Journal of

Theoretical and Applied Information Technology

Journal of Theoretical and Applied Information Technology

www.jatit.org

Efficient Distance Computation Between Natural Quadrics

[1] Mustafa S. Fawad , [2] Mustafa Atika
[1] Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences, China.
[2] National University of Computer & Emerging Sciences, Pakistan.

Abstract

The idea of our work is on one of the basic problems, namely the fast calculation of the distance between two objects. We have implemented an enhanced version of the shortest distance routine on the basis of mathematical equations, which allows the tracking of the distance between a pair of conics and quadrics in time that to be bounded by a constant is expected. Experimental results confirm this result, at least for moderate shortest distance calculation, and suggest that the enhanced algorithm might also come in handy. Computing the distance between two objects is a common problem, using the mathematical model of two objects, we find the point on each object such that the distance between the objects is minimized. To do this we took care of efficient computing, which is possible with simplifying the polynomials and their relative coefficients. Effectiveness of computing algorithms and executing time are also analyzed. We tested the calculation on various natural quadrics. The graphics part is implemented in **openGL** and **QT** library is used for user interface. The execution time is calculated by running the distance calculation routines 10,000,000 times.

Keywords Polynomial degree, equations, Lagrange formulation, real roots

INTRODUCTION

During last decade an increasing interest in virtual reality (VR) technique could be observed. Virtual reality plays a highly important part in the actual reality as can be send from movies and other scientific application (i.e. Education), and VR based industries.

However, the 'classical' challenges that come when deal with the complex objects in VR world, simulating their physical behavior.

Robustness and efficiency of such a simulation heavily depend on the quality of the collision Detection precisely distance computation algorithm. They identify the earliest time of the closest distance of two objects in a given time interval. However, one has to consider that there is a trade-off between the accuracy of the approximation and the efficiency of the distance computation. In this paper we present an efficient algorithm for computing the distance between so-called 'quadratic complexes'. We reduce the distance calculation problem to the problem of solving univariate polynomial equations.

We implemented on OpenGL and QT to check the practical approach to this idea and succeeded in having productive results. Some theoretical discussion can also be read in the paper by Christian Lennerz et al. on distance computation Reducing the high degree polynomials was indeed a challenging task, keeping in mind the approximation not affecting the quality of results. Our considerations therefore raise the question, whether or not it is worth to handle curved objects directly. It means that one has to spend more effort on the basic routines computing the distance between two patches, but can profit from higher accuracy and simplified quadratic equations to compute the closest.

Computing the distance between two objects is a common problem, using the mathematical model of two objects, we find the point on each object such that the distance between the objects is minimized. To do this we took care of efficient computing, which is possible with simplifying the polynomials and their relative coefficients. Details of computing algorithms and executing time are discussed in the following section.

Corresponding Author: Mustafa S. Fawad, Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences P.O.Box 8718 Beijing, 100080, China.
TEL:(8610)62522028(Off.), (8610)88255778 (Res.) FAX:(8610)62562533.
E-mail Address: fawad_zaidi@yahoo.com

OUR ALGORITHM

Simple Cases:

We started with the simpler cases and proceeded to gradually more complex situations.

Simpler cases are dealt with ease since there polynomial equations are simpler too. We adopted the methods for simpler cases by D. H. Eberly et al. The cases are as follows:

1. Point to Line
2. Point to Circle
3. Point to Ellipsoid
4. Point to Ellipse
5. Line to Circle
6. Line to Line

A slightly different case is of Line to Sphere, which is not as hard as is thought. It is just the extended case of line to point.

Complex Cases:

In this section we will discuss the cases of distance of different objects with a cone. Here we have assumed that the cone is centered at origin along Z -axis. Given the general equation of this cone below:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$$

We know from the general equation of conics:

$$X^T AX = 0$$

where,

$$A=\begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix}=\begin{bmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & \frac{1}{c^2} \end{bmatrix}$$

by using Lagrange formulation:

$$L=(x-p)^2+\alpha(X^T AX)$$

By minimizing the function and simplifying we get:

$$f(\alpha)=A_1 p_1^{\,2} d_2^{\,2} d_3^{\,2}+A_2 p_2^{\,2} d_1^{\,2} d_3^{\,2}+A_3 p_3^{\,2} d_1^{\,2} d_2^{\,2}=0$$

-Eq1

where $d_i = 1+\alpha A_i$ and *i=1,2 and 3*

Point to cone:

Eq (1) is the degree 4 polynomial in α for the case of distance between cone and point.

The nearest point on cone can be calculated by:

$$x=(\alpha A+I)p \quad \text{-------- Eq (2)}$$

where *I* is the identity matrix and *p* is the point.

Now considering the special case of cone, i.e circular cone:
From Eq (2) we have:

$$f(\alpha)=b^2(Bp_1^2 d_3^2+Bp_2^2 d_3^2+A_3 p_3^2 b^2)=0$$

where $A_1 = A_2 = B$ and $b=d_i$ for *i*=1 and 2, which is the degree 2 polynomial in *α*.

Line to Cone:

Line is represented as:

$$p(t)=c+tu$$

where *c* is some point on the line and *u* is the direction vector.
From Lagrange formulation, by minimizing and simplifying the function we get:

$$g(\alpha,t)=p_1 p_1' d_2 d_3 A_1+p_2 p_2' d_1 d_3 A_2+p_3 p_3' d_1 d_2 A_3=0$$

$$f(\alpha,t)=A_1 p_1^{\,2} d_2^{\,2} d_3^{\,2}+A_2 p_2^{\,2} d_1^{\,2} d_3^{\,2}+A_3 p_3^{\,2} d_1^{\,2} d_2^{\,2}=0$$

Considering special case of cone i.e. circular cone, we get:

$$g(\alpha,t)=p_1 p_1' b d_3 B+p_2 p_2' b d_3 B+p_3 p_3' b^2 A_3=0$$

$$f(\alpha,t)=b^2(Bp_1^{\,2} d_3^{\,2}+Bp_2^{\,2} d_3^{\,2}+A_3 p_3^{\,2} b^2)=0$$

With this we can easily express that:
deg(*f*, t) = 2
deg(*f*, α) = 4

deg(g, t) = 1
deg(g, α) = 2

Circle to Cone:

Circle is represented in parametric form:

$$P(t) = c + r(cos(\phi)(u) + sin(\phi)(v)) \text{ - - -- - - } Eq\ 3$$

where:

$$cos(\phi) = \frac{1-t^2}{1+t^2} \text{ and } sin(\phi) = \frac{2t}{1-t^2} \text{ and } t = [0,1]$$

$$g(\alpha,t) = b(p_1 p_1' d_3 B + p_2 p_2' d_3 B + p_3 p_3' b A_3) = 0$$
$$f(\alpha,t) = b^2(Bp_1^2 d_3^2 + Bp_2^2 d_3^2 + A_3 p_3^2 b^2) = 0$$

$$g'(\alpha,t) = (p_1 p_1' d_3 B + p_2 p_2' d_3 B + p_3 p_3' b A_3) = 0$$
$$f'(\alpha,t) = (Bp_1^2 d_3^2 + Bp_2^2 d_3^2 + A_3 p_3^2 b^2) = 0$$

where deg $(g', \alpha) = 1$

Solving *g* for *α*

$$\alpha = \frac{-[B(p_1 p_1' + p_2\ p_3') + A_3 p_3 p_3']}{A_3 B p^T p'} \text{ - - -- - - } Eq\ 4$$

Also:

$d_3 = 1 + \alpha A_3$

$$d_3 = \frac{(B - A_3) p_3 p p_3'}{Bp^T p'}$$

$$b = 1 + \alpha B$$
$$b = \frac{(A_3 - B)(p_1 p_1' + p_2 p_2'')}{A_3 p^T p''}$$

Substituting *b* and *d3* in *f'* we get:

$$p_3^2[A_3 p_3'^2 (p_1^2 + p_2^2) + B(p_1 p_1'^2 + p_2 p_2'^2)] = 0$$

We can easily express it as:

$$p_3^2 = 0$$

$$h(t) = A_3 p_3'^2 (p_1^2 + p_2^2) + B(p_1 p_1'^2 + p_2 p_2'^2) = 0$$

where *h(t)* is a polynomial of degree 4 in *t*.

Substituting *t* in *Eq. 2* gives point on a circle. Substituting *t* in *Eq. 3* gives the value of α, which can be used in *Eq. 1* to calculate the point on cone.

Solving the polynomials:

Routines from Silvia (Simulation Library for Virtual Reality and Interactive Applications) libraries are used to solve the polynomials.

The degree four or less polynomials were solved using the routines:
SolveQuadric (coeffs, roots)
SolveQuartic (coeffs, roots)

The routine returns the number of real roots of the polynomial whose coefficients are given in "coeffs" and the real roots are in "roots". For polynomials of degree higher than 4, the following routine is used:

real_roots1(coeffs, degree, roots)

where:
coeffs = coefficients of the polynomials
degree = degree of polynomial
roots = roots calculated by the routine.

The routine returns the number of real roots of the polynomial.

RESULTS

Execution Times:

The execution time (see Table 1) is calculated by running the distance calculation routines 10,000,000 times to have substantial reading and avoid micro unit measurements. The time depends on the following factors:

- Degree of polynomial
- Complexity of coefficient calculation

Table 1: Distance computation between quadrics per 10,000,000 times

S#	Objects	Time (seconds)
1	Point – Line	25
2	Point – Circle	28
3	Point – Ellipsoid	800
4	Point - Ellipse	62
5	Point – Sphere	14
6	Point – Circular Cone	117
7	Point – Cylinder	205

8	**Point – Cone**	**126**
9	**Line – Circle**	**415**
10	**Line - Sphere**	**37**
11	**Line – Circular Cone**	**119**
12	**Line – Cone**	**221**
13	**Circle - Circle**	**1764**
14	**Circle - Sphere**	**42**
15	**Circle – Circular Cone**	**1330**
16	**Circular Cone- Circular Cone**	**472**

CONCLUSION

The degrees of the resulting polynomials are usually high and hence elimination methods are often not appropriate in the context of floating-point computations. However, we could show that for all cases in which the degrees of the resulting polynomials become critical, a factorization into two lower degree polynomials can be found. This idea has proposed a way to speed up the distance computation in a dramatic way with high quality result as seen in the table above

ACKNOWLEDGMENT

The Authors would like to thank Elmar Schömer, Johannes Gutenberg Universität Mainz, Germany, for his valuable guidance and support throughout our work.

REFERENCES

Book

D. H. Eberly (2001). 3D game engine design: a practical approach to real-time computer graphics. Morgan Kaufmann, 2001.

Conference Proceedings

Christian Lennerz, Elmar Schömer. (2002) Efficient Distance Computation for Quadratic Curves and Surfaces. Geometric Modeling and Processing — Theory and Applications

Securing Multicast Groups Transactions on Ad-Hoc Networks Using Mobile Hosts

*K.Sureshkumar, ** K.S.Ravichandran
*Senior Lecturer, TIFAC-CORE, **Professor, School of Computing,
SASTRA, Deemed University, Thanjavur-613402, TN
*suresh@core.sastra.edu, **ravigirish@hotmail.com

ABSTRACT

Computing technology with its increased access has much interest in the mobile ad hoc network technology. However, the transactions strongly depend on the availability of security provisions, among other factors. In the open network environment practically any node can maliciously or selfishly disrupt and deny communication of other nodes. This paper, propose a dynamic multicast group management protocols that aims at solving problems that are specific to ad hoc networks such as mobility, unreliable links, and cost of multihop communication. The main idea is to have group members actively participate to the security of the multicast group, therefore reducing the communication and computation load on the source

Key words: Multicast, secure protocol, service mechanism, certificate

1. INTRODUCTION

The increasing diffusion of one-to-many and many-to-many network services such as e-banking solutions, stock market applications, news distribution, and video conferencing, have gained increasing popularity. They have lead to the design and implementation of the multicast primitive that provide efficient delivery of data from a source to a group of recipients [1, 2]. The communication in mobile ad hoc networks comprises two phases, the route discovery and the data transmission. In an adverse environment, both phases are vulnerable to a variety of attacks. First, adversaries can disrupt the route discovery by impersonating the destination, by responding with stale or corrupted routing information, or by disseminating forged control traffic. This way, attackers can obstruct the propagation of legitimate route control traffic and adversely influence the topological knowledge of benign nodes. However, adversaries can also disrupt the data transmission phase and, thus, incur significant data loss by tampering with, fraudulently redirecting, or even dropping data traffic or injecting forged data packets.

To provide comprehensive security, both phases of mobile ad hoc network communication must be safeguarded. The secure routing protocols, which ensure the correctness of the discovered topology information, cannot by themselves ensure the secure and undisrupted delivery of transmitted data. Since adversaries could abide with the route discovery and be placed on utilized routes. But then, they could tamper with the in-transit data in an arbitrary manner and degrade the network operation. One important research problem for secure dissemination and sharing of information over wireless multihop ad hoc networks is how to restrict the information access to the group of authorized nodes. The data information has to be encrypted and only authorized users should be able to decrypt it.

The security of the group has to be maintained when new members join/leave or when a node is revoked. The problem can be defined as follows: given one source multicasting a stream of data and multiple receivers that join and leave the multicast session, the goal is to design a low bandwidth/delay protocol that allows authorized nodes and only authorized nodes to access the data stream multicast by the source node. The underlying communication network is a multihop wireless ad hoc network with mobile nodes. Therefore, the secure multicast group management protocol has to take into account unreliable links, nodes mobility and limited communication and computation power of the nodes. Upper layer mechanisms, such as reliable transport protocols, or mechanisms currently assumed by the mobile network routing protocols, such as reliable data link or acknowledged routing, cannot cope with malicious disruptions of the data

transmission. In fact, the communicating nodes may be easily deceived for relatively long periods of time, thinking that the data flow is undisrupted, while no actual communication takes place.

2. RELATED WORK

Previous research in the area of secure multicast has mainly focused on wired networks and various techniques were proposed considering various constraints [3]. The main limitations of these algorithms are that they were not designed for multihop wireless ad hoc networks. The most known technique is the construction of a logical key tree where group members are associated with leaves and each member is given all the keys from his leave to the root. The root key is the group key. This approach allows reducing the communication cost for key update, on the event of group membership change, to O (log M) where M is the number of group members. Various extensions were proposed to deal with reliability [4]. Extensions to wireless networks were first discussed in [5] and several secure multicast protocols were proposed [6]. These protocols addressed both issues related to mobility and unreliability. However, these protocols have mainly focused on single hop wireless networks where base stations or satellite beams cover large areas. Very recently secure multicast in multihop wireless ad hoc networks was investigated in [7].

2.1 OVERVIEW OF PROTOCOLS

One way to counter security attacks would be to cryptographically protect and authenticate all control and data traffic. But to accomplish this, nodes would have to have the means to establish the necessary trust relationships with each and every peer they are transiently associated with, including nodes that just forward their data. Even if this were feasible, such cryptographic protection cannot be effective against denial of service attacks, with adversaries simply discarding data packets.

To secure the data transmission phase, design a **secure transmission protocol (STP).** The goal of STP is to ensure secure data forwarding, after the discovery of routes between the source and the destination has been already performed. In other words, STP assumes that there is a protocol that discovers routes in the ad hoc network, although such discovered routes may not be free of malicious nodes. Then, STP is to ensure routing over such routes, despite of the presence of such adversaries. In addition to STP, this paper proposes the **secure single path (SSP*)*** protocol, an end-to-end secure data forwarding protocol that utilizes a single route. Unlike STP, SSP does not incur multi-path transmission overhead. Thus, it does not require that the underlying routing protocol discover multiple routes either. As a result, SSP imposes less routing overhead per discovery than STP. Overall, we examine SSP and compare it to STP as an alternative, lower cost, more flexible protocol to secure the data-forwarding phase.

The results show that STP outperforms SSP consistently over a wide range of experiments. The advantages of STP over SSP become more pronounced in highly adverse environments: STP delivers up to 22% more data packets than SSP, and achieves up to 94% lower delays than SSP. It is also very interesting that STP imposes up to 68% less routing overhead than SSP, although overhead was expected to be lower for SSP. In contrast, SSP provides only up to 48% lower transmission overhead than STP. We especially emphasize the low-delay characteristic of STP, as we believe that one of the main applications of STP is in support of QoS for real-time traffic.

An illustrative example of a single message transmission is shown in Figure. 1. The sender disperses the encoded message into four packets, so that any three out of the four packets are sufficient for successful reconstruction of the original message. The four packets are routed over four disjoint paths and two of them arrive intact at the receiver. The remaining two packets are compromised by malicious nodes lying on the corresponding paths; for example, one packet is dropped, and one (dashed arrow) is modified.

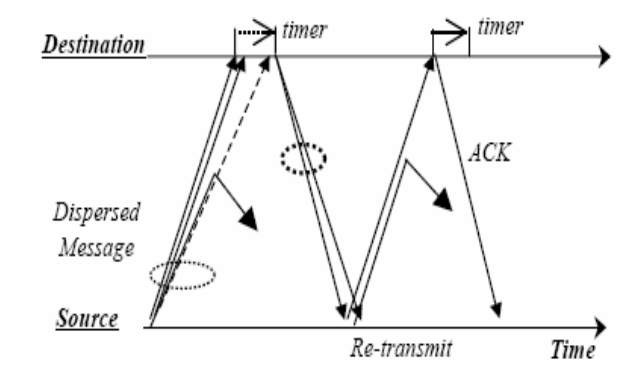


Figure.1 Example STP protocol

3. APPLICATION AND NETWORK ARCHITECTURE

The multicast group security is implemented within the secure group sessions layer. Each session has a manager that is responsible for authenticating joining members, checking service right certificates, maintaining information about attached group members, forwarding the multicast data and processing revocation requests.

3.1 PROTOCOL DESCRIPTION

In this Section we describe the join and leave process. The revocation mechanism is described in Section 4.2. The group join has three main steps: broadcast group join request, process group join replies, mutually authenticate and establish a link key with an upstream node that is already within the group. The remaining steps are for optimization and tree maintenance purpose.

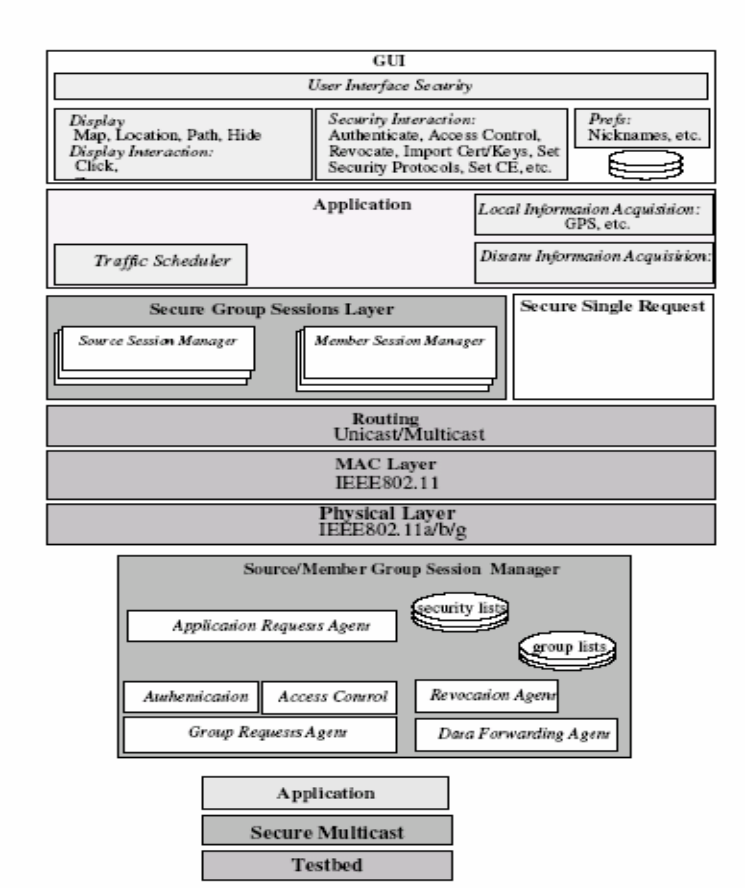


Figure 2. Application and Multicast Protocol Architecture

3.1.1. JOINING STEPS

Broadcast Group Join Request:

The parameters of this request are group_id, and TTL (Time to live). Depending on at which protocol layer this request is processed, the group_id could be a tuple uniquely describing the source, and the information service the user intends to subscribe to, or it could be an IP multicast address and port number. The TTL parameters can be first set to a small value and then exponentially increased until when a satisfactory number of replies are received. Each request has a sequence number (SN) to avoid multiple forwarding of the same request.

Receive Group Join Replies:

Nodes that are already receiving this information service send replies to the requester. Unless if the number of connected nodes to them exceeds some threshold, or if this requesting node is on their path to the source. The replies contain information about: number of hops to the source, logical path to the source (the sequence of group members that lead to the source will be used in a handover to avoid loops), path quality to the source, and number of nodes already connected to this node. Upon receipt of at least one satisfactory reply, the requesting node will initiate a registration with the sender of the most satisfactory reply. A simple criterion to accept a reply as satisfactory is if the aggregate quality of the path from the joining node to the intermediate node and path from the intermediate node to the source exceeds some threshold, and if the number of already connected nodes is below some threshold. In our basic protocol implementation the criterion is the number of hops to the intermediate node.

Authentication, registration, and key establishment:

The requester and intermediate node will first *mutually* authenticate each other. The authentication process will lead to the establishment of a shared key. Then, they will both check that they are allowed to access this information. This proof of access right is done using a service-right certificate. The authentication is complete only when the certificates are verified not to be revoked.

Tree optimization:

Once registered the newly joining node can send a path optimization message to nodes that are already in the tree but could optimize their path by attaching to the joining node. The joining node

knows the nodes that can benefit from the optimization by using the information gathered from the request replies.

Receive encrypted data:

After authentication and verification of the revocation information, the joining node can start receiving the encrypted data. The data encrypted using the secret key established during the authentication phase.

Receive tree update events:

Three types of tree update events can be received: Group Join Requests, Handover Requests, and Path Quality Drop. The Group Join Requests will be processed as explained above. The handover request will initiate a search for a better intermediate node. This search process can be done in a proactive way. When a node receives a handover request it will delay replying to incoming joins requests. The drop in path quality will also initiate a search for a better intermediary node.

3.1.2. LEAVING STEPS

Inform downstream nodes:
The depending downstream nodes should initiate a handover and send a handover complete message once reconnected to the tree.

Inform upstream node:

Once all downstream nodes are reconnected or after a *timeout*, the leaving node requests its upstream node to disconnect him from the tree.

4. SECURITY SERVICES AND MECHANISMS

In this section we describe the security services and mechanisms: authentication to join the group, revocation mechanism, and integrity protection of multicast data. We first start by describing the service right certificates that allow joining nodes to prove their right to become part of the group. Then we discuss possibilities for the revocation mechanism and detail the revocation protocol when initiated by the multicast group source.

4.1 Service right certificates

In a MANET setting where information is disseminated over the network, it is necessary to protect the access to services in a distributed manner. In the case of a multicast session, joining nodes have to prove that they are authorized to access the multicast session before being accepted in the group. The access control mechanism is distributed among the group members. A "service right certificate" allows a user/node to prove that it is authorized to access some service. Informally such a certificate is a message describing the service and signed by an authorized entity. The format of such certificates is as follows:

[DataId | Issuer | TypeOfService | ValidityPeriod|RevocationSequenceNumber | UserPublicKey | Signature]

DataID – Uniquely identify the data/session to accessed, Issuer – Entity that is signing this certificate, TypeOfService –Type of operations/requests that are allowed for this user, Validity Period – Time interval where the certificate is valid, RevocationSequenceNumber – Latest sequence number of the receiver, UserPublicKey – Public key of the user, Signature – All the above information is signed by the issuer.

4.2. Revocation Process

Several revocation models can be investigated. The revocation can be initiated by the service certifier (e.g., information source or CA), or by a third party (e.g., one or k other nodes that could belong to a revocation hierarchy). A revocation hierarchy can be used to provide rights for third parties to revoke certificates. Denial of service is an important problem when dealing with revocation. If revocation information cannot be reliably delivered to the node, because of an adversary, then legitimate nodes will not be able to verify the service right certificates, which would prevent the multicast group from correctly functioning. This might lead to forcing joining nodes to directly communicate and attach to the source. The integration of revocation information with data transmission and the requirement of reliability of delivery force the attacker to carry a full communication denial of service attack to prevent the revocation process.

4.2.1. The revocation process from the source is as follows:

1. Periodically multicast a certificate revocation list (CRL). The list is reliably multicast by requiring all downstream nodes to acknowledge its receipt. The list format is:

[MinSN| CurrentSN | ListOfRevokedCert | Timestamp | Signature]

ListOfRevokedCert contains all the certificates that were revoked since MinSN. Therefore all certificates with a SN in **[MinSN, CurrentSN]** interval and not in the revocation list are valid. A certificate with a SN lower than MinSN will have to be verified by the source unless if the intermediate node cached previous revocation lists. In order to make the revocation list more compact only the hashes of the certificates have to be included in the list.

2. Whenever, a certificate is revoked by the source, the CurrentSN is incremented and a new revocation list is issued. The source can decide to increment the MinSN if the revocation list is getting too long in order to maintain a fixed size revocation message. The source then sends the new CRL to its downstream nodes and delays sending the data to them until it gets an acknowledgement for the CRL.

5. PERFORMANCE ANALYSIS

The rationale behind the design of our protocol was to take into account mobile ad hoc network constraints to increase reliability and reduce communication cost and delay. The join/leave/revocation operations only require local updates (i.e., with upstream/downstream nodes). Therefore their cost is bounded by the maximum degree of the physical tree. Since all nodes can serve as intermediaries the degree can be kept very small while in the basic scheme the number of intermediary group key controllers is limited and therefore their degree has to be high. In comparison with the key graphs approach at each group change a multicast packet of size O (log group-size) has to be sent to all nodes while in our approach only local information has to be exchanged.

Our experiments verify that the proposed protocol can, indeed, successfully cope with a high number of adversaries, while operating only in an end-to-end manner. STP can deliver successfully more than twice the number of packets delivered by a protocol that secures only the route discovery phase but not the data-forwarding phase. Moreover, we find that STP is successful in delivering data with low end-to-end delay, low routing overhead, and limited transmission overhead, when compared to SSP. The Secure Single Path (SSP) protocol is the limiting case of STP without the dispersion of outgoing messages and the use of a single path for each message transmission. SSP is equipped with the same end-to-end feedback and the fault detection mechanisms as STP. SSP also re-transmits each failed message Retry max times, provides data integrity, authenticity, and replay protection as STP does, and selects the shortest path in hops. SSP determines, utilizes, and maintains a single path only. Once the utilized path is
deemed failed, a new route discovery may be needed in order to determine a new route.

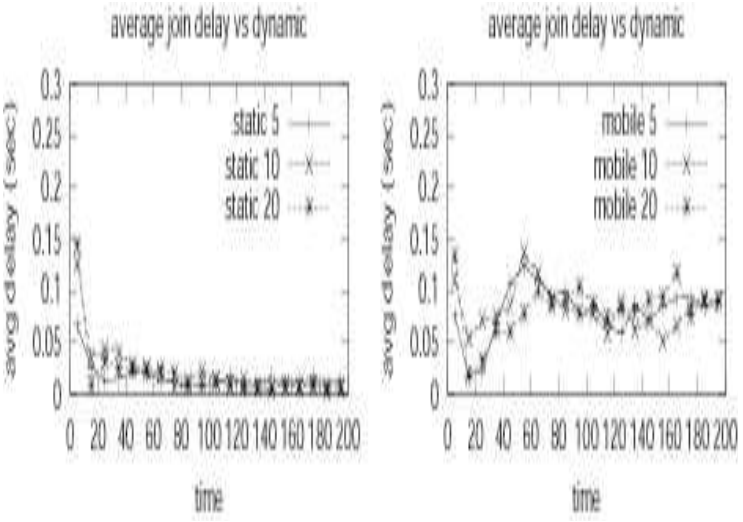


Figure 3. Average join delay for a static and mobile network with potentially 40 members average join/leave duration of 5/3, 10/6and 20/12 seconds.

6. CONCLUSION

In this paper, we have presented the secure transmission protocol to secure the data forwarding operation for mobile ad hoc network routing protocols. Our protocol takes advantage of topological and transmission redundancies and utilizes feedback, exchanged only between the two communicating end-nodes. This way, STP remains effective even under highly adverse conditions. Moreover, features such as low-cost encoding and validation mechanisms, and partial retransmissions render the scheme efficient. Nodes attach to the best closest neighbor already in the group therefore reducing the cost of join requests broadcast and reducing the communication and computation cost incurred by the source. Moreover, using shorter paths for key update increases the reliability of the secure multicast.

7. REFERENCE

1. S.E. Deering, Multicast routing in internetworks and extended LANs, in: Proc. ACM SIGCOMM'88 (1988) pp. 55–77.

2. M. Macedonia and D. Brutzman, MBONE: The Multicast BackbONE, http://www.cs.ucl.ac.uk/mice/mbone_review.html

3. Guevara Noubir and L.v. Allmen. "Security Issues in Internet Protocols over Satellite Links" in Proceedings of IEEE Vehicular Technology Conference (VTC'99 Fall).1999. Amsterdam, Holland.

4. Y. Yang, X. Li, and S. Lam. "Reliable Group Rekeying: Design and Performance Analysis" in Proceedings of ACM SIGCOMM. 2001. San Diego, CA, USA.

5. L. Gong and N. Sacham, "Multicast Security and its Extension to a mobile Environment" Wireless Networks, 1995. **1**(3): p. 281-295.

6. Danilo Bruschi and E. Rosti, "Secure Multicast in Wireless Networks of Mobile Hosts: Protocols and Issues". Mobile Networks and Applications, 2002. **7**: p. 503-511.

7. Loukas Lazos and R. Poovendran. "Energy-Aware Secure Multicast Communication in Ad-hoc Networks Using Geographic Location Information". In Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing. 2003. Hong Kong, China.

8. C.K. Wong, M. Gouda, and S. Lam. "Secure group communications using key graphs" in Proceedings of ACM SIGCOMM. 1998.

9. Guevara Noubir and L.v. Allmen. "Security Issues in Internet Protocols over Satellite Links" in Proceedings of IEEE Vehicular Technology Conference (VTC'99 Fall). 1999. Amsterdam, Holland.

10. Ran Canetti, et al. "Multicast Security: A Taxonomy and Some Efficient Constructions" in Proceedings of INFOCOMM. 1999: IEEE Press.

11. Y. Yang, X. Li, and S. Lam. "Reliable Group Rekeying: Design and Performance Analysis" in Proceedings of ACM SIGCOMM. 2001. San Diego, CA, USA.

12. P. Papadimitratos, Z.J. Haas, and P. Samar, "The Secure Routing Protocol (SRP) for Ad Hoc Networks," Internet Draft, draft-papadimitratos-secure-routing-protocol-00.txt, Dec. 2002.

13. A. Tsirigos and Z.J. Haas, "Multipath Routing in the Presence of Frequent Topological Changes," IEEE Comm. Magazine, pp. 132-138, Nov. 2001.

14. D. Maughan, M. Schertler, M. Schneider, and J. Turner, "Internet Security Association and Key Management Protocol," IETF RFC 2408, Nov. 1998.

15. M. G. Zapata and N. Asokan, "Securing Ad hoc Routing Protocols," in proceedings of the ACM WiSe 2002, Atlanta GA, Sept. 2002.

16. C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multihop Network," in proceedings of the Third ACM Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc 2002), Lausanne, Switzerland, Jun. 2002.

17. R. Canetti, J. Garay, G. Itkis, D. Minciaccio, M. Naor and B. Pinkas, Multicast security: A taxonomy and efficient reconstructions, in: Proc. IEEE INFOCOM'99 (1999).

18. A. Fiat and M. Naor, Broadcast encryption, in: Advances in Cryptology: CRYPTO'93, Lecture Notes in Computer Science, Vol. 773 (1993) pp. 480–491.

19. P. Kruus, A survey of multicast security issues and architectures, in: Proc. 21st National Information Systems Security Conference (1998).

20. M. Steiner, G. Tsudik and M. Waidner, Cliques: A protocol suite for key agreement in dynamic groups, Research Report RZ 2984 (#93030), IBM Zürich Research Lab (1997).

21. D.M. Wallner, E.J. Harder and R.C. Agee, Key management for multicast: Issues and architectures, RFC 2627 (June 1999).

Journal of Theoretical and Applied Information Technology

www.jatit.org

Priority of Active Rules and Termination Analysis

L. Baba-hamed and H. Belbachir
Laboratory: Signals, Systems and Data, USTO, B.P. 1505 El Menouar, Oran, Algeria
{lbabahamed, h_belbach}@yahoo.fr

Abstract

An active database system is a conventional database system extended with a facility for managing active rules (or triggers). Active rules may interact in complex and sometimes unpredictable ways, thus possibly yielding infinite rule executions by triggering each other indefinitely. In this paper, we propose a method of termination analysis of active rules based on Petri nets (PN) and give an object oriented representation to implement it. This approach is better than the previous ones because it takes into a count composite events and the rule priority on the one hand, and both rule representation and rule analysis are performed in the same PN on the other hand.

Key words: ECA rule, termination, Petri net, priority, path, class.

1. Introduction

Active database systems (Collet, 1996) (Widom and Ceri, 1996) (Paton and Diaz, 1999) aim at the representation of more real-world semantics in the database by supporting event-condition-action rules (ECA-rules). ECA-rules can be interpreted as "when the specified event occurs and the condition holds, execute the action". An event indicates the point in time when some sort of reaction is required from the DBMS. For primitive events, this point in time can be specified by an occurrence in the database, by an occurrence in the DBMS, or by an occurrence in the database environment. For composite events (Chakravarty and Mishra, 1994) (Gatziu and Dittrich, 1994) (Gehani et al., 1992) (Collet and Coupaye, 1996), the point in time is defined on the basis of other points in time which represent other primitive and/or composite events (called component events). These components are combined by means of event constructors as: negation, conjunction, disjunction, sequence, etc. The action describes treatments to achieve when a specific event happens and some condition holds.

