. His inventions were truly adopted. He was critical of disclosing his invention first to Royal Society rather than patenting it. Before such disclosures, he even refused to patent his inventions. The Hughes microphone is shown in Fig. 9.

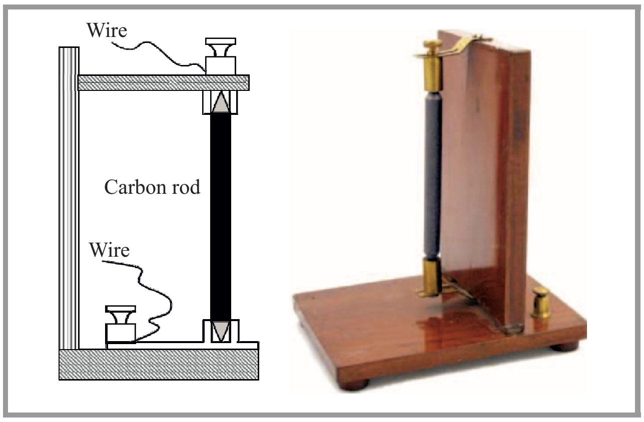


Fig. 9. Hughes carbon microphone.

Dynamic moving coil microphones were introduced by Ernst Siemens in 1874, and during 1916, at Bell labs, wide range quality condenser microphones was developed by Wente [34]. Dynamic moving coil microphones were also introduced in late 1920 by Wente and Thuran.

4.1. Carbon Microphone

Figure 10 shows the construction of carbon microphone and associated circuit. The microphone consists of cup containing carbon granules. Cup is covered with the diaphragm such that both diaphragm and cup are electrically isolated to act as two terminals. As matter of fact, they acts as two points of resistance, which varies as per the sound vibrations. Certain amount dc current is passed through the series combination of microphone and primary winding of the transformer. The dc current varies through the circuit when resistance of carbon microphone is changed due to sound waves striking the diaphragms giving the output at the secondary side.

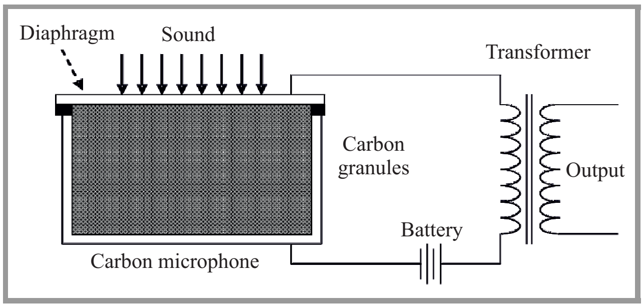


Fig. 10. Carbon microphone circuit.

Since carbon microphone uses carbon granules in the cavity between its two terminals, the granules in motion give rise to some amount of random variation in resistance leading to constant hiss affecting signal to noise performance of the microphone. However it has high sensitivity (about 100 mV of output signal). The carbon microphones were become standard in early days of telephony.

4.2. Magnetoelectric Moving Coil Dynamic Microphone

The moving coil dynamic microphone consists of voice coil attached to a diaphragm that vibrates due to variable sound pressures. The principle of working is explained by Fig. 11 which consists of two parts, one part is permanent magnet and other is a coil itself. The flux created by the permanent magnet between its north and south poles. When sound vibrations strike on the on the diaphragm vibrates, that is responsible creating the coil movements within a flux created by north and south poles of permanent magnet, intern that also creates an electrical output voltage at the ends of electric coil equivalent the magnitude of sound created near to the diaphragm.

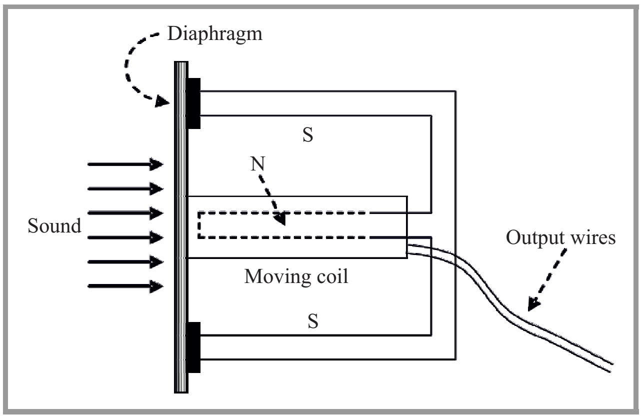


Fig. 11. Moving coil dynamic microphone.

4.3. Condenser Microphone

In condenser microphone the capacitance offered by two parallel plates is varied by sound waves, resulting in varying charge across the capacitor. The charge variation produces change of current in the circuit and further change of voltage across the resistance that is, dependent on sound waves (Fig. 12).

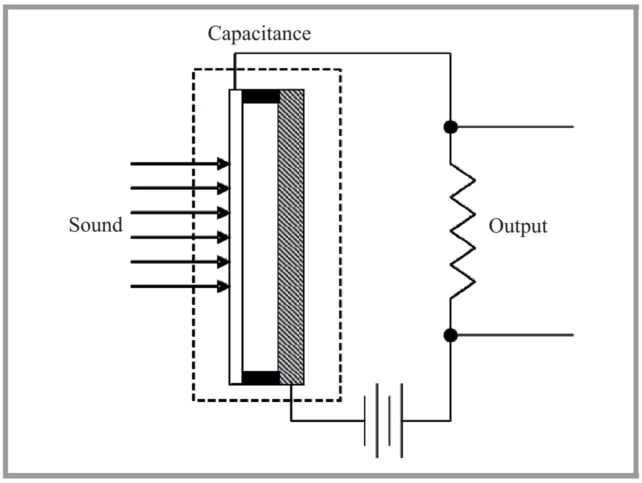


Fig. 12. Condenser microphone.

The inventions related to telegraphy and telephony made lot of impacts on the world of telecommunications. In telephony, microphone severed as critical block for the entry of voice into the telephone experiment by Bell, which made use of tuning forks.

5. Bell's Experiments for Voice Transmission

Articulation of voice using Bell's initial experiment predicted the shape and manner in which things to come in future. The efforts were on for coupling voice to circuit invented by Bell [24].

5.1. Bell's Liquid Transmitter

On March 10, 1876, Bell successfully passed sounds from one end to other end [12], [16], [22], [33], [35]– [39]. This experiment was conducted using Bell's liquid transmitter. In this experiment, instead of using tuning fork, the human voice is allowed to strike diaphragm which in turn, drives a conducting rod just making the contact with the acid placed in a cup. The cup body and conducting rod act as variable resistance whose resistance is varied depending on voice. This arrangement converts voice to electrical signals to be carried over Bell's initial experiment (Fig. 13).

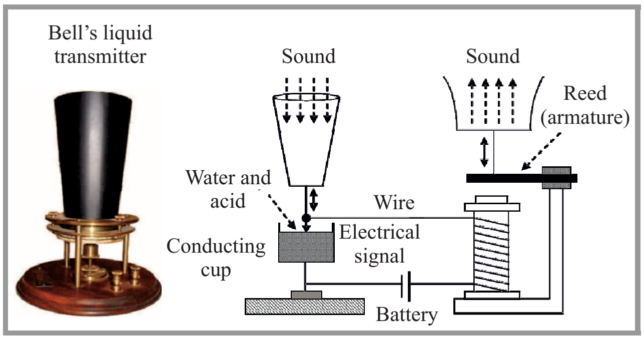


Fig. 13. Bell's voice experiment using liquid transmitter.

5.2. Bell's Centennial Telephone Model

The Bell's liquid transmitter was not convenient for use and hence he developed yet another transmitter free from use of any liquid, which was called as Bell's centennial model [16], [33], [40]–[44]. This model used electromagnetic transmitter, metal diaphragm and permanent magnet [42] (Fig. 14).

Since Bell's initial experiment served as a solid base for voice transmission, the ways for sound conversion were invented. Electromagnetic devices using soft iron and not of hard magnetized steel played important role in the conversion of sound waves to electrical signals and vice versa using such conversion as shown in Fig. 15. The voice coupled to electromagnetic device was clearly and distinctly heard at other end. Figure 15a shows the type of arrangements used while Fig. 15b shows circuit representation. The pull upon the diaphragm is provided by the flux that is

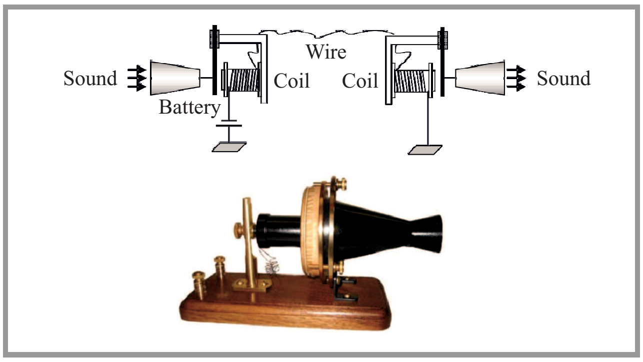


Fig. 14. Bell's centennial telephone.

generated by the loop current due to battery, which is inserted in series with the line and not by flux due permanent magnet as in earlier cases.

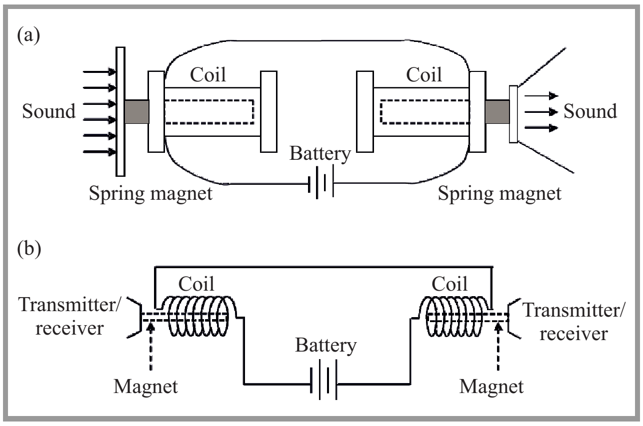


Fig. 15. Principle of voice transmission using electromagnetic coils.

At both the ends electromagnetic devices are used for conversion of voice signals to electrical signals and vice versa. The diaphragms are attached to spring magnets for such conversion leading to such novel invention that became an eye opener for the development of systems related to telephony.

6. Next Inventions

After success of Bell's experiments for voice transmission, he continued with his experiments for calling distant person. At the end of 1876, due to earlier Bell's experiment, Boston experienced intense telephone activities and served as a center of telephone activities and many private telephone lines started coming up. In many cases existing telegraph lines used for connecting equipments on these telegraph lines were replaced by early type telephone instruments.

Later on, October 9, 1876, the experiment was performed by attaching telephones to both ends of 2 mile (3.2 km) telegraph line between Boston and Cambridge that was owned by the Walworth manufacturing company for connecting their office in Boston and factory in Cambridge. In this type of experiment, nine Daniell cells were used as battery [46], [47]. However, lot was to be done for commercialization of telephone.

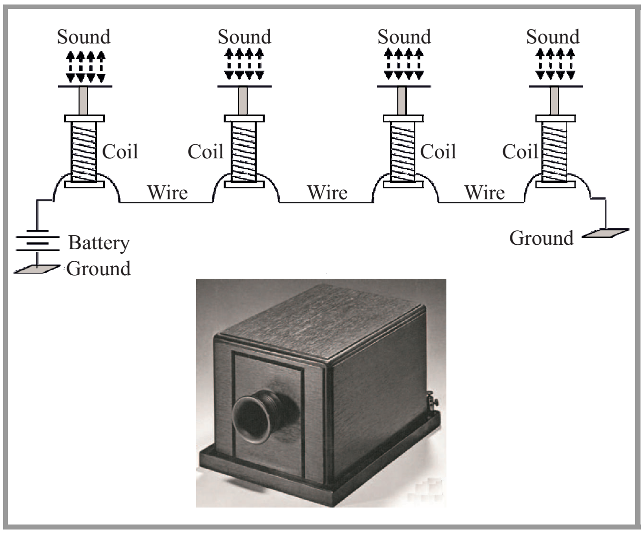


Fig. 16. Bell's box telephone connections and model.

On 30 January 1877, yet another patent [17], [48] was granted to Bell for an electromagnetic telephone using permanent magnets, iron diaphragms, and a call bell. Based on this second patent [48]–[50], a box telephone (looking like camera) was designed. A camera like opening on box served as a place for both transmitter and receiver. The box telephone was the first commercial telephone and a Boston banker was first to lease this type of instrument for connecting his office and home in Somerville, Mass. There were many difficulties using box telephone such as one can either talk or listen and signaling to the party at other end was done by tapping on the diaphragm. The figures related to patent of Bell's box telephone are shown in Fig. 16 [51].

6.1. Telephone Line Configurations

The conversation between any two customers out of any n customers requires direct wire connection between these two customers. So if a network of 2 telephones is to be formulated, it required $2(2-1)/2 = 1$ connections, similarly if a network of three telephones to be designed, $3(3-1)/2 = 3$ connections are required. By generalizing this concept for n telephones e $n(n-1)/2$ connections or links are required. This is shown in Fig. 17.

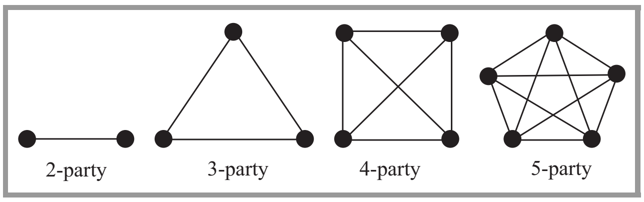


Fig. 17. General telephone networking concepts.

In order to reduce number of telephone lines, concepts of line sharing was put forward. These arrangements are called as party lines [16]. Single party line is used to serve more than two stations including the central office as one of the station if any. Some autonomous company using its own exchange for its private cause having its own lines are called as private lines. These lines are used to connect two or more isolated stations. However, private lines working in conjunction with telephone exchange may be called as party line, which may not seem to be consistent with the basic definition. There are two types of party line configurations called as series configurations and bridged configurations.

The earliest type of grounded-circuit series party line configuration is shown in Fig. 16. The box telephones worked on grounded circuit series configuration while Butter stamp telephones worked on grounded bridged configurations (Fig. 18).

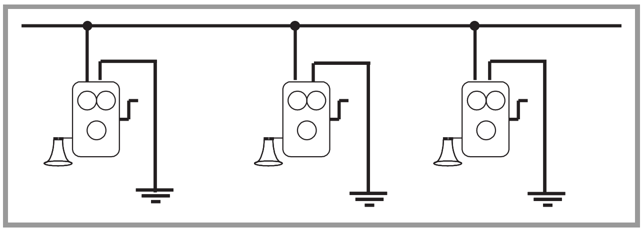


Fig. 18. Party line grounded bridged configuration.

In 1881, John Joseph Carty, an electrical engineer working in Bell's manufacturing unit called Western Electric Co. demonstrated advantages of two wire metallic circuit and after this, metallic party line configurations became popular (Fig. 19).

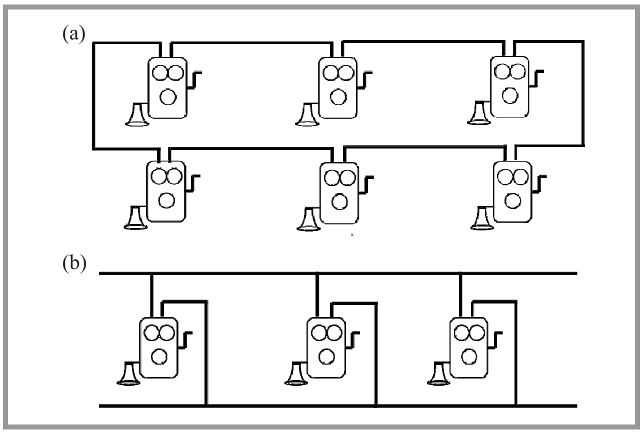


Fig. 19. Party lines metallic configurations: (a) with series connection, (b) bridge configuration.

6.2. World's First Long Distance Telephone Line

The world's first long distance telephone line was setup between French Corral mining campus and French lake (58 miles, 93 km) [52]–[54] in California, USA in 1877 and had 22 telephones. This link was built by Ridge Telephone Company [55] and operated by Milton Mining Co. Line was operated using "Edison's speaking telephones" [56]. Meanwhile, in 1877, Tivadar Puskas designed first telephone exchange for Thomas Edison [57].

In April 15, 1877, first regular 3 mile (4.8 km) telephone line was established from Boston to Somerville, Massachusetts [39], [58] which was the first commercial service in USA. Up to 1880, the customers' number grew up to 47900, while in the year 1878, workable exchange was developed for enabling call switching amongst various customers.

6.3. Butter Stamp Telephone

Bell's box telephone model (combined receiver/transmitter) was not so convenient to use because of its shape and hence it was modified to hand held telephone unit. This wooden hand held telephone came in existence in May 1877 [60]. This telephone model was attached to a base terminal stand using flexible wire cable.
To initiate a call, the operator at central place was being alerted using push button circuit [16], [61]–[63]. It used permanent bar magnet over which coil was wound to pick and reproduce the sounds. It looked like butter stamp that was common at the time and hence the name (Fig. 20). In this telephone, although both transmitting and receiving functions were performed by same mechanism, transmission of sound was poor as compared to received signal.

