Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: E

Network, Web & Security

Highlights

- Data Hiding and Extraction
- Bandwidth Estimation Methods
- Unequal Clustering Protocol
- Computational Forensic Expertise

Discovering Thoughts, Inventing Future

VOLUME 22 ISSUE 2 VERSION 1.0

Global Journal of Computer Science and Technology: E
Network, Web & Security

Global Journal of Computer Science and Technology: E
Network, Web & Security

Volume 22 Issue 2 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

Editorial Board

Global Journal of Computer Science and Technology

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajesh Kumar Rolen

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MIcrosoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

Contents of the Issue

Global Journal of Computer Science and Technology: E Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Use of Techniques and Tools for Investigative Process in Computational Forensic Expertise

By Priscilla Leão de Lima

Abstract- Computer forensics investigates and retrieves information about a fact, as well as examining digital evidence that can be decisive in any technological situation. The research will explain techniques that are used during the expertise, ensuring the integrity of the data so that the analysis is not impaired. First, a bibliographic survey will be carried out in search of concepts about data collection and analysis techniques in the for ensic area. The method includes performing procedures such as equipment identification and chain of custody in a simulated environment, in orderto determine the best mechanism for the presented scenario. Finally, the mainsteps and forensic tools were presented for a betterway in which the expert can use to perform the exactan alysis of the crime.

Palavras-Chaves: *evidências. investigação. computaçãoforense.*

GJCST-E Classification: *DDC Code: 004.62 LCC Code: T58.5*

Strictly as per the compliance and regulations of:

Use of Techniques and Tools for Investigative Process in Computational Forensic Expertise

Priscilla Leão de Lima

Abstract- **Computer forensics investigates and retrieves information about a fact, as well as examining digital evidence that can be decisive in any technological situation. The research will explain techniques that are used during the expertise, ensuring the integrity of the data so that the analysis is not impaired. First, a bibliographic survey will be carried out in search of concepts about data collection and analysis techniques in the for ensic area. The method includes performing procedures such as equipment identification and chain of custody in a simulated environment, in orderto determine the best mechanism for the presented scenario. Finally, the mainsteps and forensic tools were presented for a betterway in which the expert can use to perform the exactan alysis of the crime.**

Palavras-Chaves: evidências. investigação. computaçãoforense.

I. Introdução

E indiscutível que os avanços tecnológicos tenham se difundindo em nossa sociedade, porém esse desenvolvimento ampliou crimes ocorridos na Internet, atingindo particulares como também grandes empresas.

Pode-se afirmar que em razão desta situação, foram criadas técnicas para investigação e manipulação de evidencias eletrônicas, como a forense computacional, " é a ciência responsável por adquirir, preservar, identificar, extrair, restaurar e documentar evidências computacionais, processadas eletronicamente e armazenadas em mídias computacionaisll (ATÍLIO, 2003 apud SOUSA et al., 2006).

Para este, também há outra interpretação como "Forense Computacional: O estudo de como as pessoas usam o computador para causar danos, ferir e mesmo destruir. " (MOHAY et al., 2003, p. 1)

Em base desse estudo, também possuímos respaldo legal conforme o Código de Processo Penal nos Artigo 170 "Os peritos deverão guardar material suficiente para nova perícia" e Artigo 171 "Nos crimes cometidos com destruição ou rompimento de obstáculo a subtração da coisa, ou por meio de escalada, os peritos, além de descrever os vestígios, indicarão com que instrumentos, por que meios e em que época presumem ter sido o fato praticado". Em que deverá ser feito cópias reservas, caso acontecer algum imprevisto na análise e apresentar todos os detalhes das evidencias para compor seu laudo judicial.

Assim, este trabalho explanará técnicas para melhor manuseio de vestígios em uma análise de perícia forense computacional, para que no final da investigação possa garantir o valor judicial de uma prova eletrônica.

II. Forense Computacional

Segundo Queiroz e Vargas (2010), a forense computacional é um conjunto de procedimentos e metodologias com a função de investigar e armazenar evidências que possam responder se houve ou não um crime, tendo como base de análise equipamentos de processamento de dados (computadores pessoais, laptops, servidores, estações de trabalho ou outras mídias eletrônicas).

O principal objetivo deste tipo de perícia forense pode ser definido como a coleta de vestígios relacionados ao crime investigado, os quais possibilitem a formulação de conclusões sobre o caso (REIS, 2003).

O intuito é coletar todas as evidências, buscando indícios virtuais com a aquisição, a identificação, a extração e análise de dados que estejam em uma mídia computacional, possibilitando comprovar de uma forma legal em que realmente ocorreu o crime ilícito de acordo com a investigação realizada pelos peritos forenses.

A Computação Forense determina à dinâmica, a materialidade e autoria de ilícitos ligados à área de informática, tendo como questão principal a identificação e o processamento de evidências digitais em provas materiais de crime, por meio de métodos técnico-científicos, conferindo-lhe validade probatória em juízo. (ELEUTÉRIO; MACHADO, 2010, p. 16). Portanto, é basicamente uma área nova e está sendo desenvolvida para combater crimes cibernéticos. Com a expansão da internet, está ocorrendo a ampliação de fraudes eletrônicos, para estes casos, a perícia forense é uma ferramenta eficiente para identificação desse delito e assim reduzir riscos.

III. Processo De Investigação Da Forense Computacional

Segundo Noblett, Pollitt e Presley (2000), para que os resultados da perícia sejam válidos, é necessário que sejam postos em práticas

Author: ESP (Instituto de Especialização do Amazonas)
e-mail: prilima2413@gmail.com

procedimentos e protocolos (documentados) que garantam assim os requisitos legais e técnicos para a evidência pericial.

Portanto, o processo investigativo da Forense Computacional deve assegurar a integridade dos vestígios coletados, porém devido à volatilidade das evidências eletrônicas, essa tarefa pode ser considerada difícil. Sendo assim, para garantir a integridade e confiabilidade das evidências coletadas, o perito forense deve seguir procedimentos e protocolos reconhecidos pela comunidade científica, e a cada passo, deve detalhar e revisar a documentação desenvolvida, para que deste modo, evite erros durante o processo investigativo (EOGHAN, 2002).

De acordo com Eleutério e Machado (2001), a Computação Forense tem quatro etapas do processo de computação forenseprincipais:

Coleta: Segundo Vargas (2007), os procedimentos adotados, na coleta de dados, devem ser formais, seguindo normas internacionais de padronização e padrões de como se obter provas para apresentação judicial, como um *checklist.*

Obtendo todas as evidências, realizar cópias garantindo sua integridade, etiquetar os vestígios, como colocar adesivos com cores distintas para identificar se o equipamento estava ligado ou não, realizar vídeos ou fotos do cenário, sendo possível verificar quais programas estavam em execução e suas conexões de rede.

Itens que requerem atenção especial durante a documentação, devendo ser fotografados, são (STEEL, 2006):

- Telas do computador, com resolução suficiente para leitura de textos ali presentes, se necessário;
- Conexões de rede, mostrando quaisquer cabos de rede conectados ao computador. As duas pontas do cabo devem ser fotografadas, para o caso em que o perito tenha que provar que o computador estava conectado a algum equipamento específico;
- Conexões de periféricos, para provar que estavam conectados ao computador.

Nesta etapa, pode empregar-se a Forense *Em Vivo*, que dependendo do contexto, podemos encerrar a análise com o desligamento do computador de forma abrupta, o que classicamente é procedimento tomado pelo perito (MELO, 2009). Em que o perito utilizará suas técnicas e procedimentos que possam produzir vestígios na investigação ainda em funcionamento.

Outro fator importante, é verificar na ordem judicial se há restrição de coleta de algum equipamento que não possa ser retirado do lugar do acontecimento do crime. Alguns equipamentos possíveis de fontes de dados são computadores, maquinas fotográficas, pen drives, dispositivos de armazenamento em rede, entre outros.

E para Cansian (2000), é importante sempre fazer a coleta de dados de acordo com a ordem de maior volatilidade para a de menor, dos elementos mais utilizados:

- Registros de memória periférica, cache;
- Memória principal;
- Estado da rede, rotas, interfaces;
- Processos em execução;
- Discos e partições;
- Fitas, disquetes e outros meios magnéticos;
- Em mídias como CD-ROMs e cópias impressas.

Os peritos devem coletar esses dados voláteis o mais rápido possível para não ser perdido nenhum aspecto relacionado ao crime que possam ser de extrema importância para uma das etapas da perícia forense computacional, a análise dos dados.

Segundo BATTULA (2000), Imagem e Espelhamento são técnicas de duplicação/cópia utilizadas na fase de coleta A cópia dos dados é realizada através de ferramentas apropriadas para duplicação de dados, como o utilitário dd (Linux) que é capaz de recolher dados voláteis que estão englobados na memória e replicar dados não voláteis.

Entretanto, as coletas devem ser realizadas minuciosamente para que não ocorra nenhuma falha, resultando erros em sua análise, logo o laudo pericial não estará coerente com a investigação.

Exame: identificar, extrair, filtrar e documentar dados relevantes a análise, buscando até aquele não estarão explícitos que serão manipuladas por ferramentas forenses adequadas para perícia digital.

Segundo Kerr (2001, apud VARGAS, 2007, p. 21), um perito forense computacional deve garantir que uma evidência será manuseada e protegida de tal forma que não seja danificada, destruída ou até comprometida. Isso pode se dá, pelo mal uso e escolha dos procedimentos a serem introduzidos.

Para que não seja danificada, pode ser utilizado o HASH(MD5/SHA1/SHA256) que estabelecem a sequência de caracteres de tamanho fixo, certificando que os dados coletados não sofram nenhum dano, garantindo a sua integridade. Como de acordo com Eleutério e Machado (2010): "O que torna esse tipo de função extremamente utilizada para a verificação de integridade de dados computacionais é o fato que uma simples alteração na informação de entrada do algoritmo gerará uma sequência de bits (valor hash) completamente diferente."

Esse artefato consegue assegurar até alguns tipos de arquivos que facilitam a ocultação de dados, evitando sua descoberta, em que uma das técnicas mais conhecida e utilizada sobre este caso é o uso de esteganografia.

O processo de esteganografia consiste em esconder uma informação através de uma mensagem de menor importância, conhecida como mensagem de

cobertura. Após inserir os dados na mensagem de cobertura, obtém-se o chamado estego-objeto, que é uma mensagem inócua que contém secretamente uma mensagem de maior importância (ROCHA, 2004).

Com esses dados contidos nos equipamentos que foram capturados no local do fato ocorrido deve ser registrado toda as informações como data, hora, tamanho, descrições, observações, entre outros, relatando todo período da investigação.

