híradástechnika

VOLUME LVII.

2002/12

Selected papers

Mobile Systems

Call Admission Controll

Picture Transmission and Processing

Test Systems

Scientific Association for Infocommunications

Contents

Foreword

As usual preparing the December issue we look over the papers of the recent semester to select the articles for our English version. During this work we got dubious feelings. The researchers and developers are working continuously to make use of the new results of the basic or theoretical sciences and to fulfil the demands of the present and future costumers. In spite of these it does not reflect completely the development trends in telecommunications. Namely the Hungarian development is governed mostly by the orders of the international companies having subsidiaries in Hungary. Partly they have laboratories working with talented young Hungarian research fellows partly they co-operate with different department of the University. They results, the methods or systems are important and new but do not cover the whole spectrum of infocommunication so we feel the want of completeness.

The mobile service providers try to broaden the access to new services and to decrease the tariffs. They do research intensively to reach their goal, incurs of it sometimes they have some theoretical results, too. The practical tests lead also to new solutions. These topics were represented in a large proportion in the last 5 Hungarian issues of our journal. So it was simple to select the description of $\binom{2}{2}$ worldwide new, interesting results. The studies discussing the novelties in mobile telecommunications are leading the December issue.

The measure of congestion and the methods reducing the congestion in democratic, packet switched IP based network is also a popular research topic. Here we had also a great choice namely a lot of research work dealt with methods achieving a guaranteed QoS level applying grouping or differentiating the users. The circuit reservations seems to be a successful method, one of its recent realization can be the MPLS. The CAC (Call Admission Control) is an other possible way to reduce the congestion but it can be resulted in the loss of several calls. From this field we are publishing $\binom{2}{2}$ theoretically also appealing papers which are really actual.

Thinking more deeply on these topic come up the question that the WDM (Wave Division Multiplexing) on optical fibre will offer cheap broadband transmission possibility. So in the future important real-time information can be transmitted on fix, message associated, (or in other words circuit switched) channels. It would be the return of some old methods. In this case the convergence would be limited on common components and digitalisation but the different information should be handled in a way, which can optimally matched to the information. The synergy of information would not be hurt by this concept, and in the same time the network could be utilized more economically, namely the high amount of traffic does not reasons the common manipulation of different type of information.

The enhancement of accessible bandwidth and the decreasing price of a unit of information are increasing the possibility of the transmission of a great number of programs. It will be advantageous for the customers, and is a good business for network-owners because they can make use 10 Gbit/sec on each wavelength, which results 1 Tbit/sec on a single fibre and more than 20 Tbit/sec on a cable. This solution is an attractive possibility in the near future therefore it is also represented here by 2 articles.

We suppose that the WDM introduction and the development of items and systems to it will be determined success of telcos in the coming years. It's a pity but our development staffs is not interested in it. So in this year we had not a single new, original paper about this filed, which would be interested abroad too. We have the feeling that it would be good to encourage the young engineers to make steps in that direction. Without industrial background the development of photonic components will not have results, but the application of them have started so to operate, buy and measure photonic devices is inevitable.

So we hope that in the future these topics will not be neglected, and our next English version will contain also such or similar results.

After all 2 paper deals with test methods. One of the developed results is useful in the software testing the other is interesting in defining the material characteristics. It illustrates a further application of optical fibres. The closing paper is analysing the actual economic situation and one of its interesting symptom. Often we can read about the fusion of well-known infocom companies but the background of them is neither clear nor evident. The telecom community is saying sorrow fully good-bye to the world wide famous companies, and the number of competitor is diminishing.

Our summary on the technical and economic situation does not reflect all sector of the development and is not able to enlighten all the steps done to solve the existing problems. This is only a simplified picture that can have some distortion.

Dr. György Lajtha

Security of Mobile Ad Hoc Networks

ROLAND GÉMESI[1] gemesi@alpha.ttt.bme.hu
BALÁZS IVÁDY[1] ivady@alpha.ttt.bme.hu
LÁSZLÓ ZÖMBIK[2] laszlo.zombik@eth.ericsson.se

[1] Budapest University of Technology and Economics, Department of Telecommunications and Telematics
[2] Ericsson Hungary, ResearchLab

Reviewed

1. Introduction

In contrast with traditional networks, ad hoc networks not require any previously built infrastructure, they are distributed and fully self organized systems. In such a network there should not need special roles for control and management.

Mobile ad hoc networks have the following special attributes. Having dynamic topology means that the architecture of the network is not static, so routes can be valid just for a short time. Moving or disappearing nodes should not disturb the operation of the services. Usually nodes are small handheld devices with a limited capacity of resources (like battery power, CPU or memory). Wireless links usually have much lower bandwidth and these resources are often shared and limited. Moreover links and devices are more vulnerable than in fixed, wired networks.

Additional threats exist for mobile ad hoc environments due to their distributed conception. In the rest of this paper we will examine these networks from security point of view, and further previously presented ideas will be introduced to make these systems secure.

In section 2 we introduce routing types of ad hoc networks and we show two present protocols. Security goals and different types of threats are discussed in section 3. Section 4 gives cryptographic overview about the mechanisms needed in ad hoc networks. Security issues of routing mechanisms are discussed is section 5 and possible solutions are introduced to achieve a secure solution.

2. Routing mechanisms in ad hoc networks

In a multi-hop environment packets have to be routed to find their destinations; they should be transferred through a correct route. In a regular (non ad hoc) network dedicated points (gateways, servers) have information about the architecture of the network, so they can determine which direction the packet should be forwarded. Ad hoc networks have no fixed infrastructure, so there is no centralised knowledge or role. All nodes therefore should participate in the routing process; this should be distributed system from routing and other signalling point of view.

Types of routing

There are two main groups of routing protocols in the mobile area: pro-active and reactive.

In a pro-active protocol, routes are constantly being tracked and stored. The main advantage of these protocols is that the possible routes are continuously known. However these can cause a high quantity of data to be stored and a lot of packages have to be transmitted through the network. On the contrary reactive protocols only try to find a route, when it is needed. This may occur higher delay at the beginning of the communication, because routes have to be discovered at this time. These are also known as on-demand protocols. Hybrid protocols combine the above two methods.

In the next two sections two reactive protocols will be introduced. [bevprot]

Dynamic Source Routing (DSR)

In the Dynamic Source Routing (DSR) the sender strictly determines the full path of packets. The header of packets contain the complete sequence of hops, from which intermediate nodes will be able to figure out the address of the next hop.

Sender maintains a route cache, and it always checks first their database in order to find a path towards the destination. If a route does not exist, sender initiates a route request. This is a broadcast message that spreads through the network and locates the target host. If the destination is reachable, it will send back an answer message that will include the list of intermediate nodes of the path. If the path that the sender tries to use is broken, then an error message is being generated as the detection of failure. Moreover the sender should try to request a route again. [perfan]

In networks with low mobility this protocol can be quite effective, because the entries in the cache tables will stay usable. Its disadvantage is that all packets have to contain all the intermediate node addresses in the route, which can mean significant overhead in the signalling and in the sender's memory.

Ad hoc On-Demand Distance Vector Routing (AODV)

AODV is also a reactive protocol, which uses a broadcast route discovery mechanism similar to DSR, but instead of source routing, AODV relies on the dynamically established route table entries at intermediate nodes. Endpoints not required to know the whole path, since the intermediate nodes know only the correct direction and as a result they form the actual route. Loop freedom is reached with the application of sequence numbers.

When a node starts to communicate, it initiates a Path Discovery process, which is a broadcasted route request (RREQ) message containing the address of source and destination. Each node receiving a RREQ message sets up a reverse route back to the source, increments the hop-count and rebroadcasts the request. The destination replies to the first incoming RREQ with a route reply (RREP) message on the same path to the originator. This reply message runs back through the selected reverse route and sets the forward route in the nodes. (Figure 1.) Each route entries have a specific lifetime, consequently unused reverse route entries will be deleted. When there is no traffic in a link, nodes can perform local connectivity management. This means broadcasting Hello messages only to the neighbours to check connectivity.

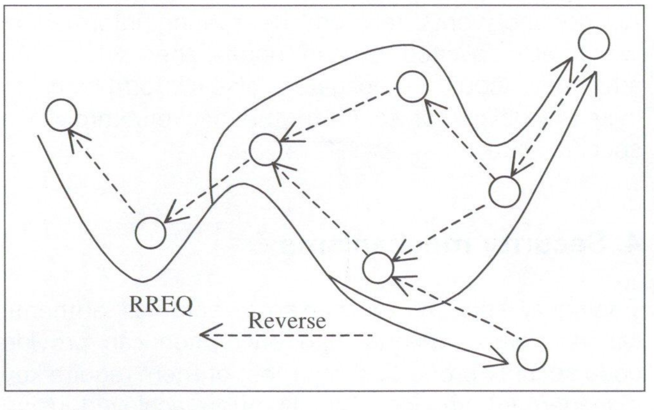


Figure 1. Established reverse routes

When a link error is detected a route error (RERR) message is generated with the list of unreachable destinations and sent back to the sources. In AODV it is possible to repair a link locally, which is called local repair. If the local repair process succeeds, the endpoints will not be informed about the action.

AODV is a scalable protocol, it offers low processing and memory overhead, low network utilization and it can determine and maintain routes effectively. This protocol can handle networks with higher mobility. [aodvdraft]

3. Security issues of ad hoc networks

Security goals

Communication networks, which have proper security provides the following security services for reliable and secure information transport.

First of all availability ensures that the network and services remain accessible and usable despite of different attacks or malfunctions. Authentication guarantees a node to ensure the source of data or identity of peer communicates with. Without this, a node could not make sure of the origin of the other side. Confidentiality means the secrecy of data, so that information never discloses to any unauthorized entities. Neither user data nor system signalling should be allowed for eavesdroppers or for not permitted parties. Integrity ensures that the information is correctly transferred, it is unaltered and errorless. Such modifications can be made by natural conditions or malicious attacks. Non-repudiation is a protection against false denial of involvement in a communication. Without this, nobody should take responsibility for his acts. Several cryptographic algorithms needs key management service to administer keys. This involves generating, distributing, storing, loading, auditing and destroying keys.

Other services might be needed, like as authorization, which can grant a system entity to access a system resource. [secissues]

Threats in mobile ad hoc environments

Wired networks with fixed infrastructure have many security threats, however mobile Ad hoc networks reach further questions. In a distributed network, without any infrastructure (fixed routers, centralised points, static neighbours) communicating peers have to rely on the whole network, therefore proper security hardly can be established. The most important questions are highlighted in this subsection. [secissues]

Denial of Service Attacks

Denial of Service Attacks (DoS) block service functioning. These type of attacks decrease service availability, prevent of authorised access to a system or generate high delays.

The adversary could scramble the physical medium, which causes interference and large number of information loss. This can make the communication channel almost unusable. Moreover it is hard to identify the data of the attacker from the natural noise of the medium. Radio channel is a shared and limited resource of mobile communication, so everyone should keep the access rules. Flooding the medium with bad messages or neglecting the access rules can prevent nodes to access the medium. If a node play destructively communication become impossible.

Mobile devices have relatively small capacity of CPU, memory and battery. Sending large amount of CPU consuming queries overloads the processor, which can prevent other operations for a while. Battery usage also should be optimised in order to increase the availability time of a service. Therefore in unused time periods processor and radio receivers are always attempted to change to power saving (sleep) mode. However, incoming requests must be still processed. A malicious node performing an energy exhaustion attack can exploit this fact. [resduckl]

Routing information is created collectively, but malicious nodes could broadcast wrong routes. They could send the others corrupt and outdated information or even could alter message paths. They can cause much longer paths or can overload nodes.

Other possible malicious behaviour is node selfishness, taking advantages of other nodes' collaboration but not taking part in. These nodes may be tempted to not relay packets in order to save their own resources. (e.g. Battery) [questfor]

Impersonation

A malicious node can simulate that it is the node the other party wanted to communicate with. However using authentication process, a verification step can reveal the identification of the communicating parties. This step is quite difficult because of the lack of any centralized entity that stores the needed information about peers.

Other type of impersonation is man in the middle attack, when a malicious node become involved in the route between the two peers.

Confidentiality Violation

By the usage of radio medium, the physical area of the shared medium is extended therefore eavesdropping is available. Passive eavesdropping is a special kind of attack where a malicious node only listens on the shared medium to catch others' communication.

Active eavesdroppers can intercept and selectively modify data. It can be man in the middle attack or attack based on routing manipulation.

Even communication is encrypted code breaking attacks are possible, mainly when weak encryption algorithms are used.

Message alteration

Malicious node can influence the message has been sent to the destination. It could insert new message to confuse the nodes or alter the meaning of the information. The removal of important information is also dangerous, this could be reached for example by route manipulation.

Message repudiation

It means denial by a system entity that it was involved in a communication. If not possible to bound message to its originator than later an investigation cannot decide that the node itself or other malicious entity behaved as logged.

Anonymity violation

Malicious nodes involved in the communication process are able to collect some information about the neighbouring nodes due to the access medium is shared and identification is readable. Therefore identification or localisation of the target host became easy.

When nodes have authentication credentials, they can identify themselves, so it is possible to collect information about the other nodes. These identifications can be disclosed to eavesdroppers too.

Physical tampering

If an unauthorized person gets access to the equipment, hardware or software modification could be made. An adversary can modify the operation of device, can install viruses, Trojans, etc.

Lack of centralised trusted element

In a mobile ad hoc environment it is assumed that there is no trusted centralised secure network element. The advantage of this is that the system is not vulnerable when a node with centralised functionality brakes down or became compromised. However this raises a lot of problems, because authentication procedure in distributed system cannot easily handled since there is no fully trusted network element.

Routing manipulation

Ad hoc networks rely on the routing information, which are solicited by the nodes themselves. An adversary could propagate false information to increase delays or to route the packets through a specific node.

4. Security mechanisms

Traditional security mechanisms, such as authentication, digital signature and encryption can provide good security properties, but most of them require key management service. This is often achieved using a centralised entity. In an ad hoc network we cannot rely on a node with centralised responsibility, these tasks should be distributed. Moreover this architecture should consist redundancy, because nodes disappear or they lost connection and even they can be compromised.

Public Key Infrastructure (PKI)

In a public key infrastructure (PKI) each node has a public/private key pair. A message encrypted with the public key is only readable with the corresponding secret key. From the public key the corresponding private key cannot be derived.

The goal of Certificate Authority (CA) is to sign certificates so to bind public keys to nodes. CA should stay on-line to reflect the current bindings. It should track changes and it can revoke certificates if the node is no longer trusted.

PKI is a main element of certificate-based authentication, because it provides the CA to ensure the identity of the nodes.

Threshold cryptography

A solution for certificate authority in ad hoc environment is distribution of trust using threshold cryptography. A (n, t+1) threshold cryptography scheme allows n parties to share the ability to perform a cryptographic operation, so that any t+1 parties can perform this operation jointly, whereas it is infeasible for at most t parties to do so, even by collusion.

In this case n servers of the key management service share the ability to sign certificates. This scheme can tolerate *t* compromised servers by dividing the private key *k* of the service into n shares (s_1, s_2,...,s_n) assigning one share to each server. We call this composition an (n, t+1) sharing of *k*. After this each server can generate a part of the signature, which are combined. This certificate still adequate even if t parties are compromised.

Besides the advantage of threshold signature, this kind of key management service is easily able to refresh the participants' shares. This is very effective against mobile adversaries that compromises a server and then moves on the next victim. This kind of attacker can take several servers, but by the end of periodical share refresh the obtained node loses the ability of service. During a share refresh a new threshold configuration can be created, which is a good occasion to adapt network changes. Moreover share refreshing is not a complicated operation, so once created share configurations can be adapted for a long time. When too many participants are compromised, an entirely new share should be created.

Threshold cryptography has a serious problem, it assumes synchrony of nodes. In reality it is not possible, because a node could disconnect for a period or a denial of service attack could slow it down, while the others would do a share refresh. When a node with older share tries to attach to the service, it will not be able to participate the communication because a new threshold configuration would be applied. [secadhoc]

The establishment of distributed service causes another issue. When only few nodes exist in the network, it is undecided which one is authorized to initiate the shared authority. Moreover, once network splits into distinct parts, these parts may operate as different authorities.

Self organizing Public key infrastructure

Trusted third parties are a very troublesome part of certificate-based authentication. Therefore it would be beneficial to avoid them and build a self-organized structure. There are some certificate-based systems (e.g. Pretty Good Privacy – PGP), in which certificates are issued by the users themselves and distribution of certificates is managed by a self-organized structure.

Each user maintains a local certificate repository that contains a limited number of certificates. When user *u* wants to obtain the public key of user *v*, they merge their local certificate repositories and u tries to find an appropriate certificate chain from *u* to *v* in the merged repository (Figure 2). For the construction of the local repositories several algorithms exists to provide that almost any pair of users can find certificate chains, even if the size of the local repositories is small compared to the total number of users of the network.

This approach realises a self-organized certificate system.

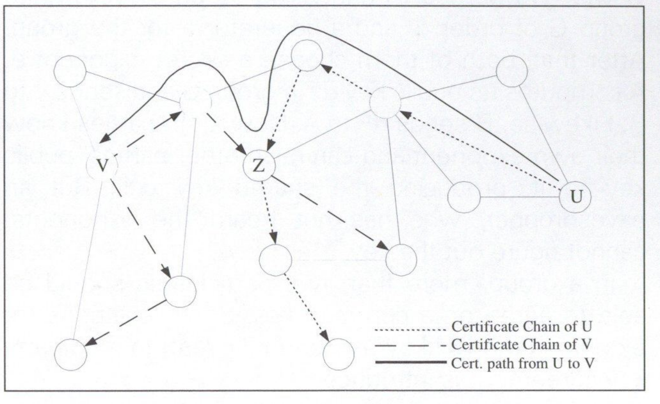


Figure 2. Self-organizing PKI certificate chain

ID-Based cryptography

The basic idea of identity-based cryptoschemes is that information used for identification (node address, user name, etc.) could act as a public key, so that CA is no longer needed to bind keys and identities. However the strength of PKI is that the secret and public keys cannot be derived from each other. In this case, only a central entity is able to derive the corresponding secret pair of a public key (identity). This operation is needed only once, at the registration of users. After this, participants are able to authenticate each other in a non-interactive manner, without communicating with a third party.

In a real scenario compromised nodes have to reregister themselves, so the registering central entity should stay available. Serious problem is that this central entity will know all the registered secret keys also, so it became a vulnerable part of the system. Moreover the registration process needs huge resources and lasts a lot of time. [questfor]

Common key architecture

A group of nodes that previously identified each other, wish to form a secure part of the network; therefore they should establish an encrypted session. Symmetric key architectures are usually used for this purpose, because of the relatively low processing overhead of the encoding and decoding mechanisms. This requires the participants to have common (shared) secret key, which they use for both encrypting and decrypting. By the application of symmetric keys, secure broadcast and multicast communication is also possible. Several key establishment protocols exist.

Diffie-Hellman (DH) key exchange

The Diffie-Hellman (DH) key-exchange algorithm is suitable to securely share a common secret between two parties. Securely means that a third party who can also hear their communication (e.g. eavesdropper) must not be able to generate that key.

The two parties (A and B) agree on a cyclic finite group G of order q, and a generator a for the group. After that, both of them choose a secret exponent e. A computes its public key α^A in group G and sends it to B. Likewise, B sends α^B to A. Now, both parties know their own exponent and can raise other parties' public key to it, producing the shared key α^{AB}. But an eavesdropper, who has not heard the exponents, cannot figure out the key.

In a group, more than two participants should be able to agree on a common key. Some solutions for extending the DH key exchange [keyest] to multiparty key agreement is introduced.

GDH.2

GDH is the acronym of generalized Diffie-Hellman. Nodes form a chain and the first member starts a DH like message to the next one, who extends it with additional information and forwards. This process goes till the last node. This node is now able to generate the common key, and also key pieces that needed by each participant. These pieces are all broadcasted, and a third party (eavesdropper) cannot generate the common secret from these. Finally the common key is successfully established. (Figure 3.)

The last (broadcaster) node should play a central role, which is not advantageous in an ad hoc environment. Moreover the relatively large amount of broadcasted data results significant overhead.

Hypercube and Octopus

The main idea for establish a common secure key between more than two participants in Hypercube is to form pairs of nodes and establish the common secret using DH protocol. Then they form pairs of pairs

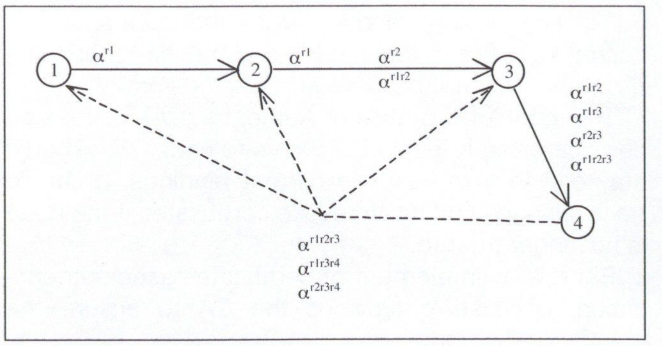


Figure 3. GDH.2 common key generation

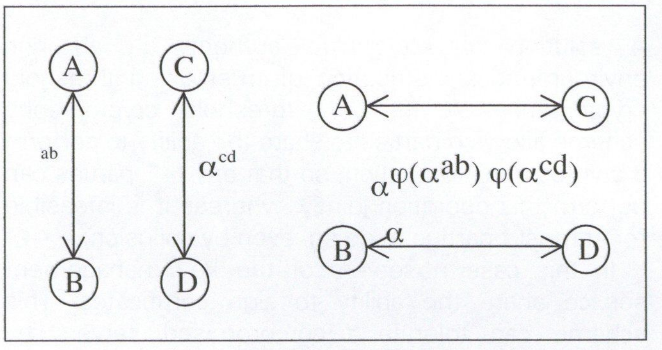


Figure 4. Hypercube protocol example

and perform key exchange by all participants, and so on. (figure 4.) The problem is that the number of members must be 2n which cannot be supposed.

An extension to this problem is the Octopus protocol. This forms a hypercube kernel of the network and extends it with tentacles. First these tentacles do a DH key exchange with their respective central nodes. After that the central nodes perform the Hypercube key exchange and inform the tentacle nodes about the new key. However Hypercube kernel plays a central role and the involvement of a new participant is difficult.

GKMP (Group Key Management Protocol)

This protocol provides management functions of symmetric keys for a set of nodes. Key generation concept used by GKMP is a cooperative generation between two entities, such as Diffie-Hellman key exchange. After generating the group key, GKMP distributes it to qualified GKMP entities. It also allows new members to join, member deletion and rekeying the group. GKMP provides a peer-to-peer review process; entities have permission certificates (PC) as part of the keying process. Therefore each entity can verify the permissions of any other GKMP entity but can modify none. GKMP supports compromise recovery and the list of compromised nodes is disseminated through the network and Compromise Recovery List (CRL) is stored in each host.

This protocol attempts to delegate as many functions to the group as possible, so it tries to not rely on any centralized entity. However some functions, such as granting privileges, creating and distributing key, creating group and re-key messages should still be centralized. [gkmp]

We can conclude, that achieving confidentiality is not a difficult problem, because there are methods to establish common secret (DH, GDH.2, Hypercube, Octopus) among participants. This secret can be used in a common key architecture.

However more serious problem is the proper authentication of the participants. In ad hoc environment, we cannot rely on a centralized Certificate Authority that could ensure certificates. We could see some possible solutions (Threshold Cryptography, Self organizing PKI, ID-Based PKI), although those are effective only in special circumstances.

5. Security issues in routing mechanisms

Security properties of ad hoc networks depends on the used routing mechanism, although in present routing protocols security is not included. An attacker may become involved in packet forwarding, moreover it is sufficient to run an intermediate node close and force a DoS attack using the shared radio channel. The scramble of the attacker is mostly indistinguishable from the natural loss caused by the noise of the channel. The goal is to establish a reliable channel, where trustiness can be a QoS parameter of the route.

Consequently a routing protocol should be able to guarantee some level of security of a path and to force packets to travel through this route. Using distributed mechanisms or having a node with additional routing information (e.g. at the source) the path can be determined.

Onion routing

In onion routing, between peers the messages traversed securely. This is realized with a public key system. The sender collects the public keys of the intermediate nodes and encrypts the sent message with all of them. Each intermediate node decrypts it with its secret key, removes the outermost lock. At the destination the encryption can disappear only if each intermediate node has used its secret key for decoding (Figure 5).

This kind of routing security is applicable mostly when the sender knows the path. (e.g. DSR).

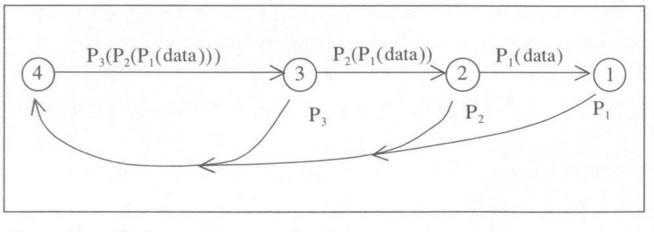


Figure 5. Onion routing mechanism

Security-aware routing

Traditionally, routing protocols try to find an optimal route. The metric for optimality is distance usually measured in hops. To improve the quality of security of an ad hoc route "Security Aware ad-hoc Routing" (SAR) has been presented, which incorporates security levels of nodes.

Different techniques exist to measure or specify the quality of security of a route.

Each host has the attributes such as "trust level" and "security level". These attributes are used by the routing algorithm, and only nodes that provide the required level can participate the routing protocol. However, these levels should be immutable, so that a node with a lower level can neither change it's own nor the requested level.

In SAR the sender who initiates a route discovery, embeds the needed security attributes into the request. Intermediate nodes forward it only if it has the proper security attributes, otherwise it should be dropped. If the destination receives a request with proper security attributes, consequently an end-to-end path with the required security attributes can be found, and a security-aware route could be established. SAR can be implemented based on any on-demand ad hoc routing protocol with suitable modification. [secaware]

These protocol modifications result in changes the nature of discovered routes. The route discovered by SAR may not be the shortest one in the terms of hop count, although SAR is able to find a route with a quantifiable guarantee of security. If one or more routes that satisfy the required properties exist, SAR will find the shortest (optimal) such route. However SAR may fail even if a network is connected, but the required security attributes cannot be provided.

A problem with SAR is that the specific levels should be authenticated; nodes should not identify their own attributes. We could see that authentication in ad hoc networks is not a simple process. On top of that comprehensible levels reach new threats, because those levels are often related with the importance of nodes.

Watchdog, Pathrater

Using static parameters of participants is not perfect approach, because they may change over time. (e.g. Compromitted nodes)

Watchdog method examines behaviour of participant continuously. Due to the shared radio channel, each node can listen for the sent packets of their neighbours, consequently a malicious node can be detected easily by its direct neighbour. This procedure does not cause additional network traffic, although it may make a mistake (e.g. Asymmetric link, collusion) or can be misleaded (eg. Directional aerial).

After detecting malicious nodes, another goal is to exclude them from communication. In the Pathrater method each node maintains a rating for every other node it knows. Ratings of nodes on actively used paths are continuously incremented; while decremented on misbehaving paths. With the help of such information it is possible to select a well behaving path.

Selfishness

Previously examined secure routing procedures tries to exclude misbehaving parties from packet forwarding, although their messages are forwarded without complaint.

This type of Selfishness behaviour can be hopeful for example to save battery resources, although in quantities it occurs the network to stop functioning. Forcing cooperation can be achieved by a virtual currency, which is called 'nuglets'. Participants buy services from each other, while selfish nodes run out of nuglets.

Punishment of untrusty nodes should provide that over time only good behaviour pay off. This means that misbehaving nodes should be totally refused by the network.

6. Conclusions

We could see that ad hoc networks have many serious threats. We introduced some possible solutions towards the possibility of secure communication in such a distributed environment.

References:

[bevprot] Földesi András, Homolya György, Horváth Cz. János, Dr. Imre Sándor. Bevezetés a mobil ad hoc útvonalválasztó protokollok világába (Híradástechnika, 2001. május)

[questfor] Jean-Pierre Hubaux, Levente Buttyán and Srdan Capkun. The Quest for Security in Mobile Ad Hoc Networks (ACM Symposium on Mobile Ad Hoc Networking and Computing, MobiHOC 2001)

[secadhoc] Lidong Zhou and Zygmunt J. Haas. Securing Ad Hoc Networks (IEEE Network Magazine, vol. 13, no.6, November/December 1999)

[keyest] Maarit Hietalahti. Key Establishment in Ad-hoc Networks (Proceedings of the Helsinki University of Technology, Seminar on Network Security fall 2000)

[secaware] Seung Yi, Prasad Naldurg and Robin Kravets. Security-Aware Ad-Hoc Routing for Wireless Networks (Technical Report UIUCDCS-R-2001-2241(ps/pdf), August 2001)

[secissues] Kata Molnár and László Zömbik. Security Issues in Mobile Ad hoc Networks. (Evolution in the military communications systems – trends and challenges in the XXI. century, 2001.)

[resduckl] F. Stajano and R. Anderson. The Resurrecting Duckling: Security Issues for Ad-hoc Wireless Networks (Security Protocols, 7th International Workshop, 1999)

[gkmp] H. Harney and C. Muckenhirn. Group Key Management Protocol (GKMP) Architecture. RFC2094. Group Key Management Protocol (GKMP) Specification. RFC2093

[aodvdraft] Charles E. Perkins, Elizabeth M. Belding-Royer and Samir R. Das. Ad hoc On-Demand Distance Vector (AODV) Routing (work in progress). Draft-ietf-manet-aodv-10.txt, January 2002.

News

Multimedia „Triple Play" Specifications Approved

ITU has approved specifications that will allow traditional 'coopper wire' telecom operators to compete cost-effectively with cable and satellite operators in providing the 'triple play' of multimedia services, namely multiple high-quality digital video streams, high-speed Internet access and voice services.

Significant interest has been shown in Very High-Speed Digital Subscriber Line (VDSL) services, with operators, which already has a triple play service proposition, plans to extend this service to apartment buildings using VDSL. It cites the service as providing new competitive opportunities to generate revenues via the enhanced interactivity that VDSL provides.

The specifications were approved at the first meeting of the ITU-T 'Full Service – Very high speed Digital Subscriber Line' (FS-VDSL) Focus Group.

Thanks to continuing advances in video compression, operators will be able to ensure that the quality of the image will match that provided by cable and satellite.

Examination of IP Macromobility in OMNeT++ Simulation Environment

CSABA CSEGEDI, SÁNDOR IMRE, RÓBERT SCHULCZ, SZABOLCS VAJDA

Budapest University of Technology and Economics, Department of Telecommunications
imre@hit.bme.hu

Our article deals with IPv6 based mobile networks. First of all we would like to introduce the major novelties of IPv6 from the aspect of mobility – compared to the IPv4. After that we describe the handling of IPv6 mobility, then we introduce the simulation environment made by us and the derived results.

Spreading of mobile computer technology

Spreading of mobile telephones was beyond any expectations of the engineers. By now there is an urging need for mobile transmission of other data than that of speech coding. Because the systems operating nowadays were planned to transmit speech coding data, they are not appropriate for transmitting bigger amount of data.

At the present time the trend in informatics is to introduce services on IP basis from user to user if possible. This technology called All IP. Because of IP being a packet switched data transmission method, it uses resources in a more efficient way than the land-based telephony or the circuit switched GSM system. Thus it needs more complicated protocol to control the packets – especially in case of mobile systems. The IPv4 systems operating today and having almost 20 years of history are not eligible for the current increased demands (appropriate big address field, integrated mobility handling, QoS (Quality of Service) parameters), that is why new systems have to be developed. The solution might be the latest version of IP protocol, the IPv6 [1] [4], which supports integrated mobility at the side of many other novelties.

IPv6 versus IPv4

IPv6 in its operation is basically similar to IPv4 but also integrates several important innovations. Among these innovations the most important are:

Bigger address field

IPv4's most emerging bottleneck is its limited address space. Considering the 4.3 billion addresses provided by 32 bit addressing, one could ask why this would be an issue since the number of computers attached to the Internet is still below 100 million. As a matter of fact, IPv4 addresses are allocated very wastefully. The problem of the limited address space can be alleviated by using address translation techniques like Network Address Translation (NAT) but these methods provoke a number of problems by sacrificing the end-to-end semantics of IP. As the number of subnetworks connected to the Internet grows, routing tables are also becoming slowly unmanageable.

IPv6's perhaps most often discussed feature is its 128 bit addressing [8], offering powerful scalability. This addressing architecture provides 2^{128} (which is 340.282.366.920.938.463.463.374.607.431.768.211.45) addresses for Internet hosts, which means that there will be thousands of IP addresses for each square meter of the Earth's surface. An address space of such an enormous size can probably meet every future demand, independently of wasteful allocation strategies. At least according to the optimistic network planners. In the future not only the computers will have IP addresses but... At the present time there are more mobile handsets than personal computers all over the world and a big segment of these is capable of connecting to any kind of data network, e.g. via WAP. This is not yet to be called real Internet connection but the number of "intelligent" devices that can connect to the network will dramatically increase in the near future. In a few years' time not only the laptops or mobile telephones but PDAs, digital cameras, different Bluetooth devices, telemetry and building informatics system elements, vehicles, and what is more, household devices can access the network with their unique IP addresses.