The potential uses of reactive behaviour are significant: active rules support data derivations, integrity maintenance, workflow management, replication management, and more.

One of the problems of using ECA rules is the inherent difficulty of analysing and controlling their behaviour. An important behavioural property is that of *termination*, since when multiple ECA rules have been defined in an active DBMS, possibly by different people at different times, there is the possibility that the rules may trigger each other indefinitely.

Most of the research efforts on methods and tools for active rules have been focused on compile-time analysis (Aiken et al., 1995), (Baralis et al., 1996), (Baralis and Widom, 2000), (Lee and Ling, 1998), (Bailey et al., 1997), (Karadimce and Urban, 1994), (Kokkinaki, 1998), (Zimmer et al., 1996). This analysis is said *static analysis*. In contrast, very little effort has been focused on run-time analysis (Baralis et al., 1998), (Kappel et al., 2001). This analysis is said *dynamic analysis*. Commercial systems detect loops by keeping counters on the number of cascading rules, and the termination is enforced by imposing a fixed limit on these counters (in Oracle 8, for example, this limit is 64).

As it is above-mentioned, many works have tried to solve the no termination problem; nevertheless, most of them present insufficiencies.

The aim of this study is to propose a rule termination analysis approach which considers the impact of both the composite events and the rule priority on the termination analysis problem. This method detects cyclic paths in the base of ECA rules, can analyse the relationships among ECA rule components and detects cases of termination which are not detected by the other approaches.

2. Related work

Among methods proposed to study rule termination, some are based on models of graphs, and others use formal basis as the systems of rewrite or the Petri nets. We present some of them in the following.

Aiken et al. (1995) are the first to introduce the notion of Triggering Graph (TG). They show that a triggering graph without cycle determines and guarantees the termination of a set of active rules in the context of relational systems.

Baralis et al. (1996) grouped the active rules into modules, termination of rule execution, within each module is assumed and inter-module termination is analysed. It is the only method that presents a modular conception of the active rules.

Baralis et al. (1998) propose a technique that exploits the information of the graph TG and other graph called Activation Graph (AG) to analyse the termination of a set of ECA rules. This analysis uses an algorithm called algorithm of reduction. This approach presents an inconvenience because it doesn't propose a method of building the AG graph which is not obvious.

Baralis and Widom (2000) try to improve these last methods. Their approach is based on a "propagation algorithm", which uses an extended relational algebra to accurately determine when the action of one rule can affect the condition of another. The termination analysis is made by building the graph AG. Their work is considered as a complementary method for the two last one.

Lee and Ling (1998) propose a path technique for reducing the graph TG. The method considers together the conditions of long triggering sequences called activation formulas. It is necessary to guarantee that the execution of rules outside the triggering sequence cannot unpredictably change the database state. Hence, only non-updatable predicates can be included in the activation formula. This condition severely limits the applicability of the technique.

Bailey et al. (1997) use abstract interpretation for the termination analysis of active rules. The idea is to reason about sequences of database states using "approximate semantics", and to use the fix point computation (over a lattice) to handle cycles. This approach is applicable to a simple and restricted rule language.

A different approach is taken in (Karadimce and Urban, 1994), where ECA rules are reduced to term rewriting systems, and known analysis techniques for termination of term rewriting systems are applied. The analysis is based on an object-oriented data model and instance-oriented rule execution model. This approach is powerful, since it exploits the body of work on Conditional Term Rewriting System, but its implementation appears to be complex even for small rule applications.

Kokkinaki (1998) uses Parameterised Petri Nets (PPN) to analyse the active rule termination in the relational model. A PPN is a Petri Net (PN) whose places are parameterised. The firing of the transition corresponds to the rule execution. If there is no cycle in the PPN model then the rules execution in the ADBS must terminate. The inconvenience of this approach is that, the use of PPN in modelling complex systems results in complex graphic representations which are very difficult to be conceptualised, and handled.

Other PN based method is presented in (Zimmer et al., 1996). To represent the triggering and activation notions in the PN, the authors give for each rule two subnets Ei (for the triggering) and Ci (for the activation). The authors detect a non-terminating behaviour of rules using a coverability graph based on the reachability graph. In this approach, the conception of the PN is too complex. In addition, the presence of a cycle in the PN does not imply that will occur an infinite rule triggering.

Li and Marin (2004) present an approach based on coloured Petri nets named CCPN (Conditional Coloured Petri Net) for modelling the active database behaviour. Incidence matrix of PN theory is used to find cyclic paths existing in the CCPN. Cycles which satisfy some theorems given by the authors are deleted. If there is no cycle in the CCPN, the termination of the corresponding set of rules is guaranteed. Nevertheless, this approach does not consider the priority of rules.

3. Extended colored petri net (ecpn)

Our approach is inspired of (Li and Marin, 2004). We improve this method by adding the notion of rules priority and showing how the termination analysis can be affected by this notion.

PN is a graphical and mathematical tool and may be applied in various areas. Active database is a promising application area of PN. Up to now, few researches have used PN as ECA specification language. In SAMOS (Gatziu and Dittrich, 1994) (Geppert et al., 1995), PN is partially used for composite event detection and termination analysis.

In our model named ECPN, the rule event is represented by a place *p1*, the condition *c* and the priority *pr* of the rule are attached to a transition *t*, and the rule action *a* is represented by a place *p2*. Relationships between the rules can be viewed in the same graph (see figure 1).

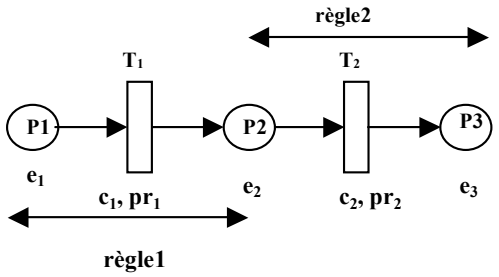


Figure 1.*Relationships between two rules r1 and r2*

3.1. The formal Definition of an ECPN

ECPN is a colored PN (David and Alla, 1989) (Jensens, 1992) defined as follows:
ECPN= {E, P, T, A, N, C, Con, Action, D,τ, I} where:
- E is a finite set of non-empty type, called colour sets. It determines all the data value, the operations and functions that can be used in the net expressions.
- P is a finite set of places. It is divided into four subsets: P_{prim}, P_{comp}, P_{virt} and P_{copy}. P_{prim} represents the set of primitive places and correspond graphically to a single circle. P_{comp} represents the set of composite events. P_{comp} includes the following events: negation, sequence, closure, last, history, and simultaneous. They correspond graphically to a double circle. P_{virt} represents the set of composite events which includes the conjunction, disjunction and any. They correspond graphically to a single dashed circle. P_{copy} is the set of places which are used when two or more rules are triggered by the same event. They correspond graphically to a double circle where the interior circle is a dashed one.
- T is a finite set of transitions; it is divided into three subsets: T_{rule}, T_{copy} and T_{comp}. T_{rule} corresponds to the set of rules. Each transition of rule type is represented graphically by a rectangle. T_{copy} is the set of transitions of copy type. They are represented graphically by a single bar. A copy transition is used when one event *e* can trigger two or more rules. A copy transition will produce *n* same events where *n* is the number of rules which are triggered by the same event *e*. T_{comp} is the set of transitions of composite type. They are represented graphically by a double bar. A composite transition is used for generating a composite event from a set of primitive or composite events.
- A is a finite set of arcs. It is divided into two subsets: the input arcs which are defined from P to T and output arcs which are defined from T to P. Inhibitor arcs are used to represent the Negation composite event.
- N is a node function. It maps each arc into a pair where the first element is the source node and the second is the destination node (N: A → PxT∪TxP).
- C is a color function. It maps each place p to a type C(p) (i.e. C: P → E).
- Con is a condition function. It is defined from either T_{rule} or T_{comp} into expression such that:
$\forall\ t \in T_{rule}$: Type (con (t)) = B, where Con function evaluates the rule condition.
$\forall\ t \in T_{comp}$: Type (con (t)) = B, where Con function evaluates the time interval of t against tokens timestamp. B is used to denote the Boolean type containing the values *false* and *true*.
- Action is an action function. It maps each rule type transition $t \in T_{rule}$ into a type C(p) which will be deposited into its output place.
- D is a time interval function. It is defined from T_{comp} to a time interval $[d_1(t), d_2(t)]$, where $t \in T_{comp}$, and $d_1(t)$, $d_2(t)$ are the initial and the final interval time, respectively. The interval is used by the Con function to evaluate transitions $t \in T_{comp.}$
- τ is a timestamp function. It assigns each token in place p a timestamp.
- I is an initialization function. It maps each place p into a closed expression which must be of type C(p).

3.2. ECPN Execution

In ECPN, a token is a triple (*p, c, time*) where $p \in P$, $c \in C$ (p), and *time* specifies the natural time when the token is deposited into the place *p*.The set of all token is denoted by TS. A marking is a multi-set over TS. The initial marking M_0 is the marking which is obtained by evaluating the initialization expressions:
$\forall\ (p, c, 0) \in TS: M_0(p, c, 0) = (I(p))(c, \tau)$.
The set of all markings is denoted by M. We note by • t the set of the input arcs of the transition t, and by t • the set of the output arcs of the transition t. Transition $t \in T$ is enabled at a marking M iff:
$\forall p \in \bullet t: | M(p) | = 0$, type(t) = Negation
$\forall p \in \bullet t: | M(p) | \geq 1$, else
When a transition t is enabled, an enabled function is needed to specify what tokens transition t is enabled about. In ECPN, composite transitions and rule transitions fire conditionally. A composite transition fires once it is enabled (by $C_{composite}$ function) and the temporal condition is satisfied. A rule transition fires once it is enabled (by $C_{enabled}$ function) and the rule condition is satisfied.

A transition t ∈T fires iff:
∀ t ∈ T_{rule}, t is enabled and Type(Con(t)) = true.
∀ t ∈ T_{copy}, t is enabled
∀ t ∈ T_{comp}, t is enabled, and ∀ p ∈• t, D(t) = [d_1(t), d_2(t)]: [d_1(t) ≤ τ (M(p)) ≤ d_2(t)]
When a transition t is enabled in a marking M_1, and it fires, marking M_1 changes to marking M_2, defined by:
If t ∈ T_{rule}, ∀ p ∈ P: $M_2(p) = M_1(p) - C_{enabled}(p, t) +$ Action(t, p)
If t ∈ T_{copy} , ∀ p_1, p_2 ∈P, p_1 ∈• t, p_2∈ t• :
- $M_2(p_1) = M_1(p_1) - C_{enabled}(p_1, t)$
- $M_2(p_2) = M_1(p_2) + C_{enabled}(p_1, t)$
If t ∈ T_{comp} :
• if Type(t) = Negation, •t= {p_1}, t• = {p_2}
- $M_2(p_1) = M_1(p_1)$
- $M_2(p_2) = M_1(p_2) + C_{composite}(t, p_2)$
• else, ∀p^1_i ∈ •t, t• = {p_2}
- $M_2(p^1_i) = M_1(p^1_i) - C_{enabled}(p^1_i, t)$
- $M_2(p_2) = M_1(p_2) + C_{composite}(t, p_2)$

When a transition t ∈ Trule ∪ Tcomp is enabled at a marking M_1, but not fired because Type(Con(t)) = false, marking change still exists, new marking M_2 is defined as following:
∀p ∈ P: $M_2(p) = M_1(p) - C_{enabled}(p, t)$.

3.3. Example of an ECPN

Let's consider the set of rules R1, R2, R3, and R4 expressed according to the following formalism:
Define rule rule-name
On event
If condition
Then action

Define rule R1	
On reduce-salary ()	(e0)
If employee.salary <1500	(T0)
Then raise-salary ()	(e1)
Define rule R2	
On raise-salary ()	(e3 a copy of e1)
If employee.children-nbr >5	(T3)
Then send-bonus ()	(e4)
Define rule R3	
On raise-salary ()	(e2 a copy of e1)
If employee.age >60	(T2)
Then be-retired ()	(e5)
Define rule R4	
On send-bonus ()	(e4)
If employee.salary <10000	(T4)
Then raise-salary ()	(e1)

These rules necessitate the class Employee which has the attributes: id-emp, salary, bonus, children-nbr, age. The ECPN which corresponds to the rules set given above is shown in figure 2. Furthermore, we consider that $R1 >_p R3 >_p R2 >_p R4$ where $Ri >_p Rj$ means that Ri has priority than Rj.

3.4. Termination Analysis

Let S = {r1, r2, r3, . . ., rn} be a set of rules, when the action of rule r1 triggers the rule r2, action of rule r2 triggers the rule r3, and so on, and finally, action of rule rn, triggers the rule r1, this process performs a cyclic rule triggering and it can produce an inconsistent state of the database when it executes these rules infinitely. To study the problem of rules termination and the impact of rules priority on it, we give an algorithm which uses the incidence matrix, the notions of paths, cyclic paths and acyclic paths of PN theory.
In **incidence matrix**, places are represented by its columns and transitions are represented by its rows, so it is possible to identify both initial and final nodes of an ECPN.
ECPN is a directed graph constituted by a sequence of nodes forming paths, where each node is either a place or a transition in an alternate way. If a cyclic path is found, then it may produce an infinite rule triggering.
A ***path*** *R* is a sequence of pairs (*r, c*) which are obtained from the incidence matrix of the ECPN, where *r* and *c* are incidence matrix indexes.
The sequence of pairs (r, c) describes the connection between places and transitions as follows: (t1, p1), (t1, p2), (t2, p2)... (tn-1, pn-1), (tn, pn-1), (tn, pn). The first pair is formed by the transition t1 and its input place p1, following the path, the next pair is formed by the same transition t1 and its output place p2, the third pair is formed by the same place p2 that now is the input place to transition t2, and so on.
Paths search starts from the coordinate of (t_1, p_1), and then a positive integer is looked for in the row corresponding to t1, finding the coordinate to (t1, p2). After, a negative integer is looked for in the column corresponding to p2, finding the coordinate to (t2, p2). Then, another positive integer is looked for in the row corresponding to t2, and so on, until either a terminal node or an existing node in the path is found.
A ***cyclic*** path CP is a path R where the last pair (r, c) has been already found.
An ***acyclic*** path AP is a path where the last pair (r, c) is different from each other in AP. If all the paths, in the ECPN, are acyclic then the termination is guaranteed. If there is at least one cyclic path CP in the ECPN, the rule triggering may not finish. But it does not mean that whenever exist a cyclic path CP the rule triggering does not terminate; there are other facts that should be taken into account. For

example, in ECPN model, the priority, the condition and the composite events of ECA rules are considered, so they can have an impact on the termination analysis problem. The composite events that affect the termination analysis are: Negation, conjunction, any, and sequence. The conditions 1, 2 and 3, given below, correspond to them. The condition 4 concerns the rule condition belonging to a cyclic path.

- Condition 1: If a cyclic path CP contains an inhibitor arc, i.e. a composite event negation is included in CP, then CP finishes its rule triggering.
- Condition 2: For composite events conjunction, sequence, and simultaneous, if any of its constituent events is not generated by the action of a rule belonging to CP then the rule triggering finishes.
- Condition 3: If composite event any (m, e1, e2,…, en) is a part of a cyclic path CP, and if k constituent events of the composite event any are not generated by the action of a rule belonging to CP and n-k < m then the rule triggering finishes.
- Condition 4: If the condition of a transition ti ∈{t | (t, p) ∈ CP} is always false according to the event produced by the action of previous rule, then the rule triggering finishes.

3.4.1. Algorithm of the Priority

The following algorithm shows the impact of the rules priority on the termination analysis for a set of rules modelled by an ECPN.

Step 1. - Convert a base of ECA rules into an ECPN
Step 2. - Create the incidence matrix from the ECPN
Step 3. - Search all the paths of ECPN
Step 4. - Create a set Rset of paths R
Step 5.-
If Rset contains no cycle
Then returns “the termination is guaranteed”
Else explore the paths in Rset one by one according to the priority of rules:
  termine ← false
  Repeat for each path R
    If R is acyclic
    Then termine ← true
    Else /*R is cyclic*/
      If R satisfies the conditions given above
      Then termine ← true
      Else consider the next path
      End
    End
  Until the end of the paths exploration or termine=true
  If termine = true
  Then returns “the termination is guaranteed”
  Else returns “the termination is not guaranteed”
  End
End.

3.4.2. Illustrative Example

To show the impact of the rules priority on the termination analysis in an ECPN, we consider the example given above. The incidence matrix generated from the ECPN of figure 2 (which corresponds to the example of section 3.3) is presented in figure 3. It can be observed that there are two paths constituted by the elements: (0,0), (0,1), (1,1), (1,2), (2,2), and (2,5) for the first one and the elements: (0,0), (0,1), (1,1), (1,3), (3,3), (3,4), (4,4), (4,1), and (1,1) for the second one which includes the cycle. Nevertheless, as R3 has priority than R2, this cycle is not considered and the termination of the set {R1, R2, R3, R4} is guaranteed according to the algorithm. The path to which belongs R3 (represented by the transition T2) is acyclic.

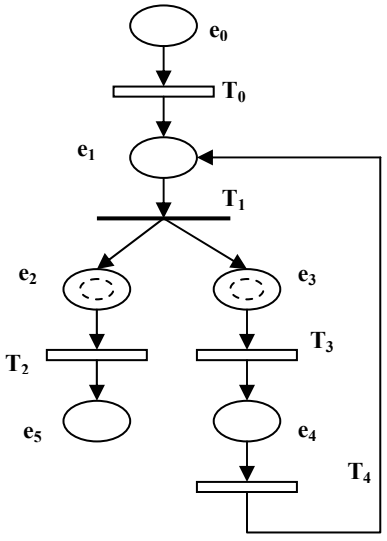


Figure 2.*Example of an ECPN.*

	e0	e1	e2	e3	e4	e5
T0	-1	1	0	0	0	0

T1	0	-1	1	1	0	0
T2	0	0	-1	0	0	1
T3	0	0	0	-1	1	0
T4	0	1	0	0	-1	0

Figure 3.*Incidence matrix from ECPN of figure 2.*

4. Implementation issues

In order to implement an ADBS, the architecture of a (passive) DBMS has to be augmented by new components like an ECA rule editor, an analyzer of rules, a rule manager, an event detector for primitive and composite events, and a rule execution component.

ECA rules are defined by ECA rule editor. After ECA rules are converted into an ECPN, ECPN is saved into ECPN base as places, transitions and arcs. ECPN rule base is used by ECPN rule manager which calls the event detector for detecting events, the rule execution component for evaluating the condition and executing the action of rules. It calls also the termination analyzer component to cheek no-termination problem in ECPN rule base. This last component includes an incidence matrix generator, a paths generator, a cyclic paths detector and an analyzer of paths which take into a count the priority of rules. As the implementation of our passive DBMS is an object-oriented implementation, and in accordance to the idea "to stay in the same world and exploit its advantages", we built the rule structure by means of object-oriented features (Atkinson et al., 1989).

In our model, the ECA rules are represented by transitions of rule type, the conditions and the priorities of rules are attached to these transitions. The events and the actions are modelled by places which are inputs and outputs of a transition respectively. The rule firing corresponds to transition firing. So we need four main classes named: PLACE, TRANSITION, ARC and TOKEN.

All places are instances of a class PLACE with attributes *index, event-name, event-body, list-rules, list-tokens* and methods *insert-place, delete-place, modify-list-rules* and *modify-list-tokens*. The *index* is an internal number assigned by the system. The *event-name* is the event name. The *event-body* is the content of the rule event part. The *list-rules* is the list of references to objects of class T-RULE which are triggered by this event. The *list-tokens* is a list of references to objects of class TOKEN; these tokens represent the actual marking of the place. Class PLACE is the super-class of the subclasses P-COMPOSITE (which includes the attributes *composed-by* and *interval*), P-PRIMITIVE and P-COPY (which includes the attribute *copy-of)*. *Composed-by* holds the events (single or composite) that comprise the event. *Interval* refers to objects of class INTERVALS (i.e. refers to time intervals). *Copy-of* holds the event for which we have made copies.

Each transition is an object of class TRANSITION with attribute *index, name, arc-in, arc-out* and methods *insert-transition* and *delete-transition*. The *name* is the transition name. The *arc-in* is the list of references to objects of class ARC (the set of the input arcs). The *arc-out* is the list of references to objects of class ARC (the set of the output arcs). Class TRANSITION is the super-class of the subclasses T-COMP, T-COPY and T-RULE. The last one regroups all transitions of rule type; it includes attributes *rule-name, events, cond, act, priority* and methods *insert-rule, delete-rule, activate-rule, modify-priority, evaluate-condition. Rule-name* represents the rule name. *Events* represents the rule event; it refers to an object of class PLACE. *Cond* represents the rule condition; it refers to an object of class CONDITION. *Act* represents the rule action; it refers to an object of class PLACE. *Priority* is a number reflecting the importance of the rule.

Also we define a class ARC with three subclasses: ARC-IN (input arcs), ARC-OUT (output arcs) and ARC-INHIB (inhibitor arcs). Class ARC has as attributes *index-p* and *index-t*, the value of which are the appropriate place and transition indexes (in relation to the attribute *index* of classes PLACE and TRANSITION).

The attribute *time* in the class TOKEN (the class which regroups all the tokens of an ECPN) indicates the time when we put the token on a place. The figure 4 presents the class hierarchy of our metabase with more details.

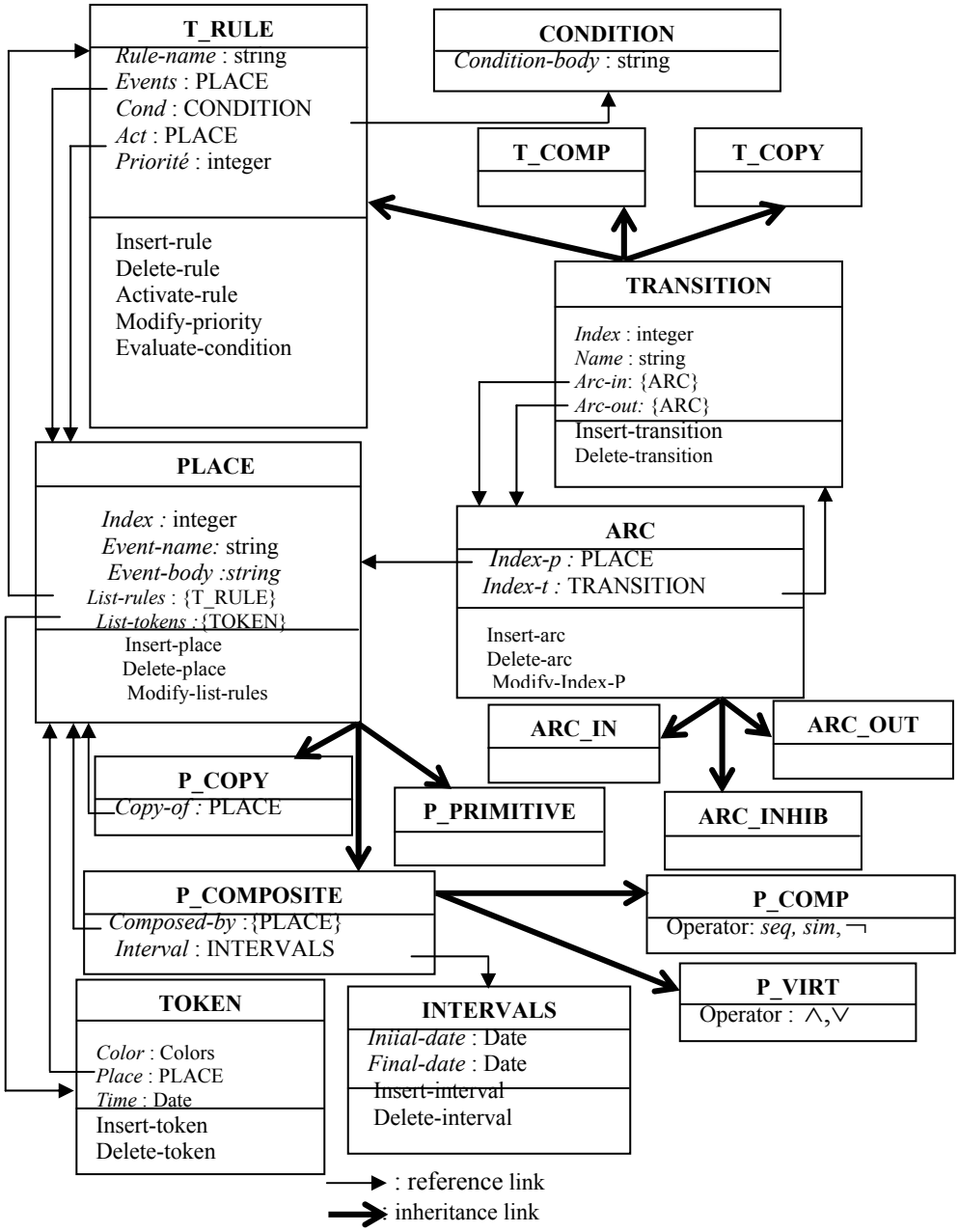


Figure 4.*Class hierarchy.*

5. Conclusion

The approach presented in this paper, is based on the Petri net model to analyse the termination of a set of rules. In this PN named Extended Coloured Petri Net (ECPN), the different components of the rules are presented such as their events, conditions, actions and priorities. ECPN can model both primitive and composite events. Furthermore, not only composite events can affect the active rules termination, but also the rules priority. Our approach is better than those presented in related work section because the ECPN is a good model for modelling, analysing and simulation of active database systems. It does not perform a simple analysis of cyclic paths but analyzes each element of the graph to determine if the rule triggering in a cyclic path finishes or not. This approach is general and can be applied not only in the database area but also in others applications which need event detection.

6. REFERENCES

1. Aiken, A., J. M. Hellerstein, and J. Widom, 1995. Static Analysis Techniques for Predicting the Behaviour of Active Database Rules. *In ACM Transactions on Database Systems*, vol. 20, N°1, pp : 3-41.

2. Atkinson, M., F. Bancilhon, D. DeWitt, K. Dittrich, D. Maier, and S. Zdonik, 1989. The Object-Oriented Database System Manifesto. *In Proc. of the 1st Int. Conf. on Deductive and Object-Oriented Databases,* Kyoto, Japan. Elsevier Science Publishers B.V (NorthHolland), pp : 223-240.

3. Bailey, J., L. Crnogorac, K. Ramamohanarao, and H. Sndergaard, 1997. Abstract Interpretation of Active Rules and its Use in Termination Analysis. *In Proceedings of the 6th International Conference on Database Theory*, January 1997, Delphi, Greece, pp : 188-202.

4. Baralis, E., S. Ceri, and S. Paraboschi, 1996. Modularization Techniques for Active Rules Design. *ACM Transactions on Database Systems (TODS)*, vol. 21 N°1, pp : 1-29.

5. BARALIS, E., S. CERI; AND S. PARABOSCHI, 1998. COMPILE-TIME AND RUNTIME ANALYSIS OF ACTIVE BEHAVIORS. *COMMUNICATION OF THE IEEE TCDE*, VOL.10, NO 3, PP : 353-370.

6. BARALIS, E. AND J. WIDOM, 2000. BETTER STATIC RULE ANALYSIS FOR ACTIVE DATABASE SYSTEMS. *ACM TRANSACTIONS ON DATABASE SYSTEMS (TODS)*, VOL 25, N°3, PP : 269-332.

7. Chakravarty, S. and D. Mishra, 1994. Snoop : An Expressive Event Specification Language for Active Databases. In *Data & Knowledge Engineering*, vol. 14, N°1, pp : 1-26.

8. Collet, C., 1996. Bases de données actives : des systèmes relationnels aux systèmes à objets. Rapport de recherche 965-I-LSR-4, Laboratoire IMAG, université Joseph Fourier, Grenoble, France, Novembre, pp : 1-102.

9. Collet, C. and T. Coupaye, 1996. Primitive and Composite Events in NAOS. *In Actes des 12ème journées de Bases de Données Avancées*, France, pp : 331-349.

10. David, R. and H Alla (1989). Du Grafcet aux Réseaux de Petri. Edition Hermès, Paris.

11. Gatziu, S. and K. R. Dittrich, 1994. Detecting Composite Events in Active Database Systems Using Petri Nets. In *Proceedings of the 4th International Workshop on Research Issues in Data Engineering : Active Database Systems*. Houston, Texas, February, pp : 1-8.

12. Gehani, N., H. V. Jagadish, and O. Shumeli, 1992. Composite Event Specification in Active Databases: Model and Implementation. *In Proc. 18th International Conference on Very Large Data Bases*, Vancouver, Canada, August, pp : 1-12.

13. Geppert, A., S. Gatziu, K. R. Dittrich, H. Fritschi, and A. Vaduva, 1995. Architecture and Implementation of the Active Object Oriented Database Management System SAMOS. Technical Report 95.29, CS Department, University of Zurich, pp : 1-39.

14. Jensens, K., 1992. Coloured Petri Nets : Basic Concepts, Analysis Method and Practical Use, vol. 1, EATCS.

15. Kappel, G., G. Kramler and W. Retschitzegger, 2001. TriGS Debugger: a tool for debugging active database behavior. Research Report, Linz University, Austria, pp : 1-12.

16. Karadimce, A. and S. Urban, 1994. Conditional Term Rewriting as a Formal Basis for

Analysis of Active Database Rules. In *4th Workshop on Research. Issues in Data Engineering (RIDE-ADS94*), Texas, February, pp : 156-162.

17. Kokkinaki, A.I., 1998. On Using Multiple Abstraction Models to Analyse Active Databases Behaviour. In: *Biennial World Conference on Integrated Design and Process Technology*, Berlin, Germany, June, pp: 1-8.

18. Lee S.Y. and T. W. Ling, 1998. A Path Removing Technique for Detecting Trigger Termination. In Proceedings of 6th EDBT, Valencia, pp : 341-355.

19. Li, X. and J. M. Marín, 2004. Termination Analysis in Active Databases: A Petri Net Approach. *International Symposium on Robotics and Automation (ISRA2004),* Querétaro, Mexico, July, pp : 677-684.

20. Paton, N. W. and O. Diaz, 1999. Active Database Systems. *In ACM Computing Surveys*, vol.1, N°31, pp : 63-103.

21. Widom J. and S. Ceri, editors (1996). Active Database Systems: Triggers and Rules for Advanced Database Processing. Data Management Systems. Morgan Kaufmann Publishers, San Francisco, California.

22. Zimmer, D., R. Unland, and A. Meckenstock, 1996. Rule Termination Analysis Based on Petri Nets. Cadlab Research Report 3, university of Paderborn, Germany, February, pp : 1-17.

Real-Time Multimedia Data Acquisition Protocol for Industrial Applications

Asad Ali Shaikh, Tabasum Abbasi, Muhammad Iqbal Bhatti
Institute of Information Technology
Jamshoro, Pakistan
Asadshaikh_56@yahoo.com, tabasum_17@yahoo.com, MI_Bhatti@yahoo.com

Abstract

In this paper the design and implementation of Real Time Multimedia Data Acquisition protocol (RTMDAP) for industrial applications is presented. The intended use of this protocol is factory environment, where the remote supervision and control of the devices (sensors, actuators) are important tasks. During control activities data is in shape of text strings and in some cases, symbols representing device state. Factory has many equipments like robots, Numerical Controlled Machines, Vision, Image processing and controllers for achieving the better quality, part tracking, and assembly. By deployment of multimedia applications at plant level factory can perform all above mentioned tasks effectively. RTMDAP extends common network communications facilities into the scope of real time data acquisition. RTMDAP (The Real Time Multimedia Data Acquisition Protocol) is based on Switched Ethernet 100 Mbps. TCP/IP and RTP Protocols are employed in the RTMDAP along with UDP with acknowledgement mechanism so that better reliability can be achieved. Well-established and widely used compression techniques H.263, JPEG are also used in the RTMDAP and timely data is. Both the design and implementation are presented together with results of from measurements on a prototype of the protocol. Sampling rate in case of JPEG (live video) from 41.383 to 151.492 packets per second are obtained for various resolutions and in case of H.263 (live video) sampling rate is achieved from 10.53 to 12.374 packets per second with low resolutions. Significant contribution is to deploy the video compression in the factory environment. Well established widely supported video compression techniques H.263 and MJPEG are included in the protocol. The protocol performs the capturing of video through webcam (live).

Key Words: RTMDAP, Administrator, Producer, consumer, switched Ethernet

1 Introduction

The main objective of every industrial system is the remote supervision and control of devices (sensors and actuators located at remote facilities)

for better quality control, assembly and better part tracking. Many factories are using SCADA (Supervisory Control And Data Acquisition) for this task Boyer, S.A 1999. The infromation provided by the devices is about few bytes. The traditionally used. Work is carried out during 2004-2005 at Institute of Infromation Technology, University of Sindh, Jamshro Pakistan

technologies (fieldbuses) are still suitable for the delivery of devices data which is about few bytes. The communication technolgies (fieldbuses)
used in industy are time-sensitive.During the two decades transmission and networking technologies are improved dramatically beside this multimedia compression techniques have also evolved in such a way that digital video and audio can now be processed by inexpensive PCs, obtaining good quality displays.

2. Review of Existing Techniques

There are many issues to be considered when dealing with computer multimedia systems
Video captured by cameras needs to be digitized so computer can handle them. Modern cameras, such as Digital Video Cameras (DVCAM), already provide the video signal in digital form, but other video devices needs a previous step called capturing to be performed before processing is done. Typically, frame grabber does video capturing, which has all circuitry such as A/D converters.
Video data when on raw digitized form, take a huge amount of space. Keeping data size on practical level, multimedia data compression is required. There are many coders/decoders (normally called codecs) are available. Few of them are highlighted in this work MJPEG, H.261, H.263 for video compression.

Sending multimedia data over computer networks is done by a process called streaming. Streaming consists of taking compressed video and chopping it off into blocks. But even when data is compressed it still requires high bandwidth links to be transmitted. A 640x480 pixels stream encoded with MJPEG codec at a frame rate of 25 fps requires 1.5 Mbps ISO/IEC 11172-2 (1993).