Freitas (2006) lista alguns exemplos de procedimentos para a preservação dos dados, para que estes não sejam comprometidos por qualquer ação:

- Inicialmente, devem-se criar imagens do sistema investigado, para que as potenciais provas possam ser posteriormente analisadas;
- Se o caso necessitar de uma análise ao vivo, salvar as evidências em dispositivos e bloqueá-los contra regravação;
- Todas as evidências deverão ser lacradas em sacos e etiquetadas;
- A etiqueta deverá conter um número para a identificação das evidências, o número do caso, a data e o horário em que a evidência foi coletada e o nome da pessoa que a está levando para a custódia, além do nome de quem coletou essas evidências;
- Etiquetar todos os cabos e componentes do computador, para que, depois, possam ser montados corretamente quando chegar ao laboratório;
- Os HDs deverão ser armazenados em sacos antiestática, para evitar danos e, também, para que os dados não sejam corrompidos;
- Durante o transporte das provas, tomar cuidado com líquidos, umidade, impacto, sujeira, calor excessivo, eletricidade e estática;
- Quando já tiverem sido transportadas, as evidências deverão ser armazenadas e trancadas para evitar a adulteração até o momento em que poderão ser examinadas e analisadas;
- Todas as mudanças feitas durante essa fase deverão ser documentadas e justificadas (cadeia de custódia);

Uma das técnicas para duplicação de uma cópia idêntica dos equipamentos aprendidos é o de imagem e espelhamento. Segundo (ELEUTÉRIO; MACHADO, 2011) relata que essas técnicas, ao serem realizadas através de softwares e equipamentos forenses, garantem uma cópia fiel dos dados e consequentemente a preservação correta do material que foi apreendido.

O espelhamento é uma técnica de duplicação que realiza uma cópia exata e fiel dos dadoscontidos em um dispositivo de armazenamento computacional para outro (ELEUTÉRIO; MACHADO, 2011). Uma reprodução realizada de bit a bit para local de destino almejado, porém esse disco deverá conter o tamanho ideal ou maior para finalizar esta cópia com êxito.

E a imagem também é uma cópia fiel de drivers, sistema operacionais, configurações, entre outros é construída arquivos de imagem de disco, contendo toda base do armazenamento realizado.

De acordo com Eleutério e Machado (2011), a técnica de imagem possui algumas vantagens se comparada com o espelhamento:

- Um dispositivo de destino pode armazenar diversas imagens de disco, se houver capacidade;
- Possibilidade de compactação dos arquivos de imagem; facilidade de replicação das imagens de disco, uma vez que podem ser copiadas por qualquer sistema operacional;
- Eventuais setores defeituosos no dispositivo de destino são tratados pelo sistema operacionais

Recomenda-se utilizar o recurso de bloqueio de escrita para acessar as mídias sem realizar nenhuma modificação no conteúdo. Pode ser feito no software colocando o computador do investigado em somente leitura. E também por hardware, que há vários equipamentos no mercado, com custo baixo, mas também há aqueles mais aprimorados.

Segundo Eleutério e Machado (2001), Os equipamentos EspionForensics e o Forensic Bridge Tableau são os mais utilizados para bloqueio de escrita em discos, já o software ICS Write ProtectCard Reader é o mais utilizado para bloqueio de escrita em cartões de memória.

Posteriormente de realizar estes mecanismos para garantir a integridade, há a extração de todos os dados etapa, que são a extração que irá remover das mídias apreendidas tudo o que for importante para a investigação, recuperando os dados que foi possivelmente excluído intencionalmente pelo suspeito do delito.

De acordo com Eleutério e Machado (2001), Data Carving, que na computação refere-se à recuperação de arquivos apagados, é uma técnica realizada através da localização de assinaturas conhecidas (por exemplo, cabeçalhos que contêm a identificação do tipo de arquivo).

Para organizar os arquivos são indexados os dados, para que em uma busca, sejam localizados rapidamente, podendo ser realizado por palavras-chave no conteúdo do material examinado.

Esta técnica muito eficiente e muito utilizada na etapa de análise é a busca por palavras-chave, sendo essa técnica disponível em muitos softwares de análise de arquivos. (ELEUTÉRIO; MACHADO, 2011).

Com todos esses procedimentos, os peritos são capazes de realizar cópias exatas do que é

pretendido, capturando até dados que possivelmente foram ocultos para camuflar uma evidência.

Profissionais habilidosos de Forense Computacional conseguem descobrir informações escondidas em espaços desalocados utilizando ferramentas adequadas, porém dados omitidos são mais difíceis de ser encontrados e também de ser utilizados como evidências em audiências não-técnicas (KESLER, 2007).

Portanto, é necessário realizar a operação com bastante atenção pois os dados que possivelmente ser úteis para o caso, possam estar em locais imprevisíveis e que o perito deve estar apto para a identificação e recuperação desses dados.

Analise: identificar (pessoas, locais e eventos) e a correlacionam, reconstruindo acena a fim de encontrar vestígios com paramentos adequados com o crime investigado.

Segundo Kerr (2001, apud VARGAS, 2007, p. 21), a análise será a pesquisa propriamente dita, em que o investigador pode se deter, especificamente, nos elementos relevantes ao caso em questão, pois todos os filtros de camadas de informação anteriores já foram transpostos.

Na análise, o perito examina os dados que foram coletados, verificando os parâmetros das evidencias interligadas ao crime, utilizando ferramentas adequadas para desvendar o caso.

De acordoEckert (1997), o valor de uma evidência é medido por quatro parâmetros:

- Relevância: que descreve a importância da evidência no contexto ou escopo dos fatos ocorridos;
- Materialidade: que descreve capacidade da evidência em ajudar a reproduzir os fatos ocorridos;
- Credibilidade: que descreve o meio pelo qual a evidência foi obtida;
- Competência: que descreve o nível de validade dos procedimentos científicos empregados na análise e teste da mesma.

Dessa forma, deverão ser aplicadas ferramentas e técnicas que efetuem uma cópia fidedigna dos dados e mantenham a integridade do material apreendido. (ELEUTÉRIO; MACHADO, 2011).

Nesta fase, há grandes chances de ocorrer equívocos durante a análise devidoa grande demanda de dados que nem sempre é evidente, como apresenta CASEY, 2006. Essa é uma fase que além de consumir muito tempo, está muito suscetível a equívocos, pois depende muito da experiência e do conhecimento dos peritos, já que são poucas as ferramentas que realizam esse tipo de análise com precisão.

Outro artefato, é proceder com o equipamento ainda ligado para não que os dados não sejam perdido se o computador reinicie ou for desligado bruscamente. Segundo Bertoglio (2008), a *Live forensics*, ou análise ao vivo, representa a perícia que é feita em um computador ou equipamento ainda em funcionamento.

De acordo comAdelstein (2006, apud BERTOGLIO 2008, p. 31) um desafio para a análise ao vivo é o fato de que o sistema não é estático, arquivos e processos estão sempre mudando. Para que a aquisição dos dados seja realizada corretamente, é adotado um conjunto de melhores práticas que busca melhorar a qualidade das evidências. É importante seguir essas práticas, já que trabalhar com um computador em funcionamento dificulta a coleta de provas, uma vez que qualquer ação pode afetar outros elementos das informações requeridas.

Dessa forma, a analise necessita de conhecimento maior do perito para realizar toda a pesquisa e escolher ferramentas propicias para ter um laudo correto sem cometer nenhuma injustiça, pois é nessa fase que ocorrerá a conclusão do fato que será entregue para o poder judiciário.

a) Resultados Obtidos

Basicamente, esta é última etapa da investigação do caso examinado em que irá ser gerado um laudo constando informações necessárias para a apuração do crime. Para Kent e outros (2006, apud PEREIRA et al., 2007) nessa fase, é necessária a organização da documentação necessária para a criação do laudo pericial, sendo necessários alguns procedimentos como: reunir toda a documentação e anotações geradas nas etapas de coleta, exame e análise, incluindo as conclusões obtidas; identificar os fatos que fornecerão suporte as conclusões descritas no laudo pericial;listar as conclusões obtidas; organizar e classificar informações recolhidas para garantir um laudo conciso e inquestionável.

"É importante que a conclusão apresentada no relatório seja imparcial e final, de forma a não favorecer alguma decisão. Por esta razão, a etapa de apresentação é a fase conclusiva da investigação". (PEREIRA *et al.*, 2007).

Com os resultados obtidos durante toda a investigação, é de fundamental importância redigir o laudo, anexando evidências detalhadas encontradas no material examinado e demais documentos encontrados na cena do crime, concluindo toda analise realizada. Porém, esse conteúdo que será adicionado ao relatório, deverá ser organizado em seções para não ocorrer informações dispersas e assim deixar o laudo inexato.

Para Kent *et al.* (2006), o conteúdo do laudo pericial torna-se um documento de fácil interpretação e, para isso, deve ser organizado em seções:

- Finalidade da investigação: explicar claramente os objetivos do laudo;
- autor(es) do relatório: listar todos os autores e co-autores do relatório, incluindo suas especialidades e responsabilidades, durante a investigação;

- Resumo do incidente: síntese explicando o incidente investigado e suas consequências;
- Relação de evidências: relacionar e descrever todos objetos, onde se encontram, estado, como, quando e por quem elas foram adquiridas no decorrer das investigações;
- Detalhes: fornecer uma descrição detalhada de quais evidências foram analisadas, quais os métodos utilizados e quais as conclusões alcançadas, descrevendo os procedimentos e as técnicas adotados, durante a investigação;
- Conclusão: os resultados da investigação devem ser somente descritos, citando especificamente as evidências que comprovem as conclusões. A conclusão deve ser clara e não oferecer dupla interpretação;
- Anexos: todas as documentações devem ser anexadas, ao final do laudo, tais como: diagramas da rede, formulários descritivos dos procedimentos utilizados, formulário de cadeia de custódia e informações gerais sobre as tecnologias envolvidas na investigação;
- Glossário: adicionar um glossário dos termos utilizados no laudo, que poderá esclarecer muitas dúvidas que possam surgir durante a leitura do juiz e/ou dos jurados.

Nessa fase do processo da perícia, é apresentada todo o procedimento produzido durante todo o ciclo para poder ser entregue ao jurídico e assim obter um julgamento dos dados examinados.

De acordo com Freitas (2006), nessa fase ocorre o enquadramento das evidências no formato jurídico, por isso, também, é conhecido como substanciação das evidências, sendo isso feito pelo juiz e/ou advogados na esfera civil ou criminal.

E com todas as evidências recolhidas, o perito descreverá tudo o que foi encontrado, relatando indícios que podem ajudar a desvendar o crime investigado, tornando-se encarregado de realizar o laudo pericial.

Como responsável pelo laudo operitodeveenvolvertodas as pessoas que achar conveniente em sua investigação, de acordo com (REINALDO, 2007)

E segundo Reis e Geus (2004), a elaboração do laudo é o último passo da perícia, neste momento o perito tem a liberdade de descrever o incidente, sendo ele o único responsável pelo documento e todo o seu conteúdo.

Dessa forma, com a conclusão do laudo, os documentos deverão ser armazenados e poderá também ser anexado a mídia com o material coletado. Neste, é necessário conter todas os requisitos que foram solicitados, bem como a descrição das provas e assim ser entregue ao solicitante.

Portanto, é de fundamental importância que o laudo contenha uma linguagem clara para que todas as pessoas que tiverem acesso a esse material possam entender e compreender o que realmente o perito concluiu da investigação.

IV. Equipe Forense

Para obtenção de um laudo coerente, é necessário a composição de uma equipe preparada e especializada para não ocorrer falhas durante qualquer investigação, do início até sua finalização.

