ПЛЕНАРЛЫҚ МӘЖІЛІС БАЯНДАМАЛАРЫ

ДОКЛАДЫ ПЛЕНАРНОГО ЗАСЕДАНИЯ

PLENARY MEETING

UDC 530.145

V. V. NIKULIN

State University of New York, Binghamton, USA

PERFORMANCE OF QUANTUM ENCRYPTED FREE-SPACE OPTICAL COMMUNICATION LINKS

Abstract

Free-space optical links offer secure communication channels with low probability of interception and detection. However, based on practical issues, such as antenna size and pointing errors, there is still a risk of unauthorized access to information, which can be significantly minimized with data encryption. A search for the "holy grail" of cryptography has led to the application of quantum mechanical principles in optical communication networks. Quantum processes at the physical layer of encryption can be used to facilitate ultra-secure quantum communications (QC) with very competitive performance characteristics. This paper discusses a system that uses keyed communication in quantum noise as an encryption mechanism. Simulation studies of the effects of quantum noise on phase estimation are conducted for quantum systems with different number of encryption bases and operating at different power levels.

1. Keyed communication in quantum noise

Alpha-Eta (αη) quantum communication protocol, also known as Y-00 protocol in Japan, is a product of NuCrypt developed under AFRL and DARPA funding [1], [2]. It is a random cipher in the sense that the cipher text output is a quantum state and it requires coherent measurements to perform decryption. Fig. 1 illustrates operation of the system where a random (seed) key is first extended into a much longer running key using, for example, a linear feedback shift register (LFSR) for combination with a communication stream (binary message) to generate ciphertext. Then quantum features of light, such as polarization or phase, are used as a physical mechanism for encrypting a signal before it is sent via an optical channel. Encryption bases shown in the middle in Fig. 1 are arranged close to each other and with alternating values of bits, such that without knowing an encryption key it is very difficult to detect bit values hidden in fundamental and irreducible quantum measurement noise of coherent states. While the ultimate goal of this research can be formulated as finding the limits of exploitation of quantum communication links and a valid proof of their level of secrecy, first we need to come up with a blueprint of steps required to achieve this objective. There are two security features that make exploitation extremely difficult. First, an eavesdropper needs to inspect many decrypted ciphertext possibilities to find a plaintext message. Second, the amount of data to be stored for subsequent decryption generally exceeds any reasonable storage capacity available to an eavesdropper. The task becomes even more formidable as the number of encryption bases N_b is increased and the mean photon-number $|\alpha|^2$ is decreased.

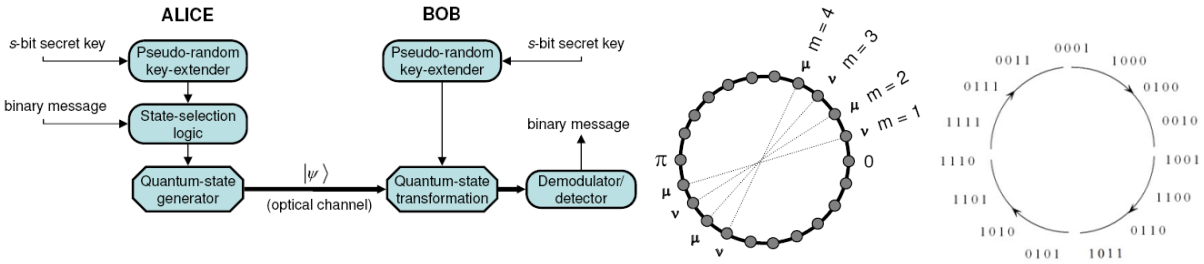


Figure 1. The Alpha-Eta stream cipher hides bits in irreducible quantum noise

In a practical Alpha-Eta system with phase-based encryption the fundamental noise is expected to bring significant measurements ambiguity. The main objective of this research is not a plaintext attack, but estimation of the values of the running key. Let the number of encryption bases be 2^m, then each value of the running (pseudo-random) key that designates a particular phase encryption base can be expressed by the following bit sequence: (a^{m-1} a^{m-2} ... a^0). Since each measurement of the phase is affected by the quantum noise of light, not all of the bits can be estimated accurately, especially the least significant bit. However, a certain number of the most significant bits (a^{m-1} a^{m-2} ... a^n...) can be measured with a certain degree of precision. The value of *n* will be studied as a function of the number of encryption bases and the mean photon-number.

2. Phase-based quantum encryption

Advancement and popularity of laser communication over the years has called the attention of data security and encryption for free-space optical links. In particular, phase-based encryption protocols are used for free space optical communication. The concept and the associated methodology is better demonstrated using Alpha-Eta αη protocol, which is a form of keyed communication in quantum noise (KCQ). The two - mode coherent states are applied as follows [3]:

$$\left|\psi_m^{(a)}\right\rangle = \left|\alpha e^{i\theta_m}\right\rangle \text{ and } \left|\psi_m^{(b)}\right\rangle = \left|\alpha e^{i(\theta_m+\pi)}\right\rangle \quad (1)$$

Here, $\theta_m = \pi m/M$, where *M* must be odd and $m \in \{0,1,2,\ldots(M-1)\}$. Hence θ_m may vary between 0 and $\pi m/M$ with increments of π/M. The 2M states make up M bases (antipodal phase pairs) that uniformly span the phase circle, as shown on the right in Fig. 1. As an example, if $M = 5$, θ_m varies between 0 and $4\pi/5$ with increments of $\pi/5$. A combination of the number of bases M and the mean photon count $|\alpha|^2$ determines the level of secrecy. Generally, with a larger M and smaller $|\alpha|^2$ we have more ambiguity when determining the measurement base.

The αη protocol begins with the use of a private *s*-bit seed (secret) key known to both Alice and Bob. Now using this pre-shared seed key, an extended key is generated using a linear feedback shift register (LFSR). The LFSR outputs (2^s-1) unique combinations of the seed key and then repeats itself. This extended key is then converted into a running key using a "mapper" or an invertible *n*-bit *to n*-bit deterministic algorithm where n=Int($\log_2$M) and n<<s.

Next, depending on the data bit to be sent and the instantiation of the running key, a phase state is chosen. The decimal representation of the binary running key called *m* is used to calculate θ_m, where $\theta_m = \pi m/M$. Since the data bits are encoded differentially using differential-phase-shift keying (DPSK) approach, the following mapping approach is used: (0, π) → ($\left|\psi_m^{(a)}\right\rangle$, $\left|\psi_m^{(b)}\right\rangle$) for even bases and (0, π) → ($\left|\psi_m^{(b)}\right\rangle$, $\left|\psi_m^{(a)}\right\rangle$) for odd bases. If 0 and π are re-labeled as μ and ν, as shown in Fig. 1, then logic '0' is encoded as $\left|\psi_m^{(\mu)}\right\rangle$ ($\left|\psi_m^{(\nu)}\right\rangle$) is the previous state was the same, i.e. { $\left|\psi_m^{(\mu)}\right\rangle$} ({$\left|\psi_m^{(\nu)}\right\rangle$}). Correspondingly, logic '1' is encoded as $\left|\psi_m^{(\mu)}\right\rangle$ ($\left|\psi_m^{(\nu)}\right\rangle$) is the previous state was the same, i.e. { $\left|\psi_m^{(\nu)}\right\rangle$} ({$\left|\psi_m^{(\mu)}\right\rangle$}).

When the signal is received the same secret key is used to decrypt the message by applying signal transformations. As a result, depending on the bit values, we have

$$\left|\psi_m^{(a)}\right\rangle' = |\eta\alpha\rangle \quad (2)$$

$$\left|\psi_m^{(b)}\right\rangle' = |-\eta\alpha\rangle \quad (3)$$

Since the bit values are encoded differentially using two subsequent symbols, detection requires some form of interferometry

3. Simulation results and analysis

A complete program was implemented in MATLAB to simulate transmission of encrypted bit symbols. Each symbol has the following properties: the number of photons is user-defined and follows Poisson distribution. Each photon's phase has three components: encryption component defined by the Alpha Eta algorithm, random atmospheric distortion component, and a random source wave front distortion component. The resultant phase state can be estimated using any of the standard approaches. True phase values are $\theta_m = \pi m/M$ or $\theta_m = \pi m/M + \pi$; however, due to the phase distortions even a perfect demodulation/detection system will estimate with an error. Then the answer is rounded to the nearest phase base. Since the spacing between any two adjacent phases is $\theta_m = \pi/(M+1)$ an error in base estimation can occur whenever a phase error exceeds $\theta_m = \pi/[2(M+1)]$. As a result, our estimate can be off by one or more phase bases. No noise of the demodulation/detection hardware is modeled to isolate the effects of encryption and to estimate the level of secrecy. The following user-defined parameters are used in our simulation study: M – the number of encryption bases, $|\alpha|^2$ – the mean photon-number per symbol, σ_{rms} – rms wave front error of the transmitted signal. The results presented in Fig. 2 show percentage of phase bases estimation errors for a beam with very little phase noise (equivalent rms wave front error is $\lambda/10$). The results are given for three different values of M.

It was discussed earlier that a particular phase encryption base can be expressed by the following bit sequence: (a^{m-1} a^{m-2} ... a^0). The number of bits used depends on the number of bases. Therefore, valuable information can be obtained by looking at the distribution of the base estimation errors to see how much the result is off from a true base number.