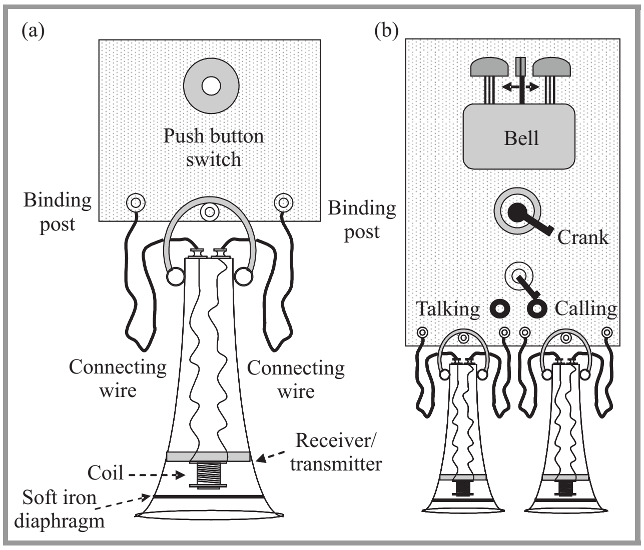


Fig. 20. Two wall units using butter stamp telephones: (a) basic model, (b) the modified version.

The single combined wooden hand held transmitter and receiver telephone was not still convenient to use for both taking and listening, hence two such units are used, one dedicated for speaking and other for listening. The model had come in the form of wall unit. Instead of push button, it used a crank to signal the operator to initiate that call is being initiated when "calling/talking lever" is set to "calling". In next version a bell was also added a to indicate incoming call when "calling/talking lever" is set to "talking" (Fig. 20b).
The circuits related to both the cases of wall units of butter stamp telephones are shown in Fig. 21. The first circuit (Fig. 21a) shows single unit while in other circuit (Fig. 21b) shows the modified device.

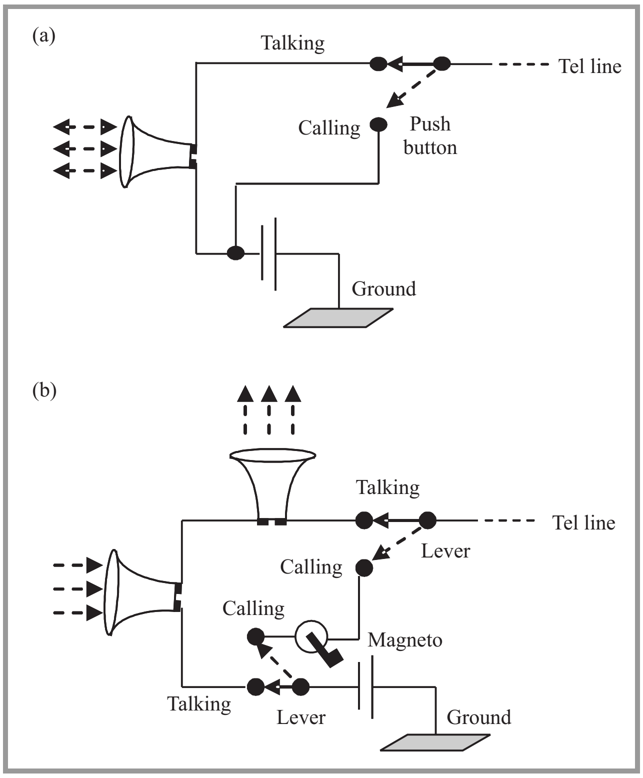


Fig. 21. Butter stamp telephone circuits: (a) single unit, (b) the modified version with separated transmitter and receiver.

6.4. 1877 – A Year of Public Acceptance of Telephone

1877 was the year of telephone business, and also an official and public recognition of telephone on existing telegraph lines, which helped the rapid spread of telephone services. On July 9, 1877 Bell transferred all the rights of this telephone to Mr. Hubbard and by July 31, 1877, as many as 778 telephones were leased to public by Mr. Hubbard [46]. By the end of year 1877 total of 5491 telephones of Bell were leased. On August 1, 1877, Hubbard organized an association called Bell Telephone Association of Boston and acted as its trustee while Sanders as treasurer. Bell, Sanders and Hubbard created in 1878 company called Bell Telephone Company (BTC) [17], predecessor of today's AT&T.
In USA, on October 24, 1877, two lease deeds for territorial rights were executed with Telephone and Telegraph construction Company, Detroit, Michigan and District telephone Company, New Haven, Connecticut. However District telephone Company, New Haven, Connecticut could take the lead in establishing world's first telephone exchange [16].

6.5. Developments at Western Electrical Manufacturing Company

Elisha Gray was also working on the invention of telephone and fought legal battle with Bell. However, Bell won the legal battle on the basis that Bell had working experimental model. Meanwhile, telecommunication industries started to take off. A Telegraph Industry Supply Company (TISC) was founded in 1869 by Elisha Gray and Enos Barton. Later on in 1872, the company changed its name to Western Electrical Manufacturing Company (WEMC).

The WEMC was the principal manufacturer for Western Union (WU). In 1877, WU decided to compete with Bell and entered in the commercial production of telephones [65]–[67]. It was also aware that although centennial model was successfully demonstrated at Philadelphia as well as Butter stamp telephone was successfully used, transmitter and receiver were notably weak in terms of voice signal conversion.

Owing to above weaknesses, it conducted extensive tests of various models of telephones designed by Edison, Gray and Phelps. The designs of Bell, Gray and George Phelps used magneto design. In earlier designs, magneto were used at both ends for speaking and listening created some sort of inconvenience to the user. After speaking, the same device was used for listening. Furthermore, telephones based on magneto were suffering from weaker signals when transmitted over longer distances.

Till 1875, Edison's concentration was on telegraphs related problems and did not bothered for working for telephony, during the same year Edison started working on carbon microphone. At the same time, Bell was concentrating on the work of sound conversion using electromagnetic principle. In 1876 Edison started working with Western Union and William Orton [28], [68], and concentrated on conversion of sound using a pair or single carbon electrodes in contact with each other that was exposed to sound. This mechanism generated strong signals as compared to the principle of electromagnetic sound conversion. Orton pursued Edison for filing of the patent for which he paid 100,000 US dollars to Edison.

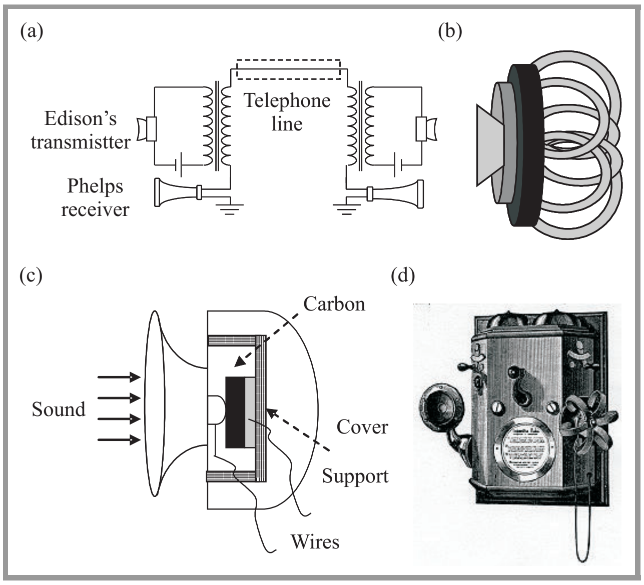


Fig. 22. Telephone model used in exchange opened in San Francisco in 1878: (a) schematic, (b) Phelps receiver, (c) Edison's carbon transmitter, (d) photo of telephone.

Edison's carbon microphone solved these problems of magneto design. In early 1878, Western Union evaluated Edison's carbon microphone for a link between New York and Philadelphia and determined that it has superior performance. While for receiving, Phelps multipole design [69] (Fig. 22) was used [28], [69], [70].

The Phelps design used number of magnets having their individual coils for efficient performance. This combined design was used in telephone exchange opened by Gold and Stock Telegraph Company (GSTC) in San Francisco in February 1878 and was marketed by Western Union of New York (run by Phelps) through their subsidiaries called American speaking telephone company (ASTC) and Gold and Stock Telegraph Company (GSTC) from 1877 to 1879.

6.6. World's First Manual Commercial Telephone Switchboard

The world's first manual commercial telephone exchange was developed by George W. Coy based on the talk of Bell regarding importance of exchanges in commercial telephone lines, Coy applied for franchise from BTC for New Heavens and Middlesex. Coy along with Herrick P. Frost and Walter Lewis established District Telephone Company of New Havens on January 15, 1878. The company grew rapidly in 1880, as it expanded. It was renamed to Connecticut Telephone and then Southern New England Telephone in 1882. This world's first manual commercial telephone exchange was went in operation in Boardman building at New Havens, Connecticut on January 28, 1878 [71], [72], which was designed by Coy himself. Initially this facility went in operation with 21 subscribers on 8 lines [73]. These exchanges were consisted of several hundred plug boards. Early switchboard was built from "carriage bolts, handles from teapot lids and bustle wire" and could handle two simultaneous conversations. Later exchanges consisted number of plug boards operated by the operators. Operator has to sit in front one to three banks of phone jacks fronted by several rows of phone cords, each of which was the local termination of a phone subscriber line. The constructed schematic switch board used in this exchange is shown in Fig. 23.

The switch board [74], [75] consists of four arms that can be rotated in circular fashion for making the contacts, party line terminations, annunciator, a strip along with contacts, operators telephone instrument and a calling device.

The switch board receives eight party lines and each party line caters to on an average twelve subscribers so that exchange had a capacity to serve to $12 \times 8 = 96$ subscribers. The exchange also supports two simultaneous calls.

The operator can manipulate the connections using four rotating brass arms [76]. Out of these four arms, two arms

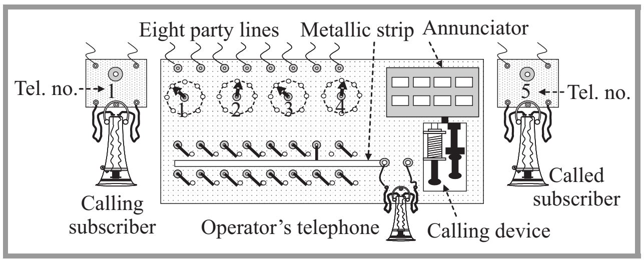


Fig. 23. Manual switch board at New Havens, Connecticut.

are used for connecting two subscriber wires (caller and called subscribers) on the switch board, third arm is used for connecting operator to circuit and fourth arm is used to ring called subscriber.

The call at that time was materialized as per the following sequence.

At first instance, caller draws the attention of switch board operator by pressing push button on his telephone that actuates a single stroke bell in the exchange. This action also releases a drop in annunciator indicating a party line to which caller is attached ("1" in Fig. 23). On hearing the bell and knowing the calling party line from the annunciator display, operator moves third rotating arm to 7 and also turn respective "single switch (no. 7)" to metallic strip to connect his telephone (operator's telephone) to a caller's telephone for inquiring the subscriber to whom he wants to connect. After knowing the identity of party to be called, "single switch" is moved off position.

During next step, calling device (also called as Watson's Squealer or Coy's chicken) is attached to a party line on which subscriber being called is attached (say party being called is attached to party line 8 with subscriber or telephone number is 5). Then calling device is attached to line number 8 using fourth rotating arm. The attached device is then used to send 5 long squeals over this line to indicate that this call is for subscriber 5 and not for others.

To establish connection between caller and called party, first and second arms are used. The arm of first circle is rotated to respective position of calling subscriber line (7) while arm related to other circle was set to calling subscriber line (8). The butter stamp telephones were to communicate through this exchange established at New Havens. In February 21, 1878. The telephone directory related to its 21 customer was published.

6.7. Origination of PSTN Concept

Although first commercial telephone exchange at New Havens was opened by BTC, as on 1877, Western union had lot of potential in terms of its already existing 250,000 miles of telegraph lines covering 100,000 miles route [77] and had unchallenged monopoly for making it a largest telecom company at that time. However, this monopoly was later on ended because of the role of Bell's shrewd lawyer. On February 17, 1878, Western Union, on the basis of its capacity, opened its first large city exchange with 18 phones in San Francisco [78], [79] that enabled any of the users to talk to any other using different lines rather than dedicated lines that generated a concept of Public Switched Telephone Network (PSTN).

The year 1878 was a year of many firsts, telephones exchanges spread quickly to other locations and was a year of commercial spread of telephones. It included opening of first telephone exchange at Albany, New York state on March 18, 1878 [80] and first telephone exchange at Lowell, Massachusetts on April 19, 1878. In telephone exchange, at Lowell, telephones were designated by the numbers [81], [82]. Earlier to this, operators simply knew the plug related to the customer telephone. The first telephone exchange outside United States was set up in Hamilton, Ontario, Canada on July 15, 1878.

These earliest telephone lines were called as grounded lines [83] in which only one wire is used for carrying currents related to telephone metal while other point is a ground point. The grounded line telephone system is shown in Fig. 22a. At later stages, two wire telephone systems were used and these circuits using two wires were referred as metallic circuits.

There were some advantages of grounded configurations such as less cost of lines due its use of only one wire. The batteries were used locally at each end and there was no central provisioning of dc voltage supply by telephone companies. The multiple users can be configured as shown in Fig. 24.

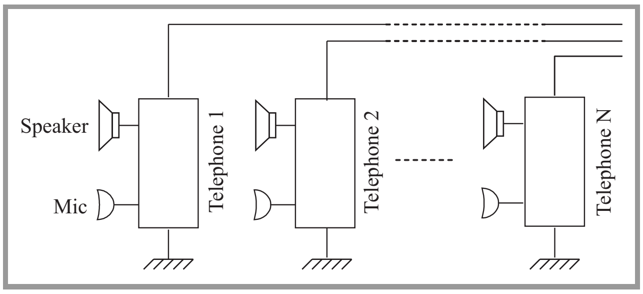


Fig. 24. Grounded telephone line configurations.

6.8. Thomas Watson Files Patent for a Ringer

On August 1, 1878 Thomas Watson files a patent for ringer [84], however, crude bells were in use in earlier telephones. It was similar to Henry's class room door bell. Hammer strikes two bells was that is attached to an armature with magnetic field strengthened by permanent magnet that moves due ringer current. The patent was awarded on Dec 17, 1878. These rings were operated by AC current produced by magneto of calling end telephone.

7. Blake's Transmitter

On other side of the developments, BTC was using a transmitter invented by Berliner [87], [88]. However, BTC was not satisfied with its performance and hence it was modified by Francis Blake Jr. However, during this period of developments, Bell continued to use Berliner's transmitter till 1879.

Although centennial model was successfully demonstrated at Philadelphia as well as Butter stamp telephone was successfully used, transmitter and receiver were notably weak in terms of voice to signal conversion and signal to voice conversion respectively. Blake promptly identified this weakness and started working towards this problem and produced the working model that was taken to BTC offices in Boston for testing on October 18, 1878 [89], [90]. After testing it, Watson had no doubt that this was better transmitter as compared to any of the transmitters available at that time. Blake used carbon and platinum resistance elements which was again improved later by Berliner and remained standard for many years. Blake provided commercially viable solution for transmitter to a BTC. Figure 25 shows photograph and schematic of Blake's transmitter.

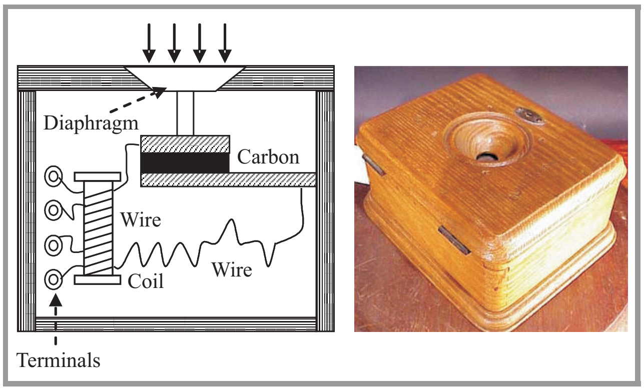


Fig. 25. The Blake transmitter.

The Blake transmitter was powered using local wet batteries [46], [92], which was not so convenient for use due to acid leaks. Blake transmitter was based on the modification of Edison's carbon transmitter patent owned by Western Union Telegraph company and that is the reason why Bell could not adopt the device invented by Edison. Subsequently Bell acquired the patent rights of Edison's patents through the successful lawsuits against Western Union Telegraph Company. This opened new avenues for BTC that accelerated the use of carbon transmitter in the form of Blake transmitter.

It is interesting to note that customer side telephone equipment was then consisted of many modules such magneto receiver (Bells's hand receiver), Blake transmitter and battery pack (at the bottom) (Fig. 28), while in some cases transmitter is located at the left side of magneto receiver.

8. The Automatic Calls Switching

During 1879, M. D. Connolly and T. J. Mc Tighe [93]–[95] discussed an idea of automatic switching with their nephew Almon Brown Strowger and later on, Strowger was able to put forward their idea practically. The application for first dial telephone exchange patent was advanced by a team of three persons M. D. Connolly, T. A. Connolly and T. J. Mc Tighe on September 10, 1879 and a team was awarded with a patent on December 9, 1879.

In 1880, the exchange in New Havens, served around 30,000 customers using 138 exchanges [63] and further till 1887, it expanded its customer base up to 150,000 customers using its 743 main exchanges and 44 branch exchanges. While in 1881, a 45 miles (72 km), a long distance telephone service between Boston and Providence, Rode Island was put up in operation. Handling of such large number of calls posed great difficulty in managing exchanges.