The new addressing structure also allows more hierarchical backbone routing based on provider topology, which can stop the expansion of routing tables in backbone Internet routers.

Security

While the use of IPSec is optional in IPv4, it is a mandatory, integral part of IPv6. Thus, it is always possible to set up an IPv6 connection in a secure fashion. The multitude of IPv6 addresses also contributes to achieving end-to-end security, since it eliminates the problem of NATs breaking the security during address translation.

IPSec security is implemented with two extension headers in IPv6: the Authentication Header [9] and the Encapsulated Security Payload header [10]. A Security Association (SA) is used to describe how the communicating entities utilize security services in their communication sessions. A SA is identified by three parameters: the Security Parameter Index (SPI), the destination IP address and the identifier of the used security protocol (AH or ESP).

Autoconfiguration

IPv6's tremendous address space would rather become a drawback as a beneficial feature if there were no advanced mechanisms for assignment and management of these addresses. As far as possible, a mechanism like this must provide an automatic, cost-effective and well manageable way for configuring addresses. IPv6 introduces an elegant approach for this task with its Address Autoconfiguration protocol [6]. Besides configuring addresses, the protocol also allows other network parameters to be set up automatically, like the default gateway address, default router, etc. These features allow a host to connect to a network in a plug-and-play manner, without the need of any manual intervention.

There are two ways of configuring a host's address automatically: a stateful and a stateless address autoconfiguration mechanism. Stateless and stateful autoconfiguration complement each other.

Neighbour Discovery

The Neighbour Discovery (ND) [7] protocol replaces the old ARP protocol used to determine the link-layer addresses of hosts from their network addresses. It uses ICMPv6 (Internet Control Message Protocol v6) messages to discover a host's surrounding network and neighbour hosts. ICMPv6 is the new version of IPv4's ICMP protocol. Basically, it includes all the messages defined in ICMP, and five additional messages for neighbour discovery.

New address types

Unicast addresses: These addresses resemble most IPv4 addresses. A packet sent to a unicast address is received by one (and only one) interface assigned to that address. A unicast address unambiguously defines an interface (link-local, site-local or global) and thereby a network host.

Multicast addresses: Multicast addresses replace IPv4's broadcast addresses. Just like anycast addresses, the multicast addresses are also assigned to a group of interfaces. However, packets sent to a multicast address are received by all hosts of the group.

Anycast addresses: Anycast addresses are assigned to groups of interfaces, possibly belonging to different hosts. A packet sent to an anycast address is received by only one of the interfaces belonging to the group, usually by the one that is located nearest to the sender.

Streamlined Header Format

IPv6 also streamlines and enhances the basic header layout of the packet. Compared to IPv4, some of the headers were dropped and others were made optional. Omitted fields include:

Fragmentation: Fragmentation can now only be done by end stations of a route. IPv4's method of fragmenting datagrams in intermediate hops requires resources and processing efforts in routers unreasonably.

Options: IPv6 defines the new "Extension header" mechanism instead of IPv4's header options. Extension headers carry optional header information, and are generally not processed by intermediate routers, which contributes greatly to faster end-to-end delivery.

Header checksum: Calculating header checksums requires a lot of processing time and offers little advantages in today's Internet. Up-to-date network technologies have considerably low error rates, while error checking and correcting functions are included in other layers as well. This makes calculating checksums at the network layer unnecessary.

The new header contains only 8 fields compared to the 14 fields of the IPv4 header. It also has a fixed size, which is a further milestone in increasing processing speed. One important additional field is the 20 bit long Flow Label that is responsible for QoS support for traffic flows. Due to the simplified header structure, the total IPv6 header size is only twice as large as the IPv4 header, even though 16-byte IPv6 addresses are four times longer than the 4-byte IPv4 addresses.

Header extension

Header extensions are located between the IPv6 header and the transport layer header, and contain optional network layer information. Usually only the destination host processes them, but some of them require hop-by-hop processing. These extensions immediately follow the base IPv6 header, preceding other extensions. Thus, intermediate routers only have to investigate the first part of a packet, which simplifies processing in comparison with IPv4, where the header length is variable depending on the included options. An IPv6 packet can carry several extension headers.

Mobility

Mobile hosts connected to the Internet via a wireless interface are likely to change their point of access frequently. A mechanism is required that ensures that packets addressed to moving hosts are successfully delivered. During handover, packet loss may occur due to delayed propagation of new location information up to the Home Agent. These losses should be minimized in order to avoid he degradation of service quality as handover become more frequent. Mobility management can be divided into two parts: micro- and macro mobility. Macro mobility handles interdomain handovers, while micro mobility responsible for intradomain handovers (see Figure 1.).

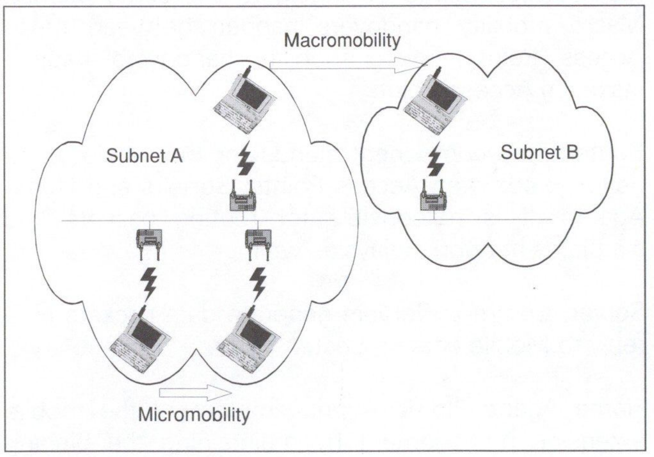


Figure 1. Micro- and macromobility

To improve performance, the frequent handoffs (due to small radio cells) inside a given domain – also so called intra domain handovers – are handled locally by the micro mobility protocols. The role of micro mobility protocols is to hide user movement from the mobile IPv4 or IPv6 protocol, by handling user mobility locally, fast, and simple inside the micro mobility domain. IPv6 or Mobile IPv4 are responsible for wide area mobility support, called macro mobility.

The Mobile IP protocol is considered to have limitations in its capability to handle large numbers of mobile stations moving fast between different radio cells. The handover frequency should typically not exceed once a second. However, Mobile IP is well suited for interconnecting disparate cellular networks effectively providing global mobility. Resulting from this fact, several micro mobility approaches have been proposed within the IETF, which are supporting mobility in a well-defined area. Such as the two most discussed micro mobility protocols: HAWAII and CIP.

The mobile IPv6 (and the mobile extension of IPv4 as well) basically solves the macromobility problem. The basic idea is the following: let us separate the identification and routing role of IP addresses! In the mobile IP each node has a static, so called home address, which identifies the device. This address remains unchanged while roaming. Devices leaving their home network get a temporary, so called care-of address in every foreign network, which topologically belongs to the given network, thus the mobile device remains reachable in the foreign network using this care-of address.

The terms introduced by Mobile IPv6 are the following:

- *Home address:* The IP address assigned to a mobile node within its home link.
- *Care-of address (CoA):* A temporary IP address assigned to a mobile node while visiting a foreign link.
- *Binding:* The association of the home address and a care-of address of a mobile node, along with the remaining lifetime of that association.
- *Mobile node (MN):* A node that can change its point of attachment to the Internet, while still being reachable via its home address.
- *Home Agent (HA):* A router on a mobile node's home link with which the mobile node has registered its current care-of address. While the mobile node is away from home, the Home Agent intercepts packets on the home link destined to the mobile node's home address, encapsulates them, and tunnels them to the mobile node's registered care-of address.
- *Access Point:* A router in the foreign network, which ensures the visiting mobile's connection to the network, through wired or radio interface.
- *Binding Cache (BC):* A conceptual data structure for storing bindings. The Binding Cache should be implemented by all IPv6 nodes.
- *Header extension:* A header extension, which can be sent along with arbitrary (even empty) packet and which contains complementary information, e.g. for handling mobility.
- *Binding Update (BU) message:* A header extension, which contains the current binding and the lifetime of the binding of the sending mobile node.
- *Binding Request (BR) message:* A header extension, in which a communication partner can ask the mobile node to send its current address.

- *Binding Acknowledge (BA) message:* A header extension, which the Home Agent acknowledges the reception of the Binding Update message with.
- *Home Address extension:* Thus the mobile node usually sets up the sending address to the foreign address while sending packages, this extension serves for telling the addressee its identifying home address.
- *Binding List (BL):* This list contains those Binding Update messages that were sent by the mobile node.

Every mobile device in IPv6 can always be addressed with its home address. When the mobile device is not attached to its home network, it obtains a temporary IP address – a care-of address – from the foreign network it is currently attached to. In order to be able to receive packages in this case the mobile always informs its Home Agent – a router in its home subnetwork – about its current care-of address. Correspondent nodes can send packages directly to the care-of address if they know it, otherwise they send them to the home address and the Home Agent forwards them to the mobile. The association between the home address and the care-of address is called binding. In IPv6 networks, every node contains a so-called Binding Cache to store binding information about mobile devices. If the correspondent node uses the home address and the Home Agent forwards the packet to the user, this routeing is called triangled routing. This of course overloads the network.

With the limited capability of mobiles and network overhead caused by triangle routing the optimization of the Binding Cache's size and the binding entries' lifetimes is very important. Our simulation demonstrates this issue in different network scenarios. We investigate different statistics like end-to-end delay time, rate of packets sent via triangle routing, rate of packet loss, handover frequency, etc.

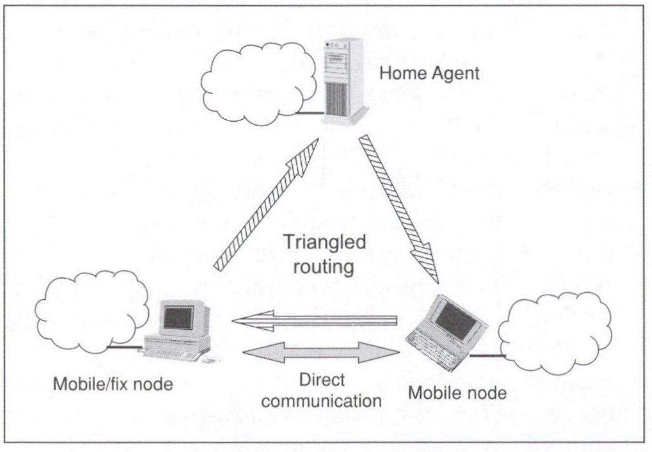


Figure 2. Triangled routing and direct communication

The simulation environment

We have developed a simulation environment to prove our concepts of Mobile IPv6 under OMNeT++. OMNeT++ (Objective Modular Network Testbed in C++) is a free, open-source discrete event simulation tool, similar to other tools like PARSEC, NS, or commercial products like OPNET. Our Mobile IPv6 model can be freely downloaded along with many other models.

Modules of the macro mobility environment

Our simulation environment deals with the IPv6 Mobility Extension, especially with the binding management methods. With the simulator, we can easily build different network scenarios by providing a few simple parameters from which the simulator constructs the network automatically.

According to OMNeT++, the structure of our simulator is modular. We defined the modules and their connections in the NED language, and implemented their functions in C++. The modules are the following:

Mobile: This component represents a mobile device, which changes its location and speed periodically, and sends data requests to servers and other mobiles, as well as receives data from these.

Air: It represents the radio interface, but now it simply connects Mobiles to the wireline network. There is only one Air module, because OMNeT++ can not handle dynamic connections properly.

Access Point: These elements represent all physical radio access points belonging to the same subnet. Macro mobility handovers happen between these Access Points, micro mobility handovers happen inside an Access Point.

Router: This component stands for the whole wired network between Access Points, Servers and Home Agents. It is responsible for routing packets and simulates network delays as well.

Server: Common Servers generate data packets as a reply to Mobile data requests.

Home Agent: Home Agents implement the mobile extension management by maintaining the binding between a Mobile's home and foreign address.

The modules are connected to each other according to the following figure (Figure 3.)

Simulation results

As seen before the Binding Cache has a very important role in mobility support. We simulated how the Binding Cache size and the bindings stored in the cache effect the ratio of the triangled packets. In the network we had 50 mobile nodes, 9 subnets, 5 servers and 7 Home Agents.

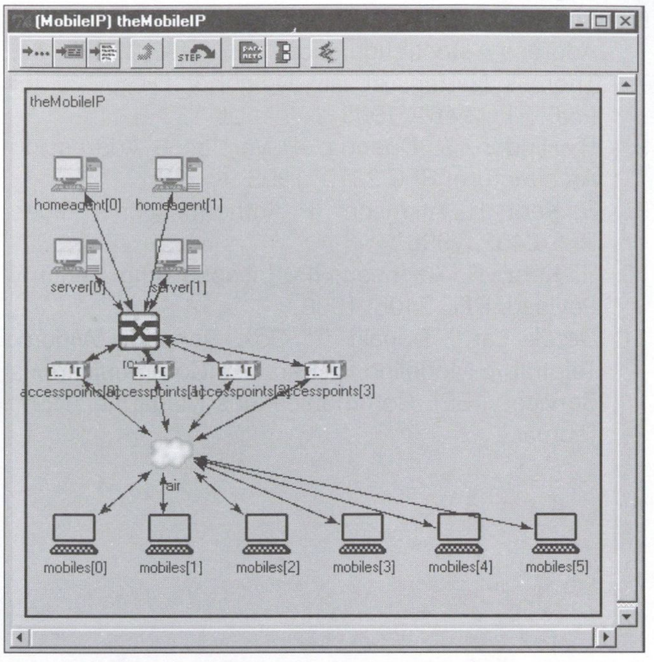


Figure 3. The Simulator

In our case two types of packets could be delivered on the triangled route, that is why we tracked these packets in the network:

Data request: the ratio of the triangled data request to the total number of data request packages. Every packet contain the Binding Update extensions the reply for these packets are delivered on the direct route.

Spontaneous data packets: in this case a remote mobile node or a server send a packet to the mobile. Here also the ratio of the triangled data packets to the total number of data packets are measured.

Examining the Binding Cache's size

By running our simulator with different Binding Cache sizes between 1 and 55 (the number of mobile devices is 50 in this case), we came to the following results seen on Figure 4.

It can be seen that increasing the cache's size linearly decreases the rate of the triangled packets. This is because bigger Binding Cache can store more binding. It is self evident as seen on the figure, that when the Binding Cache is smaller, more data packets are triangled, and this leeds to an overload in the network In case of a big Binding Cache, the size of the portable devices were increased, but this is not good in the mobility point of view.

Examining the binding entries' lifetimes

By running our simulator with different entry lifetimes between 0 and 100 seconds, we came to the following results, as seen on Figure 5.

It can be seen that by increasing the lifetimes first quickly decreases the rate of the triangled packets,

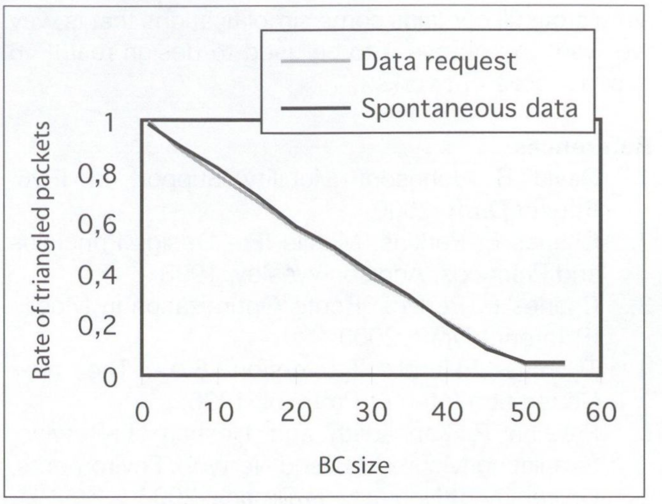


Figure 4. The effect of BC size

until it reaches the fifty percent level. (The Binding Cache's size in this scenario was 25, that is, the half of the mobile's number.) When the lifetime is 0, all the packets are delivered on the triangle rout. If we start to increase the lifetime, the ratio of the triangled packets

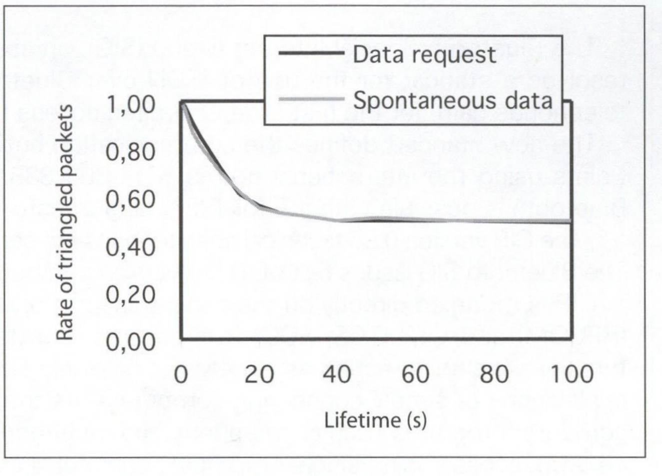


Figure 5. The effect of BC entries' lifetime

is decreasing, until a point, while at this point the new entries are replacing the old ones. In this particular case this number at around 20-25 seconds.

Summary

In our article we introduced the functions of the IPv6 needed for the mobility handling. To inspect these functions we wrote a general IPv6 simulator. We analysed the effect of the Binding Cache sizes and the lifetime of the bindings on the network performance. With our simulator the optimal Binding Cache size and lifetime can be calculated.

The simulator is capable of simulating mobile terminals operating at willing IP environments. The simulator still contains some simplifications that is way we want to improve it to be used to design real IPv6 based mobile networks.

References

1. David B. Johnson: Mobility Support in IPv6, Intenet Draft, 2000
2. Charles E. Perkins: Mobile IP – Design Principles and Practices, Addison-Wesley, 1998
3. Charles E. Perkins: Route Optimization in Mobile IP, Intenet Draft, 2000
4. Thomas Eklund: IP version 6 – The next Generation Internet Protocol, 1996
5. Preetha P. Kannadath and Hesham El-Rewino: Simulating Mobile IP Based Network Enviroments, University of Nebraska at Omaha, 2000.
6. Susan Thomson, Thomas Narten: IPv6 Stateless Address Autoconfiguration, RFC 2462, 1998
7. Thomas Narten, et. al.: Neighbor Discovery for IPv6, RFC 2461, 1998
8. R. Hinden, S. Deering: IP Version 6 Addressing Architecture, RFC 2373, 1998
9. S. Kent, R. Atkinson: IP Authentication Header, RFC 2402, 1998
10. S. Kent, R. Atkinson: IP Encapsulating Security Payload, RFC 2406, 1998
11. Derek Lam, Donald C. Cox, Jennifer Widom: Teletraffic Modeling for Personal Communications Services, IEEE Communications Magazine, 1997. február

News

The Bluetooth Special Interest Group (SIG), an association of over 2000 technology companies, has just resolved a standar for the use of ISDN over Bluetooth. Bluetooth devices such as PCs PDAs and GSM telephones gain, for the first time, unlimited access to ISDN data and telephone communications services.

The new standart defines the communication between the so-called ISDN Clients and the ISDN Access Points using the international norms ETSI 300 838 and GSM 07.08. Wireless ISDN communication over Bluetooth is possible with the full ISDN data transfer rate and a coverage of 100 meters and more

„The CIP version 0.95 technical specifications are presently available at www.bluetooth.org/specifications.htm. The Bluetooth SIG issues first of all the version number 0.95 to all officially adopted and published Profiles.

CIP is modeled directly on the basis transport protocol L2CAP can be used parallel to PAN (BNEP), DUN (RFCOM) and CTP (TCS, SCO). CIP-capable end devices can therefore avail of all ISDN capabilities and functions including acting as a network gateway (with compression), support telephony and relevant CTI applications or simply connecting computers. As the ISDN software interface CAPI has been integrated, all computer programs such as telephony and multimedia applications, answering machines, fast Internet over network access using single or multiple channel access are possible, without using any cables. Additionally when using CIP a PC or a PDA can dial via ISDN directly to a host and swap data. The highly versatile ISDN interface CAPI can now be used for Bluetooth telephones or headsets.

ISDN applications based on the CAPI interface have proven themselves in a variety of different fields over the past years. Because of the CAPI specifications software from manufacturer A is compatible with the hardware from Manufacturer B. This principle is also valid for ISDN applications over Bluetooth.

BlueFRITZ! can be used with applications from other manufacturers.

AVM will release the CMTP source code to facilitate the fast and widespread usage of ISDN over Bluetooth. This source code allows for the integration of the CMTP protocol as the basis of CIP conveniently into existing Bluetooth structures under Linux.

Using the Bluetooth protocol stack under Linux it will be possible to use CIP-capable end-devices and it will allow the use of existing CAPI based ISDN applications for data and speech without modification under Bluetooth.

BlueFRITZ! products support CIP as a central Profile, the BlueFRITZ! products update policy guarantees the implementation of the newest standards. In the same manner all other important protocols will be made available

A New Blind Signal Processing Algorithm for Decorrelation and Multiuser Detection

JÁNOS LEVENDOVSZKY, LÓRÁNT KOVÁCS, ANDRÁS OLÁH, DOMONKOS VARGA
Budapest University of Technology and Economics
Department of Telecommunications
levendov@hit.bme.hu

The paper is concerned with developing novel adaptive algorithms for decorrelating weakly stationary random processes. The proposed decorrelation technique proves to be instrumental in mobile communication, where equalizing radio channels with severe intersymbol and multiuser interference is a frequently occurring task. The algorithm can perform blind equalization, in terms of eliminating channel distortions even without a training sequence. The convergence properties of the new algorithm are proven in mean square by using the Kushner-Clark theory of stochastic approximation. The performance is studied by extensive simulations in the case of mobile communication with typical radio channels in the presence of multipath propagation.
The proposed detector structure is proven to be superior to the traditional receivers. The results are demonstrated by extensive simulations as well.

1. Introduction

Since direct sequence code division multiple access (DS-CDMA) is going to be the standard access protocol in 3G wireless multiple access communication (i.e. UMTS), to increase its spectral effciency by developing near optimal detection algorithms has been one of the central issues of signal processing theory. By eliminating the "crosstalk" between users, (called multiuser interference, MUI) the spectral effciency (and the corresponding channel effciency) can be dramatically increased.

In the case of CDMA (Code Division Multiple Access), each user is associated with a signature signal which is applied for user identi_cation. In general, for a larger population and limited signature support one cannot construct a mutually orthogonal signature-set, which leads to crosstalk between users. The optimal code-set, resulting in minimal interference between users, was developed by Gold [4]. The situation of simultaneous use of a communication channel by several users is depicted in Figure (1).

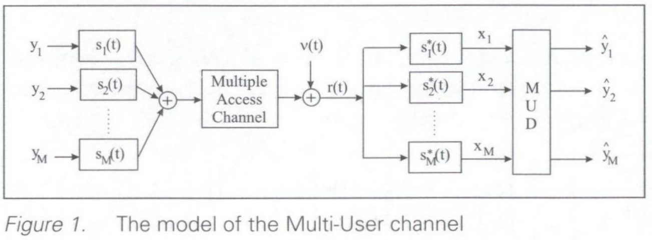


Figure 1. The model of the Multi-User channel

The most promising solution to this problem is the maximum-likelihood multiuser detector which was introduced by Verdú [1]. Although the performance of the optimal multiuser detector almost reaches the performance of the optimal single user detector, its complexity is rather high (it increases exponentially with respect to the number of users). Hence, finding suboptimal solutions with moderate computational complexity is of major importance. Unfortunately, MUI is not the only factor which impairs the performance of mobile channels. Selective fading, resulting in excessive channel distortions, is also a rather typical phenomenon in digital radio communication because of multipath propagation. Therefore, to achieve a high QoS (Quality of Service) communication with a given cell loss or bit error rate via narrowband radio access networks (or via mobile systems), one has to equalize the channel distortions. Since fading can occur randomly in time, training sequence is not available to optimize the weights of the receiver, or the training sequence must be repeated frequently enough, which yields considerable overhead and thus a loss in data speed. This prompts the development of blind signal processing algorithms which can equalize the channel by only observing the output signal of a linearly distorted noisy channel. Traditional techniques apply modified MMSE and ZF algorithms when the decisions are used to replace the training sequence. Errors in the decisions, however, can give rise to severe instability. Therefore, the aim of this paper is to propose a novel decorrelation algorithm which does not use the decisions, but only the signal sequence observed at the output of the channels. The convergence of the algorithms can be proven by the well known Kushner-Clark method [10], while simulations demonstrate that the equalized channel characteristic is close to the optimal one and can support high QoS (low error rate) communication. The results will be treated in the following structure:

- first the model and some notations will be introduced in which framework the analysis will be carried out;
- then the novel blind decorrelation algorithm will be proposed, the convergence of which will be proven in mean square;
- the performance of the newly developed detector algorithm will be analyzed numerically and the simulation results will be compared to the performance of traditional multiuser detectors.

2. System Model

In this paper we investigate an asynchronous multipath propagation model. The transmitted baseband equivalent signal of the kth user is denoted by $q_k(t)$. Applying BPSK modulation, the kth user transmits $b_k[i] \in \{-1,1\}$ binary symbols, where i refers to the time instant. The transmitted signal is given as

$$q_k(t) = A_k \sum_{i=1}^{N_B} b_k[i] s_k\left(t - iT - \theta_k\right),$$

where A_k denotes the signal amplitude associated with the kth user, T refers to the time period of one symbol, θ_k denotes the delay of the kth user, and N_B is the block size, respectively. The spreading waveform of user k is denoted by $s_k(t)$, and can be written as follows

$$s_k(t) = \sum_{i=0}^{PG-1} S_k[i] e(t - iT_c).$$

Here $S_k[i] \in \{-1,1\}$denotes the ith time-chip of user k based on the Gold code set of length 31 [4]. $e(t)$ denotes the elementary waveform, and PG refers to the processing gain (PG = 31). The chip duration is denoted by $T_c = T/PG$. Each user transmits over a specific channel characterized by its impulse response function, $h_k(t)$. Thus, the received signal becomes

$$r(t) = \sum_{k=1}^{K} h_k(t) * q_k(t) + n(t), \tag{1}$$

where K refers to the number of users, and $n(t)$ is a white Gaussian noise with a constant one-sided N_0 spectral density.

Unfortunately, synchronous high bit-rate communication cannot be maintained in practice, due to multipath propagation. For the sake of generality, we assume a general channel impulse response function $h_k(t)$. To shorten the forthcoming formulas, we define a new function as

$$\phi_k(t) = h_k(t) * s_k(t - \theta_k). \tag{2}$$

Substituting this into (1) one obtains

$$r(t) = \sum_{k=1}^{K} \sum_{i=1}^{N_B} A_k \phi_k(t - iT) b_k[i] + n(t). \tag{3}$$

The $K \times K$ cross correlation matrix $\mathbf{\Phi}(t)$, containing all correlation information between $\phi_i(t)$ and $\phi_j(t)$, is defined by the dyadic convolution product

$$\Phi_{ij}(t) = \phi_i^*(-t) * \phi_j(t). \tag{4}$$

The kth matched filter performs a convolutional product on the incoming stream with $\phi_k^*(-t)$, which results in a continuous signal $\tilde{r}_k(t)$, which can be written as

$$\tilde{r}_k(t) = \phi_k^*(-t) * r(t). \tag{5}$$

Substituting (3) into (5) the equation above becomes

$$\tilde{\mathbf{r}}(t) = \sum_{i=1}^{N_B} \mathbf{\Phi}(t - iT) \mathbf{A}\, \mathbf{b}[i] + \tilde{\mathbf{n}}(t), \tag{6}$$

where $\mathbf{A}$ = diag $[A_1; A_2; ...; A_K]$, $\tilde{r}(t) = \tilde{r}_1(t); \tilde{r}_2(t); ...; \tilde{r}_K(t)^T$, $\mathbf{b}[i] = (\mathrm{b}_1[i]; \mathrm{b}_2[i]; ...; \mathrm{b}_K[i])^T$, and $\tilde{\mathbf{n}}(t) = \phi^*(-t) * n(t)$. Sampling with (iT), expression (6) results in a discrete-time model mapping $\mathbf{b}[i]$ into $\tilde{\mathbf{b}}[i]$, namely

$$\tilde{\mathbf{b}}[i] = \mathbf{R}[i] * \mathbf{A}\, \mathbf{b}[i] + \tilde{\mathbf{n}}[i], \tag{7}$$

where $\tilde{\mathbf{b}}[i] = \tilde{\mathbf{r}}(iT)$, $\tilde{\mathbf{n}}[i] = \tilde{\mathbf{n}}(iT)$ *and* $\mathbf{R}[i] = \mathbf{\Phi}(iT)$, which is the discrete-time channel matrix. To describe the detection schemes in an easy-to-understand manner, it is helpful to write $\tilde{\mathbf{b}}[i]$, $\mathbf{b}[i]$ and $\tilde{\mathbf{n}}[i]$ into column vectors with KN_B elements each, and use block notation.

$$\begin{aligned}
\tilde{\mathbf{b}} &= [\tilde{b}_1[1], \tilde{b}_2[1], \ldots, \tilde{b}_K[1], \tilde{b}_1[2], \ldots, \tilde{b}_K[N_B]]^T, \\
\mathbf{b} &= [b_1[1], b_2[1], \ldots, b_K[1], b_1[2], \ldots, b_K[N_B]]^T, \\
\tilde{\mathbf{n}} &= [\tilde{n}_1[1], \tilde{n}_2[1], \ldots, \tilde{n}_K[1], \tilde{n}_1[2], \ldots, \tilde{n}_K[N_B]]^T,
\end{aligned}$$

respectively. The convolutional product in (7) can then be rewritten as a multiplication:

$$\tilde{\mathbf{b}} = \mathbf{R}\, \mathbf{A}^A \mathbf{b} + \tilde{\mathbf{n}}, \tag{8}$$

where $\mathbf{A}^A$ = diag $[\mathbf{A}; \mathbf{A}; ...; \mathbf{A}]$ is a $KN_B \times KN_B$ dimensional diagonal matrix, containing NB sub-matrices of $\mathbf{A}$. Moreover, the matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}[0] & \mathbf{R}[-1] & \ldots & \mathbf{R}[-N_B+1] \\ \mathbf{R}[1] & \mathbf{R}[0] & \ldots & \mathbf{R}[-N_B+2] \\ \vdots & & \ddots & \\ \mathbf{R}[N_B-1] & \ldots & \mathbf{R}[1] & \mathbf{R}[0] \end{bmatrix}$$

is a $KN_B \times KN_B$ dimensional block Toeplitz matrix and Hermitian ($\mathbf{R}[i] = (\mathbf{R}[-i])^H$ for all i) with diagonal dominance ($|\mathrm{R}_{ii}| > |\mathrm{R}_{ij}|\ \forall\ (i, j);\ i \neq j$). Furthermore, it is a sparse matrix, in most practical situations only a few submatrices differ significantly from zero ($\mathbf{R}[k], \forall\ |k| \leq 6$).

The traditional single-user detector (SUD) applies a simple signum decision function after (8), yielding

$$\hat{\mathbf{b}}^{\mathrm{SUD}} = \mathrm{sign}\left\{\tilde{\mathbf{b}}\right\} = \mathrm{sign}\left\{\mathbf{R}\,\mathbf{A}^A\mathbf{b} + \tilde{\mathbf{n}}\right\} \tag{9}$$

which cannot combat multiple access interference efficiently. (In the rest of the paper a variable with "hat" refers to the estimated value of the original variable without "hat"). If $\mathbf{R}$ is diagonal then the single-user detector defined in (9) offers an appealing solution. However, in most cases $\mathbf{R}$ is definitely not diagonal, which is mainly due to the existence of interference sources on the channel. This problem motivates the use of MUD techniques [1]. An algorithmically simple but near optimal multiuser detector architecture is the so-called stochastic neural network. Its performance is analyzed in [8]. The proof about its dynamics and stationary behavior is given in [9]. On the operation of the stochastic Hopfield neural network (SHN) is described by its state transition rule, given as

$$Y_l[\ell] = \mathrm{sign}\left\{\sum_{j=1}^{l-1} W_{lj}Y_j[\ell] + \sum_{j=l}^{M} W_{lj}Y_j[\ell-1] + V_l + \nu_l[\ell]\right\}, \tag{10}$$

where $\mathrm{Y}_l[\ell]$ is the output of the lth neuron at the ℓth iteration (in one iteration the whole network is updated, so $l = 1,2,...,M$), and M denotes the dimension of the network. V_l is the decision threshold of the lth neuron and W_{lj} is the connection weight between the output of the jth neuron and the input of the lth neuron. Parameter $\nu_l[\ell]$ is a white noise that follows a logistic distribution

$(\Pr\{\nu_l[\ell] = \nu < x\} = \phi_\alpha(x) = \frac{1}{1+e^{-\alpha x}})$

with decreasing variance (as $\alpha \to \infty$). The parameters of the network in (10) should be adjusted as follows:

$$M = K \cdot N_B, \quad \mathbf{W} = -\left(\mathbf{R} - \mathrm{diag}\left[\mathbf{R}\right]\right), \quad \mathbf{V} = \tilde{\mathbf{b}}$$

$$\hat{\mathbf{b}}^{\mathrm{SHN}} = \mathrm{sign}\left\{\lim_{\ell \to \#it} \mathbf{Y}[\ell]\right\}, \tag{11}$$

where #it is the number of performed iterations. Further quantitative specification will be given in section 4.