Segundo Ng (2007) diz que a equipe forense deve ter algumas funções como:

- Identificar atividades suspeitas e realizar o processo de investigação;
- Tratar as atividades suspeitas que não foram identificas pela equipe;
- Definir níveis de criticidade e um tempo de *report*para cada um dos eventos;
- Realizar*reports*periódicos a respeito dos processos de investigação;
- Coletar e documentar as evidências encontradas;
- Definir, manter e gerenciar um local para armazenar as evidências;
- Envolver todos os profissionais que podem auxiliar no processo de investigação;
- Realizar atividades de acordo com as políticas da organização e das leis vigentes;
- Seguir a metodologia de análise forense computacional definida e implementada.

Como todo trabalho, é essencial ter uma equipe, em que possa ser dividido as tarefas, mas na perícia essas pessoas contém um maior conhecimento, são especializadas na área forense e assim gerar buscas e analises de evidências mais eficazes.

V. Obstaculos Durante a Investigação

Segundo REIS, 2003, podem ocorrer muitas dificuldades na coleta e análise de vestígios deixados na máquina utilizada no ato ilícito, sendo que a quantidade de evidências deixadas é inversamente proporcional às habilidades apresentadas pelo criminoso.

Durante a investigação, poderão ocorrer algumas dificuldades que o perito irá enfrentar como a grande quantidade de arquivos, por isso é necessário que os requisitos do laudo estejam bem claros, não deixando sentido genérico. Então, é importante que o a autoridade solicitante busque sempre detalhar o quê procura, descrevendo no máximo de detalhes possível, ou seja, que mostre para a equipe pericial exatamente o quê deve ser buscado, para dessa forma, evitar desperdício de trabalho dos peritos. (ALMEIDA, 2011).

Também há equipamentos que contém senhas,criptografia onde as informações são escritas

em códigos, esteganografia que oculta seu verdadeiro sentido."Esteganografia é uma palavra de origem grega, onde Stegano significa escondido ou secreto e Grafia: escrita ou desenho", conforme Coelho e Bento (2004), entre outros obstáculos e para quebra desses artefatos, uso de fermentas forenses são adequados para estes procedimentos.

Entretanto, um mecanismo que pode ser útil no caso examinado é a engenharia social. De acordo com (SOCIAL-ENGINEER, 2015) engenharia social é um método de ataque, no qual alguém faz uso da persuasão, diversas vezes abusando da ingenuidade ou confiança do usuário, com a finalidade de obter informações que possam ser utilizadas no acesso não autorizado a computadores ou informações. Facilitando o acesso das senhas e o tempo da analise pericial.

VI. Ferramentas Periciais

Com o advento e disseminação da tecnologia nesses últimos anos, as infracções, invasões, venda e roubo de informações privilegiadas, pirataria, envio de e-mails falsos, tentativas de acessos indevidos à organizações ou até mesmo à pessoas comuns vêm se aumentando cada vez mais e por isso há a necessidade do auxílio de ferramentas mais modernos e incrementada para busca de infractores, além da necessidade de padronização das buscas e apresentação das evidências mais consistentes. (Vargas, 2006)

Em uma perícia, é de extrema importância a utilização de ferramentas que irão auxiliar no desenvolvimento da análise do caso, examinando e resgatando todas as evidências para êxito do laudo investigativo.

De acordo com Eleutério e Machado (2011), as ferramentas que são destaques na etapa de análise são os softwares Encase e o software Forensic Toolkit (FTK), pois eles além de serem uteis em todas as etapas do processo forense computacional, eles ainda têm diversas funcionalidades que são fundamentais para a etapa de analise, como: as buscas por palavra-chave, a navegação adequada pelos arquivos e pastas da base de dados, o KFF, entre outras.

O software ForensicToolKit (FTK) e o software Encase são soluções proprietárias compatíveis com o Windows. Ambas possuem diversas funcionalidades que possibilitam a realização de diversas técnicas para perícia forense computacional. Já como opções para Linux, o autor destaca a utilidade de alguns softwares para a etapa de preservação e coleta: DC3DD e Guymager, e destaca também os sistemas Linux CAINE e FDTK-Ubuntu para mesma finalidade. (ELEUTÉRIO; MACHADO, 2011).

Muitos softwares de computação forense trabalham com indexação de dados, dentre estes, destaca-se o FTK e o Encase. (BATTULA, 2000).

a) Encase

O Encase possui a capacidade de formar o *timeline*(linha do tempo), de forma gráfica. Este recurso ajuda muito a estabelecer a temporalidade dos dados encontrados e mostrar de forma clara as datas relativas aos arquivos encontrados e ou existentes no sistema (COSTA, 2003).

Pode-se citar como uma das principais vantagens no uso do Encase a sua documentação clara, extensa e cheia de ilustrações (LAWRENCE, 2009). (LAWRENCE, 2009). A interface organizada permite ao usuário visualizar os dados através de três maneiras diferentes, sendo que essas visões incluem galerias de fotos e imagens de evidências. O software EnCaseForensic também pode ser utilizado para analisar diferentes tipos de mídias como *Palm tops* e a maioria de unidades removíveis (LAWRENCE, 2009).

De acordo com Christófaro (2006) a ferramenta EnCase da Guidance Software surgiu no cenário da Computação Forense em 1998 nos Estados Unidos, dois anos depois ela se tornaria a principal ferramenta forense. Naquela época, a maioria dos examinadores se utilizavam do prompt dos sistemas operacionais em suas investigações; a proposta de um ambiente compatível com os populares ponteiros e janelas do Windows da Microsoft era ousada, uma vez que na visão dos examinadores forenses o prompt era mais poderoso e ainda oferecia mais opções de controle, além do mais a ferramenta irá ser composta por quatro recursos básicos que permitem sua funcionalidade abrangente. Esses recursos serão vistos a seguir (GUIDANCE, 2007)

- Análise detalhada do sistema: a ferramenta é capaz de descobrir arquivos ocultos e excluídos, detectar rootkits, procurar documentos, identificar processos de invasores, reconstruir atividade de Web e de e-mails, até decodificar certos tipos de criptografia e identificar comunicação não autorizada na rede; – análise paralela: este tópico analisa com rapidez um grande número de computadores ao mesmo tempo, reunindo informações críticas sobre seu estado e conteúdo.
- A análise paralela:é o recurso básico que permite produzir velocidades de pesquisa empresarial e reação a incidentes muito superiores às tecnologias concorrentes;
- Correção: após a identificação de um evento mal-intencionado, a ferramenta auxilia a detê-lo e controlá-lo. Em quase todos os casos de reação a incidentes, é possível ver o problema, mas não fazer algo a respeito ou não fazer nada, ou então, usar ferramentas de terceiros para remediar a situação, o que pode significar a desativação de pelo menos uma parte da rede. Com a ferramenta pode-se documentar o incidente detalhadamente,

acessar os computadores comprometidos e corrigir o problema;

- Integração: a ferramenta é capaz de ser integrada à infra-estrutura de segurança existente na empresa para proporcionar reação a incidentes em tempo real automatizada. Os alertas gerados por tecnologias de monitoração, com os sistemas IDS1 e SIG2, ativam reações automatizadas pela ferramenta, permitindo aos profissionais de segurança reagir a centenas e potencialmente a milhares de alertas por dia, logo após o evento ocorrer.

b) Accessdata Forensic Toolkit

O software AccessDataForensic Toolkit, também conhecido como FTK, é considerado de fácil utilização para profissionais que estão familiarizados com ferramentas forenses (LAWRENCE, 2009). Este conjunto comercial de ferramentas além de conter limpadores de mídias, que são utilizados para salvar imagens de discos rígidos em mídias removíveis de maneira limpa e íntegra, possui programas para recuperação de dados e discos, assim como e visualizadores de registros e outros utilitários (LAWRENCE, 2009).

O FTK pode ser utilizado apenas em plataforma Windows e Linux, sendo que desse modo apresenta uma desvantagem se comparado a outros softwares que suportam mais modelos de sistemas de arquivos. Além do software possuir sua própria ferramenta para criação de imagens, o FTK pode ler imagens produzidas pelo Encase, pelo Linux DD, Safeback e outros softwares forenses (LAWRENCE, 2009).

i. *Ftk Imager (Access Data)*

Este programa é extremamente útil para a coleta de dados em local de busca e apreensão, sendo disponibilizado gratuitamente pela empresa AccessData (ACCESSDATA, 2011).

Esta ferramenta possui apena três funcionalidades, para a linha de comando, no entanto, não é um programa de linhas de comando, mas proporcionam em que sejam automatizadas as coletas dos dados. Portanto, as linhas de comando suportadas são (ACCESSDATA, 2011):

- /CreateDirListing – Cria um arquivo de lista de diretório na pasta onde o "FTK Imager" é executado;
- /VerifyImage – verifica uma imagem quando especificado o nome do arquivo e seu caminho;
- /EnableDebuLog – permite acesso ao arquivo FTKImageDebug.log criado na pasta em que o "FTK Imager" é executado.

As ferramentas apresentadas são umas das principais utilizadas em pericia forenses, que auxiliam na recuperação de dados, buscando vestígios para compor uma investigação bem-sucedida.

VII. Considerações Finais

A forense computacional está sendo propagada cada vez mais em entidades de Justiças devido ao aumento dos crimes cibernéticos. Então, para suprir esses determinados caos, é necessária uma perícia para ser desvendado o crime almejado.

O artigo apresentou as técnicas essências como a coleta, identificação, analise e resultados obtidos em uma perícia digital, bem como algumas ferramentas que poderão contribuir para o trabalho do perito computacional.

Entretanto, pode haver dificuldades durante o período da investigação, devido a falha humana ou até mesmo ausência de evidências. Portanto, é necessária uma equipe forense qualificada, buscando a verdade nos vestígios examinados para a obtenção correta do laudo pericial.

Referências Bibliográficas

1. BATTULA, B. et al. *Techniques in Computer Forensics:* IJS. V. 3, 2000.
2. BERTOGLIO, B. B. *Análise Forense:* Estudo teórico e prático. Novo Hamburgo, junho de 2008
3. CANSIAN, A. M. *Conceitos para perícia forense computacional.* VI ESCOLA REGIONAL DE INFORMÁTICA DA SBC, 2001, São Carlos, SP. Anais da VI Escola Regional de Informática da SBC. São Carlos, SP: ICMC/USP, 2001. p. 141 - 156.
4. ELEUTÉRIO, P. M. S.; MACHADO, M. P. *Desvenda ndo a computação forense*. São Paulo: Novatec, 2010.
5. GALVÃO, R. K. M. *Introdução à análise forense em redes de computadores:* São Paulo: Novatec, 2013.
6. GALVÃO, R. K. M. *Introdução à análise forense em redes de computadores:* São Paulo: Novatec, 2013.
7. KENT, K. et al. *Guidetointegratingforensictec hniquesintoincident response:* Special publication. Gaithersburg: NIST, 2006.
8. MOHAY, George; ANDERSON, Alison; COLLIE, Byron; DE VEL, Oliver; McKEMMISH, Rod. *Compu ter andIntrusionForensics. edArtechHouse*, 2003.
9. NG, R. *Forense Computacional Corporativa*. Ed. Brasport, Rio de Janeiro, 2007.
10. REINALDO,N.G.(2007). *Forense Computacional Corporativa,* ed. Brasport, 1ª Ed.
11. REIS, M. A. dosGeus. *Forense computacional e sua aplicação em segurança imunológica*. Campinas, SP: 2003
12. REIS, Marcelo Abdalla dos, GEUS, Paulo Lácio de. *Análise Forense de Intrusões em Sistemas Comp utacionais: Técnicas, Procedimentos e Ferr amentas.* Instituto de Computação –Universi dade Estadual de Campinas, 2004, p.54
13. SOUSA, S. R. A.; GOUVEIA, B. J. *Estudo para verif icação da eficiência da aplicação de métodos de*

análise forense computacional em ambientes Windows. Rio Verde, dezembro 2006.