2.1 Multimedia on Industrial Systems

Typical transmission links and protocols used on industrial networks do not fulfill the bandwidth requirements of multimedia data deployment, as data acquired from Industrial devices are usually transmitted over low-bandwidth links using serial protocols. Another important key factor is that time-sensitive requirements on the reception of the acquired data are very rigid in the industrial environment Jasperneite,,J . Neumann,P(2001) . Furthermore transmission reliability is also very important as loss of data might cause a big impact to the system integrity. When multimedia data is deployed on the industrial environment, both multimedia data and factory will share the same physical medium where the temporal approach must be adopted. Ethernet network can integrate both on same medium, but transmission reliability must also be considered. Ethernet hub devices use CSMA/CD based medium sharing ANSI/IEEE Std 802.3 (1998), which can lead to packet loss in high traffic scenario, due to packet-to-packet collision. Therefore, switches should be starting point on the design of any multimedia industrial system. Network management techniques such as quality of service Feng, W. et al (1998), Almquist, P (192), and Zhang, L. et.al. (1993) might also be used to ensure factory data always gets the bandwidth it needs. QoS allows packet prioritization and bandwidth provision, based on one or more factors, such as the IP address of the sender.

3 Real Time Multimedia Data Acquisition Protocol

In this section the design of RTMDAP is described. The features in this design serve as the basic building block of RTMDAP. The RTMDAP deploys multimedia applications in factory environment.

3.1 Design Constraints

Uses communication model which supports industrial applications: one of the goals of RTMDAP is to be used in factory environment because of this constrain the communication model will be used must meet with the industrial communication requirements.

Based on existing hardware and software protocols: another important requirement is to use existing hardware and software whenever possible. This requires that the RTMDAP should support existing hardware. Customized hardware would drive up the price of devices considerably and take up a lot of time to design and implement.

Multimedia support: RTMDAP can perform acquisition of live video, stored video and other data type. Live video capturing will be done through webcam and stored video will be read from storage. For other data type sampling will be performed.

Non-real-time traffic: beside real-time traffic (RT) protocol should allow the non real-time traffic (NRT), with a condition non real-time traffic should not affect real time traffic. The only constraint is that preference is given to real time traffic.

QoS guarantees: To meet with quality receiver sends the acknowledgement to sender so that next packet will be transmitted.

Error recovery: the physical layer of network generally does not provide faultless delivery of all packets. When such situation will be occurred RTMDAP should recover from network faults during transmission.

3.2 Overview

The rest of this section describes the design of the protocol in detail and elaborates on all the problems, solutions and design choices made during the development of RTMDAP

3.2.1 Hybrid Model

The RTMDAP is to be applied in factory environment where the communication tasks can be implemented according to two basic models: message-based and Producer-Consumer. The message-based model has at least one master and other slaves. The cyclic data run through all the slaves as polled by master. The message-based model does not allow the slaves to access the network without the master's permission. Therefore the alarms are transmitted to the master only when the slaves are polled. In the Producer-Consumer model one producer broadcasts (multicasts) the data once to all consumers. All the consumers to see the data simultaneously and may choose whether to consume (receive) the data or not. The real–time protocols designed in conformity with Producer-Consumer model have station called either scheduler or bus arbitrator. A hybrid model is proposed which contains two different models:

Client-Server model

Producer-Consumer model and model is illustrated in Figure 1.

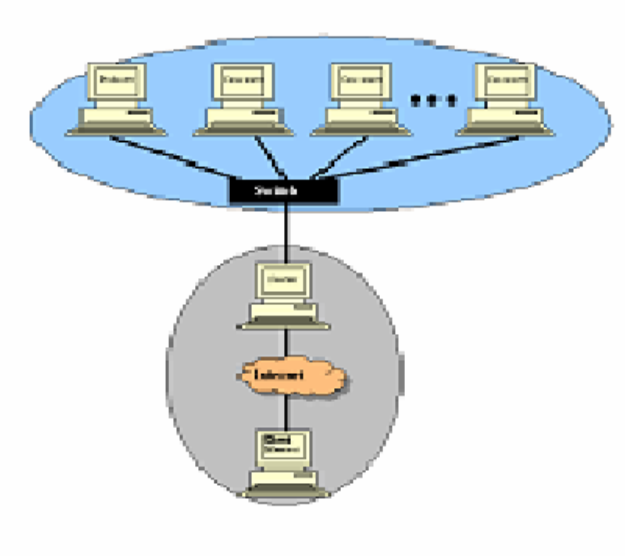


Figure 1: View of proposed Hybrid Model

3.2.2 Low-Level Hardware Protocol

The prototype uses switched Ethernet because of its design, which offers low latency. The original bus-based Ethernet is unsuitable for real-time traffic because of shared medium and back-off algorithm. Several approaches are used to make original bus-based Ethernet suitable for real-time traffic. The approaches falls in three categories: token based medium access control protocols, time-slot based protocols, and statistical. By using time-slots and token passing techniques nodes can cooperate and in result the collision among the nodes minimized and bandwidth can be allocated. Recent enhancement (i-e connecting system to full duplex Ethernet switches) in Ethernet technology has eliminated the possibility of collision occurring in an Ethernet networks. A major step towards deterministic behavior in Ethernet networks to eliminate the random CSM/CD bus arbitration. This can be achieved by using latest Ethernet switch technology instead of hub-based infrastructure. Switch technology divides into simple point-to-point connection between network component and stations. Collision no longer occurs and back-off algorithm no more required.

3.2.3 Real-Time Traffic

In the RTMDAP a mechanism is adopted to make protocol suitable for real-time traffic. Non real-time traffic is assigned to TCP and real time traffic is assigned to UDP. Since UDP does not provide any reliability. Modification is
done to UDP by including acknowledgement mechanism in it. Non real-time traffic, which is normally brusty, can affect real-time traffic. Leaky bucket smoother is used to control the non-real-time traffic.

3.2.4 Non-real-time-traffic

The protocol also has to handle non-real-time traffic. In order to do this, non real-time traffic is permitted only when no real-time streams want to transmit something. When any node ready to transmit real-time traffic, it is given the preference over the node, which wants to transmit non-real-time traffic.

4 Implementation

RTMDAP consists of four portions. One is called Administrator (client). The administrator is located on machine, which invokes the protocol, passes the model number and other parameters to server, and also it passes the information about type of the traffic whether real time or non real time traffic. It also monitors transfer of data between producer and consumer through server.
The second portion is called server, the server is located on the machine, which connects producer and consumer after establishing the connection with administrator and also sends the information to administrator for monitoring the activity between producer and consumer. Once server gets the message from the administrator machine, which contains the model number, parameters and data type RT (Real-time) or NRT (Non Real-time). Server switches between UDP and TCP as per type of data. It multicast the same message to all producers, once the server gets the response from producer(s) which intends to transmit the data. Server sends model number and parameters to producer(s) and receives acknowledgment from producer(s). Server requests the consumer for channel and gets acknowledgement from consumer. Server establishes connection with consumer. Server requests the producer to begin data transmission and server gets the data packets and sends the same to consumer and monitoring packet to administrator. If during communication between producer and consumer which is purely real time transmission and administrator sends another request for NRT transmission server avoids it and waits for completion of transmission between producer and consumer. Server uses leaky bucket for smoothing the brusty NRT traffic.

The third portion is producer; the producer is located on the machine, which performs the data sampling. Its job is to wait for request from the server to open a communication channel, provide such a channel when asked, run a requested measurement program or start capturing images through web camera, then

close the channel when data acquisition is finished. The measurement program itself sends the data packets using facilities established by RTMDAP.
The fourth portion is consumer; the consumer is located on a machine, which collects the data, which is transmitted by the producer through server. The consumer is responsible to establish a communication channel with producer through server, obtaining data packets, estimating data packets if necessary, then closing the channel when finished.

RTMDAP is developed in JAVA. It can run on Window 98, windows 2000, windows XP, Linux 7.2, Linux 7.3, Linux 8.0 and Linux 9.0 and the UNIX operating systems. It includes facilities for TCP/IP, UDP, and RTP.

Figure 2 depicts the type of data packets, which are transmitted over the communication channel during a typical invocation of RTMDAP. Initially, the administrator requests that the server open a communication channel. The server responds with acknowledgement. The administrator then sends a multipurpose packet. This packet contains the name of measurement program, which the producer is to run and transmit the data to consumer through server. It also contains the type of the data whether real-time or non real time. After receiving the multipurpose packet from administrator the server sends the request to producer(s) and gets acknowledgment. Server establishes the connection with producer(s). Server requests the consumer for channel and consumer acknowledges. Server establishes connection with consumer and sends multipurpose packet to producer. This packet contains the name of the measurement program, which producer is to run. It also contains the name of the file from which the model are to be read, as well as a code identifying the class of model whose parameters needed. The producer responds with the values of the requested parameters. These packets serve to prepare both consumer and producer for data acquisition. After the producer receives the "Setup Info Send & Receive" packet, the required Figure 2:

Measurement program is loaded. The measurement program then waits for a synchronization packet telling it when to start sampling. Similarly, the consumer initializes its internal data structures to begin data collection and estimation.

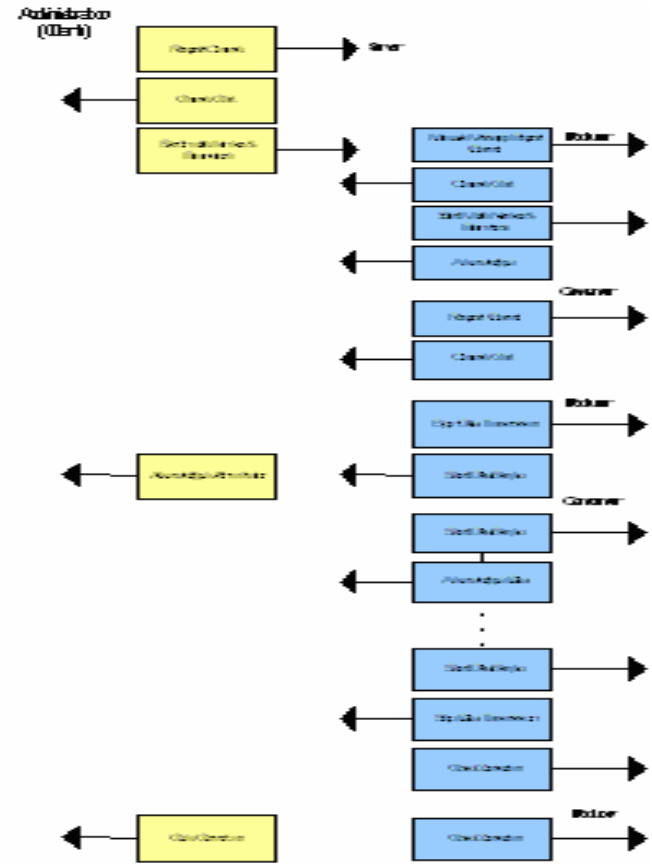


Figure 2: Typical RTMDAP Packet Traffic

When the consumer is prepared and ready to receive data, it sends a request to producer through server asking that data transmission to begin. The producer then repeatedly sends data packets. After receiving each data packets, the consumer acknowledges reception to ensure reliable packet transmission and process the data in whatever fashion necessary.

When the user's applications on the consumer machine have collected sufficient data, it then sends a packet to producer through a server informing it to stop data transmission. The measurement program on producer then stops, and RTMDAP closes the connection. The user's application on consumer may then proceed to other activities.

5. Experimental Setup for Videos

A series of experiments have been performed to examine JPEG and H.263 for collecting information for average packet size, average packet per second, average packet size, bytes, average bytes per second, and average mega bits per second. These quantities

were also examined for a variety of resolution: 160 x120, 176 x 144, 320x 240, 352x288, and 640 x 480 for both live and stored video.

The machines being used throughout the experiments are 4 machines, 2 Dell OptiPlex 400 MHz machines; one machine uses the MS-Windows 98 another uses MS-windows The communication hardware on one machine is the 3COM EtherLink XL 10/100 PCI complete management NIC (3C 905C-TX) Ethernet card. The second machine uses the Netelligent 10/100 TX PCI Embedded UTP controller. Third machine IBM IntelliStation Dual processor of 500 MHz uses RedHat Linux 7.3. The communication hardware on this machine is the 3COM EtherLink XL 10/100 PCI complete management NIC

(3C 905C-TX) Ethernet card. Fourth Machine is Compaq DESKPRO SB 400 MHz; machine uses the MS-Windows XP. The communication hardware on this machine is the 3COM 3C918 Integrated Fast Ethernet Controller (3C 905B -TX) compatible Ethernet card. Machines are connected through two CISCO 2500 routers and CISCO 2800 switch. UTP cables are also used. IOS for all CISCO equipments 11.0. LOGI TECH PC-CAM is used in setup.

The video capturing program is on producer which grabs the video through webCam and video player is on consumer. The measurement program is on the consumer machine. The Ethereal and winpcap are used for collecting the information.

The independent variables for these experiments are time in which packets to be transmitted and the resolution. The following values are dependent quantities which are measured. These values are collected at consumer

A comparison of resolutions is illustrated in Table 1. Clearly indicates that the numbers of packets received by consumer, during the transmission of live video with low resolution the numbers of packets are low and with the high resolution the numbers of packets are high. Only in case of one resolution 176 x 144 some unusual result is achieved, where numbers of packets are slightly lesser than resolution 160 x 120. It may be because of compression technique. In Table 8.19 it is clearly indicated the average packets/sec increase with increase in the resolution only slightly decrease in case of resolution 176 x 144. The average packet size shows some increase. In all cases of resolutions the numbers of bytes are increased from low to high resolutions. Only some variation is seen in case of resolution 352 x 288. Smooth increase is achieved in the measurement of average bytes per second. Similar pattern is for average mega bits per second.

JPEG					
Resolution	160 x 120	176 x 144	320 X 240	352 x 288	640 x 480
Time (Min)	2	2	2	2	2
Packets	4986	4213	10655	9917	18199
Avg. Packet/sec	41.383	34.856	88.607	82.539	151.492
Avg. Packet size	749.818	919.185	908.495	945.707	1003.309
Bytes	3738595	3872527	9680017	9378580	18259221
Avg. bytes/sec	31029.74	32039.34	80498.68	98058.08	151993.12
Avg. Mbits/sec	0.248	0.256	0.644	0.624	1.126

Table 1: Comparison of results for live video (JPEG)

H.263		
Resolution	160 x 120	176 x 144
Time (Min)	2	2
Packets	1425	1492
Avg. Packet/sec	10.053	12.374
Avg. Packet size	442.413	480.141
Bytes	630438.53	716371
Avg. bytes/sec	5253.6542	5941.052
Avg. Mbits/sec	0.045	0.048

Table 2: Comparison of results for video (H.263)

A comparison of different resolutions illustrated in Table 2. Clearly indicates that the numbers of packets received by consumer, during the transmission of live video with low resolution the numbers of packets are low and with the high resolution the numbers of packets are high. In Table 2. it is clearly indicated that the average packets/sec increase with increase in the resolution. The average packet size shows some increase, in all cases of resolutions the numbers of bytes are increased from low to high resolutions. Smooth increase is achieved in the measurement of average bytes per second. Similar pattern is for average mega bits per second. No observations are

displayed for 320 x 240, 352 x 288 and 640 x 480. It seems the H.263 does not support high resolutions.

7. Conclusion

RTMDAP is intended to deploy the multimedia applications to current factory environment in which monitoring and quality control is achieved through text mode and symbols. Video compressions H.263 and MJPEG are suitable for factory environment because of low bandwidth support. Because of deployment of compression techniques in the protocol, it eases the life of the factory operators to perform better monitoring and quality control activities.

RTMDAP is based on switched Ethernet hardware protocol and the TCP/IP suite, UDP/IP, and RTP software protocols. Significant contribution of RTMDAP is to deploy the video compression at factory environment. Well established widely supported video compressions techniques H.263, MJPEG are included in the protocol. The protocol performs the capturing of video through webcam (live). RTMDAP is implemented in JAVA 2.0, Java.net, swing, and JMF. The particular protocol is designed to run under MS-DOS, MS-Windows 98, Windows 2000, Windows XP, Linux 7.2, Linux 7.3, Linux 8.0, and Linux 9.0.

Acknowledgement

The authors are thankful to former Dean Faculty of Natural Sciences for his constant support. We are debited to pay thanks to Mr. Mazhar-ul-Siddiqui,, Honourable Vice Chancellor University of Sindh who is always encouraging scholars to carry on their research work,

8. References

1. ANSI/IEEE Std 802.3 1998 ”ISO/IEC 8802-3, Carrier sense multiple access with collision detection (CSMA/CD) and physical specifications”, IEEE, NJ.

2. Almquist, P (192) “Type of Service in the Internet Protocol Suite,” RFC 1349, July.

3. Boyer, S.A 1999 “SCADA Supervisory Control And Data Acquisition, 2nd Edition “, ISA the Instrumentation, Systems and Automation Society, New York (United States).

4. Feng, W. and Kandlur, K. et al (1998) “Adaptive packet marking for providing differentiated services on the internet”, Proc. of Int. Conf. on Network Protocols, Austin, TX, October.pp. 108-11

5. Jasperneite, J Neumann, P 2001 " Switched Ethernet for Factory Communication”, 8th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA´01), Antibes – Juan Le Pins, October ,pp. 205-212

6. International Organization for Standardization. Information technology. Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s . Part 2: Video, 1993. ISO/IEC 11172-2:1993.

7. Zhang, L. Deering, S et.al. (1993) “RSVP: A New Resource Reservation Protocol,” IEEE Network Magazine, Vol. 7. pp. 8-18 September .

Adapting Bresenham Algorithm

Mustafa S. Fawad
Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences, China
fawad_zaidi@yahoo.com

Abstract

The Bresenham line algorithm is a promising computer graphics process whose basic justification seems to have been its lack of reliance on floating-point operations. In more recent time it has sometimes been passed by for other algorithms that can support scan conversion of regions with non-linear boundaries. However, the Bresenham algorithm can be thought of as interpolating pixels, and there are ways it can be an excellent basis for other interpolation operations. In this paper, we describe several of these operations and how they can be built from the Bresenham algorithm.

Keywords: Polygonal filling, scan-line, algorithm, graphic interpolation

1. Introduction

There is a well-known algorithm for plotting straight lines on a display device or a plotter where the grid over which the line is drawn consists of discrete points or pixels. In orking with a lattice of points it is useful to avoid floating point arithmetic. One of the first published algorithms was by Jack Bresenham[1] who worked for I.B.M. (1965). The descriptions can be found in [2], Chapter 3. The main idea of the algorithm is to analyze and to manipulate the linear equation so that only integer arithmetic is used in all the calculations. Integer arithmetic has the advantages of speed and precision; working with floating point values requires more time and memory and such values would need to be rounded to integers anyway. In this paper, we describe several of these operations and how they can be built from the Bresenham algorithm.

1.1. Basic Bresenham Approach

The algorithm has been presented in many articles that we limit our discussion to graphic interpolation processes with Bresenham Algorithm here. The pseudo codes (in simple C) are totally general and covers all the cases of line orientation. Our discussions will refer to this algorithm in detail and will use the notation of the Appendix mentioned. In order to illustrate the approaches of this paper, the simpler version of this algorithm is given in Figure1. The algorithm operates in screen space, taking as its input the integer coordinates of the end-points of a line segment and identifying the pixels to be drawn to represent the line segment itself. The key to the algorithm is that it steps one unit at a time through the more quickly varying coordinate and each time it does so, it uses an error term "e" to decide whether or not to step one unit in the more slowly moving coordinate.

```
px = x1;
py = y1;
dx = (x2 - x1);
dy = (y2 - y1);
dx = abs (dx);
dy = abs (dy);
e = 2 * dy - dx;
const2 = 2 * (dy - dx);
const1 = 2 * dy;
/* set initial point on line */
setpixel (px, py, COLOUR);
while (px != x2)
{
px += 1;
if (e < 0) e += const1;
else
{
e += const2;
py += 1;
}
setpixel (px, py, COLOUR);
}
```

Figure 1
The standard Bresenham algorithm for the case of a line more horizontal than vertical

2. Our Approach

We will focus on using the Bresenham Algorithm to different graphic interpolation.

2.1. Adapting to Polygon Filling

Polygon filling or more generally, filling a boundary-enclosed space is a classical problem in computer graphics that has been taken into consideration in many ways. One of the worth mentioning approaches is taken in [3], where the boundary is built in a one-bit-deep memory that duplicates the dimensions of the marking space. Besides outlining the boundary-defined area fill problem, [3] discusses some issues in getting accurate filling. The Bresenham algorithm can be used to fill a general polygon with vertices in screen space by determining the edges of the polygon and storing appropriate points from these edges in scanline lists from which the polygon can be filled. Polygon filling needs to address several issues such as:

- ensuring that the area of a screen-space polygon is reasonably the same as the area of its model-space original, and

- ensuring that the display of two abutting polygons will be the same regardless of the sequence in which the polygons are drawn. Managing these issues is a standard part of polygon filling. For example, see [4].

We address these issues by adopting a few conventions in choosing the points to enter for each scanline. These conventions have the following effects:

- on any scanline, the filled area will extend from the leftmost pixel of the set of all pixels from the left-hand edge on that scanline to the point immediately before the leftmost pixel

- on the right-hand edge on that scanline, in any polygon, any point that can reasonably be called a top point for the polygon (that is, a point from which two edges slope downward, or any point on a horizontal top edge) will not be drawn, and

- when two edges meet in any point which is the lowest point for one edge and the highest point for the other, that point will be stored only once. There are very simple solutions to handling these problem areas. We have found that three simple conventions allow polygon filling without encountering these problems. These conventions are:

1) Drawing each scanline including the leftmost pixel but omitting the rightmost pixel.

2) Always drawing any (non-vertical) line segment from left to right, and only entering the point (px, py) in the scanline list when there is a change in py. This captures the leftmost among all the points on the line segment that lie on the scanline.

3) Entering only the points (px, py) on a line segment when that point is not the highest point on the segment. The first convention assures that the width of the polygon on any scanline is a reasonable approximation of the actual polygon width for that value of y. The second convention assures that for any one line segment and any one scanline, the leftmost point that the Bresenham algorithm places on that scanline will be the only one point entered on the scanline. It also assures that the first convention will be followed because all the endpoints from the leftmost point of the left-hand edge up to (but not including) the leftmost point on the right-hand edge can be identified for filling. The third convention assures that no shared topmost vertex will be saved (so top points of polygons will not be drawn), that any shared bottomlast vertex will be listed twice (so the vertex will not be part of the filled space, and that any vertex that is the topmost point of one line segment and the bottommost point of the next or previous line segment will only be entered once.

These conventions are quite similar to those of [3], who advocated the drawing in a fixed direction, not drawing the last pixel of a boundary segment, and not entering horizontal segments in a polygon boundary. However, Gay's[3] conventions do not speak to the question of area accuracy, which we address in the first of our conventions.

When we add the endpoint inclusion and the top exclusion as described earliest, a slight change is done to the code in Figure 1. The new version of the code can be seen in Figure 2. The ***savePixel(px,py)*** function inserts the point (px,py) into a vertex list for the scan line for later processing in filling a polygon. Such a filling process is straightforward and proceeds by sorting the points in the vertex list for the scan line and then for each pair of points in the vertex list, drawing a horizontal line segment connecting them. That horizontal line needs to avoid drawing the right-hand endpoint of the segment, but this is a simple change.

```
orderPoints( &x1, &y1, &x2, &y2);
/* ensure x1 < x2 */
ymax = max( y1, y2);
px = x1; by = y1;
dx = x2 - x1; dy = y2 - y1;
dx = abs(dx); dy = abs(dy);
```

```
e = 2 * dy - dx;
const2 = 2 * (dy - dx);
const1 = 2 * dy;
if ( py != ymax )
/* set initial point */
savePixel (px, py, COLOUR );
while (px != x2)
{
px = px + 1;
if( e < 0 ) e = e + const1;
/* no pixel saved */
else
{
e = e + const2;
py = py + 1;
if (py != ymax )
savePixel ( px, by, COLOUR );
}
}
```

Figure 2

The first octant Bresenham algorithm of Figure 1, adapted to store vertices for polygon filling. ***Italics and bold indicate changes***.

2.2. Adapting to Property Interpolation

In a more general setting, let us assume the existence of a property π that is associated with each vertex of a line segment, so that the two endpoints can be written as (X1, Y1, π1) and (X2, Y2, π 2). Then we can set up a unit increment of the property by dividing the total change in the property across the line segment by the larger of the changes in X and Y over the segment. This increment is applied for each step of the Bresenham process, and the values of the property are saved in the vertex-storing process. The approach is described in Figure 3. It saves the interpolated value of the property along with the point on the scanline list, and this interpolated value may be used as it is again interpolated across the scanline as the actual polygon is filled, as shown in Figure 4. Here the *setPixel* function is intended to perform any functions needed to determine the COLOUR for the pixel. Some applications of this general interpolation process are noted in the Section 3.

```
orderPoints( &x1, &y1, &π1, &x2, &y2, &π 2 );
/* x1 < x2 */
ymax = max(y1, y2 );
px = x1; py = y1; pπ = π1;
dx = (x2 - x1); dy = (y2 - y1);
/meta statement* property difference */
dπ = (π2 - π1)/dx;
dx = abs(dx); dy = abs(dy);
e = 2 * dy - dx;
const2 = 2 * (dy - dx);
const1 = 2 * dy;
if (py != ymax )
savePixel( px, py, pπ); /* set initial point */
while ( px != x2 )
{
px = px + 1;
pπ = pπ + dπ;

if ( e < 0 ) e = e + const1;
else
{
e = e + const2;
py = py + 1;
if ( py != ymax )
savePixel( px, py, pπ );
}}
```

Figure 3

Maintaining a property value with the algorithm of Figure 2. Italics and bold indicate changes.

```
getScanlinePoint (x1, y1, π1 1 );
getScanlinePoint (x2, y2, π1 2 );
px = x1;
py = y1;
pP = P1;
dx = x2 - x1;
dπ = π1 2 - π1 1;
xsign = dx / abs(dx);
setPixel(px, py, p□);
while(px != x2)
{
px += xsign;
p π += d π;
setPixel(px, py, p□);
}
```

Figure 4

Modification of the horizontal special case for filling with an interpolated property.

2. Discussion

The paper highlights the applicability of the approach mentioned in Section 2. to several
different graphics processes, as enlightened in the following:

Case 1: If the property π is the Z-value of each vertex of a polygon, the interpolated value is the Z-coordinate of each point in the polygon as it is being filled, so the interpolation process supports Z-buffering.

Case 2: If the property π is the colour of each vertex of the polygon, the interpolation process provides an easy way to perform Gouraud shading of the polygon.

Case 3: If the property π is the normal vector for each vertex, the interpolation process provides a normal vector for each point in the polygon, supporting Phong shading for the polygon. A variation provides bump mapping by using a 2D map of normal vector perturbations instead of a 2D texture; adding this interpolation to the normal interpolation for Phong shading provides a perturbed normal vector at each point in the polygon. Another variation provides environment mapping by computing the reflection of the eye vector at each point in a polygon based on an interpolated normal, allowing one to compute the intersections of the reflection vector with a mapped environment.

Case 4: the property π is a texture coordinate — a coordinate in a 2D or 3D texture space. Then interpolation will provide a texture coordinate for each point in the polygon, so the point may be coloured as determined by the texture at the interpolated texture coordinate.

3. Conclusion

Our paper focuses on the basis of Bresenham algorithm in graphic interpolation processes. There are doubtless other areas where straightforward interpolations across polynomials can be managed using this technique. It seems to be a reasonable approach to teaching interpolation processes, even though there may be faster algorithms for many of these interpolations. Our approach has opened the newer dimension, focusing on how adaptability can take place with Bresenham algorithm.

4. References

1. Jack Bresenham, *Algorithm for Computer Control of a Digital Plotter* IBM Systems Journal, Volume, Number, 1965, pp. 25-30.

2. J.D. Foley, A. Van Dam, S. Feiner, and J.F. Hughes. *Computer Graphics Principles and Practice.* Addison-Welsey, 2e edition, 1990, 1997.

3. Gay, A.C., *Experience in Practical Implementation of Boundary-Defined Area Fill,* in Earnshaw, R. (ed), *Fundamental Algorithms for Computer Graphics*, Springer- Verlag, Heidelberg, 1985

4. 4.Watt, Alan, *Fundamentals of Three- Dimensional Computer Graphics*, Addison- Wesley, 1989

www.jatit.org

Modeling of Packet Loss and Delay using Multi-Path Diversity

[1]Muhammad Asad Khan, [1]Yasir Irshad Abassi, [2] Imran Baig, [2] S.Tauseef-ur-Rehman
[1]Department of Computer Science, Faculty of Applied Sciences, International Islamic University, Islamabad, Pakistan
[2]Chairman, Department of Telecommunication, Federal Urdu University of Arts, science & Technology Islamabad, Pakistan
asad.khan1@hotmail.com, gallian@hotmail.com, imran_baig_mirza@yahoo.com, stauseef@fuuastisb.edu.pk

Abstract

The quality of voice in IP based networks is highly reliant on Packet loss and delay. Packet loss is the failure of packets to reach the destination [1,2]. The multi-path diversity is a phenomenon to send identical packets from diversified paths. Therefore, if certain packet is lost from one stream, still its identical packet has fair chance to arrive at the destination [3]. We first discuss the modeling of packet loss and delay using Bernoulli and Extended Gilbert Model [4]. We then evaluate the results of both models. We have developed a modified loss model for modeling of packet loss using multi-path diversity. The present research aims to increase the efficiency of Forward Error Correction (FEC) by the modeling of packet loss and delay in the real time communication. The set of examined results from our research shows that multi-path diversity bestows positive effect on the voice quality of VoIP network. The results also validate our model as comparison with actual data results in similar trend.

Key Words: Packet Loss, VoIP, Forward Error Correction, Multi-path Diversity

1 Introduction

Voice over Internet Protocol (VoIP) is the conversion of analog and digital communication into data transmission of digital packets over internet or intranet [5]. Basically, it is the toggling from circuit switching to packet switching. Voice over IP (VoIP) is susceptible to network behaviors, referred to as delay and packet loss, which can degrade the voice application to the point of being unacceptable to the average user [1,2,5].

The loss of the packets mostly occurs due to the congestion or the link failure between the nodes. The clogging with in the nodes mostly occurs due to the limited bandwidth and heavy load of voice and data traffic, which also falls out as packet loss [1,2,4]. The loss or dropping of packets results in highly noticeable performance issues and affects all other network applications [1,2].

The loss of data packets can be recovered either by resending the packets to the destination or by using some recovery algorithm. The data

packets are time independent, so they are easy to recover. On the other hand, voice packets are
based on the time limitations. As a result, if voice packets are lost or delayed, they are hard to recover for the reason that if they cross the play out time then they are futile. Therefore, it creates either the short duration calls or very appalling voice quality [5].

1.1 Multi-path Diversity

In Multi Path diversity, the copies of identical packets are sent over a network to achieve the advantage of uncorrelated packet loss and delay [3]. The advantage of multi-path diversity over single path is that there is less probability of losing the packet. If certain packet is lost from one path, still it has positive probability of reaching the destination from the second path. Another advantage of multi-path diversity is that the extent of the packet loss burst is relatively diminutive. The major shortcoming of multi-path diversity is that it needs superfluous bandwidth on both source and the destination links. In addition the destination node needs further processing to discard the already received packets. Therefore, it creates some extra delay in the arrival of voice packet to the destination, but it is better then losing the packets [3].

1.2 Loss Modeling

The analytical modeling of the packets loss and delay for the assessment of the packet loss needs to be done to hit upon the better model for estimating packet loss [4,6].

1.2.1 The Bernoulli Model

The Bernoulli loss model is based on a geometric distribution. It is the most widely used model and based on simple independent losses. It is very basic modeling algorithm for estimating packet loss. Therefore, often use in modeling of packet loss for IP voice and multicast systems [4,6]. The probability of packet loss in the Bernoulli model is represented by **p**. If there are large number of packets **n** to be transmitted over a network then

the expected number of lost packets is **n*p**. Bernoulli loss model is a two-state process. The one state (State 0) symbolizes a packet loss, and the other state (State 1) stands for a packet reaching the destination. The mean loss probability **p** represents the probability that the current packet is lost given that the last packet was also lost. **q** is the probability that the current packet is arrived provided that the previous packet was also arrived [6].

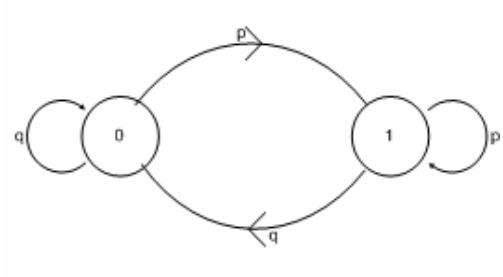


Figure 1.0: The Bernoulli Model

The Bernoulli model gives an estimation of loss probability by calculating the total number of packets that were lost and then dividing the result by the total transmitted packets. In Bernoulli model, each packet transmitted on a network has fixed and independent loss probability with constant loss rate of the link [4,6]. Networks in which time interval between the packets is very short, the loss of packets can not be estimated properly by a Bernoulli model [4,6].

1.2.2 The Extended Gilbert Model

The Bernoulli model maintains the record of all past **n** number of losses to calculate the probability of losing the next packet, where as Extended Gilbert model consider only last **n** number of consecutive loss of packets to calculate the probability of the next packet to be lost. Therefore the probability calculation and loss estimation are in the vicinity of the actual loss of packets. The extended Gilbert model needs **n+1** states to remember **n** events [4].

The Extended Gilbert Algorithm is used with a structure that maintains a counter l, which is the number of consecutive packet loss. But it is reset whenever the packet is received [4]. The extended Gilbert model is the extension of two-state Gilbert model which calculates the burst state with almost the same transition probability as of Bernoulli algorithm by considering burst state as 1 and then according to state transition probabilities P01 and P11 is the probability for the burst length of packet loss.