3. A novel blind channel equalization algorithm

In this section a novel adaptive decorrelation algorithm will be described, which is not only able to perform inverse channel identification but can also eliminate multiuser and intersymbol interference as well. The algorithm is given as

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \Delta_k\left(\check{\mathbf{b}}\,\check{\mathbf{b}}^T - \mathbf{I}\right), \tag{12}$$

where $\mathbf{W}_k$ denotes the current estimate of the channel matrix; $\check{\mathbf{b}} = \mathbf{W}_k\tilde{\mathbf{b}}$; $\mathbf{I}$ is the identity matrix, and Δ_k is the learning parameter (which was decreased linearly with time in the simulations). When analyzing the steady state of (12) in mean square, one can obviously obtain $\mathrm{E}[\check{\mathbf{b}}\check{\mathbf{b}}^T] = \mathbf{I}$. This implies that in stationary state the multiuser interference is eliminated if the algorithm is stable. Therefore, in case of stability, this method generates the inverse estimate of the original $\mathbf{R}$, which can then be plugged into the SHN algorithm:

$$\hat{\mathbf{R}}_k^{\mathrm{ADM}} = \mathbf{W}_k^{-1}.$$

Here index "ADM" refers to Adaptive Decorrelation Method. In the next subsection we analyze the convergence properties of this new algorithm.

3.1. The steady state and the convergence properties of ADM

To analyze the convergence in mean square, the following lemma will be of help:

Lemma 1 *If a random vector* $\mathbf{x}$ *is given with covariance matrix* $\mathbf{C}$, *then* $\mathbf{x}$ *can be decorrelated by multiplying with any matrix* $\mathbf{W}$ *that satisfies*

$$\mathbf{W}\mathbf{W}^T = \mathbf{C}^{-1} \tag{13}$$

PROOF. Let us denote $\mathbf{z} = \mathbf{W}\mathrm{x}$, where z is supposed to be uncorrelated. Then

$$\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right] = \mathbf{I} = \mathrm{E}\left[\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T\right] = \mathbf{W}\mathbf{C}\mathbf{W}^T \tag{14}$$

and so

$$\mathbf{C} = (\mathbf{W}^{-1})(\mathbf{W}^{-1})^T$$

which implies the statement of the Lemma.

Substituting (14) into (12) the algorithm yields

$$\mathbf{W}(k+1) = \mathbf{W}(k) - \Delta\left\{\mathbf{W}\mathbf{x}\mathbf{x}^T\mathbf{W}^T - \mathbf{I}\right\} \tag{15}$$

Taking the expectation of both sides of (15) results in

$$\mathrm{E}\left[\mathbf{W}_{opt}\mathbf{x}\mathbf{x}^T\mathbf{W}_{opt}^T\right] = \mathbf{I}$$

$$\mathbf{W}_{opt}\mathbf{R}\mathrm{E}\left[\mathbf{y}\mathbf{y}^T\right]\mathbf{R}^T\mathbf{W}_{opt}^T = \mathbf{I}.$$

Since the transmitted information sequences are supposed to be independent and due to the symmetrical property of both matrices

$$\mathbf{W}_{opt}\mathbf{R}^2\mathbf{W}_{opt} = \mathbf{I}, \tag{16}$$

$$(\mathbf{W}_{opt})^2 = (\mathbf{R}^{-1})^2. \tag{17}$$

The received vector x has a covariance matrix of $\mathbf{R}\mathbf{R}^T$, hence, in the sense of Lemma 1, it is proven that the steady-states of the ADM are the matrices that decorrelate the received vector. Now we embark on proving the convergence of the ADM recursion (12) in mean square by using the Kushner-Clark Theorem. First we brie y sight the theorem [10]

Theorem 2 (Kushner-Clark) *If an arbitrary nonlinear stochastic recursion is given in the form of* $\mathrm{w}(k+1) = \mathrm{w}(k) + \Delta(k)\Psi(\mathrm{w}(k), \mathrm{x}(k))$ *and the following properties are satisfied:*

1. $\sum_{k=1}^{\infty}\Delta(k)=\infty$, $\sum_{k=1}^{\infty}\Delta(k)^p<\infty\ \forall p>1$ *and* $\lim_{k\to\infty}\Delta(k)=0$;
2. *The following partial derivatives exist:* $\frac{\partial\Psi(\mathbf{w},\mathbf{x})}{\partial w_i}$, $\frac{\partial\Psi(\mathbf{w},\mathbf{x})}{\partial x_i}$ $\forall i$;
3. *The following limit exist:* $\Psi(\mathbf{w})=\lim_{k\to\infty}\mathrm{E}_x\left[\Psi(\mathbf{w}(k),\mathbf{x})\right]$

then $\mathrm{l.i.m}_{k\to\infty}\mathbf{w}(k)=\mathbf{w}^*$ *where* $\frac{d\mathbf{w}(t)}{dt}=\Psi(\mathbf{w}(t))$ *and* $\mathbf{w}^*:\ \ \Psi(\mathbf{w}^*)=0$.

In this way the convergence analysis of nonlinear stochastic recursion can be reduced to the analysis of ordinary differential equations.

Theorem 3 *The stochastic recursion (12) converges to* $\hat{\mathbf{R}}_k^{\mathrm{ADM}}=\mathbf{W}_k^{-1}$ in mean square.

Proof. Let us apply the Kushner-Clark theorem to our recursion. First we should check whether recursion (12) satisfies the conditions of the theorem. Property 1. can be easily met by setting $\Delta = 1/k$. The function $\Psi(\mathbf{W},\mathrm{x}) = \mathbf{W}\mathrm{x}\mathrm{x}^T\mathbf{W}-\mathbf{I}$, which is differentiable in both W_{ij} and x_i. Hence, property 2. is satisfied. And, finally, the limit in Property 3. amounts to

$$\Psi(\mathbf{W})=\mathbf{W}\mathbf{R}^2\mathbf{W}-\mathbf{I}, \tag{18}$$

which leads us to the analysis of the following matrix Riccati differential equation [11]:

$$\frac{d\mathbf{W}(t)}{dt}=\mathbf{W}(t)\mathbf{R}^2\mathbf{W}(t)-\mathbf{I}. \tag{19}$$

The asymptotic solution of (19) is given by Kwakernaak-Sivan theorem [11], stating that as the terminal time t_1 approaches infinity, the solution $W(t)$ of the matrix Riccati equation

$$-\dot{\mathbf{W}}=-\mathbf{W}(t)\mathbf{R}^2\mathbf{W}(t)+\mathbf{I}.$$

Here the dot above $\mathbf{W}$ refers to the differentiation with respect to time. Starting from the initial condition $\mathbf{W}(0) = \mathbf{W}_0$ $\mathbf{W}(t)$ generally approaches a steadystate solution $\bar{\mathbf{W}}$ that is independent of W_0. The steady-state solution $\bar{\mathbf{W}}$ is the nonnegative-definite solution of the algebraic Riccati equation, given as

$$\mathbf{0}=\mathbf{I}-\bar{\mathbf{W}}\mathbf{R}^2\bar{\mathbf{W}} \tag{20}$$

(For further details on the Kwakernaak-Sivathe theorem see [11]). Using elementary matrix algebra one obtains

$$\bar{\mathbf{W}}^2=\mathbf{R}^{-2}. \tag{21}$$

Matrix $\mathbf{R}$ is positive definite, therefore it has only eigenvalues with positive real part. In the sense of the Kwakernaak-Sivan theorem, $\bar{\mathbf{W}}$ is positive semi-definite, which implies

$$\bar{\mathbf{W}}=\mathbf{R}^{-1}. \tag{22}$$

This completes the proof that the ADM recursion (12) converges to the inverse of the channel matrix in mean square.

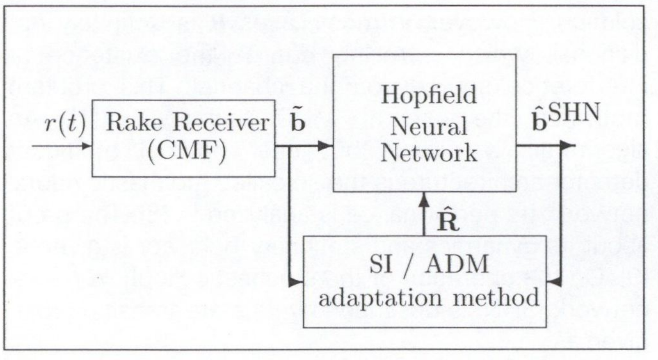


Figure 2. The Structure of the Proposed Detector Schemes

3.2. The effect of Additive Gaussian Noise

So far we have only investigated the noiseless case, but in practice a channel noise ($N(0, \mathbf{K})$) is to be taken into account. This noise will slightly modify (14), resulting in

$$\tilde{\mathbf{y}}=\mathbf{Q}\mathbf{y}+\mathbf{W}\mathbf{n}, \tag{23}$$

where $\mathbf{Q} = \mathbf{RW}$ implying that the identity matrix is the steady state.

$$\begin{aligned}\mathrm{E}\left[\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T\right] &= \mathrm{E}\left[\mathbf{Q}\mathbf{y}\mathbf{y}^T\mathbf{Q}^T+\mathbf{W}\mathbf{n}\mathbf{y}^T\mathbf{Q}^T+\mathbf{Q}\mathbf{y}\mathbf{n}^T\mathbf{W}^T+\mathbf{W}\mathbf{n}\mathbf{n}^T\mathbf{W}^T\right]= \\ &= \mathbf{Q}\mathbf{Q}^T+\mathbf{W}\mathbf{K}\mathbf{W}^T\end{aligned} \tag{24}$$

The second term of this equation can be seen as a bias term, which can be removed by altering the original ADM algorithm (12) as

$$\mathbf{W}(k+1)=\mathbf{W}(k)-\Delta\left\{\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T-\mathbf{A}\right\} \tag{25}$$

where

$$\mathbf{A}=\mathbf{I}+\mathbf{W}\mathbf{K}\mathbf{W}^T.$$

The covariance matrix $\mathbf{K}$ in most of the cases is assumed to be diagonal with elements N_0.

4. Performance Analysis

For the sake of better comprehension, the proposed architectures are depicted in Figure 2. Note that the perfect operation of the channel matched filter is assumed, although only its soft output values are used in the detector. The estimation of $\mathbf{R}$ is the task of the adaptation method, depicted at the bottom of the picture. The estimation method relies on the output of the detector ($\hat{b}$) and the output of the channel matched filter ($\tilde{b}$). At the same time the stochastic Hopfield network was operating with the estimated values of $\mathbf{R}$, denoted by $\hat{\mathbf{R}}$.

In this section the performance of the new algorithm is analyzed by simulations. The following notations and scenario was adopted:

- Number of users (u)
- Block length (b)
- Channel model: urban
- Spreading codes: Gold codes
- Disturbing effects: ISI, MUI, AWGN

The performance is investigated by plotting the Bit Error Rate (BER)against the Signal to Noise Ratio (SNR). In the first figure, the BER(SNR) curve is exhibited in the case of applying the SHN detection algorithm. In Figure 3

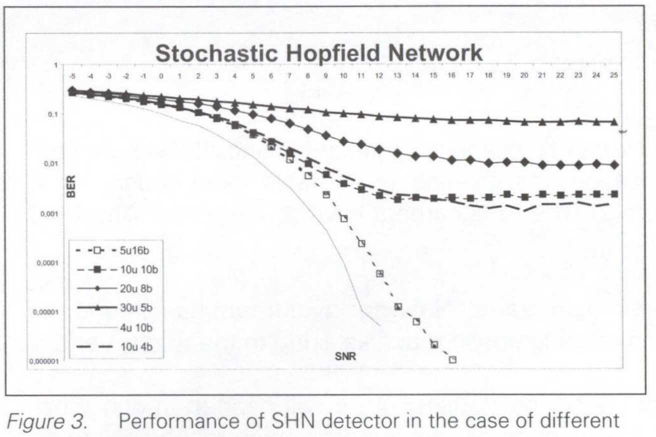


Figure 3. Performance of SHN detector in the case of different channel models

one can see that the larger the number of users are in the system the worse the performance is due to the increased MUI. On the other hand, the block length does not have a great impact on the performance (the results obtained for channel 10u10ba and for channel 10u4b almost coincide).

In the next figure the BER-SNR is plotted in the case of different detection algorithm based on the channel model 4u10b. The list of simulated detection algorithms is given as follows:

- Single User Detector (SUD)
- Blind Additive Decorrelation Method (ADM)
- Stochastic Hopfield Neural Network (SHN)
- Simple Identificator (SI).

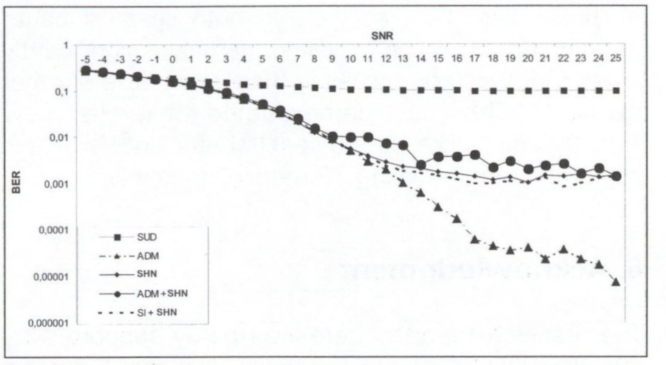


Figure 4. Performance of different detectors in the case of different noise powers for an 4 user 10 block length channel

In Figure 4 one can see that the best performance can be obtained when the channel is assumed to be known and the detection is done by SHN. On the other hand, when the channel is identified by ADM only a slight performance degradation can be observed. The performance is further reduced by only applying ADM as decorrelator followed by the simple threshold detector. Nonetheless, the performance is relatively good given in this case which further prompts the application of ADM. Of course the worst result was obtained by applying simply SUD.

The next figure shows the performance of the same detector algorithms in the case of 10u4b channel. The same conclusions can be drawn form this figure as form the previous one.

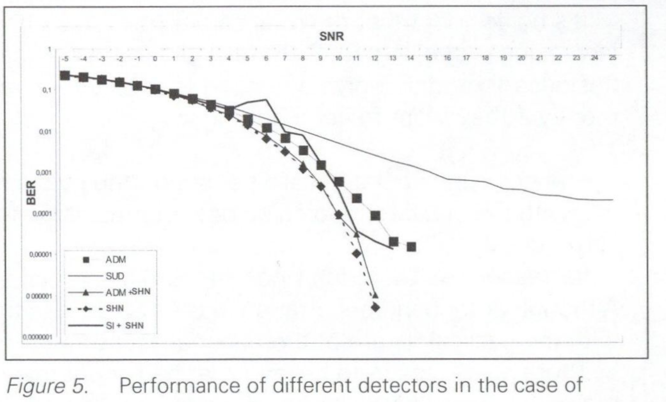


Figure 5. Performance of different detectors in the case of different noise powers for an 10 user 4 block length channel

5. Conclusions

In this paper a novel blind signal detection algorithm was introduced capable of both channel identification and multiuser detection. When the new algorithm was used as a blind channel identifier, then stochastic Hopfield net carried out the detection based on the identified channel characteristics. Since the new algorithm is capable of decorrelating weakly stationer stochastic sequences it has been directly applied to multiuser detection, as well. In both cases superior performance could be achieved, namely a significantly lower BER was accomplished than in the case of other methods. This has demonstrated that the new algorithm can increase the spectral efficiency which is imperative in introducing 3G mobile standards.

6. Acknowledgment

The research reported here was partly supported by project T025964 of the Hungarian National Scientific Foundation, and Project GOA/98/06 of Research Fund K.U.Leuven.

References

1. S. Verdu, Multiuser Detection, Cambridge University Press, 1998.
2. J. G. Proakis, Digital Communications, McGraw-Hill, 1995.
3. T. Kohonen, Self-Organizing Maps, Springer, 2000.
4. R. Gold, "Optimum Binary Sequences for Spread Spectrum Multiplexing," IEEE Trans. Information Theory, vol. 14, pp. 154-156, 1967.
5. J. J. Hopfield, D. W. Tank, "Neural Computation of Decision in Optimization Problems" Biological Cybernetics, vol. 52, pp. 141-152, 1985.
6. A. Hottinen, "Self-organizing Multiuser Detection," IEEE Symp. Spread Spectrum Techn. and Appl., vol. 1, pp. 152-156, 1994.
7. J. Levendovszky, L. Kovács, "A new blind signal processing algorithm for decorrelation and channel equalization," Accepted for presentation at the VIPromCom2001 Symposium Zadar, Croatia, Jun13-15, 2001.
8. G. Jeney, J. Levendovszky, "Stochastic Hopfield Network for Multi-user Detection" European Conf. Wireless Techn., pp. 147-150, 2000.
9. G. Jeney, J. Levendovszky, S. Imre, "Near Optimal MUD Applying Stochastic Hopfield Network," Submitted to IEEE Trans. Circuits and Systems.
10. Kushner, H. J., and Clark, D. S., Stochastic Approximation Methods for Unconstrained Systems, Springer 1978.
11. H. Kwakernaak and R. Sivan, Linear Optimal Control Systems, Wiley-Interscience, 1972.

News

Telcos situation (in USA) by Dan O'Shea

It's no accident that the rural carrier market is one of telecom's healthiest markets financially. Vendors are increasingly seeking RUS (Rural Utility Service) certification and chasing the smaller independent telco business they didn't give a second look to a few years ago. The rural carriers have to keep spending RUS money if they want to keep getting it.

The non-rural operators, are not supported by such special programs. Network investment has ground toa halt with the capital spending spigots of most large telcos offering not so much as a drip to the thirsty vendor community.

Its reason is: They don't need it. Broadband is a wonderful inevitability, but it will happen in due time. Although broadband penetration in the U.S. is less than many countries, the overwhelming reason has to do with the cyclical nature of the national economy.

There is no reason to believe that bad economic conditions will last forever. After the bad days are over, network investment will pick up again because it will have to.

The rural carriers are in a different situation entirely. They serve customer bases that are costly to serve because they are so spread out, and they provide service to customers that bigger telcos, for the most part, want absolutely nothing to do with. Though some people may see the RUS program as subsidizing this country's rural infrastructures, it's really not much more than an incentive that – effectively or not – is designed to keep rural users in the game.

Certain industry interests will continue to fight for the broadband subsidy, but it is only in the interests of saving themselves from their own financial problems. A debate will also only serve to complicate and draw attention away from the real issues affecting telecom regulation. The temptation is there, and increasingly harder to resist, but we must resist it if we want to have a balanced, non-monopolistic industry. Subsidizing a broadband buildout for the biggest companies will effectively put them that much further ahead of their potential competitors when the dust of the downturn dissipates.

Analytical Approach to a Large Deviation-Based CAC Algorithm

TAMÁS JAKABFY,
Eötvös Loránd University, Faculty of Science

ÁRPÁD SZLÁVIK, GERGELY SERES
Traffic Analysis and Network Performance Laboratory, Ericsson Research Hungary,
{Tamas.Jakabfy, Arpad.Szlavik, Gergely.Seres}@eth.ericsson.se

Reviewed

This paper deals with analytical traffic models that attempt to construct models that take the limited burst length into consideration. Based on these models the paper attempts to create admission rules that can be used to further increase the utilisation of the network without sacrificing service quality.

1. Introduction

Connection admission control (CAC) is a key resource management function in multiservice networks. Its role is twofold. It must limit the amount of traffic admitted into the network, so as to ensure that all existing and newly admitted flows get their contracted level of service quality. Yet, the admission rule can not be excessively strict, as this would result in low network utilisation and consequently in lost profit for the network operator. Thus the right balance has to be found between raising the utilisation of the network and protecting traffic from congestion through leaving spare network capacity. The indicator of a well performing connection admission control algorithm is that the observed service quality and the targeted level of service quality are very close to each other.

The core component of a connection admission control algorithm is the quantification of the resource status of the network with the assumption that the new flow is already added to the system. This can be done either by estimating the amount of used service capacity orby the determination of the expected loss or delay in the system.

There are two markedly different approaches to connection admission control. In the first approach, a traffic model is used to describe the general behaviour of the traffic. The parameters of the traffic model are then used to close the gap between the model (the general template) and the actual traffic. These imply that there has to be a trust in the model, i.e. it is assumed that there is a good alignment between the model and the reality. Also, the parameters that fit the model to the real traffic must be made available somehow. Either it can be assumed that there are a small number of traffic classes that can be described in advance, or the traffic characterisation has to be left to the user, i.e. they must provide the parameters upon initiating a new traffic flow. The alternative to this strategy is the method of measurement-based admission control. This method exploits the fact that the resource status of the network (utilisation, anticipated loss & delay) can be estimated by measuring the statistical properties of the traffic. This approach does not need neither a traffic model, nor parameters.

The traffic model and parameter based approach has been a popular approach for constructing CAC algorithms, especially for asynchronous transfer mode (ATM) networks. However, most of the algorithms developed in the literature make overly limiting assumptions that increase the gap between the traffic model used and the real traffic. Such an assumption is that the bursts in the traffic can be infinitely large. Most current multiservice systems use some form of traffic policing, an essential component of which is the token bucket. This function is responsible for limiting the length of the bursts the traffic source is allowed to send into the network. This way, traffic bursts encountered inside the network are truncated at a certain maximum length. This fact is ignored by most current CAC algorithms.

The rest of the paper is organised as follows. Section 2 gives the mathematical theoretical background. Next, in Section 3 the scaled cumulant generating function is calculated for some traffic models. Section 3.1 describes the known results for general Markovmodulated ON-OFF sources. Section 3.2 introduces a general approach for analysing continous-time ON-OFF models. The models of Sections 3.3.1 and 3.3.2 use uniformly and truncated exponentially distributed ON period lengths, respectively. The new approach is demonstrated in Section 3.3.3 for a very simple, almost deterministic system. Conclusions are given in Section 4.

2. Mathematical Background

The large buffer asymptotics of the large deviation theory [1, 2, 3, 4, 5] is concerned with the estimation of the buffer overflow probability when the buffer size gets very large. Consider a single-server queueing system with buffer size b and service rate c. Let Q denote the queue length in the system. The large buffer asymptotics state that the decay rate of the logarithm of the buffer overflow probability is asymptotically linear in b, as b approaches infinity [2]:

$$\lim_{b\to\infty}\frac{1}{b}\ln P(Q>b) = -\delta(c), \tag{1}$$

where $\delta(c)$ is the rate function of the tail probability. It is computed as

$$\delta(c) = \sup\{s>0 : SCGF(s) \le sc\} \tag{2}$$

and is parameterised by the total amount of arriving work through $SCGF(s)$ and by the service rate c. The scaled cumulant generating function (SCGF) of the arrival process is defined as

$$SCGF(s) := \lim_{t\to\infty}\frac{1}{t}\ln E\left[e^{sX[0,t)}\right], \tag{3}$$

where $X[0, t]$ is the total amount of work arrived in the time interval $[0, t]$.

Once the decay rate is determined from the optimisation defined in (2), the buffer overflow probability can be computed according to (1) for large buffers:

$$P(Q>b) \approx e^{-b\delta(c)}.$$

The equivalent capacity, defined as the minimal service rate for which the QoS constraint is fulfilled, can be calculated analytically [2]:

$$c_{equ} := \inf\left\{c : e^{-b\delta(c)} \le \epsilon\right\} = \\ = \frac{SCGF(\delta(c_{equ}))}{\delta(c_{equ})} = \frac{SCGF(\gamma/b)}{\gamma/b}. \tag{4}$$

where the target overflow probability is $\epsilon = e^{-\gamma}$

The CAC algorithm uses this equivalent capacity for its decision: if a new connection arrives, it checks whether the sum of this value and the peak rate of the new connection is less than the service rate of the outgoing link. If $c_{equ} + p \le c$ holds, the new connection is accepted, otherwise it is rejected. It can be seen from (4) that the CAC algorithm uses the scaled cumulant generating function of the arrival process. This function can be estimated from measurements or it can be calculated analytically when assuming specific traffic models.

3. Calculating the Scaled Cumulant Generating Function for Some Traffic Models

The goal is to calculate the scaled cumulant generating function for traffic models as realistic as possible. A general analytical result is available for the scaled cumulant generating function of the superposition of Markov-modulated ON-OFF sources (see Section 3.1), but this model involves the assumption of unbounded ON periods (arbitrary long bursts can occur with positive probability) which is not realistic enough in practice. This is why analytical studies were mainly concerned with arrival processes with bounded burst behaviour.

An overview of the scaled cumulant generating functions computed for previous traffic models is presented in [6] and [7].

The central idea is that the expected value in the definition of the scaled cumulant generating function can be computed from the distribution of $X[0, t]$ the total amount of work arroved in the time interval $[0, t]$. It is enough to compute the scaled cumulant generating function of one source because the SCGF of the sum of more independent sources is equal to the sum of the SCGF of the sources.

3.1. Markov-modulated ON-OFF Model

The scaled cumulant generating function for the superposition of multi-class Markov-modulated ON-OFF sources can be computed from an analytical expression [8]. If the arrival process $X[0, t]$ is described by atwo-state Markov chain with transition rates μ and η and instantaneous arrival rates p and 0, then its scaled cumulant generating function (3) can be expressed [8] as:

$$SCGF(s) = \lim_{t\to\infty}\frac{1}{t}\ln\left(\left[\frac{\mu}{\eta+\mu}, \frac{\mu}{\eta+\mu}\right]\times \right.$$

$$\left.\times \exp\left[\begin{pmatrix} -\mu+ps & \mu \\ \eta & \eta \end{pmatrix} t\right]\times\begin{pmatrix}1\\1\end{pmatrix}\right] =$$

$$= \frac{1}{2}\left(ps-\mu-\eta+\sqrt{(ps-\mu-\eta)^2+4\eta\mu}\right).$$

In the discrete-time equivalent of this continuous-time case we replace the exponential distribution with the geometric distribution in the two-state Markov chain. An example of such a discrete model is that in each time-slot p amount of traffic is generated with probability q and no traffic is generated with probability $1-q$. In this model $P(X[0,n]=kp)=\binom{n}{k}q^k(1-q)^{n-k}$, and the scaled cumulant generating function is

$$SCGF(s) = \lim_{n\to\infty} \frac{1}{n} \ln \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} e^{kps} =$$

$$= \lim_{n\to\infty} \frac{1}{n} \ln \sum_{k=0}^{n} \binom{n}{k} \left(qe^{ps}\right)^k (1-q)^{n-k} =$$

$$= \lim_{n\to\infty} \frac{1}{n} \ln \left(qe^{ps} + 1 - q\right)^n = \ln\left(qe^{ps} + 1 - q\right).$$

3.2. Continuous-time ON-OFF Models in General

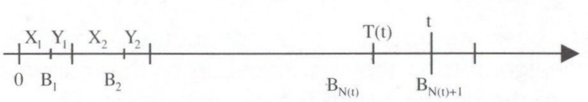
Let X_1, Y_1, X_2, Y_2,... (or Y_1, X_1, Y_2, X_2,...) denote the lengths of the successive ON and OFF (or OFF and ON) periods of the arrival stream. The lengths of the ON and OFF periods are identically distributed –X and Y respectively and all of them are independent. Consider the random variable $B_i = X_i + Y_i$, the length of the i^{th} ON-OFF (or OFF-ON) block $(B=X+Y)$. Two important random variables can be defined:

$$N(t) := \sup\left\{k : \sum_{i=1}^{k} B_i \le t\right\}$$

the maximal number of blocks that can be fitted in the time interval $[0, t]$ and

$$T(t) := \sum_{i=1}^{N(t)} B_i$$

the ending time of the last block that still fits into



The distribution function of $X[0, t]$ can be ex-pressed using these definitions as

$$P\left(X[0,t) < x\right) = P\left(X^{*N(t)} + R(t) < x\right),$$

where $R(t)$ is the amount of work arriving at the system in the last t–$T(t)$ long time interval. The distribution of $X[0, t]$ can be approximated as

$$\sum_{k=1}^{\infty} P\left(X^{*N(t)} > x \,|\, N(t) = k) P(N(t) = k\right) \le$$
$$\le P\left(X[0,t) > x\right) \le$$
$$\le \sum_{k=1}^{\infty} P\left(X^{*(N(t)+1)} \ge x \,|\, N(t) = k) P(N(t) = k\right), \qquad (5)$$

provided that the distribution of $N(t)$ is known. Note the connection between the random variable $N(t)$ and the length of the blocks:

$$P(N(t) \ge k) = P\left(\sum_{i=1}^{k} B_i < t\right) =$$
$$= P(B^{*k} < t) =: F_{B^{*k}}(t).$$

If all the convolutions of B are known, $B^{*k} = X^{*k} * Y^{*k}$, the distribution of $N(t)$, also used in (5), is attained as

$$P(N(t)=k) = P(N(t) \ge k) -$$
$$- P(N(t) \ge k+1 = F_{B^{*k}}(t) - F_{B^{*(k+1)}}(t) \qquad (6)$$

The random variable $N(t)$ can be used to arrive at the distribution function of t–$T(t)$ with:

$$P(t - T(t) > x | N(t) = k) =$$

$$= \int_0^{t-x} (1 - F_B(t-y)) f_{B^{*k}}(y) dy.$$

In this approach the key is the knowledge of the convolutions of B.

3.3. Random and Bounded ON Bursts

In the following scenarios the length of the OFF period is kept constant (l_{OFF}), but the length of the bursts is assumed to be either a uniformly or truncated (bounded) exponentially distributed. The reasoning below follows the general approach outlined in the previous section.

As seen in (6) the distribution of $N(t)$ requires the convolutions of B, and because of the constant OFF period lengths, the distribution of B^{*k} is:

$$P(B^{*k} < t) = P(X^{*k} < t - kl_{OFF}) \qquad (7)$$

3.3.1. Uniformly Distributed Burst Lengths

If the lengths of the ON periods are uniformly distributed in [0, 1], for the sake of simplicity and for establishing basic results reused in more complex cases. $F_{B^{*k}}(t)$ can be computed as

$$F_{B^{*k}}(t) = \frac{1}{k!} \sum_{i=0}^{\lfloor t - kl_{OFF} \rfloor - 1} (-1)^i \binom{k}{i} (t - kl_{OFF} - i)^k,$$

using the density function of the k th convolution of an arbitrary uniformly distributed random variable U (see [9]):

$$f_{U^{*k}}(x) = \frac{1}{(k-1)!} \sum_{j=0}^{i} (-1)^j \binom{k}{j} (x-j)^{k-1}, \qquad (8)$$

for $x \in [i, i+1]$ and $i \in \{1\ 2\ ..., k-1\}$.

It is expected that the limit $(t \to \infty)$ of the corresponding lower and upper bounds from the approximations of $P(X[0,t] < x)$ in (5) lead to closed-form expressions. These result in analytical lower and upper bounds for the scaled cumulant generating function. The other required quantity in (5), besides the probability distribution of $N(t)$, is $P(N^{*N(t)}>x|N(t)=k)=P(X^{*k}>x)$. This can be calculated directly from the k^{th} convolution of the uniformly distributed X by (8) with $U=X$.

3.3.2. Truncated Exponentially Distributed ON Period Lengths

Truncated exponential distribution is a bounded variant of the exponential distribution. Its density function is

$$f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x}}{1-e^{-\lambda l_{\mathrm{ON}}}} & \text{if } x \in [0, l_{\mathrm{ON}}) \\ 0 & \text{otherwise} \end{cases}$$

Remark: If we have a random number generator producing exponentially distributed random samples (X'), we can generate truncated exponential random numbers with the modulo operation:

$$P(X < x) = P(X' \; mod \, l_{\mathrm{ON}} \; < \; x) =$$

$$= P(X' \in [0, x) \cup [l_{\mathrm{ON}}, l_{\mathrm{ON}} + x) \cup \ldots) =$$

$$\sum_{n=0}^{\infty} P(X' \in [nl_{\mathrm{ON}}, nl_{\mathrm{ON}} + x)) =$$

$$= \sum_{n=0}^{\infty} \int_{nl_{\mathrm{ON}}}^{nl_{\mathrm{ON}}+x} \lambda e^{-\lambda t} dt = \sum_{n=0}^{\infty} \left[-e^{-\lambda t} \right]_{t=nl_{\mathrm{ON}}}^{nl_{\mathrm{ON}}+x} =$$

$$= \sum_{n=0}^{\infty} \left(e^{-\lambda n l_{\mathrm{ON}}} - e^{-\lambda (n l_{\mathrm{ON}} + x)} \right) =$$

$$= \frac{1}{1-e^{-\lambda l_{\mathrm{ON}}}} - e^{-\lambda x} \frac{1}{1-e^{-\lambda l_{\mathrm{ON}}}} = \frac{1-e^{-\lambda x}}{1-e^{-\lambda l_{\mathrm{ON}}}}.$$

Although this time the ON period lengths have truncated exponential distribution (instead of the uniform distribution in the previous subsection), $F_{B^{*k}}(t)$ can be computed by the convolutions of the uniform distribution from (8) and by (7):

$$f_{X^{*k}}(x) = \frac{\lambda^k e^{-\lambda x}}{\left(1-e^{-\lambda l_{\mathrm{ON}}}\right)^k} f_{U^{*k}}(x/l_{\mathrm{ON}}). \tag{9}$$

In order to determine the upper and lower bounds of the scaled cumulant generating function, the bounds in (5) have to be computed. The first part of the bounds, $P(X^{*k(t)}>x|N(t)=k)=P(X^{*k}>x)$ and $P(X^{*N(t)+1)}>x|N(t)=k)=P(X^{*k}>x)$, can directly be calculated from (9). The determination of the second part, $P(N(t)=k)$, involves the conversion of (9) to a distribution function and a series of substitutions into equations (7) and (6). Unfortunately at the moment this procedure does not lead to closed form expressions, thus numerical methods have to be used from equation (7).