Global Journal of Computer Science and Technology: E Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Blockchain and Smart Contracts as Complex Self-Organizing Frameworks: Theoretical Perspective

By Neyole Misko, Dr. Rono Kipronoh & Dr. Muchelule Yusuf

Abstract- Blockchain offers unprecedented opportunities for innovation in financial transactions with a whole new world of opportunities for banking, lending, insurance, and money transfers. Through its algorithms, digital security by decentralization, form smart contracts. Smart contracts allow the performance of credible transactions without third parties, the transactions premised by trackable and irreversible processes are superior to traditional contract law and greatly reduce other transaction costs associated with contracting. Globally, enterprises are undergoing a major transformation towards smart businesses that use intelligent systems integrated into planning for their daily routine. Blockchain technology and smart contacts termed disruptive technologies provide innovative solutions that cannot be ignored due to their inherent complexities. Regarded as complex systems, there is a need to have a theoretical view to understanding the hidden order to the evolution of these systems to bring out traits that are common and have a combination of independent actors behaving as a single unit responding and adapting to their existent setting, as self-organizing systems.

Keywords: blockchain, smart contracts, complex systems, self-organizing frameworks, theoretical perspective.

GJCST-E Classification: DDC Code: 332.178 LCC Code: HG1710

Strictly as per the compliance and regulations of:

Blockchain and Smart Contracts as Complex Self-Organizing Frameworks: Theoretical Perspective

Neyole Misko [α], Dr. Rono Kipronoh[σ] & Dr. Muchelule Yusuf [ρ]

Abstract- **Blockchain offers unprecedented opportunities for innovation in financial transactions with a whole new world of opportunities for banking, lending, insurance, and money transfers. Through its algorithms, digital security by decentralization, form smart contracts. Smart contracts allow the performance of credible transactions without third parties, the transactions premised by trackable and irreversible processes are superior to traditional contract law and greatly reduce other transaction costs associated with contracting. Globally, enterprises are undergoing a major transformation towards smart businesses that use intelligent systems integrated into planning for their daily routine. Blockchain technology and smart contacts termed disruptive technologies provide innovative solutions that cannot be ignored due to their inherent complexities. Regarded as complex systems, there is a need to have a theoretical view to understanding the hidden order to the evolution of these systems to bring out traits that are common and have a combination of independent actors behaving as a single unit responding and adapting to their existent setting, as self-organizing systems. This study significantly plays a unique role in contemporary science by explaining how blockchain and smart contracts unify run as nonlinearity complex system that adapts to their environment to bring about consistency hence their applicability.**

Keywords: blockchain, smart contracts, complex systems, self-organizing frameworks, theoretical perspective.

I. Introduction

Blockchain [1] is essentially a distributed database of records in the form of a public ledger that holds transactions and digital events that have been executed, verified by consensus, and shared among participating parties [2]. Defined as a foundation of Bitcoin [3], it ensures that every transaction is set public, consensus verified, and permanently stored in chains of blocks simply put Blockchain creating automated trust through the decentralized ledger that is distributed to all participants in the blockchain.

These underlined technology-entrenched undertakings have demonstrated through the power of the trust, consensus mechanisms, and deliberate veneration of the social contract that it is possible to use the internet to make a decentralized value-transfer system [4] and since it is a chain, the technology is easy to trace back an event shared across geographical boundaries. These processes are essentially unregimented, open, and rather not dominated by a tightly organized social or economic system.

The advent of blockchain has led to the development of important technological drivers for the world economy such as the concept of the smart contract. In a blockchain context, smart contracts mean transactions that go beyond simple buy and sell currency transactions, that have more extensive instructions embedded into them [5]. Technically, a contract is a set of IFs and THEN statements that give conditions and actions. Swan (2015), further indicates that much as the contracts feature the same kind of agreement to act or not act, they remove the need for one type of trust between parties as the smart contract is both defined and executed by the code automatically without discretion making a contract autonomy, self-sufficiency and decentralization.

Smart contracts in turn enable a decentralized application that accomplishes more than a transfer of value [6]. Its efficient automation of decentralized application can therefore be transitioned into solving real-world problems such as in managing real-time secure health records in medical facilities, in agriculture to provide experts and farmers greater interaction and sharing of data that affect land produce and in government ministries such as the lands ministry to update and manage lands record.

The smart contract has been established as a program deployed within a distributed network that can acquire outside information and update the internal state automatically. With the emergence of Ethereum as a built-in Turing-complete scripting language, the popularity and use of smart contract has been rising rapidly [7]. Currently, smart contracts are mainly embedded and used in blockchain distributed ledgers utilizing Ethereum as a platform [8].

II. Applications And use Of The Technology

Globally, many enterprises are undergoing a major evolution and transformation towards smart businesses where intelligent systems are being

Authore α σ ρ: e-mails: jmisiko@umma.ac.ke, ronohkipronoh@gmail.com, mucheluly@gmail.com, ymuchelule@umma.ac.ke

developed alongside a plethora of smart programs, services, and transactions integrated into planning for their daily routine. Startup firms and major enterprises alike are now developing smart contract solutions for an assortment of markets, alleging to offer a digital bypass around traditional mechanisms. The concept is to make smart contracts offer a superior solution to the challenges and limitations businesses face. The potential and the limitations of smart contracts have been and are being tested by offering novel possibilities to significantly alter traditional methods of doing thing spositively.

Karamitsos et.al (2018) presented a paper with an overview of Blockchain technology as a disruptive technology for the real estate industry. This study was designed to determine the effect of the smart contract with the various components for the implementation specifically to Real Estate development with the adoption of Blockchain to optimize the current processes. The paper observed among others the following recompenses of using smart contracts and blockchain technology in real estate: - different parties easy to modify databases in the real estate ecosystem, the advantage of disintermediation, and transactions advantage among other [9].

Smart contracts[10] have found applications in two main aspects: electronic monetary transactions and applications for storing information [11]. Bogner et al. [2016] developed a smart contract application aiming to rent devices. The food supply field has applied smart contracts to achieve food trade [12], [13] since smart contracts decrease the costs of traders on the food supply chain while providing a decentralized, accountable and transparent architecture based on blockchain technology.

Al-Bassam [2017] built a system for identity management, and organizational process for identifying, authenticating, and authorizing individuals' people to have access to applications, systems, or networks by associating user rights and restrictions with established identities. Shangping et al. (2019) proposed a product traceability system based on blockchain technology in which all product transferring histories would be perpetually recorded in the unchangeable ledger by using smart contracts, and the process of product registration, transferring and tracking is realized through the collaboration of smart contracts.

Through the supply chain process, customers would participate in the process as nodes to maintain information flows. The system described by decentralized characteristics would significantly reduce issues of privately tampering with data within enterprises and the system would have an event response mechanism to verify the identity of both parties of the transaction and by verifying the signature contained in the event to determine whether the event is valid. All events can be listened to and permanently stored in the blockchain in the form of a log. The results of the security analysis show that the system would be characterized by data accessibility, tamper-proofing, and resistance to man-in-the-middle attacks [14].

Hasan & Salah (2018) presented a blockchain-based solution and framework for the PoD of digital assets in which the framework was generic enough that it could be used to orchestrate and govern the sale and delivery of any digital asset and content including streaming video and audio. Based on the novel system, all transactions and interactions for the sale and download of digital items were to be controlled by Ethereum smart contracts. The demonstration showed that implementing smart contract code is safe and free of known exploitable security vulnerabilities and bugs. The solution also addressed some key security requirements such as security against popular attacks such as MITM and replay attacks [15].

Salah et al. (2019) proposed a framework leveraging Ethereum blockchain and smart contracts to trace, track, and perform business transactions by removing intermediaries for soybean traceability across the agricultural supply chain. The novel presentation presented details and aspects related to the system architecture, design, entity-relation diagram, interactions, sequence diagrams, and implementation algorithms. They showed that blockchain as a disruptive technology could provide an innovative solution for product traceability in agriculture and food supply chains. That could leverage the supply chain process to efficiently perform business transactions for soybean tracking and traceability across the agricultural supply chain.

III. Complex Self-Organizing Frameworks

Complexity theory premises that there is a hidden order to the evolution of complex systems, whether that system is a national economy, an ecosystem, an organization, or a production line. The Protagonists of the theory believe that certain traits are common to most complex systems where the systems are a combination of many independent actors behaving as a single unit and responding to their existent setting. To help understand the use of the theory in the modeling and analyzing complex systems, some frameworks contribute as core sets of commonality generic frameworks to understand the theory. They include: - Self-organizing, the adaptive system, the network theory, and nonlinear system theory.

a) Self-Organizing and Emergence Theory

Self-organization theory describes characteristic relationships in emergent systems complex systems as created by feedback mechanisms that either amplify an

effect to give positive feedback or dampen an effect to give negative feedback [16]. By definition, a complex system is a system composed of a large number of different interacting elements [17] it is therefore composed of many parts without centralized control. Emergence, described as a key concept in complexity science is the growth and evolution of more complex forms through simple rules. It, therefore, refers to how new levels of an organization are formed as parts are assembled.

Models of self-organization theory draw upon information theories to help understand the organization in terms of information and entropy. The theory also draws upon ideas of physics surrounding synchronization and pattern formation. Researchers try to model complex systems by capturing the local rules and using complex computational tools such as solidity for Ethereum framework or agent-based modeling to try and simulate the process through which order emerges out of initially homogenous or disordered states.

b) Complex Adaptive System

According to Sammut-Bonnici (2015), complex adaptive systems (CAS) consist of diverse components that are interdependent, act as a unified whole, and can learn from experience and adapt to change in the environment. Consequently, adaptive systems are classical examples of complex systems. They consist of many parts acting and reacting to each other's behavior. They are highly dynamic and developed through an evolutionary-like process the central issues being that of the process of adaptation and evolution.

Adaptation forms a central part of cybernetics that contributes to ideas surrounding control systems and how systems regulate themselves in their environment to maintain themselves (homeostasis). The key issue here is the dynamic surrounding cooperation and competition that form as adaptive agents interact and try to pursue their goals collectively. When ideas of adaptation are generalized to a whole population of agents and take place over a series of life cycles it can be termed evolution. This is supported by other theories such as evolution game theory, replicator equation, fitness landscapes, and genetic algorithm among others.

c) Network Theory

Network theory can be explained as the study of graphs as a representation of either symmetric relations or asymmetric relations between discrete objects, for computer science and network science, network theory is a part of graph theory characterized by nodes and edges that have attributes [18]. All complex systems can be modeled and analyzed as networks.