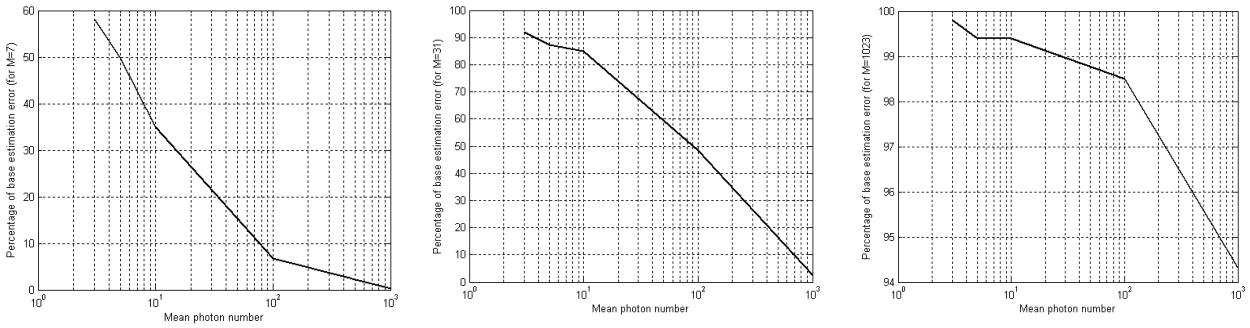
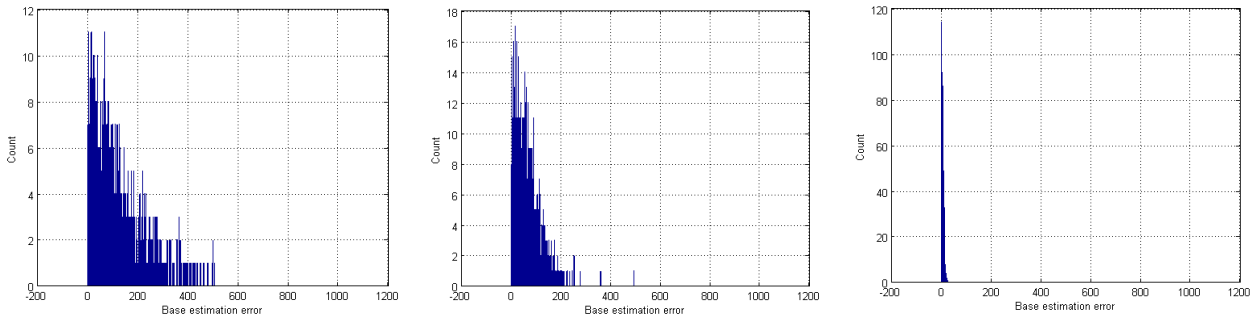


Figure 2. System performance as a function of photon count for $\sigma_{rms} = \lambda/10$

The last case study is presented in Fig. 3 when 1023 bases are used (10 bits).



$|\alpha|^2=3$ $|\alpha|^2=10$ $|\alpha|^2=1000$

Figure 3. System performance as a function of $|\alpha|^2$ for M=1023, $\sigma_{rms} = \lambda/10$

With 3 photons per bit the maximum error is 506 bases and no information on the running key can be obtained. When $|\alpha|^2=10$, the largest error is 493 bases; therefore, 9 out of 10 bits remain unknown and only the MSB can be estimated. With 1000 photons per bit, the maximum error is 27

bases, which makes the last 5 bits indistinguishable. However, we are able to estimate $a^9 \ldots a^5$ from the running key.

4. Conclusion

A simulation testbed presented in this paper has provided a model for a free-space Alpha Eta optical encryption link. Our case study shows the level of security for different combination of design parameters. The developed model can be used to estimate performance of a practical KCQ link prior to building a hardware prototype.

References

1. S. Kartalopoulous, "Effect of Noise in Quantum Communications," *Proc. SPIE*, Vol. 6603, 66030L, 2007.

2. H. Yuen, "KCQ: A new approach to quantum cryptography I. General principles and qubit key generation," quant-ph/0311061, 2003.

3. E. Corndorf, "Quantum *Cryptography Using Coherent States: Randomized Encryption and Key Generation*," Ph.D. Dissertation, Northwestern University, June 2005.

SILAMU W., HUI W., HUIQIN J.[1, 2]

[1]*College of Information Science and Engineering, Xinjiang University, Urumqi*
[2]*Key Laboratory of Multilingual Information Technology, Xinjiang, Urumqi*

AN OVERVIEW OF UYGHUR SPEECH RECOGNITION

Abstract: This paper reviews the progress of the Uyghur speech recognition research from the Uyghur acoustic features, speaker identification, keyword spotting, continuous speech recognition, etc. Firstly, this paper described the key technologies of the Uyghur speaker recognition from the feature extraction and transformation, the application of the hybrid model. Secondly, implement the Uyghur keyword spotting (KWS) by using the filler model, and tell the function of syllable lattice and confusion network. Finally, construct the Uyghur acoustic model and language model in continuous speech recognition based on continuous hidden Markov model. It also explores the applications of the Uyghur acoustics, prosody features in the post-processing.

Keywords: Speech Recognition; Acoustic Models; Language Models; Uyghur

1. Introduction

The automatic speech recognition (ASR) technology began in the 1950s. With the rapid development of computer hardware and software technology, and the gradual improvement of digital signal processing, machine learning, statistical learning theory, pattern recognition theory to lay the foundation for the research of speech recognition. In early 1970s, the speech signals linear predictive coding technology (LPC), dynamic time warping (DTW), vector quantization (VQ) was emerged. In 1980s, the hidden Markov model (HMM) based on the statistical theory replaced the method of traditional template matching. After that, there have been many successful applications based on the hybrid model of hidden Markov model with the Gaussian mixture model (GMM) or with the artificial neural network (ANN).

Uyghur belongs to the Turkic and Altaic, Uyghur is an agglutinative language, a wealth of affixes append to the same stem may constitute a very large scale vocabulary, and the lack of text and speech corpus, restricting the development of the Uyghur language speech recognition. The Uyghur speech recognition research started in the 1990s, with the establishment of the Uyghur speech acoustic parameters corpus, Uyghur speech recognition research has been made some

progress. In 1994, we developed a Uyghur isolated-word speech recognition system with the recognition units of 1200 syllables, and the vocabulary of 40000 words [3]. Since 2000, the Uyghur continuous digital recognition system based on the method of DTW and VQ has been developed by Xinjiang Multi-lingual Information Key Laboratory. A HMM-based Uyghur continuous speech recognition system [4][6], a Uyghur telephony speaker recognition system based on a hybrid model of the Gaussian and support vector machine (SVM), and the Filler model-based Uyghur keyword spotting system have been born in our laboratory. According to the different tasks of speech recognition, this paper will illustrate the progress of the Uyghur speech recognition research from the Uyghur speaker identification, keyword detection and continuous speech recognition.

2. Acoustic Features of the Uyghur

Now China's official Uyghur is based on Arabic alphabet and Latin Uyghur as a supplement. Uyghur Phonemes include 8 vowels (V) and 24 consonants (C), and the Latin Uyghur is described in table 1.

Table 1: The Latin Uyghur

V	a	E	i	O	u	ö	ü	ä
C	n	M	l	K	j	h	ğ	g
C	f	D	č	B	ž	z	y	x
C	w	T	š	S	r	q	p	ñ

Figure 1 show the vowel chart in type of Joos, the vertical axis is first formant, the horizontal axis is second formant and the ellipses represent vowels' degree of dispersion. We completed this statistics with manually annotated continuous reading speech data.

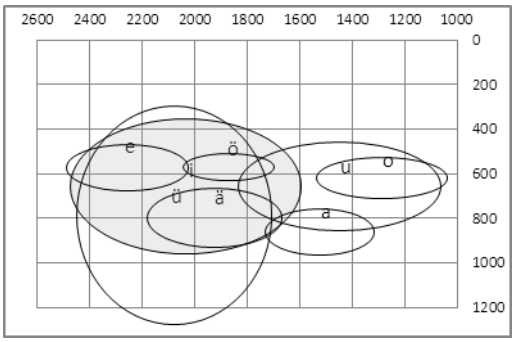


Figure 1: Uyghur acoustic vowel chart

With the constitutive rules of Uyghur Syllable, each syllable has one and only one vowel, and one vowel can form a separate syllable. Uyghur vowel harmony based on articulation harmony, and rounded harmony is appended, so the former is the most serious part, and the latter is relatively unimportant.

In the word stem, if the vowel in the first syllable is a front (back) vowel, the vowel in the syllable after the first one is also a front (back) vowel. While the harmony phenomena in affixes part are often affected by the last syllable of the word stem, and does not appear with [o], [ö], [e]. On the effect of the allophone and the vowel reduction in unrounded front vowel, such as [i], [e], which can take both front and back vowel in the harmony of these vowels, and they are also impacted by the consonant before them. Uyghur is a kind of accented language, word accent is usually on the last syllable, when append affix to a word, the accent will be moved to the last. Consonant loss and vowel reduction phenomena in the Uyghur word lead to accent moving forward.

3. Uyghur Speaker Recognition

The speaker's speech is able to demonstrate person's physiological characteristics. Speaker recognition can be divided into two tasks: speaker verification and speaker identification. According to the test speech, the former's task is to determine whether the speaker and the alleged reference speaker are matched or not. The latter will use test speech to judge the speaker belong to which one of the multiple reference speakers. Speaker identification can be divided into an open set identification and a closed set identification.

3.1 Feature Extraction and Transformation

The acoustic features that can characterize the speaker's personality characteristic mainly includes: the linear prediction coefficients and its derived parameters, which are consistent with the channel parameter model. For representing of speaker's physiological differences (such as the excitation source, channel characteristics),speaker recognition also choose features from short-term spectrum, pitch contour, formant bandwidth and its trajectory, Robust parameter cepstral coefficients, the dynamic characteristics result from the difference of static characteristic, etc. Each parameter given above has its own merit on the characterization of speaker's features, so feature fusion is required, but directly superimposed would increase the characteristic dimension and redundancy, and resulting in lower recognition performance. The transformation of the Uyghur language feature combines with principal component analysis (PCA) firstly, and then uses Fisher criterion for getting a set of high discrimination features. We also use the linear discriminant analysis (LDA), linear heterogeneous discriminant (HLDA) analysis to implement the hybrid feature dimension reduction processing.

3.2 Models of Speaker Recognition

Speaker recognition models, generally divided into the probability of statistical models (such as GMM, HMM), and the decision model (such as SVM, ANN), the former reacts the similarity of similar data, the latter reacts differences of difference data.

SVM is commonly used to solve the problems which encountered in nonlinear, high dimension, and limited sample of the pattern recognition problem in practical applications. In speaker identification system's training phase, regard each speaker as a class, extract each speaker's speech feature vectors as the input of the various types, after training, the results of which forms a multi-class SVM support vector and generates the classification function. In the identification stage, it is also extracted from the speech signal of the test feature vector sequence as the input of the SVM model. The SVM model classifies each vector, and then does statistics of the feature vector to assign which class of vector is the most. Finally, the model chooses the speaker corresponding to this class as the identification result.