3.3.3. Random-start Deterministic ON-OFF Model

This model works with constant OFF period duration l_{OFF} and constant duration of ON bursts of length l_{OFF}. The only random part in this model is that the arrival process begins with an ON burst with probability q and with an OFF period with probability $(1-q)$. The values of $N(t)$ and $T(t)$ in this case are deterministic:

$$N(t) = \left\lfloor \frac{t}{l_{\mathrm{ON}} + l_{\mathrm{OFF}}} \right\rfloor \text{ and}$$

$$T(t) = N(t)(l_{\mathrm{ON}} + l_{\mathrm{OFF}}).$$

The distribution function of $X[0, t]$ consists of three parts: it is $N(t)l_{\mathrm{ON}}$ with probability $(1-q)\frac{l_{\mathrm{OFF}}}{l_{\mathrm{ON}}+l_{\mathrm{OFF}}}$, it is $(N(t)+1)l_{\mathrm{ON}}$ with probability $q\frac{l_{\mathrm{OFF}}}{l_{\mathrm{ON}}+l_{\mathrm{OFF}}}$ and it is uniformly distributed in $[N(t)l_{\mathrm{ON}}, (N(t)+1)l_{\mathrm{ON}}]$ with probability $\frac{l_{\mathrm{ON}}}{l_{\mathrm{ON}}+l_{\mathrm{OFF}}}$.

After some calculations the scaled cumulant generating function (3) can be determined:

$$SCGF(s) = \frac{sl_{\mathrm{ON}}}{l_{\mathrm{ON}} + l_{\mathrm{OFF}}}.$$

4. Conclusion

A very basic component of a connection admission control algorithm is the quantification of the resource status of the network with the assumption that a new connection is already added to the system. This can be done, for instance, by estimating the amount of used outgoing link capacity. One possible approach is to measure the traffic to obtain this estimate, another is to set up a parameterised traffic model for the real system in order to calculate this estimate analytically. One of the main advantages of the measurement-based approach is that it does not need preliminary information on the traffic. The analytical models, on the other hand, have the advantage of having closed-form and thus easily applicable formulae. Therefore their application is usually less complex and less time consuming than their measurement-based counterparts.

As most of the existing analytical results for the large deviation-based connection admission control methods work with the very unrealistic assumption of having unbounded traffic burst sizes, we attempted arrive at results for the case of policed sources. The closed-form bandwidth requirement estimators presented in the paper were derived from the results of the large buffer asymptotics. The estimator in each case is the equivalent capacity (the bandwidth requirement) of the traffic and it can be calculated from the scaled cumulant generating function of the arrival process. A general probabilistic approachwas applied to some simple traffic models in order to arrive atan analytical expression for this scaled cumulant generating function. The idea turned out to work well for the simplest cases, as it led to some difficult problems yet to be solved for more complex, but still simple models. However, the general approach delivered a promising lower and upper bound for the scaled cumulant generating function. Future work will focus on expressing these bounds in closed form. Steps towards more complex models will be taken in the future in order to obtain the target model of the OFF period lengths having a general distribution and the ON burst lengths having a truncated (bounded) general distribution through the model of exponentially distributed OFF and exponentially distributed bounded (truncated) ON period lengths.

References

1. John T. Lewis, Raymond Russell: An Introduction to Large Deviations for Teletraffic Engineers, Technical Report, Dublin Institute for Advanced Studies - Applied Probability Group, 1996
2. Simon Crosby, Ian McGurk, John T. Lewis, Raymond Russell, Fergal Toomey: Statistical Properties of a Near-Optimal Measurment-Based CAC Algorithm, Proc. IEEE ATM'97, Lisbon, Portugal
3. N. G. Duffield: Economies of scale in queues with sources having power-law large deviation scalings, J. Appl. Prob., no. 33, pp. 840-857, 1996
4. Peter W. Glynn, Ward Whitt: Logarithmic asymptotics for seady-state tail probabilities in a single-server queue, Studies in Applied Probability, 1993
5. Dembo-Zeitouni: Large Deviations Techniques and Applications, Springer, 1998
6. Frank Kelly: Notes on Effective Bandwidths, Stochastic Networks: Theory and Applications, vol. 4, pp. 141-168, Oxford University Press, 1996
7. Tamás Jakabfy: A nagy eltérések elméletén alapuló hívásengedélyezési algoritmusok viselkedése néhány forgalomtípus esetére, Students' Scientific Conference, Eötvös Loránd University, 2002, in Hungarian
8. George Kesidis, Jean Walrand, and Cheng-Shang Chang, Effective bandwidths for multiclass Markov fluids and other ATM sources, IEEE/ACM Trans. Networking vol. 1, no. 4 pp. 424-428, 1993
9. Prékopa András: Valószínűségelmélet műszaki alkalmazásokkal, Műszaki Könyvkiadó, Budapest, 1962

News

Seclutions, a Swiss information security software company, introduced a powerful and easy-to-deploy „application security gateway" that protects Internet-connected companies from damages that can occur from both application-level attacks and service outages. Companies are deploying an increasing number of business applications for the Web that were designed to handle business transactions, rather than security tasks.

The security gateway, called AirLock 3.0 enables companies to provision secure, performance-optimized, and always-on Web services over any type of network. The core security functionality of its predecessor AirLock 2.0 by adding application load balancing and failover support, critical to ensuring applications are fully responsive even during difficult situations such as denial of service attacks, changing load requirements, and short application maintenance intervals.

AirLock offloads user authentication and authorization enforcement from business applications and provisions user identities and their entitlements to any application on the backend. Doing so optimizes application performance, creates a single point of integration to authentication mechanisms, and reduces complexity of applications all of which translate into cost savings.

Invariant Associative Image Memory Based on Stir and MPEG4

JÁN TURÁN[1] JR., KÁLMÁN FAZEKAS[2], JÁN TURÁN[1], ĽUBOŠ OVSENÍK[1]

[1]Technical University of Košice, Slovakia; [2]Budapest University of Technology and Economics, Hungary

Reviewed

Invariant Associative Image Memory is an essential part of various multimedia, security and image processing systems. To create an Invariant Associative Image Memory two key problems have to be solved: selection of effective invariant features and effective coding of the original, reference images. The paper gives results of the development work related to design an Invariant Associative Image Memory based on the use of STIR (Scale, Translation and In-plane Rotation) invariant transform and MPEG4-like codec. The memory was implemented as a software package on a powerful PC and tested in various experiments with recognition a large set of human face photographs, which were arbitrary shifted, rotated, scaled, partly covered and corrupted by noise.

1. Introduction

Invariant Associative Image Memory is an essential part of various multimedia, security, and image processing systems [1,2]. Similar memories can be applied also in various image indexing systems [3].

To create an Invariant Associative Image Memory two key problems have to be solved: selection of effective invariant features and effective coding of the reference images. Effective coding of the original, images is important for the minimalising of the image memory capacity if there is a large number of image classes (which is a case at the most of practical applications). Selection of effective invariant features: or embedding invariant pattern recognition to the associative memory design is important for the obtaining desired invariance, robustness, immunity to presence of distortions, occlusions and noise in relation to the input images. One solution of this problem is the use of a class of transforms which are simultaneously invariant under shift, scaling and rotation, so called STIR (Scale, Translation and In-plane Rotation) invariant transforms [4]. The algorithms of these transforms are based on a general transformation where the kernel itself contains the function to be transformed. Thus the invariance is achieved by a kind of self mapping. Pre-processing of the centroid or co-ordinate transformation is not necessary, which simplified of possible electronics and optoelectronics hardware implementation. The second key problem of the invariant associative memory design was solved with application of a MPEG4-like codec [6,7,8] based on an intensively MESH-based scheme, already applied for low bit-rate multimedia communications [5].

On the base of STIR invariant transform and MPEG4-like codec a new Invariant Associative Image Memory was developed [5,6]. The paper gives the results of the development work related to design of this memory. The memory was implemented as a software package on a powerful PC and tested in various experiments with recognition a large set of human face photographs, which were arbitrary shifted, scaled, partly covered and corrupted by noise. The obtained experimental results are very satisfied and the memory can be easy implemented in hardware.

2. Scale, translation and in-plane rotation (STIR) invariant transform

The STIR invariant transform is based on a general transformation where the kernel itself contains the function to be transformed [4]. In this approach a function $f(I(x,y))$ of intensity $I(x,y)$ is nonlinearly mapped into a function $F(\omega)$ of a one- dimensional frequency space

$$F(\omega)=\iint_{\Omega} f_1(\mathrm{x})h[\omega;f_2(\mathrm{x})]d\mathrm{x} \tag{1}$$

where $\mathbf{x}$ is a 2D vector and $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ describe two functions of the 2D pattern $f(\mathbf{x})$ where $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are defined over the $\Omega=\{\mathrm{x}|f_1(\mathbf{x})>0 \cup f_2(\mathbf{x})>0\}$ range and $h[\omega;f_2(\mathbf{x})]$ is a transform kernel.

If the input signal is shifted and rotated simultaneously, the spectrum is obtained as follows:

$$F_T(\omega)=\iint_{\Omega'} f_1(\mathrm{x}')h[\omega;f_2(\mathrm{x}')]d\mathrm{x}' \tag{2}$$

where $\mathbf{x}'=\beta(\mathbf{x}_0+\mathbf{T}.\mathbf{x})$. Here $\mathbf{x}_0$ denotes a translation of the signal, $\mathbf{T}$ is a rotation matrix, and its determinant $\det(\mathbf{T})=1$, and β is a positive number which describes possible scaling. The function $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ after coordinate transformation are defined over a new range of $\Omega'=\{x'|f_1(\mathbf{x}')>0 \cup f_2(\mathbf{x}')>0\}$.

The transform can be represented in the old coordinates using relations

$$F_T(\omega)=\iint_{\Omega_0} f_1(\mathrm{x}')h[\omega;f_2(\mathrm{x}')]d\mathrm{x}$$

$$F_T(\omega)=\iint_{\Omega_0} f_1(\mathrm{x}')h[\omega;f_2(\mathrm{x}')]\frac{d\mathrm{x}'}{\beta.\det(T)}$$

$$F_T(\omega)=\frac{1}{\beta}\iint_{\Omega_0} f_1(\mathrm{x})h[\omega;f_2(\mathrm{x})]d\mathrm{x} \tag{3}$$

If $\Omega_0 \supset (\Omega \cup \Omega')$ is chosen for all possible shifts $\mathbf{x}_0$, scaling factors β, and rotation $\mathbf{T}$, than

$$F_T(\omega)=\frac{1}{\beta}\iint_{\Omega_0} f_1(\mathrm{x})h[\omega;f_2(\mathrm{x})]d\mathrm{x}=\beta.F(\omega) \tag{4}$$

If both spectra are normalised, for example, at $\omega=0$, the spectra of the original pattern and of its corresponding shifted and rotated pattern with different size are identical. This means that the transform displays shift, rotation, and scale invariance at the same time.

3. Architecture of the invariant associative image memory

On the base of STIR invariant transform and MPEG4-like codec a new architecture of Image Associative Memory was developed (Fig.1). Digital input images after pre-processing (segmentation, filtering, etc.) [1,9] enter to the STIR Transform Processor where invariant transform coefficients (features) are computed. In the Learning Mode of the memory this features together with the MPEG4 coded Original Images (Reference Images) are feeded to the Feature and Image Memory. In the Recognition Mode the features of an unknown (corrupted) picture enter to the Classificator where the Image Class is determined. According to the determined Image Class, MPEG4 coded Reference Image is feeded to the MPEG4 Decoder and displayed at the output of the Invariant Associative Image Memory in the original input quality memorised during the Learning Mode. The memory was implemented as a software package (STIR-AIM) on a powerful PC and tested with recognition of a large set of human photographs [6,7,8]. In our experiments we used following parameters of the memory:

$g(\mathbf{x})\in[0,1,2,...,255]$, $g(\mathbf{x})=g(x,y)$ represents an image intensity,

$\mathbf{R}=(\omega_1,\omega_2)$

$h[\mathbf{R};\mathbf{f}_2(\mathbf{x})]=\exp[-i\mathbf{R}.\mathbf{f}_2(g(\mathbf{x}))]$

$I(\mathbf{x})=g(\mathbf{x})/n\in<0,1>$

$f_I(I(\mathrm{x}))=1$ for each $I(\mathrm{x})\neq 0$ and (5a)

$f_I(I(\mathrm{x}))=0$ for each $I(\mathrm{x})=0$ (5b)

$f_2^X(I(\mathrm{x}))=16I(\mathrm{x})\sin(14I(\mathrm{x}))$ – function in x direction (5c)

$f_2^Y=L_{M7}(I(\mathrm{x}))$ (5d)

$L_{M7}(I(\mathrm{x}))=429\, I(\mathrm{x})^7-693\, I(\mathrm{x})^5+315\, I(\mathrm{x})^3-35\, I(\mathrm{x})$

– function in y direction,(6)

where $\nu=255$ is the normalisation factor and where $L_{M7}(I(\mathbf{r}))$ is the modification of seventh-order Legendre polynomial. We choose to work with a real field by considering only the amplitude of the STIR transform:

$$F_T(\omega_1,\omega_2)=\left|\sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f_1(I(x,y))\exp\{-i\lambda\}\right| \tag{7}$$

where

$$\lambda=[16\,\omega_1 I(x,y)\sin(14I(\mathrm{x},\mathrm{y}))+\omega_2 L_{M7}(I(x,\mathrm{y}))] \tag{8}$$

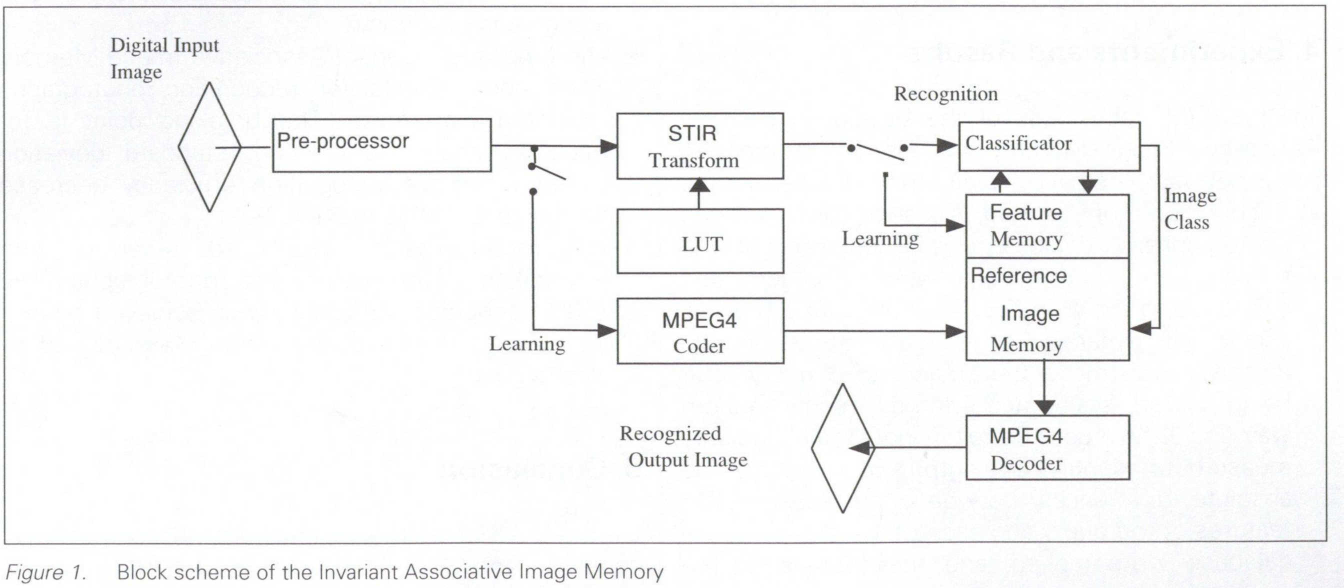


Figure 1. Block scheme of the Invariant Associative Image Memory

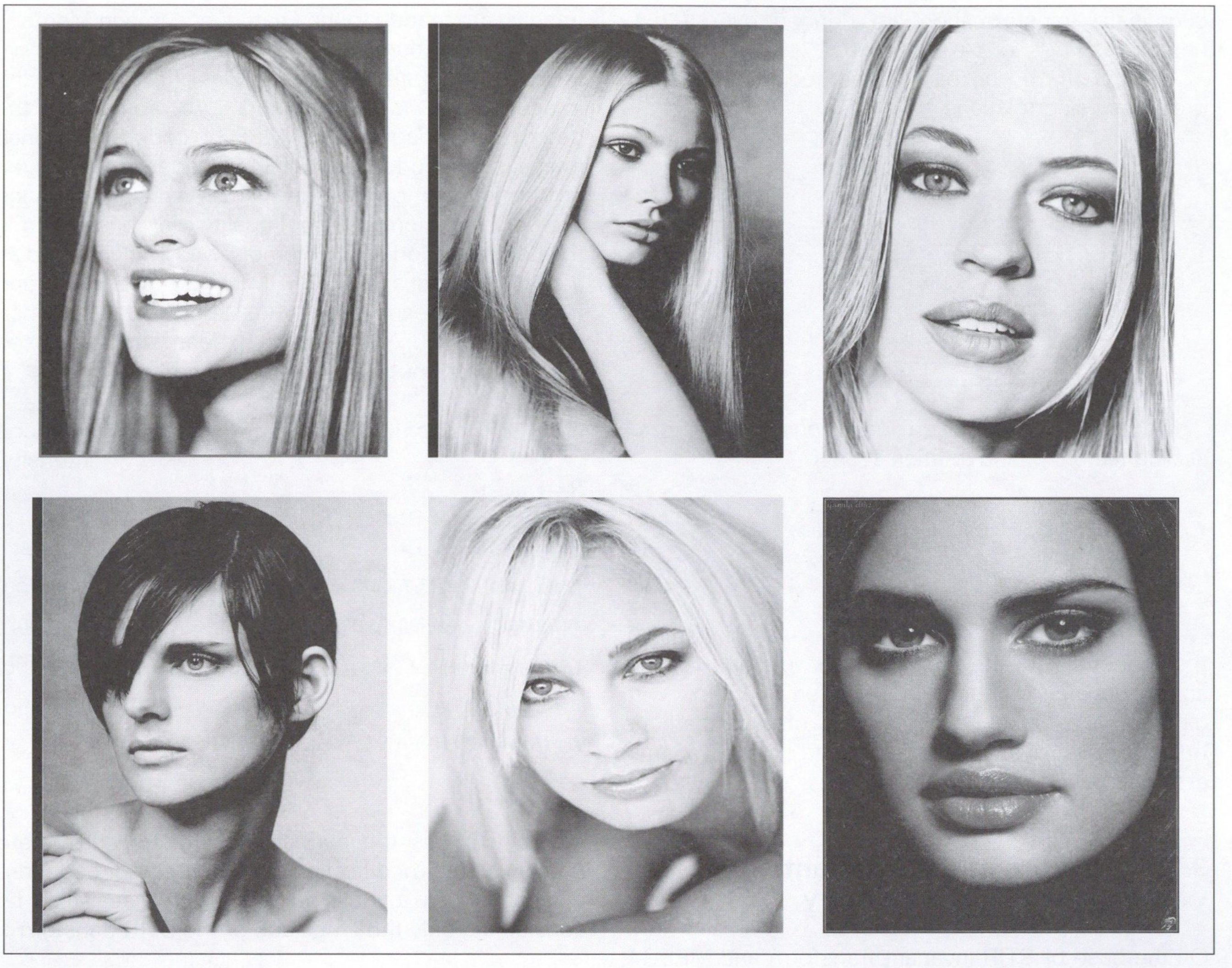


Figure 2. Part of testing sets of human face photographs

In STIR-AIM programme package a fast algorithm using Look-up tables (LUT) was used. Therefore, the computation speed of the memory in the Recognition Mode is for typical PC application (clock 1.2 GHz) at the level of 0.37s per human face photograph.

4. Experiments and Results

To check the robustness of the developed Invariant Associative Image Memory and identify its breaking points several systematic experiments were performed:

1. Recognition of shifted, scaled and rotated photographs: A database (Fig.2) consisting of 150 human face photographs was choosed and converted to digital images each of 256x256 pixels and 8 bit dynamic range. Each image in the database was made the subject of a query after being shifted, scaled and rotated in some random way (Fig.3). We use the following simple similarity measure to identify the output of memory. The absolute differences between the scaled STIR features of the query and each of the entries of the database were ranked, and the average of the smallest two thirds of then was calculated as the similarity measure. Total number of tested photographs was up to 1000. The proposed memory is capable of identifying a large set of shifted, rotated and highly scaled (up to 0.4) input human face photographs with practically 100% recognition efficiency.
2. The proposed Invariant Associative Image Memory gives good results for recognition photographs corrupted with noise. The breaking point is for Gaussian noise (zero mean, standard deviation 50-60) when the recognition efficiency decrease from original 100% to 96-91%.
3. Very good results was also received with recognition of partly covered photographs. The 100% recognition efficiency was achieved even if the 12% of the photograph area was covered by black signal.

5. Conclusion

On the base of STIR invariant transform and MPEG4 codec a new architecture of Invariant Image

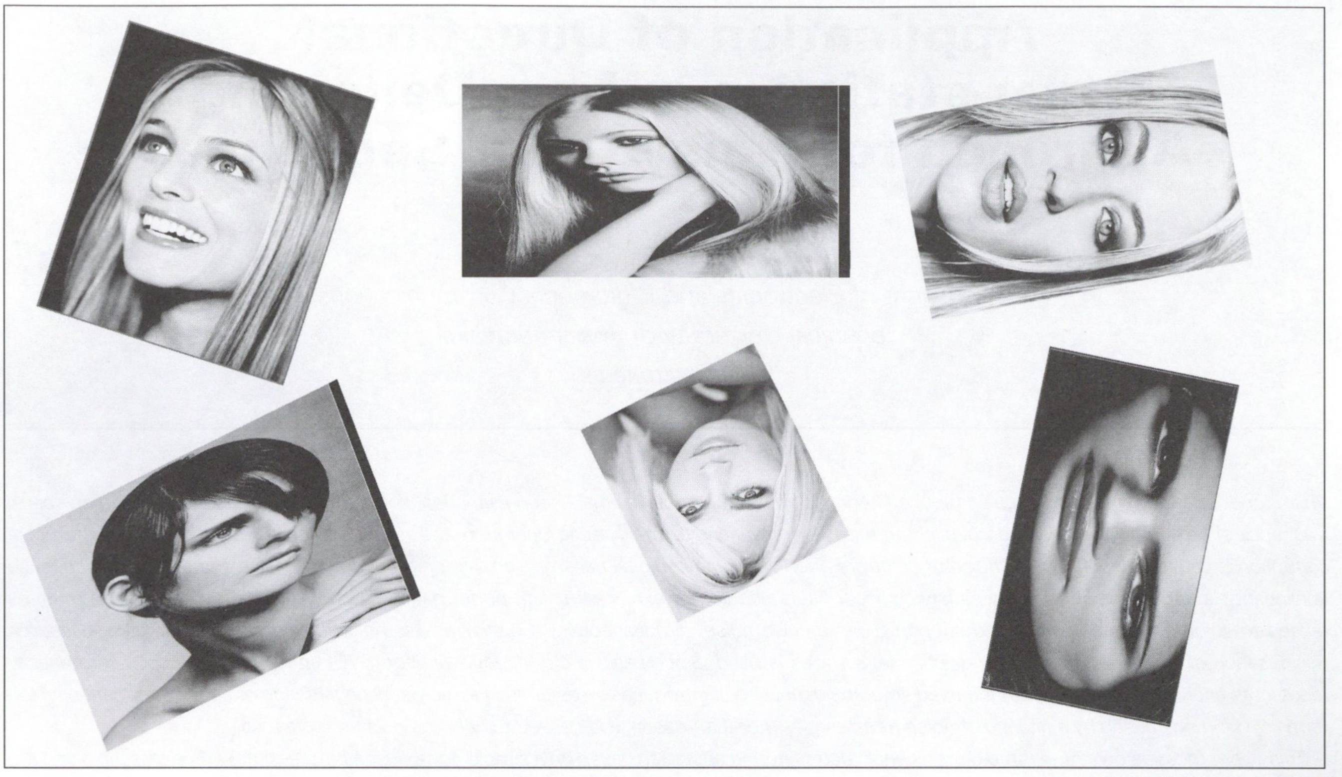


Figure 3. Part of testing sets of shifted, rotated and scaled photographs

Associative Memory was developed. The memory was implemented as a software package (STIR-AIM) on a powerful PC and tested with recognition of human face photographs. Results of our experiments show, that the proposed memory is capable of identifying a large set of arbitrary shifted, rotated and highly scaled input photographs (and partly covered photographs), even if they are corrupted by noise.

6. Acknowledgements

The authors are thanking for the financial supports from the COST 276, VEGA grant No. 1/7444/20 and NATO grant No. 1035/NATO/01.

References

1. Davies,E.R.: Machine Vision. Artech House, London, 1995.
2. Turán,J.-Kövesi,L.-Kövesi,M.: CAD System for Pattern Recognition and DSP with Use of Fast Translation Invariant Transforms. J. on Communications, Vol.XLV, 1994, 85-89.
3. Turán,J.: Fast Translation Invariant Transforms and Their Applications. Elfa press, Košice, Slovakia, 1999.
4. Fang,M.-Hausler,G.-: Class of Transforms Invariant Under Shift, Rotation, and Scaling. Applied Optics, Vol.29, No.5, February 1990, 704-708.
5. Tran,S.M.-Fazekas,K.-Gshwindt,A.-Benois-Pineau,J.-Turán,J.: MPEG4-Like Codec Scheme for Advanced Multimedia Communication. In: Proc. COST276 Workshop: Inf. and Knowledge Management for Integrated Media Communication, Leganés, Spain, November 21-23, 2001, 241-247.
6. Turán,J.-Kövesi,L.-Kövesi,M.: Invariant Associative Memory Design. ICOMT'97, Budapest, Hungary, Oct. 23-29,, 1997, 31-35.
7. Turán,J.-Kövesi,L.-Kövesi,M.: STIR Transform Based Invariant Image Memory Design. Proc. of SBT/IEEE Int. Telecommunication Symposium (ITS-98), Sao Paulo, Brazil, August 9-13, 1998, 512-520.
8. Turán,J.-Kövesi,L.-Kövesi,M.: Experiments with STIR Based Associative Image Memory. IWSSIP'99, Bratislava, Slovakia, June 2-4, 1999, 135-139.
9. Umbaugh,S.E.: Computer Vision and Image Processing. Prentice Hall, New York, 1998.

Application of Directional Order-statistics in New Denoising Approach for Noisy Color Sequences

RASTISLAV LUKÁČ

Department of Electronics and Multimedia Communications,
Technical University of Košice, Slovakia
lukacr@tuke.sk

An important problem in color image sequence filtering is how to near the optimal filtering situation, so that the desired features including spatial frequencies, temporal changes and color chromaticity are invariant to the filtering operation and noise would be affected, only. If the noise corruption is caused by the impulse noise or bit errors often occurring in digital transmission, on the ground of the noise and the image sequence nonlinearity, a superposition property satisfying linear filters should not exhibit sufficient performance. In order to preserve all desired features related to color image sequences and suppress the impulse noise, simultaneously, a nonlinear class of vector order-statistic filters based on minimisation of distance function of multichannel samples is used. Since basic vector directional filter (BVDF) preserving the color chromaticity represents filtering approaches with the fixed smoothing function, it introduces an estimation error, visual degradation of useful information, in the form of a blurring caused by a filtering of noise-free samples. For that reason, it is desired to filter the noisy samples, only.

This paper focuses on a new three-dimensional vector filtering approach based on a simple comparison between the angle threshold and the operation value depends on neighbourhoods. As the proposed operation value serves an angle between the central sample and the mean of several directional order-statistics with the smallest angle distances to input samples. Note is that the computation of angle distances between input samples is necessary in BVDF filter structure, too. If the filter operation value is greater than or equal to the angle threshold, the central sample of a filter window is corrupted and it will be estimated by BVDF. If the sample is noise-free, the proposed method will perform an identity operation, i.e. the central sample is passed to a filter output without change. In comparison with BVDF, the proposed method is characterised by significant improvement, especially in the term of mean absolute error criteria that is a mirror of the signal-details preservation.

Keywords: color image sequences, impulse noise, directional order-statistics, adaptive filter

1. Introduction

Modern communication and multimedia system such as videoconferencing, videophoning, digital television, Internet and etc. incorporate the recent advances in the field of hardware, software, telecommunications, digital signal and image processing, graphics and computer vision to an integrated system and extend the possibilities of conventional communications. In order to provide the high quality of multimedia signals [7], [9], [10], [12] represented by a set of visual information, audio information and data, there is necessary to ensure the correct processing of useful information in all multimedia components and determine the accurate relationships among them.

In the case of visual information often expressed in the form of motion color video [16], [25] the visual quality of processed image signals often depends on the performance of preprocessing methods such as linear and nonlinear adaptive filters for signal denoising. If the noise corruption has an impulse character, e.g. the impulse noise, bit errors, especially introduced during the signal scanning and the transmission over the information channel (defiance the robust prediction algorithms for transmission error), the best performance will be provided by nonlinear filters based on order-statistics.

According to a high dimensionality of color image sequences [16], [25] that are three-dimensional (3-D) image signals or time sequences of two-dimensional color (mutichannel) images, besides the spatial frequencies in the frames, it is necessary to consider the temporal correlation of an image sequence and the correlation between the color channels, too. For that reason, the filtering of color image sequences still represents a very important and interesting problem. In addition, the mathematical attraction of the color image sequences filtering lies in the vector processing of 3-D image data that takes advantages against scalar methods, since it utilises the inherent correlation that exists between color channels. Thus, the vector approaches produce an output image friendly for human visual system that is sensitive to color distortion, i.e. color artefacts or atypical image points. According to the fact that each multichannel sample represents a vector determined by its magnitude and direction in the vector space, in general, vector filters can be separated into two classes, namely vector-median based filters and vector directional filters.

In the term of signal dimensionality and considered correlation present in the input set, the filtering for image sequences can be divided into three classes [2],[11],[14] such as temporal filters, spatial filters and spatiotemporal filters. The class of temporal filters (Figure 1a) is referred to temporal correlation of frames. One-dimensional filters remove noise without impairing the spatial resolution in stationary areas. In order to improve a filter performance, the temporal filtering is usually connected with motion compensation [11] so as to filter objects along their motion trajectory. However, this way is very computationally complex, and because of spatial warping and scene changes, the motion compensation [2] often does not work well. In general, spatial (2-D) filters represent the most frequently used filtering techniques in the image processing. Since these filters process the each frame independently (Figure 1b), spatial filters can be designed to give good filtering results in frames, however, respecting no motion trajectory results in a motion blurring. Three-dimensional filters (Figure 1c) utilise both temporal correlation and spatial correlation present in image sequences. For that reason, the spatiotemporal filters represent a natural filter class for noisy image sequences, where it is possible to observe the significant reduction of the spatial and motion blurring, in the dependence on filtering methods.

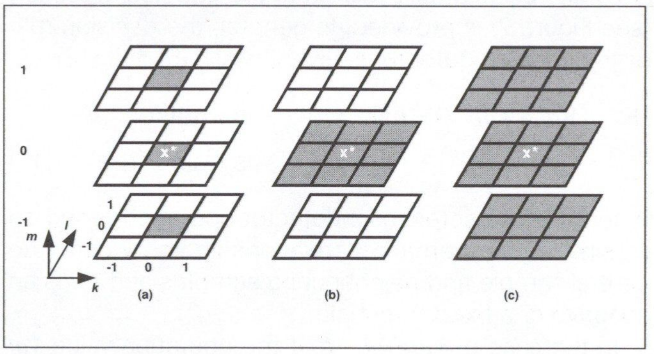


Figure 1. Filtering methods for the image sequences divided according to a dimensionality of the input set.
(a) Temporal (1-D) filters
(b) Spatial (2-D) filters
(c) Spatiotemporal (3-D) filters

This paper focuses on the non-motion-compensated directional processing of color image sequences corrupted by the impulse noise, i.e. a distance function of vector filters is based on the directions between the vectors spanned by a cube filter window (three 3×3 windows applied to past, actual and future frame on each spatial position). The novelty of the proposed method lies in the utilising of a number of the smallest directional-ordered input vector samples for the determination of the operation value used in the switching between the basic vector directional filter (BVDF) and the identity filter. An adaptive alternation between identity filter and BVDF is controlled by comparison of the operation value (given by the angle between the central sample of the input set and the mean of the several smallest directional order-statistics) and the threshold angle. If the operation (detector) value is greater than or equal to the threshold angle, it is the case of a noisy sample, since the central sample is very different from the 'smallest' vector directional order-statistics that are noise-free samples with a high probability. If the operation value is smaller than the threshold angle, the proposed method is equivalent to identity filter.