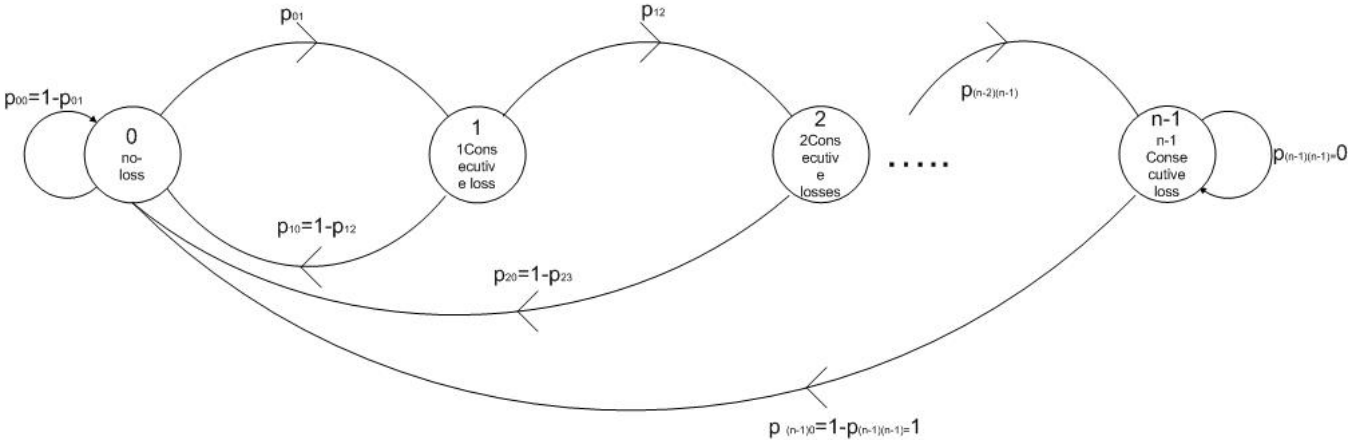


Figure 1.2: The Extended Gilbert Model

2 Literature Review

The work has been done on the algorithms of recovering packets and modeling of voice and data packets loss. The basic modeling algorithm of packet loss and delay are Bernoulli model, Markov model, Gilbert Model and Extended Gilbert Model.

Wenyu Jiang effort was based on the efficiency and perceptual quality of FEC. They started with the modeling of packet loss and delay. Then they proposed the joint use of Extended Gilbert model and inter-loss distance (ILD). Then they implemented the internet traces to validate the necessity and effectiveness of these models. The next phase was the assessment of the consequences of this reliance on VoIP. But their research lack in finding out that how burstiness relays to quality of voice [4].

Yi J. Liang, Eckehard G. Steinbach, and Bernd Girod contributed in the reconstruction of the voice transmission over internet. They worked to improve the tradeoff among delay, late loss rate, and speech quality using multi-stream

transmission of real-time voice over the Internet, where multiple redundant descriptions of the voice stream are sent over independent network paths. Scheduling the play out of the received voice packets is based a Lagrangian cost function to trade delay versus loss. They observed significant reductions in mean end-to-end latency and loss rates as well as improved speech quality when compared to FEC protected single-path transmission at the same data rate. In addition to our Internet measurements, they analyzed the performance of the proposed multi-path voice communication scheme using the *ns* network simulator [3].

Jean Yves Le implemented a joint play out buffer and Forward Error Correction (FEC) adjustment scheme for Internet Telephony that incorporates the impact of end-to-end delay on the perceived audio quality. Their results publicized that it offers a better quality than the adjustment schemes for play out buffer and FEC. This is important because of a threshold effect when the end-to-end delay of interactive audio is around 150 ms. They represented the perceived audio quality as a function of both the end-to-end delay and the distortion of the voice signal. They validated their approach by simulation that their scheme allows a source to increase its utility by avoiding increasing the play out delay when it is not really necessary. But the problem in that development is the static size of buffer, which again creates the loss of the packets and the delay by the bursting of traffic on certain conditions. The problem remains the same, as when the packets cross the size of the buffer, they start to loss or creates the delay [7].

In the earlier research, different loss models were designed to calculate the packet loss and delay for the single path. The emphasis was on estimating the packet loss and delay using diverse loss models. But no one has compared the loss models to unearth the better among them and subsequently implement the resulted better model using multi-path diversity phenomenon to observe any advantage of multi-path diversity for the modeling of packet loss. Mostly the packet loss and delay are due to the congestion or the link failure. Therefore, it requires a technique to recover the lost data. Forward Error Correction (FEC) is use to recover the packets. The snag in FEC for the single path is that it cannot recover lengthy burst of packet loss, as a result all the interlinked packets are also lost, because the information of each packet is added in its subsequent packet. So the sender needs to send the packets again, hence creating more delay.

We have proposed a solution to have better quality of service using the multiple paths. We will compare the extended Gilbert model with Bernoulli model for the single path. The better model will be implemented using the multi-path diversity phenomenon for the modeling of the packet loss and delay. The whole emphasis is on the efficiency of Forward error correction by the modeling of packet loss and delay in the real time communication and to make it more useable to recover the lost packets. By the implementation, we will validate any change in the efficiency of forward Error Correction (FEC) by the introduction of multi path diversity.

3 Methodology

3.1 Estimating Packet Loss:

The implementation for estimating packet loss and delay has been done using network Simulator *ns*. The source of the nodes generates either tcp packets or udp packets to the destination. The tcp packets generating nodes are boasting ftp at the application layer, whereas the cbr is running at the application layer of the udp generating nodes. The capacity of the links is capable enough to handle massive voice and data traffic.

The queues used in the links are SFQ, the stochastic fair queue. The advantage of SFQ over DropTail Queuing is that it does fair queuing. Therefore, the dropping of tcp and udp packets will be same [8].

The first step is the implementation of packet loss modeling using Bernoulli and Extended Gilbert Model in single path environment. The implementation is based on sending the packets form source to destination for specific period of time and calculating total number of voice packets that were sent over the network in single stream. The outcome of the Bernoulli probability and Extended Gilbert probability leads to the estimation of the packet loss using Bernoulli model and Extended Gilbert model respectively. The results will be compared with the actual loss of the packets to obtain the better loss model.

3.2 Bernoulli Loss Model for Single Path

The Implementation was based on the following mentioned steps:

- Design a complex network which comprises of both data and voice traffic from different nodes to the destination nodes.
- Assign starting and ending time to the simulation to extract relevant data during specific time.
- Dispense starting and ending time to all the nodes which are generating the packets.
- Start sending the packets.
- Calculate the sent time, received time and the total delay that packets required to reach the destination.
- Calculate the total voice packets sent by specific source to destination.
- Calculate the actual loss of voice packets for the above mentioned link.

- Calculate the actual length of bursts of packet loss.
- Start calculating the mean Probability using under mentioned method of Bernoulli algorithm.

$$\prod = (\sum_{i=1}^{n-1} Li) / L0$$

Where
- Li is the loss bursts numbers with length i.
- The value of i=1, 2, 3…….n-1. i=1 means that single packet loss whereas n-1 is the longest burst of packet loss.
- L0 is the total number of packets sent from source to the destination [4].

- Now as we have the Mean Probability using Bernoulli model, we need to approximate the Probability of losing the packets for the different lengths of bursts using following probability formula.

$$\prod k = \prod^{k-1} * (1- \prod)$$

Where
- ∏k is the actual probability for estimating the packet loss.
- k is the length of burst of packet loss. e.g. If k=1 that means ∏1 is calculating the probability of single packet loss [1].

- The result of probability of packet loss will lead us to the estimation of packet loss by using following algorithm.

L(est) = Total packet loss * ∏k

Where
- ∏k is the probability taken from the above step.
- L(est) is the estimated packet loss using Bernoulli model of packet loss.

3.3 Extended Gilbert Loss Model for Single Path:

The Implementation was based on the following mentioned steps:

- The steps till the calculation of actual length of burst are same as that of Bernoulli Model.

- Start calculating the mean Probability using under mentioned method of Extended Gilbert Algorithm

$$\mu = 1- (\sum_{j=1}^{n-1} Lj * (j-1)) /(\sum_{j=1}^{n-1} L(j-1) * (j-1))$$

Where
- Lj is the loss bursts numbers with length j.
- The value of j=1, 2, 3…….n-1. j=1 means that single packet loss whereas n-1 is the longest burst of packet loss.
- Now as we have the Mean Probability using Extended Gilbert model, we need to approximate the Probability of losing the packets for the different lengths of bursts using following probability formula.

$$\mu k = (1- \mu(k-1)(k))^{k-1} * \mu$$

Where
- μk is the actual probability for estimating the packet loss.
- k is the length of burst of packet loss. e.g. If k=1 that means μ1 is calculating the probability of single packet loss [4].

- The result of probability of packet loss will lead us to the estimation of packet loss by using following algorithm of extended Gilbert Model.

$$L(est)k = \sum_{j=1}^{k-1} Lj * \mu k$$

Where
- μk is the probability taken from the above step.
- L(est)k is the estimated packet loss using Extended Gilbert model of packet loss.
- k is the length of packet loss for which estimation has been calculated.

3.4 Extended Gilbert Model for Multi-path

The major change in implementing the multi-path diversity over single path is that we designed a network in which there were two streams from the source to the destination for the specific voice communicating link. The duplicate streams of voice packets from source to destination start sending the packets at the same time. All the packets in these streams are duplicate packets of each other. The sending time and the packet fid of duplicate packets are same. Therefore, if there will be no packet loss in the two streams that means the destination node receives every packet twice. So there is a need of discarding the packets which are already arrived from the other stream. When any packet is arrived to the destination, its unique id is compared with all earlier received packets. If the id of the packet is matched with any of the earlier received packet's id that means it is already received and there is no need to receive that

packet again [4]. Therefore, we made a mechanism to drop any packet which was earlier received. The end result by implementing the multi-path diversity was that there were very limited packet losses and even the length of packet loss bursts was not lengthy [3,4].

Rest of the implementation of the packet loss was similar to that of single path implementation.

- Design a complex network which comprises of both TCP and UDP packets that means that the network should comprise of data and voice traffic from different nodes to the destination nodes.
- Assign starting and ending time in the simulation to extract relevant data for the analysis.
- Assign starting and ending time to all the nodes which are generating the packets.
- Start sending the packets.
- Calculate the sent time, received time and the total delay that packet required to reach the destination.
- Calculate the total voice packets sent by some specific source to destination.
- Calculate the actual loss of voice packets for the above mentioned link.
- Calculate the actual length of bursts of packet loss.
- Start calculating the mean Probability using under mentioned method of Extended Gilbert Algorithm.

$$\mu = 1 - \left(\sum_{j=1}^{n-1} Lj * (j-1)\right) / \left(\sum_{j=1}^{n-1} L(j-1) * (j-1)\right)$$

Where

- Lj is the loss bursts numbers with length j.
- The value of j=1, 2, 3…….n-1. j=1 means that single packet loss whereas n-1 is the longest burst of packet loss.
- Now as we have the Mean Probability using Extended Gilbert model, we need to approximate the Probability of losing the packets for the different lengths of bursts using following probability formula.

$$\mu k = (1 - \mu(k-1)(k)) * \mu^{k-1}$$

Where

- μk is the actual probability for estimating the packet loss.
- k is the length of burst of packet loss. e.g. If k=1 that means μ1 is calculating the probability of single packet loss.

- The result of probability of packet loss will lead us to the estimation of packet loss by using following algorithm of extended Gilbert Model.

$$L(est)k = \sum_{j=1}^{k-1} Lj * \mu k$$

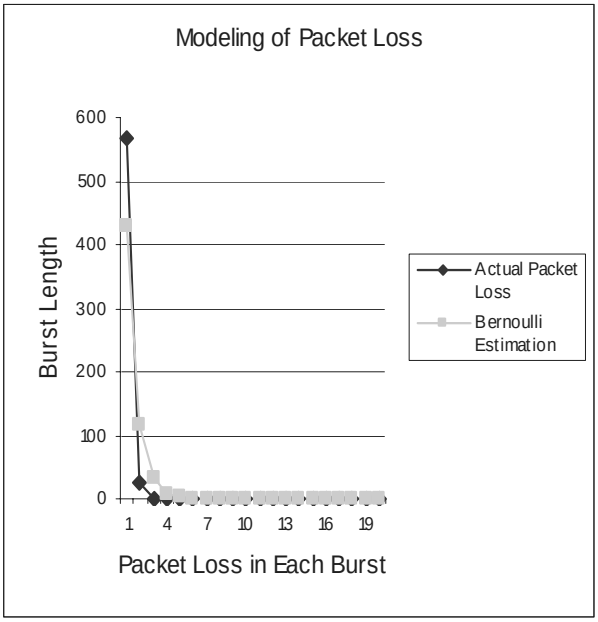
Where

- μk is the probability taken from the above step.
- L(est)k is the estimated packet loss using Extended Gilbert model of packet loss.
- k is the length of packet loss for which estimation has been calculated [4].

Another advantage of modeling of packets using multi-path diversity is that the Forward error correction works better if the bursts of packet loss are not very lengthy. Multi-path packet forwarding gives the advantage to forward error correction to even recover very few lost packets [3,4].

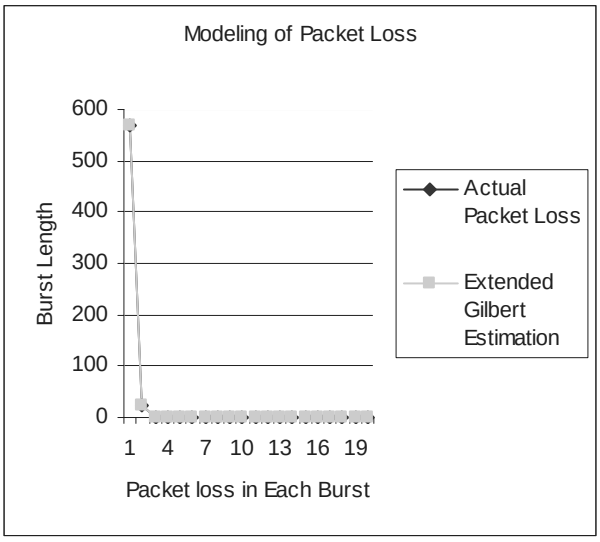
4 Results and Conclusion

The major advantage of Extended Gilbert model over Bernoulli model results is that the probability distribution of every packet depends upon only the **n** consecutive loss packets, whereas in the Bernoulli model of probability distribution the future probability loss of the packet is dependent on all **n** packets. Therefore, the probability in Extended Gilbert model is well calculated to the actual loss of the packets as compare to the Extended Gilbert model [4,6].



Graph 4.0 Bernoulli Loss Model in Single Path (Trace 1)

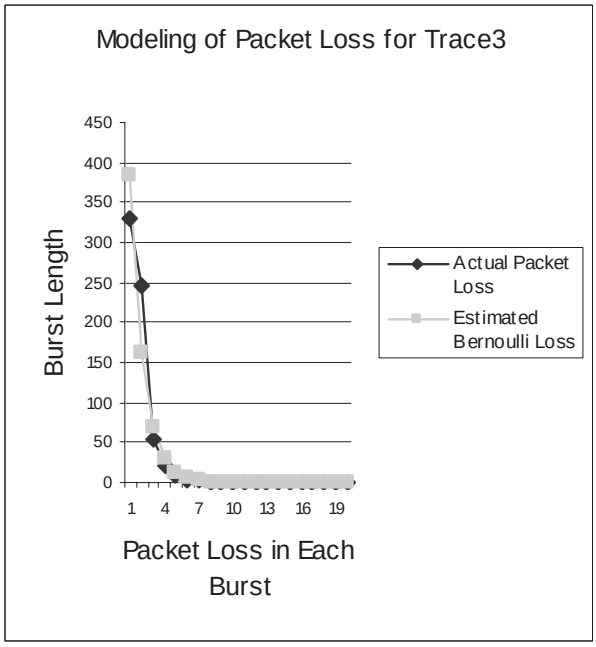
The implementation of packet loss modeling in single path and multi-path using Bernoulli and Extended Gilbert model ensures that the quality of service can be extendable and the voice quality over the packet network can get better and it can be estimate able better.



Graph 4.1 Extended Gilbert Loss Model in Single Path (Trace 1)

The above mentioned graphs prove that if the total packets lost in each burst (except burst length of one and two) are short then Extended Gilbert Model accurately estimates the packet loss to the actual loss of the packets. On the other hand, the Bernoulli loss model over estimates the number of packet loss in every burst, therefore gives results far away from the actual loss of packets.

The following tables present results regarding the simulation for both single path and multi-path phenomenon and the estimation of packet loss using the two loss models.



Graph 4.2 Bernoulli Loss Model in Single path (Trace 3)

Trace	Data Rate	Packet Size	Total nodes	Total Links	Total Links	Delay	Band width	Packets	Packets	Packets Lost
1	1mb	1K	15	15	14	30ms/10	2mb	2237	1646	591
2	1mb	1K	15	15	15	30ms/10	2mb	2237	1605	632
3	1mb	1K	15	15	15	30ms/10	1.5mb	2237	1572	665
4	1mb	1K	15	15	15	30ms/10	1.1mb	2237	1690	547
5	2mb	1K	15	15	15	30ms/10	2mb	4475	XXX	XXX
6	2mb	1K	15	15	14	30ms/10	2mb	4475	3041	1434

Table 4.0 Single Path Network Structure

Trace	Data Rate	Packet Size	Total nodes	Total Links	Total Links UP	Delay	Band width	Packets Sent	Packets Received	Packets Lost

1	1mb	1K	15	15	14	30ms/10ms	2mb	2237	2115	122
2	1mb	1K	15	15	15	30ms/10ms	2mb	2237	2041	196
3	1mb	1K	15	15	15	30ms/10ms	1.5mb	2237	1973	264
4	1mb	1K	15	15	15	30ms/10ms	1.1mb	2237	1865	372
5	2mb	1K	15	15	15	30ms/10ms	2mb	4475	3158	1317
6	2mb	1K	15	15	14	30ms/10ms	2mb	4475	3211	1264

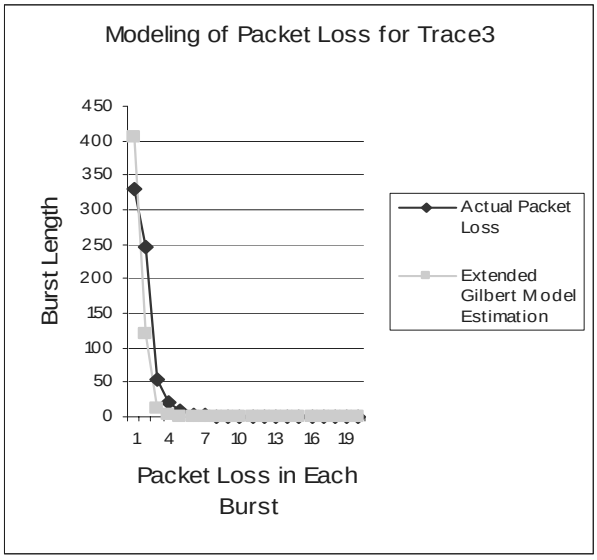
Table 4.1 Multi-Path Network Structure

Table 4.2 Comparison of Loss models using Trace 1

Table 4.3 Comparison of Loss models using Trace 3

Length of Loss Burst	Actual Packet Loss (Single Path)	Estimated by Bernoulli Model (Single Path)	Estimated by Extended Gilbert Model (Single Path)	Actual Packet Loss (Multi-Path)	Estimated by Bernoulli Model (Multi-Path)	Estimated by Extended Gilbert Model (Multi-Path)
0	1572	1572	1572	1973	1973	1973
1	330	385	404	249	223.7	249.8
2	247	162	121	15	26.6	13.4
3	54	68	12	0	3.1	0
4	20	28.6	3	0	0.3	0
5	10	12	1	0	0	0
6	2	5	0.21	0	0	0
7	2	2	0.07	0	0	0

Length of Loss Burst	Actual Packet Loss (Single Path)	Estimated by Bernoulli Model (Single Path)	Estimated by Extended Gilbert Model (Single Path)	Actual Packet Loss (Multi-Path)	Estimated by Bernoulli Model (Multi-Path)	Estimated by Extended Gilbert Model (Multi-Path)
0	1646	1646	1646	2115	1996	2115
1	567	430.3	567.9	121	215.64	121
2	24	116.9	22.1	1	23.4	0.9
3	0	31.8	0	0	2.5	0
4	0	8.6	0	0	0.03	0
5	0	2.3	0	0	0.003	0
6	0	0.6	0	0	0	0
7	0	0.17	0	0	0	0

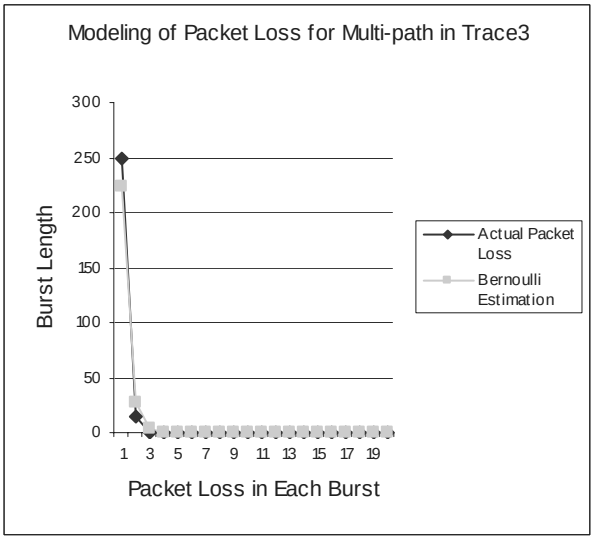


Graph 4.3 Extended Gilbert Loss Model in Single path (Trace 3)

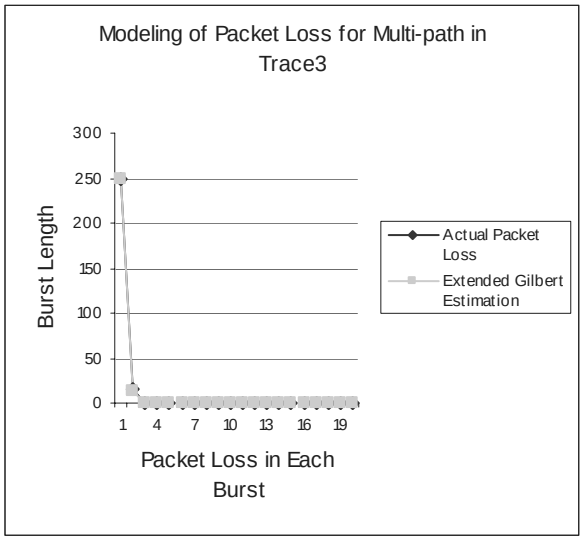
The mentioned tables and graphs give very interesting results regarding the loss models. The results of the research prove that Bernoulli model also provides effective results as Extended Gilbert do, but for some specific Scenarios. The analysis of Table 4.2 and Table 4.3 explains that if the actual burst of packet loss is very lengthy, each burst includes massive frequency of packet losses and comparatively limited number of single packet loss, then Bernoulli Loss model can give better results as compare to Extended Gilbert Model, because it always over estimates the loss of packets, hence close down the estimation of packet loss to the actual packet loss. So we have used both the models for the estimation of packet loss in multi-path diversity as well.

The above mentioned comparison table shows that the multi-path diversity provides better quality of service, as it allows more packets to reach the destination. The simulation proves that using multi-path phenomenon provides packets more options to reach the destination, hence less number of packet loss.

In the above mentioned tables, the length of the lost burst describes the actual length of the lost packets. If the length of the lost burst is zero then it describes the total number of packets that are received at the destination.

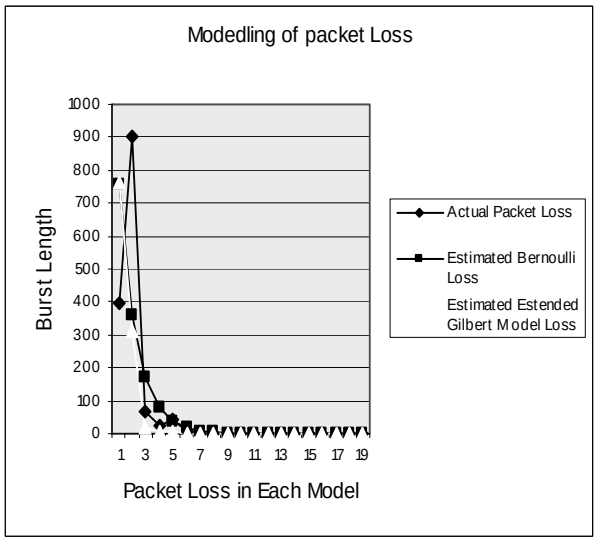


Graph 4.4 Bernoulli Loss Model in Multi-path (Trace 3)

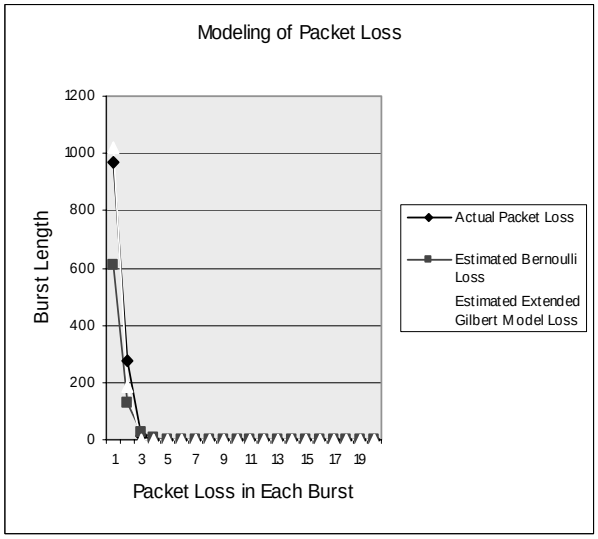


Graph 4.5 Extended Gilbert Loss Model in Multi-path (Trace 3)

The Graphs of Trace3 for the single and multi-path diversity demonstrate that the results of the expected number of Packet Loss by the Extended Gilbert model are very close to the actual loss of the packets only if the number of single packet loss is high. If there is huge number of packet loss with burst length more then one creates under estimation by the Extended Gilbert model. Thus the trace3 graphs prove that multi-path confers better results.



Graph 4.6 Comparison of Loss Models in Single path (Trace 6)



Graph 4.7 Comparison of Loss Models in Multi-path (Trace 6)

The above mentioned results have also shown that the implementation of multi-path diversity has given better quality of service because we have very limited number of actual packet loss. Therefore, the estimation of packet loss using Extended Gilbert model is very near to the real packet loss. The extended Gilbert model is better option for estimating the packet loss specifically for multi-path, as the burst length in multi-path diversity not exceeding then two consecutive packet losses; also extended Gilbert Model estimates the packet loss better then Bernoulli in most of the scenarios.

It is better to use both packet loss models for estimating the packet loss and delay while implementing them in single path and multi path. But if the network is very congested and there is huge probability of losing the lengthy burst of packets in high frequency then Bernoulli model estimates the packet loss more accurately as compare to Extended Gilbert model, juts because Bernoulli model always over estimates the packet loss. Conversely, if the network is less congested and the frequency of packet loss in each burst is limited then Extended Gilbert model provides better results. As Multi-path has less number of packet loss, therefore Extended Gilbert always estimates the packet loss accurately.

5 Future Work

The research is based on the modeling of packet loss using single path and multi-path phenomenon. We plan to research that what will be the advantage of using FEC in Multi-path diversity and its effect of the integration on bandwidth and performance of network in terms of packet loss. We have also plan to integrate both the loss models

for the estimation of packet loss and see their performance.

6 References

1. *Packet Loss*, http://en.wikipedia.org/ wiki/Packet_loss

2. Yoshiaki Hori, *udp packet loss, Effect of TCP Traffic*, http://www.isoc .org /inet97/ proceedings/F3/F3_1.HTM,

3. Yi J.Liang, Eckehard G. Steinbach, and Bernd Girod, *Real-time Voice Communication over the internet Using Packet Path Diversity*, Information Systems Laboratory, Stanford University, Stanford.

4. Wenyu Jiang, Henning Schulzrinne {Wenyu,Schulzrinne} @cs.colombia.edu. *Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality* Department of Computer Science, Columbia University., May 2000.

5. VoIP vs. IP telephony http://vertical.com/voip/voip-vs-ip-telephony.html.

6. Understanding the Bernoulli Model, http://www.commsdesign.com/design_corner/showArticle.jhtml?articleID=18901597.

7. *Jean Yves Boudes*, Adaptive Joint playout Buffer and FEC Adjustment for Internet Telephony, 2001.

8. Manual NS.

9. D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. The case for resilient overlay networks. In Proceedings of the 8th Annual Workshop on Hot Topics in Operating Systems (HotOS-VIII), May 2001. Online at: http://nms.lcs.mit.edu/projects/ron/.

10. J. G. Apostolopoulos. Reliable video communication over lost packet networks using multiple state encoding and path diversity. In Proceedings Visual Communication and Image Processing, pages 392–409, Jan. 2001.

11. M. Arlitt and T. Jin. Workload characterization study of the 1998 World Cup web site. IEEE Network, 14(3):30–7, May 2000.

12. J.-C. Bolot. End-to-end packet delay and loss behavior in the Internet. Computer Communication Review, ACM SIGCOMM ’93, 23(4):289–298, Sept. 1993.

13. J.-C. Bolot, S. Fosse-Parisis, and D. Towsley. Adaptive FEC-based error control for Internet telephony. In Proceedings of IEEE INFOCOM ’99, volume 3, pages 1453–1460, Mar. 1999.

14. V. Hardman, A. Sasse, and A. Watson. Reliable audio for use over the Internet. In Proceedings of INET ’95, pages 171–178, June 1995.

15. R. V. Hogg and E. A. Tanis. Probability and statistical inference. Macmillan Publishing Company, 4th edition, 1993.

16. ITU-T Recommendation G.726. 40, 32, 24, 16 kbit/s Adaptive differential pulse code modulation (ADPCM), Dec. 1990.

17. ITU-T Recommandation P.862. Perceptual evaluation of speech quality (PESQ), an objective method for End-to-end speech quality assessment of narrow-band telephone networks and speech codec, Feb. 2001.

www.jatit.org

Agent Based Framework for Anomaly Detection in Distributed Environment

[1] Hadi Ejaz Ahmed, [1]Muhammad Hafeez, [2]Zaigham Mahmood, [3]M.A.Ansari, [3]S.Tauseef ur Rehman

[1]Department of Computer Science, Faculty of Applied Sciences, International Islamic University, Islamabad, Pakistan

[2] Faculty of Business, Computing and Law, University of Derby UK

[3]Chairman, Department of Telecommunication, Federal Urdu University of Arts, science & Technology Islamabad, Pakistan

hadiejaz@hotmail.com, hafiiez@hotmail.com, z.mahmood@derby.ac.uk, drmaansari@fuuastisb.edu.pk, stauseef@fuuastisb.edu.pk

,

Abstract

In this paper we discuss our research in mounting a distributed architecture for intruder detection. The key inspiration is to use data mining techniques to ascertain patterns of system features that can describe the program and user behavior and to use machine learning algorithm to analyze audit data and to extract features that can distinguish anomalies from legitimate activities. Auditing agents are used in this research to assemble concise and accurate classifiers to detect anomalies. To increases computability we used a Web Service based distributed architecture which will provide high accuracy and low computational cost. WinMine toolkit is used to build the statistical model and to facilitate model construction and incremental updates.

Keywords: Anomaly Detection; Data Mining; Machine Learning; Network Protocol.

1. Introduction

Intrusion detection has become an essential component of computer security in recent years. Security administrators are complementing existing security measures with intrusion detection systems (IDSs) to achieve defense in depth [22].

Most anomaly detection systems are based on hand-crafted signatures that are developed by manual encoding of expert knowledge. These systems match activity on the system being monitored to known signatures of attacks. The major problem with this approach is that these anomaly detection systems fail to generalize to detect new attacks or attacks without known signatures. Recently, there has been increased interest in data mining based approach to build detection model for anomaly detection systems. These models generalize from both known attacks and normal behavior in order to detect unknown attacks. They can also be generated in a quicker and more automated method than manually encoded models that require difficult analysis of audit data by domain experts. Several effective data mining techniques for detecting intrusions have been developed [0, 2, 4, 5, 6], many of which perform close to or better than systems engineered by domain experts [1].

However, successful data mining techniques are themselves not enough to create deployable anomaly detection systems. Despite the promise of better detection performance and

generalization ability of data mining-based anomaly detection systems, there are some inherent difficulties in the implementation and deployment of these systems [1]. We can group these difficulties into four general categories: accuracy (i.e., detection performance), extensibility (i.e., extensible domain), computability (i.e., low computational cost), and *functional-usability* (i.e., practical utilization). Typically, data mining-based anomaly detection systems (especially anomaly detection systems) have higher false positive rates than traditional hand-crafted signature based methods, making them unusable in real environments. Also, these systems tend to be inefficient (i.e., computationally expensive) during both training and evaluation. This prevents them from being able to process audit data and detect intrusions in real time. Finally, these systems require large amounts of training data and are significantly more complex than traditional systems. In order to be able to deploy real time data mining-based anomaly detection systems, these issues must be addressed [1].

In this paper, we discuss several problems inherent in developing and deploying a data mining-based anomaly detection system and present an overview of our research, which addresses these problems. These problems are independent of the actual learning algorithms or models used by an anomaly detection system and must be

overcome in order to implement data mining methods in a deployable system [1].

An effective data mining-based anomaly detection system must address each of these four groups of issues. Although there are tradeoffs between these groups, each can generally be handled separately. We present the key design elements and group them into which general issues they address [1].

1.1 Related Work

Several influential research ADSs were developed from mid-80's to mid-90. STAT [8] and IDIOT [9] are misuse detection systems that use the “signatures" of known attacks, i.e., the patterns of attack behavior or effects, to identify a matched activity as an attack instance. By definition, misuse detection is not effective against new attacks, i.e., those that do not have known signatures. NIDES [10] have an anomaly detection subsystem that uses established normal profiles, i.e., the expected behavior, to identify any unacceptable deviation as the result of an attack. Anomaly detection is capable of catching new attacks. However, new legitimate behavior can also be falsely identified as an attack, resulting in a false alarm. These systems and most of the later research and commercial systems are developed using a pure knowledge-engineering process.