8.1. Metallic Circuits for Improvements in Quality of Signals

Mean while, in the year 1881, American Bell Telephone company purchased Western Electrical Manufacturing company and it become Bell's manufacturing arm. It was further renamed to Western Electric Co. During 1881 itself, John Joseph Carty [96], an electrical engineer working in Bell's manufacturing unit called Western Electric Co., demonstrated advantages of two wire metallic circuit over conventional single wire grounded telephone system in terms of quality of audio signal as it rejects common mode unwanted signals. This type of service was then introduced commercially in October 1881 and first metallic circuit multiple switch board was put in commercial service. The metallic circuit configuration is shown in Fig. 26.

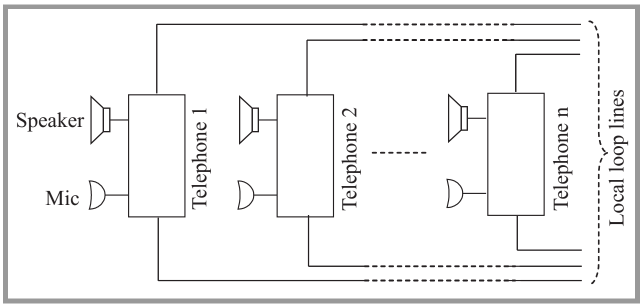


Fig. 26. Metallic circuit configurations (2 wire transmission).

There was a use of local battery in metallic circuit configurations and use of such local batteries was limited to only customer ends. These types of configurations had many drawbacks from the point of maintenance and continuous functioning of the telephone network.

Metallic circuits have become significant and iconic landmark in for circuit switching technologies and inventors started searching for the ways of metallic circuit switching in the following years. On January 17, 1882, Leroy B. Firman received the first patent [97] for a telephone switch board (Fig. 27).

This exchange is used to switch a pair of wires, for example, jumper/patchcord J_1 is used to connect B and D wires while

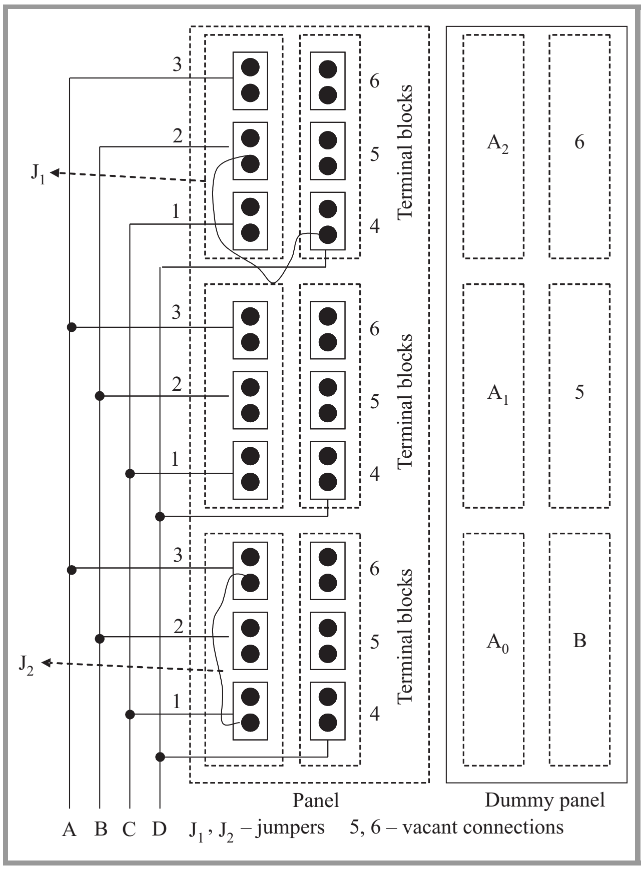


Fig. 27. Metallic circuit switching.

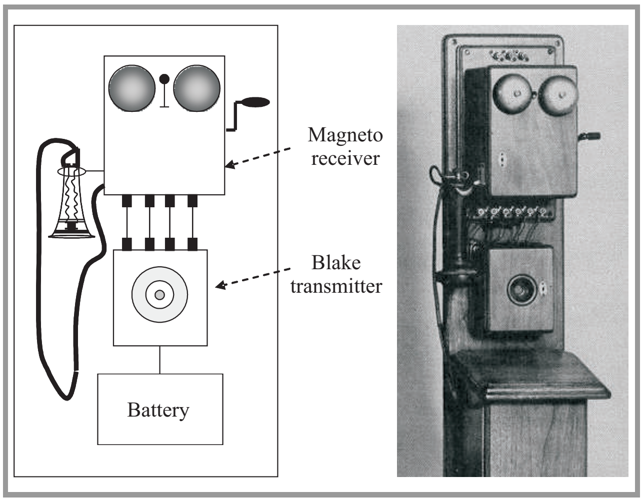


Fig. 28. Magneto wall set and its photograph.

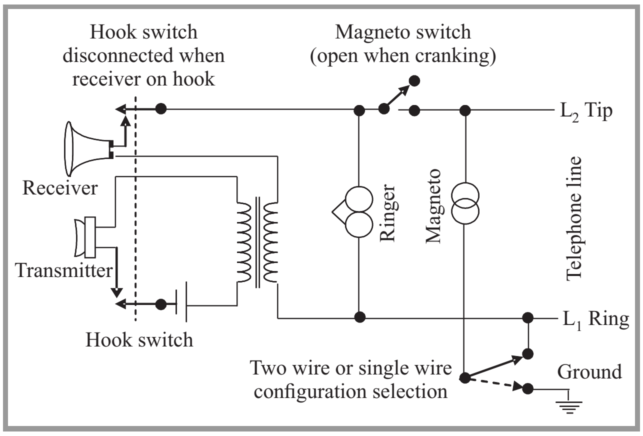


Fig. 29. Circuit for wall magneto telephone.

jumper/patchcord J_2 is used to connect A and C wires. The dummy terminal blocks are used to indicate vacant places on dummy board.

During 1884, a copper based overhead transmission link covering a distance of 235 miles (378 km) was established between New York and Boston. Twisted wires helped to minimize disturbances on the lines and allowed to maintain the balance on the line.

8.2. Commercial Use of Blake Transmitter

From 1880, Vail started buying Western Electric instruments and during November 1881, virtually market was controlled by Bell. After February 26, 1882, Western Electric gave up its remaining patent rights. During this time the telephone instrument of Fig. 28 was used by Bell which was an integration of Bell receiver and Blake transmitter on a wooden panel using wooden cases to form magneto wall unit [98], [99] made supplied by Western Electric to Bell System.

Circuit diagram of wall mounted magneto telephone unit is shown in Fig. 29 [100], [101].

The circuit consists of receiver, transmitter, transformer, magneto, ringer magneto switch and hook switch. While calling caller opens magneto switch to disconnect receiver and transmitter circuit and signals the operator by cranking the magneto (when cranked, magneto generates 80–100 V AC at 20 Hz). After signaling the operator, caller closes the magneto switch. Operator, after locating and connecting the party to the caller's line, signals the caller by activating the ringer. After hearing the ring, caller lifts the receiver from the hook switch to talk to the concerned party.

The magneto can be connected between tip and ring or tip and ground depending on number of lines used for carrying the signals to the operator. Normally two wire loop consisting of tip and ring is used, which is also called as metallic circuit. In such case magneto is connected between tip and ring, while in order to maintain the compatibility with earlier single line transmission, magneto is connected between tip and ground.

8.3. Introduction of phantom circuits

Increase in telephone customer base was resulted in crowding of skylines of every city with telephone wires, and it was high time for reduction in crowding of telephone lines. To some extent, phantom circuits were able to serve this purpose.

During 1882, Frank Jacob proposed to connect three telephones circuits using two pairs of telephone lines. This allows to connect three telephone circuits using 4 wires in-

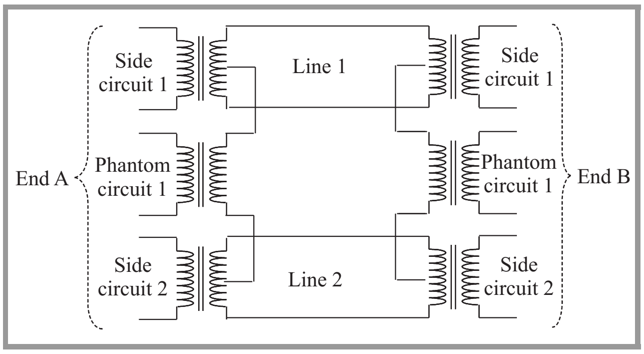


Fig. 30. Phantom circuit.

stead of 6 wires saving a pairs of wires which otherwise was required. It was possible by combining two simplex circuits. The additional circuit derived is called as phantom circuit (Fig. 30). In this arrangement, third circuit is derived by center tapping repeat coils at both the ends of two telephones lines. The phantom circuits were first used by John J. Carty in 1886 [102].

8.4. Formation of AT&T

In 1879 the Bell Telephone Company was renamed as National Bell Telephone Company [17]. It became American Bell Telephone Company in 1880. In the same year it started Long Distance Telephone Line (LDTL) project to create nationwide telephone network, where all most all wires were running above the ground before the year 1900 [103]. It required separate company to be established for such stupendous task. Hence, during 1885, BTC launched its new subsidiary called American Telephone and Telegraph (AT&T) which completes its first line, between New York and Philadelphia. Acquisition of Western Electric in 1891, paved a way for subsequent formation of Bell Laboratories. In 1892, AT&T established long distance telephone network from New York to Chicago.

8.5. Need of Common Battery Telephone System

Till now, the batteries were used locally, caused lot of irregularity in maintaining batteries at local ends and hence it was felt that batteries at some fixed place owned by telephone company can reduce the down time of the network. This will also facilitate the future signaling mechanisms. In wake of these developments, common battery system was developed by Hammond V. Hayes in 1888, in which, customers are connected locally to central office through local lines supplying current to the customers without use of any power source at the customer premises.

8.6. Telephone Numbering

Prior to 1889, subscribers were designated by the names, which acted as unique identification of the telephone subscribers. It was then thought that instead of using wide ranging A to Z characters, one can use 0 to 9 digits. The main advantage is hierarchical and modular periodicities that can very well address the issue of identification of subscriber. During 1889, telephone subscribers began to be designated by numbers rather than names. The basic purpose of designating subscriber by number is to uniquely identify the subscriber. The telephone number is a string of decimal digits that uniquely identifies network termination point. The telephone number contains information necessary to route the information to that particular telephone number. It is a uniform resource identifier (URI) for a particular telephone related to a particular subscriber [104].

8.7. Extending Telephone Lines Using Loading Coils

During 1889 to 1899 many telephones lines went in to service and many cities were connected. The cities such as New York-Chicago (1892), Boston-Chicago (1893), New York-Cincinnati (1893), Chicago-Nashville(1895), Kansas City-Omaha (1896), New York-St. Louis (1896), New York-Charleston (1897), New York-Minneapolis (1897), New York-Norfolk (1897) and New York-Kansas City (1898) got connected. The telephone wire running distances went up. This needed signal boosting and conditioning on the telephone lines called regeneration using line repeaters. In 1899, Mihajlo Idvorski Pupin [105]–[108] (also known as Michael I. Pupin), working at Columbia University invented loading coils that permitted construction of longer telephone lines (Fig. 31).

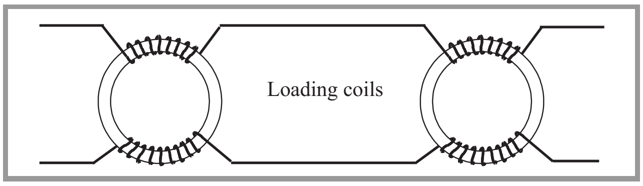


Fig. 31. Loading coils in telephone lines.

Loading coils are inserted at the distance of every 1.24 mile (2 km). Analog telephone line consists of a balanced pair of twisted wires with characteristic impedance terminations of 600 or 900 Ω. These terminations are engineered to match amplitude and frequency characteristics of voice signals carried by telephone lines.

First time, AT&T installed Puppin coils in 1906 on its 10.5 mile (17 kilometers) line between New York and Newark, New Jersey in the year 1902, while introduction of 250 mH loading coils [37] of at every 1.14 mile (1.85 km) interval on other cables between New York-New Havens and New York-Philadelphia followed. Before introduction of amplifiers, the longest loaded cable was between Boston and Washington through New York. In 1920, all loading coils were removed to give a way for use of amplifiers.

9. Strowger Switch – the Facilitator for Automatic Exchanges

Almon Strowger, an undertaker in Topeka, Kansas invents automatic telephone switch and on March 12, 1889, he filed an application for patent [109]–[111] which was granted on March 10, 1891. The Strowger switch (Fig. 32), was a device that led to the automation of the telephone circuit switching. It was the first automatic commercial dial system. In simple words, initial Strowger switch can be viewed as 1 pole 10 position switch that can connect one of the 10 destinations. To operate such a switch electrically using electromagnetic devices, it requires two relays for advancing the contacts depending on the destination telephone number to be contacted while other relay is used for resetting the contacts home position. This represents one pole ten position automated contact selector switch, also called as uniselector switch. Simple demonstration of such a switch is shown in Fig. 32.

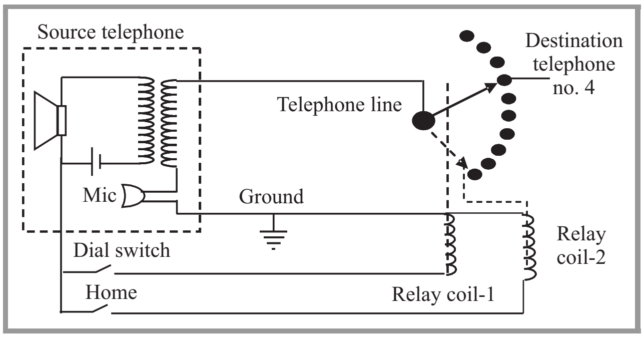


Fig. 32. Uniselector automatic Strowger switch.

To increase the capacity to more destinations, 1 pole 10 contacts wafers are stacked together [112]. If 10 wafers are stacked, one can increase capacity to 100 destinations, but instead of 10 poles, the pole contact is moved vertically up or down to select destination is called as group selector switch. Each wafer can be viewed as uniselector in which contacts are moved laterally/rotated in one plane. The switch consisting of 10 such wafers involving vertical and lateral movements is called as biselector switch (Fig. 33).

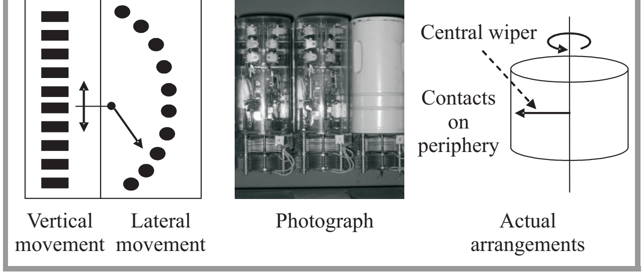


Fig. 33. Biselective Strowger switch.

The biselector switch was housed in a metal can for purpose of use in actual commercial exchanges. It was operated using three electromagnetic coils, first coil is used to step relay upwards to next deck of contacts, and second coil is used to rotate the contacts in a deck and third coil is used for resting the contacts to home position in which lateral contacts rotate to extremely left position in the decks and vertical contact falls down to a bottom position. The following detailed figure (Fig. 34) may clear all doubts that reader may have in his mind.

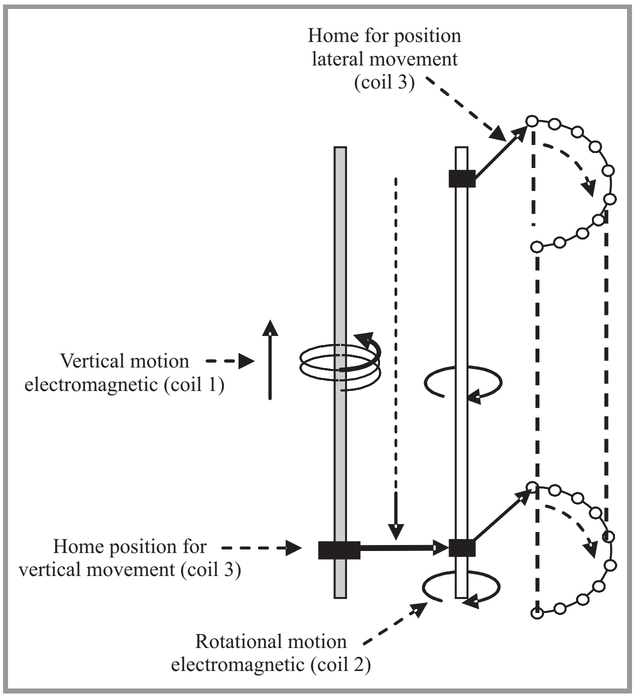


Fig. 34. Biselector switch functioning details.

Strowger was the first person who really invented automatic means of communication between two telephone users.

9.1. First Commercial Exchange Using Strowger Switch

Strowger Automatic Telephone Exchange Company (SATEC), Chicago was formed in October 30, 1891. Early of 1892, A. E. Keith joined the SATEC and installed and opened the world's first automatic commercial exchange [110], [113] in then home town of Strowger, La Porte, Indiana on November 3, 1892, with about 75 subscribers and capacity for 99, which indicates that using single biselective switch one can control 99 connections using two push button switches. First push button caused the switch to move its position up and down to select a wafer in a particular row dictated by a 10th place digit of telephone number while using second push button, in a selected wafer, a particular contact is selected based on a unit place digit of telephone number. Thus, if we want to connect telephone number 63, first switch is pushed 6 times to select 6th wafer in 6th row while other switch is pushed three times to move the contact at 3rd place. The third push button switch is used for disconnection.