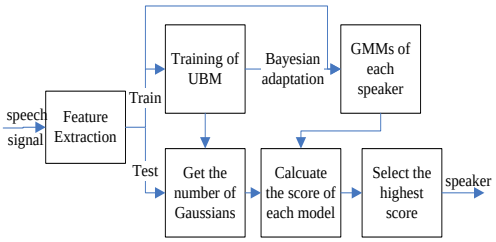


Figure 2: Depiction of GMM-UBM approach

The Universal Background Model (UBM) is a large GMM trained to represent the speaker-independent distribution of features. In the GMM-UBM based system, if the speaker's training data can cover all pronunciations, it could be use the speaker's data to modeling. If the data cannot cover all cases, it should be use the set of other speakers' data to training UBM and adapting. Figure 2 shows these steps in GMM-UBM modeling. In Uyghur speaker recognition, we combining the advantages of both approach, describe as in table 2 [7].

Table 2: Results of the Uyghur speaker recognition

length of time	10s	20s	30s
GMM	58.3	81.2	90.8
GMM-UBM/SVM	63.5	92.1	93.4

4. Uyghur Keyword Spotting

Keywords spotting technology is divided into: the spotting method which based on Filler model or continuous speech recognition. According to the method of continuous speech recognition, the document retrieval and content detection technology based on the syllable lattice and confusion network has become the focus research of keywords spotting.

4.1 Filler Model

Filler model absorbs all outside of the keywords in languages (such as out of vocabulary words) and non-linguistic phenomena (such as noise, etc.). The search network includes filler, keyword, which can reduce the insert and delete errors in syllable recognition. In order to prevent keywords filler model engulfed by the keywords, the search structure often sets a certain degree of reward to the keywords models or punishment to filler models.

Table 2 shows the results of the Uyghur keyword spotting based on filler model with different sample size and number of keywords.

Table 3: Word accuracy and length of recognition time

num of keywords	1	5	10	20
sample size	100	200	300	500
word acc (%)	97.0	93.5	91.0	89.6
average time	0.11s	0.54s	0.91s	1.61s

4.2 Syllable Lattice and Confusion Network

An approach of keywords spotting based on continuous speech recognition uses the decoder's output to form syllables lattice or confusion network, and these are stored as text files. And then the KWS system selects the candidate keyword from the syllables lattice or confusion network by keyword search algorithm. At last, it verifies and outputs by using the confidence assessment. Such spotting methods, without repeated recognition of speech documents, it only needs to keyword search from the syllable lattices or confusion networks, so it is efficient and suitable for large vocabulary keywords spotting.

Speech has not only associated with spectrum of each acoustic unit segment, but also affected by a variety of suprasegmental factors. Suprasegmental factors refer to a level beyond the context of the local speech, such as accent, intonation, speed, pause, etc. Uyghur keyword spotting research tries to put the prosody features into syllables lattice or confuse network for the secondary decoder to improve the identification accuracy.

5. Continuous Speech Recognition

With the establishment of the Uyghur continuous speech corpus (telephone speech and reading speech), laid the foundation for Uyghur large vocabulary continuous speech recognition (LVCSR). Figure 1 shows the main steps of LVCSR system. The Uyghur continuous speech recognition research, mostly based on hidden Markov model, combined with the characteristics of Uyghur language. We have made a lot of work in the choice of recognition units, the creation of pronunciation dictionary, the design of question sets and the construction of language models.

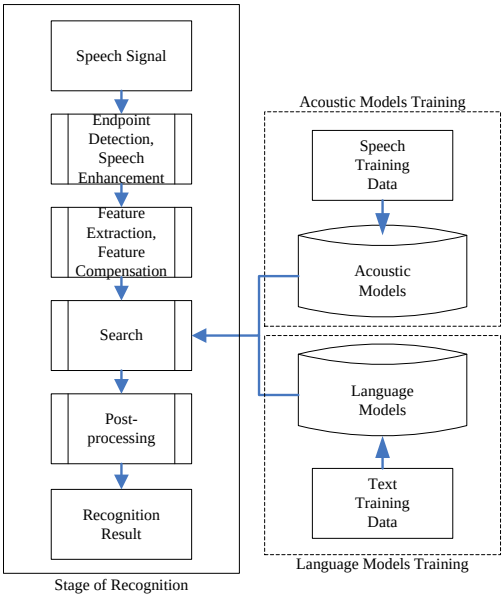


Figure 3: Main flowchart of LVCSR

5.1 Acoustic Models

At present, the mainstream speech recognition system generally uses statistical modeling frameworks based on Bayesian decision theory [1]. Uyghur contains more than 5000 commonly used syllables, so when use syllable as the unit of recognition, totally have 5000 monophone models, and the triphones will be extended to over ten million models, which will be lead to the sparseness of training data, resulting in parameter estimation insufficient. Uyghur continuous speech recognition generally uses the 34 phonemes (including silence and short pause) as the recognition units. Taking into account of the impact of context phoneme coarticulation, when Uyghur monophone is extended to triphone models, it would generate nearly 40,000 triphone models. It will also lead to a serious shortage of training data for each model (especially some models' training with less occurrences or not) and reduce the training efficiency, and too much storage space occupied. The model parameters shared strategy can reduce the scale of parameters to be estimated.

Uyghur continuous speech recognition is generally used the mechanism of decision tree clustering to binding similar state. The strategy of decision tree provided a classification by doing a combination of the top-down data-driven method and the experts' knowledge. For the splitting of decision tree, we need to combine the Uyghur acoustic and linguistic knowledge to build the problem sets. Uyghur questions are generally constructed in accordance with the vowel, consonant, the difference of articulation manner and articulation place and simple questions.

In the hybrid model of HMM/GMM [5], the Gaussian mixture models calculate the output probability of each model state in the HMM framework. Via the EM algorithm based on the probability of the GMM and maximum likelihood estimate (MLE) criteria to achieve the best analog on the probability distribution of each of the categories. With the increase of the Gaussian mixture component, the recognition rate there will be some improvement, but it also increase the reestimation of required parameters. So the increased number of Gaussian has associated with the size of the training corpus.

5.2 Statistical Language Models

The statistical language model provides the context and semantic information between words. As for Uyghur, it includes over 20,000 common roots and more than 300 affixes. They are formed an ultra-large-scale vocabulary through the combination. We lack speech corpus of the Uyghur language [2]: Firstly, a large number of balanced texts are difficult to obtain, cause the low coverage of commonly used words. Secondly, the field of the text corpus is limited. The size of the dictionary relates to the real system's recognition effect. Generally speaking, the vocabulary of the practical recognition system betweens 5-10 orders of magnitude. The Uyghur statistical language

model to sort through four areas of the text corpus (the medical, news, magazines and novels) , and do statistics of high-frequency appeared words， then make dictionaries with the vocabulary of 10-100 thousand words.

Because of the splice between the different classes of speech signal, the person is difficult to distinguish without context. The language model can improve the discrimination of the acoustic model. At present, more sophisticated statistical language model is the N-gram, the greater n value makes the higher precision and complexity of the language model. If the value of n is small, word sequence would make some certain words ignored, and result in some combinations cannot be rolled back. General choice of n = 3 to construct trigram language model.

Table 3 shows the recognition results of the Uyghur LVCSR, we use 6497 sentences for training and 90 sentences for testing with different number of the Gaussian mixture in 3-gram language models.

Table 4: Word and sentence accuracy of the recognition and length of recognition time

mixture	18	24	28	30
word acc (%)	86.70	89.60	92.02	92.14
sent acc (%)	53.33	58.89	66.67	64.44
length of time	35s	38s	41s	42s

Language model training text size and its distribution has some limitations and one-sidedness, reasonable words context does not appear in the training text. By expanding the size of the training corpus, the low-frequency words still cannot get enough statistical properties of parameter probability estimation. Smoothing is to adjust the parameter values in the language model, the increase of low probability and zero probability, reducing high probability of parameter values. Prevent the emergence of zero probability to improve the accuracy of the language model. Commonly used smoothing techniques (such as the Good-Turning Katz, linear interpolation smoothing) are selected in Uyghur language model.

6. Conclusion

This paper described the status of Uyghur speech recognition development, although some progress has been made, compared with major domestic and foreign-language research also apart very far. In order to promote the development of the Uyghur speech recognition, we needs to combine acoustic, linguistic characteristics, analysis the factors restricting correct recognition rate, make the application of new modeling techniques, and study the research experience from the speech recognition technology of large multilingual.

Reference

1. Wushour Silamu, Nasirjan Tursun. Speech Processing Technology of Uyghur Language. Speech Database and Assessments, Oriental COCOSDA International Conference, 10-12 Aug. 2009.

2. I Dawa, Hao huang, Nasirjan, Wushour. An expansion of speech technology in languages and tasks. CN NMYS-NLP2010&MLKD2010, 2010.6, 205-210.

3. Wushour Silamu. Segmentation and Self-Adaptation Techniques in Uyghur Arabic Syllable Speech Recognition. 1st National Conference of Intelligent Machine, 1993.

4. Wushour Silamu, Nasirjan Tursun. HMM-Based Uyghur Continuous Speech Recognition System. CSIE2009, IEEE Computer Society, 31 March-2 April 2009, Los Angeles, California USA, 243-247.

5. Nasirjan Tursun, Wushour Silamu, TAO Mei. Uyghur Continuous Speech Recognition Based on HTK. Recent Advance of Chinese Computing Technologies, Chinese and Oriental Language Information Processing Society, Singapore, 2008.4, 377-380.

6. Nasirjan Tursun, Wushour Silamu. Large Vocabulary Continuous Speech Recognition in Uyghur: Data Preparation and Experimental Results, 2008 6th ISCSLP, IEEE eXpress Publishing, Dec.16-19, 2008, Kunming, China, 197-200.

7. LI Xiaoyang, I Dawa, Wushour Silamu, Yoshinori Sagisaka. Telephone Speech Monitoring System Based on GMM-UBM/SVM for Uyghur Language. Computer Applications and Software, Jan.2012, Vol.29, No.1, 46-48.