In recent years, there has been several learning- based or data mining-based research efforts in intrusion detection. Warrender et al. [11] showed that a number of machine-learning approaches, e.g., rule induction, can be used to learn the normal execution profile of a program, which are the short sequences of its run-time system calls made. These learned models were shown to be able to accurately detect anomalies caused by exploits on the programs. Lane and Brodley developed machine learning algorithms for analyzing user shell commands and detecting anomalies of user activities [12]. A team of researchers at Columbia University have been working on data mining-based intrusion detection since 1996 (see Stolfo et al. [13] for an overview). The main capabilities developed in this research include: pattern mining and feature construction, cost-sensitive modeling for efficient runtime model execution, anomaly detection, learning over noisy data, and correlation analysis over multiple of data streams. The ADAM project at George Mason University is developing anomaly detection algorithms based on automated audit data analysis [7].

2. Goals

Several important criteria affect anomaly detection systems, we can group these difficulties into four general categories: accuracy (i.e., detection performance), extensibility (i.e., extensible domain), computability (i.e., low computational cost), and *functional-usability* (i.e., practical utilization).

2.1 Accuracy

Crucial to the design and implementation of effective data mining-based anomaly detection system is defining specifically how detection performance, or accuracy, of these systems is measured. Because of the difference in nature between a data mining-based system and a typical anomaly detection system, the evaluation metrics must take into account factors which are not important for traditional anomaly detection systems [1].

At the most basic level, accuracy measures how well an anomaly detection system detects attacks. There are several key components of an accuracy measurement. One important component is detection rate, which is the percentage of attacks that a system detects [1]. Another component is the false positive rate, which is the percentage of normal data that the system falsely determines to be anomalous. These quantities are typically measured by testing the system on a set of data (normal and anomalous) that are not seen during the training of the system in order to simulate an actual deployment.

Data mining-based anomaly detection systems are only useful if their detection rate is higher than a hand-crafted method’s detection rate with an acceptably low false positive rate. Given this framework, our goal is to develop a data mining-based anomaly detection system that is capable of outperforming hand-crafted signature-based systems at the tolerated false positive rate [1].

We have developed and applied a number of algorithms to improve the performance of data mining-based anomaly detection systems. In our research, we focus on a few particular techniques that have been proven to be empirically successful. We first present a generic framework for extracting rules from audit data which help discriminate attacks from normal data, by using data mining algorithms. These rules can then be used to build a detection plan. Finally, we present a method for filtering anomalous data from our audit data using model generated from the detection plan.

2.2 Functional-Usability

A data mining-based anomaly detection system is significantly more complex than a traditional system. The

main cause for this is that data mining systems require large sets of data from which to train. The hope to reduce the complexity of data mining systems has led to many active research areas [1].

First, management of both training and historical data sets is a difficult task, especially if the system handles many different kinds of data. Second, once new data has been analyzed, models need to be updated. It is impractical to update models by retraining over all available data, as retraining can take weeks, or even months, and updated models are required immediately to ensure the protection of our systems. Some mechanism is needed to adapt a model to incorporate new information. Third, many data mining-based anomaly detection systems are difficult to deploy because they need a large set of clean (i.e., not noisy) labeled training data. Typically the attacks within the data must either be manually labeled for training signature detection models, or removed for training anomaly detection models. Manually cleaning training data is expensive, especially in the context of large networks. In order to reduce the cost of deploying a system, we must be able to minimize the amount of clean data that is required by the data mining process [1].

We present an approach to each of these problems. We use the *detection plan* builder technique, which is a generic mechanism for adding new information to a model without retraining. We employ *distributed anomaly detection* which is a new class of anomaly detection algorithms that do not deeply rely on agent hardware strength.

2.3 Computability

We have implemented a distributed system that is capable of evaluating a set of models in real-time without requiring Herculean computability at agent machines. This system uses a network protocol scanner for extracting "primitive," features from raw network traffic data to produce connection records at agent machine, and then posts audit data to web service based server for model evaluation and higher level feature computation. The motivation for posting this audit data for computation and evaluation is that it is quite costly and we do not wish to overburden the agent machine.

We have currently focused this system at TCP/IP protocol suite as the business domain, although SOAP is used for communication between the network protocol scanner and DAD server.

2.4 Extensibility

DAD uses concurrent-path approach for anomaly detection, by using multiple rule-sets. Each rule-set may represent a different domain, obviously having a different model as well. This allows the addition of new rule-sets without effecting existing rule-sets. Thus, providing domain extensibility, without having a need for code extensibility.

3. Methodology and Implementation

The overall system architecture is designed to support a data mining-based anomaly detection system with the properties described throughout this paper. As shown in Figure 1, the architecture consists of DAD Agents running anomaly detection agents and DAD Servers running web services, audit data warehouse, rule generation component, plan building component and a model generation component. This architecture is capable of supporting not only data gathering, sharing, and analysis, but also model generation.

The system is designed to be independent of the network protocol scanner data format and anomaly detection server. A single network protocol scanner data can contain an arbitrary number of features. Each feature can be continuous or discrete, numerical or symbolic. In this framework, we can have as many anomaly detection servers as we will like. To deal with this heterogeneity, an XML encoding is used so each component can easily exchange data and/or anomaly detection server.

The key advantage of our architecture is its high performance and scalability. That is, all components can reside in the same local network, in which case, the work load is distributed among the components; or the components can be in different networks, in which case, they can also participate in the collaboration with other anomaly detection systems in the Internet; or the components can be distributed within the same network as well as the Internet at the same time.

3.1 Anomaly Detection Agents

Anomaly Detection Agents observe raw data on a monitored system and compute features for use in rule evaluation. Anomaly Detection Agents insulate the rest of the anomaly detection system from the specific low level properties of the target system being monitored. It then posts the audit data through the web service interface exposed at DAD Server. Types of Anomaly Detection Agents:

a. Network Monitoring Agent

b. Process Monitoring Agent

c. File System Monitoring Agent

d. Event Log Monitoring Agent

The data warehouse also facilitates the integration of data from multiple scanners. By correlating data/results from different anomaly detection systems or data collected over a longer period of time, the detection of complicated or large scale attack become possible [1].

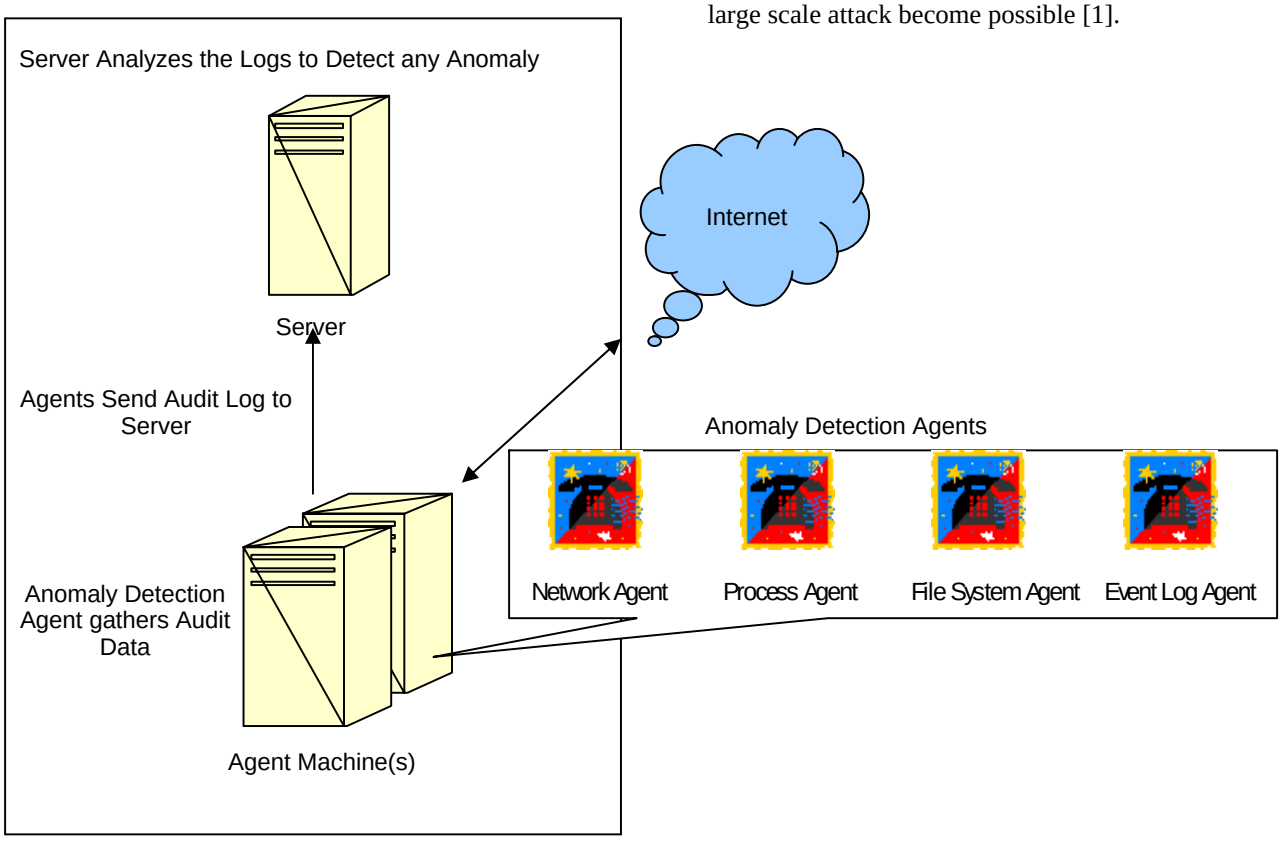


Figure 1 DAD Architecture

3.2 Distributed Data Collector

DAD Server collects audit data from protocol scanner through an exposes web service interface called the 'Data Collector'. Data Collector not only gathers audit data from the agents, but it also arranges data according to information provided by the scanner.

3.3 Data Warehouse

The data warehouse serves as a centralized storage for data, rules, plans and models, although each server can also have a separate data warehouse as well. One advantage of a centralized repository for the data is that different components can manipulate the same piece of data asynchronously with the existence of a database, such as off-line training and manually labeling. The same type of components, such as multiple rule generators, can manipulate data concurrently.

On the other hand, we can perform efficient load-balancing by having different data warehouses for different servers.

But the real beauty of DAD architecture is, that we can have a centralized audit data warehouse for a selected number of protocols, and separate audit data warehouse for others at the same time.

3.4 Data Mining based Rule Generator

DAD Server take audit data from agents and use a data mining based rule generator to evaluate and extract information from a bulky audit data. The generated rules and then sent back to the data warehouse for further analysis and report.

There can be several (or multiple layers of) rule generators at the same system. For example, work loads can be distributed to different rule generators to analyze events in parallel. But, it is always performed behind the scene for effective load balancing.

Aprori Association rule algorithm is used to compute rule which satisfy the minimum confidence requirement. The algorithm used is outline here [15]

```
Begin
(1)  scan database D to form L_1 = {frequent 1-itemsets};
(2)  k = 2; /* k is the length of the itemsets */
(3)   while L_{k-1} ≠ ∅  do begin /* association generation */
(4)      for each pair of l^1_{k-1}, l^2_{k-1} ∈ L_{k-1} and l^1_{k-1} ≠ l^2_{k-1} where
           their first k − 2 items are the same  do begin
(5)        construct candidate itemset c_k such that its first k − 2
             items are the same as l^1_{k-1}, and the last two items
             are the last item of l^1_{k-1} and the last item of l^2_{k-1};
(6)        if there is a length k − 1 subset s_{k-1} ⊂ c_k
             and s_{k-1} ∉ L_{k-1}
           then
(7)           remove c_k; /* the prune step */
           else
(8)           add c_k to C_k;
        end for
(9)     scan D and count the support of each c_k ∈ C_k;
(10)    L_k = {c_k | support(c_k) ≥ minimum_support};
(11)    k = k + 1;
     end while
(12)  forall l_k, k > 2  do begin /* rule generation */
(13)     forall subset a_m ⊂ l_k  do begin
(14)       conf = support(l_k)/support(a_m);
(15)       if conf ≥ minimum_confidence  then begin
(16)          output rule a_m → (l_k − a_m),
              with confidence = conf and support = support(l_k);
           end if
        end for
     end for
end
```

3.5 Data For

As rule gen
effective loa
obviously be
devise a plar
this drawbac

Data Formatter takes raw rules as input and generates Rules in XML format, to server as a basic detection plan, to be used for model generation.

3.6 Detection Planner

Detection Planner takes XML formatted basic detection plan as input, and provides the DAD Server Administrator with a user friendly tool for Anomaly Detection Planning. The detection planner then sends back the result to the data warehouse for further analysis.

3.7 Model Generator

The main purpose of the model generator is to facilitate the rapid development and distribution of new (or updated) anomaly detection models [1]. By using, the detection planner, we have minimized false alarms, by allowing the DAD Server Administrator give his feedback in plan generation. In this architecture, an attack detected first as an

anomaly may have its exemplary data processed by the model generator, which in turn, using the archived (historical) normal, intrusion data sets from the data warehouse and detection plans devised by the administrator, automatically generates a model that can detect the new anomalies

3.8 Anomaly Detection

Once, a model has been generated, it now time for anomaly detection. The DAD Server detects anomalies from different models as well as it allows the administrator to search for anomalies using the same models.

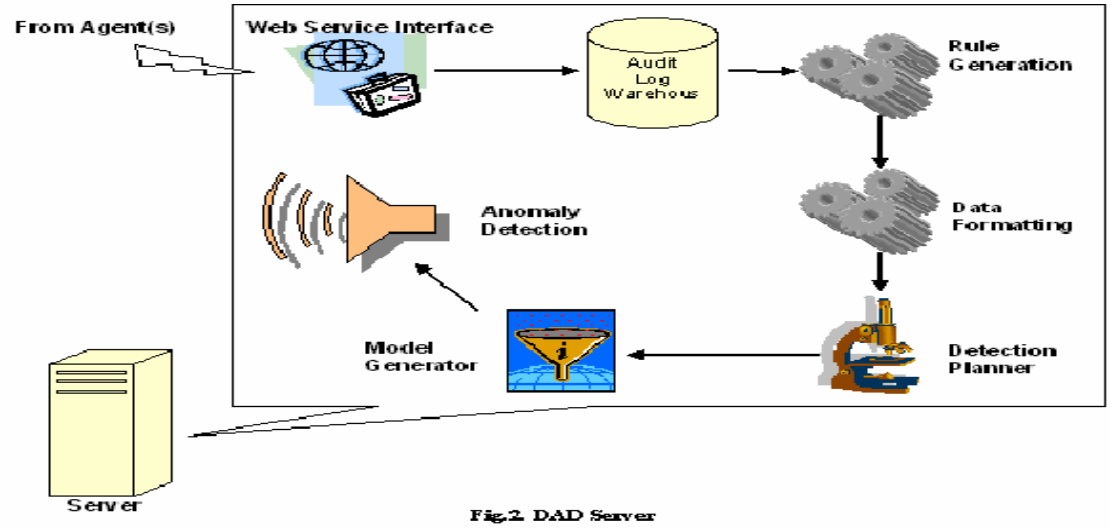


Fig.2. DAD Server

ıve
aly

4. Classification of the Surveyed System according to System Characteristics

Below table shows the classification of the surveyed system according to system characteristics. It is important

to mention that these are some of the attributes of the characters tics and not cover all the functionality.

Name of System	Publ. Year	Time of Detection	Granularity	Type of Response	Security	Interop.
DPEM [16]	1994	Real	Batch	Passive	Low	Low
GRIDS [17]	1996	Non-Real	Batch	Passive	Low	Low
CSM [18]	1996	Real	Continuous	Active	Low	Low
EMERALD [19]	1997	Real	Continuous	Active	Moderate	High
BRO [20]	1998	Real	Continuous	Passive	Higher	Low
DAD	2006	Real*	Continuous	Passive	Moderate	High

1

DPEM[16] , CSM[18] utilizes host based auditing for anomaly detection where as BRO[20] utilizes network auditing but in DAD like GRIDS[17] we have utilized the concepts of agents running on both host as well as network these agents are called Auditing Agents. The use of auditing agents gives us more flexibility and more diverse traffic. This can be elaborated more from the figure 4.1

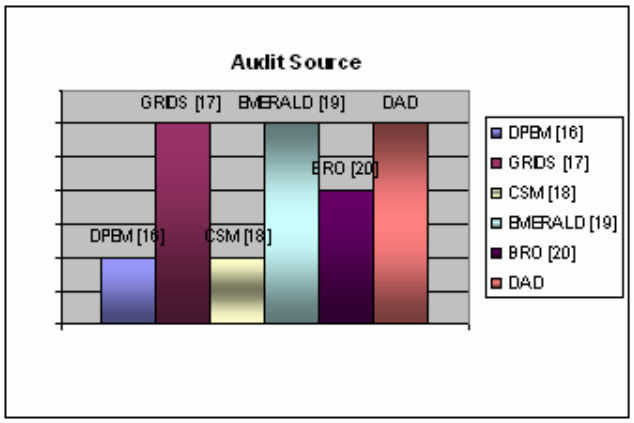


Figure 4.1

Part of detection is done in real time on the host computers through the set of predefined rules these rules consist of events regarding the security and system events of the local system. At this point we have only monitored events from the following logs

Application log: The application log contains events logged by applications or programs.

Security log: The security log records events such as valid and invalid logon attempts, as well as events related to resource use such as creating, opening, or deleting files or other objects.

System log: The system log contains events logged by Windows system components.

Directory service log: The directory service log contains events logged by the Windows directory service. For example, connection problems between the server and the global catalog are recorded in the directory service log.

File Replication service log: The File Replication service log contains events logged by the Windows File Replication service. For example, file replication failures and events that occur while domain controllers are being updated with information about sysvol changes are recorded in the file replication log.[23]

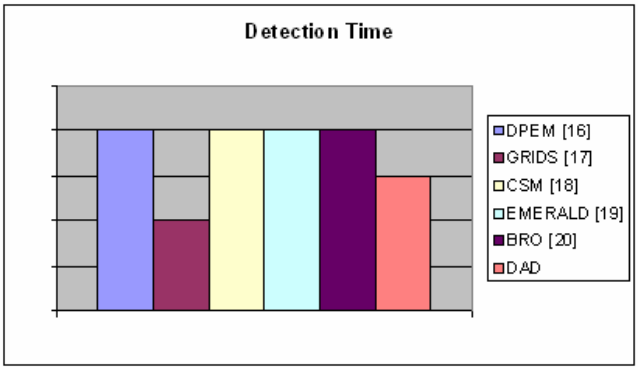


Figure 4.2

DAD support granular discovery and reaction capabilities so that specific attacks on a specific host can be detected and a specific response can be enacted this refined level of granularity produces supplementary control on the protection policy.

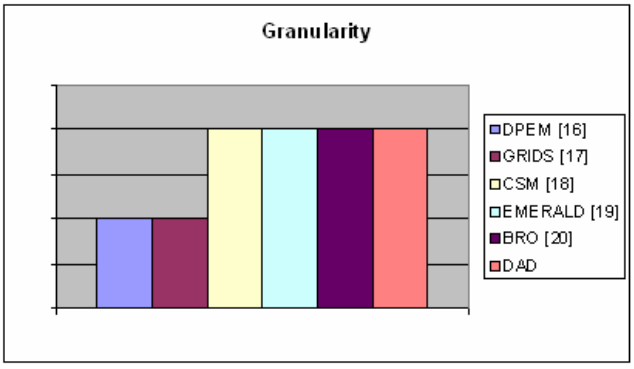


Figure 4.3

Like DPEM [16], GRIDS [17] and BRO [20], DAD response time is passive.

[1] * Part of the detection is made at real time on the host computer.

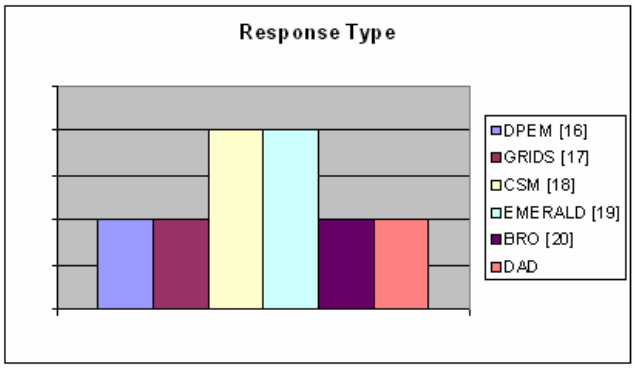
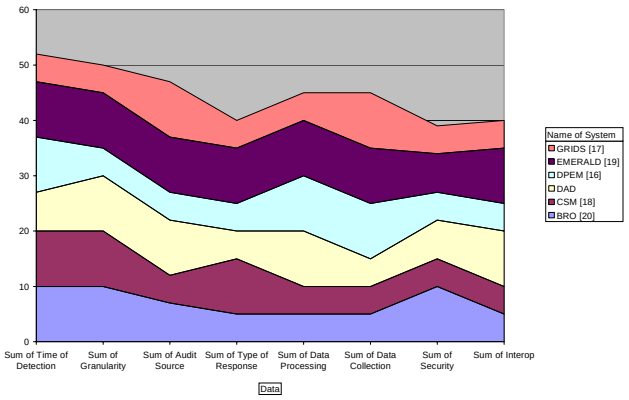


Figure 4.4

There are also many other factors which could be used to compare DAD with other IDS. Figure 4.5 shows the stacked area chart which display the trend of the contribution of each value over different categories.



4.2 Conclusion

Anomaly detection – useful if data set is linearly separable without more complicated processing and right user specified thresholds [3].

We have tried to build up a cost effective and extendable solution for intrusion detection and by utilizing the Asynchronous web service architecture we have tried to provide Interoperability between different OS/platforms.

The focus of this work is to improve the network anomaly detection process, especially data processing by using multiple servers running web services. Widespread adoption of a framework such as that described in this paper will permit anomaly detection by consuming web services from many parties, while individual developers can benefit from the ability to leverage others work while moving forward independently. By opening this field previously largely restricted to major anomaly detection vendors. It is hoped that the industry as a whole can progress forward more rapidly.

5. Future Work

There is currently no formal reasoning behind intrusion detection techniques. We don't know a priori whether a method would work. And when an anomaly detection model fails, we can't prove that it is the fault of the modeling algorithm, the lack of “meaningful" data, or both. We are limited by the current state of design and implementation practices of operating systems and network services. There are no (or very little) formal specifications for our computing components and the interactions among them. Since we have no way of reasoning the possible behavior of a system, we can only observe, which involves a lot of imprecise guesswork. We need to address these problems by working with the operating system and networking communities. Although it is unlikely that a system component will be designed and implemented using formal methods in every single step, we can require that its test plans (i.e., what are its intended and well-tested usages) be released to the ADS community, along with a specification of its auditing mechanisms (what can or cannot be recorded, and from which data sources). We can then reason the best possible models for such a system component [14].

6. References

1. Wenke Lee , Salvatore J. Stolfo, Philip K. Chan, Eleazar Eskin, Wei Fan, Matthew Miller, Shlomo Hershkop, and Junxin Zhang, “Real Time Data Mining-based Intrusion Detection,” In DARPA Information Survivability Conference and Exposition II. June 2001.

2. E. Eskin, “Anomaly detection over noisy data using learned probability distributions,” In Proceedings of the Seventeenth International Conference on Machine Learning (ICML-2000), 2000.

3. Dae-Ki Kang, Doug Fuller, Vasant Honavar, “Learning Classifiers for Misuse and Anomaly Detection Using a Bag of System Calls Representation,” This paper appears in: Systems, Man and Cybernetics (SMC) Information Assurance Workshop, 2005. Proceedings from the Sixth Annual IEEE.

4. A. Ghosh and A. Schwartzbard, "A study in using neural networks for anomaly and misuse detection." In Proceedings of the Eighth USENIX Security Symposium, 1999.

5. W. Lee, S. J. Stolfo, and K. Mok, "Data mining in work flow environments: Experiences in intrusion detection," In Proceedings of the 1999 Conference on Knowledge Discovery and Data Mining (KDD-99), 1999.

6. C. Warrender, S. Forrest, and B. Pearlmutter, "Detecting intrusions using system calls: alternative data models," In In1999 IEEE Symposium on Security and Privacy, pages 133–145. IEEE Computer Society, 1999.

7. Wenke Lee , Wei Fan, "Mining System Audit Data: Opportunities and Challenges," SIGMOD Record 30(4): 35-44 (2001) ISSN:0163-5808

8. K. Ilgun, R. A. Kemmerer, and P. A. Porras. State transition analysis: A rule-based intrusion detection approach. IEEE Transactions on Software Engineering, 21(3):181{199, March 1995.

9. S. Kumar and E. H. Spafford. A software architecture to support misuse intrusion detection. In Proceedings

of the 18th National Information Security Conference, pages 194{204, 1995.

10. D. Anderson, T. Frivold, and A. Valdes. Next-generation intrusion detection expert system (NIDES): A summary. Technical Report SRI-CSL-95-07, Computer Science Laboratory, SRI International, Menlo Park, California, May 1995.

11. C. Warrender, S. Forrest, and B. Pearlmutter. Detecting intrusions using system calls: Alternative data models. In Proceedings of the 1999 IEEE Symposium on Security and Privacy, May 1999.

12. T. Lane and C. E. Brodley. Temporal sequence learning and data reduction for anomaly detection. ACM Transactions on Information and System Security, 2(3):295{331, August 1999.

13. S.J. Stolfo, W. Lee, P.K. Chan, W. Fan, and E. Eskin. Data mining-based intrusion detectors: An overview of the Columbia IDS project. ACM SIGMOD Record, 30(4), December 2001.

14. Wenke Lee, "A Data Mining Framework for Constructing Features and Models for Intrusion Detection Systems," ACM Transactions on Information and System Security (TISSEC). Volume 3, Issue 4 (November 2000) ISSN: 1094-9224

[15] Wenke Lee, Salvatore J. Stolfo, Kui W. Mok

"Algorithms for Mining System Audit Data" Data mining, rough sets and granular computing (Pages: 166 - 189) Year of Publication: 2002

ISBN: 3-7908-1461-X

16. Calvin Ko, George Fink, and Karl Levitt. Automated detection of vulnerabilities in privileged programs by execution monitoring. In *Proceedings of the 10th Annual Computer Security Applications Conference*, volume xiii, pages 134-144. IEEE, IEEE Computer Society Press, Los Alamitos, CA, USA, 1994.

17. S. Stani ford Chen, S. Cheung, R. Crawford, M. Dilger, J. Frank, J. Hoagland, K Levitt, C. Wee, R. Yip, and D. Zerkle. GrIDS-A graph based intrusion detection system for large networks. In *Proceedings of the 19th National Information Systems Security Conference*, 1996.

18. G. White and V. Pooch. Cooperating security managers: Distributed intrusion detection systems. *Computers & Security*, Vol. 15(No. 5):441-450, 1996. Elsevier Science Ltd.

19. Philip A Porras and Peter G Neumann. EMERALD: Event monitoring enabling responses to anomalous live disturbances. In *Proceedings of the 20th National Information Systems Security Conference*, pages 353-365, Baltimore, Maryland, USA, 7.10 October 1997. NIST, National Institute of Standards and Technology/National Computer Security Center.

20. Vern Paxon. Bro: A system for detecting network intruders in real-time. In *Proceedings of the 7th USENIX Security Symposium*, San Antonio, TX, USA, January 1988. USENIX, USENIX Association. Corrected version, 7 overstated the traffic level on the FDDI ring by a factor of two.

21. WinMine Toolkit

http://research.microsoft.com/~dmax/winmine/tooldoc.htm

22. J. Allen, A. Christie, W.Fithen, et al., "State of the Practice of Intrusion Detection Technologies," CMU/SEI-99-TR-028

http://www.sei.cmu.edu/publications/documents/99.reports/99tr028/99tr028abstract.html

23. Microsoft Windows XP Professional Product Documentation.

http://www.microsoft.com/resources/documentation/windows/xp/all/proddocs/en-us/event_overview_01.mspx?mfr=true

Indian Information Technology Industry: Past, Present and Future& A Tool for National Development

Somesh.K.Mathur[1]

Abstract

The present study examines the growth performance of India's IT industries, with particular attention paid to the role of policy in this process. The study recognizes that emergence of a strong Indian IT industry happened due to concerted efforts on the part of the Government, particularly since 1980s, and host of other factors like Government-Diaspora relationships, private initiatives, emergence of software technology parks, clustering and public private partnerships. In this study we further look at the major parameters of the Indian IT industry and give justification for including the main factors responsible for the IT boom in India. The study has looked into the past and present trends of the Indian IT industry and has considered further needs of IT sector to act as a catalyst of growth and development. The study has examined whether the Indian IT growth does have enough lessons for other countries to model their IT policy which may help them to shape their IT industry as driver of growth and development.

Keywords: Indian Information Technology, Offshore Outsourcing, Software Exports, E-Commerce, E-Governance

Introduction

Riding high on the outsourcing wave, India is likely to witness software and services exports growth of 25-28% clocking revenues of $36-38 billion in fiscal year 2007. IT- ITES (Information Technology enabled services) exports are likely to grow by 27-30% in FY 06-07, posting revenues between $29-31 billion, according to National Association of Software and Service Companies (Nasscom), which stated that exports for FY 05-06 had risen 33% to register revenues worth $23.6 billion as compared with export revenues of $ 17.7 billion in FY04-05.FY 05-06 also saw the overall Indian IT-ITES industry (including domestic market) grow by 31%, revenues of $29.6 billion up from $22.5 billion in 04-05.

Over a period of time, India has established itself as a preferred global sourcing base in these segments and they are expected to continue to fuel growth in the future. These segments have been evolving over the years into a sophisticated model of operations. Indian IT and ITES companies have created global delivery models (onsite-near shore-offshore), entered into long term engagements with customers, expanded their portfolio of services offerings, built scale, extended service propositions beyond cost savings to quality and innovation, evolved their pricing models and have tried to find sustainable solutions to various issues such as risk management, human capital attraction and retention and cost management. A key demand driver for the Indian IT services and ITES industry has been the changing global business landscape which has exerted performance pressures on multinational enterprises. The IT industry and IT-enabled services, which are rapidly growing offer opportunities for FDI as well. India has emerged as an important venue for the services sector including financial accounting, call centers, and business process outsourcing. There is considerable potential for growth in these areas. Biotechnology and Bio informatics, which are on Government's priority list for development, offer scope for FDI.

[1] Fellow ,RIS. Email: someshmathur@ris.org.in,som3@vsnl.com.First Draft. Most of the figures in the paper are from NASSCOM and International Data Corporation(IDC)

The industry has crossed $I billion dollar mark, with a growth rate of 36.55 per cent**,** compared to only 8 percent for the economy as a whole in 2005. Software exports accounted for 20% of Indian export revenues in 2003-04.By 2008 it would account for 7% of India's GDP and would contribute 30% of total Indian export revenues. The IT sector is likely to give employment to 9 million people in India by 2008 and also generate $87 billion in annual revenues & $225 billion in market value by 2008 (see McKinsey Report, quoted in the Department of Information Technology webpage http://www.mit.gov.in/dbid/eproduction.asp-appendix Table I). In addition to the nearly 1.3 million-strong workforce employed directly in the industry, Indian IT-ITES[2] is estimated to have helped create an additional 3 million job opportunities through indirect and induced employment. Indirect employment includes expenditure on vendors including telecom, power, construction, facility management, IT, transportation, catering and other services. Induced employment is driven by consumption expenditure of employees on food, clothing, utilities, recreation, health and other services. Against the level of $9.5 billion achieved in 2002-03, software and IT services exports are expected to grow to $87 billion by 2008. While the software export target is set at $50 billion, the target for export of hardware has been kept at $10 billion by 2008.India's share in the overall global software market is expected to increase from the present 2 per cent to 7 per cent by the terminal year of the Tenth Plan.

India's strength has emerged through large client wins, cross border mergers and acquisitions, and the movement of the industry towards a stable pricing model. With low costs no longer being the deciding factor for foreign companies looking for developing software in India, research, chip design and financial analytical modeling are some of high-end services increasingly coming to India. In an industry which has been one of the flywheels of robust economic growth in India gaining a reputation for being able to handle complex contracts, the country top firms are now looking at large-sized, multi-year orders to boost revenue stability. Asia's third-largest economy has become a hub for global firms like Motorola Inc. and International Business Machines Corp. for services such as handset software and supply chain management. India's large English-speaking engineering workforce and cheaper wages of nearly one-fifth of western salaries have helped to attract outsourcing[3]. The top three Indian software exporters, flagship Tata Consultancy Services Ltd, Infosys Technologies Ltd. and Wipro Ltd., each boast more than a billion dollars in annual revenue. The United States is the biggest market

[2] [2] Information technology essentially refers to the digital processing, storage and communication of information of all kinds. IT can be defined as computing and telecommunication technologies that provide automatic means of handling information. IT includes software and hardware. The role of IT in services i.e., IT-led services includes the following : Product support Process outsourcing Hardware & software maintenance Training & education IT outsourcing System integration & application development .Therefore, IT can potentially be used in every sector of the economy. The true impact of IT on growth and productivity continues to be a matter of debate, even in the United States, which has been the leader and largest adopter of IT. However, there is no doubt that the IT sector has been a dynamic one in many developed countries, and India has stood out as a developing country where IT, in the guise of software exports, has grown dramatically, despite the country's relatively low level of income and development. An example of IT's broader impact comes from the case of so-called IT-enabled services, a broad category covering many different kinds of data processing and voice interactions that use some IT infrastructure as inputs, but do not necessarily involve the production of IT outputs. IT enabled services include call centres, medical transcription, back office operations, revenue accounting, insurance claim processing ,legal database, payroll records, logistic management, content development/animation, entertainment software, graphics and design, computer animation, among others

[3] Outsourcing is a business strategy that many corporations have used for decades. The most common processes that carriers are outsourcing are software development, system maintenance, core systems hosting and other systems hosting. In the last decade, the IT and BPO industries have seen substantial offshoring. India has been the leading offshore destination during this period, and now accounts for 65 per cent of the global industry in offshore IT and 46 per cent of the global Business Process Offshoring (BPO) industry. Traditionally, off shored functions have involved IT enterprise applications, infrastructure management, and IT support. However, recently we have increasingly seen off shoring extending to more mainstream functions such as HR back office, payroll, benefits administration, and inbound call centers. New entrants in this market include recruitment and staffing functions, tax planning, budgeting and reporting, risk management, and also the operations planning and scheduling functions. According to a report on global sourcing published by Gartner, the leading provider of research and analysis on the global IT industry, India remains the undisputed leader in terms of offshore centers, with China and Russia emerging as strong contenders. Countries such as India, Malaysia, and the Philippines are popular because they offer well-educated, English-speaking workers at a fraction of what they would cost in the US, and are therefore reaping rich dividends. Other countries benefiting from outsourcing are Mexico (where the first jobs from the US moved), Bangladesh and Sri Lanka. India holds the upper hand with a perfect time zone difference that enables a 24/7 service to US firms. The areas of greatest global sourcing expansion over the next three years are expected to be Eastern Europe and Southeast Asia.

for Indian software firms, accounting for as much as 70 percent of revenue. The domestic software sector is dominated by ready-to-use products and packages, which account for 40 per cent of the market followed by projects, around 30 per cent. Domestic companies account for less than 20 per cent of the total market, indicating a high demand for imported products.