9.2. Management of First Commercial Strowger Switch Exchange

Although, first installed commercial exchange was having capacity of 100, the patent filed by Strowger, explains the functioning of Strowger system having capacity of 1000 using four push buttons switches [110]. First three push buttons were used for selection of 3 digit telephone number and fourth switch is used for disconnecting the call after conversation is over. However, first commercial Strowger exchange having a capacity of 100 lines was implemented using biselector switch (Fig. 35), in which biselector switch moved in both vertical and horizontal direction to reach any one of 100 lines those are arranged in 10 groups having 10 lines each.

First set of dialed pulses (represented by first digit of destination telephone number) move the contact upwards in vertical direction to select a particular wafer and within that particular wafer, contacts in the selected are continuously rotated horizontally to higher contact positions from home position till free trunk line is located. This arrangement now can be controlled using three push button switches having defined functions of controlling of vertical contacts, lateral contacts and returning to a home position the broad view of such exchange is shown in Fig. 36.

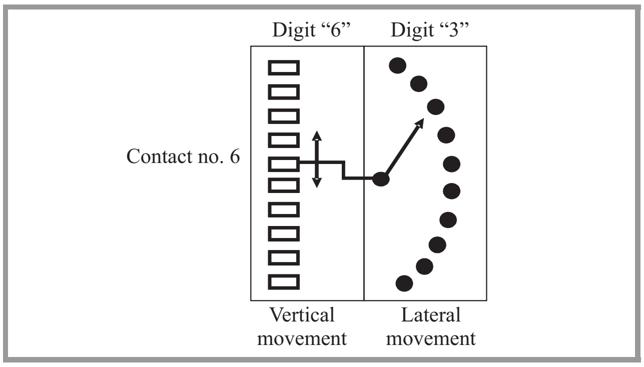


Fig. 35. Use of biselective Strowger switch for connecting destination telephone.

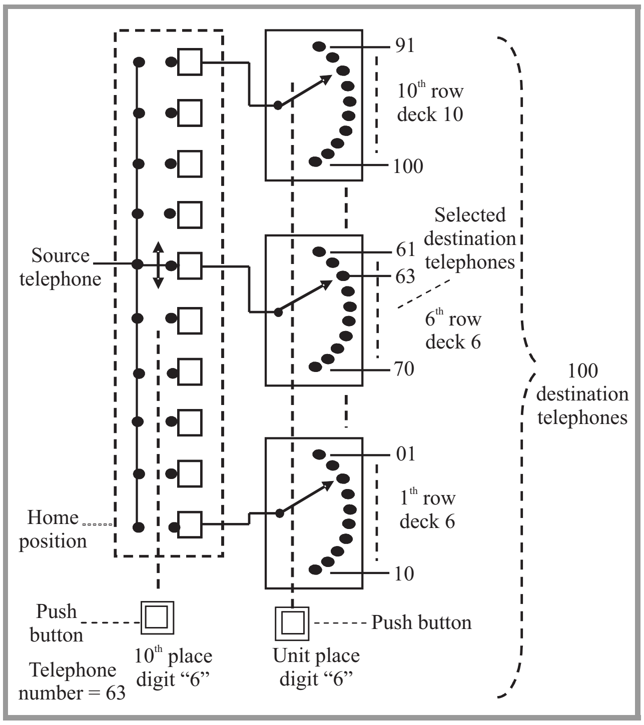


Fig. 36. First Strowger exchange principle.

9.3. Five Wire Telephone Interface

In this exchange, electromagnetic relays were operated using push button switches arranged at tens place and units place. This exchange used two axes Strowger switch, a relay having three electromagnetic coils, first coil is responsible to step-up through various wafers arranged vertically selecting a particular wafer acting as deck of 10 contacts, while second coil rotates contact in a selected wafer. Finally a third coil resets the relay at home position. To have such arrangements, it required 5 wires interface between telephone and exchange, leading to a complicated wiring between telephone office and customer premises. In the example shown in Fig. 37, source telephone is connected to destination telephone (telephone number 63).

10. Opening of Long Distance Line by AT&T

During 1892, AT&T reaches its initial goal, opening a long distance line connecting New York and Chicago covering a distance of 950 miles (1528 km).

Few long distance circuits [115] introduced thereafter were New York to Havana, New York to New Orleans, New York to St. Louis, Boston to Chicago, Chicago to Denver, Chicago to Los Angles, Chicago to San Francisco, Denver to San Francisco, Kansas City to Denver, Dallas to St. Louis.

Opening of this long distance line exposed the problems of long distance voice communication. An English physicist Heaviside developed the theory of transmission of signals over two wire metallic circuits and published a series of papers in 1892. Based on these papers, Michael I. Pupin and George A. Campbell realized that periodic use of inductive loading coils can reduce the loss of signals [107]. First application of loading coils occurred in 25 miles (40.2 km) Boston circuit during 1900. A line between New York and Denver with loading coils was opened by AT&T in 1909 [116], [117].

11. Introduction Common Battery Systems in Exchanges

The first common battery (CB) system for subscriber's telephone exchange was introduced in December, 1893 at Lexington, Massachusetts, before that subscribers required their own battery and hand operated magneto generator [37].

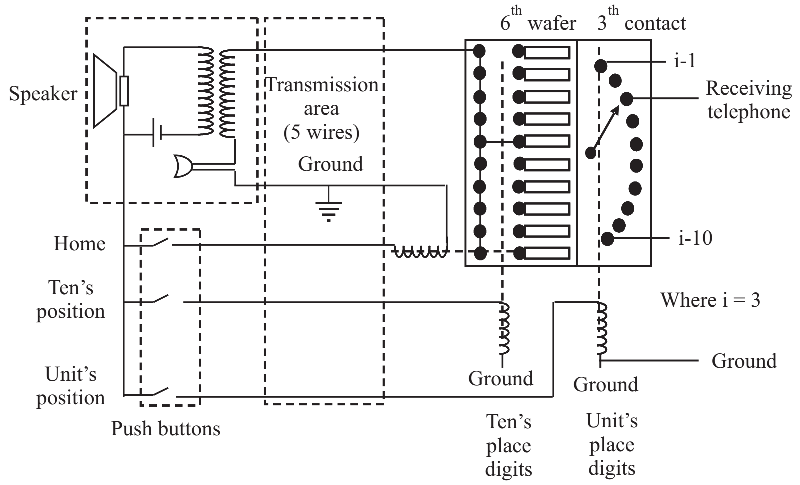


Fig. 37. Five wire telephone interface.

While, in Europe, common battery was introduced in the year 1900 at exchanges at Bristol, England and Adlershof near Berlin, Germany. Initial CB systems had the problems of crosstalk, which was solved by using bridged impedance system introduced by Western Electric. After that, there was great acceptability for CB systems. By 1901 there were 14 exchanges in United States operated based on CB systems [76]. In following years CB systems were installed at Hague, Netherlands in 1903, Kyoto exchange in Japan (1903) and exchange serving 60,000 subscribers at Moscow (1910). During 1905, first automatic exchange operated entirely on common battery system was opened at South Bend, Illinois.

12. Improvements for Successive Strowger Exchanges

The various mechanisms were designed to improve the performance of switches. The first design was piano wire. The design went in problem and inventors were forced back to follow the original design for improvements.

In 1893, Strowger switch was exhibited in a World Fair held at Chicago pushing this product for more and more commercial use [110]. At the same time, during 1893 itself, automatic connection release concept was introduced by A. E. Keith. This led to a present day hook switch concept. During beginning of 1894, Frank A. Lunquist and Erickson brothers (John and Charles) joined the services of Strowger company to work with Keith to work for the improvements of Strowger switch that was capable of both a rotary and a longitudinal movement called as Piano wire design [110]. This type of design was installed at La Porte, Indiana in the year 1894, and at the same time, a simple concept of indication of busy signal was introduced. The free line was indicated by ringing the bells at both the ends, while line busy status is indicated by no bell signal. These installations also used automatic connection release invented by Keith. The piano wire design could not prove to be popular.

The earlier piano wire type of switches was replaced by this improved design in the exchange at La Porte, Indiana. It still required 5 lines for interconnecting telephone. During 1895, Keith and Erickson brothers filed a patent application on the basis of improved design based on Strowger's work, that resembles to modern step by step switch and patent was granted to them in 1899. These types of arrangements continued in exchanges till the 1900.

12.1. Rotary Dial Mechanism

Till 1896, in Strowger system, selection of the digits was done by a complicated system involving push button switches involving five wire telephone interfaces. Rotary dial was facilitator for automatic selection telephone line leading to destination telephone or listener's telephone. Many people worked toward success of automatic telephone exchanges. M. D. Connolly of Philadelphia, T. A. Connolly of Washington DC and T. J. of Pittsburgh worked jointly for automatic telephone exchanges for dial telephones and first telephone patent was awarded to then for their exchange. Many patents for dial telephone switching were issued between 1879 to 1900. However only Strowger patent from March 10, 1891 resulted in successful commercial automatic system [118].

Later on, the system of timed pulse (TP) dialing was invented by Keith and Erickson brothers using a rotary dial in 1895 [119]. The patent was filed by them on August 20, 1896 and was granted on January 11, 1898. This work replaced push buttons by a rotary dial having digits 1 to 9 and X impregnated on dial base. While holds on the finger plate aligned with digits is used for dialing that particular digit producing the pulses equal to that digit. Most modern rotary dials had holes instead of holds, and now we have push buttons representing 0–9 digits and many other functions. The rotating wheel was a simple make and break arrangement to produce the pulses related to a num-

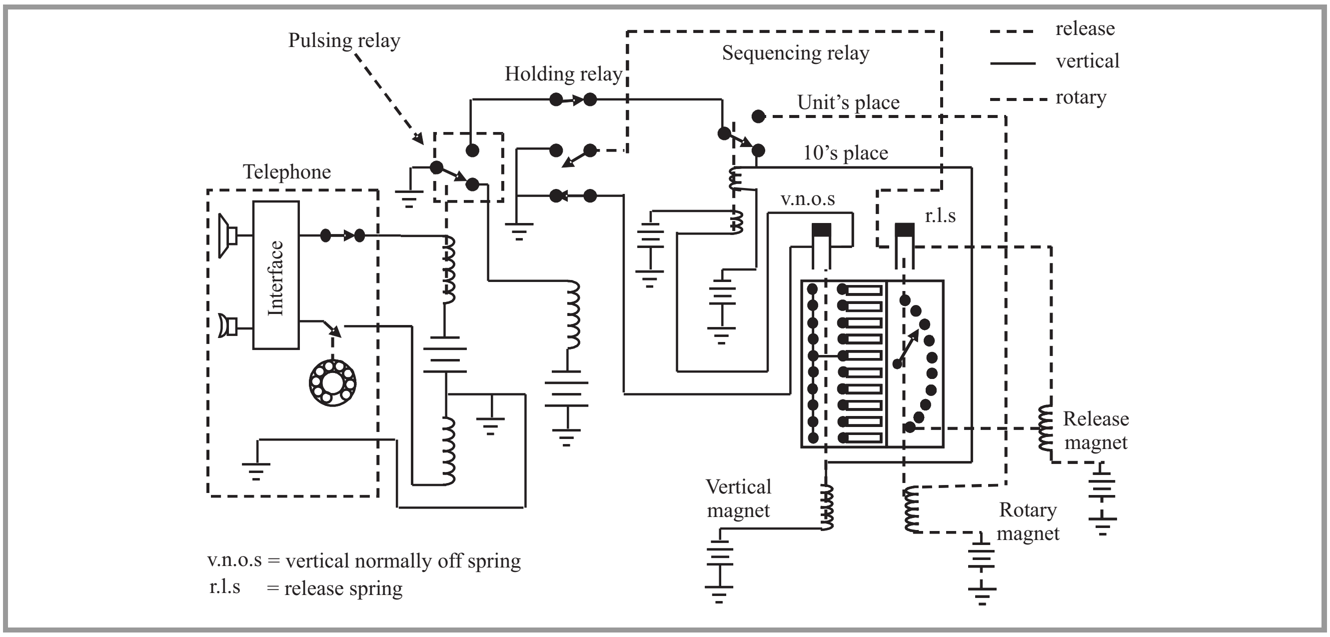


Fig. 38. Three-wire interface between rotary dial telephone and telephone office.

ber dialed. These pulses took over the control of operations at central office. The first rotary telephone dial used during 1896.

The pulses produces by rotary dial are transmitted over 2 wire interface [120], [121]. Telephone instrument, consists of microphone, speaker, hybrid converter interface, dial and hook switch. The hook switch represents two conditions called on-hook (idle condition) and off-hook (telephone is connected to telephone company office). When cradle is lifted, hook switch makes physical connection between telephone instrument and central office allowing central office battery to get connected to telephone instrument forming a current loop between CO to telephone instrument making telephone instrument ready for dialing required telephone. The telephone connection to telephone company's office is shown in Fig. 38.

When cradle is lifted, telephone instrument at customer side is powered by a battery at central office and telephone instrument at customer is activated to complete local loop leading to the customer. These happenings allows customer to dial the destination telephone number. When first digit is dialed (digit at 10's place), pulses produced by pulsing relay are carried forward through holding relay and sequencing relay for the purpose of moving vertical contacts of biselector switch. The contacts are moved till desired vertical position is attained. At this point, sequencing relay to shifts its position to unit's place allowing next digit to be dialed for activation of rotary magnet to select a desired contact. When contact is selected, conversation to selected phone takes place and when conversation is over caller releases the line, this also releases holding relay, release of holding relay operates release magnet, which intern rotates shaft to home position and wiper falls down to lowest position due to gravity.

13. 1000 Line Augusta Exchange

Also during 1896, Keith and Erickson brothers started working towards 1000 line telephone system based on call transfer concept or trunking (or selector trunking) for multiplying subscriber lines at each switch. Using trunking concepts, large number of users can be served using smaller number of communication paths. Various COs at various places are connected using trunk lines (pairs of copper wires). A selector trunking system of 1000 lines based

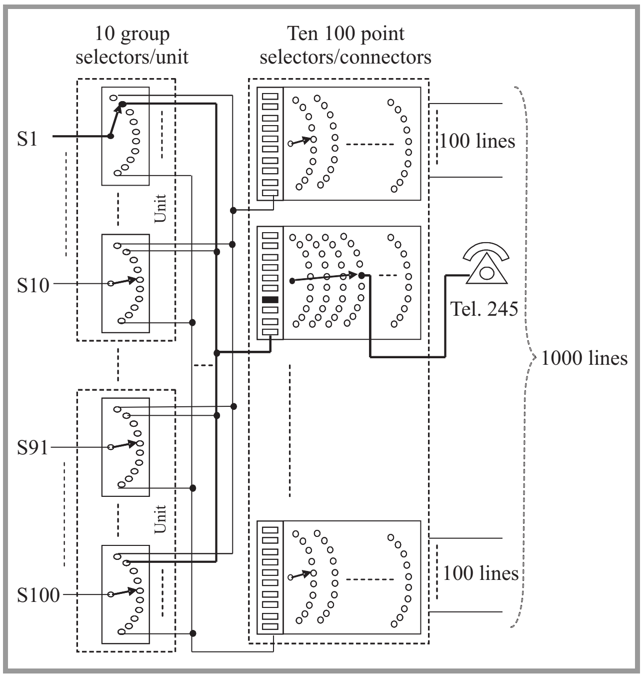


Fig. 39. Trunking scheme employed at Augusta, Georgia, USA, during 1897.

on dialing system developed by Keith and Erickson. The system of 400 lines was installed at Augusta, Georgia in 1897 [76]. Individual subscriber has his own group selector. Ten subscribers had ten group selectors. A primary stage (first selector) gave the access to a group of 100 subscribers while secondary stage (connectors) has 100 point selectors (concentration ratio of 1:10). The basic principles employed in exchange at Augusta are shown in Fig. 39.

In above mentioned scheme, internal blocking occurred when two subscribers in the same group, calling the subscriber belonging to same hundred group leading to impractical systems. Owing to these difficulties, Keith improved the design (Fig. 40) and first selectors near subscriber were modified to include 1 out of 10 search

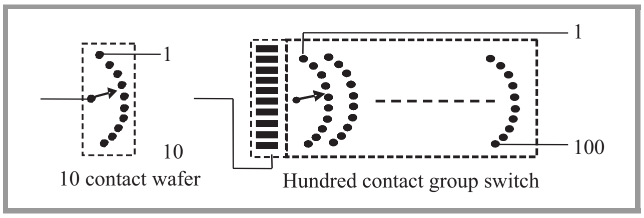


Fig. 40. Improvement in selector switches.

for finding idle line to approach to next selector that is connector [76]. With this modification, group selector switches no longer considered to be a set of ten but considered to be a set of 100. This arrangement drastically reduced number of selector switches required in the exchange.