This paper is organised as follows. In the next section, the mathematical models of the impulse noise and bit errors are provided. Mathematical preliminaries related to directional processing are presented in Section 3. Section 4 focuses on a design of a new method. Objective criteria including well-known mean absolute error, mean square error and color difference criterion for the color chromaticity preservation and cross-correlation for the motion preservation and the experimental results including a number of figures and tables are presented in Section 5 and Section 6, separately. Finally, the properties of the proposed method are concluded in conclusion, where the possibilities of the future research on the field of color image filtering are suggested.

2. Impulse noise

In general, mathematical model of the impulse noise for the color images can be expressed as [17]

$$\mathbf{x}_{i,j,t} = \begin{cases} \upsilon & \text{with probability } p \\ \mathbf{o}_{i,j,t} & \text{with probability } 1\text{-}p \end{cases} \qquad (1)$$

where i, j characterise sample position, t is frame index, $\mathbf{o}_{i,j,t}$ is the sample from the original image, $\mathbf{x}_{i,j,t}$ represents the sample from the noisy image, p is a corruption probability and $\upsilon=(\upsilon_R,\upsilon_G,\upsilon_B)$ is a noise vector of intensity random values. Since, single components of υ are generated independently (Figures 3e,f), the gray impulse ($\upsilon_R=\upsilon_G=\upsilon_B$) can occur in the special case, only.

The second model of the impulse noise corruption can be expressed through the model of bit errors [3]

$$^{*}\mathrm{k}^{m}_{i,j,t} = \begin{cases} \mathrm{k}^{m}_{i,j,t} & 1-p \\ 1-\mathrm{k}^{m}_{i,j,t} & p \end{cases} \qquad (2)$$

where i, j and t characterise spatial position and time position, m is index of bit level forced to be between 1 and B, B is a number of bits per sample, p is a bit error probability and finally $\{\mathbf{k}\}$ and $\{^{*}\mathbf{k}\}$ characterise original and corrupted bit level. Note that original sample is expressed as

$$\mathrm{o}_{i,j} = \mathrm{k}^{1}_{i,j}2^{B-1} + \mathrm{k}^{2}_{i,j}2^{B-2} + \ldots + \mathrm{k}^{B-1}_{i,j}2 + \mathrm{k}^{B}_{i,j} \qquad (3)$$

whereas a sample from noisy image is defined by

$$x_{i,j} = {}^*k^1_{i,j}\, 2^{B-1} + {}^*k^2_{i,j}\, 2^{B-2} + \ldots + {}^*k^{B-1}_{i,j}\, 2 + {}^*k^B_{i,j} \qquad (4)$$

To suppress the impulse noise, usually the robust order-statistics theory [3], [15], [18], [22] is used. In the case of vector-valued signals, observed samples are ordered according to the distance function, where both magnitude [4], [17] and direction [8], [19], [23], [24] of multichannel samples can be considered. If the vectors' direction in the vector space between the input samples serves as ordering criteria then it is the case of directional processing. The second ordering criteria is based on the distance in the vector space between the input samples, the typical representatives are vector median-based filters [4], [13], [20]. In general, vectors' magnitude takes a measure of their brightness, whereas the direction of vector samples wreaks their chromaticity [19], [23].

3. Vector angle-based filters

Let $y(x){:}Z^l \rightarrow Z^m$ represent a multichannel image, where l is an image dimension and m represents a number of channels. If $m \geq 2$, then it is the case of m-channel image processing. In the case of standard color images $l = 2$ and $m = 2$. Let $W=\{\mathbf{x}_i \in Z^l;\ i=1,2,...,N\}$ represent a filter window of a finite size N, where $\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N$, is a set of noised samples. Note, that the position of the filter window is determined by the central sample $\mathbf{x}_{(N+1)/2}$. Each input vector $\mathbf{x}_i$ is associated with the angle distance α_i that is defined by [19], [20]

$$\alpha_i = \sum_{j=1}^{N} A(x_i, x_j) \quad \text{for } i=1,2,...N \qquad (5)$$

where

$$A(x_i, x_j) = \cos^{-1}\left(\frac{x_i, x_j}{|x_i| \cdot |x_j|}\right) \qquad (6)$$

represents the angle between two m-dimensional vectors $\mathbf{x}_i = (\mathbf{x}_{i1},\mathbf{x}_{i2},...,\mathbf{x}_{im})$ and $\mathbf{x}_j = (\mathbf{x}_{j1},\mathbf{x}_{j2},...,\mathbf{x}_{jm})$. If angle distances (5) serve as an ordering criterion, i.e.

$$\alpha_{(1)} \leq \alpha_{(2)} \leq \ldots \leq \alpha_{(r)} \leq \ldots \leq \alpha_{(N)} \qquad (7)$$

then it means that the same ordering is implied to input set which results in ordered input sequence

$$x^{(1)} \leq x^{(2)} \leq \ldots \quad x^{(r)} \leq \ldots \leq x^{(N)} \qquad (8)$$

If a filter output is given by the sample from the input set that minimises the sum of angles with other vectors, then filter performs BVDF filtering operation, i.e. [23],[24]

$$y_{BCDF} = x^{(1)} \qquad (9)$$

where sample $\mathbf{x}^{(1)}$ is associated with minimal angle distance $\alpha_{(1)}$.

In the case, where a filter output can be expressed as the set of the first r terms of (8) with simultaneous valid (7), the filter is called general vector directional filter (GVDF). Mathematically, GVDF output is defined by [23], [24]

$$y_{GDVF} = \{x^1, x^2, ..., x^N\} \qquad (10)$$

GVDF output the set of r vectors whose angle α_i (for $i=1,2,...,N$) from all other vectors is small. Usually, the output set of GVDF is used in the second level as an input for additional filter [20], e.g. α-trimmed average filter, multistage median filter and some morphological filters, where samples $\mathbf{x}^{(1)},\mathbf{x}^{(2)},...,\mathbf{x}^{(r)}$ will be processed according to their magnitude, since these vectors have approximately equal direction in a vector space. Simply, GVDF produces a set of vectors with similar directions in color space, and thus samples with atypical directions are eliminated. It follows that GVDF differentiate the processing of color vector on directional processing and magnitude processing.

4. The proposed method

Now, the basic theory related to the impulse detection (see Figure 2) is provided. In general, the decision rule of the impulse detector is given by [1], [5], [6], [13]

$$\text{IF} \quad Val \geq Tol \quad \text{THEN} \quad \mathbf{x}_{(N+1)/2} \text{ is imulse}$$
$$\mathbf{x}_{(N+1)/2} \text{ is noise-free} \qquad (11)$$

where Val characterises a detector operation based on a simple mathematical relationship between the central sample and neighbouring samples and Tol is an adaptive or a fixed threshold.

In the case of $Val \geq Tol$, i.e. if the operation value Val is greater than or equal to the threshold value Tol, then the central input sample $\mathbf{x}_{(N+1)/2}$ is corrupted and processed by a smoothing filter, consecutively. If $Val < Tol$, the central sample $\mathbf{x}_{(N+1)/2}$ is noise free. Then, it is retained without change and thus, a blurring introduced by a filter influence is reduced.

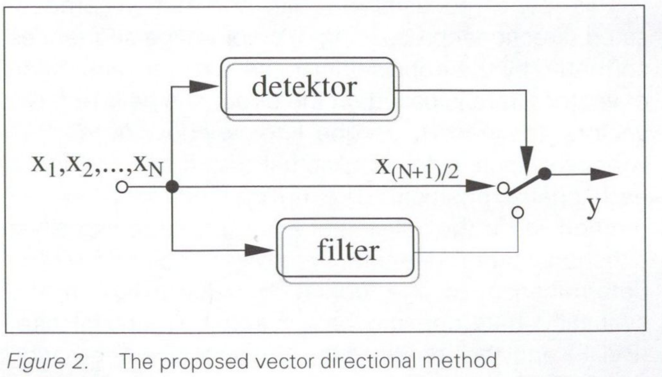


Figure 2. The proposed vector directional method

Now, the proposed method is presented. In order to obtain information about an impulse occurrence, the following steps should be performed. According to (11), each input vector sample $\mathbf{x}_i$, $i=1,2,...,N$, is associated with the angle distance α_i. In the next step, there is necessary to sort the set $\alpha_1,\alpha_2,...,\alpha_N$, resulting in (6). Then, the same ordering is applied to input set $\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N$.

Likewise the definition of GVDF (10), let $\mathbf{x}^{(1)},\mathbf{x}^{(2)},...,\mathbf{x}^{(r)}$ be a set of the r smallest vector directional order statistics (8), i.e. the input samples with similar directions in color space. Note that r is forced by $1\geq r\geq N$. Then, the decision rule of the vector impulse detector can be expressed by

$$\text{IF} \quad A\left(\bar{\mathrm{x}}(r),\mathrm{x}_{(N+1)/2}\right)\geq Tol \quad \text{THEN} \quad \mathrm{x}_{(N+1)/2} \text{ is corrupted}$$
$$\mathrm{x}_{(N+1)/2} \text{ is noise-free} \quad (12)$$

where Tol is threshold angle and $A(\bar{\mathbf{x}}(r),\mathbf{x}_{(N+1)/2})$ characterises the angle between the central sample and the mean $\mathbf{x}_{(N+1)/2}$ given by

$$\bar{\mathrm{x}}(r)=\frac{1}{r}\sum_{i=1}^{r}\mathrm{x}^{i} \quad (13)$$

In the case of the noise detection, i.e. if the angle between $\bar{\mathbf{x}}(r)$ and $\mathbf{x}_{(N+1)/2}$ is greater than or equal to the threshold angle Tol, the noisy central sample is replaced with the sample $\mathbf{x}^{(1)}$, i.e. with the BVDF output. On the other hand, if the mean $\bar{\mathbf{x}}(r)$ and the central sample $\mathbf{x}_{(N+1)/2}$ have similar direction in the vector space, then the central sample will be probably noise-free and thus, it will be passed to the resulting image without change and no additional processing will be performed.

The proposed method includes both identity operation with no smoothing and BVDF with maximum amount of smoothing (according to directional processing) and it reduces blurring that should be introduced by BVDF. Switching control between the identity filter and the BVDF is performed in the dependence on the threshold angle Tol and parameter r corresponding to a number of the considered smallest vector directional order-statistics.

5. Objective criteria

As a measure of the noise corruption and the filter performance, too, four objective criteria, namely mean absolute error (MAE), mean square error (MSE), cross correlation (ΔR) and color difference (CD), are used. In general, MAE is a mirror of the signal-details preservation, MSE evaluates the noise suppression well, ΔR expresses the preservation of the motion trajectory in the image sequence and CD is a measure of the color chromaticity preservation. Thus, the quality of the processed image sequences is quantified with a high accuracy related to the signal dimensionality.

Mathematically, the 2-D definitions of MAE and MSE for monochromatic images are given by [14]

$$MAE=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left|o_{i,j}-x_{i,j}\right| \quad (14)$$

$$MSE=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(o_{i,j}-x_{i,j}\right)^2 \quad (15)$$

where $\{o_{i,j}\}$ is the original image, $\{x_{i,j}\}$ is the filtered (noisy) image, i,j are indices of sample position and characterise an image size. Note is that in the case of color image sequence, MAE and MSE criteria are understood as a mean over color channels and all frames.

The motion trajectory preservation criteria is derived from the cross correlation coefficient [14] given by

$$R=\frac{\left|\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}x_{i,j}^{t}x_{i,j}^{t+1}-E^{t}E^{t+1}\right|}{\sigma^{t}\sigma^{t+1}} \quad (16)$$

where E^t is the mean value and σ^t is the standard deviation of the t th frame and E^{t+1} is the mean value and σ^{t+1} is the standard deviation of the $(t+1)$ th frame. Then, it is necessary to apply (16) on the original sequence, too. The best motion preservation is achieved by the smallest absolute difference of the mean cross correlation coefficients between the original noise-free sequence and the filtered sequence.

Finally, the color distortion or the color chromaticity preservation is evaluated by CD that requires transformation from RGB to Luv color space [21]. For the color image, the CD is expressed as

$$\Delta E_{Luv}=\sqrt{(\Delta L)^2+(\Delta u)^2+(\Delta v)^2} \quad (17)$$

where ΔL, Δu and Δv represent the difference between original and noisy images in L, u and v color channels. The overall value of CD is a mean value over all frames.

6. Experimental results

The test sequence "Grandmom" (Figure 3a) consists of 99 frames with a resolution of samples and a 8 bit per sample representation for each color channel. In Figure 3a and Figure 3b are showed 5th and 95th frames of the test sequence. The second sequence called "Osu-1" (Figures 3c,d) is characterised by high motion. For that reason, it will be very interesting to observe the filter behaviour in these complex conditions. The resolution and number of frames related to sequence "Osu-1" is identical with sequence "Grandmom".

Method	MAE	MSE	CD	ΔR
identity	4.150	529.2	21.061	0.281
BVDF	4.173	55.5	30.517	0.008
the proposed method (r =7, Tol=0.09)	1.163	40.2	12.787	0.142
the proposed method (r=10, Tol=0.09)	1.112	39.7	12.631	0.137
the proposed method (r=15, Tol=0.09)	1.172	40.0	12.792	0.145

Table 1. Performance on sequence Grandmom corrupted by the impulse noise p = 0.05

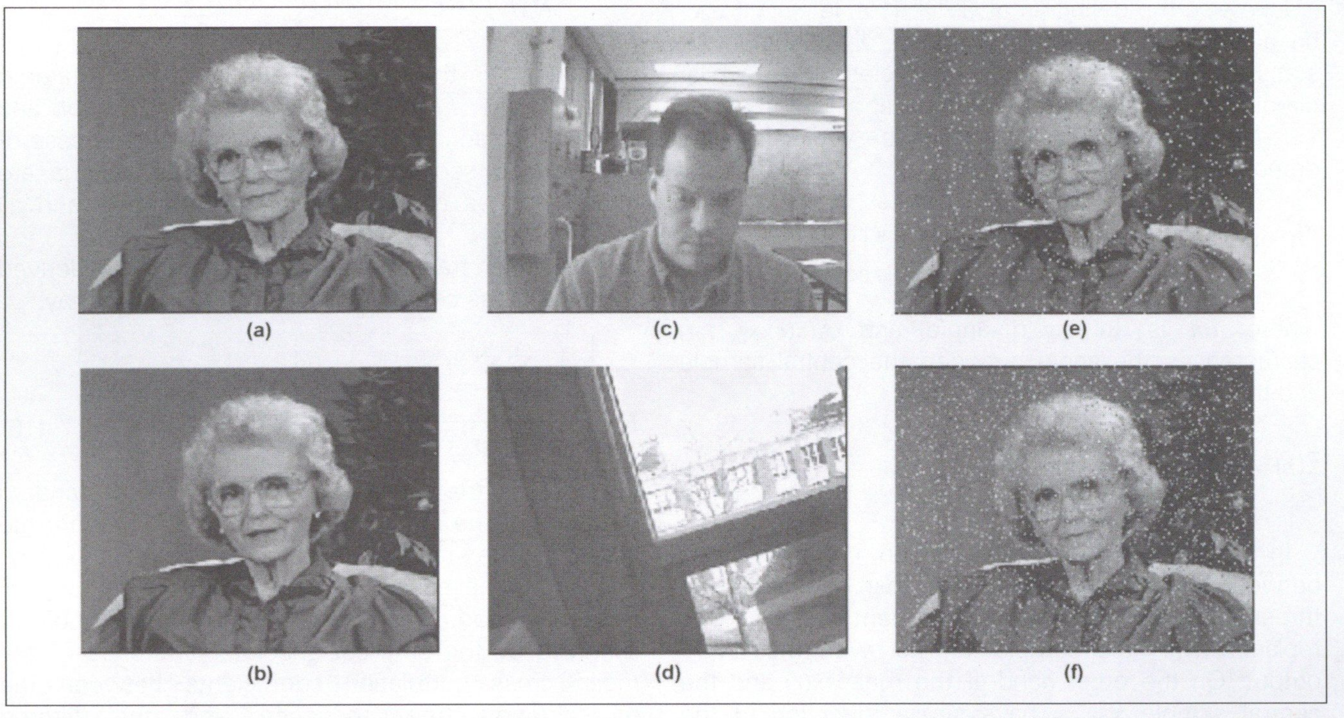


Figure 3. Test color image sequence. a) Original 5th frame of sequence "Grandmom" b) Original 95th frame of sequence "Grandmom" c) Original 5th frame of sequence "Osu-1" d) Original 95th frame of sequence "Osu1" e) 5th frame corrupted by the impulse noise (p=0.05) f) 5th frame corrupted by the impulse noise (p=0.1)

Now, the performance of the proposed method is compared (Figure 4, Table 1 and Table 2) with the performance of BVDF [20],[25]. From these results it can be seen that the proposed method produces the output with spatial signal-details more close to desired image than BVDF. The above-mentioned capability is expressed through MAE criteria that is a mirror of the signal-details preservation, where the achieved improvement is equal to 60%, approximately. However, the results corresponding to MSE criteria are not so good. In the case of error criteria CD, the improvement of the proposed method in comparison with BVDF is equal to 60% again. If the motion preservation capability of used vector filters is compared, the proposed method produces worse results than BVDF. From these results and output video sequences, too, it can be seen that the proposed method preserves the signal-details and the color chromaticity, excellently. This preservation capability is supported by the filtering of probably corrupted samples, only, and thus the estimation error depends on the detection accuracy. Some false detection of noise-free samples tends to preserve impulses in the output image. The mentioned drawback is caused by the fixed determined optimal parameter r and the angle threshold Tol. For that reason, the additional improvement and better MSE and ΔR related to the performance of the proposed method should be achieved by fully adaptive control of the threshold parameter Tol.

Method	MAE	MSE	CD	ΔR
identity	8.133	1043.6	40.896	0.447
BVDF	4.280	62.7	31.039	0.010
the proposed method (r =7, Tol=0.09)	1.376	43.7	13.326	0.016
the proposed method (r=10, Tol=0.09)	1.321	43.2	13.211	0.014
the proposed method (r=15, Tol=0.09)	1.357	44.1	13.455	0.015

Table 2. Performance on sequence Grandmom corrupted by the impulse noise p = 0.1

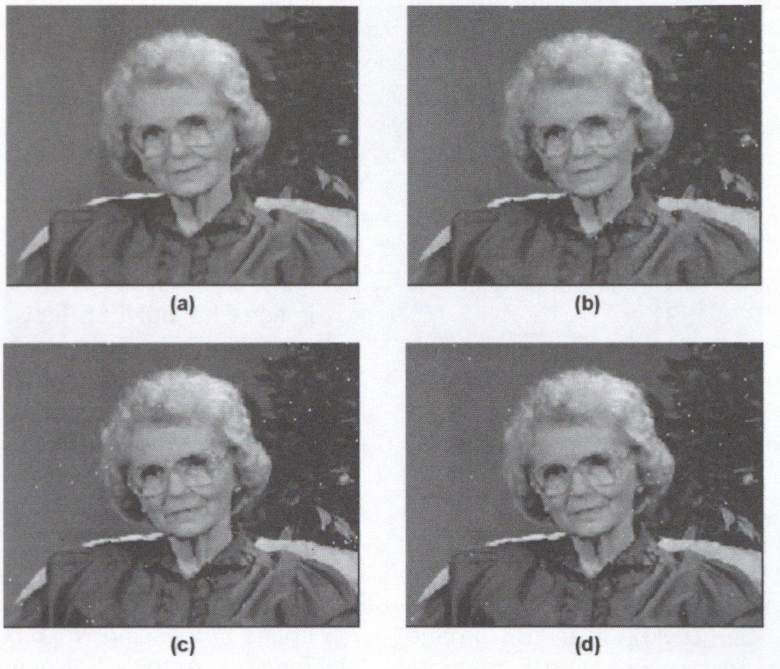


Figure 4. Filtered 5th frame corrupted by impulse noise
a) Output of 3-D vector median. b) Output of 3-D BVDF c) Output of the proposed method (*r*=7, *Tol*=0.09) d) Output of the proposed method (*r*=10, *Tol*=0.05)

Method	MAE	MSE	CD	ΔR
identity	3.808	443.5	15.843	0.161
BVDF	8.722	442.6	14.328	0.050
the proposed method (r =7, Tol=0.09)	1.068	87.3	1.587	0.020
the proposed method (r=10, Tol=0.09)	1.081	88.2	1.607	0.020
the proposed method (r=15, Tol=0.09)	1.078	87.8	1.615	0.020

Table 3. Performance on sequence Osu-1 corrupted by the impulse noise p = 0.05

Method	MAE	MSE	CD	ΔR
identity	7.341	853.3	30.535	0.267
BVDF	8.789	450.4	14.443	0.050
the proposed method (r =7, Tol=0.09)	1.629	126.3	2.362	0.027
the proposed method (r=10, Tol=0.09)	1.641	127.1	2.384	0.027
the proposed method (r=15, Tol=0.09)	1.642	127.1	2.386	0.027

Table 4. Performance on sequence Osu-1 corrupted by the impulse noise p = 0.10

7. Conclusion

The new three-dimensional adaptive vector directional filter based on the adaptive alternation between the basic vector directional filter and the identity filter has been provided as the improvement of well-known basic vector directional filter. The proposed method has been designed especially for the impulse noise suppression in color image sequences.. As the decision rule dividing the sample into noise-free samples and corrupted samples have served the comparison of the threshold angle and the angle between the central sample and the mean of the smallest vector directional order-statistics. The experimental results showed the excellent signal-details and color chromaticity preservation of the proposed method. According to the character of decision rule, the future research tasks are related to the searching for the adaptive control of the threshold angle.

References

1. Abreu, E., Lighstone, M., Mitra, S.K., Arakawa, K.: A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images, IEEE Trans. on Image Processing, Vol.5 (1996) 1012-1025
2. Arce, G.R.: Mulstistage Order Statistic Filters for Image Sequence Processing. IEEE Trans. on Signal Processing, Vol.39, (1991) 1146-1163

3. Astola, J., Kuosmanen, P.: "Fundamentals of Nonlinear Digital Filtering," CRC Press, 1997.
4. Astola, J., Haavisto, P., Neuvo, Y.: Vector median filters. Proc. of the IEEE, Vol.78, (1990) 678-689
5. Beghdadi, A., Khellaf, K.: A Noise-Filtering Method Using a Local Information Measure, IEEE Trans. on Image Processing, Vol.6 (1997) 879-882
6. Bernstein, R.: Adaptive Nonlinear Filters for Simultaneous Removal of Different Kinds of Noise in Images, IEEE Trans. on Circuits and Systems, Vol.Cas-34, (1987) 1275-1291
7. Bojkovic, Z., Milovanovic, D.: Challenges of Information Processing in Multimedia Communications, Proc. of ECMCS'2001, pp.133-140
8. Gabbouj, M., Cheickh, F.A.: Vector median-vector directional hybrid filter for color image restoration. Proc. of EUSIPCO-96, (1996) 879-881
9. Grgic, M., Ghanbari, M., Grgic, S.: Texture-based Image Retrieval in MPEG-7 Multimedia System, Proc. of IEEE Region 8 EUROCON'2001, pp.365-368
10. Grgic, M., Grgic, S., Zovko-Cihlar, B.: Image Retrieval Based on Texture Features Extraction, Proc. of ECMCS'2001, pp. 152-155.
11. Kleihorst, R.P., Lagendijk, R.L., Biemond, J.: Noise Reduction of Image Sequences Using Motion Compensation and Signal Decomposition. IEEE Trans. on Image Processing, Vol.4, (1995) 274-284
12. Kotropoulos, C., Pitas, I., Intelligence in Modern Communication Systems. Proc. of COST 254, Neuchatel (1999) pp. 1-12
13. Lukáč, R.: Vector LUM smoothers as impulse detector for color images. Proc. of ECCTD '01, Vol.3, (2001) 137-140
14. Lukáč, R., Marchevský, S.: LUM Smoother with Smooth Control for Noisy Image Sequences. EURASIP Journal on Applied Signal Processing, Vol.2001, (2001) 110-120
15. Lukáč, R. Marchevský, S.: Boolean Expression of LUM Smoothers. IEEE Signal Processing Letters, Vol.8, (2001) 292-294
16. Lukáč, R. Marchevský, S.: The Use of Threshold LUM Smoothers in Noised Color Sequences. Proc. of IEEE REGION 8 EUROCON'2001, pp.373-376
17. Lukáč, R., Marchevský, S.: Adaptive Vector LUM Smoother. Proc. of IEEE ICIP 2001, Vol.1, (2001) 878-881
18. Pitas, I., Tsakalides, P., Multivariate Ordering in Color Image Filtering, IEEE Trans. on Circuits and Systems for Video Technology, Vol.1 (1991) 247-259
19. Platanoitis, K.N., Androutsos, D., Venetsanopoulos, A.N.: Color image processing using adaptive vector directional filters. IEEE Trans. on Circuits and Systems II, Vol.45, (1998) 1414-1419
20. Plataniotis, K.N., Venetsanopoulos, A.N.: Color Image Processing and Applications, Springer Verlag, 2000
21. Sharma, G.: Digital Color Imaging. IEEE Trans. on Image Processing, Vol.6 (1997) 901-932
22. Tang, K., Astola, J., Neuvo, Y.: Nonlinear Multivariate Image Filtering Techniques, IEEE Trans. on Image Processing, Vol. 4 (1995) 788-798
23. Trahanias, P.E., Venetsanopoulos, A.N.: Vector directional filters - a new class of multichannel image processing filters. IEEE Trans. on Image Processing, Vol.2, (1993) 528-534
24. Trahanias, P.E., Karakos, D., Venetsanopoulos, A. N.: Directional processing of color images: theory and experimental results. IEEE Trans. on Image Processing, Vol.5, (1996) 868-881
25. Viero, T., Oistamo, K., Neuvo, Y.: Three-Dimensional Median Related Filters for Color Image Sequence Filtering. IEEE Trans. on Circuits and Systems for Video Technology, Vol.4, (1994) 129-142

News

To maximize revenue and avoid losing call records is vital for each GSM operator. Due to the growing number of subscribers, services, roaming partners and the frequently changed tariffs and services make the billing process extremely complex.

To keep the billing process under full control and reduce the number of uncharged calls as low as possible in this rapidly changing environment without an effective tool is almost impossible.

By utilizing CeDaR GSM operators can have a powerful application in their hands for checking and analyzing the errors in the billing process and recognize malfunctions in time.

❖ ❖ ❖

Comcast has officially swalowed AT&T Broadband, forming a new cable behemoth with about 22 million subscribers in 41 states, including 6.3 million digital cable customers, 3.3 million high-speed data subscriber and 1.3 million mostly AT&T-based telephony users.

Tool for automatic test selection based on formal specification

ZOLTÁN ANDRISKA, GÁBOR BÁTORI, DUNG LE VIET, ANTAL WU-HEN-CHANG AND DR. GYULA CSOPAKI

Budapest University of Technology and Economics
Department of Telecommunications and Telematics
High Speed Network Laboratory
{andriska,batori,leviet,wuhen}@david.ttt.bme.hu

We present a new method that uses formal specification as a starting point for automatic test generation for telecommunications software, typically protocols. In the recent paper we utilize the specification to generate test suites automatically, by the help of mutation analysis. Mutants of a specification are used as selection criteria to pick out adequate test cases. We developed a software tool based on the presented method to ease the process of test case generation.

1. Introduction

One of the most important criteria that apply to telecommunications software is compatibility with systems from different vendors. This is usually achieved by the means of standardization. These standards or recommendations define the specifications of systems. Manufacturers of the actual products ensure compatibility by applying these specifications. At the end of the development process, the final and most important step is to test the actual products to guarantee that they work as required by the specification. This is called conformance testing, which provides the means to ensure that systems from different companies are compatible, and are able to interoperate correctly according to the standard.

There is a problem in the telecommunication system to make software that operates quick and reliably. The test process is very time consuming and requires the manual effort of many well-trained developers. Therefore, its automation is an important challenge. In this paper we present a possible solution for this problem.

2. Formal description of protocols

Communication protocols are the rules that govern the communication between the different components within a distributed computer system.

Protocol engineering consists of several steps. The most importants are:

- design, specification
- validation
- implementation – making software/hardware product
- checking for errors in the implementation (conformance and co-operation examination etc.)

Protocols must be described for many purposes. Early descriptions provide a reference for co-operation among designers of different parts of a protocol system. The design must be checked for logical correctness. The protocol must be implemented, and if the protocol is in wide use, many different implementations may have to be checked for compliance with a standard. Although informal parts (narrative descriptions and informal walkthroughs) are invaluable elements of this process, experience has shown that by themselves they are not sufficient for descriptions purposes.

The informal techniques traditionally used to design and implement communication protocols have been largely successful, but have also yielded a disturbing number of errors or unexcepted and undesirable behavior in most protocols. The use of a specification written in natural language gives the illusion of being easily understood, but leads to lengthy and informal specifications which often contain ambiguities and are difficult to check for completeness and correctness.

Many different formal description techniques have been proposed for the protocol engineering cycle, including finite state machines (FSM), Petri nets, formal grammars, high-level programming languages, process algebras, abstract data types, and temporal logic. Formal description techniques ensure that the specifications are unambiguous and that they can be processed automatically.

On the standardization for OSI (Open Systems Interconnection), special working groups on "Formal Description Techniques" (FDT) were established within ISO (International Organization for Standardization) and CCITT (Comité Consultatif International Téléphonique et Télégraphique) in the early eighties with the purpose of studying the possibility of using formal specification for the definition of the OSI protocols and services.

Their work led to the proposal of three languages, Estelle, LOTOS (Language Of Temporal Ordering Specification), and SDL.

SDL (Specification and Description Language) is basically a description language, it cannot be used for implementing an application. In SDL, behavior and composition of the functional modules of a system can be represented on arbitrary abstraction level. Typically complex, event-driven, real-time communicating systems can be effectively described in SDL.

The most important property of SDL is that it describes the dynamic behavior of a system as a CEFSM. Communication (between processes or processes-environment) is represented via signals that travel on certain predefined FIFO (First in First out) channels and signal routes. These channels are located between processes or between processes and the environment of the system model.

In SDL state machines determine the system behavior. Formally the behavior of a state machine can be described by state graphs. SDL also provides the possibility to describe a system in a hierarchical way. We can distinguish system, block and process levels.

MSC (Message Sequence Chart) is recommending, by the ITU-T recommended Z.120 [6] to provide a trace language for the specification and description of the communication behavior of system components and their environment by means of message interchange. Thereby MSCs may be used for requirement specification, interface specification, simulation and validation, test case specification and documentation of real-time systems. Since in MSCs the communication behavior is presented in a very intuitive and transparent manner, particularly in the graphical representation, the MSC-language is easy to learn, use and interpret.

3. Test selection method based on formal description

The method, we present in this paper, creates optimal test suites using a formal specification and a finite size, unstructured and highly redundant test set by the help of mutation analysis. Firstly this method estimates the mutant detection ability of every test case. After that an optimization process – based on the previously earned results – ensures that only adequate test cases will be selected, and thus, a more effective test set is produced.

A mutation analysis system defines a set of mutation operators [11], where each operator represents a type of atomic syntactic change. Using these operators is practical for two reasons. On the one hand, they enable the formal description of fault types. On the other hand, operators make automated mutant generation possible. By applying the operators systematically to the specification, a set of mutants can be generated. A mutation analysis system consists of three components (Figure 1):

- Original system.
- Mutant system – it is a small syntactic variation of the original. Mutants can be created by applying mutation operators, where each operator represents a small syntactic change.
- An oracle – a person or in most cases a program to distinguish the original from the mutant by their interaction with the environment.

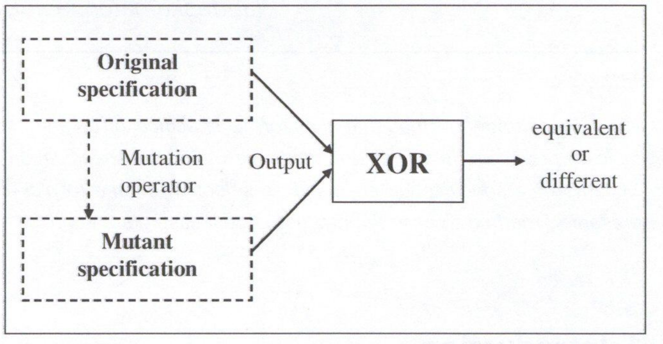


Figure 1. Components of a mutation analysis system

The test selection method needs a formal specification and an unstructured and highly redundant test set as a starting point. We only stimulate the system using inputs from the environment, and only check the outputs to the environment for inconsistency. Figure 2 shows the test selection process consisting of the following steps:

Test selection algorithm:

1. Apply a mutation operator to the specification, that is, create a mutant.
2. Run the test set on the mutant specification and check for inconsistency. If differences can be found between the responses of the original and the mutant systems, then the actual test case kills that mutant.
3. Repeat steps 1-2 until all possible mutants, and corresponding test cases are generated.
4. Select an optimal test suite from the original set, using a linear programming method
5. As a result we get a set of test cases.