КЛАРК Д., ЛАСТОВЕЦКИЙ А., РЫЧКОВ В.

Университетский колледж Дублина, Ирландия

РАЗРАБОТКА И РЕАЛИЗАЦИЯ АДАПТИВНЫХ ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ ДЛЯ НЕОДНОРОДНЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ

В докладе описывается алгоритм оптимального распределения независимых единиц вычисления между неоднородными процессорами, скорость которых характеризуется априори неизвестными функциями объема вычислений, назначенных этим процессорам. Высокая эффективность алгоритма, позволяющая использовать его при расзработке адаптивных программ для выполнения на неоднородных вычислительных системах, демонстрируется на примере задач вычислительной линейной алгебры.

1. Введение

Традиционные алгоритмы распределения вычислений между неоднородными процессорами [1-4] основаны на константной модели производительности (КМП), которая представляет скорость процессора положительным константой. При этом вычисления между процессорами распределеяются пропорционально этим постоянным скоростям. Таким образом, традиционные алгоритмы предполагают, что скорость процессора не зависит от размера вычислительной задачи, решаемой ни этом процессоре. Это предположение, как правило, справедливое при решении вычислительных научных задач среднего размера на неоднородных кластерах рабочих станций, становится неверным в следующих случаях:

- В результате разбиения вычислений, некоторые задачи либо не помещаются главной памяти своего процессора, вызывая подкачку страниц с дисковой памяти, либо, наоборот, полностью умещаются в верхних, более быстрых, уровнях памяти.
- Некоторые вычислительные устройства, участвующие в выполнении программы, являются не традиционными процессорами общего назначения, а, например, ускорителями, такими как графические процессоры или специализированные ядра. В этом случае, относительная скорость традиционного и нетрадиционного процессоров может отличаться для двух разных размеров одной и той же задачи, даже если для этих размеров задача полностью умещается в основной памяти устройств.
- Разные процессоры используют разные программные для локального решения одной и той же вычислительной задачи.

Перечисленные случаи характерны для современных и, в особенности, для перспективных высокопроизводительных неоднородных вычислительных систем. В результате, применимость алгоритмов распределения вычислений, основанных на КМП, становится все более ограниченной. Действительно, чем выше уровень неоднородности и количество различных различных вычислительных устройств, тем уже становится диапазон рамеров вычислительной задачи, для которого их относительные скорости можно считать

постоянными. Дла этих систем нужны новые алгоритмы, способные оптимально распределять вычисления для всего диапазона размеров вычислительных задач.

Функциональная модель производительности (ФМП) [5] намного реалистичней КМП, поскольку учитывает влияие на производительность процессоров многих важных факторов, таких как неоднородность архитектуры и ее прогрммно-аппаратной реализации, неоднородность структуры памяти, эффекты подкачки страниц и т.д. [6-7]. ФМП представляет скорость процессора функцией размера задачи. Скорость определяется как число единиц вычисления, выполняемых процессором в единицу.

Проблема распределения вычислений, использующая ФМП, была сформулирована и решена в [6-7] следующим образом. Общий размер задачи n представляет собой число единиц вычисления, которые нужно распределеить между p ($p << n$) процессорами $P_1,\ldots,P_p$. Скорости процессоров представляются положительными непрерывными функциями размера задачи $s_1(x),\ldots,s_p(x)$: $s_i(x) = x / t_i(x)$, где $t_i(x)$ – это время выполнения x единиц вычисления процессором i. Функции скорости определены в диапазоне $[0, n]$. Алгоритм решения этой проблемы возвращает распределение единиц вычисления, $d_1,\ldots,d_p$, между процессорами так, что $d_1 + d_2 + \ldots + d_p = n$. Нагрузка будет сбалансирована, если все процессоры завершат вычисления одновременно: $t_1(d_1) \approx t_2(d_2) \approx \ldots \approx t_p(d_p)$. Это можно выразить следующим образом:

$$\begin{cases} \dfrac{d_1}{s_1(d_1)} \approx \dfrac{d_2}{s_2(d_2)} \approx \ldots \approx \dfrac{d_p}{s_p(d_p)} \\ d_1 + d_2 + \ldots + d_p = n \end{cases}$$

Геометрически, решение этой системы уравнений, $d_1,\ldots,d_p$, можно представляется финкций скорости прямой линией, проходящей через начало системы координат (рис. 1).

Геометрический алгоритм решения этой проблемы, предложенный в [6-7], суммируется следующим образом. Любая прамая, выходящая из начала системы координат и пересекающая функции скорости представляет оптимальное распределение вычислений для некорого конкретного размера проблемы. Поэтому, пространство решений проблемы распределения вычислений состоит из всех таких прямых линий. Две прямые выбираются в качестве начального приближения. Верхняя прямая представляет оптимальное распределение $d_1^u,\ldots,d_p^u$ для проблемы размером $n_u < n$, $n_u = d_1^u + \ldots + d_p^u$, а нижняя прямая дает решение $d_1^l,\ldots,d_p^l$ для $n_l > n$, $n_l = d_1^l + \ldots + d_p^l$. Область между этими прямыми последовательно делится пополам. На шаге k, размер проблемы, соответствующий новой прямой, пересекающей функции скорости в точках $d_1^k,\ldots,d_p^k$ вычисляется как $n_k = d_1^k + \ldots + d_p^k$. В зависимиости от того, является ли n_k меньше или больше, чем n, эта прямая становится новой верхней или нижней границей. Приближая n_k к n, алгоритм находит оптимальное разбиение проблемы данного размера $d_1,\ldots,d_p$: $d_1 + \ldots + d_p = n$.

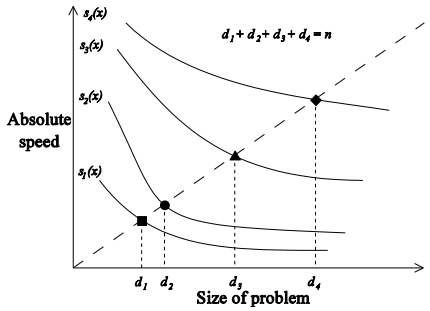


Рис. 1. Оптимальное распределение вычислений, демострирующее геометрическую пропорциональнсть числа единиц вычисления скорости процессора

Алгоритмы решения задач линейной алгебры на неоднородных вычислительных системах, использующие приведенный алгоритм в качестве основного строительного блока, опубликованы в [8-10].

Построение полной ФПМ может быть очень дорогой операцией, поскольку требует выполнения вычислительного ядра для большого числа размеров задачи. Проблема минимизации стоимости построения полных ФМП изучалась в [11], где предложено ее относительно эффективное приближенное решение. Однако, даже в случае оптимального решения этой проблемы стоимость построения полной ФМП будет слишком высокой для того, чтобы использовать алгоритмы распределения вычислений на основе полных ФМП в адаптивных научных прикладных программах.

В докладе описывается другое решение проблемы минимизации стоимости использования ФМП для распределении вычислений между неоднородными процессорами. Вместо использования полных ФМП, описываемый алгоритм строит и использует их частичные приближения, достаточные для оптимального распределения вычислений.

2. Алгоритм распределения вычислений, основанный на частичных ФМП

В математической форме, проблема распределения вычислений между неоднородными поцессорами, которую мы пытаемся решить, выглядит следующим образом:

Дано:

- Множество n независимых и одинаковых единиц вычисления;
- Множество p ($p << n$) процессоров $P_1,\ldots,P_p$, выполняющих x единиц вычисления за время $t_i(x)$;
- Требуемая относительная точность решения, ε.

Найти: разбиение n единиц вычисления между p процессорами, такое что $\sum_{i=1}^{p} d_i = n$ и $\max_{1\le i,j\le p}\left|\frac{t_i(d_i)-t_j(d_j)}{t_i(d_i)}\right| \le \varepsilon$.

Алгоритм решения:

1. Все p процессоров параллельно выполняют n / p единиц вычисления. Время выполнения $t_1(n / p),\ldots,t_p(n / p)$ посылается процессору P_1.

2. IF $\max_{1\le i,j\le p}\left|\frac{t_i(n/p)-t_j(n/p)}{t_i(n/p)}\right| \le \varepsilon$ THEN равномерное распределения вычислений является приемлемым решением и алгоритм останавливается;
ELSE процессор P_1 вычисляет абсолютные скорости всех процессоров, $s_i(n / p) = (n / p) / t_i(n / p)$ для $1 \le i \le p$, и строит первое приближение их ФМП в виде КМП, $s_i(x) = s_i(n / p)$.

3. Применяя геометрический алгоритм распределения вычислений к этому приближению ФМП, процессор P_1 вычисляет новое распределение, $d_1,\ldots,d_p$, и затем посылает каждому процессору P_i новое число единиц вычисления, d_i , которое тот должен выполнять.

4. Процессор P_i , параллельно с другими процессорами, выполняет d_i единиц вычислений, $1 \le i \le p$. Время выполнения $t_1(d_1),\ldots,t_p(d_p)$ посылается процессору P_1.