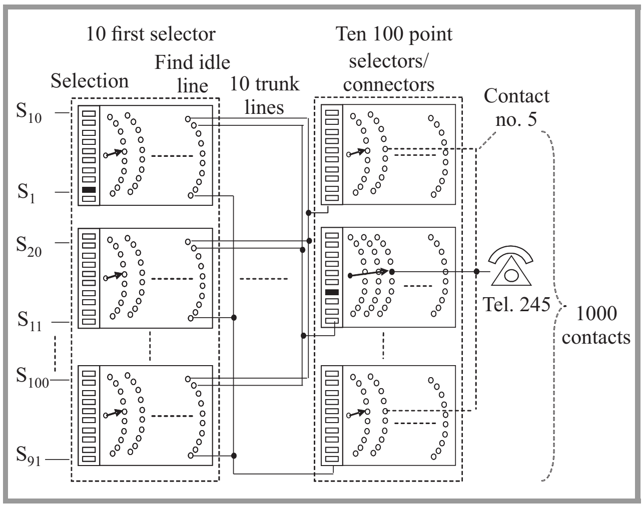


Fig. 41. Improved Strowger exchange at Augusta, Georgia, USA.

The modified Strowger exchange based on above modifications at Augusta is shown in Fig. 41.

14. Telephone Interface

In 1896, push buttons in Augusta exchange, were replaced by large rotary dial having three wire interface (2 wires and a ground), while large dial (which was invented in 1907) with two wire interface without ground was introduced at Pontiac, Illinois in 1908 [122]. In the same year, a small rotary dial using 2 wire interface without ground was introduced at Lansing, Michigan. These two wires were used for sending dialing pulses, supplying power to subscribers and transmission of speech signals. The basic principle of two wire interface was based on the bridge circuits consisting of four impedances connected within the arms of the square leading to four junction points A, B, C and D. The pairs of opposite ends (A, C) and (B, D) are used for connecting the source (mic) and receiver (speaker) respectively as shown in Fig. 42.

Ideal DC feeding circuit is required to ride voice signal on the DC current of higher magnitude, which is blocked at the voice signal receiver to separate the voice signal from the composite signal (DC and voice signal) serving the purpose of raising the energy of signal to be transmitted on the telephone lines. The ideal DC feeding circuit is well described in a series of European standards called EN 300001 v1.5.1 [123], and is shown in Fig. 43.

The feeder voltage values differ from country to country and range from minimum of 24 V (Czech Republic and Norway) to maximum of 104 V (France). However in most of the countries feeder voltage is 48 V.

The tolerable range of variation of values of feeder current, inductors, resistances and capacitances are as follows. Feeder voltage variation ΔV_f: 15 to 64 V while in loop condition $\Delta V_f \leq 64$ V. Feeder current variation ΔI_f: 1 to 60 mA, while loop condition ΔI_f = 1 to 60 mA while loop feeder resistance values ΔR_f: depends on V_f and I_f values (300 to 5000 Ω). The feeder components such as C_f and L_f affect ac signal performance and must meet all requirements of signal of interest. Usually $C_f \geq 20$ μF and $L_f \geq 5$ H.

The popular feeder voltages are 48 and 60 V. The feeder voltage of 48 V (–10% , +20%) requires bridge resistance of 800 Ω while for feeder voltage of 60 V (10%) requires bridge resistance of 1000 Ω.

15. 1000 Trunk Line Automatic Strowger Exchange

During 1897, engineers at Strowger Automatic Telephone Exchange Company started working on problem of 1000 trunk line automatic exchange in which selector have several sets of contacts in each row, leading to several connectors that can serve a group of 100 lines. In this project the idle trunk line was automatically selected as subscriber goes on dialing and patent was awarded to Frank A. Lundquist and first practical automatic exchange was put in service on November 1900 at New Belford, Mass, having its 4000 subscribers. In 1901, name of Strowger Automatic Telephone Exchange

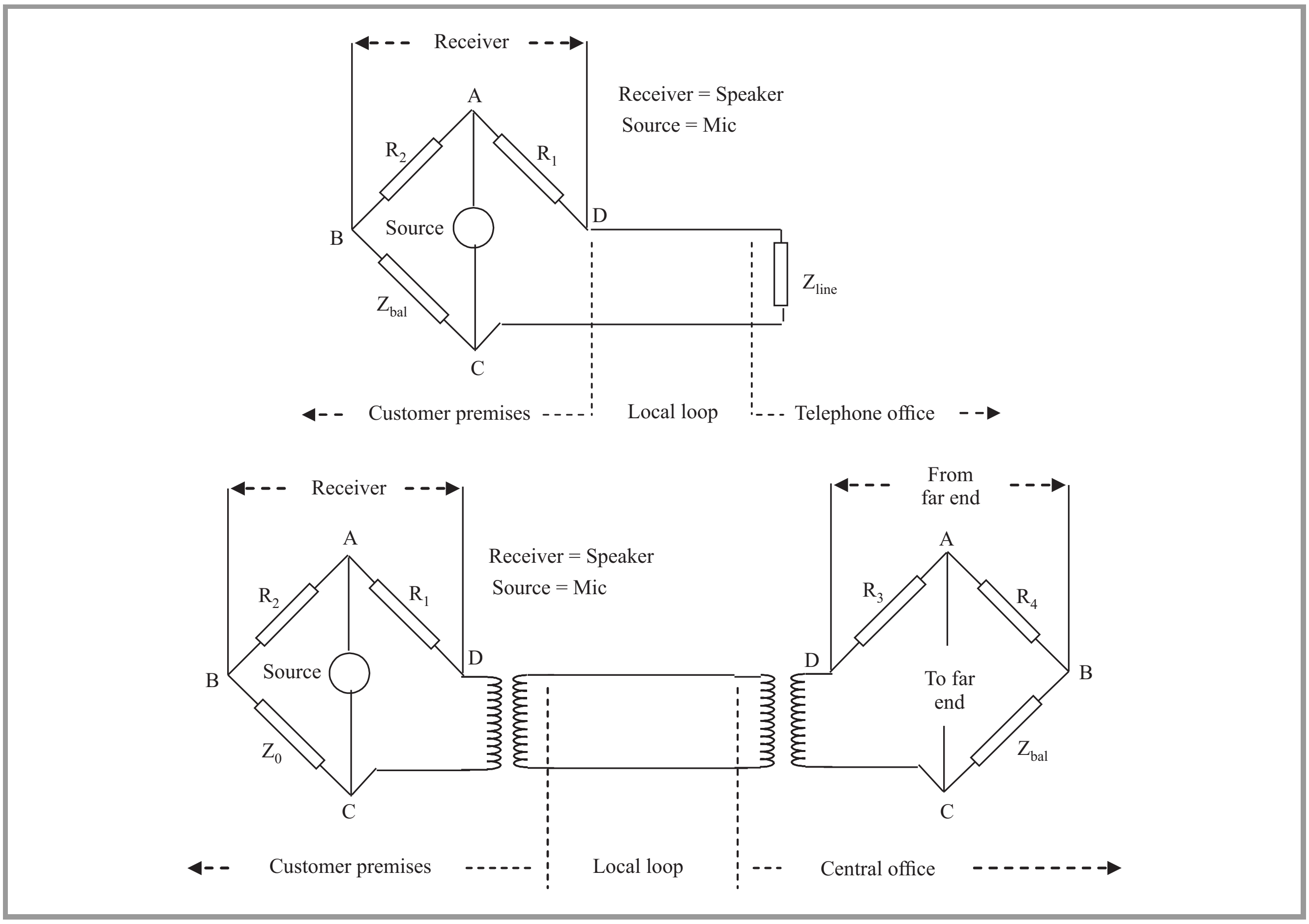


Fig. 42. Basic bridge configurations used in 2 wire local loop.

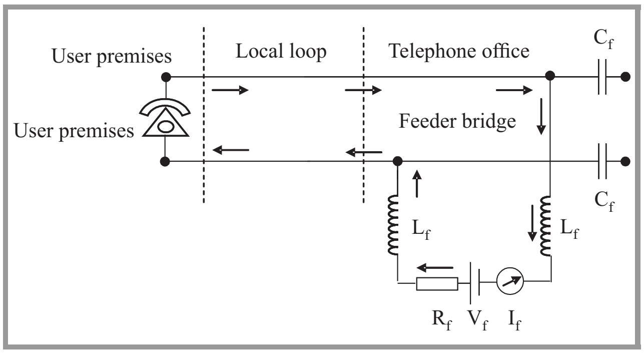


Fig. 43. Ideal DC feeding bridge.

Company was changed to Automatic Electric Company (Autelco) to completely exploit the Strowger system under the Keith technical leadership. In new company, engineering staff consisted A. E. Keith, T. G. Martin, J. Erickson, C. J. Erickson and E. C. Dickenson. The similar exchange as that of earlier, having capacity of 10,000 lines was installed at Chicago in 1903 [37], [76]. During following years, to cater number of customer/subscriber lines and to reduce number of group selectors, a principle of preselection was introduced at exchanges installed at Delaware.

15.1. Ten Thousand Line Automatic Strowger Exchanges

In automatic exchanges, the three configurations of Strowger relays (Fig. 44) called preselector switches for

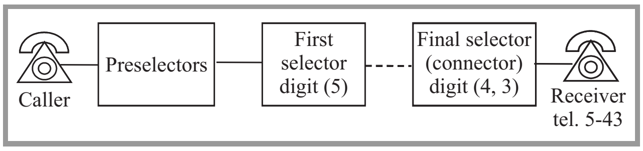


Fig. 44. General principle of routing in step-by-step automatic Strowger exchange.

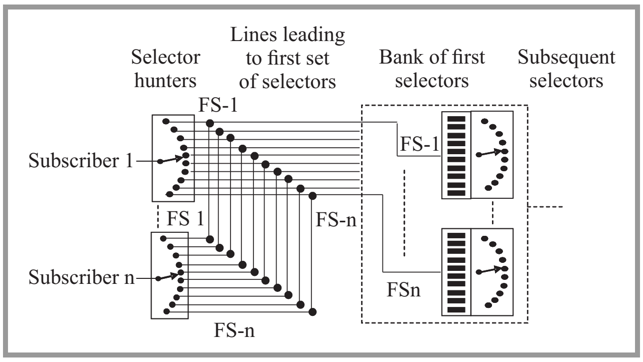


Fig. 45. Hunting scheme using uniselector switches.

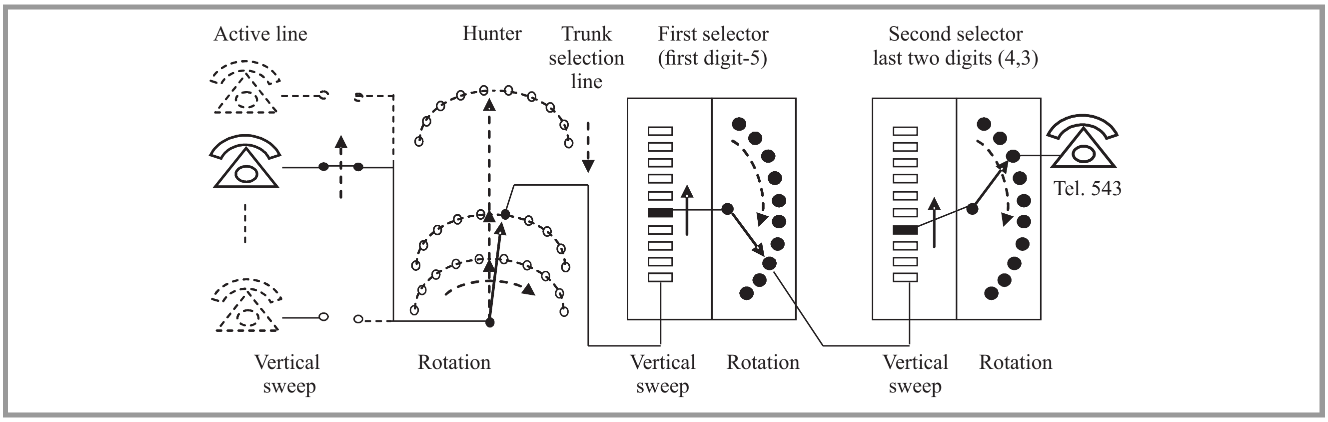


Fig. 46. Hunting scheme using biselector switch.

connecting to subscribers, two biselector switches for step by step routing, and connector switches for connecting destination telephones were used.

The preselectors are used in the form of two configurations called as hunters or line finders. The hunters or line finders are front relay selectors used for connecting the caller's end, while there can be number of selectors depending on number of digits assigned to a telephone number in the system and finally using connector relay, receiving telephone is connected. For example, is system has three digit telephone numbers and if destination telephone number is 5643 is to be contacted, the hunter relay continuously sweeps vertically across all the incoming lines till active line (off-hook) is located, dial tone tells subscriber that he may start dialing and also relay rotates repeatedly till free trunk line is found. After free trunk line is found, this free trunk line connects next relay (selector) that handles the first digit 5 of telephone number. At this point, when first digit 5 is dialed, contact is moved to 5th deck of contacts, when relay rotates to find next switch in chain. When next digit of destination telephone number 6 is dialed, contact is moved to 6th deck of contacts when relay rotates to find next switch in chain, which acts a connector which uses last two digits (4 and 3) of telephone number for connecting destination telephone. Digit 4 points to a 4th deck of contacts and rotational motion selects 3rd contact to establish the contact between source and destination telephones.

15.2. Preselector Schemes

Caller side interface is implemented using two types of approaches called hunter and line finder schemes [112], [124], [125].

In hunter based configuration, path search is initiated by the active subscriber line or a caller has an access to 23 or 24 first selectors through his own hunter selector. Hunter based scheme using uniselector switches is shown in Fig. 45.

In this case, each subscriber is connected to one uniselector switch representing a configuration such that number of uniselectors is equal to number of subscribers. In this scheme, when subscriber lifts his telephone for calling, hunter gets activated, and wiper steps continuously till first selector (FS) group can approached. The first selector switch then sends a dial tone to the active subscriber through selector hunter indicating that subscriber can call the party of his interest.

The another scheme can be implemented using single biselector switch for a group of 10 subscribers, in which active subscriber is located vertical motion of switch, while first group selector switches are approached using continuous rotational movement of selected wiper. The example of hunter scheme using biselector is shown in Fig. 46.

The main difference in hunting schemes and line finding is that in line finding scheme, first selector has direct approach to line finder, while in hunting scheme active subscriber line is responsible to locate idle first selector.

In line finder, all the subscriber lines are terminated to all the line finder connections, indicating that every subscriber is approachable through every line finder switch, which is evident from line finder scheme is shown in Fig. 47. A line from every finder is extended to its corresponding switch in a bank of first selector switches.

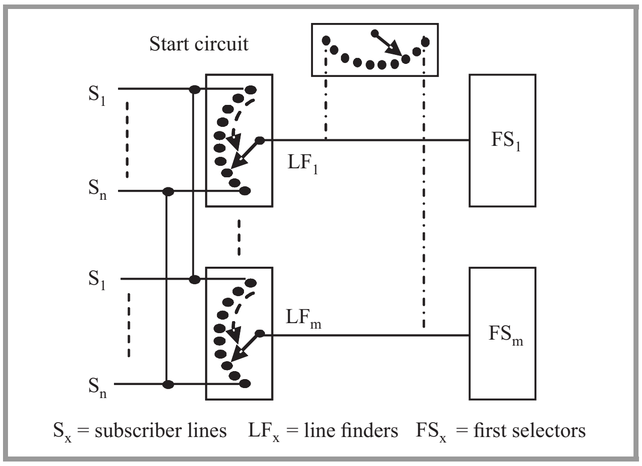


Fig. 47. Line finder scheme.

Line finder mechanism continuously scans the incoming subscriber lines to see that there is any active line approach-ing from the subscribers and once active line is located by the line finder, the existing connection between line finder and first selector is stated using start circuit. In short, in hunting scheme, path to first selector is located and connected, while line finding schemes existing path to first selector is activated once active line is located.

15.3. Ten Thousand Line Automatic Strowger Exchange Scheme

10,000 line automatic exchange schemes using above concepts were implemented, using preselectors consisting of line finders, hunters, selector switches and connectors [37], [112], [125], [126]. The scheme is shown in Fig. 48.

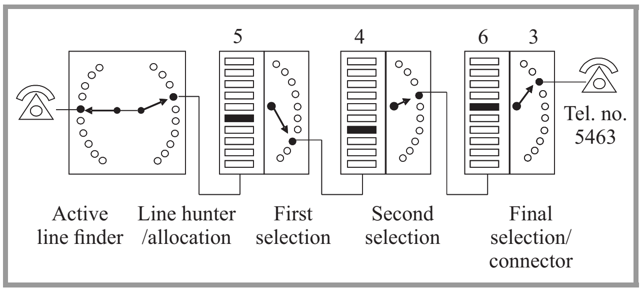


Fig. 48. 10,000 line automatic exchange configuration.

In following years, many important automatic systems were installed at Lincoln, Dayton, Columbus, Tampa, Grand Rapids, St Paul, Sioux, Jackson Ville, Buffalo, Spokane, Portland, Omaha, Edmonton, Regina in the USA, and Saskatoon in Canada, Havana, Cuba and most of the Australia.

16. After the Expiration of Bell's Patent in 1894

In 1894, Bell's patent on telephone expired, but due to valuable services rendered by company, continued to be in focus. Till 17 years of patent protection, Bell's and its associated companies enjoyed monopoly over the telephone services. After this point of expiration of Bell's patent, various independent (non-Bell companies) were free to offer telephone services without paying any license fees to Bell. Just year after expiration (1895), 87 independent companies entered in the telephone business [127]. In 1902, throughout US, 40,000 companies were in the telephone business. In 1907, the business of share independents grew to 51% while AT&T share was dropped to 49% from nearly 94% in 1894.