We store the mutant detection ability of the test cases in a matrix. If a test case detects a mutation, the corresponding element of the matrix is marked true. This matrix describes for each test case the mutations the given test case is able to detect. When only one test case kills a mutant then we call that test case critical and we select it into the optimized test suite. There are some simplifications to reduce the matrix of criteria (see Figure 3). On the one hand, if there is a row with only false elements representing that the corresponding (MSC) test case did not find any of the mutants, the row can be omitted. On the other hand,

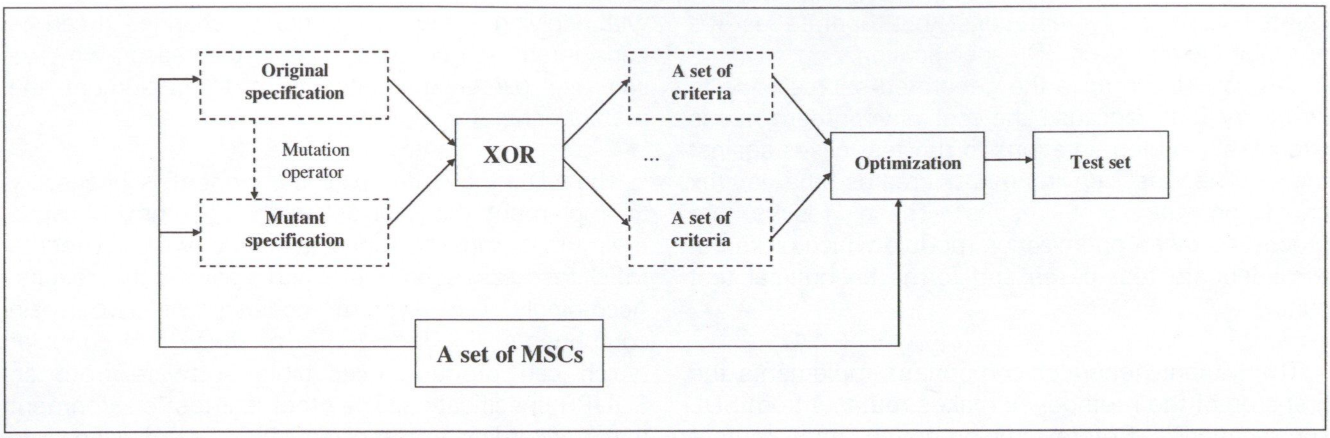


Figure 2. Structure of the automatic test selection system

if there is a column with only false elements, it represents either that the given mutant is an equivalent, or that there was no (MSC) test case in the original set that could find the difference.

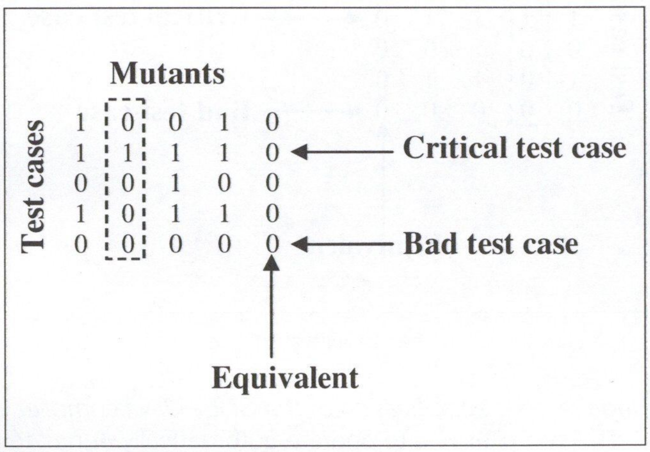


Figure 3. Matrix of criteria

4. Structure of the tool

We developed a Java software based on the previously presented method. As formal description language we used SDL. For test case representation we applied the MSC language.

The first step of the test selection method is to create a large number of test cases using the Test case generator component. After that we create mutants from the original specification. In our implementation this task is performed by the mutant generator component (Figure 4). This module applies mutant operators to a formal specification thus it creates mutant specifications.

The next step involves executing all MSC test cases on the mutant system. To fulfil this task we need a component which is able to create executable code from a formal specification. In our case this challenge is solved by the SDL→Java compiler. This module

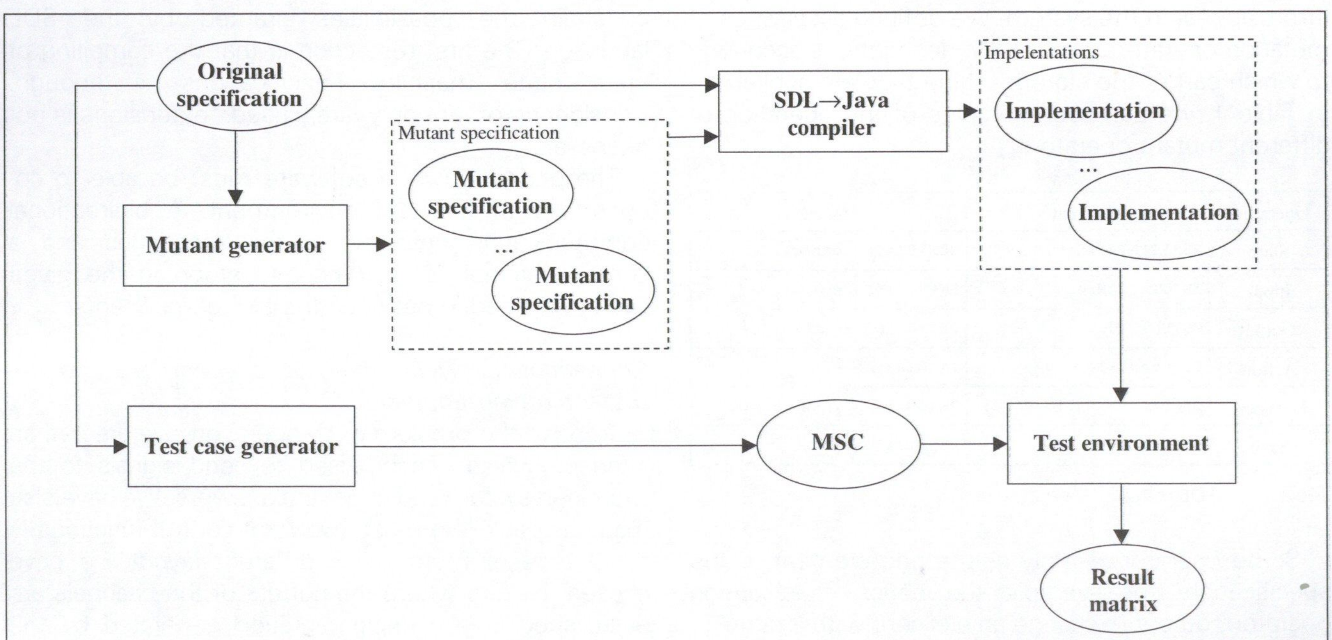


Figure 4. Implementation of the test selection algorithm

creates Java implementations that operate as it's described in the given SDL specification.

We have to compile the generated source code to its binary form because the test environment needs executable implementations to run test cases against them. The test environment creates the matrix containing the test results. These results are processed by an optimization module which picks out the adequate test cases and forms an optimal test suite.

The Mutant generator component implements the first step of the method – it makes mutants from SDL specifications. Before implementing the mutant generator, we have to define the mutant operators and we have to set a couple of principles:

- Mutation operators should model atomic faults.
- Only first-order mutants should be created.
- Only syntactically correct mutants should be generated.
- To be able to create test cases, only semantically correct mutants should be generated.
- Operators should create finite – and as small as possible – number of mutant specifications.

Why are the first two principles necessary? The reason is if test sets detecting changes to the original system created by a simple error would also detect complex changes created by applying a sequence of simple mutations. That's why we only apply first-order faults – i.e. we apply exactly one mutation at a time. To select conformance test cases, it is essential to generate syntactically correct mutants. Syntactic correctness is necessary to be able to execute the mutant system.

We defined the mutation operators – regarding the principles mentioned above - in order that all sorts of errors appear in the system. We defined six classes of mutation operators for CEFSM descriptions according to which part of the state machine they are applied to. In Table 1. we can see examples of the operation of different mutant operators.

Operators	Original	Mutant
State	`NEXTSTATE wait;`	`NEXTSTATE connected;`
Input	`INPUT ICONreq;`	`INPUT IDISreq;`
Output	`OUTPUT CC;`	`OUTPUT DT (number, d);`
Action	`Counter:=1;`	`/* Missing */`
Decision	`DECISION sdu!id = CC;`	`DECISION NOT(sdu!id=CC);`
Save	`SAVE IDATreq(d);`	`/* Missing */`

Table 1. SDL mutation operators

Some operators simply delete one element of the specification (for example the action modification operator), others exchange an element with another.

Since mutant generator produces syntactically correct mutant specifications from an SDL specification with applying a few small syntactic changes, it can be considered as a compiler. This is the reason why we can use parser generator tools to implement the mutant generator.

The SDL→Java compiler component is necessary to implement the test selection algorithm, because the mutants must be runnable in order we can execute MSC test cases on them and analyse the results. Accordingly our system consists of two main components. (Figure 5.) One of these is a compiler which can produce executable software from an SDL/PR specification. The other is a test environment that – according to the given MSC – sends signals to the simulated system and processes the answers arriving in return.

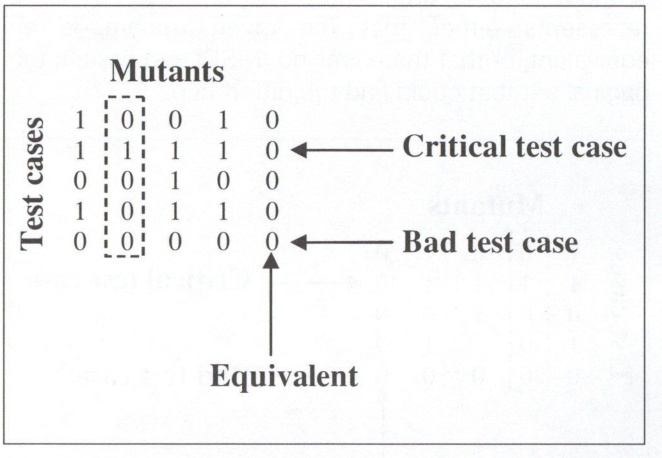


Figure 5. The role of the SDLÆJava compiler

Requirement specification of the SDL→Java compiler

The compiler has to operate automatically. It has to produce the code of the simulated system without human intervention. For this reason we have to constrain the possibilities granted by the SDL language. The first restriction is that the compiling of finite state machine specifications is enough. Considering other - only rarely used - extensions is not necessary.

The generated Java software must be able to co-operate with the test environment. A bidirectional communication channel for the messages and a control channel for starting and stopping the given system is needed between the two components.

Connection between the test environment and the mutant implementation

To solve the problem of co-operation, we created an interface which can be used to send signals to the examined system and to get the answers. We have also made available the most important control functions in this interface. From many different layout we have chosen the one where the puffers of the channels are established in the implementation generated by the compiler, because we have found it simple but effective. The composition can be seen on Figure 6.

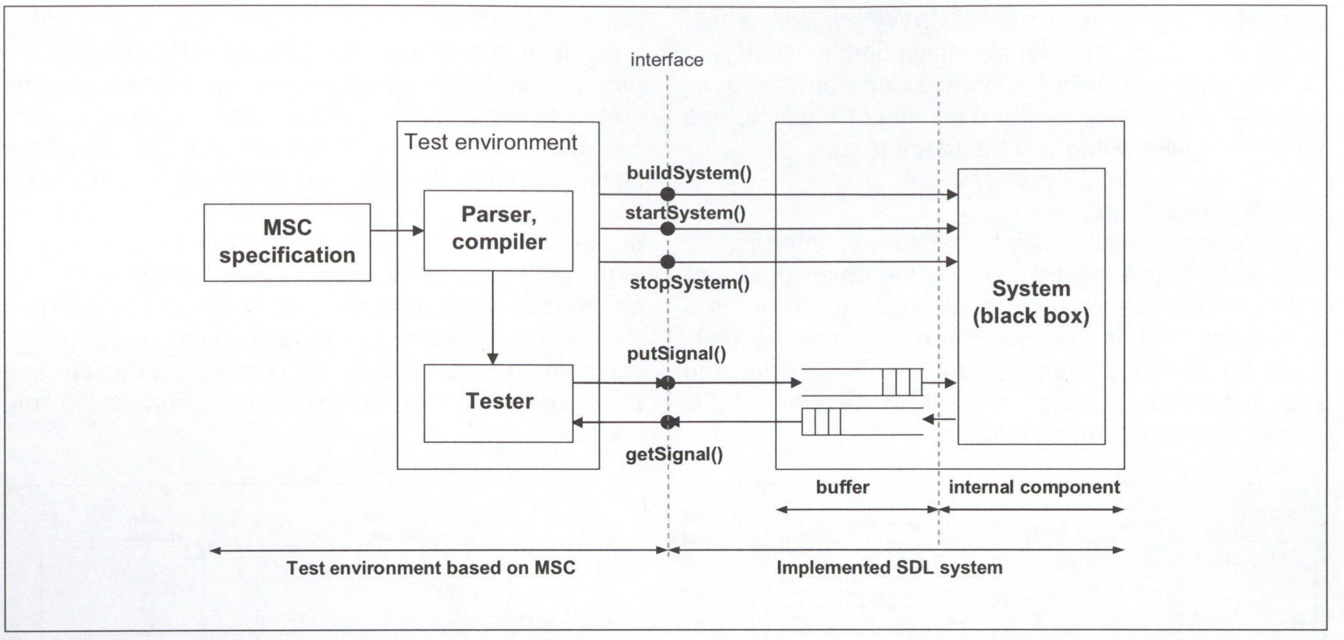


Figure 6. Co-operation between the test environment and the mutant implementation

An other fundamental point of the communication is that both sides have to understand the sent messages. In SDL the signals are defined with their names and the types of their parameters. Accordingly we created a universal Java class that can implement every kind of signals.

The implementation of the SDL→Java compiler

Before implementing the compiler it's practical to assign a Java code pattern for every SDL syntactic element. These patterns implement the functionality of the given elements. After developing the mappings between the source (SDL) and destination (Java) languages we can build the code generator part of the compiler on this basis.

We relied on the [4, 5] sources during the elaboration of the SDL→Java transformations, but we had to modify a few suggested patterns because of the requirements. For example in the context of our tool, timers don't need to be implemented. Timing events are controlled from the environment. This helps on executing a large number of MSC test cases in a shorter time, therefore we can save time.

The SDL components (signals, processes, channels, blocks etc.) are object-based, hence it's worth assigning a Java class to each. These classes must contain the characteristic properties and functions of the corresponding SDL component. We set up a library from these classes to reduce the complexity of the source code and to increase the reliability with code reuse.

The MSC parser, test execution component is the core of the tool. The inputs of this component are the program codes generated before and MSC test cases given in textual form. This component outputs a boolean matrix, which is the input of the last component, the optimizer. The final output of the tool the names of the MSC test cases found adequate.

The structure of the MSC is sequential, there are no conditions so we can parse the input files dynamically. After parsing an MSC line we make a decision about continuing the processing, because when inproper answer arrives we don't need to execute the remaining part of the test case. Consequently we can use the same parser with various MSC test cases, what makes it possible to test the system with different test cases at the same time.

After parsing a line on the basis of the parse information we either send a signal to the mutant system or wait for an incoming signal which we compare with the signal we created during the parsing.

Besides the ordinary signal we can send a special signal. This signal is created when a timer expires (**timeout** keyword in the MSC). To be able to test timer transitions, timeout events are made controllable from the environment. That is, whenever a test case reaches a timeout (for example timeout T3; in an MSC test case), a corresponding input is sent directly to the owner machine of the timer from the environment, and after its consumption the corresponding timer transition is executed. The timer signal is a message with a name and without any parameters; the exact handling of this message is the task of the tested system. We can save a lot of time from the test execution with the test environment handling the timers because the system doesn't wait until a timer expires.

Two messages are considered identical if their names and their parameters are the same. We compare the received signals with the test signal was created on the basis of the MSC. We consider a test successful if all the signals described in MSC are received correctly. When an invalid signal arrives to the test environment we stop the parsing of the MSC and we write a false value into the result matrix.

Graphical user interface

We need a well-structured graphical interface to control the test execution. During the development of the graphic environment we had to take into consideration that the user interface must be well-arranged and easy-to-use. That's why we created the sub functions (testing, mutant generation, MSC generation) in different windows.

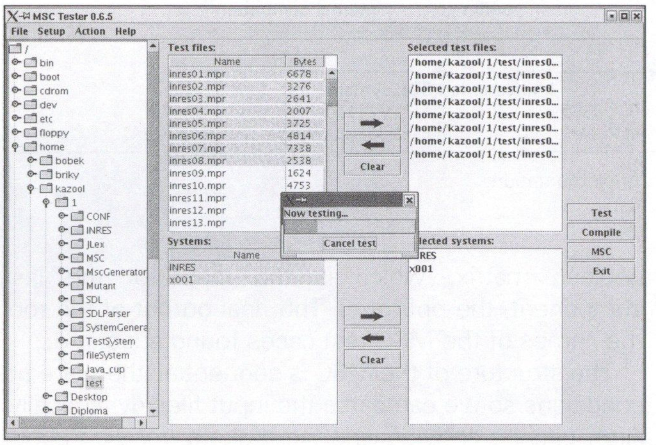


Figure 7. The graphical user interface

Figure 7 shows the main window of the tool. We can select test cases and mutant systems on the left. After we have selected the test cases and the mutants we press the Test button in order that the test environment starts test selection process. After the examination ended a table representing the boolean matrix appears. When we press the Compile button a new window turns up where the mutant system can be generated after the SDL specification has been selected.

5. Experiment on the system InRes

The InRes system (Initiator-Responder), is not a real system, although it is often used by demonstration of layered communication systems. It does contain many basic OSI (Open System Interconnection) concepts and is therefore very suitable for illustrative purposes. Besides it is easy to understand and not too big. The InRes system is similar to real systems and its specification contains most of the syntactic elements of SDL.

The InRes protocol implements a reliable, connection-oriented data transfer service, the *InRes service*, between two users. The InRes service is not symmetrical: it offers only one way data transition from an initiating process to a responding process. The protocol itself operates above a medium which offers a connectionless unreliable data transfer service. The protocol entities communicate by exchanging the protocol data units (PDU). The communication between the two protocol entities takes place in three distinct phases: the connection establishment phase, the data transmission phase, and the disconnection phase. The description consists of four processes, two at the both ends of the protocol (see the Figure 8). Two of the processes *Initiator* and *Responder* implement the service by exchanging protocol data units between each other. Two other processes, *coder_ini* and *coder_res* are used to hide the interface to the medium.

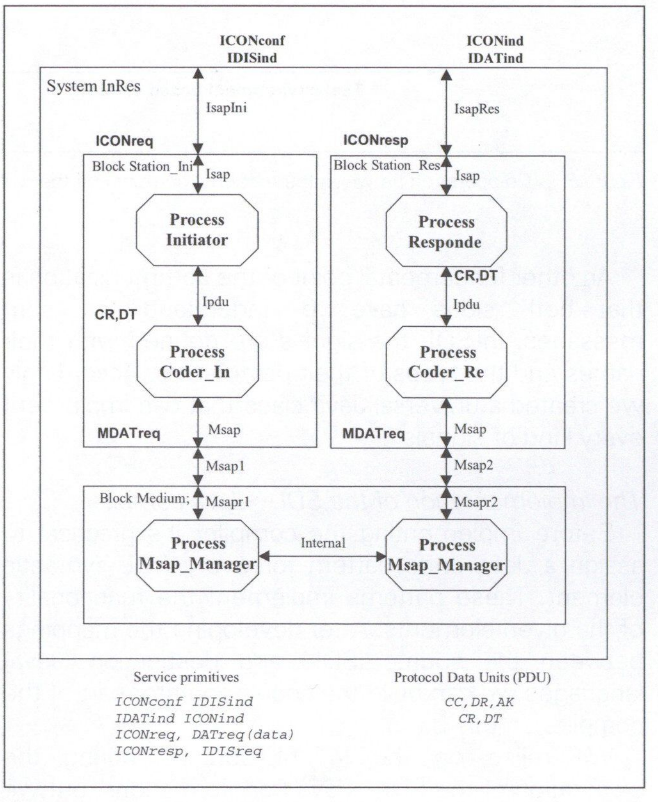


Figure 8. The InRes system in SDL

For the experiment and empirical analysis we generated 126 MSC test cases using the MSC generator component for the InRes protocol. This set intentionally included some appropriate and some useless test cases, and most of them were generated randomly. By applying the operators to the specification, 125 mutants were generated automatically. Finally, nine test cases of the original set were selected. The selection included some of the test cases we initially considered appropriate based on our knowledge of the system.

As the data shows, twenty mutants have not been discovered by any of the MSC test cases. The worst mutation uncover ratio is, in the case of state mutation operators, and the best is in the case of input exchange. More than half (12) of the cases, where no difference has been found is in the Initiator process

Mutation operators	Mutants generated	Detected	Detection ratio[%]
State	30	19	63
Input	28	23	82
Output	36	30	83
Action	221	15	68
Predicate	8	5	63
Save	1	1	100

Table 2. Mutation Analysis Applied to the SDL specification of InRes

that is the largest in the system. Six of the mutants of the Responder process have not been killed. Both in case of processes Coder_Ini and Coder_Res, one mutation has not been revealed.

There have been some mutations discovered by only one test case, we call them critical mutants. The test cases detecting these critical mutants have to be included in the resulting set by all means. Interestingly, these test cases give eight of the nine selected. Out of the eight critical mutants two has been generated using predicate, two using output, three using state and one using action mutation operator. Applying input mutation generated no critical mutants.

Both this, and the mutation uncover ratio indicate that input mutation (and input like mutations, e.g. missing transition) operators result in very rough mutant systems. That is, they produce radical changes in the behavior of the – InRes – system that can be detected by most of the test cases. State mutation, on the other hand, produces errors that can be discovered by only a small percentage of the test cases, but it has provided three critical mutants, more than any other operator.

6. Conclusions

Fault-based testing requires special fault models and tries to detect faults in the implementation, with respect to the specification. In this paper, we described how to apply mutation analysis, a white box method, to formal SDL specifications, and use mutant systems to automatically select adequate test cases for black box testing. For this purpose we created and formally specified a set of mutation operators for SDL specifications. Using these operators we don't need the knowledge of the typical errors of an implementation. Instead the method generates erroneous specifications that make automatic test selection possible.

The recent method and the tool are useful not only for simple protocols, but also for real, complex telecommunications systems described in SDL. It's easy to examine different implementation with the automatically generated test sets.

References

1. G. Kovács, Z. Pap, dr. Gy. Csopaki: "Automatic test selection based on CEFSM specifications", 7th Symposium on Programming Languages and Software Tools, pages 200-212, Szeged, Hungary, 2001.
2. Le Viet Dung, Wu-Hen-Chang Antal: „Automatikus tesztkiválasztás SDL specifikáció alapján mutációs elemzéssel", TDK dolgozat, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2001.
3. Andriska Zoltán, Bátori Gábor: „Tesztkörnyezet kialakítása MSC alapú mutáció elemzőhöz", TDK dolgozat, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2001.
4. Horváth G.: „Formális nyelven leírt protokollok Java nyelven történő megvalósítása", Diplomaterv, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2000.
5. Kovács G.: „Az SS7-MTP2 protokoll megvalósítása Java nyelvi környezetben", Diplomaterv, Budapesti Műszaki és Gazdaságtudományi Egyetem, 2000.
6. ITU-T recommendation Z.120 Message Sequence Chart, 1996
7. ITU-T recommendation Z.100 Specification and Description Language, 1992

Laboratory Equipment Type Fiber Optic Refractometry System

JÁN TURÁN[1], L'UBOŠ OVSENÍK[1], EDUARD F. CAROME[2], KÁLMÁN FAZEKAS[3]

[1]Technical University of Košice, Slovakia; [2]John Carroll University Cleveland, Ohio USA;
[3]Budapest University of Technology and Economics, Hungary

Reviewed

Using fiber optics and microoptics technologies it is possible to design innovative fiber optic index of refraction transducer that has unique properties. On the base of this transducer a laboratory equipment type fiber optic refractometer was developed for liquid index of refraction measurements. Such refractometer may be used for medical, pharmaceutical, industrial fluid, petrochemical, plastic, food, and beverage industry applications. For example, it may be used for measuring the concentrations of aqueous solutions: as the concentration or density of a solute increase, the refractive index increases proportionately. The paper describes development work related to design of laboratory type fiber optic refractometer and describe experiments to evaluation of its basic properties.

1. Introduction

Almost all currently available classical refractometers employ a prismatic element on which the liquid sample is placed [1,2]. These instruments yield an output that is based on the degree of a light beam at the liquid-prism interface. In the simplest refractometers, the index is determined visually, via a lensed viewing tube, by observing where the output beam intersects a graduated scale. In more modern, digital type instruments, the degree of bending is measured automatically using a linear photodetector array [1,4].

Using fiber optic technology and microoptic is possible to design innovative fiber optic index of refraction transducer that has unique properties [3]. Basically it consists of input – output pair of simple multimode fibers that interrogate a small lens. The cone of light injected into the lens from the input fiber is internally reflected from inside surface of the lens and focused back into the output fiber. When the outer surface of the lens is in contact with a liquid, the attenuation of the light reaching the output fiber depends strongly on the refractive index of the liquid. Depending on the construction and design of such a transducer, this transducer may have a wide dynamic range. Its relative sensitivity, i.e. the ratio of fractional change of the optical intensity for a given change in index of refraction, is substantial, (of the order of 5 to 10) over a wide range of index, e.g. n = 1.3 to 1.6.

To use one of the many commercially available bench type and hand held index of refraction instruments, or refractometers, a few drops of the liquid under test must be withdrawn and placed on a planar measuring surface, while to use one of the few in-line refractive index monitors that are currently available requires a complex installation, including two or more relatively large ports and/or flow bypass tubes.

The new fiber optic small, probe type structure, transducer elements, can be easily inserted into the top of liquid containers or through a simple fitting can be simple to inserted in a flow line. It is essential to combine these very practical and versatile transducers capable of detecting small changes of index of refraction with "smart" data acquisition and signal processing technology to develop new versatile and practical instruments.

The paper describe development work of a flexible fiber optic refractometer and its experimental evaluation. Two types of fiber optic refractometers are considered: a basic fiber optic refractometer (based on the use of one fiber optic transducer) and a most sophisticate differential fiber optic refractometer (based on the use of two fiber optic transducers). Both types of refractometers are developed as a laboratory equipment and tested in experiments in laboratory and field conditions.

2. Architecture of fiber optic refractometer

On the base of combination of fiber optics and microoptic technologies new small, flexible and probe type index of refraction transducer element was fabricated [3] and for practical (industrial) application packaged into transducer module (Fig.1) with light source, photodetector and thermistor based temperature sensor element.

Depending on the construction and design this fiber optic transducer module may have high relative sensitivity (of the order of 5 to 10) over a wide range of index of refraction, e.g. n = 1.3 to 1.6 and can be used to measure at relative high distance placed measured points (up to 1.5 km).

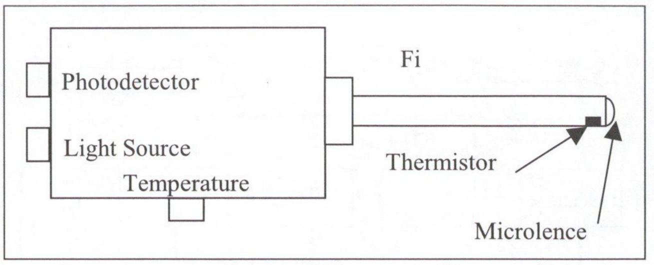


Figure 1. Fiber optic transducer module

The basic transducer properties are illustrated on Fig. 2 and 3, which show linear and semi-logarithmic plots, respectively, of the optical power output of a typical transducer as a function of index of refraction. For convenience, usually employed micro-lenses have diameter from 5 to 1 mm. In principle, it should be possible to go to even smaller diameters, e.g. 250 to 300 microns – so that relatively small catheter type transducers are feasible. Since the basic design is quite simple it also is possible to produce extremely rugged transducer elements that are suitable for field/industrial applications, even for remote sensing.

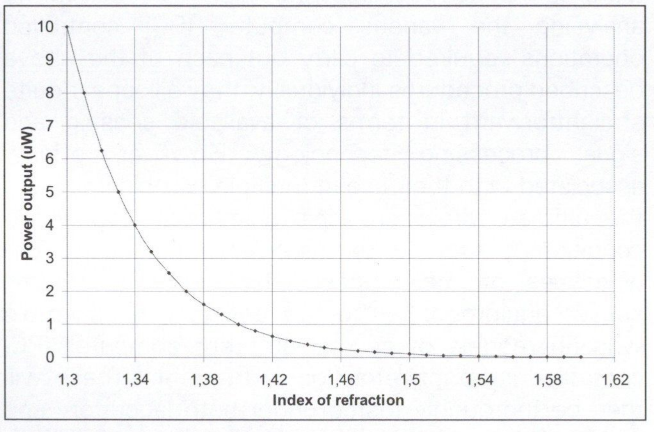


Figure 2. Linear plots of the transducer output vs. refractive index

Since the index of refraction is strongly dependent on temperature and also, to a lesser degree, on wavelength, these effects must be corrected for in designing and/or using such instruments. They are of major importance when one attempts to determine an index of refraction to 1 part in 10.000 or better, which is the desired sensitivity for a high quality, laboratory type instrument. Their importance can easily be seen in the case of water, for example, which has a temperature coefficient Dn/DT of 1,5 parts 10.000 per °C.

As with any other refractometer, any instrument that employs this fiber optic based transducer module must be capable of correcting for the intrinsic temperature dependence of a liquid's index of refraction [2,3,4]. In addition, however, for an intensity type fiber optic sensor, other corrections and precautions must be taken especially if, as already mentioned, a precision of 1 part in 10.000 is to be attained. It will be necessary to employ low noise electronic circuitry and/or correct for photodetector dark current, especially at higher indices, where the output light intensity is strongly attenuated. It also will be necessary to correct for light source and photodetector temperature sensitivities and for any stray light that might affect the photodetector. In one sense, in terms of capabilities of today's microprocessor controlled "smart" sensor technology, it should be straightforward to design instruments that automatically "massage" the raw transducer data to correct them for each of these effects [2,3,4].

Several architectures for design optic refractometer equipment are possible:

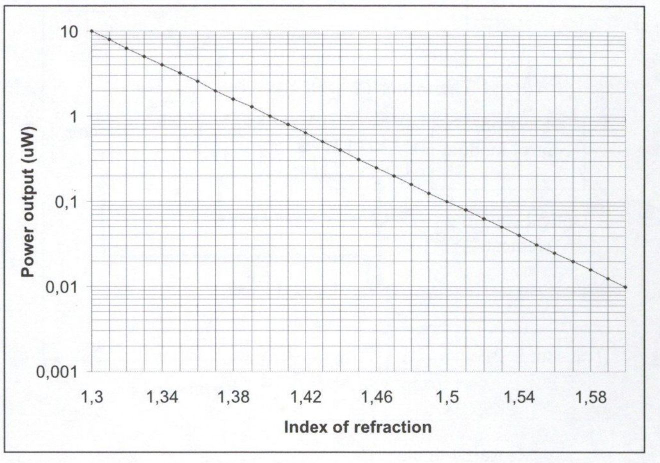


Figure 3. Semi-logarithmic plots of the transducer output vs. refractive index

One type of the basic equipment architecture proposed is to produce computer (PC) controlled laboratory instrument capable of determining the index of refraction of various liquids to 1 part 10.000 or better. Another type is to produce microprocessor controlled, handhold and other similar types of small, portable refractometers, probably of somewhat reduced accuracy, that can easily be used for field type measurements. A third type is to produce in-line refractive index monitors for use in chemical, industrial, food processing, and other similar facilities. Refractometers are frequently used for medical, pharmaceutical, industrial fluid, chemical, petrochemical, plastic, food, and beverage industry applications. For example, they are used for measuring the concentrations of aqueous solutions: as the concentration or density of a solute increase, the refractive index increases proportionately. Included in such measurements is the percentage of sugar in fruits, soft drinks, canned syrups and other solutions. They also are used to determine the salinity of aquariums and of solutions used in food processing, the freezing point of coolants and deicing fluids, the charge status of acid batteries, the serum protein and urine specific gravity, etc.

In the following part of this paper we consider of design of laboratory type equipment. Referring to the block diagram on Figure 4, the basic architecture, as presently conceived, consists of the following elements: a light emitting diode (LED) or semiconductor laser diode (LD) light source and its electronic driver/pulser circuit; a monitor photodetector to determine the output light level of the used light source (D1), a second photodetector (D2) to record the return light from the fiber optic transducer module (P1); and a PC to automatically control the system and process the data from the various elements. PC will compute the ratio of the intensity of the transducer output for an unknown liquid and that recorded earlier for a standard liquid, e.g., water.