5. IF $\max_{1\le i,j\le p}\left|\frac{t_i(d_i)-t_j(d_j)}{t_i(d_i)}\right| \le \varepsilon$, THEN текущее распределение, $d_1,\ldots,d_p$, решает проблему и алгоритм останавливается;
ELSE процессор P_1 вычисляет абсолютные скорости, демонстрируемые процессорами для этого распределения, $s_i(d_i) = d_i / t_i(d_i)$ $(1 \le i \le p)$, и использует эти вновь полученные точки функциональных моделей процессоров P_i, $(d_i, s_i(d_i))$, для построения их более точного кусочно-линейного приближения. А именно, пусть $\{(d_i^{(j)}, s_i(d_i^{(j)}))\}_{j=1}^{m}$, $d_i^{(1)} < \ldots < d_i^{(m)}$, - экспериментально полученные точки $s_i(x)$, использованные для построения ее текущего кусочно-линейного приближения. Тогда

- IF $d_i < d_i^{(1)}$, THEN отрезок $(0, s_i(d_i^{(1)})) \to (d_i^{(1)}, s_i(d_i^{(1)}))$ этого приблидения заменяется парой отрезков $(0, s_i(d_{(i)})) \to (d_i, s_i(d_i))$ и $(d_i, s_i(d_i)) \to (d_i^{(1)}, s_i(d_i^{(1)}))$;

○ IF $d_i > d_i^{(m)}$, THEN прямая $(d_i^{(m)}, s_i(d_i^{(m)})) \to (\infty, s_i(d_i^{(m)}))$ заменяется отрезком $(d_i^{(m)}, s_i(d_i^{(m)})) \to (d_i, s_i(d_i))$ прямой $(d_i, s_i(d_i)) \to (\infty, s_i(d_i))$;

○ IF $d_i^{(k)} < d_i < d_i^{(k+1)}$, THEN отрезок $(d_i^{(k)}, s_i(d_i^{(k)})) \to (d_i^{(k+1)}, s_i(d_i^{(k+1)}))$ парой отрезков $(d_i^{(k)}, s_i(d_i^{(k)})) \to (d_i, s_i(d_i))$ и $(d_i, s_i(d_i)) \to (d_i^{(k+1)}, s_i(d_i^{(k+1)}))$.

6. GOTO 3.

Описанный алгоритм динамически строит функциональные модели процессоров в требуемом диапазоне и с требуемой точностью, и возвращает решение, совпадающее или близкое к решению, основанному на полных функциональных моделях. Данные затем распределяются в соответствии с найденным решением и программа затем выполняет вычисления над этими данными. В докладе приводятся примеры применения описанного алгоритма для реализации адаптиных научных прикладных программ для современных неоднородных вычислительных систем, демонстрирущие эффективность предложенного решения. Другие приложения можно найти в [12-15].

Литература

1. A. Kalinov, A. Lastovetsky, Heterogeneous distribution of computations while solving linear algebra problems on networks of heterogeneous computers, in: HPCN Europe 1999, LNCS 1593 (1999) 191-200.

2. O. Beaumont, V. Boudet, F. Rastello, Y. Robert, Matrix Multiplication on Heterogeneous Platforms, IEEE Trans. Parallel Distrib. Syst. 12 (2001) 1033-1051.

3. A. Kalinov, A. Lastovetsky, Heterogeneous Distribution of Computations Solving Linear Algebra Problems on Networks of Heterogeneous Computers, J. Parallel Distrib. Comput. 61 (2001) 520-535.

4. M. Cierniak, M. Zaki, W. Li, Compile-Time Scheduling Algorithms for Heterogeneous Network of Workstations, Computer J. 40 (1997) 356-372.

5. A. Lastovetsky and J. Twamley, Towards a Realistic Performance Model for Networks of Heterogeneous Computers, in: IFIP TC5 Workshop, World Computer Congress, Springer, 2005, pp. 39-58.

6. A. Lastovetsky and R. Reddy, Data Partitioning with a Realistic Performance Model of Networks of Heterogeneous Computers, in: 18th International Parallel and Distributed Processing Symposium (IPDPS 2004), 2004.

7. A. Lastovetsky, R. Reddy, Data Partitioning with a Functional Performance Model of Heterogeneous Processors, Int. J. High Perform. Comput. Appl. 21 (2007) 76-90.

8. A. Lastovetsky and R. Reddy, Data Distribution for Dense Factorization on Computers with Memory Heterogeneity, Parallel Computing 33 (2007) 757-779, 2007

9. A. Lastovetsky, R. Reddy, Two-dimensional Matrix Partitioning for Parallel Computing on Heterogeneous Processors Based on their Functional Performance Models, in: HeteroPar 2009, LNCS 6043, Springer, 2010, pp. 112-121.

10. D. Clarke, A. Lastovetsky, V. Rychkov, Column-based Matrix Partitioning for Parallel Matrix Multiplication on Heterogeneous Processors Based on Functional Performance Models, in: HeteroPar 2011, LNCS 7155, pp. 450-459, 2012.

11. A. Lastovetsky, R. Reddy, and R. Higgins, Building the Functional Performance Model of a Processor, in: 21st Annual ACM Symposium on Applied Computing (SAC 2006), ACM Press, pp. 746-753, 2006.

12. D. Clarke, A. Lastovetsky, V. Rychkov, Dynamic Load Balancing of Parallel Computational Iterative Routines on Highly Heterogeneous HPC Platforms, Parallel Processing Letters. 21 (2011) 195-217.

13. A. Lastovetsky, R. Reddy, V. Rychkov, D. Clarke, Design and Implementation of Self-adaptable Parallel Algorithms for Scientific Computing on Highly Heterogeneous HPC Platforms. Arxiv preprint arXiv:1109.3074 (2011)

14. D. Clarke, A. Ilic, A. Lastovetsky, and L. Sousa, Hierarchical Partitioning Algorithm for Scientific Computing on Highly Heterogeneous CPU + GPU Clusters, 18th International European Conference on Parallel and Distributed Computing (Euro-Par 2012), 2012.

15. Z. Zhong, V. Rychkov, and A. Lastovetsky, Data Partitioning on Heterogeneous Multicore Platforms, in: 2011 IEEE International Conference on Cluster Computing (Cluster 2011), pp. 580-584, 2011.

УДК 519.6

РОМАНОВ В.Г., ИСКАКОВ К.Т., БАЙМУРАТОВА Г.Г.

Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан

РАЗВИТИЕ ИССЛЕДОВАНИЯ ОБРАТНЫХ ЗАДАЧ ЭЛЕКТРОДИНАМИКИ, УПРУГОСТИ В НЕОДНОРОДНЫХ СРЕДАХ И ИТЕРАЦИОННЫЕ МЕТОДЫ ИХ РЕШЕНИЯ

Проект «Развитие исследования обратных задач электродинамики, упругости в неоднородных средах и итерационные методы их решения» направлен на изучение обратных задач для дифференциальных уравнений, описывающих эволюционные процессы деформирования неоднородных сред (связанные с распространением волн, изменением естественных физических полей и конфигурации подземных объектов), на развитие методов их исследования и решения. Суть изучаемых обратных задач заключается в определении переменных коэффициентов, входящих в дифференциальные уравнения и граничные условия. Подобные задачи возникают в механике, ядерной физике, электродинамике, медицине, геофизике, механике горных пород и других естественных науках, и находят приложения в промышленности для решения проблем прогноза динамических явлений, неразрушающего контроля и диагностики состояния объектов различного масштабного уровня. Одной из таких задач, например, является фундаментальная проблема геофизики – изучение недр Земли. Среди других важнейших применений – поиск полезных ископаемых, прогноз землетрясений и горных ударов, медицинская и промышленная томография.

Планируется исследовать новые прикладные обратные задачи, получить оценки устойчивости решения задач, а также найти оценки скорости сходимости численных алгоритмов решения этих задач, основанные на методах граничного управления, оптимизации, использования спектральных свойств матриц и операторов, и метода М.К. Крейна.

Полученные результаты будут использованы в геофизических методах поиска и разведки полезных ископаемых, эксплуатации месторождений, интерпретации результатов исследования скважин.

- **Научная новизна:** доказательство новых теорем единственности и устойчивости решения класса обратных задач для уравнений гиперболического типа; обоснование алгоритмов численного метода их решения; получение оценок скорости сходимости итерационных методов решения класса обратных задач; разработка комплекса прикладных программ и издание научной монографии.

– **Значимость**: эти исследования существенно пополнят общую теорию некорректных задач как в теоретическом плане: вопросы единственности и устойчивости для рассматриваемого класса обратных задач, а также пополнят базу численных методов их решения.

– **Влияние полученных результатов на развитие науки и технологий:** использование полученных результатов в таких областях как: прикладная математика, геофизика, сейсмика, дистанционное зондирование, электродинамика, прикладные задачи археологии (в использовании неразрушающих методов в поисках локализованных археологических объектов геофизическим приборами), а также во многих прикладных задачах естествознания.

– **Ожидаемый социальный эффект:** для исследования рассматриваемых прикладных задач естествознания используются неразрушающие технологии. В задачах по идентификации физических параметров используются геофизические приборы - георадары, основанные на исследовании процессов возбуждаемые электромагнитными источниками.

– **Экономический эффект:** рассматриваемые задачи решаются методами математического компьютерного моделирования, которое не требует дорогостоящего оборудования для проведения экспериментов.

– **Обзор предшествующих научных исследований, проведенных в мире, относящихся к исследуемой теме и их взаимосвязь с настоящим Проектом:** - Основы теории обратных и некорректных задач были заложены академиком А. Н. Тихоновым. В дальнейшем по данному направлению сформировалось несколько научных школ, в том числе школа сибирских ученых под руководством академика М. М. Лаврентьева, которая получила широкое признание, как в нашей стране, так и за рубежом. Работы М.М. Лаврентьева, его сподвижников и их учеников создали новое научное направление в математической физике, имеющее большое теоретическое и прикладное значение. Центральным объектом исследований являются обратные задачи, связанные с дифференциальными уравнениями, описывающими процессы распространения волн (упругих, электромагнитных и т.д.). Под обратными задачами здесь понимаются задачи определения переменных коэффициентов дифференциальных уравнений по информации о решениях этих уравнений. Как правило, эта информация задается на границе некоторой области, внутри которой определяются искомые коэффициенты. Обратные задачи возникают при интерпретации данных геофизических измерений в сейсморазведке, электроразведке, и рентгеновской томографии, геотомографии, ЯМР-томографии. Методы исследования обратных задач включают как необходимый элемент детальное исследование свойств решений обычных начально-краевых задач математической физики. Последние задачи принято называть прямыми задачами.