17. Integrated Transmitter and Receiver Handset

During 1904, Bell developed a handset that integrated both transmitter and receiver for human convenience. However it was made available to subscriber during 1927 due to its cost and desk set were used till 1927. The phones having integrated transmitter and receiver were called French phone.

18. Establishment of Bell Telephone Laboratories

In 1907, Vail N. Theodore was again asked to return as president of AT&T. During his term, he developed guiding principles for the company in terms of philosophy, strategy and organizational structure. During 1908 he took nationwide advertising campaign with a slogan "One system, one policy, universal service". AT&T being a parent company of Bell System put an end to harassment to independent telephone companies. It also perused the policy of standardization. The terms of collaboration made friendly for attracting independent telephone companies. During 1913, AT&T settles first federal antitrust suit called Kingsbury (AT&T vice president was Nathan Kingsbury) commitment, establishing AT&T as Government approved monopoly. It also allowed non competing independent telephone companies to connect AT&T network and no more taking over of independent telephone companies. In 1915, AT&T's long distance service reached San Francisco. In June 1919, Vail N. Theodore retired as president and became the chairman of board. During 1925 AT&T establishes Bell Telephone Laboratories as its research and development subsidiary, and its systematic contributions and policies were awarded in terms rise in revenue share to 79% in the year 1934.

19. An Era of Automatic Telephone Exchanges

After introduction of automatic telephone exchanges, in 1910, it was an era of automatic telephone exchanges and automatic telephone started superseding manual telephones [128] and US observed highest telephone penetration rate numbering 7 million subscribers as compared to Russia having 155,000 subscribers. This also made telephone numbering complicated due to highest growth leading to millions of subscribers. This demanded large number of digits representing telephone numbers. In 1930, New York introduced three literals and four numerals to represent telephone number (LLL-NNNN) and till mid 1960's London and Paris followed the same rule to represent telephone number.

The wide spread of telecommunication was led to united approach for solving the problems of telecommunication workers. During 1911, an international conference of postal, telegraph and telephone workers was held in Paris established federation of International Post, Telegraph and Telephone (IPTT). However, its formal launch was delayed until the 1920, when congress meet was held in Milan and it was registered in the form of its present form Posts, Telegraph and Telephones International (PTTI). The headquarter was located in Vienna and remained at Vienna till 1933. This secretariat was transferred to Bern, Brussels and finally to Geneva in 1969.

20. Introduction of Cables Services for Telephone Lines

Early long distance telephone lines were open wires, use of cables started when loaded cables became available [129]. In 1902, a telephone using loaded cable was laid between New York to Newark covering the distance of 11 miles (17.7 km), while in 1906 loaded cable from New York to New Heaven (79 miles, 127 km) and New York to Philadelphia (89 miles, 143 km) were installed. The loads of these lines were in the form of coils of 250 mH and wires were of 14 AWG diameter (2 mm^2).

21. Underground Cable Services and Vacuum Tube Amplifiers

The large numbers of telephone lines on the poles were posing lot of hurdles for civic administrations forcing them to make local ordinances in many town and cities called for elimination of such lines from main streets, forcing a way for placing cables underground [130]. In the winter of 1910 work moved for placing underground telephone cables between Washington and Boston. First long distance underground cable service between Boston to Washington via New York was started in the year 1914. To make up for signal loss, use of loading coils was evident. The telephone engineers extended long distance service westward from the Atlantic seaboard to Denver by 1911 and to Salt Lake City by 1913. More effective means, better than loading coils were invented by Lee De Forest in the form of 3 element "vacuum tube" or "Audion" (October 30, 1912), which could amplify the signals for long distance transmission to make up for signal losses. The use of such amplifiers periodically along the telephone lines were known as repeaters. AT&T brought the patent rights from De Forest. During 1913, AT&T tested the technology of vacuum tube amplifiers on a long distance network. On June 27, 1914 the company completed the line from Denver to Wendover, Utah and also on January 25, 1915, AT&T opened a first coast to coast transcontinental telephone line from San Francisco to New York (3500 miles, 5632 km) and it was the first transcontinental line using such amplifiers [131].

22. Transatlantic Telephone Services

During 1840–1850, for telegraphy, marine cables with gutta-percha insulation were laid in rivers and harbors [132]. In 1851, a underwater electric cable telegraph link was laid between England and France, and later on European countries, USA and Canada were connected together. The first transatlantic cable was laid in 1858 between Ireland and Newfoundland, which was hardly operated for a 26 days and failed after that [132], [133]. The failed transatlantic cable was then replaced by new transatlantic cable in 1866. Yet another underwater cable service went operation in 1884 from San Francisco to Oakland. The transatlantic cable service remained problematic and hence, the later part of long distance communication was dominated by radio transmission.

During 1901, wireless link between Cornwall and Newfoundland covering a distance of 2112 miles (3400 km) was established [134]. Similarly during 1904 and 1908 transmission of photograph between Nuremberg and Munich as well as radio telephone link between Paris and Brittany (west side area of France) respectively.

During 1915, Bell System (AT&T being parent company of Bell system) engineers were able to achieve voice transmission across the Atlantic connecting Virginia and Paris. A year later, two way communication was held using ship at sea. During 1926, there was first successful transatlantic two way communication and finally a commercial service using two radio stations was started between New York and London on January 1927. The year 1935, was a year of celebration for AT&T marked by a successful "Around the globe" communication by president and vice presidents of AT&T named W. S. Gifford and T. Miller respectively.

The global communication using submarine cable began on September, 1956 when first transatlantic under sea telephone system (TAT-1) from Clarenville, Newfoundland to Oban, Scotland went in to service [135]. It was the result of effort put forward by AT&T Bell Laboratories and British post office that provided reliable service with relatively fragile components in hostile environment. The two cables were used for each direction stretching over the distance of 1950 nautical miles and each cable had fifty one repeaters providing a 65 dB gain with 144 kHz bandwidth. TAT-1 provided 29 telephone circuits between New York and London. This system was operational till 1978.

23. Development of Condenser Microphone

Meanwhile, during 1916, E. C. Wente at Bell Labs developed the condenser microphone [136]–[138]. This invention enabled use of vacuum tube amplifiers. The patent of condenser microphone was filed on December 20, 1916. The device was constructed using two parallel plates forming a capacitor whose value changes due to variation of distance between them as a result of sound pressure striking on one of the thin plate in the form of diaphragm. The diaphragm was constructed using steel having thickness of 0.002 inches (50 μm) which was placed near to another plate maintaining the distance of 0.001 inch (25.4 μm) containing an air gap between them. In idle condition, capacitor having its value as C_0 is changed to

$C_1 = C_0 \pm \triangle C$ due to dynamic pressure exerted by the sound waves. The change in capacitor value is responsible for generating a voltage proportional to C_1/C_0 converting sound vibrations to an electrical signal.

24. Bell System's First Automatic Exchange

Bell System did not install their first automatic exchange until 1919 in Norfolk, VA. The one feature which was provided with the first Bell System exchange in Norfolk was dial tone [122]. Dial tone was first introduced by Siemens in Germany in the year 1908 but it took nearly nine years to get it popularized. Bell System was first to make it commercially popular in it's above mentioned Norfolk exchange.

25. Crossbar Telephone Exchanges

Idea of crossbar exchange was conceived as early as 1901 by Homer J. Roberts, but it took 16 years to be practical [76], [140]–[142]. The pioneering efforts by John G. Roberts and John N. Reynolds (1913), engineer with Western Electric, USA was responsible for patenting the practical crossbar system in 1916 but was never exploited, while Gotthilf A. Betulander (engineer in Swedish administration called Televerket) and Nils G. Palmgren were responsible for designing first satisfactory cross bar mechanism. Betulander was responsible for separating selection circuits from connecting circuits. The selection circuit consisted of user dial and devices for finding the called subscriber line. The logic of separation is shown in Fig. 49.

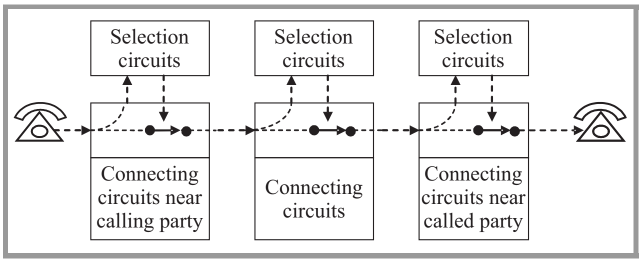


Fig. 49. Separation of selection circuits and connecting circuits (logic of Betulander).

That selection circuits are used only during the process of connection establishment and connection release. As digits related to destination telephone number are dialed one by one, the connections are successively established along the path leading to destination telephone and remain active only for short duration (selection time) till connection is established. Once connection is established, selection circuits have no role to play. The connected path is then used for voice communication between calling and called parties. The connecting circuits were in the form horizontal vertical lines in which cross points do represent a matrix of switches.

The basic principle of cross bar switch is shown in Fig. 50. To connect input line to output line, the switch at cross point is activated. Using such configurations any input line can be connected to any output line. Main advantage of cross bar switching is to eliminate sliding contacts in the earlier systems and was based entirely on the extensive use of relays. This design used one reed relay at each cross point leading to expensive design. This design was modified later on to use fewer relays. In modified design, single relays were used for each column and row, minimizing the number of relays in the design.

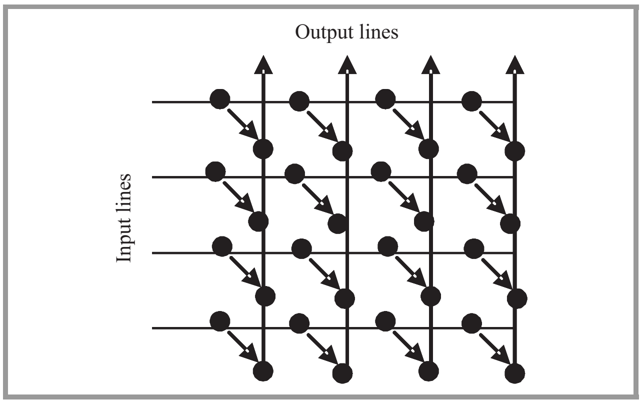


Fig. 50. Basic principle cross bar switch.

25.1. Foundations of Crossbar Switches

The basic foundation of crossbar switch was laid in a small company of Betulander called Nya Autotelefon Betulander. This company worked on registers controlled automatic telephone switches based on relays leading to a patent in the year 1912. The first 100 line subscriber demonstration exchange based on Betulander's all relay system was exhibited in Marconi House in London during 1913. Televerket kept rights for such a system in Sweden, while overseas right was transferred to an another English firm called The Relay Automatic, which installed exchanges at London (in 1916), Lancs (in 1922), India and France [143]. In 1926, these switches were used initially in rural exchange at Sundsvall. By 1944, around 1100 exchanges were installed in Sweden.

In 1919, due to order placed by Televerket to Betulander for supply of test station using such system, Betulander and Palmgren started working on new selector design based on crossbar switch using fewer relays than earlier design [142]. During 1919, Nya Autotelefon Betulander was acquired by Ericsson. The Ericsson started marketing of this switch. After sale of his company, Betulander resumed his work at Televerket. One year later, in 1920, Televerket selected 500 switches (sometimes referred as work horse) for automated equipment in Stockholm and Gothenburg. The first station with switch 500 was opened in 1923. The idea of 500 switching system was conceived by the Axel Hultman, director of Televerket, Swedish Post, telegraph and Telephones (PTT), Stockholm. The success of these Swedish switches received worldwide attention during 1920s. During 1926, the work of Betulander and Palmgren resulted in a country's first experimental exchange in Sweden using crossbar switching [144].

25.2. Crossbar Link Trunking

Using switch matrices of different size compression of expansion of number of lines can be for example, if switch matrix of 3×4 (3 inputs and 4 outputs) is used, 3 to 4 line expansion is achieved. The 9 line to 16 line expansion is shown in the Fig. 51.

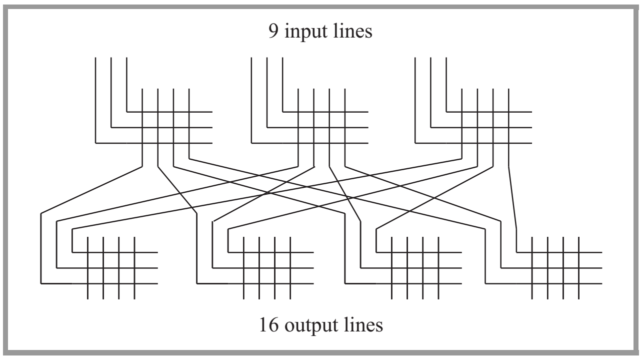


Fig. 51. Crossbar link trunking principle.

25.3. Spread of Crossbar Exchanges in United States

The Bell Labs Engineers visited Sweden to study the economics and traffic characteristics of these switches in the year 1930. Bell Labs ordered few switches for analysis and worked secretly for their own developments which led to the development of switch called Xbar #1 (developed by Western Electric). These developments by Bell Labs were not exposed till 1937 and finally Xbar #1 was first used in an installation at Troy Avenue central office, Brooklyn, NY in 1938, which was closely followed by East 30 street Manhattan. The crossbar switch designed by Bell Labs was similar to Swedish design in which Bell Labs used link connection principle with resister and mark control of Betulander and Palmgren. The faster developments took place during World War II and Xbar #1 was modified to more advanced Xbar #5 switch, which became key element in crossbar exchanges during 1950s and 1960s. The Xbar #5 switch was existed till 1969 after that its production was stopped. The developments of Xbar #5 was significant from the angle of handling DTMF tone dialing, which was introduced during early 1962. Xbar #1 and Xbar #5 were the switches to be used in local central offices. Xbar #1 switch was used in early crossbar systems while Xbar #5 switch was used in latter period (during 1962).

During early crossbar exchange era (late 1930s), it was realized that Xbar #1 was not sufficient to handle short haul (regional) and long haul (toll) calling. This resulted in a development of tandem crossbar switch (XBT) by Western Electric in 1941. XBTs were mainly used for regional or city wide tandom switch. In early day's context, XBT served the purpose of long haul switches. Further, during 1940s it was realized that even XBT was not enough to serve the purpose of long haul communication that gave a way for development of Xbar #4 switch which was installed at Philadelphia PA in 1943. During 1950s Xbar #4 switch was further modified (called Xbar #4A) to accommodate dynamic routing rather than hardwired routing and were in use from 1953 to 1968.

25.4. Modifications in Exchanges Due to Introduction of Crossbar Switches

The early telephone exchangers used the step by step (SXS) switching, in which, progressive connections were made as digits are dialed. Invention of crossbar switches allowed switching over from progressive control to common control. The common control systems are divided in to three categories such as trunk supervision, routing control and data storage control. In common control, the dialed digits are first stored in a register in the form of a complete telephone number which is then analyzed and acted upon by a marker. Marker is a hardware processor used in analysis of a call and once call is setup, register and markers are freed to handle other call setups [145]. The markers are treated as brain in the crossbar exchange systems. The crossbar matrix is controlled by common control. The control signals from transmission lines are detected and used for the managing the connections between input and output lines of crossbar matrix. Modern signaling systems emulate common control signaling principles.

25.5. Mark and Register Controlled Crossbar Exchange

Typical crossbar trunking exchange consists of line units at sender's and receiver's ends, group units, registers and markers as shown in Fig. 52.

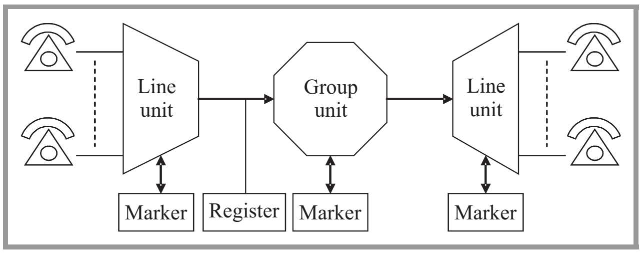


Fig. 52. Trunking crossbar exchange.

When telephone number is dialed, dialed digits are stored in register, and once complete number is dialed, marker analyzes the dialed number and connection is made though group unit to a line unit related to destination telephone. Markers at input and output line units analyze the input active lines and destination telephone lines for connection. As on today, there are three generations of switching exchanges called as step by step exchanges, cross bar exchanges (common control exchanges) and processor controlled exchanges or recent digital exchanges. The Strowger exchanges are the classic examples of step by step exchanges in which connections are made progressively and step by step to make connections to far end. In crossbar exchanges, the number of relays got increased drastically, which really required the miniaturization of these relays (reed relays) that allowed number of simultaneous connections. The crossbar exchanges being second generation, demanded more elaborative arrangements exchanges in the form of various classes of exchanges.

26. Introduction of Toll Switching Plans by AT&T

With the intension of creating hierarchical telecommunication network in USA, in 1929, AT&T implemented its general toll switching plan. As per this plan, AT&T set up a national network by connecting its eight regional centers. The regional centers were accessible to 2000 toll offices throughout the country through more than 140 primary centers (approximately one primary center in each state). The hierarchical arrangement of various switching centers is shown in Fig. 53. The additional direct connections were provided to various heavily loaded traffic centers for smooth functioning of total system.