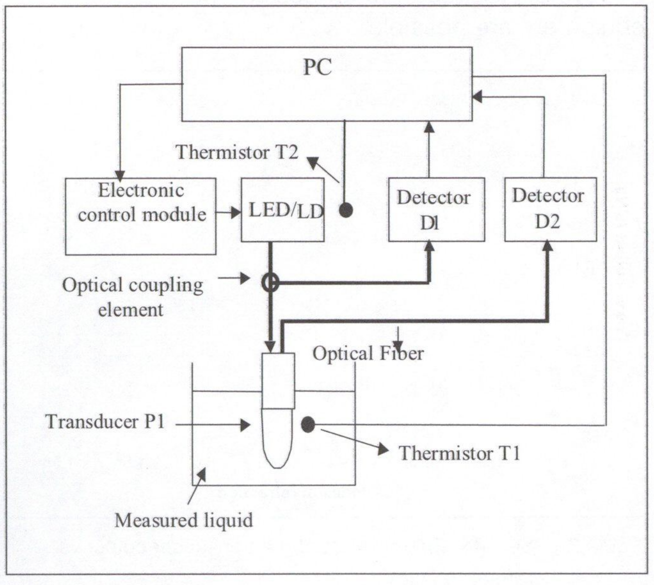


Figure 4. Block diagram outining the design of basic fiber opticc refractometer

The PC will then determine the index of refraction, either by a comparison and interpolation process between this ratio and those in a calibration data lookup table, or by computation using a transducer response equation. In addition, the temperature of the liquid sample and of the source/detector module will be measured simultaneously with the index of refraction, to correct for their temperature dependence. The liquid sample temperature will be determined using a thermistor, as indicated on Figure 4, or using a fiber optic temperature sensor in applications requiring an all dielectric transducer, e.g., for use in explosive or high voltage environments [5].

A second possible architecture of the equipment system, as outlined on Figure 5, was also developed. Basically, it employs a differential technique and would allow measurements/comparisons of index of refraction to a very high precision. Instead of taking comparative readings of index of refraction of a known and an unknown liquid at separate time, as would be done with the architecture outlined on Figure 4, both readings would be taken simultaneously, using two index of refraction transducers (P1 and P2), one in the unknown and the other in a standard liquid sample.

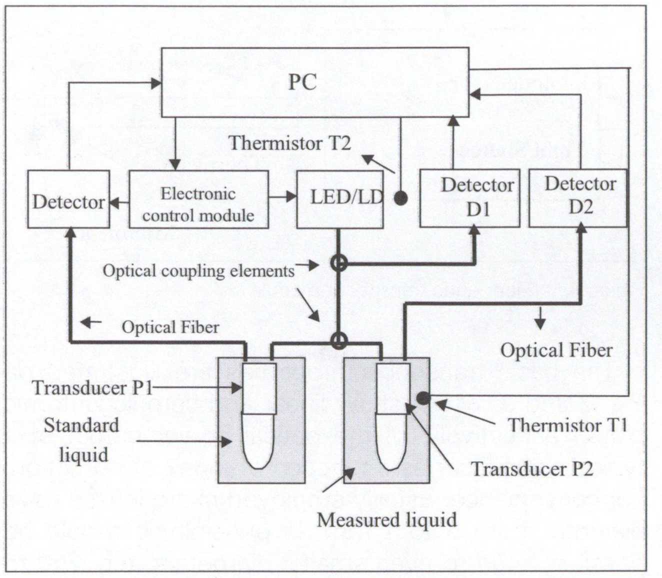


Figure 5. Block diagram outlining the design of a differential type fiber optic refractometer

This technique would be especially useful when extreme relative accuracy/precision is required. In analysing the various computer (PC) controlled operations required to carry out each of the above described procedures individually, they all appear quite straightforward in terms of available sensing and signal processing technology. However, effects associated with their use in multiple combination, and the ultimate precision achievable under real world conditions, must be examined experimentally. The objectives of the proposed laboratory fiber optic sensor equipment will be to design and assemble working models of several different computer (PC) controlled index of refraction instruments. These will then be thoroughly tested under both laboratory and typical field conditions to determine the ultimate precision and accuracy achievable with various designs.

3. Laboratory equipment design

The block scheme of the laboratory fiber optic refractometer equipment is on Figure 6. Equipment is able in hardware and using dedicated software realise both architectures (Fig. 4 and Fig. 5) of basic and differential type of fiber optic refractometers. The front panel of the equipment is through BNC type connectors connected to fiber optic transducer modules (P1 and P2).

There are two types of transducers available: a LED sensor module and a LD sensor module, which can be used with the described laboratory refractometer. The LED diode sensor module is based on a pair of LED (HFE4020) as photoemitter and photodiode (CLD42) as photodetector. The LD sensor module consists of LD (HFE4080-321) as photoemitter and phototransistor (CD1440) as photodetector.

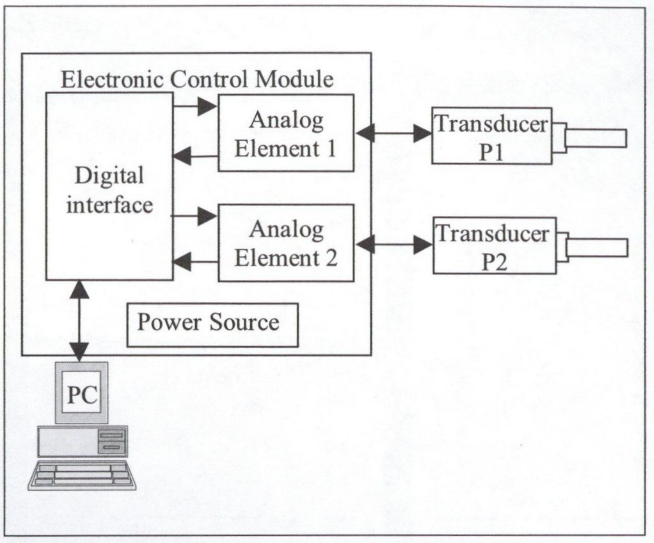


Figure 6. Block scheme of the laboratory fiber optic refractometer

The analog element contains a driving circuit for LED and/or LD based transducer, which produces appropriate current for specific type of transducer, respectively. There is a circuit for back measurement of driving current for LED and/or LD based transducer, for more accurate driving/detection rate in the analog element. A detection circuit of photodiode for LED sensor module and/or circuit of phototransistor for LD sensor module, respectively, is responsible for amplifying and adjusting of measured currents within range of ADC (Analog Digital Converter). A detection circuit of thermistor adjusts a voltage from thermistor to ADC range (0-10V).

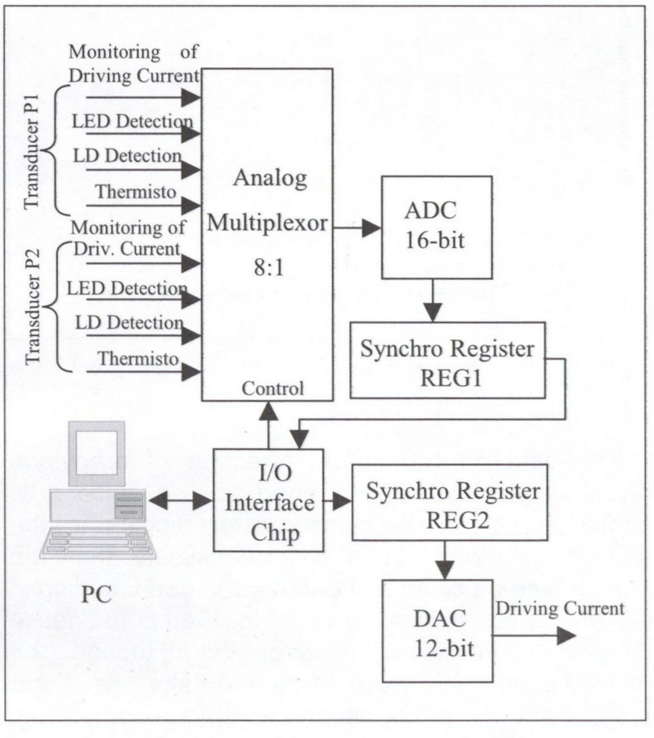


Figure 7. Block scheme of the digital interface to PC

The described instrument is designed as a laboratory instrument controlled by computer (PC) which is powered from the standard supply network ~230V/50Hz. There are transformers and stabilizators for stabilization of necessary voltages (+5V, ±15V) inside of the control module.

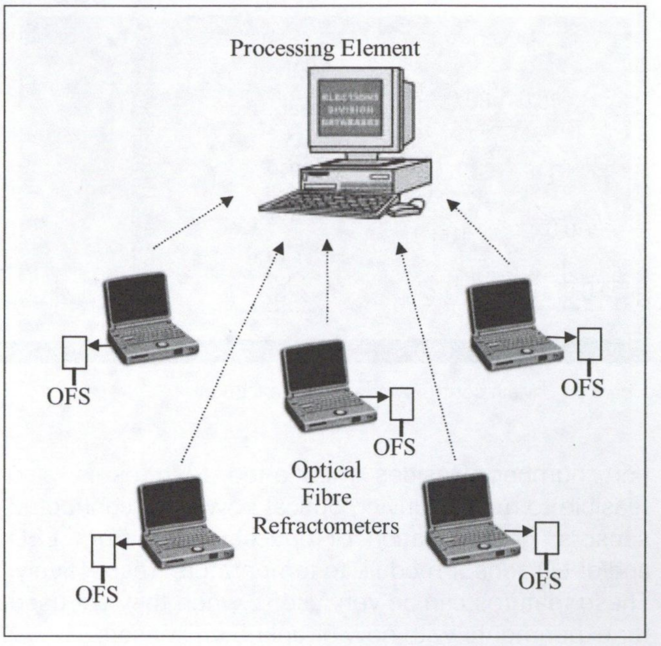


Figure 8. Optical fiber refractometry industrial telemetry systems

The interface sensor-computer is developed with the help of the I/O interface chip (MHB8255), Fig.7. The interface chip is responsible for communication of control module with computer via parallel PC interface. A part of I/O chip controls a 16-bit ADC, starting of conversion and reading out of data to computer. There is implemented a synchronization register REG1 because of difference in width of data line between ADC (16-bit) and I/O chip (8-bit). Analog multiplexor 8:1 is addressed and controlled by I/O chip. There are 4 analog channels for each transducer (driving current, LED diode sensor detection, LD diode sensor detection and thermistor data). The I/O interface chip generates a data word for 12-bit DAC (Digital Analog Converter), for driving current of LED and/or LD diode transducer, respectively. There is implemented another synchronization register REG2 since data line of I/O chip (8-bit) is different to width of DAC data line (12-bit).

The program package for operation with laboratory refractometer is written in design environment of the LabWindows/CVI version 4.0.1. The program allows performing of refractive index measurements with basic or differential architecture. In the main control window (Fig. 8) we can see value of measured refractive index, temperature of measured liquid and value of actual driving current for each measurement. Within a window menu is possible to change a method of measurement, type of sensor and communication

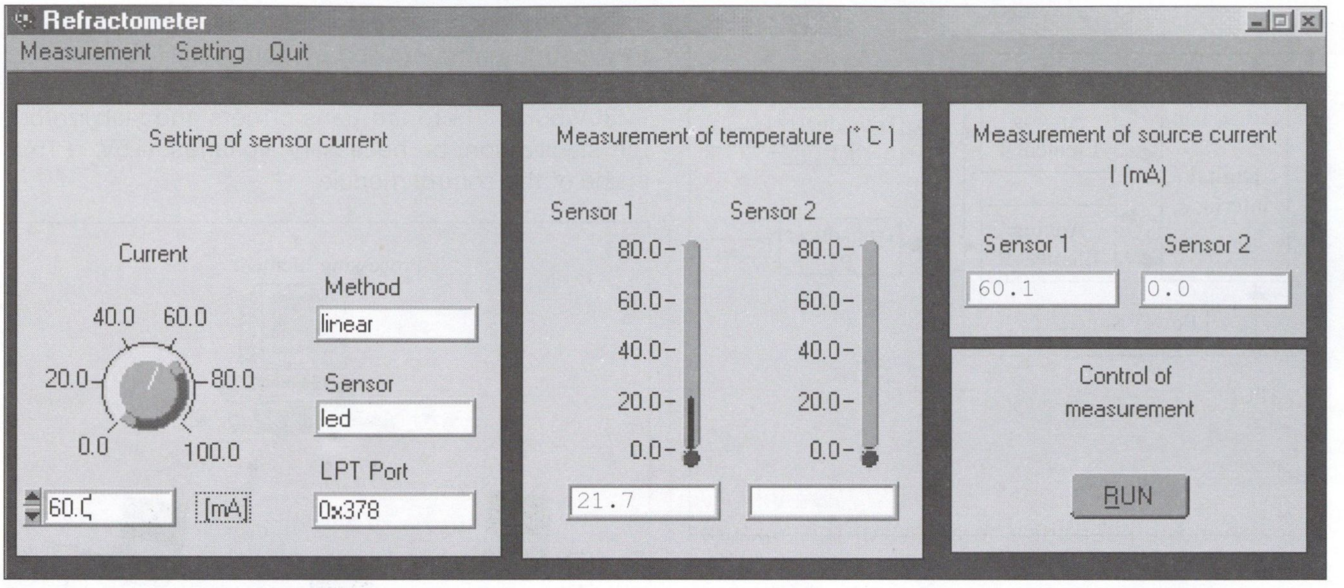


Figure 9. Main control window for the laboratory

port number. Besides mentioned features is also feasible to modify driving optical power for connected sensors, get a relation of optical output from LED and/or LD sensor module to temperature, respectively. These features can be very useful when they are used for experiments with new or unknown sensors.

Interconnecting several fiber optic refractometry equipments using fiber optic networking techniques (Fig. 9) it is possible to develop a fiber optic refractometry system. Such system may by suitable for various industrial and scientific laboratory applications.

4. Experiments and Results

Performance of the developed laboratory fiber optic refractometer equipment was evaluated by various testing measurements, this is here demonstrated by results of two basic laboratory experiments and measurements of index of refraction of various petrochemical products:

A) Basic laboratory experiments

- dependency of the refractive index of propylene glycol on temperature (Fig.10),
- dependency of the refractive index of water propylene glycol solution on propylene glycol concentration (Fig.11).

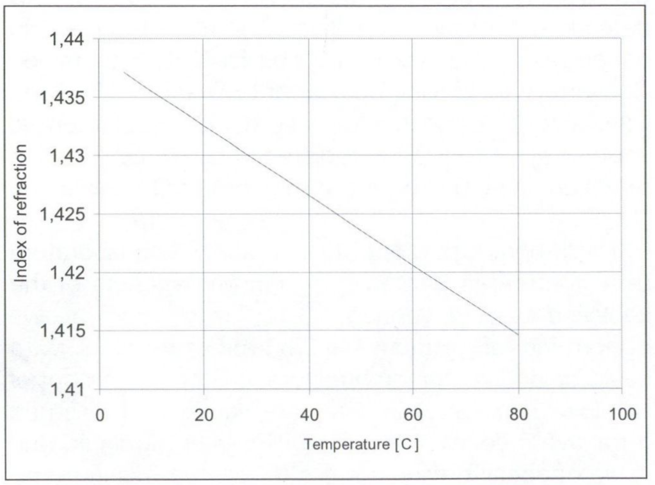


Figure 10. The refractive index of propylene glycol vs. temperature (°C)

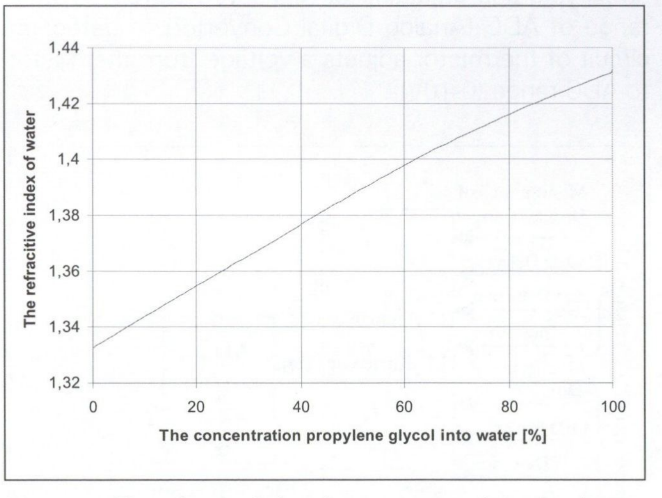


Figure 11. The refractive index of water vs. propylene glycol concentration

The data on the index refraction of propylene glycol/water mixture, are good to use since it is simple to, clean the sensor after dipping in the various solutions, since it is necessary that the sensor face be clean and dry to get a good reading in air and also before dipping in water to get a calibration/comparison reading. Even though the index varies with wavelength we take the water reading at 20°C to correspond to an index of 1.3330, and refer all other readings to this value. The dn/dt for water in the range 15 to 30°C is 0.0001 per degree °C, while that for 100% propylene glycol is 0.0003. Thus for glycol/water solutions one could assume a linear dependence of dn/dt, that is, for example assume dn/dt = 0.0002 for a 50% solution. In all instruments with automatic temperature compensation, it is assumed that they are to be used with water solutions, and it is assumed that dn/dt = 0.0001. We may use also propylene glycol, either 100%, taking its index to be 1.4312, or 100% plus a few mixtures.

B) Measurements of petrochemical products

As field tests we use some petrochemical products of known index of rafraction. They make it difficult to clean the sensor face however. We recommend to keep the containers of the mixtures sealed when not in use since they tend to drift in index, though this is not much of a problem. We frequently make up relatively large sample, e.g., 100 to 200 cc, and then use one half for our measurements for while and after a week or two, take a reading in the other half that has been kept well sealed. That way we can determine if there has been any shift in the index of our measure ment sample. The results of measured index of refraction are depicted on Fig. 12, and are in very good success as compared with results obtained by classical methods.

5. Conclusion

During the development work PC controlled laboratory type fiber optic refractometer for liquid index of refraction measurements was constructed. The device was able in hardware and with appropriate software realise both basic and differential type of proposed fiber optic refractometer architectures. Basic properties of the equipment was evaluated in laboratory and field type experiments with measurements of standard and petrochemical product liquids.

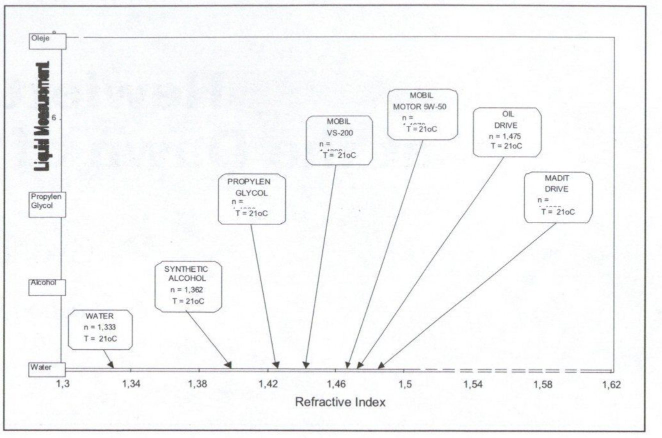


Figure 12. Result of petrochemical products measurements

6. Acknowledgements

The authors are thanking for the financial supports from the COST 276 project and VEGA grant No. 1/7444/20.

References

1. DAVIS,Ch.M.-CAROME,E.F.-WEIK,M.H.-EZEKIEL, S.-EINZING,R.E.: "Fiber Optic Sensor Technology Handbook. Optical Technologies", Herndon,1986.
2. TURÁN,J-PETRÍK,S.: "Fiber Optic Sensors". Alfa, Bratislava, 1990.
3. TURÁN,J-CAROME,E.F.-OVSENÍK,Ľ.: Fiber Optic Refracto-meter for Liquid Index of Refraction Measurements. In. Proc. of TELSIKS2001, Niš, Yugoslavia, Sept. 19-21, 2001, 489-491.
4. JASENEK,J.-ČERMÁK,O.: "Optical Refractometry with Synthesized Coherence Function", In: Photonics, Devices and Systems, Proc. of SPIE, Vol.4016, 2000, 204-210.
5. TRAVICA,S. et al.: "Optically Powered Fiber-Optic Temperature thick Film NTC Sensor", Proc. of LASER'98, Tucson, Arizona, USA, December 1998, 562-567.

We wish a merry Christmas

and a happy New Year for every reader

Editorial Board

Hewlett-Packard: at the Dawn of the New Decade

GYÖRGY BŐGEL

gybogel@kfki.com

The news on the planned merger of Hewlett-Packard and Compaq were published first on 4. September 2001 on the home pages of the news agencies. For a couple of days this event was on the headlines of all the papers, then came the terrorist attacks on 11 September in New York when the world's attention was turned to another direction.

Both the professional public opinion and the press received the news of the merger calmly. At Boston Hotel Marriott where the executives of both companies were present, hardly any applause could be heard. „Here we have two companies – Sanjay Jhavery the technological analyst of Vontobel Asset Management of Switzerland – commented to the press – who simply want to survive. They back the wall to protect themselves." Right after the announcement the share prices of HP drooped by 21.5 percent while those of Compaq increased by 15.7 percent. The joint loss of value amounted to USD 13 million in two days.

According to the opinion of Mr. Carly Fiorina CEO of HP they are excellently positioned to seize the opportunities when the companies – after the recession emerging in 2001 –start again to spend more on IT. The better focused competitors – like IBM, Sun, or EMC – have stronger market positions at the upper end of the market, still, the customers who are fed up with the promises of the stock exchange balloon era – will appreciate the complex solutions offered by the two companies. The costs can be reduced through the rationalization following the merger, thousands of jobs can be terminated, and the companies will be able to negotiate with certain powerful suppliers from the position of strength, while risks will be kept at manageable level. Excellent, new opportunities will open up in the areas of consulting and outsourcing, the large consulting companies inevitably need partners who are able to construct the necessary infrastructure. The demand for PCs will increase again when new PCs are designed for new purposes. The two companies have excellent innovative abilities and market leading position depends on this feature in the long run.

The fusion – if it takes place – will be no simple procedure. The product lines of the two companies are in overlap, both of them are present in 160 countries of the world with a total staff of 150 thousand. HP has no experience in the implementation of by-outs of this volume Compaq has not been able to step over in every respect the 1998-year merger of Digital Equipment. Both companies face problems of internal reengineering.

From the garage to the top

The Hewlett-Packard company was established in 1938 by two young American engineers, Mr. William Hewlett and David Packard – both of them were the students of Fred Termannak of the Stamford University – who set up the company with a total capital input of cca USD 5 hundred. We can say that at that time not only one of the leading companies of our world economy was born but also the typical American story of the Silicon Valley.

This was a typical „garage story" in the closest sense of the word as Hewlett and Packard really developed their first market products in a garage. They could easily have developed any other product: to the regret of their professors, in the lectures they held for business schools they often repeated that they had had no plan to the start-up of a company, they simply seized the opportunities they came across. They wanted to establish a company, and did whatever they thought they were able to do, and whatever they expected to bring profit: they developed meters for a telescope, and constructed automated water lubrication for a washbasins and so on.

The start was extremely successful: the revenues of the company doubled in each year of the forties, Hewlett and Packard strictly insisted on self-financing for quite a long time, they did not take loans, and offered their shares to the public only in 1957 for the first time, and even at that time they sold not more than 10 percent of the shares. They maintained the opinion that financing from loan would undermine the entrepreneur spirit. They were similarly careful to accept risk capital, claiming that it would inspire too rapid growth of the companies, and if you develop too rapidly you may suffer losses.

In the following decades HP became the famous model company of the American economy, featured by the practice of progressive human policy innovative mind, entrepreneur attitude, flooding the market with excellent technical products ever since its existence. No doubt the management practice and corporate culture established by Hewlett and Packard (the „HP way" as it is normally called) has survived long after the death of the founders, it was easy to detect, and it helped the company to survive even the most difficult times.

Although HP grew over the notable garage soon, the managers committed themselves to maintain the bounce of mall enterprises, and its norms of behavior. David Packard writes in his book on organizational issues published in 1995, titled „The HP Way":

„At HP we did not devote too much attention to organizational issues until he fifties. There was no need to do so. We had a well-defined choice of products closely related to each others, we did everything at one single location, developed our own sales and agency network; we were a strongly centralized, functionally organized company, with vice-chairmen responsible for marketing, production and R&D. As the company grew started to diversify, Bill (Hewlett, the other founder) and me started to realize that we must find some sort of decentralization strategy to maintain the personal accountability and interest in profitability. We had the concern that personal involvement will disappear from HP that used to be the integer part of our behavior before.

The idea of splitting up the company to divisions came up first in connection with the product development laboratories, where we formed four groups under the control of the vice-chairman managing the R&D activities, and each of the groups were responsible for a particular product family. It happened in 1957. Later on further steps followed in this direction, that were motivated by the establishment of production plants in Germany and the acquisitions started at that time.

Mid sixties we had a dozen of operative divisions, each of them organized as an integrated, self-sustaining organizational unit responsible for the development, production and marketing of its own products. The reason why we established these units is to provide for their independence, operate them in an environment that motivates individual performance, initiative and creativity, to allow them quite a freedom of maneuvering while working for the attainment of the mutual goals. We wanted to eliminate bureaucracy, and had the objective of making the decisions at the place where the problems incur. We wanted to have divisions that can maintain the intimate atmosphere, where people receive special attention, communication is free, that is, everything is the same as it used to be in the early days of the company....

...As time passed some of the divisions started to flourish,: they manufactured large variety of products, the number of employees reached 1 500. At this stage the lines of communicants got saturated, management becomes troublesome, people loose their direct connection to their products and the results of their division. At this stage we implemented our policy that has been applied even today, that is, we have separated certain parts of the over-sized divisions, assigned a successful and profitable product line to them and geographically removed them to a remote location.

...The rapidly growing companies often undergo organizational changes In the case of HP the following, dramatic change took place in 1968. As the number of product lines and divisions increased constantly, we gradually implemented the group structure. It meant that we formed groups from the divisions manufacturing close products, servicing the related markets, that were controlled by group leaders having their own staff. Each group was responsible for the coordination of the division and its financial performance. This decision was governed by dual objectives: On the one hand, we made attempts to increase the cooperation among the related divisions, on the other hand, we wanted to decentralize some of the activities performed by the top management before, among others, for example, part of the engineering and designing work was shifted to the groups...

...In the beginning of the nineties HP had as many as 65 divisions, arranged to 13 groups".

Strategy and corporate culture

In a constantly growing company the values and norms of the founders were difficult to transfer in an indirect, personal and informal way. In the book of Packard we can find the cooperate objective formally published in 1966:

1. Profit. Profit is on the one hand the best tool of measuring our contribution to the welfare of the society, on the other hand it is the source of the capital of our company as well. We have to maximize our profit that can be still harmonized with our other cooperate objectives.
2. Customers. We have to make efforts to improve the quality of our products and services, and increase their usefulness.
3. Focus. While we are constantly looking for new growth opportunities we can focus only on the areas where we really have the skills and abilities, and where we are able to reach substantial results.
4. Growth. Growth is the indicator of power to us, which is an absolutely necessary condition of survival.

5. Employees. We have to provide the HP employees with opportunities to work to share with them the success of the company which is due to these employees. Based on their performance we have to provide for the working environment, and we are convinced that the results they achieve in work will bring them personal satisfaction.
6. Organization. We have to maintain an organizational environment that provides for personal motivation, initiative, creativity and high level of personal freedom in course of implementation of the corporate objectives.
7. Social responsibility. We have to be good citizens of our company, and meet our obligations towards the community creating our labor environment, and towards the social institutions as well.

The goals listed here did not remain mere slogans at HP. Lets see, for example the responsibility for taking care of the employees. In 1940 a profit sharing scheme was introduced, in the frame of which the director was entitled to the same percentage of bonus as the warden and simultaneously with that an insurance scheme was rolled out to all the employees of the company. At that time these were completely extraordinary solutions especially among the small companies. When in the fifties the company became a public joint stock company, all the employees with an employment exceeding 6 months was entitled to shares, that was supplemented with a shares option program. Later on the employees' share program was subsidized with a 25 percent discount.

The company rejected some of the large scale government orders as on completion of the project they would have had to dismiss the employees becoming redundant. In the case of vacancies the divisions were expected to recruit staff in-house, from within he company, this method provided for the protection of the company culture as well. At the time of recession all the employees were requested to go for an extraordinary holiday and agree to the 10 percent reduction of their wages to avoid termination of employments. The company was among the first American companies to introduce flexible working hours at each level of the organization and conduct employee satisfaction surveys. They were seriously committed to the open doors' policy, the employees were free to contact even the highest level management with their problems. Despite the common American practice of the fifties none of the managers could have an own office with closed doors. In view the above, no wonder that the trade union campaigns failed at HP from time to time.

Despite the practice of that age the managers of HP did not allow to the HR department to get involved in the HR affairs of the management: they had the opinion that all the managers must assume responsibility for the employees, and no one is supposed to transfer the tasks related to the employees.

The managers made all the efforts to keep the company in the mainstream of technical development. Starting from the fifties they recruited their engineers from among the smartest graduates of the best schools having the best results, rather than selecting less talented, still, experienced professionals. The graduates were proud of getting an employment from HP. HP provided for accommodation to the values of the company by means of sophisticated interviews, and methodological „stimulation" of newcomers. HP came up with new and advanced products each year, even if the old products were in an emerging life cycle which could have been exploited from both market and financial aspects. Only the innovation representing excellence and outstanding engineering work could pass the exam.

At decision-making the key role was played by the engineers while marketing staff played secondary role only. The heros were those who invented something rather than those who sold it. Career tracks also reflected the technical approach: most of the heads of divisions had a technical qualification.

In 1959 David Packard summarized in a company presentation the essence of corporate philosophy of maintaining the entrepreneur spirit in the following: „Set a clearly defined goal, allow maximum freedom to those who make efforts to reach the goal, and provide for the acknowledgment of the performance at corporate level". The policy of organizational decentralization outlined in the above was also aligned to this philosophy.

The most innovative divisions got the largest stakes from development budgets. The production units of the company could get a full division status only in case they could develop some kind of new product and were able to successfully launch it at the market. Unlike a number of other companies, HP stimulated its international subsidiaries to the development of their own Research and Development abilities, to be more than just a sales and distribution point.

According to one of the most popular theories of the American business world in the seventies the companies should acquire the largest possible market stake as this is a precondition of progressing on the „learning curve", reducing the costs, that is, at the end of all, market stake is the key factor of profit maximization. This approach could not gain floor in the strategy of HP. If a product is not good enough to generate an outstanding profit margin in the very first year, then it isn't representing a significant technological edge and HP should not deal with that" – Packard explained it to the managers in 1974. The company had a clear objection against the employees who boasted about the large market stake of their product, or started to think about the opportunities for increasing the market share.

Looking for ways in the nineties

HP developed its first mini-computer in 1964 to provide automated control of mechanical instruments. In the beginning of the nineties the company had substantial interests in the computer business and the related services and supporting activities. Despite the promising start the company was not able to develop an organizational structure that is in accordance with the rapidly developing industry. As Packard explained, they made trials with divisions, various group structures, action groups, committees to improve coordination, as a result of which a complicated system of bureaucracy was established, the labyrinths of which slowed down the process of decision-making. The share prices – sensitive to the troubles – started to drop.

Hewlett and Packard – at that time the active directors of the company – got down to solve the problem themselves, and in 1992 they nominated Lew Platt to the CEO 's position. They dissolved a number of committees, and allowed more freedom to the computer units. They made efforts to reinforce the philosophy that at HP the flexibility of the small companies must be merged with the power of a unified large company.

HP has developed successfully for a couple of years under the control of Platt, The growth was constantly 20 percent above the plan. All the business lines of the highly diversified company had a good performance. During mid nineties the company advertised its Internet strategy. They stated that Internet affected all the business lines of HP therefore no separate unit would be organized to that purpose: everybody should include the Internet ideas in his own strategy, in an appropriate way, under certain central coordination. Nevertheless, around 1996 the growth started to slow down again.

The computer division of HP was reorganized in June 1997 due to its weak financial performance. The project was controlled by Rick Belluzzo a veteran, who arranged the company to five business lines:

1. Laser printers
2. ink-jet printers
3. personal systems (workstations and PC-s);
4. company systems (Unix -based);
5. software and services

During mid nineties some of the competitors – including among others Sun Microsystems – made benefit of the contradicting tendencies, and the uncertain messages sent to the market and they hold stronger foothold at the upper edge of the market.

Belluzzo assigned with the reorganization work left the company in 1998 when he went to Silicon Graphics. At that time Lew Platt was 56 and there was a custom at HP, that the key executives retire at the age of sixty. As it became clear, the company had no strategy, as the various plans were elaborated and rejected at top level from time to time. Platt started to talk openly about the failure and frustration of the management. In 1998 the revenues grow substantially – primarily due to printers, Unix servers, and desktop computers, while the costs raised rapidly, too, financial indicators worsened, the company could not reach the goals they set. Certain analysts blamed the organizational structure and cost structure of HP for the failures.