Network theory is a formal mathematical language but it has proven a very practical tool for analysis and found widespread application in many areas. Network theory has become an important study for complex systems, especially with the advent of information technology. Network theory is driven less by models and equations but more by real-time dense datasets, meaning we no longer stare at the models but are accessible to visualization to get us a much richer intuitive real source of exactly what the complex systems are like. Areas of contribution here are mathematics (graph theory) and computer science [19].

d) Nonlinear System and Chaos Theory

Non-linearity is an inherent feature and major theme that crosses all areas of complex systems. Chaos theory which is the study of non-linear dynamical systems was one of the major challenges to the Newtonian paradigm that was accepted into the mainstream of scientific knowledge [20] [21]. By definition, Chaos theory is a part of mathematics that looks at certain systems that are very sensitive in that any small change may make the system behave completely differently [22]. The modern scientific framework is based upon linear systems theory and this places significant constraints upon it.

Linear systems theory is dependent upon the concept of a system having an equilibrium [23]. All linear systems work as an approximation the fact is that many of the phenomena are nonlinear processes of change that happen far from equilibrium and are governed by dynamics of feedback loops and not linear equations. Trying to model complex systems by use of the traditional linear system is not effective hence areas of nonlinear systems and dynamics are a major part of complexity theory borrowed largely from physics, mathematics, and equilibrium processes from chemistry [24] [25].

IV. Conclusion

The Newtonian paradigm is a clockwork universe concept, a scientific model that supports science characterized by a materialistic and atomistic vision of sequestered inert objects interacting in a linear cause and effect fashion- the linear systems theory. This gives a concept of the universe that is analog to a system that is orderly, coherent, and predictable. Linear systems theory deals with relatively simple systems, systems that have a finite amount of independent, homogeneous elements interacting in a well-defined fashion with a relatively low level of connectivity. In its essence, linear systems theory fundamentally describes closed systems at or near equilibrium.

The context and significance of complexity theory are that it plays a unique role in contemporary science. Defined as an emerging post- Newtonian paradigm, it is a unifying point of view that views nonlinear complex systems running and adapting to their environment to bring about consistency. the current

society is surrounded by complex nonlinear systems that are well described by nonlinear theories such as the complexity theory. The theory, therefore, borrows support from the four models namely Network Theory, Complex Adaptive Theory, Nonlinear Systems Theory, and Self Organization Theory to try and explain complex systems such as the blockchain Ethereum platform for smart contracts and their applicability.

Acknowledgements

I acknowledge the support I was given by the UMMA University Computer science department for their academic critique and professionalism that guided my writing of the paper. Am fully indebted to them.

References Références Referencias

1. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," pp. 1-9, 2008.
2. M. Crosby, P. Pattanayak, S. Verma and V. Kalyanara, "Blockchain technology: Beyond bitcoin," Applied Innovation, pp. 6-19, 2016.
3. Z. Zibin, X. Shaoan, D. Hongning, C. Xiangping, and W. Huaimin, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," IEEE 6th International Congress on Big Data, pp. 557-566, 2017.
4. G. Wood, "Ethereum: A secure decentralized generalized transaction ledger.," Ethereum project yellow paper, pp. 1-32, 2014.
5. M. Swan, Blockchain: Blueprint for a new economy, USA: O'Reilly Media, Inc, 2015.
6. H. Yining, L. Madhusanka, A. Manzoor and K. Thilakarathna, "Blockchain-based Smart Contracts - Applications and Challenges," University of New South Wales, Aruna Seneviratne, 2019.
7. F. Tianyu, Y. Xiao, C. Yueting and L. Yi, "Smart contract model for complex reality transaction," International Journal of Crowd Science, pp. 184 - 197, 2019.
8. V. Buterin, "Ethereum: a next-generation smart contract and decentralized application platform," 2014. [Online]. Available: https://www.weusecoins.com/assets/pdf/library/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf.
9. I. Karamitsos, M. Papadaki and N. B. Al Barghuthi, "Design of the Blockchain Smart Contract: A Use Case for Real Estate," Journal of Information Security, pp. 177-190, 2018.
10. B. M and P. L, "An empirical analysis of smart contracts: platforms, applications, and design patterns," ' in Proc. Int. Conf. Financial Cryptographic Data Security, p. 494–509, 2017.
11. M. Dianhui, W. Fan, W. Yale, and H. Zhihao, "Visual and User-Defined Smart Contract Designing System Based on Automatic Coding," IEEE, pp. 73131 - 73144, 2019.
12. G. Peters, E. Panayi, and A. Chapelle, Trends in Crypto-Currencies and Blockchain Technologies: A Monetary Theory and Regulation Perspective vol.3, New York, NY, USA:: Social Science Electronic Publishing, 2015.
13. R. Adams, G. Parry, P. Godsiff and P. Ward, "The future of money and further applications of the blockchain," International journal for Strategic Change, pp. 417 - 422, 2017.
14. W. Shangping, L. Dongyi, Z. Yaling and C. Juanjuan, "Smart Contract-Based Product Traceability System in the Supply Chain Scenario," IEEE, pp. 115122 - 115134, 2019.
15. H. R. Hasan and K. Salah, "Proof of Delivery of Digital Assets Using Blockchain and Smart Contracts," IEEE, pp. 65439 - 65449, 2018.
16. ESPA, "Emergence and Self-Organization," 5 October 2019. [Online]. Available: http://www.complexity.soton.ac.uk/theory/_Emergence_and_Self-Organization.php.
17. G. Weisbuch, Complex systems dynamics, Boca Raton London New York: CRC Press Taylor &Francis Group, 2018.
18. E. Estrada and P. A. Knight, A First Course in Network Theory, London: OXFORD University Press, 2015.
19. J. Baez, Network Theory, USA: University of Califonia, 2015.
20. W. J. Rugh, Nonlinear System Theory The Volterra/Wiener Approach, USA: The Johns Hopkins, 2002.
21. J. P. Noël and G. Kerschen, "Nonlinear system identification in structural dynamics: 10 more years of progress," Mechanical Systems and Signal Processing, pp. 2 - 35, 2016.
22. A. T. Azar and S. Vaidyanathan, Advances in Chaos Theory and Intelligent Control, Switzerland: Springer International Publishers, 2016.
23. J. P. Hespanha, Linear System Theory 2nd. Edition, New Jersey: Princeton University Press, 2017.
24. H. Ozbay, Introduction to feedback control theory, Boca Raton London New York Washington, D.C: CRC Press, 2000.
25. J. M. Epstein, Nonlinear Dynamics,Ma them atical, Biology, and Social Science, Boca Raion London Mew York: CRC Press Taylor & Francis Group, 2018.

Global Journal of Computer Science and Technology: E
Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Comparison of Effective Bandwidth Estimation Methods for Data Networks

By José Bavio, Carina Fernández & Beatriz Marrón
Universidad Nacional del Sur

Abstract- The purpose of this work is to apply techniques to estimate the Effective Bandwidth, from traffic traces, for the Generalized Markov Fluid Model in data networks. This model is assumed because it is versatile in describing traffic fluctuations. The concept of Effective Bandwidth proposed by Kelly is used to measure the channel occupancy of each source. Since the estimation techniques we will use require prior knowledge of the number of clustering clusters, the Silhouette algorithm is used as a first step to determine the number of classes of the modulating chain involved in the model. Using that optimal number of clusters, the Kernel Estimation and Gaussian Mixture Models techniques are used to estimate the model parameters. After that, the performance of the proposed methods is analyzed using simulated traffic traces generated by Markov Chain Monte Carlo algorithms.

Index Terms: *effective bandwidth; markov fluid model; parameter estimation; data networking.*

GJCST-E Classification: *DDC Code: 388.31 LCC Code: HE336.T7*

Strictly as per the compliance and regulations of:

Comparison of Effective Bandwidth Estimation Methods for Data Networks

José Bavio [α], Carina Fernández [σ] & Beatriz Marrón [ρ]

Abstract- The purpose of this work is to apply techniques to estimate the Effective Bandwidth, from traffic traces, for the Generalized Markov Fluid Model in data networks. This model is assumed because it is versatile in describing traffic fluctuations. The concept of Effective Bandwidth proposed by Kelly is used to measure the channel occupancy of each source. Since the estimation techniques we will use require prior knowledge of the number of clustering clusters, the Silhouette algorithm is used as a first step to determine the number of classes of the modulating chain involved in the model. Using that optimal number of clusters, the Kernel Estimation and Gaussian Mixture Models techniques are used to estimate the model parameters. After that, the performance of the proposed methods is analyzed using simulated traffic traces generated by Markov Chain Monte Carlo algorithms.

Index Terms: effective bandwidth; markov fluid model; parameter estimation; data networking.

I. Introduction

WHEN working with several aggregated services in a telecommunications network, we must resort to a digital network of integrated services. Integration means that the network can carry many types of information, such as voice, video, and data, all in digital form, using a single infrastructure. The general problem in a multiplexing system is that resources are shared by a set of heterogeneous sources. Multiplexing poses a mathematical and statistical problem: estimating the resource requirements of a font or set of fonts and, as sources are variable, statistical gain is to be expected. The different requirements of each service during the connection can be explored by statistical multiplexing. Since the notion of Effective Bandwidth (EB) introduced by Kelly in 1996 [4], the development of statistical tools that allow finding expressions to estimate the probability of loss in a link has emerged strongly.

The problem with multiplexing is that the probability of many sources deciding to dispatch the maximum rate, in which case there would be an overflow, is not zero. Admission control mechanisms must be in place to accept a new connection, which minimizes the effects of data loss while maintaining the quality of service (QoS) for both current and future sources. Therefore, it will be essential to have mathematical models that describe the behavior of the sources. Traffic modelingin network services is necessary in order to dimension their components and evaluate their performance. Traffic models can be used both to find the appropriate descriptors that characterize the service, and facilitate management tasks, such as the establishing of control admission criteria (CAC). In particular, we are interested in applying estimation techniques from the traffic traces of a data network to monitor and predict network performance so that more opportune and effective control decisions can be made. This paper is structured as follows: Section II introduces the Generalized Markov Fluid Model and provides an expression to determine the EB for this model. This tool is used to measure the channel occupancy of each source. Section III studies the kernel estimation techniques and the Gaussian mixture model, to use these tools to estimate the EB for our model. Section IV presents the parameters of the simulated model and the estimation methods of the EB of the GMFM from traces. A comparison of the estimation methods is presented in section V, and conclusions are drawn in Section VI, along with some considerations for future work.

II. Traffic Description

a) The Model

Modulated Markovian models have been developed for more than two decades. They have been especially useful for modeling, with relative accuracy, many real data sources because they can capture the temporal correlation of the data. These processes are characterized by a set of states, which form a Markov chain, and the transition times. In this work, we use the Generalized Markov Fluid Model, introduced in [1], due to its properties.

This model is modulated by a continuous-time, homogeneous and irreducible Markov chain. In each state of the chain, the generation rate is a random variable, distributed according to a probability law, depending on the state, that does not change during the time interval in which the Markov chain is in that state.

The interpretation of the model is that each state in the chain is identified as the activity performed by a user, such as chatting or videoconferencing. Hence, a sudden change in transfer rate reports a change of state in the chain. Within a state, the data transfer rate assumes values that depend specifically on that activity, according to some probability distribution.