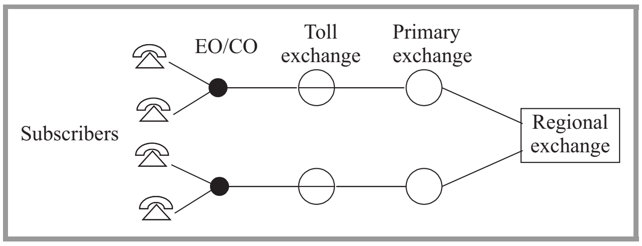


Fig. 53. AT&T's hierarchical switching plan.

27. Analog Voice Communication Before Digitization of Voice

The basic analog telephone interface circuit (Fig. 54) describes how analog telephone is interfaced to telephone line.

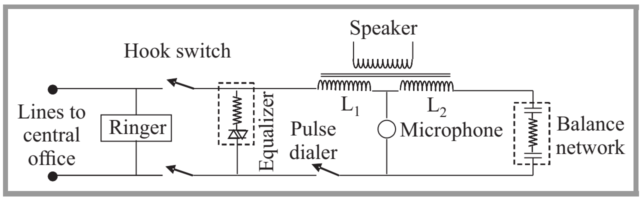


Fig. 54. Analog telephone interface.

The interface circuit shown is a simple bridge circuit that can be compared with a circuit shown in Fig. 42 where $R_1 \equiv L_1$, $R_2 \equiv L_2$, $Z \equiv$ balance network, $Z_{line} \equiv$ equalizer.

28. Logic of End to End Connection Management Through Various Exchanges

To have a telephone conversation, first connection is established and once connection is established, the users at both ends can talk, is called as connection oriented technology. Automatic exchanges allows establishment of connection between two end parties to have voice conversation. The telephone call consists of establishment of connection between two end parties, voice conversation, breaking of connection. Connection and disconnection between end parties is called as a signaling part, while actual voice service availed by user is a user service part. The automatic exchanges enable above mentioned services.
When hand set is lifted, power supply from central office gets connected to the phone and dial tone can be heard indicating phone is ready for dialing. Now the telephone number of other end could be dialed. If other telephone is busy then the busy tone is heard otherwise ring goes to that end indicating phone is connected. If handset of other end is lifted then a talk starts and after talk is over then handsets are put in on hook position.

28.1. Both Ends Are Idle

When telephones at both the ends are on-hook, indicates both ends are idle. This condition is expressed in the form of Fig. 55a.

28.2. Telephone at Caller End Lifted

When customer/caller decides to make call, he lifts from handset from cradle, closes the loop between the CO switch and the telephone at caller's end that allows current to flow in local loop at caller's end. CO switch detects this current flow and transmits a dial tone (350 and 440 Hz tones played continuously) to the telephone set of caller (Fig. 55b).

28.3. Dialing Phase

The dialing phase allows the caller to enter a telephone number at another location. This generates pulses or a touch-tone (push-button) phone that generates tones. Soon after first digit is dialed, dial tone is removed. These telephones use two different types of address signaling in order to notify the telephone company where a subscriber calls: dual tone multi frequency (DTMF) dialing and pulse dialing. These pulses or tones are transmitted to the CO switch across a two-wire twisted-pair cable (tip and ring lines), see Fig. 55c.

28.4. Switching Phase

In the switching phase, the CO switch translates the pulses or tones into a port address that connects to the telephone set of the called party. This connection could go directly

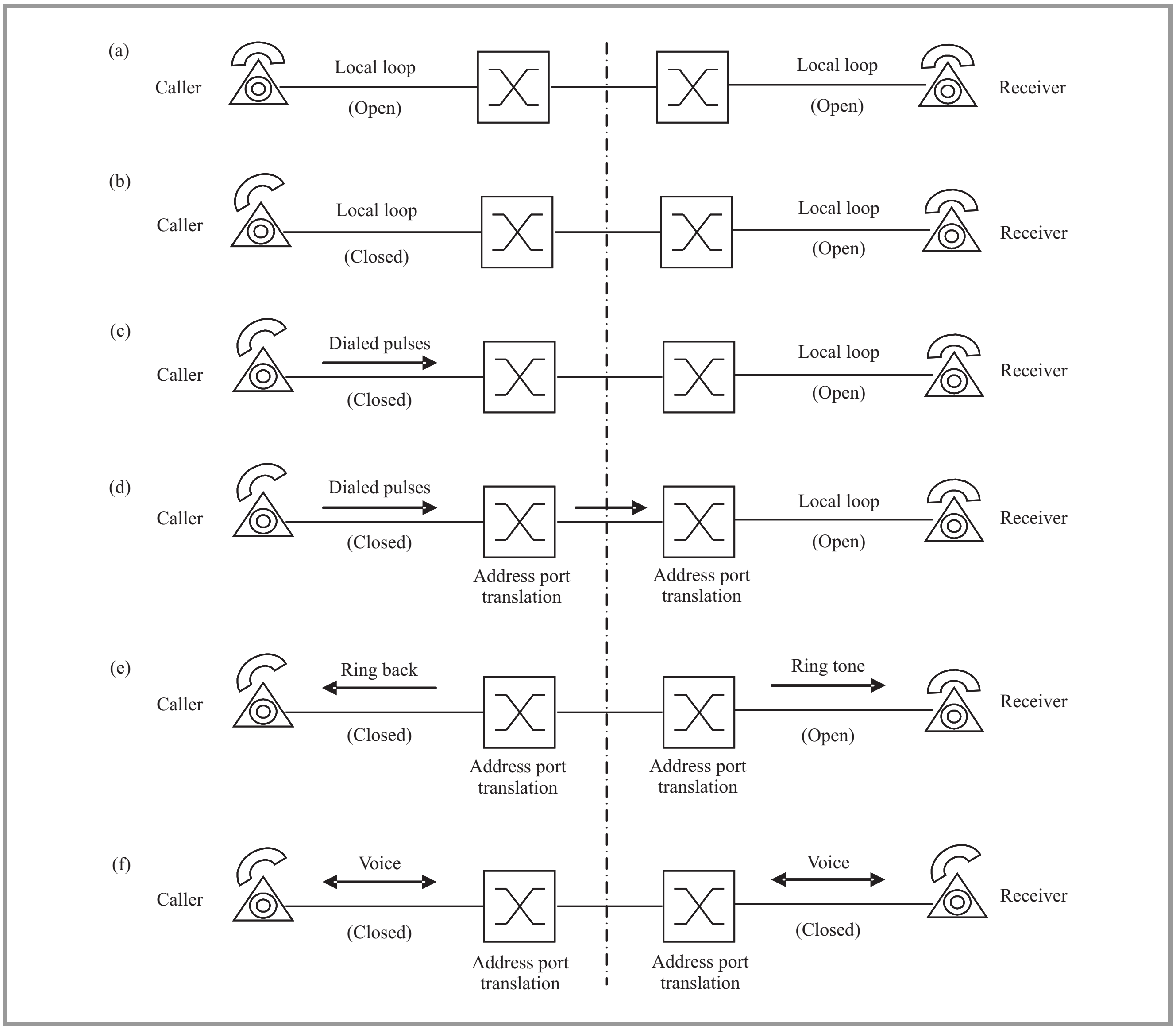


Fig. 55. Voice communication between caller and called party.

to the requested telephone set (for local calls) or go through another switch or several switches (for long distance calls) before it reach its final destination (Fig. 55d).

28.5. Ringing Phase

Once the CO switch connects to the called line, the switch sends a 20 Hz, 90 V signal to this line. This signal rings the phone of the called party. While ringing the phone of the called party (Fig. 55e), the CO switch sends an audible ring-back tone to the caller. This ring-back lets the caller know that ringing occurs at the called party. The CO switch transmits 440 and 480 Hz tones to the caller phone in order to generate a ring-back. These tones are played for a specific on time and off time. If the called party phone is busy, the CO switch sends a busy signal to the caller. This busy signal consists of 480 and 620 Hz tones. Figure 55f shows the talking phase.

28.6. Talking Phase

In the talking phase, the called party hears the phone ringing and decides to answer. As soon as the called party lifts the handset, an off-hook phase starts again, this time on the opposite end of the network. The local loop is closed on the called party side, so current starts to flow to the CO switch. This switch detects current flow and completes the voice connection back to the calling party phone. Now, voice communication can start between both ends of this connection (Fig. 55f).

Acknowledgements

Author wishes the acknowledge the management of Trinity College of Engineering and Research, Pune for encouragement in taking up such self styled work for the benefit of students to know how interesting are developments of antique telephone systems got with respect to historical timelines. I also acknowledge my colleagues of Trinity College of Engineering and Research, Pune, for very fruitful suggestions and discussions time to time. I also acknowledge the critical review by Dr. G. N. Acharya, former director, Central Electronics Engineering Research (CEERI), Pilani, India.

References

[1] F. M. Sorrentino and J. Krase, *The Review of Italian American Studies*. Lanham, Maryland, USA: Lexington Books, 2000.

[2] G. E. Schiavo, *Antonio Meucci Inventor of the Telephone*. New York: The Vigo Press, 1958, pp. 171–181.

[3] Andrea Sandre, *Italy@150 – Celebrating the Friendship Between Italy and the United States*, published by Embassy of Italy in Washington (DC), 2011.

[4] B. Catania, "Il governo degli stati uniti contro Alexander Graham Bell", Supplemento di *AEI Automazione Energia Informazione*, no. 10, 1999 (in Italian).

[5] B. Catania, "Un documento inedito rivaluta l'opera di Antonio Meucci", *AEI – Automazione, Energia, Informazione*, vol. 82, no. 2, 1995, pp. 32–40 (in Italian).

[6] B. Catania, "Antonio Meucci Revisited", *Antenna*, Newsletter of the Mercurians, in the Society for the History of Technology, vol. 9, no. 1, pp. 4–5, 1996.

[7] M. D. Fagen, *A History of Engineering and Science in the Bell System: The Early Years (1875–1925)*. New York: Harper and Row Publishers, 1975.

[8] F. Winwar, "Il Monumento di Staten Island: Meucci, Garibaldi e il telefono", *Poste e Telecomunicazioni*, Apr. 1958, p. 307 (in Italian).

[9] C. Bourseul, "Transmission electrique de la parole", *L'Illustration*, Aug. 26, 1854 (in French).

[10] R. Camboulives, "Un Occitan d'adoption: Charles Bourseul, inventeur du téléphone", communication faite à l'Académie des sciences, inscriptions et Belles-lettres de Toulouse dans sa séance du 8 novembre 1978 (in French).

[11] S. P. Thompson, *Philipp Reis: Inventor of the Telephone*. New York: Arno Press, 1974.

[12] F. I. Green, "Telephone", *The Bell System Tech. J.*, vol. 37, no. 2, 1958, pp. 289–338.

[13] F. William and M. D. Channing, "The telephone, with a sketch of its inventor, Philipp Reis", *Popular Science Monthly*, vol. 23, 1883, pp. 540–551.

[14] R. J. P. Weber and D. N. Perkins, (Eds.), *Inventive Minds: Creativity in Technology*. New York: Oxford University Press, 1992, pp. 48–79.

[15] "Fundamentals of Telephone Communication Systems", Western Electric Co., New York, NY, USA, 1969.

[16] D. Mercer, *The Telephone: The Life Story of a Technology*. Westport, CT, USA: Greenwood Publishing Group, 2006.

[17] The Telephone in America, American Telephone and Telegraph Company, New York, NY, USA, 1936 & 1952.

[18] "Alexander Graham Bell – American inventor", Encyclopædia Britannica [Online]. Available: http://www.britannica.com/biography/Alexander-Graham-Bell

[19] L. Coe, *The Telephone and It's Several Inventors: A History*. Jefferson, NC: McFarland, 1995.

[20] H. N. Casson, *The History of the Telephone*, 10th ed. Chicago, USA: A. C. McClurg & Co., 1922.

[21] B. S. Finn, "Bell and Gray: just a coincidence?", *Technology and Culture*, vol. 50, no. 1, pp. 193–201, 2009.

[22] T. Farley and M. van der Hoek, "Telephone History: The Inventors Gray and Bell", Jan, 2006 [Online]. Available: http://www.privateline.com/mt_telephonehistory/2006/01/the_inventors_gray_and_bell.html

[23] M. E. Gorman and W. B. Carlson, "Interpreting invention as a cognitive process: THE CASe of Alexander Graham Bell, Thomas Edison and the telephone", *Science Technol. & Human Values*, vol. 15, no. 2, pp. 131–164, 1990.

[24] A. G. Bell, "Experiments made by Alexander Graham Bell", Bell Notebooks Pages, vol. I, p. 26, 1876.

[25] C. Wheatstone, *The Scientific Papers of Sir Charles Wheatstone*. London: Taylor & Francis, 1879.

[26] F. W. Wile, *Emile Berliner – Maker of the Microphone*. New York: Arno Press, 1974.

[27] E. Berliner, "Improvement in electrical-contact telephones", U.S. Patent no. 222652, Dec. 1879.

[28] F. L. Dyer and T. M. Commerford, *Edison: His Life and Inventions*, vol. 1. New York: Harper & Brothers Publishers, 1910, p. 226.

[29] B. Heil, "The microphone: a short illustrated history", *QST*, vol. 90, no. 6, p. 50, 2006.

[30] I. Wills, "Instrumentalizing failure: Edison's invention of carbon microphone", *Annals of Science*, vol. 64, no. 3, pp. 383–409, 2007.

[31] I. Hughes, "Professor David Edward Hughes", *AWA Review*, vol. 22, pp. 111–134, 2009.

[32] D. M. Worrall, "David Edward Hughes: Concertinist and Inventor", Papers of the International Concertina Association (PICA), pp. 41–57, vol. 4, 2007.

[33] Early Telephone Apparatus [Online]. Available: http://www.sparkmuseum.com/TELEPHONE.HTM

[34] J. Klapholz, "History and development of microphones", *Sound & Commun. Mag.*, Sept./Oct., 1986.

[35] W. H. Martin, "Seventy-five years of telephone: an evolution in technology", *Bell System Technical J.*, vol. 30, no. 2, pp. 215–238, 1951.

[36] A. G. Bell, "Bell's Experimental Notebook", vol. 1, p. 40, 1876 (Manuscript Division, Library of Congress, USA).

[37] A. A. Huurdeman, *Worldwide History of Telecommunications*. Hoboken, NJ, USA: Wiley, 2003, pp. 237–240.

[38] T. Farley, Telephone History Series (Rev. 1), Aug. 2010, pp. 1–45 [Online]. Available: http://www.privateline.com/TelephoneHistory/History1.htm

[39] A. L. Albert, *Electrical Communication*, 3rd ed. New York: Wiley and London: Chapman & Hall, 1954.

[40] A. G. Bell, "Invention of telephone", *Elec. Rev. & West. Electrician*, vol. 59, no. 20, pp. 972–974, 1911.

[41] A. E. Evenson, *The Telephone Patent Conspiracy of 1876*. Jefferson, NC, USA: McFarland, 2000.

[42] C. Beauchamp, "Who invented the telephone?: lawyers, patents, and the judgments of history", *Technology & Culture*, vol. 51, no. 4, pp. 854–878, 2010.

[43] S. Shulman, *The Telephone Gambit*. New York: W. W. Norton & Company, 2008.

[44] R. O. Meyer and E. S. Grosvenor, "Did Alexander Graham Bell steal the telephone patent?", *American Heritage*, vol. 58, no. 4, p. 52, 2008.

[45] A. G. Bell, "Experiments made by Alexander Graham Bell", Bell Notebooks Pages, vol. I, p. 85, 1876.

[46] F. De Land, "Notes on development of telephone service. II", *The Popular Science Monthly*, vol. 69, no. 3, pp. 490–501, 1906.

[47] K. MacDonald, "Innovation in Cambridge – First Phone Call", Cambridge Historical Society, 2002 [Online]. Available: http://www.cambridgehistory.org/discover/innovation/First%20Phone%20Call.html

[48] S. E. McMaster, *The Telecommunication Industry*. Greenwood Press, 2002.

[49] J. Brooks, *Telephone: The First Hundred Years*. New York: Harper & Row, 1976, pp. 40–41.

[50] Telephone History, Bell Homestead National Historic Site, Brantford, ON, Canada [Online]. Available: http://www.bellhomestead.ca

[51] Telephone Instruments 1877–1900, Illinois Bell's Telephony Museum in Chicago, IL, USA [Online]. Available: http://www.telephonymuseum.com/telephones.htm

[52] Worlds First Long Distance Telephone Line in French Corral, California Office of Historic Preservation, California State Parks, CA, USA, 1937.

[53] Nevada County, California Reformation Prayer Network, Western Region, Live Oak, CA, USA.

[54] Worlds First Long Distance Telephone Line 1878, Nevada County Historical Landmarks Commission, Nevada City, CA, USA.

[55] C. Wolff, "The Ridge Telephone Company and the California Electric Works CD 130.1 Cobalt Blue Insulator", March 13, 2007 [Online]. Available: http://www.insulators.info/articles/ridge_telephone.htm

[56] W. Preece, Hand-written notes by Sir William Preece from 1877 for Thomas Edison's newly invented "Speaking Telegraph", The IET Archives [Online]. Available: http://www.theiet.org/resources/library/archives/featured/edison.cfm

[57] L. B. Firman, "Manual Service Exchanges: Multiple switch boards for Telephone Exchanges", U.S. Patent no. 252576, Jan. 1882.