Ниже излагаются основные предшествующие научные результаты, полученные в сибирской научной школе, которая стояла у истоков этого научного направления. Теория многомерных обратных задач систематически начала изучаться сибирскими учеными с середины 60-х годов 20-го столетия. Ими обнаружена тесная связь обратных задач для гиперболических уравнений с новыми задачами интегральной геометрии на семействах геодезических и римановых эллипсоидах. На этой основе получены теоремы единственности и оценки устойчивости решений обратных задач для гиперболических уравнений второго порядка, систем уравнений упругости, электродинамики и электроупругости. Ими разработаны метод исследования локальной разрешимости обратных задач в классах функций, аналитических по части переменных. Этот метод использован для обоснования численных алгоритмов решения ряда обратных задач. Исследована задача определения римановой метрики внутри некоторой ограниченной области через расстояния между точками границы области (обратная кинематическая задача сейсмики) и ее линейный вариант (задача интегральной геометрии). Найдены оценки устойчивости решения этих задач и разработаны алгоритмы численного решения соответствующих задач. Сибирскими учеными исследована сходимость численных методов решения обратных и некорректных задач. В частности, для методов Ньютона-Канторовича, итераций Ландвебера, градиентных методы минимизации целевого функционала найдены оценки скорости их сходимости, рассмотрены два из важнейших направлений развития теории обратных и некорректных задач: 1) оценки условной устойчивости решения, 2) оценки скорости сходимости

градиентных методов. Уточнены оценки скорости сильной сходимости градиентных методов решения сильно некорректных задач (задачи Коши для уравнения Лапласа и для параболического уравнения с обратным временем и др.), для которых ранее удавалось оценить лишь скорость сходимости по функционалу. На основе полученных оценок, исследованы градиентные методы решения некорректных задач Коши для уравнений гиперболического, параболического и эллиптического типов. Сформулировано новое правило выбора номера остановки итераций в градиентных методах – правило условной устойчивости. Ими же предложены метод численного решения уравнений теории упругости и уравнений Максвелла для горизонтально-слоистых анизотропных сред. Решение систем сводилось к решению дифференциального матричного уравнения Риккати, которое в каждом слое имеет аналитическое решение. Представленный численный метод решения отличается от существующих аналогов и может быть применен для слоистой среды с любым видом анизотропии. На основе предложенного метода решения прямой задачи численно решена обратная задача для системы упругости по определению параметров среды пачки тонких анизотропных слоев. Исследованы случаи орторомбической и трансверсально-изотропной симметрии. Приведен один из возможных конкретных путей численного решения поставленной задачи. Теория обратных и некорректных задач успешно развивалась в Москве - Тихоновым А.Н. и его учениками: Бакушинским А.Б., Морозовым В.А., Денисовым А.М., Яголой А.Г., в Санкт-Петербурге (Ленинграде) - Благовещенским А. С., Белишевым М.И., в Екатеринбурге – Ивановым В.К., Васиным В.В., Тананой В.П. Подобные исследования ведутся также в США, Франции, ФРГ, Австрии, Швеции, Италии, Японии, Китае и ряде других стран. Предлагаемый проект по своим целям, задачам и предполагаемым результатам соответствует мировому уровню исследований. Планируемые исследования связаны с теорией обратных задач для дифференциальных уравнений. В этом направлении у коллектива имеется хороший научный задел: более 400 научных публикаций и около 20 монографий. Авторами проекта (и консультантами) предложены и развиты методы исследования обратных задач, основанные на редукции их к задачам интегральной геометрии, операторным уравнениям вольтеровского типа, интегро-дифференциальным уравнениям и изучении последних методами функционального анализа, обоснован метод линеаризации. Адаптирован к теории обратных задач целый ряд методов численного решения, найдены оценки их сходимости. Многие из полученных результатов имеют приоритетный характер. Имеющиеся результаты дают хорошую основу для успешного выполнения проекта

– **Цель Проекта**: Проект направлен на изучение обратных задач для дифференциальных уравнений, описывающими процессы распространения волн (упругих, электромагнитных и т.д.) в неоднородных средах, на развитие методов их исследования и решения.

Суть изучаемых обратных задач заключается в определении переменных коэффициентов, входящих в дифференциальные уравнения. Как известно, искомые коэффициенты характеризуют свойства среды, в которой происходит распространения волн. Подобные задачи возникают в геофизике, механике, ядерной физике, электродинамике, медицине и других естественных науках.

Типичная постановка обратной задачи заключается в том, что на границе рассматриваемой области задается информация о решении краевой задачи, требуется по этой информации найти коэффициенты дифференциального уравнения внутри области. Давно было замечено, что решения многих наиболее интересных для приложений обратных задач являются неустойчивыми в классическом смысле. Для их устойчивости необходимо привлекать дополнительную априорную информацию о классе решений. Эта информация должна учитываться как при математическом исследовании обратных задач, так и при построении численных алгоритмов их решения. Разнообразие априорной информации определяет и разнообразие постановок обратных задач. Для одного и того же дифференциального уравнения существуют, как правило, многочисленные постановки

обратных задач, которые зачастую сильно различаются методами их исследования. И в этом контексте, есть достаточно много важных прикладных недостаточно полно изученных обратных задач, интересных и в математическом отношении. Для их исследования авторы проекта предполагают использовать ранее развитые ими методы. Один из этих методов основан на сочетании асимптотического разложения решения в окрестности волнового фронта, оценки продолжения решения задачи Коши с границы области, а затем оценки искомых коэффициентов через значения решения и его производных на волновом фронте. Среди других методов: использование априорных оценок, интегральных уравнений, карлемановских оценок, аппарат функционального анализа, интегральных преобразований.

- **Описание методов исследования**: Методы исследования обратных задач разработаны и представлены в ряде монографий. Для численного решения обратных задач будут использованы разработанные ранее варианты методов градиентного спуска, в сочетании с методом Ньютона-Канторовича и использования спектральных свойств матриц и операторов. Планируется получить оценки устойчивости решения новых постановок обратных задач, разработать численные методы их решения. Попутно будут исследованы и необходимые свойства решений соответствующих прямых задач. В частности, будут рассмотрены следующие задачи:

- Изучение разрешающих способностей обратной задачи электроразведки об одновременном определении диэлектрической проницаемости и проводимости среды, а также только диэлектрической проницаемости при известной проводимости.

- Построение новых алгоритмов и комплексов программ численного решения двумерных и трехмерных обратных задач упругости и электродинамики на основе методов Ньютона-Канторовича, М.К. Крейна, граничного управления и конечных разностей, проведено исследование скорости сходимости алгоритмов.

- Выполнение исследований условной устойчивости и разрешающей способности обратных задач геоэлектрики и упругости на основе использования спектральных свойств матриц и операторов, построены новые алгоритмы решения и получены новые оценки скорости их сходимости, предложены способы выбора параметра регуляризации.

Планируется применение аппарата использования спектральных свойств матриц и операторов к исследованию коэффициентных обратных задач геоэлектрики и упругости. С этой целью на первом этапе будет проведено исследование линеаризованных постановок. На втором этапе планируется использовать полученные результаты для анализа сходимости метода Ньютона-Канторовича. Будет проведено теоретическое и численное исследование разрешающей способности алгоритмов для типичных моделей сред.

Планируется разработка новых комплексов программ для выполнения расчетов.

- **Управление рисками**: Для *управления рисками* с целью снижения отклонений фактических показателей проекта от их запланированных значений будет разработан комплекс мероприятий, включающих идентификацию, анализ, снижение и мониторинг рисков. Для анализа риска будут использованы средства ресурсного планирования Open Plan программного продукта Project Expert, который позволит управлять всеми видами ресурсов: людьми, оборудованием, материалами, финансами.

Научным консультантом Проекта является член-корреспондент Российской академии наук, профессор Романов В.Г.

Литература

1. Tanana V. P., Bokov A. V. “On estimating the precision of an approximate solution to an inverse thermodiagnostics problem with free boundary”, *J. Appl. Industr. Math.*, 5:1 (2011), 104–109.

2. Серебрянский С. М. “Об оценках погрешности методов приближенного решения одной обратной задачи”, *Сиб. журн. индустр. матем.*, 13:2 (2010), 135–148.

3. Танана В. П. , Боков А. В. “Об оценке точности приближенного решения одной обратной задачи тепловой диагностики с подвижной границей”, *Сиб. журн. индустр. матем.*, 13:1 (2010), 133–139.

4. Денисов А.М. обратные задачи для квазилинейного гиперболического уравнения в случае движущейся точки наблюдения. Дифференциальные уравнения. 2009 г., т. 45, № 11, стр.1543-1553.

5. Денисов А.М., Захаров Е.В., Калинин А.В. Калинин В.В. Численные методы решения некоторых обратных задач электрофизиологии сердца. Дифференциальные уравнения. 2009 г., т. 45, № 7, стр.1014-1022.0

УДК 004.

ШАРИПБАЕВ А.А.

Евразийский национальный университет им. Л.Н. Гумилева, Астана, Казахстан

КОНЦЕПЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ ЭНЕРГОСИСТЕМЫ КАЗАХСТАНА

Известно, что в жизни человека энергия является самым важным фактором существования. Становление человеческого общества напрямую связано с характером использования энергии: огонь, физическая сила, ветряные и водяные мельницы, паровые машины и электрогенераторы. Социально-экономическое развитие любой страны неотделимо от развития энергетики, которая определяет и решает проблемы создания и использования энергии.

В настоящее время в мире происходят большие перемены. Неуклонно растёт численность населения Земли и объем удельного потребления энергии. Сокращаются запасы основного источника энергии - углеводорода. Производство энергии негативное воздействует на окружающую среду, происходят изменение климата и различные природные аномалии.

Человечество стоит перед лицом глобальных вызовов, одним из которых является концепция «умная энергосистема» - SmartGrid, которая позволит повысить надежность энергоснабжения потребителей, а также снизить энергопотери и расход энергоресурсов.

В управлении электросети возникает много задач, которых должен решать оперативный персонал. К ним относятся следующие задачи, решаемые на основе онлайнового анализа топологии электросети и онлайновой диагностики нештатных ситуаций в электросети [1]:

1. Мониторинг состояния электрической сети и ее элементами.
2. Контроль и координация действий оперативного персонала.
3. Контроль загрузки работающего оборудования.
4. Сбор оперативной информации при технологических нарушениях.
5. Поддержание уровней напряжения в контрольных пунктах точек сети.
6. Подготовка технологических режимов для обеспечения возможности вывода оборудования из работы.
7. Долгосрочное и краткосрочное планирование ремонтов линий электропередачи, оборудования и устройств объектов электросетевого хозяйства.
8. Выполнение функций режимной проработки диспетчерских заявок на вывод из работы элементов и оборудования.
9. Разработка мероприятий по снижению технологических потерь в сетях с новыми элементами управления.
10. Контроль и выполнение технических мероприятий, обеспечивающих безопасное производство работ на линии электропередачи, оборудовании и устройствах объектов электросетевого хозяйства.