Author α σ ρ: Departamento de Matemática. Universidad Nacional del Sur, Bahía Blanca, Argentina. e-mails: jmbavio@yahoo.com, carina.fernandez@uns.edu.ar, beatriz.marron@uns.edu.ar

b) The Effective Bandwidth

Multiplexing variable rate sources on a link result in reserving for each source a capacity more significant than the average transmission rate, which would be a very optimistic measure, but less than the maximum transmission rate, which would be a pessimistic measure and lead to a waste of resources. For this purpose, the EB defined by F. Kelly in [4] is a valuable and realistic measure of channel occupancy.

To estimate EB for a given GMFM from traffic traces, formulas of the type obtained by Kesidis, Walrand, and Chang [5] were obtained.

Let us consider $\{X_t\}, t \geq 0$ a GMFM. Then the effective bandwidth has the following expression

$$\alpha(s,t) = \frac{1}{st} \log \{\pi \exp [(Q + sH)\, t]\, \mathbf{1}\}, \qquad (1)$$

where 1 is a column vector with all entries equal to 1, π and Q are the invariant distribution and the infinitesimal generator for the modulating chain, respectively. H is a diagonal matrix whose dimension is equal to the number of states of the chain and whose non-zero elements are the first moments, μ_i, of the law governing the generation rate in state i.

This expression provides a way to estimate the EB from traffic traces. Estimators of the infinitesimal generator of the modulator chain, its invariant distribution, and the average transfer rate, and their properties can be found in [1].

III. Estimation Methods

In this section, the estimators $\hat{Q}$, $\hat{H}$, and $\hat{\pi}$ of the parameters involved in (1) are calculated using different techniques.

a) Kernel Density Estimation

One of the most common methods for nonparametric estimation of a density is the well-known kernel estimator. Given a simple random sample X_1, ...,X_n of the random variable of interest X with density f, the expression of the estimator is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad (2)$$

where h > 0 is a smoothing parameter, and K is the kernel function which is non-negative, bounded, real-valued, unimodal, symmetric around 0, and its integration is 1.

It is possible to choose different types of functions for the K-kernel. We choose to work with Gaussian kernels. The smoothing parameter h must tends slowly to zero; this means that h $\rightarrow$0, nh $\rightarrow \infty$, to ensure that $\hat{f}$ tends to the true density f. The window size parameter h is an essential aspect of these techniques, as seen in [3]. Density estimation will be smoother the larger the window. The peaks are associated with the states of the chain, if they are close together, a "large" h could lead to merging close peaks and drawing incorrect conclusions, but a "small" h could show too many peaks leading to spurious maxima.

The invariant distribution of the modulate chain was calculated using a suitable window size in (2). With the number of peaks, we estimate the number of states of the chain, because it is a multimodal density. The values where the maxima are reached represent the average sending rate in each state. We estimate the range of dispatch values associated with each state, with the values of the minima adjacent to each maximum. Next, with the area under the density, within each range, we estimate the probability π_i of each state i.

Thus, π can be reconstructed because both spatial and temporal behavior converge at the GMFM, as seen in [3].

b) Gaussian Mixture Model

In the statistical context, we can define a mixture model as a probabilistic model to represent the incidence of subpopulations within the same population. For this reason, they are helpful for estimating the invariant probability of the modulate chain and the average dispatch rates in each state.

As the Gaussian distribution is fitted to model the distribution of the dispatch rate in each case, we will use Gaussian Mixture Models (GMM). These models are an example of a parametric probability density function which can be represented as a weighted sum of all Gaussian component densities.

Let us assume that there are k clusters. Equation that defines a Gaussian Mixture is

$$\pi(x) = \sum_{i=1}^{k} \pi_i N(x|\mu_i, \Sigma_i).$$

Each Gaussian explains the data contained in every single cluster, and the mixing coefficients are themselves probabilities and must satisfy the following condition

$$\sum_{i=1}^{k} \pi_i = 1.$$

To determine the optimal values for these, we need to establish the maximum likelihood of the model. A general technique for finding maximum likelihood estimators in these models is the Expectation-Maximization, or simply the EM algorithm. This is widely used for optimization problems where the objective function has complexities, such as the one we have just found for the GMM case.

To apply this iterative process, we must initialize the parameters of our model $\theta = \{\pi, \mu, \Sigma\}$, with some value. In our case, we use the results obtained by a previous K-Means algorithm as a good starting point for our algorithm. EM algorithm consists of two steps; the first one finds an expression for the expected value of

the log-likelihood, given the initial valuesor the prior estimation of the parameters, and the second step maximizes that expectation over the parameter space. More details about these algorithms are discussed in [2] and [6].

IV. Simulation and Numerical Results

In this section, we will perform the analysis using simulated traffic traces generated by simulations to estimate with both methods. Simulations were performed in Python 3.7 using sklearn.neighborslibrary, [7] and codes can be provided by asking the authors. Several traffic simulations were performed according to the presented model presented in (II-A), where the modulating Markov chain has k = 9 states, and each state is associated with a data transfer rate interval shown in the table below.

To design the infinitesimal chain generator, we consider that the chain can pass from one state to another, with the same probability. Even it is not a very realistic model, we believe that it is the most adequate for comparing different estimation methods, so

State	Transfer speed (Mbps)
1	(0,1024]
2	(1024,2048]
3	(2048,3072]
4	(3072,4096]
5	(4096,5120]
6	(5120,6144]
7	(6144,7168]
8	(7168,8292]
9	(8292,10240]

$$Q=\begin{pmatrix} -8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -8 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -8 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -8 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -8 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -8 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -8 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -8 \end{pmatrix}.$$

Within each of these intervals, the amount actually dispatched comes from a Gaussian probability distribution.

The simulated trace is a succession of pairs $(v_i$ $t_i)$ with i from 0 to 20000, where v_i is the transfer speed, t_i is themoment when the chain jumps to another state, and 20000 is the number of jumps in the chain, so the link transfers at the speed v_i while $t_{i-1} < t < t_i$.

a) Determining the optimal number of clusters: Silhouette method

Many clustering techniques are based on a first estimation as a starting point, which requires prior knowledge of the number of clusters.

The Silhouette coefficient, a popular method of measuring the clustering quality, which combines both cohesion and separation [8], is rather independent of the number of clusters, k. For object x_i, the silhouette coefficient is expressed as follows:

$$s_{x_i} = \frac{b_{x_i} - a_{x_i}}{max(a_{x_i}, b_{x_i})},$$

where a_{xi} is the average distance of object x_i to all other objects in its cluster; for object x_i and any cluster not containing it, calculate the average distance of the object to all the objects in the given cluster, and b_{x_i} is the minimum of such values for all clusters. It is possible to obtain an overall measure of the goodness of clustering by calculating the average silhouette coefficient of all objects. For one clustering with k categories, the average silhouette coefficient of the cluster is taking the average of the silhouette coefficients of objects belonging to the clusters; that is:

$$\bar{s}_k = \sum_{i=1}^{n} s_{x_i},$$

where n is the total number of objects in the data set. Value of the silhouette coefficient can vary between –1 and 1. Higher value indicates better clustering quality.

The sum of squared errors (SSE) and the silhouette coefficient are combined to measure the quality of clustering and determine the optimal clustering number, k_{opt}.

Data objects in the same cluster are similar, and objects from distinct clusters are different from each other. This distribution minimizes the SSE of each data object from its cluster center. SSE is a commonly used criterion in measuring the quality of clustering; lower SSE indicates better partition quality for partitions with the same k. This criterion is defined as follows:

$$SSE = \sum_{i=1}^{k} \sum_{x_j \in C_i} \|x_j - c_i\|^2,$$

where x_j is the j-th object in cluster C_i, and c_i is the center of cluster C_i.To determine the optimal clustering number, we introduce the silhouette coefficient to work in conjunction with the SSE criterion because the SSE criterion is sensitive to the number of clusters, k.

In this paper, we conduct a k-means clustering analysis of the dispatch rates under different k values. We plot the curves of the SSE and average silhouette coefficient against the number of clusters to analyze the two curves and identify the optimal number of clusters, $k_{opt.}$ Below we show comparative results on the k-data clustering configuration from 7 to 11.

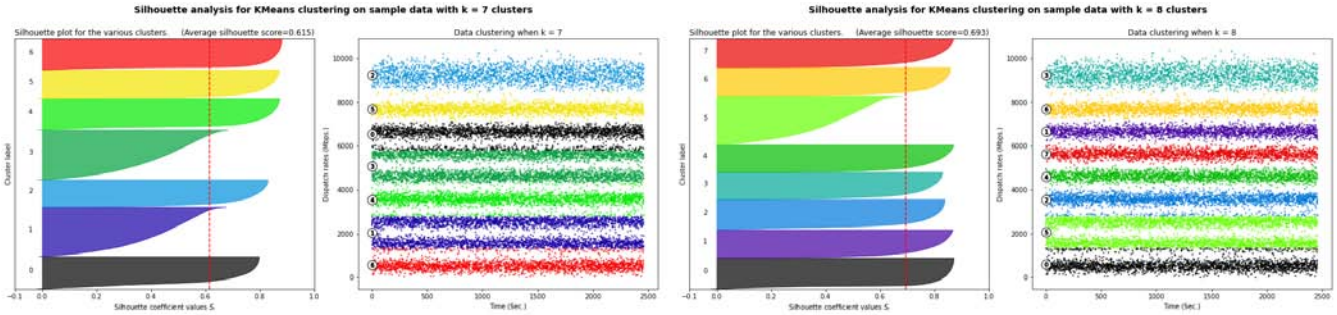

Fig. 1: Silhouette analysis for k = 7 and k = 8.

Figure 1 and Figure 3 show that choices of k equal to 7, 8, 10 or 11 would not be appropriate due to the presence of groups with below-average silhouette scores and wide fluctuations in the size of the silhouette plots. The value of k = 9 looks to be optimal one, as shown in Figure 2. The silhouette score for each cluster is above average silhouette scores. In addition, the thickness of the silhouette, plot representing each cluster is similar.