It is a surprise that India has been able to achieve as much as it has in IT development. Many scholars have termed this as leapfrog development. Even data on information and communication indicators tend to support the hypothesis. The number of personal computers per 1000 people in India in 2002 at 7.2 is just about the average of 6.9 for low income countries and a fourth of China's 27.6.Internet users per 1000 people at 17 is just about the average of 16 for low income countries, but still a quarter of China's 64.India spent 3.7% of GDP on IT compared to China's 5.3%(World Bank World Development Indicators on CDROM,2005).This points out to the concerted policies and vision by the government and industry to promote software exports and transfer of technology and telecommunication links in 1980s and particularly in 1990s.

Since 1984 under the rule of Prime Minister Rajiv Gandhi, India has been pursuing liberalization policies that have helped the IT industry develop. More specifically the computer policy of 1984 and 1986 policy on computer software export and software development and training[4] gave a much needed fillip to the software industry. On the other hand, the People's Republic of China provided little state support for this endeavor until the late 1990s. Now lagging behind, China is trying to catch up by replicating India's model.

The government recognizes the significant economic opportunity that the information technology (IT) explosion represents to India and is committed to the policies, infrastructure development and education investment to maintain the growth. The Government of India is providing for more liberal policy framework for the IT sector. As stated above one of the major factors of excellent and consistent growth of Indian software industry can be attributed to continuous liberalization of policies of the Government of India. NASSCOM and the government have worked together in close co-operation over a long time for forming and implementing these policies. During 1991, NASSCOM lobbied with the Government and for the first time, secured income tax exemption from profits of software exports. Later, Government, systematically and gradually, reduced import duty on computer software from a high 114 percent to nil. Copyright laws were also amended.

The Ministry of Information Technology is meant to act as a nodal institution for the promotion of the sector, facilitating and coordinating the various initiatives of the central and state governments and the private sector. Priority is given to e-governance, development of software in Indian languages, IT for the masses, distance education, e-commerce, cyber security and HRD. Postgraduate education and research in IT is pursued for promoting R&D in the emerging areas of Bluetooth technology[5], e-commerce, and nano-technology and bioinformatics solutions. Foreign investment in the sector is encouraged by simplifying policies and strengthening and upgrading telecommunication and IT infrastructure. Establishing an interface of computers with diverse Indian languages. The endeavor will be to develop suitable software and technologies to enable the people to use computers in local languages. Attempts to take IT to the masses will be accelerated by promoting Internet accessibility[6], content creation in local languages, IT applications for various disabilities, empowerment of the masses with special thrust on

[4] The policies reflected in various government documents emphasized that : Effective software export promotion on a sustained basis can be effective in the long run only if it is planned as a part of an overall software promotion scheme covering both export and internal requirements including import substitution. Also planning for software development is integrally connected with the plan for hardware development and system engineering (Government of India,1985).The policies ,for example, called for the setting up of a separate Software Development Promotion Agency under the erstwhile Department of Electronics. The import of inputs needed for software development was also made more liberal.

[5] Bluetooth connectivity — one that helps you talk handsfree with a wireless headset, or send data from one phone to the other wirelessly. Bluetooth is not a technology owned by one company, but by a conglomerate called the Bluetooth Special Interest Group (SIG), led by promoter companies comprising Microsoft, Intel, Nokia, IBM, Motorola, Toshiba, Ericsson, and Agere. Bluetooth-enabled products include printers, accessories, presentation systems, MP3 players, and mobile phones.

[6] India has an estimated 40 million Internet users, making it the country with the fifth-largest number of Internet users. Yet, that number only represents 3.6 per cent of the total population. The USA, with over 100 million Internet users, has a penetration rate of 68.7 per cent and Australia, with approximately 14 million users, has a penetration rate of 68.2 per cent.

women and children, rural healthcare systems, digital libraries in order to preserve the country's cultural heritage and social identity. Enrolment in Indian technology schools is expecting to reach 600,000 by 2008. The government has set a target of 20 million broadband users by 2010. The Indian federal and state governments are committed to developing and broadening e-governance. Fourteen state governments have IT-specific priority policies and many have implementing IT-related projects. Thus, we see that India's proactive government played an instrumental role in encouraging the IT industry. The present study examines the growth performance of India's IT industries, with particular attention paid to the role of policy since 1970s in this process. **The study recognizes that emergence of a strong Indian IT industry**[7] **happened due to concerted efforts on the part of the Government, particularly since 1980s, and host of other factors like Government-Diaspora relationship, private initiatives, emergence of software technology parks, patterns of spatial agglomeration in the IT sector and public private partnerships**. In this study we further look at the major parameters of the Indian IT industry in particular and give justification for including the main factors responsible for the IT boom in India. The study will look into the past and present trends of the Indian IT industry and consider further needs of IT sector to act as a catalyst of growth and development. The study will also examine whether the Indian IT growth does have enough lessons for other countries to model their IT policy which may help them to shape their IT industry as driver of growth and development A number of studies have looked into different aspects of India's IT software export boom(Schware,1987,1992;Sen,1995;Heeks,1986; Kumar,2000a,b,2001;Arora et.al,2000;Joseph and Harilal,2001;Kumar and Joseph,2005;Parthsarthi and Joseph,2002;Joseph,2002).Arora and Athreye(2002) argue that the software sector has contributed to Indian economic performance well beyond the macroeconomic indicators of GDP share, employment and foreign exchange earnings. In particular, they argue that software companies have come to represent models of good corporate governance that other enterprises can and increasingly emulate. This creates productivity spillovers to other sectors. Among the practices the authors cite are: (i) increased investment in staff training;(ii) incentive pay linked to corporate performance;(iii) flat hierarchies and team organization, designed to encourage knowledge sharing(iv)IT sector in promoting entrepreneurship

INDIAN IT INDUSTRY: Study of the Past till Present

1960s and 1970s: Indigenization and Self Sufficiency

India was motivated to try to develop self-sufficiency in computers and electronics largely by national security concerns related to border conflicts with China and Pakistan. The government created an Electronics Committee which devised a strategy for achieving self-sufficiency in electronics within ten years by "leapfrogging" ahead to absorb the most advanced products and technologies available. The goal was eventually to achieve indigenization of technology, whereby India would move away from dependence on foreign technology and produce its own. This approach not only responded to the perceived security risks, but also fit the ideology of self-sufficiency which drove much of India's post-independence political and economic agenda.

The main vehicle chosen to gain access to advanced computer technologies was negotiation with multinationals, primarily IBM, which dominated the computer market in India (from 1960-1972, IBM accounted for over 70% of all computers installed in India). From 1966 to 1968, the Indian government tried to get IBM to share equity with local capital in its Indian operations.

[7] Indicators of the strength of India's software export capabilities include the depth of its base, and the breadth of its global reach. There are over 2,500 Indian software exporters, and while only the top five (TCS, Infosys, Wipro, Satyam and HCL) are – or are approaching the status of – global brands, they together account for only about 35% of software exports. The United States remains by far the largest market for India's software exports, its share of India's software exports being 63%, with Europe coming in at 26%, and Japan and the rest of the world accounting for the remaining. Going forward, the more traditional IT outsourcing service lines such as hardware and software maintenance, network administration and help desk services will account for 45 per cent of the total addressable market for offshoring and are likely to drive the next wave of growth. While the addressable market for the global offshore IT and BPO industries is quite large, industry evolution will largely be shaped by the interplay of three major forces: (1) supply (the capacity and quality of offshore locations); (2) demand ramp-up (realistic adoption of offshoring by companies); and (3) industry conduct (the actions taken by industry players.

IBM said it would leave India before agreeing to equity sharing, and the government let the matter drop.

In an attempt to satisfy the government's interest in developing domestic production, both IBM and British-owned ICL began to refurbish used computers in Indian plants and sell or lease them to Indian customers. IBM felt that India should evolve technologically from one level of sophistication to the next. However, a 1966 report by the government's Electronics Committee stated that such step-by-step technological evolution should be avoided and that India should leap ahead to the latest technologies. But at this point, the government was unable to impose its will on IBM. The government's early attempts to regulate the IT sector actually worsened the degree of technological backwardness as Indian users installed the domestically refurbished machines rather than importing newer models.

The government's inability to effectively regulate the MNCs was due partly to institutional weaknesses in the agencies assigned the task. In 1966, responsibility for implementing the Electronics Committee Report strategies had been given to the Department of Defense Supplies, with monitoring by a new agency, the Electronics Committee of India. However, the committee lacked support staff and had no authority to compel action by other agencies. This lack of authority and technical competence left the government unable to negotiate with the MNCs or to regulate the IT sector effectively.

By 1971, the Department of Defense Supplies had a backlog of over 150 license requests for IT projects. After much criticism of the Department by other agencies and the private sector, the government announced the formation of a Department of Electronics and a new Electronics Commission. The Commission was responsible for policy formulation and oversight and the Department was responsible for day-to-day implementation of policies.

The Electronics Commission was given authority to direct other government units and to regulate private and public electronics enterprises, and it developed a professional staff capable of providing the necessary technical support to effectively regulate the sector. In 1975, the Department of Electronics was given power over the licensing of computer imports. The new Committee and DOE had the authority and capability to establish control over the development of IT in India and they did exactly that.

One of the first steps taken was the establishment of the Santa Cruz Electronics Export Processing Zone (SEEPZ) near Bombay. Foreign and Indian investors were offered incentives to establish an export base in India, including tax breaks, cheap land, duty-free import of inputs, and a streamlined permit process. In return, the government required that all or most of the production be exported and that Indian components be used as much as possible.

A second step was the creation of the state-owned ECIL (Electronics Corporation of India Ltd.) as a national champion in minicomputer production. ECIL got almost all of the government's computer development funding and the DOE made it very difficult for private competitors to get operating licenses. The government's plan was to allow imports of mainframes and large minis, give the small mini market to ECIL, and allow private firms to compete in the micro sector. Thanks to this support, ECIL's market share ranged from 40% to 53% of the computer installations in India between 1973 and 1977. However, by the end of the decade, ECIL had failed to make a computer that was technologically sophisticated, price competitive or which could be delivered on time.

The third action taken by the Electronics Department and Commission was to once again challenge the position of the multinationals. Using FERA regulations, the government began to pressure IBM and ICL to dilute their equity to 40% in their Indian operations. ICL agreed to combine its two Indian operations and reduce its equity to 40%, but IBM refused.

Negotiations with IBM went on through 1976 and 1977, but before they took place, two important developments occurred. In 1975, U.S. computer maker Burroughs entered into a joint venture with Tata Consultancy Services to export software and printers from SEEPZ. This meant the government had two MNCs (ICL and Burroughs) in the country on its own terms, which probably encouraged the government to take a hard line toward IBM.

Also in 1975, the Indian cabinet approved a proposal to set up the state-owned Computer Maintenance Corporation (CMC) with a legal monopoly on the maintenance of all foreign computer systems in the country. This reduced the

advantage IBM had with users as a result of its superior service capabilities. Now users would have to depend on CMC no matter whose system they purchased.

With its bargaining position substantially enhanced, the government continued to demand that IBM dilute its equity to 40% for all Indian operations. IBM responded with a proposal to share equity in its non-computer operations, meet export goals, and fund an Indian science center and an electronics testing facility. The government refused. After two years of negotiations, IBM decided it could not back down on the equity issue and in 1978 it quit India altogether.

IBM's exit was a seminal event, and illustrated the extent of the government's ability to exert its power over multinational corporations and direct the development of the IT industry in India. The question which naturally arises is why the government chose a showdown strategy with IBM. It seems that the Indian government did not originally set out to drive IBM away, but felt that it could not allow IBM to be exempt from the FERA without jeopardizing its ability to negotiate with other multinationals and implement its nationalistic policy objectives(Dedrick and Kraemer,1993).

One effect of IBM's departure was to open up the market to a number of competitors, including ECIL, ICL, and the Tata-Burroughs joint venture. ECIL dominated the market for a time, thanks to strong government support, but by the end of the 1970s, local private firms such as HCL, DCM and ORG had emerged to control most of the market

The decline of ECIL was partly due to its own inability to produce competitive products, but it was exacerbated by changes in policy. The DOE had come under criticism in the late 1970s for blocking the efforts of private sector firms to produce hardware and for protecting ECIL at the expense of users and domestic competitors. The government responded by giving permission to several private companies such as HCL, DCM and ORG to produce data processing systems and import parts and components. Soon these companies had supplanted ECIL as the major computer suppliers to the Indian market.

The 1970s and 1980s – Software Exports

During the 1950s and 1960s, there was no Indian software industry. Software came bundled with hardware provided by multinational hardware companies like IBM(from the US) and ICL(from UK). IBM's unbundling of software from hardware in the late 1960s is seen as a generic global catalyst for the existence of independent software firms (*Financial Times* 1989).

In the 1970s too, there was no separate software industry. Multinationals such as IBM and ICL were the largest providers of hardware to the industry, which used to be bundled with the operating systems and a few basic packages that were generally written in FORTRAN and COBOL languages.

Larger enterprises (including the Indian defense and public organizations) that needed customized applications employed in-house teams that did everything from installing systems to writing software. In fact, when specific software applications became popular, stand-alone boxes were made for them. In 1970s, the concept of stand-alone word processing software did not exist. Later, when local companies grew (after IBM's exit in early 1980s), these companies also had their own proprietary operating systems that generally executed only their computer programs.

India exported its first software services and products in the mid-1970s.Although India was among the first developing nations to recognize the importance of software, the key driver behind exporting software was foreign exchange. To export software, Indian companies had to design it for hardware systems that were the standard worldwide, which in the 1970s were the IBM mainframe computers. However, Indian import duties on this hardware were extremely high (almost 300 percent) and hence during the late 1960s and early 1970s, IBM used to sell old, refurbished and antiquated machines (because that is all that Indian companies could afford). Fortunately, within a few years, the Indian Government lowered import duties on all IT equipment but with a pre-condition that the exporters would recover twice the value of the foreign exchange spent on importing computers within five years – a clause that was modified in the 1980s. Hence, overall, the regulatory scenario was not very favorable for software exporters and this constitutes the beginning of the Indian software industry.

The first software exporting company from India was Tata Consulting Services (TCS) that started operations in 1968. Fortunately, after a few local orders, TCS bagged its first big export assignment

in 1973-74, when it was asked to provide an inventory control software solution for an electricity generation unit in Iran. During this period, TCS had also developed a hospital information system in UK along with Burroughs Corporation (which was at that time the second-largest hardware company in the world) and it became a role model for other Indian IT companies to follow in the 1980s.

Despite the tough policy with respect to imports, by early 1980s, India was the only developing nation to have any significant software exports – USD 12 million – a substantial leap over the 1979 level of USD 4.4 million and 30 companies were already beginning to export software.

The main competitive advantage for Indian companies was obviously the cost and the ability to communicate using the English language. The total charges for a software developer in India varied from USD 16,000 to USD 24,000 annually whereas the corresponding charges of sending the same developer to the US varied between USD 32,000 and USD 42,000 annually. Comparing this to the total cost of a US software developer (USD 60,000 to USD 95,000 yearly) in 1980, the savings were clearly quite significant.

Inspite of the cost advantages and a relatively good proficiency in English, the Indian software industry continued to face the following challenges in 1970s and 1980s:

• Lack of availability of hardware: Import of hardware – especially mainframe computers was very tedious and expensive.

• Shortfall in trained manpower: Although the education system was producing substantial number of engineers who were very talented, very few colleges were offering any computer training or IT courses.

The following three unrelated incidents contributed heavily in shaping the Indian IT industry(Sarla V Nagala,2005):

• In late 1970s the Indian Government passed a controversial law (which was later repealed in 1992) that forced all multinationals to reduce their equity share in their Indian subsidiaries to less than 50 percent. Since IBM did not want to reduce its equity in its subsidiary, it decided to leave India, thereby, making Indian companies less reliant on mainframe computers.

• The advent of Personal Computers in 1980s reduced the cost of importing hardware substantially, thereby, spawning an industry that has over 2,700 companies today.

• Realizing that the Indian college system was unable to provide any computer training or IT courses, three Indian entrepreneurs (living in India) took it upon themselves to provide tutorials and training classes in Information Technology. Their early days were often marked with one person driving a scooter or a motorcycle and the other riding behind with a PC in his lap so that they could impart this training in rented college and school spaces in the evenings. The training institute (NIIT) started by them is today a USD 167 million company and it continues to be number one in providing IT courses and training to Indians . Infosys, Satyam, Mastek, Silverline and Polaris, among numerous others, were started by software professionals and engineers with small savings and loans at very modest scales to begin with (Kumar 2001).N R Narayan Murthy acknowledges that several private banks refused to fund the setting up of Infosys, and it was the Public Sector financial institution that came forward and gave the seed capital. The example shows the critical role of government support in generating local entrepreneurs in the initial stages of a high technology industry.

With these as the humble beginnings, the Indian IT industry witnessed the Indian Government policies becoming more favorable in late 1980s, representative industry associations getting formed (one of which eventually became NASSCOM – the National Association of Software and Service Companies) and the IT training and education level gradually becoming strong enough for creating a full-fledged industry.

Finally, in the initial years, export of software initially meant a physical transfer – either of the programmer himself -sometimes called 'body-shopping(the provision of labor intensive ,low value added programming services, such as coding and testing at client sites') or of software on floppies. However, in 1985, Texas Instruments (TI) set up an office in Bangalore with a direct satellite link to the US and, in 1989, an Indian Government Telecom Company (VSNL) commissioned a direct 64-kbps satellite link to the US, thereby, offering software exporters a completely new way of functioning.

In terms of products and services, there have been continuous exports of software products since the early 1980s. These include enterprise systems, design software, and database management tools. However, such exports have consistently formed less than about 5% of total exports. Indian software exports have been, and remain, dominated by services .

Within the overall segment of software services exports, though, trends of change are detectable. Indian firms began with a strong emphasis on 'bodyshopping' In the late 1980s, around 75% of export earnings came from bodyshopping. By the early 2000s, this had dropped to nearer 60% (Dataquest 2001), indicating a slow but steady trend towards offshore working. The absence of reliable telecommunication links in 1980s forced Indian firms to be primarily " body shoppers" ,who provided programming services on site, typically in the US, to customers under contract.

This has been paralleled by a second trend: that of moving up the value chain from supply of programming services to addition of design/analysis services to complete turnkey project services. As with offshore working, the trend of change has been greater within individual client—vendor relationships than in the industry overall.

The 1990s – The Emergence of Offshore Outsourcing

In 1993, the US Immigration and Naturalization Service made changes that made it difficult to get B-1 visas and the new H-1 visa required a certification from the US Department of Labor that prevailing market wages were being paid to immigrant workers. As a result, US companies had less incentive to hire software engineers from India. Also, Indian software professionals who were brought under the umbrella of the Immigration Act, had to pay social security and related taxes to the US government, which added additional burden on the employees and the companies.

The two factors mentioned above led a few IT companies in India to gradually move to a mixed model, wherein some software programmers would work at the Client's premises (in the US) whereas others would continue to work in the IT company's back-office in India. As the Indian IT industry adapted to this new business model, Indian IT exports boomed from USD 128 million in FY 1990 to USD 485 million in FY 1994. It is worth pointing out that the shift to the new business model was gradual because the savings even after sending Indian IT programmers to the US were quite large and many IT companies continued to follow the old model and send their programmers to the US, the UK, and Canada.

And then came the 'Y2K problem', the Internet-Telecom boom and the Dot.com boom. All these forced companies in the US, UK, and Canada to hire lot of computer programmers and this caused such a shortage in the US that the US government had to increase its H-1 quota from 65,000 in 1998 to 130,000 in 1999 and then to 195,000 soon thereafter. Indeed, this was a very good opportunity for the Indian IT industry, which thrived by sending more and more IT professionals to the US, thereby creating a larger and larger Indian IT Diaspora.

In particular, the 'Y2K problem' presented a unique opportunity to Indian firms. Owing to this problem, the US firms needed software professionals with COBOL programming skills. COBOL had already become obsolete in 1990s and was no longer a part of university curriculum in the US. However, in India, COBOL was still taught, even in the 90s, since most of the local computer science curriculum was quite obsolete. This provided significant advantage to Indian IT services vendors, particularly because working on Y2K contracts helped Indian firms in entering new markets and building trust with their client enterprises.

Also, the use of alternative operating system since mid 1980s, UNIX , gave the Indian entrepreneurs better environment to deal with the Y2K problem. In fact, the 1986 report of the Rangarajan Committee on Modernization of India's largely state owned banking sector recommended standardized banking systems on UNIX . Subsequently government floated a tender for 400 UNIX systems setting off a scramble among Indian companies to come up with UNIX platform. . Later, in the 1990s,UNIX turned out be ideal for networked computing, and UNIX based systems still dominate the Internet server realm

By the end of 1999, the Indian IT industry was on an all-time high and the Initial Public Offerings (IPOs) of Indian software companies (in India) were getting oversubscribed. This, in turn, led to the creation of a venture capital industry in India.

Significance of Outsourcing Business & Millennium Years Performance of Domestic Market

While producing hardware in the 1980s was part of the manufacturing sector, the high technology jobs of the 1990s and present require a sophisticated enough skill set to write software and maintain computer systems. Only a few select countries have a ready supply of workers who are both technically trained and proficient in English to accept the opportunity American companies offer. For such reasons, China, Russia, and

Vietnam are also prime locations; India, however, by far has become the leader of what has come to be known as the "outsourcing" revolution, as it captures a commanding 70% of the total spending on outsourcing

Outsourcing has been defined by two types of activities: (1) foreign companies launching "liaison, project, or branch" offices in India that retain the name of the founding corporation; and (2) foreign companies contracting out stages of their production processes to already-formed Indian companies as "a joint venture or wholly-owned subsidiary."
It is important to distinguish between these two types of outsourcing because the requirements that foreign companies pursuing offices in India must meet differ significantly from those placed upon multinational partnership firms. These types of offices are limited in scope and Indian law specifically prohibits branch offices of foreign companies from carrying out manufacturing activities on its own. Rather, it encourages the subcontracting of these manufacturing tasks to established Indian manufacturers.

This transnational work is made possible by technology. High-speed data connections and software tools have allowed for great distances to be bridged, making possible the collaboration between geographically disparate groups. This technology also changed the structure of the production process; rather than a few large vertically-integrated corporations in which hardware and software are produced together, a "more fragmented industrial structure" now allows for production processes to be performed in different locations. Global communication has thereby assisted the growth of the IT industry.

Domestic Market

India has emerged as the fastest growing and the fourth largest IT market in Asia Pacific, according to an IDC(International Data Corporation) study. The result has been that – for many years – India has been the developing world's software leader. There are few large firms that control much of the exports of the Indian Software industry.The top five firms account for 32 % of total software exports.The IT industry is concentrated in TN, Karnataka and AP. Almost 90% of the software development and export activity are confined to four metropolitan areas in India namely Mumbai, Banglaore, Chennai and Delhi but slowly and steadily increasing in other cities as well. The Indian software industry has grown at a compound annual rate of over 50% in the 1990s,the highest for any country during this period. The revenues have risen from $ 175 billion to $ 8.7 billion during the decade. Indian nationals account for 45% of HI visas issued by the USA every year and a large proportion of them go as software engineers. India is home to some 650000 software developers or about 10%b of the world's developers population. The Indian software developer population is growing at an annual compound growth rate of 32% which means that in next three years the Indian developers will be the highest in the world. Among the Fortune 500 companies over 250 outsource their software's related work to India.

The industry has grown in depth and scope. It is no longer confined to producing and exporting low-end software products and services. Several multinational companies (MNCs), including many leading ones, have established software development centers in India. DataQuest (2004) reports that such MNC centers are filing for patents in large numbers It suggests that intellectual property revenues would constitute a major chunk of a software company's revenue in the future and Indian companies (other than MNCs), including some of the large ones, have not yet started preparing for it. Leading Indian IT firms, such as Infosys and Wipro, are multinational and have offices around the world and employ nationals in these countries. Infosys has alliances with the world's leading firms, including IBM, Intel, Microsoft and Oracle, and also has made strategic acquisitions of foreign firms.

NASSCOM (2004, p. 9) documents the increasing maturity of the industry following a large number of mergers and acquisitions It noted that traditional IT service players have added ITES-BPO portfolios to their existing offerings in order to provide a complete umbrella of end-to-end services. Multi-vendor and build-operate-transfer (BOT) contracts which offer customers advantages such as low risks, scalability and competitive pricing have increased. Indian vendors (IVs) are expanding the spectrum of their service offering in client locations and even setting up facilities in other low cost ITES-BPO destinations such as China and the Philippines in order to tap these markets. They are also moving up the value added ladder to offer high-end services such as equity research and analytics, insurance and technology support and development.

Moreover, Indian vendors have moved far beyond call centers into financial services, telecom, retailing and automotive segments of the ITES-BPO sector. In financial services, Indian companies are offering customers services centered around accounting, billing and payment services and transaction processing. Over the past few years, some Indian service providers have also been offering higher value services to customers in the areas of insurance claims processing and equity research support. They expect to gain from offshore-outsourcing: of customer and technical support and product development by global telecom industry; of process of transaction processing, billing, telemarketing and inventory management of large retailers; and of engineering activities, such as computer aided product and tool

design, claims processing and accounting processes of automobile industry (NASSCOM 2004, p. 10).

The report also benchmarked the performance of Indian industry on key operational issues with global benchmarks. It finds that Indian industry is able to deliver at levels comparable to their international counterparts on parameters such as quality, customer satisfaction and people satisfaction. Complementing the continued growth in IT-ITES exports is a steadily evolving domestic market. (See Table I below)

Table I: IT Industry-Sector-wise break-up(% GROWTH)

USD billion	FY 2004	FY 2005	FY 2006E
IT Services	**10.4**	**13.5**	**17.5**
-Exports	7.3	10.0	13.2
-Domestic	3.1	3.5	4.3
ITES-BPO	**3.4**	**5.2**	**7.2**
-Exports	3.1	4.6	6.3
-Domestic	0.3	0.6	0.9
Engineering Services and R&D, Software Products	**2.9**	**3.9**	**4.8**
-Exports	2.5	3.1	3.9
-Domestic	0.4	0.7	0.9
Total Software and Services Revenues	**16.7**	**22.6**	**29.5**
Of which, exports are	**12.9**	**17.7**	**23.4**
Hardware	**5.0**	**5.9**	**6.9**
Total IT Industry (including Hardware)	**21.6**	**28.4**	**36.3**

Total may not match due to rounding off

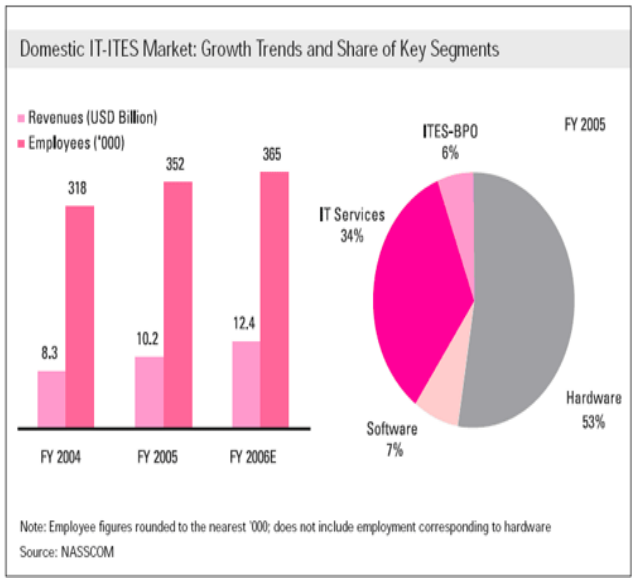
** NASSCOM estimates have been reclassified to provide greater granularity*
- Revenues from Engineering and R&D services and Software Products reported separately (erstwhile clubbed with IT Services / ITES-BPO)
- Historical values for a few segments have changed
- For ease of comparison, details for two preceding years have been restated as per the new classification

As stated above strong demand over the past few years has placed India amongst the fastest-growing IT markets in the Asia-Pacific region. While hardware still accounts for a majority share, with spending on services and the outsourced model gaining noticeable traction, growth in the domestic market is witnessing the early signs of service line depth that characterizes maturing markets.

BFSI, Telecom, Government and Manufacturing are the key vertical markets driving growth across categories including hardware systems, networking, storage, security, enterprise application products and related services. Education and healthcare are a few emerging areas expected to drive additional growth. ITES-BPO demand in the domestic market, though at a nascent stage with contact centre activities for customer care, and sales and marketing accounting for over two-third of current demand, is also witnessing increased levels of activity. BFSI, telecom and consumer durables are the early adopters of ITES-BPO in the domestic market and currently account for nearly three-fourths of the business in this space.

Recognizing its potential, leading global players (Indian as well as MNC) are also focusing some of their attention towards tapping the domestic market - with significant success. Revenue aggregate earned from the domestic market by the leading, predominantly export-focused Indian service providers have grown and several of the key IT outsourcing contracts awarded in the past year were won by MNCs. Global product companies are also looking to introduce localized versions of their software products to drive usability and penetration. This specific focus on the domestic business opportunity is helping create an environment of healthy competition in the industry that augurs well for the development of the domestic market.

As depicted in the following chart, the domestic IT-ITES market was valued at USD 10.2 billion in FY 2004-05 and is expected to exceed USD 12.4 billion, growing at nearly 22 per cent in the fiscal (FY 2005-06).



The observed growth in the domestic market reflects the strength of the Indian economy, which has grown at an annual rate of nearly 7 per cent since 2002 and at more than 8 per cent over the first three quarters of the fiscal year (FY 2005-06).

Future of IT industry

In India, the software boom started somewhere in the late 1990s. Most of the Indian software companies at that moment offered only limited software services such as the banking and the engineering software. The business software boom started with the emergence of Y2K problem, when a large number of skilled personnel were required to fulfill the mammoth database-correction demand in order to cope up with the advent of the new millennium

The profile of the Indian IT Services has been undergoing a change in the last few years, partly as it moves up the value chain and partly as a response to the market dynamics. Ten years ago, most US companies would not even consider outsourcing some of their IT projects to outside vendors. Now, ten years later, a vast majority of US companies use the professional services of Indian Software engineers in some manner, through large, medium or small companies or through individuals recruited directly.

The market competition is forcing organizations to cut down on costs of products. The professional IT services on the other hand are becoming increasingly expensive. The offshore software development model is today where onsite professional services were ten years ago. There is a high chance that in less than ten years, the vast majority of IT services (software development being just one of them) from developed countries, will be, one, outsourced and two, outsourced to an offshore vendor. The Indian IT software and services industry is maintaining a steady pace of growth. Software development activity is now not confined to a few cities in India. Software development centers, such as Bangalore, Hyderabad, Mumbai, Pune, Chennai, Calcutta, Delhi-Noida- Gurgaon, Vadodara, Bhubaneswar, Ahmedabad, Goa, Chandigarh, Trivandrum are all developing quickly. All of these places have state-of-the-art software facilities and the presence of a large number of overseas vendors. India's most prized resource is its readily available technical work force. India has the second largest English-speaking scientific professionals in the world, second only to the U.S. It is estimated that India has over 4 million technical workers, over 1,832 educational institutions and polytechnics, which train more than 67,785 computer software professionals every year. The enormous base of skilled manpower is a major draw for global customers. To become a global leader in the IT industry and retain that position, India needs to constantly keep moving up the value chain, focusing on finished products and solutions, rather than purely on skill sets and resumes. India needs to be able to package their services as products, rather than offering them as raw material

It was earlier feared that the Indian software industry might witness a slow-down in the post-Y2K period, but corporate results and other industry data point to robust growth in the future.

The net cost of hiring Indian programmers is still less than one-third the equivalent cost in either Europe or North America, though there has been a gradual bridging of the salary gap between India and developed countries. Demand for Indian software professionals is increasing rapidly in existing and new client countries. For instance, the US is likely to absorb further 50,000 software engineers every year from India once the proposed Bill to increase visa quotas for technical workers is passed by the US Congress, and both Germany and Japan are now seeking to hire as many as 30,000 yearly workers.

Rapid growth of Internet will create a whole new category of demand in software and allied services in future, such as in internet service applications, web design, internet call centers, validations systems, yellow pages and data mining. India is well positioned to capture a significant share of global business in these areas, even with its relatively low-tech skill base.

Domestic demand will perhaps become another key driver of future growth. Increasing penetration of IT and Internet, pro-active government support and changes in technology will all combine to lower costs in computer hardware, software, telecom and Internet

access in the short term. Personal computer sales are already registering double-digit growth in segments such as Government, Insurance, Banks, Public Tax System, Education and Small Office Home Office. India is expected to cross the one million PC shipment mark in 2000 and will have a total base of over 12 million PCs by 2006. These developments will create large domestic demand over the next few years for programmers, training institutes, web designers, system administrators and network engineers. In fact, domestic IT revenue is already increasing faster than exports and currently accounts for almost 30 percent of total industry turnover. Over the next 2-3 years, exports will continue to be the main pillar of the Indian software industry but domestic sales will also attain critical mass in importance and macro-economic impact.