[58] T. Petruzzellis, *Telephone Projects for Evil Genius*. New York: McGraw-Hill, 2008.

[59] I. N. McFee, *Famous Americans for Young Readers: The Story of Thomas A Edison*. New York: Barse & Hopkins, 1922.

[60] T. Falzons Collection, Historic Butter Stamp Telephones, Australian Telephone Collectors Society (ATCS) Newsletter, July 2009 [Online]. Available: http://www.telephonecollecting.org/Newsletter/July09.pdf

[61] J. K. La Rue, Early Telephone Apparatus, John K. La Rue's (JKL) Museum of Telephony, Dedicated to Preserving Telephony [Online]. Available: http://jklmuseum.com

[62] Early Telephones, The Museum of Technology, Museum Libraries and Archives (MLA), Cambridge, UK, 2007.

[63] S. N. D. North, Telephones and Telegraphs 1902, Department of Commerce Labor, US Bureau of the Census, Special Reports Government Printing Office, Washington, WA, USA, 1906.

[64] Invention the Telephone, Intellectual Property of American Telephone and Telegraph Company (AT&T), 2014.

[65] S. Adams and O. R. Butler, *Manufacturing the Future: A History of Western Electric*. New York: Cambridge University Press, 1999.

[66] A. B. Iardella, "Western Electric and Bell System – A Survey of Service", Western Electric Company, New York, NY, USA, 1964.

[67] R. Blodgett, *Timeless Values, Enduring Innovation: The Graybar Story*. Old Saybrook, CT, USA: Greenwich Publishing Group, 2009.

[68] G. Adair, *Thomas Alva Edison: Inventing the Electric Age*. New York: Oxford University Press, 1996, pp. 59–65.

[69] W. S. Pretzer, *Working at Inventing: Thomas A. Edison and the Menlo Park Experience*. Baltimore, Maryland, USA: The Johns Hopkins University Press, 2002.

[70] G. M. Phelps, "Phelp's Crown Series of Telephones", U.S. Patent no. 218684, Aug. 19, 1879.

[71] W. Lemstra, "Innovation and landscape in telecommunication industry – the paradigm shifts", in *Proc. 2nd Int. Conf. Compet. Regul. in Netw. Industry*, Brussels, Belgium, 2009, pp. 5–12.

[72] L. Smith, "First Commercial Telephone Exchange – Today in History: January 28", New Haven – National Historic Landmarks, New Haven, New Haven County [Online]. Available: http://connecticuthistory.org/the-first-commercial-telephone-exchange-today-in-history/

[73] R. Burlingame, *Engines of Democracy*. New York: Arno Press, 1976 (Reprint Edition, Ed. Charles Scribners Sons 1940, Reprinted from copy available in Newark Public Library, USA).

[74] F. Deland, "Notes on the development of telephone service", *The Popular Science Monthly*, vol. 70, pp. 229–232, 1907.

[75] V. Green, *Race on the Line: Gender, Labor, and Technology in the Bell System, 1880ℓ1980*, 1st ed. Durham, NC, USA: Duke University Press Books, 2001.

[76] R. J. Chapuis and A. E. Joel, *100 Years of Telephone Switching*, 2nd ed. Amsterdam: IOS Press, 2003, pp. 48–49.

[77] K. D. Peterson, *Business Telecom Systems: A Guide to a Choosing Best Technologies and Services*. New York: CMP Books, 2000, pp. 23–24.

[78] J. Cooley, "Viral Triggers: Telephony", Viral Networking Working Group, MIT Communications Futures Program, CFP Bi-Annual Meeting, MIT, Cambridge, MA, USA, Jan 25–27, 2005.

[79] The Telephone offices: San Francisco Region Central offices [Online]. Available: http://www.thecentraloffice.com/calif/SFO/San Francisco.htm

[80] C. Reynolds, Albany Chronicles, *A History of the City Arranged Chronologically, from the Earliest Settlement to the Present Time; Illustrated with Many Historical Pictures of Rarity and Reproductions of the Robert C. Pruyn Collection of the Mayors of Albany, Owned By the Albany Institute and Historical and Art Society*. J. B. Lyon Company, Albany, NY, USA, 1906, pp. 56–57.

[81] History: AT&T Federal Internet Law and Policy – An Educational Project, Cybertelecom, 1997 [Online]. Available: http://www.cybertelecom.org/notes/att.htm

[82] C. H. Sterling, P. W. Bernt, and M. B. H. Weiss, *Shaping American Telecommunications: A History of Technology, Policy, and Economics*. Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers, 2006.

[83] K. B. Miller *et al*., *Encyclopedia of Telephony and Telegraphy*. Chicago American School of Correspondence, St. Lansing, IL, USA, 1919.

[84] J. Gibson (Ed.), *Mobile Communications Handbook*. Boca Raton, FL, USA: CRC Press/IEEE Press, 1996.

[85] A. Stone, *Public Service Liberalism – Telecommunication and Transition in Public Policy*. Princeton, NJ, USA: Princeton University Press, 1991.

[86] The Bell – Western Union, Patent Agreement of 1879: ECCH the Case for Learning, INSEAD, Fountainbleau, France, 2005.

[87] B. Estreich, "Emile Berliner" [Online]. Available: http://telephonecollecting.org/berliner.pdf

[88] E. Berliner, "The improved gramophone", *Trans. American Instit. of Electr. Engin.*, vol. 8, no. 1, pp. 19–30, 1891.

[89] Edison vs. Blake vs. Voelker vs. Irvin, Telephone Interface, Voelker's Specifications and Papers, Statements of Francis Blake Jr (Jul. 1878), Thomas A. Edison (Sep. 29, 1879), William I. Voelker (Nov. 1879), J. H. Irvin (Dec. 18, 1879), United Patent Office, IEEE Global History Network [Online]. Available: http://ieeeghn.org

[90] E. Hall, *Francis Blake: An Inventor's Life (1850–1913)*. Boston, MA: Massachusetts Historical Society, 2005.

[91] F. Blake, Speaking Telephone", Patent no. 250126, United States Patent Office, 1881.

[92] M. Martin, *"Hello, Central?" – Gender Technology and Culture in the formation of Telephone Systems*. Montreal, Quebec, Canada: McGill Queen's University Press, 1991, pp. 18–20.

[93] W. J. Hammer, "An automatic telephone operator", *Trans. American Instit. of Electr. Engin.*, vol. 21, pp. 31–54, 1903.

[94] W. D. Lockwood, "Evolution of telephone switchboard", *Trans. American Instit. of Electr. Engin.*, vol. 21, pp. 3–30, 1903.

[95] C. M. Maxfield and A. Brown, *An Illustrated History of Electronics and Computing*, 2nd ed. Oxford, United Kingdom: Newnes Press, 2003.

[96] F. B. Jewett, *Biographical Memoir of John Joseph Carty 1861–1932*. National Academy of Sciences of the United States of America Biographical Memoirs, Volume XVIII, Fourth Memoirs, 1936 [Online]. Available: http://www.nasonline.org.

[97] L. B. Firman, "Multiple Switchboard for Telephone Exchanges", Western Electrical Manufacturing Company, U.S. Patent no. 252576, Jan. 1882.

[98] "The Communications Era (1879–1905) – Extending Man's Voice by Wire and Radio" [Online]. Available: http://customers.hbci.com/∼wenonah/history/edpart3.htm.

[99] Bicentennial issue of *Electronic Design* for engineers and engineering managers, vol. 24, no. 4, Feb. 16, 1976.

[100] J. Race, "Old telephones: how to repair and rebuild them", Technical Notes, Cambridge Electronics Laboratories, Somerville, MA, USA, 2001.

[101] D. Linn, R. Larese, and J. Pan, "E&M Signaling", White Paper, FAE Telecom, Acton, MA, USA [Online]. Available: www.faetelecom.com

[102] "Principles of Electricity applied to Telephone and Telegraph Work", Bell Training Course Material, American Telephone and Telegraph Company, January, 1953 [Online]. Available: http://wedophones.com/TheBellSystem/pdf/attbook.pdf

[103] R. MacDougall, "The Telephone on Main Street: Utility Regulation in the United States and Canada before 1900", *Business and Economic History*, vol. 4, 2006 [Online]. Available: http://www.thebhc.org/sites/default/files/macdougall.pdf

[104] H. Schulzrinne, "Request for Comments 3966 (RFC-3966)", Columbia University Network Working Group, Dec. 2004 [Online]. Available: http://www.ietf.org/rfc/rfc3966.txt

[105] J. E. Brittain, "The introduction of the loading coils: George A. Campbell and Michael I. Pupin", *Technol. & Culture*, vol. 11, no. 1, pp. 35–57, 1970.

[106] J. E. Brittain, "Amazing biography of Michael I. Pupin", *IEEE Industry Appl. Mag.*, no. 1, Jan.-Feb., 2006.

[107] H. F. Campbell, "Metallic Circuit", Patent no. 411137, United States Patent Office, Sept. 1889.

[108] T. Shaw and W. Fondiller, "Developments and applications of loading coils for telephone circuits", *Trans. American Instit. of Electr. Engin.*, vol. 45 , no. 3, pp. 291–292, 1926.

[109] Almon B. Strowger 1839-1902 (Inventor), Kansas Sesquicentennial, 150 Years and counting Science in Kansas, Kansapdeia, Kansas Historical Society, Ad Astra Kansas Initiative, 2010 [Online]. Available: http://www.adastra-ks.org/150scientists/150scientists_Oct_3.pdf

[110] R. B. Hill, "Early years of the Strowger system", *Bell Lab. Record*, vol. 31, no. 3, pp. 95–103, 1953.

[111] R. B. Hill, "Early work on dial telephone systems", *Bell Lab. Record*, no. 1, pp. 22–29, 1931.

[112] S. J. Bigelow, J. J. Carr, and S. Winder, *Understanding Telephone Electronics*, 4thed. Oxford, United Kingdom: Newnes, 2001.

[113] P. J. Povey and R. A. Earl, *Vintage Telephones of the World*. London: Peter Peregrinus Ltd., 1988.

[114] "Alexander Graham Bell Speaking on the Phone, 1892", Gilbert H. Grosvenor Collection of Photographs, Prints and Photographs Division, Library of Congress, 1892.

[115] H. H. Nance, "Some very long telephone circuits of the Bell system", *Bell System Technical J.*, vol. 3, no. 3, pp. 495-507, 1924.

[116] A. Sarlemijn and P. Kroes, *Between Science and Technology: Proceedings of the International Conference Between Science and Technology, Eindhoven University of Technology, The Netherlands, 29-30 June 1989*. Amsterdam: North-Holland, 1990.

[117] J. J. Pilliod, "Transcontinental telephone lines", *Bell System Technical J.*, vol. 18, no. 1, pp. 235–245, 1939.

[118] R. B. Hill, "Early work of dial telephone system", *Bell Lab. Record*, vol. 31, no. 1, pp. 22–29, 1953.

[119] E. Lindquist, The invention and development of the dial telephone: the contribution of three Lindsborg inventors", *Kansas Historical Quarterlies*, vol. 23, no. 1, pp. 1–8, 1957.

[120] M. Csele, "Telephone Switches", Niagara College (US), July 22, 2014 [Online]. Available: http://192.197.62.35/staff/mcsele/TelephoneSwitch.html

[121] J. H. Stevens, "Basic Relay Circuits" [Online]. Available: http://www.kadiak.org/tel/index.html

[122] "Switches, PBX's, and Central Offices", Survey of Telephone Switching, The Pacific Telephone and Telegraph Company, 1956 [Online]. Available: http://www.telephonetribute.com/switches.html

[123] "Attachments to Public Switched Telephone Network (PSTN); General technical requirements for equipment connected to an analogue subscriber interface in the PSTN; Chapter 1: General", series EN 300 001 v1.5.1 (1998-10), European Telecommunications Standards Institute, Jan. 1997 [Online]. Available: http://www.etsi.org

[124] J. Bathgate, "Introduction to Strowger", The Dean Forest Railway Telecoms, Norchard Centre, Lydney, Forest of Dean, Gloucestershire, no. 131, UK, Jan. 2008 [Online]. Available: http://www.dfrtelecoms.org.uk/pdf/basicex.pdf

[125] T. Viswanathan, *Telecommunication Switching Systems and Networks*. Delhi, India: Prentice Hall of India, 2004, pp. 40–43.

[126] N. Wilkinson, *Next Generation Network Services*. Chichester, UK: Wiley, 2002.

[127] I. Vogelsang and G. Woroch, "Local telephone service: a complex dance of technology, regulation and competition", in *Industry Studies*, 2nd ed., L. L. Deutsch, Ed. New York: M. E. Scharpe, 1998 [Online]. Available: http://eml.berkeley.edu/∼woroch/dance.pdf

[128] A. Lebedev, "A Short History of Telephone Numbers" (Mandership project, graphic and industrial design, interface engineering, typography, semiotics, and visualization, since 1997), Art Lebedev Studio, New York, Moscow, Kiev, June 18, 2002 [Online]. Available: http://www.artlebedev.com/mandership/91

[129] A. B. Clark and H. S. Osborne, "Long distance telephone circuits in cable", *Bell System Technical J.*, vol. 11, no. 4, pp. 520–545, 1932.

[130] R. Juch, "AT&T Long Lines", The Bell System Unit For Nationwide and Worldwide Communications, 1969 [Online]. Available: http://www.beatriceco.com/bti/porticus/bell/longlines.html

[131] R. MacDougall, "Long lines: AT&T, long distance telephony and corporate control", *Business and Economic History*, vol. 3, 2005 [Online]. Available: http://www.thebhc.org/sites/default/files/macdougall_0.pdf

[132] L. Carter *et al.*, "About Submarine Telecommunication Cables", International Cable Protection Committee Ltd, Lymington, United Kingdom, Oct. 2011 [Online]. Available: https://www.iscpc.org/documents/?id=132

[133] "1927: Transoceanic Telephone Service" [Online]. Available: http://www.corp.att.com/attlabs/reputation/timeline/27atlan.html

[134] "History of Telecommunications", CNES – Information Systems Directorate, Toulouse, France [Online]. Available: https://cnes.fr/en/web/CNES-en/1105-history-of-telecommunications.php

[135] B. Burns, First Submarine Transatlantic Telephone Cable System TAT-1, FTL Design [Online]. Available: http://atlantic-cable.com/Cables/1956TAT-1/

[136] E. C. Wente, "The condenser transmitter as a uniformly sensitive instrument for the absolute measurement of sound intensity", *Phys. Rev.*, vol. 10, no. 1, pp. 39–63, 1917.

[137] E. C. Wente, "The sensitivity and precision of the electrostatic transmitter for measuring sound densities", *Phys. Rev.*, vol. 19, no. 5, pp. 498–503, 1922.

[138] E. C. Wente, "Contributions of telephone research to sound pictures", AT&T Archives, Bell Telephone Laboratories, New York, Oct. 1935.

[139] Federal Internet Law and Policy – An Educational Project, Cybertelecom, 1997 [Online]. Available: http://www.cybertelecom.org/notes/telephone.htm

[140] G. William and A. Drummer, *Electronic Inventions And Discoveries: Electronics From Its Earliest Beginnings To The Present Day*. Boca Raton, FL: CRC Press, 1997.

[141] P. J. Covey, "The Telephone and the Exchange", Post Office Publication, pp. 75–84, 1917.

[142] C.-F. Helgesson, "The history of Eriksson: The crossbar switch from concept to success", Ericsson [Online]. Available: http://www.ericssonhistory.com/products/the-switches/The-crossbar-switch—from-concept-to-success/

[143] P. Walker, "Crossbar Switching Systems", Telecommunication Heritage Group for preservation and collection of the heritage of communications, London, UK [Online]. Available: http://www.thg.org.uk/index.php/telecoms/115-crossbar-switching

[144] "Swedish Telecom-Company History", Funding Universe, Farsta, Sweden [Online]. Available: http://www.fundinguniverse.com/company-histories/swedish-telecom-history/

[145] R. L. Freeman, *Fundamentals of Telecommunications*, 3rd ed. Hoboken, NJ, USA: Wiley, 2005, pp. 73–77.

Vinayak L. Patil, received his B.Sc. in 1972, B.E. in 1975, M.E. in 1979 and Ph.D. in 1990. He worked as Scientist at Central Electronics Engineering Research Institute (CEERI), Pilani from 1977 to 1991 for 14 years and during his service at CEERI he worked on Electronics System Designs based on microprocessors and computers for agriculture and agro-based industries. The work done at CEERI resulted many national and international awards confirmed on the author. From 1991 to 1997 he worked as a Director at Centre for Electronics Design and Technology (CEDT), Aurangabad, India (now National Institute of Electronics and Information Technology), where he promoted electronics product design for agriculture and agro-based industries. From 1997 he also worked in the capacity of directors and principals of various engineering colleges and presently working as a professor in Trinity College of Engineering and Research, University of Pune, Pune, India.
E-mail: patilvl.works@gmail.com
Trinity College of Engineering and Research
University of Pune
Ganesh Khind
411007 Pune, India

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The author's JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX author's style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, "Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres", *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, "Informational representability: Abstract models versus concrete models", in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

Biographies and photographs of authors. A brief professional author's biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishing's databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*Contents Continued from Front Cover*)

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.nit.eu

INSTYTUT ŁĄCZNOŚCI
PAŃSTWOWY INSTYTUT BADAWCZY

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY 3/2015