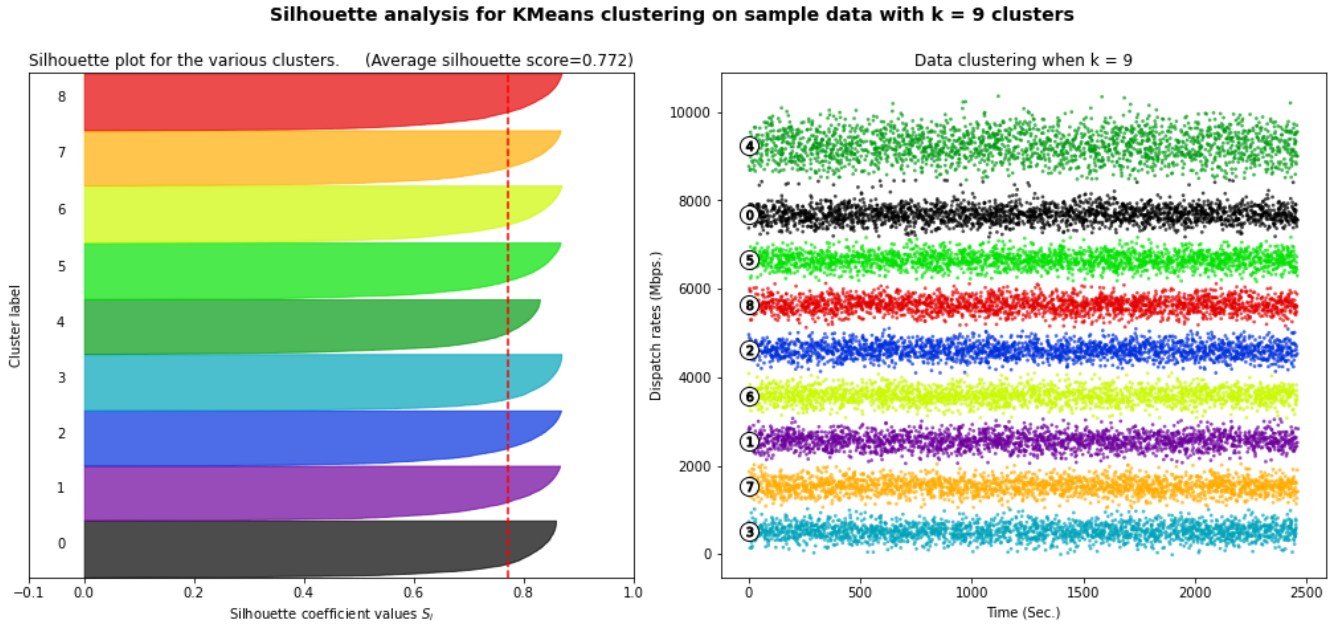

Fig. 2: Silhouette analysis for k = 9.

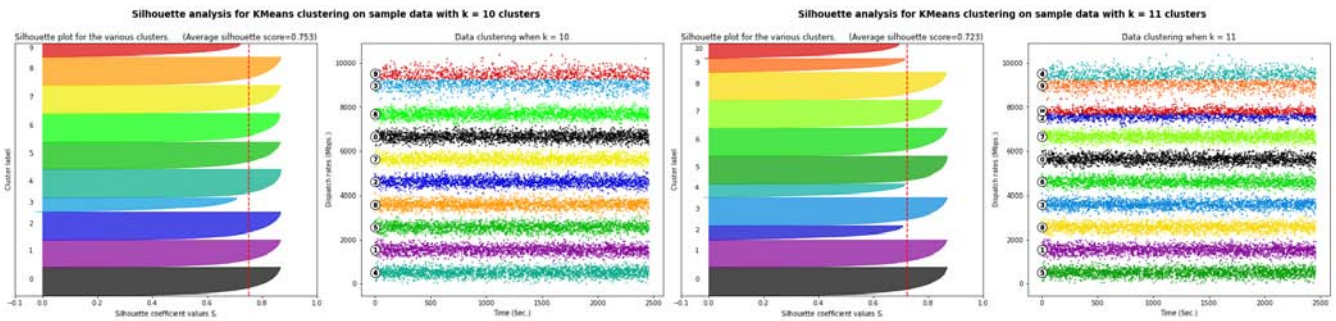

Fig. 3: Silhouette analysis for k = 10 and k = 11.

Finally, we plotted the curves of the SSE and average silhouette coefficient against the number of clusters to analyze the two curves to determine the optimal number of clusters, k_{opt}.

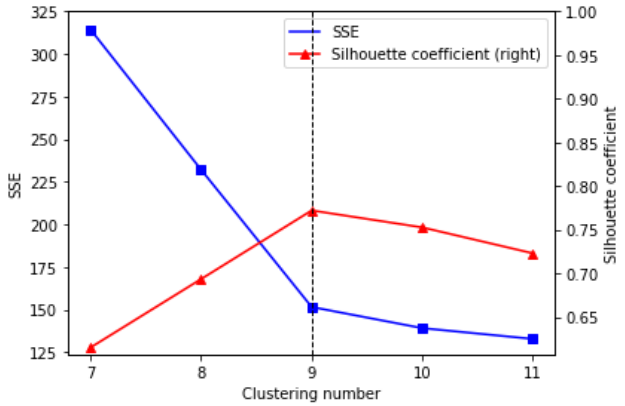

Fig. 4: SSE and average silhouette coefficient versus the number of clusters.

Figure 4 draws a curve of scores, one score, Silhouette coefficient, for each alternative number of clusters. Elbow marks the point where the line exhibits its maximum curvature. Let us note that, before reaching this point, an increase in the number of clusters helps to reduce the sum of squared errors SSE. It is to be expected that behind the elbow, we find diminishing returns: incremental reductions of the SSE, by adding more clusters, would become more negligible the farther we go beyond the elbow and would do so relatively faster after having passed the inflection point of the curve: its elbow. This value is reached when the data are grouped into 9 clusters, so we conclude from this analysis that the optimal number of clusters is $k_{opt} = 9$.

b) Estimations from Traces

We categorize the dispatch rates according to the clustering result with $k_{opt} = 9$ as the clustering number. We consider that within each interval, the dispatched comes from a Gaussian distribution centered to its midpoint and deviation equal to one sixth of the length of the interval.

i. *Kernel Estimation Method*

For the simulated trace, we estimated the EB through the following steps:

1. Apply a Gaussian kernel to all v_i, $0 \leq v_i \leq 20000$, with $h = 200$, to obtain $\hat{\pi}(x)$ for $0 < x < 10240$. This is possible because GMFM is ergodic, and time and spaces averages converge. See Figure 5.
2. Find minima for $\hat{\pi}(x)$. These minima are an estimate for the extremes of the dispatch ranges, which in turn allow us to determine the state of the modulating chain. As Gaussian distribution is symmetric, we determine rate averages using the estimated rank middle points. Finally, area under $\hat{\pi}(x)$ between two consecutive minima estimates $\hat{\pi}_i$.
3. Go through the trace comparing each v_i with the rank estimated to assign the corresponding state, to obtain theestimated chain $(\hat{c}_i, t_i)$, $0 \leq i \leq 20000$.
4. Estimate infinitesimal generator from $(\hat{c}_i, t_i)$ where t_i are cumulative so first order difference of t_i gives permanence time in state c_i.
5. Calculate the estimated EB with $\hat{H}$, $\hat{\pi}$, and $\hat{Q}$, as in (1).

We chose the window width of 200 as it provides a reasonable estimate at a low computational cost.

Figure 5 shows the theoretical and estimated density.

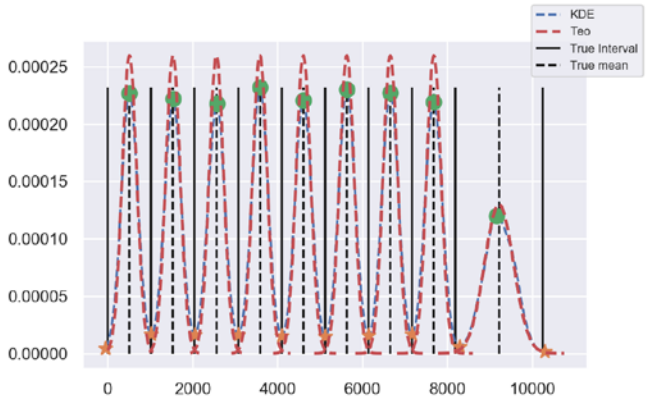

Fig. 5: Theoretical and estimated density, using Kernel Estimation techniques.

The estimation of the infinitesimal generator is as follows.

$$\hat{Q} = \begin{pmatrix} -8.2507 & 0.9618 & 1.0355 & 1.0650 & 1.0355 & 1.0723 & 0.8991 & 1.0355 & 1.1350 \\ 0.9603 & -7.9626 & 0.8585 & 1.0331 & 0.9385 & 1.0403 & 1.1204 & 1.0585 & 0.9458 \\ 1.0164 & 0.8857 & -7.8808 & 1.0527 & 1.0999 & 0.9184 & 0.9983 & 0.8785 & 1.0309 \\ 1.0265 & 1.1078 & 1.0302 & -8.3599 & 0.9601 & 1.0782 & 1.0339 & 1.0154 & 1.1078 \\ 1.0004 & 0.9279 & 1.0185 & 1.0149 & -7.9377 & 1.0620 & 1.0366 & 0.9786 & 0.8989 \\ 1.0654 & 1.0474 & 1.0403 & 1.0259 & 0.9506 & -8.0746 & 0.9649 & 1.0726 & 0.8860 \\ 0.9801 & 1.0532 & 0.9654 & 1.0569 & 1.0642 & 1.0459 & -8.2244 & 1.0678 & 0.9874 \\ 1.0078 & 1.0448 & 1.0078 & 1.0411 & 1.0115 & 1.0226 & 1.1518 & -8.3248 & 1.0115 \\ 1.1135 & 0.9761 & 0.9947 & 0.9761 & 0.9576 & 0.9910 & 1.0170 & 1.1060 & -8.1320 \end{pmatrix}.$$

.

The heat map in Figure 6 shows the percentage estimation error of the infinitesimal generator, obtained as the difference between each element with its estimate.

Negative values in the heat map indicate an underestimation, and positive values an overestimating the infinitesimal generator elements.

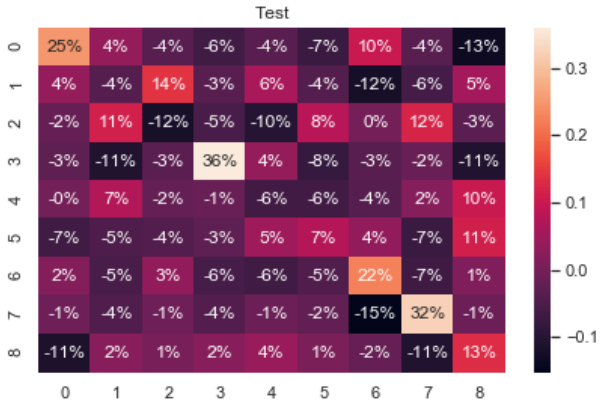

Fig. 6: Heat map for infinitesimal generator estimation errors.

We use the confusion matrix M to evaluate the performance in state estimation.

$$M = \begin{pmatrix} 2235 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 2182 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 2167 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2257 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 2184 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2240 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2243 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2236 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 2200 \end{pmatrix}.$$

.

Rows are the actual states, and columns are the predicted or estimated states. For example, the 4 in matrix M at row 3, column 2 indicates that four times the chain was in state 3 but was estimated to be in state 2.

ii. *Gaussian Mixture Model*

Through the following steps, we estimate the EB for the simulated trace.

1. Apply GMM to all $v_i, 0 \leq v_i \leq 20000$ and obtain the means μ_j, which will be the centers of each of the 9 clusters, the variances ${\sigma_j}^2$, and the weights of the mixture, $\pi_j, 1 \leq j \leq 9$.
2. Reconstruct the dispatch rate intervals, using 99% of the area of each Gaussian distribution, centered at μ_j and with ${\sigma_j}^2$ variance, $1 \leq j \leq 9$. The value 0 is token as the lower limit of the first interval.
3. Go through the trace comparing each v_i with the rank estimated to assign the corresponding state, to obtain the estimated chain $(\hat{c}_i, t_i)$, $0 \leq i \leq 20000$.
4. Estimate infinitesimal generator from $(\hat{c}_i, t_i)$ where t_i are cumulative so first order difference of t_i gives permanence time in state c_i.
5. Calculate the estimated EB with $\hat{H}$, $\hat{\pi}$, and $\hat{Q}$, as in (1).