Based on current and future trends, we expect overall revenue growth in software and allied services to be in the region of 40-50 percent annually for the next five years.

Studies of IDC points out that India will be a potential star in bioscience field in the coming years after considering the factors like bio-diversity, human resources, infrastructure facilities and government's initiatives. According to IDC, bioscience includes pharma, **Bio-IT** (bioinformatics), agriculture and R&D. IDC has been reported that the pharmaceutical firms and research institutes in India are looking forward for cost-effective and high-quality research, development, and manufacturing of drugs with more speed. Bioinformatics has emerged out of the inputs from several different areas such as biology, biochemistry, biophysics, molecular biology, biostatics, and computer science. Specially designed algorithms and organized databases is the core of all informatics operations. The requirements for such an activity make heavy and high level demands on both the hardware and software capabilities.

This sector is the quickest growing field in the country. The vertical growth is because of the linkages between IT and biotechnology, spurred by the human genome project. The promising start-ups are already there in Bangalore, Hyderabad, Pune, Chennai, and Delhi. There are over 200 companies functioning in these places. IT majors such as Intel, IBM, Wipro are getting into this segment spurred by the promises in technological developments.

E-Governance

E-government is the application of Information and Communication Technology (ICT) by government agencies. Its use promises to enhance the effectiveness and efficiency of government and alter its relationship with the public. E-government is evolving through four stages: from posting information to a two-way communication, and from exchange of value to an integrated service and exchange. Introduction of Digital Governance is a way to ensure that common citizens have equal right to be a part of decision-making processes which affect them directly or indirectly, and influence them in a manner which best improves their conditions and the quality of lives. The new form of governance will ensure that citizens are no longer passive consumers of services offered to them and would transform them to play a decisive role in deciding the kind of services they want and the structure which could best provide the same. ICT can influence the process of Governance in various ways and in varying degrees, from improving the current mechanisms of delivery of services to transforming the entire mechanism and the nature of services themselves. The role could be:

ð Technical role, in terms of automation of tedious tasks earlier done by humans.

ð Facilitating role, leading to participatory and all encompassing decision-making and implementation processes.

ð Innovative role, involving new services and mechanisms to deliver these

Computerisation of the Indian Railways' Passenger Reservation System may be
branded a success (Heeks, 1996). Not only did it significantly increase the efficiency
of the reservation process, but it also reduced corruption (though did not eliminate it),
increased rail staff morale, and improved the quality of customer service. Beyond these
reform components, it also gave Indian Railways (and India more widely) a more
modern image, and it helped to build information age capabilities within the country.

E-Parliament as a tool for fostering parliamentarian networks

The reasons why ICTs and the new media environment matter for parliaments and governments alike are the following : the declining confidence in political institutions, including legislature and laws calls for the establishment of a Parliament and government's public relation strategy; then, this new dynamically changing environment will require new strategies for political planning and action. It is quite obvious that the press and electronic media make a contribution to making political decision making process more transparent as well as simplifying the political and legislative information to the ordinary citizen.

Electronic Commerce

Increasingly the development of electronic marketplaces will become an intrinsic feature of any government's success in a global economy. Electronic commerce allows efficient interactions among customer, suppliers and development partners cutting down on transaction time and reducing the costs of doing business.

The role of government will be to enable its business community to obtain the most valuable information and apply it in a timely manner to the production and sale of goods and services. However there are a number of legal issues related to E-commerce which need to be immediately addressed. Issues relating to taxation of goods and services traversing over electronic networks have to be resolved without further delay.

The Ministry of Commerce is supporting "Electronic Commerce (EC) / Electronic Data Interchange (EDI) for Trade" project for facilitating international trade. The community partners of this project are various trade regulatory and facilitating agencies like the Customs Department, the Directorate General of Foreign Trade (DGFT)Ports, Airports, the Reserve Bank of India (RBI), Export Promotion Organisations (EPOs), Exporters, Importers, Agents, Container Corporation of India (CONCOR) and Banks. The objectives of this project are to (i) facilitate electronic delivery of services; (ii) simplify procedures; (iii) provide 24 hour access to users with their partners; (iv) make procedure transparent; (v) reduce the transaction cost and time, and (vi) introduce international standards and best practices.

Overall Forecast and Internet Revolution

From a macro perspective, the IT & Internet revolution in India is quite real. With a whole array of knowledge-based skills and legacy to draw upon, India is very well suited to integrate Internet into its industry, public institutions and education system. Perhaps more so than any other country in Asia, Internet will have a profound impact on India's progress towards a more open and accessible business environment. However, the potential of Internet in the corporate sense of profit making has being overstated, and many pure web-based Indian businesses are likely to fail. But even though Indian society is not yet ready to adapt to the 'new economy' in a consumer sense, the increasing ubiquity of Internet technology will create new global opportunities for India on the supply side, and especially in software. There will very likely be an acceleration in the pace of domestic IT penetration, and this will help in increasing productivity and efficiency in the larger economy. At a minimum, continued growth in software exports and inbound investments will provide a comfortable source of hard currency, which in turn will act as a hedge against any adverse changes in India's balance of merchandise trade. Even more, the direct and indirect impact of all Internet-related benefits could be as high as an extra 1 to 1.5 percentage point GDP growth over the medium term.

In the near future, India may become a front runner in the age of information revolution to be a Global IT superpower. The following are the some of the emerging trends in IT services:

(a)Providing access to Internet through authorised cable TV without additional licensing

(b)Provision of *'last mile'* linkages either by fibre optic or radio communication for IT application enterprises, IT promotional organisations and ISPs

(c)Setting up of *spread spectrum* based non-interference type wireless data/multimedia communication equipment viz., public wireless, subject to a maximum of 4 Watts EIRP with a radio frequency band in the range of 2.4-2.483 GHz

(d) Public Tele Info Centres (PTIC) having multimedia capability specially for ISDN services, remote database access, Government and community information systems, market info, desk top videoconferencing, tele info and internet/web access services

(e)Promotion of *'Hi-tech Habitats'* of high quality in the rural hinterland of the cities

(f)Technological advancements in wide area computer-communication networks which may result into *'Virtual Technology Parks'* where IT software and IT services are developed through on-line integration of software and services sub-systems from widely separated locations in the country

(g)Using *International Credit Cards (ICC)* to paying for IT software and IT services purchased over internet or extranets and also for registering domain names

(h)Maximum flexibility in organising the marketing of package software from India through internet

(i)Advanced network to enable organisations and companies to identify, explore and plan strategies for large *Niche markets* like *Euro*, corporate-wise as well as nationally

(j)Emergence of *'India Pavilions'* in several major IT exhibitions around the world through the initiative and co-ordination of *Electronics & Software Export Promotion Council (ESC) & NASSCOM*

(k)Creation of *'Mega Web-sites'* on Internet for promoting marketing and encouraging Indian software products and packages under multiple initiatives

(l)Encouraging creation & hosting of web-sites on servers located in India

(m) Setting up of value-added network services including ATMs, electronic Kiosks, telephones, smart cards etc. for providing a *'one-stop non-stop'* service to the public

(n)Mobilising IT in Indian languages based on multi-pronged approach, involving fiscal and other incentives for R&D, production, marketing and popularisation of IT products in Indian languages

(o)Development of indigenous technologies in wireless tele-communication such as *CorDECT, remote access switch*, etc. to achieve the national objective of rapid, low-cost expansion of telephone & Internet connectivity in rural and remote areas

(p)Promotion of *electronic commerce* in sea ports, Airports Authority of India, DGFT, banks, container services, customs & Indian railways

(q)On-line upgradation of Court Information System (COURTIS), Parliament Information System (PARLIS), Computerised Rural Information Programme (CRISP) and other such databases over NICNET

(r) Adoption of telecommuting, a new modality of doing work, in the framework of *'Management by Objectives (MBO)'*

(s)Creation of Information security agency at the national level

(t)Wider dissemination of *Cryptology & Cyber security* knowledge in the country to improve information security, network security and bring about a greater degree of secure use of EFT, digital signature, etc.

(u)Software products: emerging 'slivers', productization, and embedded software

(v) Web enabling legacy systems, e-commerce/extended enterprise applications, standards based application integration, knowledge management and convergence applications

(w*) IT enabled services*: HR services, remote customer interaction, data search ,integration and analysis, and engineering and design services

In addition, while online *electronic newspapers* have already become a reality in India on a limited scale, its wide-spread impact would be felt only with the technological advances in telecom services

Multimedia technology enabling simultaneous exchange of voice, text and data would prove to be a major medium of advertisement. Spending on advertisement is expected to be around 5 per cent by 2000 and to reach 12-15 per cent by 2020 AD. Ultimately the market would see an increase from Rs. 350 million to Rs. 120-150 billion by 2020. Around 50-75 million households are expected to be potential users of multimedia by 2020.

Geographical Breakdown of Exports

An interesting industry trend that has been noticed in recent years is the expansion of the Indian IT industry's presence from beyond traditional destinations, to newer geographies. The industry's focus is no longer on English-speaking countries alone, and a key strategy for Indian IT majors has been to harness local talent to tap domestic markets and de-risk the revenue model by reducing their dependence on one geographical region.

Americas and Europe remain the key markets, accounting for over 90 per cent of IT-ITES exports. However, export earnings from markets other than the US and the UK are also witnessing significant double-digit year-on-year growth.

While Indian service providers have built delivery centers in key source markets (e.g. US), they are expanding their footprints in specialist locations like China for engineering and design; South Africa for insurance, and near-shore locations like Eastern Europe and Mexico. Apart from companies in the US, organizations from Europe, South East Asia, Australia, Japan, Hong Kong, New Zealand, etc. are also reaching out for Indian software expertise, supported by the conducive policy environment and incentives for software exports offered by India(see Table II).

Table II: Geographical Breakdown of Exports

Location	FY 2004	FY 2005
Americas	69.4%	68.4%
Australasia	22.6%	23.1%
Europe	7.4%	8.0%
Others	0.6%	0.5%

IT Services Exports by Verticals

Services exports in FY 2004-05 witnessed continued strength in traditional vertical market segments

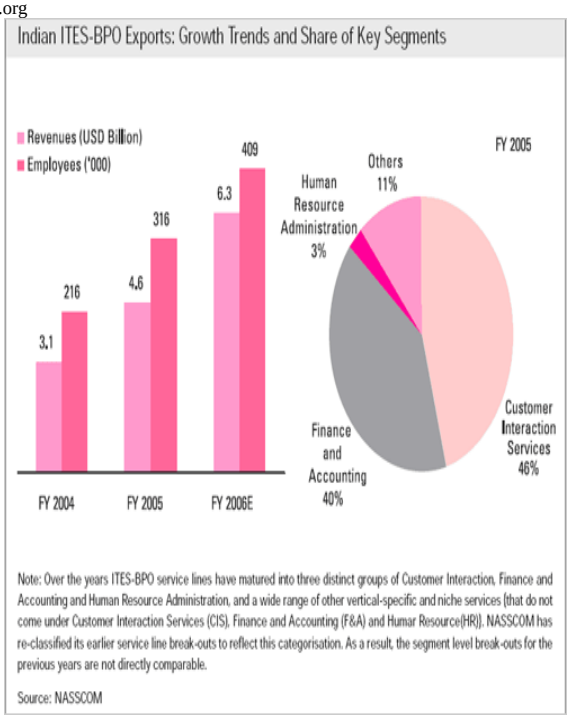
including BFSI, manufacturing and telecommunications.

Underlying the increasing geographic and vertical market penetration is the continuing supply-side maturity of the Indian industry. This is reflected in the ongoing scale expansion without compromising on quality or productivity and growing deal sizes; demonstrated abilities of India-based firms to broaden their service portfolio, leverage productivity and utilization levels to sustain competitiveness and enhance their global service delivery capabilities - while maintaining high levels of growth.

The Indian IT-enabled Business Services (referred to as ITES-BPO) segment continues to chart strong year-on-year growth, estimated at 37 per cent for FY 2005-06. Growth is being driven by a steady increase in scale and depth of existing service lines, and by the addition of newer vertical-specific and emerging, niche business services.

Indian ITES-BPO (re-classified) exports are estimated to have grown from USD 3.1 billion in FY 2003-04 to USD 4.6 billion in FY 2004-05, recording a growth of nearly 48 per cent, and are estimated to reach USD 6.3 billion by the end of the current fiscal year (FY 2005-06)[8].

Net employment in the ITES-BPO segment is estimated to have grown by approximately 100,000 in FY 2004-05, taking the total direct employment to 316,000. Based on hiring trends observed over the year, this segment is likely to end the current financial year (FY 2005-06) with total employment projected to reach 409,000. Employee turnover/ attrition levels appear to be stabilizing with the talent acquisition, development and retention initiatives being undertaken by the players, beginning to deliver results.



Source NASCOM

Factors Responsible for IT Boom in India: A Closer Look

The study will discuss in detail three initiatives which were responsible for the present IT status for India: policies that mobilize the Indian Diaspora, the formation and work of the erstwhile DoE and Ministry of Information and Communication Technology,& software technology parks .The fourth initiative is the pattern of spatial agglomeration in the IT sector.

Government-Diaspora Partnership

The Role of Diaspora in the Emergence of the Indian IT Industry –The Indian Diaspora has been very successful in Knowledge-intensive sectors in the US, and more so in the IT sector. Almost simultaneously, a very competitive and successful IT industry emerged in India. This section analyses various factors that helped in the emergence of the successful IT industry in India (during the last 35 years) and the role that the Diaspora community played in this evolution. Very few attempts and investments that were made by PIOs(Persons of Indian Origin) in the 1970s and early 1980s were quickly abandoned because of bureaucratic obstacles by the Indian government and the limited capabilities of Indian partners. Hence, the only obstacles by the Indian government and the limited capabilities of Indian partners. Hence, the only crucial role played by these PIOs was limited to being tolerant mentors of early Indian software development companies. limited to being tolerant mentors of early Indian software development companies.

[8] To better reflect how the industry and customer markets view the portfolio of services sourced from India, NASSCOM has re-classified the manner in which it reports the various segments included within IT-ITES. For instance, this year onwards, engineering and R&D services are being identified as an independent service line and will be reported separately. Consequently, some of the services (e.g. GIS), earlier included under ITES-BPO, will now be reported under engineering and R&D services. Further, NASSCOM has increased its overall estimate of industry exports for the previous year (FY 2004-05), based on the details reported to NASSCOM and STPI by individual companies. As a result of the re-classification and the revision of estimates, the historical values for a few segments have changed. In addition to the projections for FY 2005-06, to help ease comparisons we have re-stated the details for the preceding years FY 2003-04 and FY 2004-05 as per the new classification.

In early 1980s, several small Indian companies came to Silicon Valley in search of low-end contract software development work. Several PIO executives were willing to help but most found the Indian companies' work to be unsatisfactory and many suffered from deficient development tools and computers. This is partly because even until 1985-86, the Indian government was promoting Russian computers over American computers and Indian companies had just started working with PCs; hence, the companies' professionals could not meet, or sometimes even understand, US standards for quality and timeliness. To mitigate this problem, the Diaspora executives sometimes created programs within their US companies whereby Indian programmers could work in the US and with US technology (at Indian wages plus travel related costs). Further, they coached and guided the Indian companies to enable them in improving their quality and performance standards.

Hence, during the 1970s and 1980s, the role of Indian Diaspora in the evolution of Indian IT industry was limited to that of a patient mentor and brand ambassador in most of the cases.

1990s

Many Indian engineers, who had started moving to the US in 1960s, had by now either become entrepreneurs, or Venture Capitalists or high-level executives in large and medium sized companies. And, these professionals had started to coalesce especially because many had graduated from the same top-notch colleges in India (such as the IITs) and most of them also knew their counterparts in India (who were often also alumni of the same colleges). Some of these relationships quickly matured in forming non-profit associations such as TiE and SIPA (Silicon Indian Professional Association).

TiE, originally designed as a Silicon Valley organization provided mentoring to promising young expatriate IT professionals, soon developed into a worldwide network of Indian professionals which has had a substantial influence on the Indian IT industry and government policies towards it. Currently, TiE[9] has 38 chapters and over 6,800 members, worldwide.

Since many in these people knew their counterparts in India and since most were closely observing the growing Indian IT industry, in the mid and late 1990s, some of them started their own IT companies in India (e.g., Cognizant, Techspan, Mphasis) whereas others invested in nascent IT and Dot.com companies in India. Further, since US, Canada and UK were facing a shortage of IT professionals during 1996-1999, many in the Indian Diaspora convinced their companies to hire Indian IT professionals and this resulted in the 'Indian IT Diaspora' becoming stronger and the Indians constituting 24 percent of the entire Silicon Valley IT professional population by late 1999.

All these developments in turn permitted another crucial Diaspora role. Some Indians had become senior executives at many major US corporations, like IBM, GE and American Express. In nearly every instance where these companies invested in or outsourced work to India a well placed expatriate executive crucially influenced the decision. In part the individual's own success supported the emerging positive reputation of Indian engineers. And in part the individual's direct experience of India gave them credibility in vouching that the well-known problems of India's infrastructure and bureaucracy could be overcome. The US investment and outsourcing partly drove Indian software industry annual growth to 40 percent during the 1990s .

There were other Diaspora roles as well. Some younger Indians in the US moved to India as 'Expatriates' and started IT Research and Development Laboratories (e.g., IBM India Research Laboratory was started in April 1998) whereas others moved to supervise US investments, outsourcing contracts, and to train and manage Indian professionals to US efficiency and standards.

2000 & Beyond and Government Programs to Strengthen Diaspora Relationships in Promoting IT sector

By 2000 Indian engineers were at the helm of 972 Silicon Valley-based technology companies, which accounted for approximately $5 billion in sales at 25,811 jobs. Moreover, the pace of Indian entrepreneurship accelerated rapidly in the 1990s: while Indians were running only 3 percent of the technology companies started between 1980 and 1983, they were running 10 percent of those started between 1995 and 2000.

The success of these former Indian nationals is evident, and the Indian government recognizes that connections with these individuals can help promote the domestic market for IT. The Indian Diaspora of IT professionals in Silicon Valley has established social networks like The Indus Entrepreneur (TiE) and the Silicon Valley Indian Professionals Association. The existence of these organizations shows that Indian immigrants

9 The Indus Entrepreneur.-The organization currently has established chapters in Bangalore, Bombay, Delhi, Hyderabad, Calcutta, and Chennai. The global connections have paid off, as the non-resident Indians "in turn invested in promising start-ups and venture funds and have begun to serve as role models and advisors for local IT entrepreneurs."

maintain close ties to those of their own origin and value the professional connections that such a network can offer. To address the effects of brain drain, then, the Indian government began with mechanisms that strengthened the ties between the diaspora and its roots.

India's Ministry of Science and Technology formed a High Level Committee on the Indian Diaspora in 2000 in order to facilitate communication and interaction between the expatriates and their home nation. The organization's 2002 report recognizes that scientists of Indian origin abroad "are keen to contribute to their country of origin." One obvious way in which these immigrants can contribute to outsourcing in the Indian IT industry is to encourage their companies to partner with Indian firms for software development or other production processes, thus alleviating the effects of brain drain. If individuals hold leadership positions within their corporations and make managerial decisions, there is no better way to encourage outsourcing than to engage it firsthand.

The government is pursuing a number of different bilateral programs simultaneously. The first of these programs is an exchange program called "The Transfer of Knowhow Through Expatriate Nationals" (TOKTEN). The concept, formed by the United Nations Human Development Program (UNDP), encourages expatriate nationals to undertake short-term consultancies in their home countries. In India, TOKTEN has enabled 650 professionals to visit 250 institutions from 1980-2001. The alumni networks of the government-funded Indian Institutes of Technology (IITs) have been another method of encouraging interaction. Alumni from America have recently given 60 crores in Indian Rupees to IIT Kanpur and 30 crores in Indian Rupees to IIT Karagpur "for upgrading infrastructure and human resource development" and "tracking…alumni to other Indian premier academic-cum-research institutes." Investing in the IITs has helped in some way in tackling the problems caused by the shortage of computer professionals that has been predicted for the next several years.Two other notable initiatives have been taken by the government to connect with the diaspora. One of these is advisory panels with eminent non-resident Indians that have spurred investment and led to several IT joint ventures. The next is the placement of many non-resident Indians in honorary fellowships at universities, funded by professional scientific and technical societies.

Through these various programs and incentives, India has found a viable method for fighting the "brain drain" which could so easily strip the country of one of its most valuable resources, a skilled labor force. Short of working extremely hard to entice Indian nationals back to the country, fruitful interaction with the Indian Diaspora is an excellent way to push the IT industry in India to even higher levels and mitigate the shortage of IT professionals there.

Indian IT Industry: Role of Government Institutions and Ministry of Information Technology

Before 2000

In India, the Department of Electronics (DOE) was the primary agency overseeing government IT policy formulation and implementation. Three government-funded computing organizations played important roles in new technology development: the National Centre for Science and Technology (NCST) in Bombay, the Centre for Development of Advanced Computing (C-DAC) in Pune, and the National Informatics Centre (NIC) headquartered in New Delhi. C-DAC is now one of the most advanced IT development centres in India. The NIC was the second major Indian computer-related project funded by the UNDP in 1977. It operates the largest data communications network (NICNET) in India with more than 600 earth stations linking government agencies at all levels. There are many lessons to be learnt from the two decades of NIC operations in the country. NIC has done a pioneering work in popularising the use of computers in the government sector, breaking the geographical boundaries and encompassing all sectors of economic activity. In the process, it has carved out a niche for itself among the public sector organisations. It has taken upon itself the job of creation of IT applications for different government departments. In 1990,DoE devised a scheme called DOEACC jointly with the All India Council for Technical Education(AICTE) to provide accreditation according to specified level of course. The Government has supplied trained manpower for software development.

It appears that potential of software exports from India was recognized as early as 1968(as stated above).The Electronic Committee group's report in 1968 recommended the promotion of the industry for exports(Subramaniam,1992).

In the early 1970s department of electronics was created. A few software vendor firms came into the market mainly to serve the public sector: the investment by the department of electronics in the public sector R & D projects, which involved software development. Public Sector contracts for custom software development were given to private sector firms. The government also initiated software related training courses in the Indian Institutes of Technology and universities around this time. The Software Export Scheme was launched in 1972. In addition, there was an emphasis on computer and software education and

training. Any institutions that focused on training were allowed to import hardware at much lower import duties . In parallel, the DOE began to encourage public sector projects that dealt with software development. Public procurement of software gave priority to Indian companies.

As far as Hardware is concerned in the 1970s, the state-run Electronics Corporation of India (ECIL) was the main beneficiary as very few licenses were issued to private producers of microcomputers. However, the microcomputers developed by ECIL were too expensive for general consumption and lacked an adequate range of software. To meet with the growing demand for computers, the DOE gave permission to ICL's Indian subsidiary, ICIM to produce microprocessor-based computers. This decision was extremely unpopular with the domestic private sector. In 1978, the Sondhi Committee Report on the state of the computer industry pushed for the issuance of more licenses for the private sector . As a result, by the time Rajiv Gandhi became Prime Minister in 1984, there were a few very strong domestic computer companies including HCL and Wipro[10].

The DOE's role underwent a transformation in the mid-1980s It became increasingly apparent to high level policy makers at the DOE and in the Government of India that the public sector was not able to supply the computer hardware and software that was needed by the domestic market . The new approach was more supportive of the domestic software industry rather than the previous more restrictive and regulatory one. It also encouraged software exports and export-oriented foreign investment. In response to the success of companies like HCL and Wipro, a new Computer Policy was introduced in November 1984 which reduced many constraints on the industry . Further liberalisation came in the form of the new Software Policy, which was implemented in 1986. The new policy advocated what was called a `flood in - flood out' approach, allowing imports to `flood in' in the hopes that eventually exports would `flood out.

Further evidence of this changed attitude came in 1986 when Texas Instruments (TI), proposed to establish a 100 per cent export-oriented, foreign owned and operated subsidiary (its first outside the United States). The DOE and the GOI were fairly quick in the processing of the license

The DOE's support for exports did not, however, extend to the export of know-how in the form of Indian programmers going abroad to provide onsite services to clients, which in 1989 accounted for over 90 per cent of software revenues (Schware, 1992, p. 151). In 1987, a decision was taken to impose a 15 per cent tax on foreign exchange expenditure on travel. This had substantial implications for the software industry because of the "body shopping" activities (Heeks, 1996, p. 47). The importance of body shopping in terms of establishing the Indian software industry's reputation and allowing Indian programmers and engineers to gain first-hand knowledge of the latest technologies.

TI's(Texas Instruments) fully equipped software development centre, inclusive of satellite connectivity, made it clear to the DOE and the GOI that in order to foster the development of a vibrant software industry in India, it was necessary to provide an environment which would facilitate such activities. This realization acted as a catalyst for the establishment of the Software Technology Parks of India Scheme in 1988. These STPs were envisioned to be like export processing zones, where the government provided infrastructure, buildings, electricity, telecommunications facilities and high speed satellite links .

Until this time, telecommunications policies came under the auspices of the Department of Telecommunications. The DOE, however, saw an opportunity to boost the software industry and took the responsibility and the risk to install the appropriate telecommunications equipment so that Indian software companies would have an easy access to their clients and so that the delivery of software exports would be expedited.

While the Department of Telecommunications took commercial approach to the provision of telecommunications , the DOE took a development approach.

In the early 1990s the Finance Ministry made concerted attempts to consciously give thrust to software exports (Kumar and Joseph,2005). Accordingly, the following measures were initiated (a) the removal of entry barriers against foreign companies (b) the removal of restrictions on foreign technology transfers (c) the participation of the private sector in policy making (d) provisions to finance software development through equity and venture capital (e) measures to make available faster and cheaper data

[10] Another way to get around the high duties and elaborate licensing procedures was to set up a training or educational facility dedicated to software and hardware training. The Delhi-based company, NIIT got its start in this way in 1981. In order to have access to computer hardware, NIIT got into the market as a training facility and essentially entered the market where there was no market. It was only in 1985, when the law with respect to the purchase of hardware was amended that NIIT changed its strategy somewhat. It added to its portfolio computer software development and technical support, initially for the domestic market but in 1988 the company earned its first dollar from on-site professional services, and in 1990/91 the emphasis on foreign markets became a core activity.

communication facilities and (f) the reduction and rationalization of taxes, duties and tariffs.

Recognizing the potential of IT related industries and software for India's development, the Prime Minister created the National Task force[11] on information technology and software development (NTITSD).NTITSD submitted a report outlining a national IT Action plan comprising 108 recommendations for software and 87 recommendations for hardware(India,NTITSD,1998).These recommendations have since been notified by the Government in the Gazette of India dated 25 July 1998(India,MIT,2000)

The Indian Action Plan makes stimulation of the software and IT services industries a basic objective, with a goal of $50 billion in exports and a commensurately large domestic IT market. These measures will have a direct impact upon business use of the Internet since networking is integral to the activity of software and IT service companies. The Action Plan lists 38 steps to assist IT firms with venture capital, credit, subsidies, reduced taxes, duties, and fees and fewer bureaucratic roadblocks. (These measures will of course have secondary effects on all of our dimensions.).The Action Plan also supports offshore programming services with a call for diplomatic pressure to make it easier for Indian programmers abroad to obtain visas. This is a double-edged sword. If the United States or other nations ease visa and work permit restrictions, programmers emigrate. While this brain drain hurts the domestic software industry, professional nonresident Indians are an important source of hard currency and business contacts.

While most of the business emphasis in the Action Plan is in support of software and IT service, there are also measures to encourage other electronic business, such as by ordering the Department of Telecommunication to meet "communication requirements" for electronic commerce and electronic data interchange, expediting electronically based export orders, and mandating bar coding. To the extent that this traffic flows on the Internet, it will add to penetration; however, related legislation concerning privacy, digital signatures, and encryption is not yet specified.

Intellectual property laws and customs will also have a major impact In both nations, poverty and limited familiarity with credit cards and other banking services among the general public will also constrain the level of consumer-oriented electronic commerce, but this should be less of a problem in business-business transaction processing. Still, it should be noted that Internet commerce is in early stages of development in India, and it will be some time before it significantly impacts the Internet.

In 2000 and Beyond

In 2000 India set up a science and technology bureaucracy to coordinate government-administrated projects relating to information technology. A number of different government agencies, formerly under the Ministry of Science and Technology that are concerned with IT, were brought together into an integrated Ministry of Information Technology including the DOE (also referred to as MIT). It has since undertaken a large number of projects aligned with its vision of "making India an IT Super Power by the year 2008."Among the objectives identified are "creation of wealth, employment generation, and IT-led economic growth." As a policymaking body, the organization's leadership consists of a minister, a minister of state, a secretary and additional secretary, a controller, and several group coordinators and senior directors. This ministerial hierarchy fills a previous void in the Indian government; until the Ministry of Information Technology was formed, there was no "single apex institution or focal point for formulating national policies and strategies for the IT sector…and the lack of any central oversight and a critical mass of in-house expertise in the public sector often hinder[ed] the sharing of information…and the development of information standards and protocols and common information infrastructures." It is further recognized that "in general, the institutional framework [was] underdeveloped for dealing with systemic problems of computer and software requirements, planning, procurement, coordination among agencies, and IT diffusion." The Ministry of Information Technology has been India's solution for strengthening that framework.

Software Technology Parks of India : A Business, Academia & Government of India Initiative

The Department of Electronics introduced the scheme of Software Technology Parks(STP) in the early 1990s.An STP is an analogue of an export processing zone. Firms in STPs were allowed tax exemptions, guaranteed access to high speed satellite links and were provided with reliable electric power, including core computer facilities, ready to use office space, as well as communication facilities. They were allowed to import equipment duty free and without licenses

[11] The 18-member task force had representatives of industry (telecommunication, software, and IT), government, and education, but not health. This may explain the lack of emphasis on health care in the action plan

.Full(100%) foreign ownership was permitted in exchange for an export obligation. Firms were allowed to repatriate capital investment ,royalties and dividends free once they paid the taxes due. China on the other hand provided all these incentives and more in its special economic zones to attract investment ,particularly foreign direct investment that was oriented towards export markets.

The concept of software technology parks (STPs), which provide the technical infrastructure necessary for IT development, emerged from a growing problem that policymakers began to notice in the early 1980s. The New Computer Policy of 1984 and the 1986 Policy on Computer Software Export, Software Development, and Training set out the objective of expanding the Indian software export and development through data communication links. The policies' aim was to develop software in India using Indian expertise on sophisticated computers that were being imported duty-free. Protectionist policies had, until then, stunted the growth of the industry by imposing duties of up to 200% on the imported hardware needed for advancement of both the hardware and software components of the industry

The next and more significant difficulty firms faced was the high cost of the data communication links needed for software development. The poor telecommunications infrastructure India had at the time was inadequate. Foreign corporations were looking to expand their global production networks to India because the country offered a skilled, English-speaking workforce, but the corporations could not be accommodated at a reasonable cost. Though companies were allowed by law to establish the data communication link through their own initial investment, few companies could pay the high price without other incentives. From this necessity, the idea of software technology parks was born. The Ministry of Information Technology developed the concept of STPs and lists the following as the objectives of the project:

To establish and manage infrastructure resources such as Data Communication facilities, Core Computer facilities, built-up space and other common amenities; to provide 'single window' statutory services such as project approvals, import certification, software valuation, and certification of exports for software exporters; to promote development and export of software services through technology assessments, market analyses, market segmentation and marketing support; to train professionals and to encourage design and development in the field of software technology and software engineering.

The STPs now serve as intersections where a viable business model, strong Internet infrastructure, and government interface come together for a successful enterprise: "The infrastructure facilities include modern, high-speed, broad-band telecom links, powerful computers and network systems beyond the reach of individual firms, consultancy, and training support." The first of these parks were established in 1991 at Bangalore, Pune, and Bhubaneswar. In nearly all of the literature on the subject STPs have been heralded as one of the most profitable institutional initiatives for developing the IT industry.

Institutional Infrastructure and Patterns of Spatial Agglomeration in the IT sector

The Indian experience has shown that the cities with a high concentration of software development activity enjoy a disproportionate share of innovative infrastructure, the skill base and other resources for technology development. Because of significant agglomeration economies present in skill and knowledge intensive activities, such as software development, the fact that these cities have this disproportionate share of the national innovative infrastructure has served as a magnet for software development activity. These cities also had the highest concentration of public sector R&D establishments (especially defense) as well as publicly funded engineering colleges With the Indian states realizing the importance of improving the urban and rural Infrastructure sector the IT industry has been spreading to other smaller cities of India.

Causes of Indian IT Industry BOOM today and India's edge

This boom is largely fuelled by the following factors: a)The Role GOI played in providing impetus to promotion of software exports since 1980s-in industrial development and in creating an enabling environment for the integration of local industries in the global economy, b) India has a vibrant news and entertainment industry, and a large domestic audience which is hungry for content, c) Indian success stories in Silicon Valley, which by now are legion, are inspiring entrepreneurs and young people in the homeland, and d) venture financing to India's information technology sector has grown dramatically in the last year, aided in part by Indian net-millionaires from overseas playing a key role (e) its low cost -high quality-scalability model, which gives it an edge over other emerging ITES-BPO destinations such as Ireland, the Philippines, China and some Latin American countries. (f) a high quality, pool of knowledge workers who have English Speaking and relevant domain skills give India an edge over other offshore outsourcing locations.(g) The ability to focus on core competencies and use offshoring to access new technologies and talent to strengthen and expand

existing business offerings.(h) high quality and competitive pricing of Indian software and services(i) availability of a large pool of knowledge workers with requisite computing and language skills(j)flexibility and adaptability of Indian professionals(k)ability to undertake offshore software development through data com links, availability of local enterprise and skills, the Indian diaspora , especially the active and productive two way links between Indians and Indian Americans in the US(Baru,2006)
Major global players who have recently invested in Indian IT and Internet firms include Walden, Draper, Chase Manhattan Bank, Citibank, Microsoft, Intel, Pacific Century Group, News Corp and Kerry Packer Ventures. Approximately US $ 1 billion has already been invested or pledged by these and other venture capital firms, and industry reports suggest that a further US $ 10 billion is waiting to be tapped in the next 4 years.