11. Контроль выполнения согласованного диспетчерского графика перетоков электроэнергии по каждому контролируемому сечению.

12. Производство расчетов режимов электрической сети с новыми элементами управления.

13. Регулярное обучение и тренажерная подготовка оперативного персонала на специализированных тренажерах.

14. Выдача разрешений на подготовку рабочих мест и допуск ремонтного персонала на линии электропередачи, оборудование подстанций и учет работающих бригад.

Эти задачи характеризуются важностью и трудоемкостью для персонала, который должен принимать решения в условиях дефицита времени, при неполной оперативной информации, оперируя большими объемами нормативной информации. В этих условиях последствие «человеческих» ошибок могут быть весьма тяжелыми. Поэтому разработка и внедрение интеллектуальных информационных систем управления электросетью является актуальной проблемой современного общества. Эти системы, работая в режиме диалога с персоналом, обеспечат разгрузку персонала, правильно (безошибочно), своевременно и эффективно решить возникшие задачи.

Известно, что любую информационную систему (ИС) можно рассматривать как фабрику, производящую информацию, в которой: сырье - база данных, заказ – информационный запрос, продукт – выходные данные (требуемая информация), технология – программы обработки данных, оборудование – компьютеры. При этом если в ходе эксплуатации ИС выяснится потребность в модификации одного из компонентов программы, то возникнет необходимость ее переписывания, так как полным знанием проблемной области обладает только разработчик ИС, а программа служит “недумающим исполнителем” знания разработчика. Кроме того, ИС имеет простые интерфейсы в виде информационных запросов и сообщений. Эти недостатки устраняются в интеллектуальных информационных системах.

Интеллектуальная информационная система (ИИС) - это ИС, которая основана на концепции использования базы знаний (базы данных + правила вывода) для генерации программ решения задач различных классов в зависимости от конкретных информационных потребностей пользователей. База знаний отличается от базы данных возможностью выборки по запросу информации, которая может явно не храниться, а выводиться из имеющейся информации в базе данных. Примером такого запроса может быть “Вывести список клиентов, потребность в электроэнергии которых выше среднеудельной потребности”.

Для ИИС характерны следующие признаки: *интеллектуальные интерфейсы*; *умение решать сложные плохо формализуемые задачи*; *способность к самообучению*.

Интеллектуальность интерфейса подразумевает возможность поиска неявной информации в базе данных или тексте для произвольных запросов, составляемых на ограниченном естественном языке, возможно с использованием речевой технологии.

Умелость решать сложные плохо формализуемые задачи означает возможность построения оригинального алгоритма в зависимости от конкретной ситуации, для которой могут быть характерны неопределенность и динамичность исходных данных и знаний.

Способность к самообучению означает наличие возможности автоматической классификации примеров на основе обучения, которые накапливаются за некоторый период и составляют обучающую выборку. Обучающая выборка может быть “с учителем”, когда для каждого примера задается в явном виде значение признака его принадлежности некоторому классу ситуаций, или “без учителя”, когда по степени близости значений признаков классификации система сама выделяет классы ситуаций.

В результате самообучения автоматически строятся обобщенные правила или функции, определяющие принадлежность ситуаций классам, которыми можно воспользоваться при интерпретации новых возникающих ситуаций. Таким образом, по мере накопления опыта реальных ситуаций автоматически формируется и корректируется база

знаний, используемая при решении задач классификации и прогнозирования. Это позволяет сократить затраты на ее создание и обновление.

Действующие электросети страны не подходят по разным причинам, чтобы создать Интеллектуальную Энергосистему Единой Электрической Сети Республики Казахстан (ИЭЕСРК). Для устранения этих причин необходимо решить следующие задачи:

1. Обоснование необходимости и стоимости разработки требований и технических решений к объектам ИЭЕСРК на основе международных стандартов технического регулирования объектов и процессов электроэнергетики;

2. Обоснование необходимости и стоимости строительства цифровых подстанций, укомплектованных интеллектуальными вторичными оборудованиями, работающими на едином стандартном протоколе обмена информацией IEC 61850.

3. Обоснование необходимости и стоимости создания единой интеллектуальной информационной системы мониторинга и управления национальной электрической сетью Республики Казахстан;

4. Обоснование необходимости и стоимости создания системы электронного обучения (e-learning) по ИЭЕЭС на основе Web-технологии.

Разработка требований и технических решений к объектам ИЭЕЭС РК состоит из трех этапов:

1-этап. Разработка требований и технических решений к оборудованиям ИЭЕЭС РК;

2-этап. Разработка требований и технических решений к услугам ИЭЕЭС РК;

3-этап. Разработка требований и технических решений к аппаратно-программным средствам ИЭЕЭС РК;

Строительство цифровых подстанций должно осуществляться в три этапа:

1-этап. Установка высоковольтных цифровых измерительных оптических трансформаторов тока и напряжения, многофункциональных приборов измерений и учета, станционной шины и шины процесса, системы синхронизации, новой системы отображения и управления подстанцией.

2-этап. Установка терминалов релейной защиты и автоматики, использующие в качестве входных сигналов токов и напряжений цифровые информационные потоки, соответствующие международному стандарту IEC 61850-9.2.

3-этап. Строительство экспериментальной цифровой подстанции.

Создание единой интеллектуальной информационной системы мониторинга и управления ИЭЕЭС РК состоит из трех этапов:

1-этап. Обследование состава и структуры технологических и бизнес-процессов ИЭЕЭС РК, определение информационных потоков;

2-этап. Онтологическое моделирование и построение формальных спецификаций информационных, функциональных и экспертных подсистем;

3-этап. Проектирование базы данных, разработка и испытание программного обеспечения интеллектуальной информационной системы мониторинга и управления ИЭЕЭС РК;

ИЭЕЭС РК будет электроэнергетической системой нового поколения, основанной на мультиагентном принципе организации и управления ее функционированием и развитием с целью обеспечения эффективного использования всех ресурсов (природных, социальных, производственных и человеческих) для надежного, качественного и эффективного энергоснабжения потребителей за счет гибкого взаимодействия всех ее субъектов (всех видов генерации, электрических сетей и потребителей).

ИЭЕЭС РК как технологическая инфраструктура будут предоставлять возможность создания единого информационного пространство электроэнергетики на основе современных технологических средств и интеллектуальной информационной системы мониторинга и управления.

В ИЭЕЭС РК должны быть следующие интерфейсы информационного взаимодействия:

№	Наименование интерфейса	Типовые элементы, обладающие данным интерфейсом	Информационные протоколы
1	Объект технологического управления	Подстанция (генератор, распределительная сеть, крупный потребитель)	Сервер IEC 61850-90-1; Сервер IEC 62445-2; Сервер IEC 61850-90-5; Сервер МЭК 60870-5-104
2	Субъект технологического управления	Диспетчерский центр; Центр мониторинга	Клиент IEC 61850-90-1; Клиент IEC 62445-2; Клиент IEC 61850-90-5; Клиент МЭК 60870-5-104
3	Объект операционной деятельности	Информационные порталы, обеспечивающие финансовую, коммерческую и административно-хозяйственную деятельность субъектов ИЭС ААС	WEB сервисы (IEC 62541)
4	Субъект операционной деятельности	Организации, задействованные в ИЭС ААС; Малые генераторы (включая DER) и потребители электроэнергии	WEB клиент (IEC 62541)

ИЭЕЭС РК должна выполнять следующие функции реального времени:

- оценка состояния с учетом новых элементов управления;
- контроль пределов и загрузки оборудования;
- динамическая модель энергосистемы с элементами адаптивного управления;
- советчик диспетчера-тренажер;
- советчик диспетчера по инструкциям;
- прогноз потребления;
- расчет потерь с учетом элементов адаптивного управления;
- управление ремонтами с учетом новых элементов управления;
- проверка по критериям надежности;
- человеко-машинный интерфейс;
- управление видеоэкраном;
- установка диспетчерских пометок;
- оптимизация электрического режима с учетом новых элементов управления в рамках компетенции.

ИЭЕЭС РК должна содержать следующие комплексы программ, ориентированных на непосредственное управление электрической сетью:

- управление переключениями, блокировка неправильных переключений;
- тренажер оперативных переключений с учетом новых элементов управления;
- анализ нештатных ситуаций на подстанциях;
- долгосрочная оптимизация ремонтов с учетом новых элементов управления;
- диспетчерский журнал;
- формирование отчетов об эффективности использования новых элементов управления;
- анализ аварийных ситуаций в сетях;
- оптимизация управлением ремонтными бригадами.

ИЭЕЭС РК в первую очередь должна обеспечить доступ персонала к необходимой информации, которая сосредоточена в вспомогательных БД. Разрозненность информации затрудняет интеграцию разнообразных задач и усложняет развитие и техническое обслуживание комплекса. Использование CIM стандарта представления текущей, архивной и нормативно-справочной информации позволяет, путем серьезных усилий, решить проблему несовместимости задач в системе.

От реализации ИЭЕЭС РК ожидаются следующие экономические эффекты:

- снижения потерь электроэнергии;

– сглаживания графиков нагрузки;

– повышения пропускной способности линий электропередачи и выдача мощности «дешевой» генерации;

– снижения вероятности системных аварий;

– снижения недоотпуска электроэнергии потребителям;

– снижения требуемого резерва мощности и снижения необходимого;

– прироста установленной мощности электростанций.

Интеллектуальная информационная система мониторинга и управления ИЭЕЭС РК должна иметь:

– модульную структуру, позволяющую последующее добавление модулей по числу добавляемых объектов автоматизации и функций без изменения уже имеющихся;

– открытую архитектуру, позволяющую добавлять аппаратно-программные средства диспетчеризации (контроля и управления параметрами) вновь устанавливаемого электрооборудования;

– возможность интегрироваться с другими системами мониторинга и управления энергосетью.