At the end of the decade Lew Platt announced his retirement simultaneously with the separation of test and measuring business of the company to be performed in the frame of an independent company (Agilent). The name of Hewlett-Packard was carried forward by the following large units:

1. Enterprise Computing Solutions;
2. Personal Systems Group ;
3. Ink-Jet Products Group ;
4. LaserJet Solutions Group.

The company received more and more criticism for separating the business lines, for the so-called silo-approach. The critics claimed that the company handles separately its particular products and services, (for example the PCs and unix-based systems), manages its investments separately, and sales is decentralized, the company is not able to utilize the large, integrated business opportunities, where computers, network solutions and printers should be delivered simultaneously. Decisions affecting more than one business unit were made very slowly, the resources allocated for development were scattered, while excellent business opportunities, that were not integrated to the silos, did not receive the necessary CAPEX.

According to certain opinions by 1999 HP actually was split into four independent companies each of which was excellent, still, they were not able to combine their power. The chairmen of the groups managed their own business, they made their own decision on their internal, organizational issues, strategy and politics. According to the opinion of Ann Livermore the head of corporate systems group the time when Internet was used as sales channel -as it is done, for example, by Amazon.com – has passed. The future is in electronic services (e-service) like, for example procurement, billing, security electronic payments, data storing and operation of data warehouses. These services will be accessible via Internet, HP is excellently positioned to become a major player in the area of e-services. According to the opinion of Livermore the e-service strategy will affect all the units of HP. She has submitted her plans to the top management of the company as the strategy of its own group which was finally approved as a general approach relevant to all the areas. The company seemed to have a corporate strategy, while lacking the necessary organizational and management structure.

In the beginning of March 1999 Lew Platt published the electronic services strategy of Hewlett-Packard employing some 60 thousand people. In the same time he announced that the company is looking for a new CEO. This year the revenues increased by 7 percent only, which growth lags far behind the results of the competitors, IBM or SUN. Profit growth also lagged behind the industrial average.

Changes under the management of Carleton Fiorina

Carleton Fiorina the new CEO came to HP from Lucent Technologies. He has been the first CEO to come from outside the company in the history of HP. He immediately got down to work with full enthusiasm on the HP advertising campaign: appeared on the front page of magazines, lots of interviews were made with him.

It did not take too long to make decisions on reorganization, either. In October 1999 he announced that HP would be re-organized to four divisions: Two of them will focus on customers (business and residential customers) while two more divisions focus on the product lines (computers and printers) of the company. All the product focused activities are controlled by two managers: One of them controls personal computers, corporate systems storing and software applications, while the other one controls laser and ink-jet printers. Customer focused campaigns are also controlled by two managers throughout the company: one of them (Ann Livermore) is responsible for e-services and intra-company (business -to-business, B2B) transactions while the other is responsible for digital imaging strategy and the consumers (business -to-consumer, B2C). Within the new system sixteen product lines are classified to two groups (computers and printers) that make „product generating background" which have to work close with the front-line units servicing the B2B and B2C segments organized in the four geographical regions (Asia, Pacific, Europe, Latin America and North America).

The staff persons working in different positions are highly dependent on each other, more than before, as a manager no longer controls the entire product value chain.

Naturally, it is not indifferent, how the cooperation is realized in the new system with the over 40 thousand value added re-sellers and retail agents of different volume, whether HP will follow the way of Dell, i.e. develop an Internet-based, direct sales channel to avoid mediators/ agents.

Number of the managers of the company have the opinion that time has come for model change at HP. In the present environment – they claim – there is a real need for corporate strategy, and the appropriate tools must be found to the implementation of this strategy. Nevertheless, the actions of Fiorina receive a lot of criticism. According to some opinions he does not involve the managers properly in the decision-making relating to key issues, eventually the managers have to find out themselves, what they are actually expected to do. The managers who got accustomed to independence are reluctant to accept such re-shaping of the authorities. While earlier the bonuses were subject to the performance of the particular units, currently the emphasis is shifted to the direction of overall performance of HP. According to some opinions HP has lost the balance in respect of bonuses: Although the sales staff of the particular business lines earlier focused only on the products of their own unit, which made the sales of complex systems difficult, sometimes the customers did not understand why they have to carry on talks simultaneously with more than one HP representative, while currently the problem is that an „integrated" sales staff cannot be an expert of all the areas.

The stock exchange had a favorable response to the first months of operation of Fiorina., HP share prices increased. The IDSC report published in April 2000 characterized HP as the „most dynamically developing PC manufacturer of the world". Then came the blow-out of the Internet stock exchange balloon in the American economy, and within that the IT sector started to show the first signs of the nearing recession. HP was not able to avoid the impact of the general trends, either, and its share prices started to drop.

The news on the merger with Compaq put the company in the focus of attention beginning of September 2001. A Business Week professional assessed in its issue dated 17 September (published before the New Year attack of terrorists) the business lines of the to companies undergoing fusion according to the following:

PERSONAL COMPUTERS. The world's largest + PC manufacturer will be created with a 19 percent market stake. (Table 1.) Nevertheless, since PC sales and profit margins hit a negative record, the two companies made a joint loss of USD 500 billion in 2001.

PRINTERS. The 50% share of HP is expected to increase, together with the turnover of printer cartridges. Nevertheless, the demand and profit margin are expected to go downhill while price competition with Lexmark, Canon and other competitors will increase.

LOW-END SERVERS. Compaq has a leading position in this area, in respect of windows-based products its cumulated market share amounts to 37 percent. Nevertheless, Dell and IBM must be considered as strong competitors.

HIGH-END SERVERS. Both HP and Compaq are backward in this key area. Compaq runs out its Alpha servers, the sales of Unix-based servers of HP are stagnating.

SERVICES. Both companies are longing for the glory of IBM, whereas 62 percent of the joint servicing staff of the two companies deal with simple machine repair work, rather than top management consulting, which is in the focus of the „Big Blue". The earlier idea relating to PricewaterhouseCoopers buy-out failed.

STORAGE. 5,2 billion -dollar storage-business of Compaq may get a new impetus if HP also starts to sell Compaq equipment to the customers. Nevertheless, it will be difficult to conquest customers from the giants of the market, like EMC.

SOFTWARE. Complex company solutions require special middleware-software applications. In this area HP lags behind its competitors, while Compaq is not on the top list.

Company	Revenues in 2000 (billion USD)	Net income in 2000 (million USD)
IBM	88,4	8.100
Sony	58,5	134
Hewlett-Packard	48,8	3.700
Toshiba	48,0	776
NEC	45,4	478,5
Fujitsu	44,2	69
Compaq	42,4	1.700
Dell	31,8	2.180
Samsung	26,7	4.700
Sun Microsystems	15,7	1.800

Table 1. The largest computer manufacturer companies in 2000

Source: company data

The journalist finds this situation dangerous as development programs targeting the upper segments of the market increasingly require investments when milking cows dry out due to slow-down of expansion and the killing price competition. In 2000 the PC business experienced a dynamic growth while - according to the analysis- in the first and second quarters of 2001 all the profit of the company came from the printer business. The market is weakening, profit zones are shifted in the direction of upper market segments, while there is hardly any chance to abandon these two businesses.

The September 17 issue of Newsweek (also published before the terrorists' attack) compares Jack Welsh the leaving head of General Electric with Carly Fiorina. If the share prices go uphill, – Allan Sloan writes – all the faults of the executive are forgiven. If they go downhill, the executive will be blamed for all this actions and steps. The secret of successes very simple: share prices must be increased. After all, it's all about money.

Compaq at the millennium

Michael Capellas the existing CEO of Compaq was assigned in July 1999 as the top manager of the company. He was primarily expected to restore the old glory of Compaq brand.

After its foundation the Huston -based company used to be the favorite of tradesmen and usersfor long years: He was the first to come up with a series of advanced computers with increasing capabilities. It even succeeded to overtake IBM: In the mid nineties it was considered as the third pillar of IT industry, along with Microsoft the software giant and the chip-manufacturer Intel. Nevertheless, by the end of the decade the star of the company started to fade, which was reflected by the profit figures as well.

The analysts mention a couple of reasons: It seams that Compaq responded too late to the Internet revolution: Compared with Dell, it insisted too long on the traditional forms of sales, that resulted in storing problems at the time of rapid changes of the demands. In the area of server purchases serious troubles were caused by the purchase and integration of Digital Equipment Corp. In 1998 the revenue growth amounted to 5% only, while in the „golden ages" the investors got used to a 45-65 percent growth. After January 1999 the share prices dropped to less than their half within five months.

Capellas got down to solve the problems energetically: he reorganized the company, decreased the costs, gave a new impetus to electronic sales and innovation. In the second quarter of 2000 – after the year of stagnation – the revenues increased by 7.5 percent. The growth was especially dramatic in Japan and Latin-America. The PC business continuously generating losses for several quarters, again turned to profitable. The share prices did not reach their earlier peak, still, they clearly started to go uphill. The competition with Dell seemed to be more promising, too.

Capellas, who earlier filled in over twenty various management positions, made efforts to implement the popular approaches of management course books. He reorganized Compaq to half dozen market units, in place of the earlier functional units. He made the managers responsible for the results achieved by their business units. He built the objectives related to customer satisfaction into the management incentive system with a 30 percent ratio at the top level. His lectures on the actual position of the company and the plans thereof can be followed by all the employees of the company (exceeding 60 thousand) either on-spot or through a closed circuit TV. „Management must be personal" – Capellas says. We must call people, and convince them that they are participants of something they must be proud of. According to certain press opinions the head of Compaq is excellent in implementing smart methods invented by others, still, he is no visionairy type.

At the end of year 2001 the clouds started to gather in the sky, again. The New York events of 11 September and an unusually strong typhoon in the far-east seriously hit the supply channels of the company. The sales in the third quarter dropped by 33 percent, compared to the same period of the previous year. In case of desktop computers a 42, while in the case of PC servers a 44 percent decrease was experienced, respectively. The losses of Compaq amounted to 499 million USD while Dell made a profit of 429 million USD at the same time. Based on this situation the analysts have strongly cut down the 2002-year forecasts relating to profit and turnover. The share prices dropped, there was a real threat that market value goes under the book value. „HP has alternatives to chose while Compaq has problems" summarized his opinion Don. M. Young, UBS Warburg expert in December 2001.

The report on the last quarter of 2001 gives reason to some optimism. The PC business is highly volume-sensitive, while end-of-the year sales have proved better than expected. The company managed to restore its supply channels, and some signs suggest that the end of American recession is nearing.

Epilogue

Starting from the end of 2001 the debates on the future of HP and its marriage with Compaq have become more intensive. On one side of the front line Carleton Fiorina is heading the team: they insist on the changes started and support the planned fusion. The leaders of the other side are the members of the Hewlett and Packard families, who – through various channels – have substantial – although no defining – business stakes in the company marked with the names of the founders, representing pride for the nation.

Today the founding members are represented only by Walter B. Hawlett, the son of Wiliam Hewlett in the Board of Directors. He announced – causing serious upheaval – that he would vote against the merger, and in this respect he may rely on the support of David and Susan Packard, and – interestingly – even Lewis Platt the earlier CEO who put Fiorina in this position. William Hewlett made all he efforts to convince the other shareholders about the truth of his position, while – naturally – the members of the other side did the same.

In Spring 2001 the public may expect the most exiting voting battle of the economy, relating to the HP-Compaq case. The last sentences of this article were put down just some hours before the starting of the shareholders' meeting called for mid March 2001. We deliberately did not wait for the result of the voting, the readers must obviously know about it. Clearly, whatever happens, the war will not be won in this battle. HP continues to be for many the symbol of creative thinking, entrepreneur attitude and human corporate culture,. Nevertheless, the world has changed dramatically around the company. Earlier, special and unique products are sold as mass goods in the supermarkets. Competition has been extended to the entire world, rigid economic rules are prevailing, there is a huge pressure to reduce the costs. It is increasingly difficult to find the means by which a company can protect itself. In the long run, the structures that simultaneously meet the requirements of engineering creativity, customer focus, economic rationality and human thinking, which are often in contradiction with each others, and which provide for stability in the shorter or longer run.

The financial benefits of the fusion of two companies are easy to calculate for the experts. It is munch more difficult to asses the culture inside the company, the shocks people suffer and its consequences. „In fact the fight is on for the soul of Hewlett Packard" – Mr. Thomas J. Perkins – a 69-year old member of the Board of Directors told last December.

Literature

1. Bőgel, Gy.: Verseny az elektronikus üzletben. Műszaki Könyvkiadó, 2000.
2. Burrows, P.: The Radical. Business Week, 19 February 2001
3. Burrows, P.: Carly's Last Stand? Business Week, 24 December 2001
4. Collins, J. – Porras, J.: Built to Last. HarperBusiness, 1997.
5. Fiorina, C.: Making the Best of a Mess. The New York Times, 29 September 1999
6. Hagel III, J. – Brown, S.: Your Next IT Strategy. Havard Business Review, 2001. Október
7. Hámor, Sz.: Ezüstnyelv, vasakarat. Népszabadság, 26 August 2000.
8. Kaplan, D.: The Silicon Boys. William Morrow and Company, 1999.
9. Mandel, M.: The Coming Internet Depression. Basic Books, 2000.
10. Packard, D.: The HP Way. HarperCollins, 1995.
11. Perkins, A. – Perkins, M.: The Internet Bubble. HarperBusiness, 1999.
12. Peters, T. – Waterman, R.: A siker nyomában. Kossuth Könyvkiadó, 1986.
13. Sloan, A.: Now Playing!!! The Celebrity CEO. Newsweek, 17 September 2001
14. The Economist: In the Family's Way. 15. December 2001
15. Williams, M.: H-P Employees Don't Share Enthusiasm for Compaq Deal. The Wall Street Journal Europe, 19 November 2001

The Realization of DVB-T

Lothar Urmann

Spinner GMBH

(x)

In quite a number of countries terrestrial digital video broad casting (DVB-T) is about to be introduced, or operation of the first stations and chains have already started. The channel allocation which had been made for analogue television has been kept up. Since the channel bandwidth depends on the respective standard in use there are various DVB bandwidths. In Europe the 8 MHz channel bandwidth is common. The modulation method which is applied generates an out-of-band spectrum which must not reach the antenna without prior attenuation. Bandpass filters with high edge steepness are required to avoid mutual influence between channels and interference with existing wireless services. The out-of-band spectrum allowable for a transmitter is defined by masks (fig. 1) standardized by ETSI.

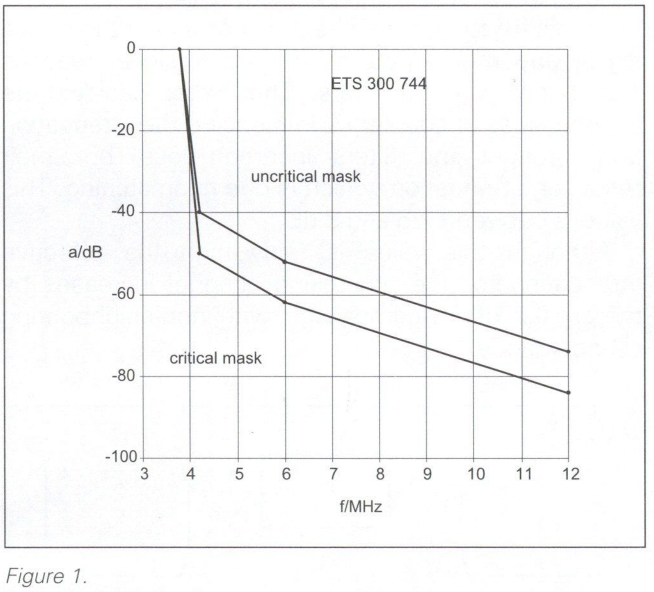


Figure 1.

A distinction is made between critical and uncritical masks. The spectral masks describe the attenuation of the out-of-band radiation relative to the total DVB channel power capacity. For measurement purposes it makes more sense to define the power spectrum of a channel for a measurement bandwidth of 4 kHz. Therefore there are representations in which the mask is shifted towards negative values by

$\Delta\alpha$ = 10*log (7600/4) = 32.8 dB
7600kHz = effective channel band width

Despite pre-equalisation measures in the transmitter the mask requirements can usually only be fulfilled if the filter pass band graph is adapted to the power spectrum of the respective transmitter.

The 7.6 MHz multiplex DVB signal is formed by about 7600 carriers, so that short-term power peaks can occur which exceed the mean level by a factor of 10. This is described by the crest factor. For the filters it means that the demands on electrical strength must be significantly higher than for analog transmission.

Spinner established already in 1946 has been known to anybody as a leading international supplier of DVB combiners (also for adjacent channel operation), filters and accessories like patch panels, switches, loads, rigid lines and connectors.

UHF DVB-T Dual Mode Waveguide Filters

The filters are designed for 5 kW of DVB power.The Dual Mode waveguide filters are completely made of invar and feature an elliptic performance graph. The patented manufacturing process and the avoidance of flanges make for the superior temperature stability. Every filter is fully tunable through at least 4 channels, with the lowest attenuation achieved in the highest respective channel. Fig. 2 shows the effective attenuation (mean attenuation in the range of fo ± 3.8 MHz) for 6 cavity filters. In addition to the 6 cavity filters for adjacent channel operation also 8-cavity filters (4 physical resonators) are available for maintaining the critical mask. If tunability is not required (channel filter) the filter capability increases to 10 kW DVB, or to 20 kW DVB power (with forced cooling).

DVB-T combiners for VHF and UHF

The DVB-T filters can be used to build 2-way or multiple combiners. We distinguish between constant impedance combiners (with directional couplers) and star point combiners. The principle of a constant impedance combiner is shown in fig. 3. The combiner

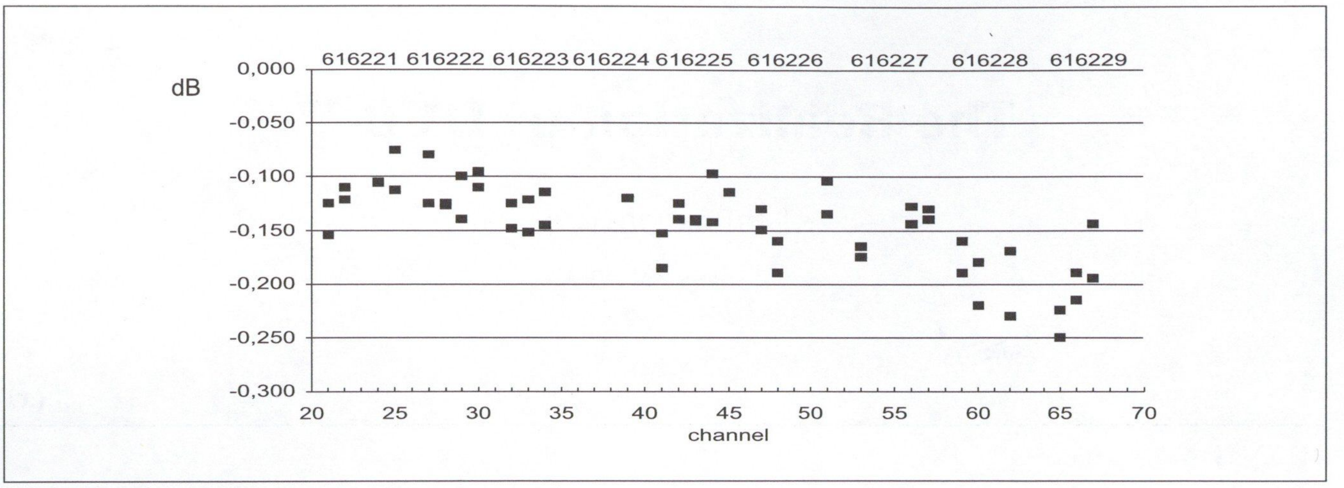


Figure 2.

consists of two DVB-T filters and two 3 dB couplers. These combiners can be adjusted to any desired channel combination by simply retuning both filters within the whole VHF resp. UHF frequency range (with coaxial filters) resp. within 4 channels (with Dual Mode filters) in the UHF range. The construction of a star point combiner is less complicated. The principle is shown in fig.4. Both channels first pass through a filter each and are then combined in the star point. The lengths L1 and L2 are designed for the respective channels. However, the channels must not be directly adjacent.

Unlike a constant impedance combiner with directional couplers a star point combiner is usually not frequency tunable, i.e. it can only be used for the channels once selected.

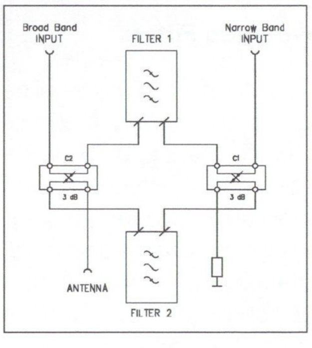


Figure 3.

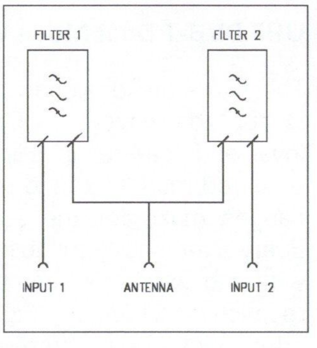


Figure 4.

DVB-T adjacent channel mode with filter diplexers (see Fig.5)

Signals arriving at the narrowband input 1 are split by the 3 dB coupler K1 so that equal amplitudes and a 90° phase difference are generated at filter inputs 2 and 3. Filters F1 and F2 are tuned to channel Km. The 3 dB coupler K2 units the signals again so that the attenuated input signal arrives at the antenna. The attenuation value is created by the insertion loss of the filter and the couplers.

Signals which arrive at the diplexer via broadband input 6 are split by the 3 dB coupler K2 so that partial signals with equal amplitudes and a 90° phase difference are generated at filter inputs 4 and 5. In the event of channels Km and Kn not being directly neighbouring channels the filters F1 and F2 reflect almost 100% of both partial signals of the broadband signal. So the input signals are available again at antenna output 7 after attenuation by twice the coupler insertion loss.

In neighbouring (adjacent) channel mode the frequency fno-3.8 is already near the transmission band of the filters F1 and F2, i.e. fmo+4.2.

At this frequency the filters do not reflect the whole signal any more. Part ot the input signal reaches absorber A via the filters. The factor besides the insertion loss of coupler K2 influencing the attenuation is no longer the filters insertion loss, but their reflection attenuation, which is due mismatching. This value is between 2.5 and 3 dB.

Although this value is fairly high the effective attenuation for the broadband channel increases by merely 0.2 dB in comparison with non-neighbouring channel mode.

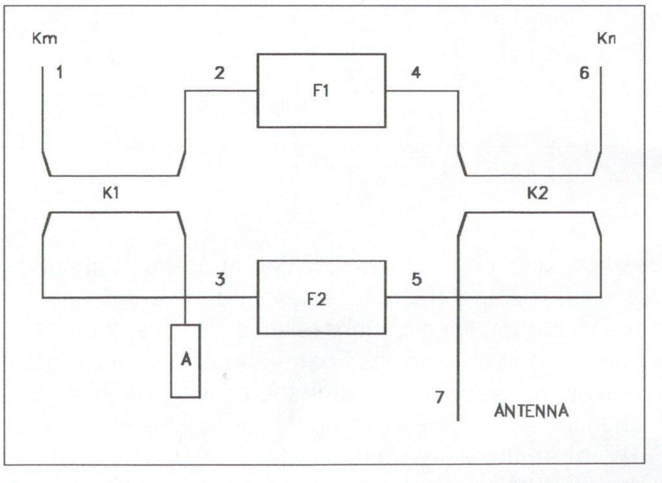
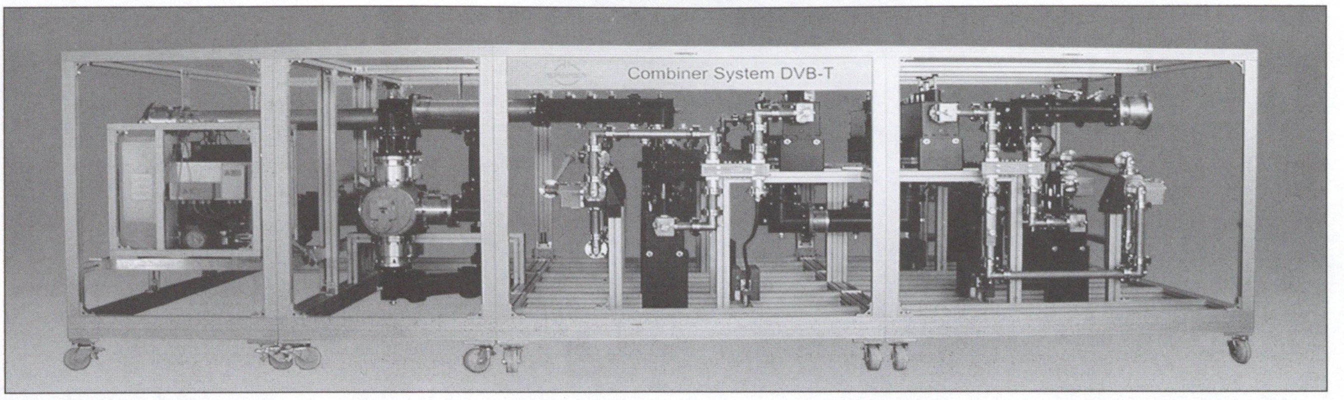


Figure 5.



DVB-T combining network Bordeaux Bouliac (France)

The most sophisticated DVB-T project was realised by Spinner for France last year.

TDF needed a combiner network for its French TV-Station in Bordeaux Bouliac in order to combine the 3 existing analogue TV transmitters with 25 kW each and 6 additional DVB-T transmitters with 2.5 kW each.

The major technical requirements were:

- combining a total of 9 transmitters and radiation from two half antennas (ch 23 through ch 62)
- the lowest possible attenuation loss for the 3 existing analogue transmitters
- combination of adjacent channels
- broadband UHF special 2-way splitter for waveguide input for 95 kW of effective input power with de-coupled 6 1/8" coaxial outputs
- remote switching between redundant systems with bridging feature to increase the system availability

For this application Spinner had developed a combiner network concept including two identical combiner paths, splitters and RF power switches, which allow regular operation as well as remote control switching to emergency operation.

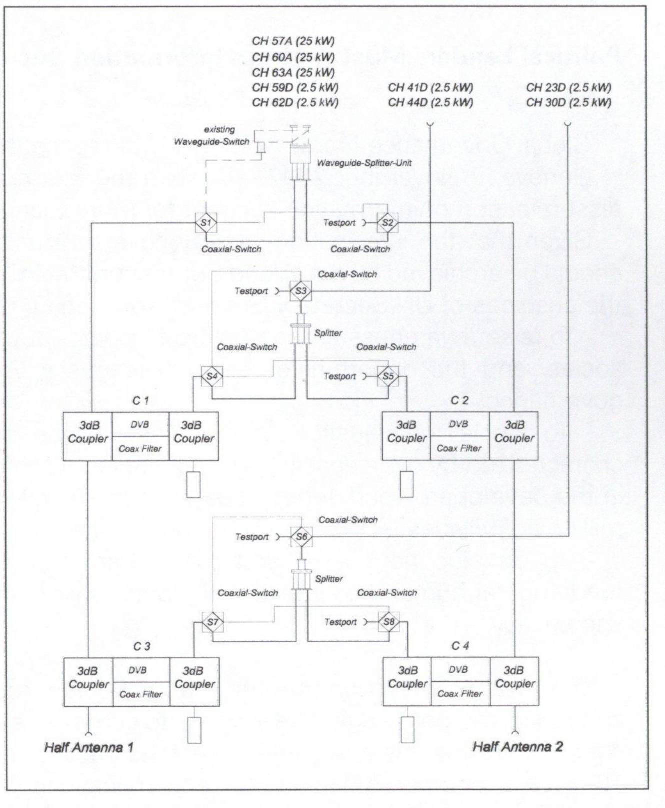
Technical details (see also block diagram):

The total analogue transmitter power of 75 kW is routed through a special waveguide to the input of a 2-way splitter, the outputs of which are de-coupled by 20 dB. The balancing load required for de-coupling is cooled by a re-cooling aggregate included in the scope of delivery.

Following that two 6 1/8" 2-way switches route the two partial signals via identical combiner paths to the half antennas.

For the purpose of combining the DVB-T channel groups the individual directional coupler combiners were equipped with specially developed coaxial DVB-T filters in order to increase the bandwidth of the narrow band input to 32 MHz and 64 MHz, respectively.

By combining channels in the upstream 2-way combiners and by feeding these channel groups into the narrow band input of the directional coupler combiners the number of the combiner steps in series was reduced to no more than 4, and the attenuation loss could be reduced significantly. This design also facilitates all installation and tuning work.



Blockdiagram DVB Combining network

In order to increase availability the two identical combiner paths were equipped with a total of 8 coaxial 2-way switches. They are used to transmit the overall signal via both half antennas in regular operation.

In emergency operation 60% of the total power can be transmitted via one combiner path and one half antenna.

A key requirement in the design of the combiner network was the demand for the same electrical length for all channels so that the total signal is radiated with identical phases.

In the acceptance test the maximum phase shift measured was 3° between the combiner paths.

(x)

News

Internet Photonics has introduced video multiplexing and conversion gateways to extend legacy cable's infrastructure for video-on-demand and other narrowband services by converting signals between legacy Digital Broadcast Video-Asynchronous Serial Interface (DVB-ASI) protocol and Gigabit Ethernet.

❖ ❖ ❖

Political Leaders Must Address Information Society Issues

Global Governance Framework for 'Cyberspace' to be issue at World Summit

Geneva, 7 November 2002 – Despite the fact that activities based around the creation, processing and dissemination of information account for more than 80 per cent of employment in the developed world.

Given that the information society covers virtually every aspect of our lives, Mr. Utsumi outlined what should be archieved by the World Summit on the Information Society, which is being organized by ITU under the auspices of UN Secretary-General, Kofi Annan. Utsumi noted the Summit is needed:

- To raise awareness among political leaders, at the highest levels, of the implications of the information society and the opportunities and challenges it will bring in the area of economics, development and governance.
- To tackle the injustice of the digital divide between developed and developing nations which is characterized by overcapacity, falling prices and profits for telecommunication operators and manufacturers in the developed world, while in parts of the developing world investment can't keep up with demand and millions of villages lack even a basic phone line.
- To develop new legal and policy frameworks, appropriate to cyberspace, in order to guarantee fundamental human and intellectual rights while addressing issues such as cyber-crime, security, taxation, and privacy.

The World Summit on the Infromation Society represents a unique opportunity to address these issues and to set the policy agenda for years to come. The unique 'two phase' structure of the Summit provides a chance to refine this policy agenda and its implementation. The first phase of the Summit will be in Geneva 10 to 12 December 2003 and the second phase in Tunisia in 2005.

In recent years, there have been many initiatives designed to tackle the digital divide number of these have been successful, we must seek new and innovative ways of mobilizing investment, by seeking a global perspective and securing justice.

❖ ❖ ❖

Sending messages to MMS-capable mobile phones is currently possible within the D2 and D1 networks in Germany. Furthermore, dispatch to mobile phones that are compatible with Nokia Smart Messaging is provided for in all four German networks. The partnership between the Californian-based company FunMail and conVISUAL, which started at the beginning of the year, has begun to bear fruit. FunMail's connections to the American film industry, its rights to use known characters such as Garfield and South Park, as well as its own designer pool, have assisted the company in successfully establishing its services worldwide. The messaging expert conVISUAL is FunMail's exclusive partner for promoting its services in Europe.

Tartalom

Editorial Office
Scientific Association for Infocommunications
H-1055 Budapest, V., Kossuth Lajos tér 6–8.
Phone: 00 36 353 1027 • Fax: 00 36 353 0451
e-mail: hte@mtesz.hu

Advertisement rates
1/1 (205 x 290 mm) 4C 120 000 Ft + áfa
Borító 3 (205 x 290 mm) 4C 180 000 Ft + áfa
Borító 4 (205 x 290 mm) 4C 240 000 Ft + áfa

Subscription rates for 2002
12 issues 150 USD, single copies 15 USD

Articles can be sent also to the following address
BME Department of Broadband Infocommunication System
H-1111 Budapest, XI., Goldmann Gy. tér 3.
Phone: 00 36 463 1559 • Fax: 00 36 463 3289
e-mail: zombory@mht.bme.hu

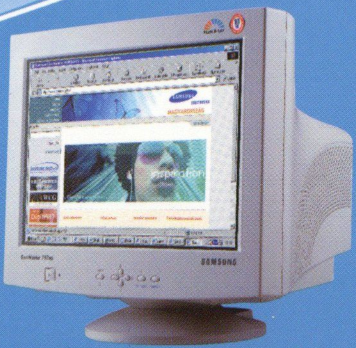
Subscription
Scientific Association for Infocommunications
H-1055 Budapest, V., Kossuth Lajos tér 6–8.
Phone: 00 36 353 1027 • Fax: 00 36 353 0451
e-mail: hte@mtesz.hu

www.hte.hu

Publisher: MÁRIA MÁTÉ
Manager: András Dankó

Design by: Kocsis és Szabó Kft.
HU ISSN 0018-2028

Printed by: Regiszter Kft.

Szövegszerkesztés →  ← Internet

↑

Film

Varázslatos változatok.

A SyncMaster MB monitorok MagicBright technológiájukkal gombnyomásra változtatnak a munkastílusodon. Szövegszerkesztés? Internet? Film? A MágikusRagyogás varázsütésre biztosítja a felhasználás módjához leginkább alkalmas fényerőt, kímélve ezzel szemeidet és kihozva a maximumot monitorodból. Három típus, egy árért? Kár lenne kihagyni.

SAMSUNG DIGITall
everyone's invited™
www.samsung.hu