Indian businesses are also moving aggressively to have a web presence, and over 200,000 large and medium sized firms are expected to launch net-based operations in the next year. Some private banks have already started on-line banking and advisory services, and regulatory authorities are expected to allow on-line stock brokerage in the next few months. Many of these companies are motivated more by a fear of losing out than by any cogent business strategy, but a positive spin-off which is likely is that it will bring into focus the importance of service quality in business. On-line transactions and customer relations are just the beginning, and in time this will lead to bigger and better ideas. Internet is already aiding a gradual process of de-intermediation in many areas, such as in recruitment, business research, travel, real estate and insurance, and e-governance initiatives at different levels of government are now being planned. In future, interacting with various authorities for routine permits and information will become simpler and quicker, and MNCs will probably waste less time in low-level tasks.

Private sector developments have actually gone hand in hand with official measures to give a boost to the IT sector, including rapid adoption of Internet in various government departments, removal of irritants in tax rules for venture capital, reduction of import duty on computer parts, duty free import of software, laying of 8,000 kilometers of optical fiber cables between cities, and strengthening the domestic Internet backbone. Just recently, the Indian parliament passed a new cyber law that provides legal sanction to e-commerce.

Many of these initiatives are still many months away from being fully implemented, but their overall impact will be to sustain and perhaps even accelerate the IT momentum over the medium term by lowering costs and increasing access. For instance, more than 80 percent of India's corporate websites are currently located in USA because Indian servers are costlier and less reliable, but many of them are expected to shift to India once optical fibers and broadband become a reality.

India: ICT as a Tool for Development and Researchable Issues

Any government will have a dual role in the ICT sector and India is no exception. It acts as a regulator and formulates long term policies for the promotion and development of various industrial and service sector reforms including the ICT related activities; and at the same time, it deploys these services for the governance and improving the efficiency of its decision making and administrative control. The government's reform agenda is also affected by bilateral and multilateral agreements. This is particularly true of the ICT because of its international outreach and impact. The recent advances in ICT will have a profound impact on the way the governments function in the coming years. While the move towards decentralized planning and management will gain momentum, the need for high quality of information for decision making, control, monitoring and evaluation will increase in all sectors of social and economic activity. To what extent the Government of India can benefit from the development and application of emerging technologies for information storage, processing and communications? What has the Government of India done in this context? Has it kept pace with global trends in ICT and to what extent it has been active in policy formulation? Has it been an effective user of ICT in its day-to-day administration and decision making? These are some of the questions, the answers to which are difficult to get.

India is a predominantly rural, agro-based country with about two thirds of its population based in villages. Rural transformation was led by the green revolution of the late sixties and the white revolution of late seventies. The next step phase of rural transformation can be brought about by a Knowledge Revolution facilitated by ICT (Information and Communication Technology).

We have seen emergence of ICT-based rural e-Governance applications in the recent past, that have demonstrated the important role ICT can play in the realm of rural development. Several of these ICT projects have attempted to improve the reach, enhance base, minimize processing costs, increase transparency and reduce cycle times. Several states have initiated the creation of SWAN (State Wide Area Networks) to facilitate electronic access of the state and district administration services to the citizens in villages.

Information and Communications Technology (ICT) has become the most exciting set of technologies today, thanks to developments like Internet growth, e-commerce proliferation, the mobile computing revolution and the arrival of broadband. At this juncture, it is appropriate to study the use of ICT in rural India and how it addresses the needs of the common man.

We discuss below a few case studies of ICT applications in a development context. Most of the cases are pilot projects undertaken either directly by Government departments or done with heavy Government support and funding. The cases concentrate on applications of ICT for rural development.

Computer Assisted Deeds Registration system of Andhra Pradesh which attempts to reengineer the Registration Process involving about 120 million documents a year. The pilot project, already implemented, has reduced registration time from several days to just 60 minutes and Encumbrance Certificate issue time from five days to 10 minutes.

Integrated Certificate Distribution across Andhra Pradesh is another high-profile experiment being implemented. The technology is simple; yet the benefits immense. The system builds up Social Security Identification (SSI) for most citizens–a unique way of identifying every individual citizen, cost-effectively.

Rural Postal System in Hyderabad, Andhra Pradesh represents an interesting way to reengineer the multiple functions in a Post Office counter in India using simple, PC-based technology. Thanks to near-zero investments in Post offices, such relatively simple and straightforward applications work.

IT at Milk Collection Centers in Cooperative Dairies in Gujarat brings real benefits to more than 60,000 farmers daily, who are involved in the milk collection project spread over 600 milk collection centers. Using relatively simple technology (PC, Weighing machine with PC interface and online milk tester), this system has delivered results over many years.

Same Language Sub-titling experiment in Gujarat has been another novel, intuitive way of improving literacy among children. In this experiment, synchronized sub-titling of film music song lyrics has been undertaken. As a result of this implementation, children have shown significant improvement in literacy. The fact that children like to co-sing the songs and read the text of the lyrics, indicates the success of the experiment.

The Warana Wired Village Project in Maharashtra is a very recent experiment which provides Internet connections using satellite communication to rural India. This experiment leverages the relative prosperity gained by the region over the past few decades through a Sugar Factory. Spreading over 70 villages, the project attempts to use Web technology for self-improvement through skills development and employment generation.

To facilitate and bring about change as a part of transformation of India's villages, the NASSCOM Foundation has taken initiatives in the States of Orissa, West Bengal, Tamil Nadu, Karnataka, Kerala, Andhra Pradesh and Maharashtra to build capacity for ICT through the creation of VKCs (Village Knowledge Centers). A part of a Rural Knowledge Network (RKN), the project engages the industry, civil society, multilateral agencies and the governments in synergistic interventions through the creation of partnerships for change. The ICT interventions from the foundation have been focused on enabling the ICTs for the three Ps: Purpose, People and Prospects.

At the end of the day, it is clear that citizens must find the ICT services relevant and beneficial to them. It would be better if the user interfaces are in local languages and there is more localized content available. This will minimize the need of the local communities to look outwards for the information. The objective is to build on partnership networks by engaging and involving major private sector players and a network of grassroot civic society organizations to engage more and more villages and enable ordinary people to access ICT.

India is in a relatively better position as compared to many South Asian/SAARC countries with regard to the development and applications of ICTs. Nevertheless, there are large imbalances in the development and use of ICTs within the country. China, which has a larger population base than India, has better availability of ICTs. Sri Lanka stand out clearly as compared to other countries of the region. The high level of socio-economic development are associated with the high availability of ICTs. The World Development Report has also shown that there is a positive relationship between literacy and the application of ICTs for development purposes.

The Software Export Success Model

We analyze the experiences of three successful major software exporters

A summary of the export success(Table III)

Table III: Export Success of the Three Leading Software Exporters

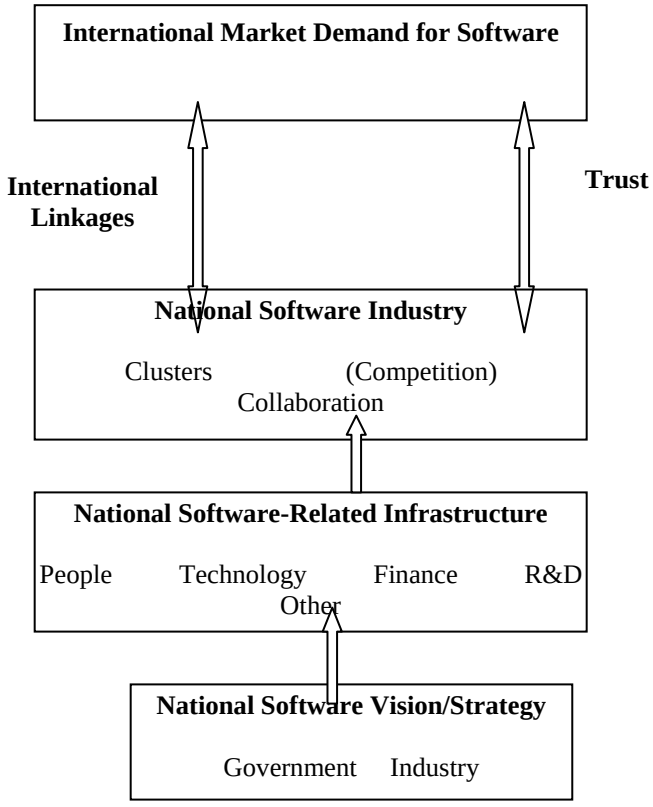
	India	Ireland	Israel
Demand	High external demand; weak domestic demand	High external demand; weak domestic demand	High external demand; strong domestic demand
National Vision and Strategy	Vision and strategy present: software services, then climbing the value chain	Vision and strategy present: product-related services for multinationals, then diversification	Vision and strategy present: home-grown product exports, then innovation and differentiation
International Linkages and Trust	Diaspora and state-funded links; reputation and trust, partly through ISO and anti-piracy	Diaspora and state-funded links; reputation and trust, partly through ISO and anti-piracy	Diaspora and state-funded links; reputation and trust, partly through ISO and anti-piracy
Software Industry Characteristics	Some competition; clustering and collaboration	Some competition; clustering and collaboration	Strong competition; clustering and collaboration
Domestic Input Factors/ Infrastructure	Strong, low-cost human capital; catching-up in telecoms; access to capital; limited R&D success	Strong human capital; strong telecoms; access to capital; some R&D base	Strong human capital; strong telecoms; access to capital; strong R&D base

Source: Heeks and Nicholson(2002)

On the basis of this analysis, the dimensional model presented earlier can now be drawn in greater detail. It is shown in Figure IV as the 'Software Export Success Model'. It shows drivers at top (pull) and bottom (push), and enablers in the middle.

In a general sense, the software export success model has proven useful as a way of understanding the experiences of developing and transitional economies. It offers a template against which to analyze national strengths and weaknesses. It also offers some more general guidance for countries seeking to increase their software exports.

Figure IV: Export Success Model



Source: Heeks and Nicholson(2002)

Indian IT model can be emulated by other governments of the NAM countries and industry collaborations assisted by the free flow of venture capital and talent flow .As a start focus can be on 5 main areas that lay foundation for a vibrant IT economy:

1) **E-infrastructure:** Success will depend on the links between the main city hubs. The focus is to strengthening such links through the development of Trust Marks and Public Key Infrastructure, which could support the greater use of online transactions and facilitate E-Commerce.

2) **E-Markets**: The focus would be on strengthening business relationships between the IT industries of different cities through the promotion of internet based facilitators, such as e-business interoperability standards. For example, the Pan Asian E-Commerce

Alliance is an industry alliance among private companies from Korea, China, Taiwan, Singapore and Japan. They are collaborating in areas such as secure cross border transaction services, Mutual Recognition of Public Key Infrastructure, and the creation of pan-Asian portal focusing on trader directories.

3) **E-Capital**: This focuses on free investment flows to give entrepreneurs greater access to funding support from across South countries. In the longer term, it may be feasible to create incubation centres funded by venture capitals for innovative start ups through the South countries. Thus, for example, Temasek Capital's InCubate Programme in China and India provides start ups with seed funding, facilities, management services, branding and network support.

4) **E-Talent**: The focus here is on facilitating of cross border IT talent movement, such as mutual recognition of IT qualifications, in order to facilitate free flow of talent and develop human resources of South countries. One example is cross certification of IT skills.

5) **E-Ideas**: Leveraging on South's large talent pool and cultural diversity, this focus is on South Countries as a hotbed for new ideas and as a credible, future though leader in international IT. An example would be where industry associations form networks to stimulate progress in areas such as ASP/IDC, wireless applications and E-Learning

HURDLES AHEAD the sector would have to overcome several problems, including inadequate quality and skills of graduates, rising salaries and weak infrastructure, which resulted in frequent power shortages, low level of PC use and internet penetration ,low level of domestic technology development, limited bandwidth, inadequate availability of venture capital and limited domestic market for knowledge based technology and products. Indian Government has a continued role to play in addressing such issues. The new IT policies enunciated through the new Indian IT Bill(May 2000) should help create the legal framework within which these issues can be addressed better. The bottom line, however, is new investment in high speed data networks, improved telecom and power infrastructure and sustained improvement of the human capital infrastructure- in the schools and technical institutions(Baru,2006).

There are several constraints besides a potential manpower shortage that could preclude their realization in full. Reliable electric power, efficient and inexpensive telecommunications and access to venture capital are essential infrastructures for the IT sector. Although telecommunications infrastructure has vastly improved there are still some unresolved issues relating to the authority of the regulatory agency (Telecommunications Authority of India (TRAI)) vis-à-vis Department of Telecommunication and the state-owned providers. The situation regarding electric power continues to be abysmal. In fact, the large IT firms, like other large enterprises, had to invest in their own captive power generation facilities. To the extent that the unit cost of power from small scale captive plants is much higher than what it would have been had their supply come from an efficient large-scale utility, the failure of India's public power system adds an avoidable cost and dampens the competitiveness of its IT sector.

India's labor and bankruptcy laws could be counterproductive in the IT sector as in other sectors of the economy. A report in 2000 of the Subject Group of Knowledge-Based Industries in the Prime Minister's Council had recommended the exemption of the IT sector from some of the draconian provisions of labor laws. Whether or not it is wise to exempt one sector from dysfunctional law rather than repealing it is arguable. In any case, political support for a repeal is not there yet. However, some de facto exemptions do exist. For example, in ITES, states often exempt call centers and the like from working hour type restrictions, allowing women to work at night. Also, it is likely that programmers in large firms are not subject to the same provisions as industrial workers. Along with increased presence of Indian IT companies across the globe, new services lines are emerging and the Industry has reached the next level in services offered. Mergers and acquisitions by Indian players is also a key trend. Inspite, of the growth seen so far, it is estimated that less than 10 percent of the addressable market for globally sourced IT-ITES has been captured till date, indicating significant headroom for growth

India is not the only country with a pool of English speaking workers for employment in the IT/ITES sector. Other countries include Bangladesh, Ireland, Pakistan, Sri Lanka and the Philippines. Except for Bangladesh, the wages costs are higher than India's in the other countries. English speaking ability can, of course, be acquired and, as such, potential future competition for India from countries currently without a significant pool of English speaking workers cannot be ruled out.

Despite India's emphasis on import-substituting industrialization, it has not developed a robust, world-class manufacturing industry, and this includes IT hardware. Much of India's hardware industry consists of assembly tasks, almost entirely for the domestic market. India's software industry is, of course, more robust – at least in certain areas. While selling packaged software to consumer (and most business) markets requires economies of scale and scope, as well as marketing and customer support muscle, project-oriented components of software development do not do so, to quite the same degree. To some extent, therefore, India's software industry remains narrowly

focused. For example, of India's 2001-02 software and services exports of Rs. 365 billion, two thirds came from IT services, and close to 88% of that amount came from custom application development and application outsourcing

Conclusions, Discussion and Suggestions

Since Independence in 1947, the goal of self-reliance has guided all spheres of policy making in India. Until the mid-1980s, India's development strategy was characterized by import substitution policies, which were aimed at nurturing domestic industry, including the computer hardware and software industry. These included extensive quantitative restrictions and high tariffs on imports, elaborate import licensing procedures, export subsidies, controls on foreign direct investment and an overvalued exchange rate . The goal of self-reliance also led to a strong commitment to the role of science and technology in India's development strategy. These were areas that were emphasised in industrial policy as well as in the field of education. In addition to establishing the Indian Institutes of Technology, which were educational institutions located in various cities around India aimed at creating a large pool of technical skills, the Government of India (GOI) has had a computer policy since the creation of the Department of Electronics (DOE) in 1970. It was the first developing country to do so and to explicitly target software as a "thrust area", for its high skill requirements, its labour intensity, and its foreign exchange earnings potential.

India's presence in the software industry dated back to 1970 when the TATA consultancy services entered the IT business sector. The foundation for the intellectual capital for software industry was laid by the establishment of the IISc in 1909 ,IITs, IIMs. However, in the 1970s the deployment of mainframe computers was interrupted when IBM withdrew from India in response to a 1976 law limiting foreign ownership of business to 40 percent. Rajiv Gandhi assumed leadership in 1984 and identified telecommunications and information technology as a "core sector" along with traditional industries such as power, steel, oil, and automobiles. The presence of a national strategy for software exports, particularly since 1980s is therefore be recognized as a vital part of software export success (Balasubramanyam & Balasubramanyam 1997). Indeed, it goes beyond this – critical to each country's success has been a vision of what software could achieve for the country; a vision shared by a relatively small but committed group of government officials and private entrepreneurs. Such visions first emerged in the 1970s, were sustained through lean early years in the 1980s, and only truly came to fruition in the 1990s.The study believes that the initiatives on three different levels(as discussed above) served and continue to serve as the backbone of the government's approach in promoting the Indian IT sector since its formative years[12]. The fourth initiative is the pattern of spatial agglomeration in the IT sector. The patterns of concentration in the software development industry in and around select cities illustrates the key importance of institutional infrastructure for the activity. With the Indian states realizing the importance of improving the urban and rural Infrastructure sector the IT industry has been spreading to other smaller cities of India. Also, the public good nature of production in a digital economy, along with the presence of network externalities, may suggest a larger role of government with a sure role in handling security matters.

The first of these initiatives is the Ministry of Information Technology, formed as an umbrella organization to coordinate the activities of the multiple government agencies that deal with the information technology. The second is the implementation of software technology parks where business, government, and academia can come together both for networking and production. The establishment of the Software Technology Parks (STPs) scheme in the late 1980s gave export-oriented software firms in designated zones tax exemptions for five years and guaranteed access to high-speed satellite links and reliable electricity. The national economic liberalization that began in 1991 greatly improved the climate for the software industry as well. Last, there is the set of initiatives that urge communication and interaction with the global Indian diaspora with hopes of encouraging investment in the country from those who have emigrated and become successful in other nations. Indian immigrants are a significant presence in the US information technology industry. There are at least 30,000 Indian professionals working in Silicon Valley today and they have developed strong collective identity based on their common educational and

[12] Complete definition of IT includes telecom sector. Realizing the importance of telecommunications in the overall development of the economy in general and that of the IT and software service sector in particular ,the Government of India initiated a number of policy reforms that helped create a highly competitive environment ,leading to a drastic reduction in telecom costs but also increased access and better services(Kumar and Joseph,2005). To promote the use of Information Technology, the following mission targets have been spelt out. PC penetration of 10 per 1000 of the population .All colleges and schools are to be hooked to internet .Presence of Internet kiosks in every Panchayat ward and Modernisation and integration of Government functions using Information Technology are policy decisions to favor IT sector . At present, the Internet is confined to large cities in both nations. While these groups are far from saturated, high levels of pervasiveness will require service to the lower urban classes and villages, which raises issues of public access, service in villages, education and language, and affordability .Please refer to the study paper of the Indian Telecommunication Regulatory Authority which compares the Indian and the Chinese Telecommunication Market(Appendix II)

professional backgrounds. Many were graduates of the prestigious Indian Institutes of Technology. Professional associations such as The Indus Entrepreneur (TiE) and the Silicon Valley Professionals Association (SIPA) provide opportunities for networking and information sharing as well as role models and sources of finance for entrepreneurs. However this community remains local rather than a transnational one. Very few US-educated Indian engineers have returned to start companies or work for established firms.

This study is not comprehensive of the efforts taken by the government of India to promote outsourcing in the IT industry for one realizes that a myriad other public policies and private initiatives have fueled the growth of the sector as well. Nor does study attempt to argue that the four policies discussed are the best or most successful institutional projects with regards to promoting outsourcing. Rather, studying these three specific policies sheds light on potential components of a model that possibly could be replicated in other developing countries operating under similar conditions to those India faces

As evident from this study much before the first generation of reforms, that is, 1991, the government was pursuing a structuralist approach toward economic development. After liberalization in 1991, the government embarked on pro-active economic policies for the diffusion and production of IT although the platform was laid for the development of vibrant software industry in the 1970s and 1980s(explicit in Kumar and Joseph,2004) Consequently, the IT industry experienced an unprecedented growth rate in domestic as well as export markets. However, foreign direct investment (FDI) policies have not been successful in attracting the desired level of foreign investment, which is very important for a high-tech sector such as IT hardware manufacturing. The study suggests that immediate corrective measures need to be taken to augment the IT manufacturing industry, which can significantly contribute to national economic development and employment generation. To achieve sustained growth in the IT sector, high-quality professionals in adequate numbers are required. The new policy envisages continuous upgradation of standards at the school level with emphasis on physics, mathematics and English; make microelectronics and biology the new focus areas of tertiary education, updating the syllabus of computing engineering, electronics and IT in various technical institutions in line with the demands of industry. There were several initiatives taken up in the IT sector. The Ministry of Information Technology, set up in October 1999, was rechristened as the Ministry of Communication and Information Technology in September 2001. Setting up of National Task Force on HRD in IT, creating an IT Venture Capital Fund of Rs. 100-crore [US$22.22 million], upgradation of the Education and Research Network (ERNET) connecting various universities and regional engineering colleges (RECs) through a high-speed network and upgrading all RECs to the level of National Institutes of Technology, computerization of government departments by spending up to 3 percent of the budget on IT are among some key initiatives that were implemented.

Post-graduate engineering education and innovative research in IT are imperative in order to maintain quality and face new challenges in this dynamic sector. To maintain skilled knowledge-workers with the right mix of technical, business and functional skills, the workforce needs to increase by at least 10-fold by 2008. As per the NASSCOM-McKinsey report 1999, India needs to have at least 2.2 million knowledge workers in IT software and service-related areas by 2008.

The Media Lab Asia project was initiated in 2001 for taking IT to the masses. Enactment of a comprehensive law called the Information Technology (IT) Act, 2000, provides legal recognition for transactions through electronic data interchange. Many e-governance applications were initiated and a number of government portals were hosted. Technology development and content creation in Indian languages was promoted. The government initiated moves to set up 487 Community Information Centres at the block headquarters in the Northeastern states and Sikkim for bridging the digital divide.

Information and Communications Technology (ICT) can be used as an effective tool for rural development in India. One of the best examples of this is the adoption of ICT by a rural community in Warana, as part of a "Wired Village" project, in the state of Maharashtra. The Warana Group of Co-operatives (WGC) is using ICT to streamline operations connected with sugar cane growing and harvesting. The "Wired Village" project, launched in 1998, as a collaboration between the National Informatics Center (NIC), the Government of Maharashtra, the Warana Vibhag Shikshan Mandal (Education Department) and the WGC, was aimed at bringing agricultural, market and educational information to 70 villages around Warana Nagar. It also intended to simplify other business operations of the co-operative.

Moving on to a much sustainable model is the "e-Choupal." Launched in June 2000, it has already become the largest initiative among all Internet-based interventions in rural India."e-Choupal" services today reach out to more than 3.5 million farmers growing a range of crops, including soya bean, coffee, wheat,

rice, pulses, and shrimp, in over 31,000 villages through 5,200 kiosks across six states (Madhya Pradesh, Karnataka, Andhra Pradesh, Uttar Pradesh, Maharashtra and Rajasthan).

ITC's "e-Choupal" empowers over 3.1 million farmers by enabling them to access crop-specific, customized and comprehensive information in their native village habitat and language. Vernacular web sites, relating to each agricultural crop that ITC deals in, provide even marginal farmers with ready and real-time information on the prevailing Indian and international prices and price trends for their crops, expert knowledge on best farming practices, and micro-level weather forecasts. The "e-Choupal" model and movement has helped aggregate demand by creating a virtual producers' co-operative, thus facilitating access to higher quality farm inputs at lower costs for farmers.

Various corporates like Wipro have also been undertaking programs for the rural communities. The company has launched its Applying Thought In Schools Program, even as giant chip-maker Intel has introduced its Intel Innovation in Education initiative and Microsoft its project Shiksha. With all these interventions, corporates are now targeting school students in a major way, in order to leverage technology in education.

The CBFL program, developed by Tata Consultancy Services, is another case in point. It operates under the aegis of the Tata Council for Community Initiatives and uses a mix of methods-such as teaching software, multimedia presentations and printed material-to teach an uneducated person. The method is implemented using computers, which deliver the lessons in multimedia format to the learners. Supplementing computers in this process are reference textbooks of the National Literacy Mission. Today, the CBFL project is operational in more than 1,000 centers in Andhra Pradesh, Tamil Nadu, Madhya Pradesh, Maharashtra, Uttar Pradesh and West Bengal, and it has touched the lives of over 20,000 people. More centers are in the process of being set up.

E-government is the application of Information and Communication Technology (ICT) by government agencies. Its use promises to enhance the effectiveness and efficiency of government and alter its relationship with the public.

Electronic commerce allows efficient interactions among customer, suppliers and development partners cutting down on transaction time and reducing the costs of doing business. The role of government is to facilitate the development of E-Commerce.

Towards the goal of IT for all by 2008, government policies are provided for setting base for a rapid spread of IT awareness among the citizens, networked Government, IT-led economic development, rural penetration of IT applications, training citizens in the use of day-to-day IT services like tele-banking, tele-medicine, tele-education, tele-documents transfer, tele-library, tele-infocentres, electronic commerce, etc.

Another important consideration is whether lessons we can learn from India's success can be applied to other developing countries. The study outlines an export success model which can be adopted by other countries. India's skilled and English-speaking workforce was a clear contributory factor in the development of the IT industry. A starting point for other countries may be to launch education programs with focus on mathematics and science subjects and which can teach their citizens marketable skills for this and other industries, while being careful to distribute training among industries in case there are economic changes.

Furthermore, an interesting future area of research into this subject could examine if the growth of a white-collar industry has ramifications for India: drastic poverty, poor healthcare, national security and widespread hunger. If such growth is found to have positive effects, other developing countries would be well-advised to study the policies that the Indian government has undertaken and emulate them within the contexts of their own national goals. Though information technology is certainly not a panacea for the problems of developing countries , it may be a good first step.

The new structure of competition in IT provides unprecedented opportunities to leapfrog economically and the brain drain can be turned from a curse into an advantage in this process. Indian policymakers have a tremendous resource in the US to draw upon for policy advice, technical expertise, and managerial and entrepreneurial know-how. The challenge is to create an environment that will attract these immigrants back home as entrepreneurs, investors, and advisors. This will require actively addressing the regulatory, infrastructure and institutional constraints that have thus far limited their return a process in which they can also play a role.

References

1. Arora A.,V.S.Arunachalam,J Asundi and F.Ronald(2000),"The Indian Software Service Industry",Research Policy,Volume 30,pp1267-1287

2. Arora,A and S .Athreye(2002),"The Software Industry and India's Economic Development",Information Economics and Policy,14,pp 253-273

3. Balasubramaniam and lasubramaniam(1997),"Trade in Services: India's Computer Software", The World Economy, Volume 20,No.6

4. Baru, Sanjaya(2006),"IT and the E-Economy: The Balast for India-US Relations",in his book entitled Strategic Consequences of India's Economic Performance, Academic Foundation, New Delhi

5. Dedrick,J and Kraemer,K.L.(1993),"India's Quest for Self Reliance in Information Technology: Costs and Benefits of Government Intervention",Centre for Research on Information Technology an Organizations, Univerity of California,Irvine

6. Heeks, R., (1996). India's Software Industry: State Policy, Liberalization and Industrial Development, New Delhi: Sage Publications.

7. Heeks, Richard and Brian Nicholson(2002),"Software Export Success Factors and Strategies in Developing and Transitional Economies", Institute for Development Policies and Management Development Informatics Working Paper no. 12,University of Manchester ,UK

8. India,(1985),"New Computer Policy", The Gazette of India,31 August, pp 679-681

9. India,NTITSD,1998.IT Action Plan(three volumes),New Delhi, National Taskforce on Information Technology and Software Development

10. India, MIT,(2000),"Action taken Report of the National Task Force on Information Technology and Software Development", IT Action Plan: Part I, Ministry of Information Technology, New Delhi

11. Kumar, Nagesh and Joseph,K.J(2004),"National Innovation Systems and India's IT Capability: Are there Any Lessons for ASEAN New Comers?" ,RIS Discussion Paper No 72/2004,RIS,New Delhi

12. Kumar, Nagesh and Joseph,K.J(2005),"Export of Software and Business Process Outsourcing from Developing Countries:Lessons from the Indian Experience, Asia Pacific Trade and Investment Review, Volume 1,No.1,April 2005

13. Kumar, Nagesh (2000a). 'New Technology Based Small Service Enterprises and Employment: The Case of Software and Related Services Industry in India', paper presented at the National Seminar on Strategic Approach to Job Creation, February 2000, organized by ILO-SAAT and Ministry of Labour, at Suraj Kund, Haryana.

14. Kumar, Nagesh (2000b). Developing Countries in International Division of Labour in Software and Service Industries: Lessons from Indian Experience, background paper for World Employment Report 2001, Geneva, ILO

15. Kumar, Nagesh (2001). National Innovation System and the Indian Software Industry Development, Background Paper for World Industrial Development Report 2002, Vienna: UNIDO.

16. Joseph, K.J(2002),"Growth of ICT and ICT for development:Realities of the Myths of Indian Experience",Discussion Paper,No 2002/78,WIDER-UNU,August

17. Joseph,K.J and K.N.Harilal(2001),"Structure and Growth of India's IT Exports: Implications of an Export-Oriented Growth Strategy",Economic and Political Weekly,volume 36,No.34,pp3263-3270

18. Nagala,Sarla V(2005),"India's Story of Success: Promoting the Information Technology Industry", Stanford Journal of International Relations, Volume 6,Issue 1

19. Nasscom website (www.nasscom.in)

20. NASSCOM. 2004. *Indian ITES-BPO Industry Handbook 2004.* New Delhi: National Association of Software and Service Companies.

21. Parthasarthy, A and K.J.Joseph(2002),"Limits to Innovation Set by Strong Export Orientation:The Experience of India's Information and Communication Technology Sector",Science,Technology and Society,volume 7,no.1,pp13-49
Magazines-Dataquest

22. Schware,R(1987),"Software Industry in the Third World:Policy Guidelines,Institutional Options and Constraints",World Development,Volume 15,No 10/11,pp 1249-1267

23. Schware, R. 1992. "Software Industry Entry Strategies for Developing Countries: A Walking on

Two Legs' Proposition. In World Development, Vol. 20

24. Sen.P(1995),"Indian Software Exports:An Assessment",Economic and Political Weekly,volume 30,No 7& 8,pp 2053-2058

25. Subramaniam, C R(1992):India and the Computer, Oxford University Press, New Delhi.

31. Financial Times, Various Issues

32. Data Quest, Various Issues

Appendix Table I: Targets for the Year 2008

Opportunities in Software Sector by the Year 2008 (McKinsey Report, in Department of IT webpage http://www.mit.gov.in/dbid/eproduction.asp)

Software Sector		
	Total Market	**Exports**
IT Services	$ 28-30 billion	$ 28-30 billion
Software Products	$ 8-11 billion	$ 8-11 billion
IT Enabled Services	$ 21-24 billion	$ 21-24 billion
Domestic Market	$13-15 billion	
Total	**$ 70-80 billion**	**$ 57-65 billion**

Rs. Crore

	Realistic Scenario	**Optimistic Scenario**
2001-02	32,750	32,750
2002-03	39,500	41,600
2003-04	45,000	50,500
2004-05	52,000	61,500
2005-06	60,000	74,700
2006-07	69,000	90,900
CAGR	**15**	**22**

IT Industry by the year 2008

Employment Generation Year 2003-04	
IT Exports • **35% of India's Total Exports in 2008** • **from 21.3% during 2003-04**	
Share of IT Software & Services Industry in GDP • **Likely to be 7% of GDP in 2008** • **from 2.64% of GDP during 2003-04**	

Employment Generation Year 2003-04	
Software Sector & Service Sector of which ITES& BPO Sector	813,500 245,000
Hardware Sector Direct Employment Indirect Employment	395,000 250,000

Hardware Production Projections for Tenth Plan

Employment Generation Year 2008	
Software Sector	2.2 million
ITES / BPO Industry	1.1 million
Indirect Employment	2-3 million
Hardware Sector Direct Employment	1.6 million
Indirect Employment	3.2 million
Total Employment Generation	**over 9 million**

Appendix Table II
Telecom Regulatory Authority of India Press Release No59/2006
Study Paper on Financial analysis of Telecom Industry of China and India

Salient features of Study Paper
i. The Growth of mobile services in India over the past few years has been phenomenal. Mobile subscribers are growing at a CAGR of around 85% since 1999. Now over 4 million mobile subscribers are added every month. On the other hand China is registered a growth of 16% in the mobile subscriber base in the year 2005 with monthly addition of 5 million subscribers every month.

ii. Total telecom revenue of Chinese telecom companies increased from $ 65 billion to $ 72.70 billion during the calendar year 2005. Telecom revenue in India during 2005-06 was $19.50 billion

iii. ARPU (Average Revenue Per User) in India and China is comparable in GSM pre-paid segment but ARPU for post paid segment in China is much higher. ARPU for CDMA services are also higher in China in comparison to India.

iv. ARPU for Basic Telephone Services is higher in India when compared to ARPU for Basic Telephone services in China.

v. Minutes of Usages (MOU) of cellular mobile Telephone services are much higher in India when compared to China's Cellular mobile telephone services.

vi. Minutes of Usage of GSM and CDMA based cellular mobile telephone services in India are 32% and 70% respectively higher when compared to Chinese cellular mobile telephone services.

vii. Lower ARPUs in India inspite of higher usage due to much lower tariffs in India.

viii. The capital employed per subscriber for the Basic Service is much lower when compared to India. However capital employed for the cellular segment is lower in India.

ix. Chinese companies are able to generate higher rate of EBITDA margin than Indian companies.
x. Chinese Companies earn higher rate of return on the capital employed (RoCE) than Indian companies

xi. Indian mobile market is much more competitive when compared to the Chinese mobile market.