В результате внедрения ИЭЕЭС РК позволит анализировать и прогнозировать изменение состояния сети и получить следующие эффекты:

1. *Повышение энергетической безопасности.* Повышение надежности энергоснабжения потребителей, в том числе в сфере «цифрового» спроса, повышение уровня локальной энергонезависимости;

2. *Повышение производительности и безопасности труда.* Снижение количества персонала и объемов участия человека в эксплуатации и обслуживании технических устройств;

3. *Оптимальное распределение электроэнергии.* Возможность распределения электроэнергии в нормальном и аварийных режимах (отсутствие нулевых перетоков и перегрузки в различных режимах сети) в зависимости от различных критериев;

4. *Улучшение условий для экономической интеграции и конкуренции.* Снижение существующих инфраструктурных и информационных барьеров для объединения рынков, формирование массового активного потребителя на оптовом и розничном уровнях;

5. *Увеличение пропускной способности.* Применение цифровых технических средств в узлах электрической сети позволяет увеличить пропускную способность и как следствие повысить уровень надежности электроснабжения потребителей без существенных капитальных затрат на сооружение новых объектов электросетевого комплекса.

6. *Инновационный импульс для экономики.* Массовый спрос на инновационные продукты энергомашиностроения, электротехнической промышленности, информационных и коммуникационных технологий.

7. *Снижение экологической нагрузки.* Снижение выбросов парниковых газов, прочих загрязняющих веществ, электромагнитного излучения, отчуждаемой площади под энергетические объекты.

Литература

1. Филатов А.А. Обслуживание электрических подстанций оперативным персоналом. Энергоатомиздат, Москва, 1990.

2. Дьяков В.Ф., Любарский Ю.Я., Моржин Ю.И., и др. Интеллектуальные информационные системы в управлении эксплуатацией электроэнергетического комплекса. МЭИ, Москва, 1995.

КОВТУН С.В.

Директор по развитию EPAM Systems

ИНФОРМАТИЗАЦИЯ ОБРАЗОВАНИЯ ЗА СЧЕТ ТЕСНОЙ ИНТЕГРАЦИИ МЕЖДУ ВУЗАМИ И ПРОИЗВОДСТВЕННЫМИ ПРЕДПРИЯТИЯМИ

Развитие казахстанского содержания является важным и актуальным на современном этапе реализации индустриальной политики. Это задача не только государства, но и всех компаний Республики Казахстан, любой формы собственности, включая иностранные компании.

Основополагающим фактором роста Казахстанского содержания является наличие отечественных квалифицированных кадров. Рост отечественного содержания можно обеспечить несколькими путями. Однако, в такой наукоемкой и в то же время динамично развивающейся отрасли как создание программного обеспечения основными поставщиками отечественных кадров становятся высшие учебные заведения. В то же время классическое ВУЗовское образование на сегодняшний день зачастую не поспевает за самыми современными тенденциями рынка и технологическими инновациями, а теоретические знания и небольшой опыт, полученные в стенах большинства отечественных ВУЗов не могут дать нужных практических навыков для начала эффективной работы выпускника в современной производственной компании. Таким образом, выпускник ВУЗа, пришедший в современную компанию вынужден дополнительно еще многому учиться. При этом, эффективной работы на уровне начинающего специалиста он в компании достигнет не ранее года, а в некоторых случаях и двух лет.

Данная ситуация не устраивает ни компанию, которая получает квалифицированного специалиста спустя долгий период переподготовки, ни ВУЗ, который заинтересован в выпуске востребованных на рынке труда специалистов, ни самого выпускника, который хочет получить современное и актуальное образование и эффективно проявить себя сразу после окончания ВУЗа. А в общем это снижает динамику роста рынка труда квалифицированных специалистов и рост отрасли в целом.

Наиболее очевидным решением данной проблемы является более тесная интеграция ВУЗов и производственных предприятий с целью создания совместных учебных программ. Одним из успешных примеров такой интеграции является совместный с ВУЗами учебный процесс, организованной компанией EPAM Systems. Компания EPAM Systems является одним из крупнейших производителей программного обеспечения в Восточной Европе, имеет штат более чем 7500 человек в 30 офисах, расположенных в Белоруссии, Украине, России, Казахстане, нескольких странах Западной и Восточной Европы, США. Учебная программа была запущена несколько лет назад и сейчас она охватывает все крупнейшие ВУЗы Белоруссии, некоторые ВУЗы на Украине и в России. В Казахстане программа реализуется совместным участием с Карагандинским Государственным Техническим Университетом.

Программа состоит из 3-х основных этапов подготовки студентов, начиная от старших курсов и заканчивая молодыми специалистами, работающими в проектных командах компании. Основными принципами программы является обучения студентов наиболее современным и востребованным на рынке технологиям, участие в реальных проектах, последовательность обучения от основ до особенностей специфических технологий.

На первом этапе студенты старших курсов изучают основы современных технологий. Для организации процесса обучения на территории ВУЗа создается лаборатория, которая

оснащается необходимым оборудованием, вычислительной техникой, программным обеспечением. Схема обучения в виде тренингов, состоящих из теоретических лекций и практических работ, проходящих несколько раз в неделю. Общее время обучения от полутора до трех месяцев, в зависимости от специализации. По специализации, в настоящее время проводятся тренинги:

- по разработке и программированию на основе Java;
- по разработке и программированию на основе технологий Microsoft и платформы .Net;
- по тестированию программного обеспечения;
- по внедрению и сопровождению информационных систем.

Учебная группа состоит из 12-17 человек. Перед стартом обучения проводиться отбор из числа желающих. К желающим обучаться предъявляются первичные требования по знанию основ современных информационных систем, языков программирования в рамках программы ВУЗа.

Важным моментом является то, что преподавание ведут сотрудники ВУЗа из числа преподавательского состава. Для подготовки преподаватели ВУЗа сами проходят весь тренинг. Для этого применяется удаленное обучение с использованием Интернета, средств голосового общения, видеоконференции, совместное удаленное использование ресурсов персональных компьютеров.

Во многом, залогом успешного прохождения данных тренингов являются учебные материалы, используемые в процессе обучения. Материалы готовятся наиболее квалифицированными преподавателями, вовлеченными в программу в различных ВУЗах, а так же ведущие специалисты компании, которые привносят свой богатый опыт, знания и практический взгляд.

По окончании тренингов студенты выполняют индивидуальное проектное задание, в решении которого они могут показать успешность усваивания материала и навыки его практического использования. Студентов, наиболее успешно закончивших тренинг, приглашают на второй этап обучения.

Второй этап имеет целью углубить знания, полученные на тренингах в лаборатории, научиться применять их на практике, а так же научиться работать по проектным принципам в составе проектных команд. Форм организации может иметь несколько. Это может быть организация производственной или преддипломной практики, создание филиалов кафедр на предприятии, организация тренингов и курсов повышения квалификации. Суть остается одинаковой. Принцип заключается в формировании обучающихся команд, где каждая из команд реализует определенный проект по созданию программного обеспечения. Команда организуется по проектному принципу. Во главе проекта обычно находится квалифицированный сотрудник компании из производства. Важным условием является то, что данный проект имеет целью создание реального программного обеспечения, которым будут пользоваться реальные люди в жизни. Таким образом, есть заказчик со своими требованиями, условиями реализации, функционирования и использования. Это дает каждому участнику проекта чувство ответственности за результаты своего труда и требования довести свою работу до состояния реального использования. В качестве таких учебных проектов могут использоваться задачи, формируемые ВУЗом или социальные проекты, реализуемые компанией, где компания не усматривает получение прибыли.

Поскольку команды организованы по проектному принципу, участники узнают основные принципы проектной организации труда, основные стадии реализации проекта, учатся принципам общения и коммуникаций в проектах. При создании команд участники подбираются таким образом, что бы при реализации проекта они использовали свои знания, полученные на тренингах первого этапа, при этом углубив их с точки зрения изучения специфических особенностей технологий. Например, студент, изучавший на первом этапе Java, на втором этапе будет участвовать в проектах, в которых используется тот или иной

J2EE сервер. В ходе реализации проекта, руководитель проводит дополнительные тренинги по особенностям применяемых в проекте технологиях.

На втором этапе обучение проходит от 3-х до 5-ти месяцев. Срок очень сильно зависит от индивидуальных способностей обучающегося. Успешно проявивших себя на данном этапе компания приглашает на работу в штат. Зачастую это совпадает с временем окончания студентами ВУЗа.

Третий этап подразумевает повышение квалификации молодых специалистов, которые уже, будучи в штате компании принимают участие в реальных производственных проектах. Основой третьего этапа является так называемая менторская программа. В ней используются принципы наставничества. Определяются наставник (ментор) и один-два молодых специалиста (менти). Ментор использует ранее разработанные программы и материалы для того, что бы молодые специалисты, с которыми он работает, усвоили их в определенный срок и могли использовать в своей производственной деятельности. Для управления всем процессом менторинга из производства выделяется наиболее квалифицированный специалист, который наблюдает за прогрессом всех групп, организует подготовку материалов для тренингов по определенным тематикам. Тренинги проводятся одним из менторов для всех молодых специалистов, участвующих в программе по данной специализации. Цель и задача менторской программы повысить квалификацию молодых специалистов до среднего квалификационного уровня принятого в компании. Оценка успешности проводиться в рамках общекорпоративного регулярного процесса ассесмента. Срок менторской программы варьируется от полугода до 9 месяцев в зависимости от специализации и динамики успеваемости.

Кроме основной трехэтапной программы сотрудничества хочется отметить следующие желательные мероприятия, которые могут повысить общее качество образования и сблизить классическую ВУЗовскую программу обучения с реальными рыночными требованиями:

– участие наиболее квалифицированных специалистов производственных предприятий в создании учебных материалов для основных и специализированных предметов в рамках ВУЗовского процесса обучения;

– организация курсов повышения квалификации или семинаров по обмену опытов среди преподавателей ВУЗов на базе производственных предприятий.

При должной организации и достаточных усилиях упомянутые выше совместные процессы обучения и мероприятия дадут возможность быстро, но в то же время планово, прогнозируемо и управляемо готовить квалифицированных специалистов со знаниями и умениями, наиболее востребованными в современных условиях.