Figure 7 shows the theoretical and estimated density.

In this case, the estimation of the infinitesimal generator is as follows.

$$\hat{Q} = \begin{pmatrix} -8.2397 & 0.9618 & 1.0392 & 1.0613 & 1.0355 & 1.0723 & 0.8991 & 1.0318 & 1.1387 \\ 0.9603 & -7.9553 & 0.8585 & 1.0294 & 0.9421 & 1.0403 & 1.1131 & 1.0622 & 0.9494 \\ 1.0155 & 0.8849 & -7.8699 & 1.0481 & 1.0989 & 0.9176 & 0.9973 & 0.8668 & 1.0409 \\ 1.0275 & 1.1088 & 1.0238 & -8.3568 & 0.9610 & 1.0792 & 1.0312 & 1.0164 & 1.1088 \\ 1.0004 & 0.9279 & 1.0185 & 1.0149 & -7.9412 & 1.0620 & 1.0402 & 0.9750 & 0.9025 \\ 1.0654 & 1.0474 & 1.0403 & 1.0259 & 0.9506 & -8.0531 & 0.9649 & 1.0690 & 0.8896 \\ 0.9788 & 1.0557 & 0.9678 & 1.0594 & 1.0667 & 1.0448 & -8.2297 & 1.0631 & 0.9934 \\ 1.0109 & 1.0479 & 1.0035 & 1.0331 & 1.0109 & 1.0257 & 1.1442 & -8.3057 & 1.0294 \\ 1.1111 & 0.9708 & 0.9966 & 0.9819 & 0.9560 & 0.9893 & 1.0188 & 1.1074 & -8.1318 \end{pmatrix}.$$

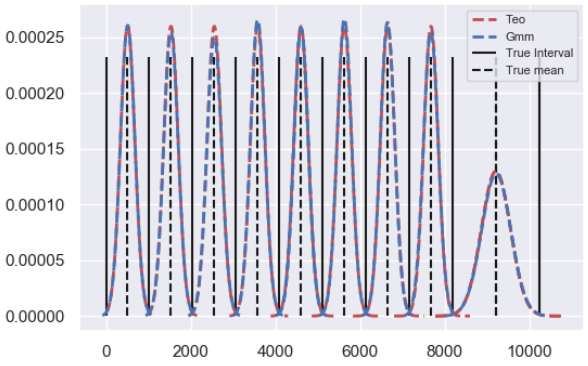

Fig. 7: Theoretical and estimated density, using Kernel Estimation techniques

To evaluate the performance of the estimator, we show in Figure 8 the heat map for the estimation error of the infinitesimal generator.

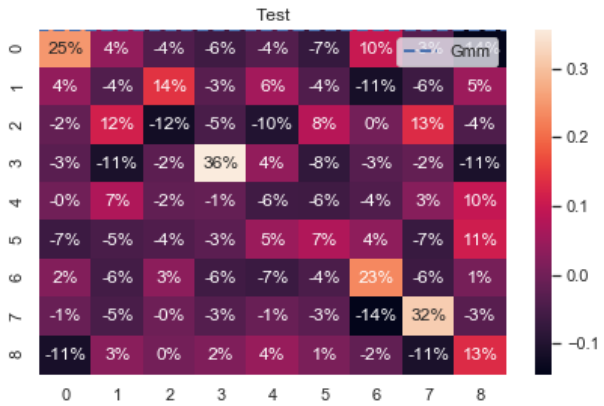

Fig. 8: Heat map for infinitesimal generator estimation errors.

The confusion matrix M that we show below allows us to evaluate the performance in the estimation of the states.

$$M = \begin{pmatrix} 2235 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 2182 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 2167 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 2254 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 2185 & 8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2240 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 2241 & 8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2236 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2203 \end{pmatrix}.$$

V. Comparison of the Results Obtained

To compare both methods, in this section, we present the theoretical parameters and their respective estimates.

Table I shows the estimated values of the average dispatch rates in each method and the value with which the traces were simulated, and the same is done for the class ranges of dispatch in Table II.

Both methods allow for to reconstruct of the modulate chain very well, and this can be seen in their confusion matrices. However, in the EB estimation, the Gaussian Mixture Model is more efficient, as can be seen in Figure 9 and zoom of it in Figure 10.

Table I: Theoretical and estimated ranges of dispatch.
Theoretical range Estimated range

Theoretical range	Estimated range	
	KDE	GMM
0	0	0
1024	1026.1086	1024.1057
2048	2047.0149	2044.5478
3072	3069.9408	3078.4938
4096	4099.5986	4098.4828
5120	5121.1781	5124.6486
6144	6143.7674	6140.8196
7168	7163.6639	7156.2757
8192	8360.9483	8189.6392
10240	10240.0000	10240.0000

Table II: Theoretical and estimated average dispatch rates.

Theoretical range	Estimated range	
	KDE	GMM
512	513.8042	513.2390
1536	1535.7202	1535.4954
2560	2561.3389	2559.7032
3584	3585.6112	3585.3959
4608	4609.2103	4611.1025
5632	5633.8191	5634.6542
6656	6654.0522	6653.1858
7680	7676.3049	7676.4876
9216	9209.5156	9223.2794

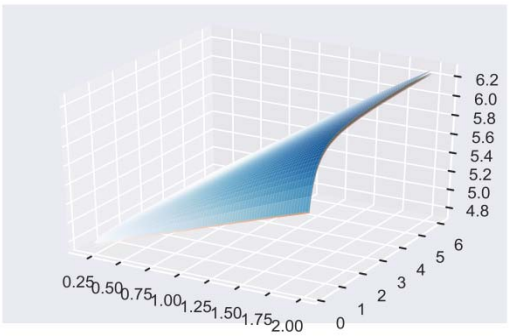

Fig. 9: Theoretical bandwidth (blue), bandwidth estimation using KDE (red) and mixed Gaussian (green).

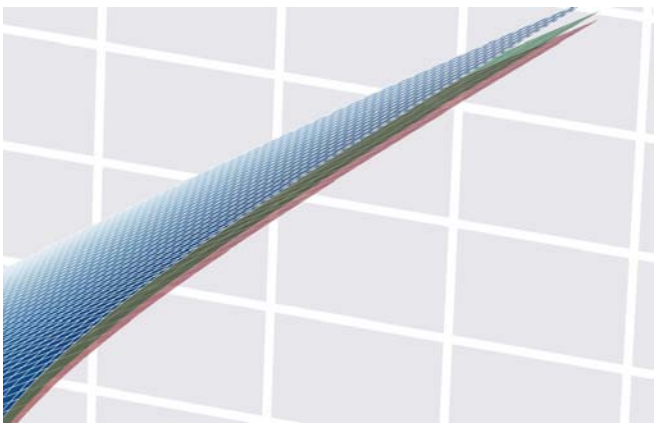

Fig. 10: Figure 9 zoom.

VI. Conclusion

In this paper we have proposed two methods to estimate effective bandwidths from traffic traces of a GMFM source.

Numerical examples of simulated traces were presented showing the results obtained. Estimation process worked much better in the Gaussian Mixture model, as seen in the graphics presented.

It is expected to extend the statistical calculation using an infinitesimal generator that models a more realistic behavior of the source and also in which the supports of each probability law have a greater intersection to develop the estimation to real data scenarios.

References Références Referencias

1. Bavio, J.M and Marron, B.S. (2018). Properties of the Estimators for the Effective Bandwidth in a Generalized Markov Fluid Model. Open Journal of Statistics, Vol.8 No.1, pp. 69-84.https://doi.org/10.4236/ojs.2018.81006.
2. Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer-Verlag, Berlin.
3. Coudret, R., Durrieu, G., Saracco, J. (2013). Comparison of kernel density estimators with assumption on number of modes. Communications in Statistics - Simulation and Computation, Taylor Francis, 2015, 44(1), 196-216. https://doi.org/10.1080/036 10918.2013.770530.
4. Kelly, F. (1996). Notes on effective bandwidths. Stochastics Netwoks: Theory and applications. Oxford University Press. Oxford. pp. 141-168. http://www.statslab.cam.ac.uk/~frank/eb.ps
5. Kesidis, G., Walrand, J. and Chang, C.S. (1993). Effective bandwidth for multiclass Markov fluids and other ATM sources. IEEE ACM Trans. Networking 1,4, pp. 424-428.
6. Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
7. Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. JMLR, 12, 2825-2830.
8. Tan, P.N.; Steinbach, M.; Kumar, V. Introduction to Data Mining. Pearson Addison Wesley: Boston, MA, USA, 2006; pp. 487–556.

Global Journal of Computer Science and Technology: E
Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Best Fit Method of Sample Selection in Data Hiding and Extraction

By Virendra P. Nikam & Dr. Shital S. Dhande
SIPNA College of Engineering and Technology

Abstract- Today data security and its transmission over the wireless network need special attention. Intruder always has a watch on sensitive data transmitted over a wireless network. This work proposes an approach that minimizes the quantization error between the original and result carrier by selecting optimize samples during Data Hiding. Propose work find out best matching carrier components during the data hiding process. Results also imply that achieved results are far better than any other steganographic method.

Index Terms: best fit strategy, steganography, quantization error, carrier object, transmission media, data hiding, data extraction etc.

GJCST-E Classification: D.2.11

Strictly as per the compliance and regulations of:

Best Fit Method of Sample Selection in Data Hiding and Extraction

Virendra P. Nikam[α] & Dr. Shital S. Dhande[σ]

Abstract- **Today data security and its transmission over the wireless network need special attention. Intruder always has a watch on sensitive data transmitted over a wireless network. This work proposes an approach that minimizes the quantization error between the original and result carrier by selecting optimize samples during Data Hiding. Propose work find out best matching carrier components during the data hiding process. Results also imply that achieved results are far better than any other steganographic method.**

Index Terms: best fit strategy, steganography, quantization error, carrier object, transmission media, data hiding, data extraction etc.

I. Introduction

Information security is a primary focus for every IT industry. Most of the industries are grown up by analyzing the data that they have. Data is any sort of raw material which can be processed to generate valuable information. Millions of dollars are spent on data security in almost all industries in India and all over the world. In the recent past 10 years, information security is a vital domain that needs special attention in every sector. Raw data is a base pillar of any IT industry. Recent history shows that most of the industries have failed to recover themselves because of not having a proper backup facility. In 2005 India, a major flood in Mumbai itself stops the functioning of more than 2000 small scale IT industries. This implies that the security of information or raw data is very much important. Without security, it's not possible for any IT industry to grow fast and within the expected time.

Now the major issue that comes into focus is how to provide security to sensitive data of an industry? There are many techniques available that are best to provide security to data which is stored either on a separate server system or on a local server system. Many IT Industries prefer to store their data on the server system. Server systems have their own security features and protocols which are enough to protect data. But from the money point of view, it's not convenient to maintain a separate server to stored data especially for those industries which have an annual turnover between 1 to 5 lakh.

The common techniques which are used to provide security to data are cryptography, steganography and watermarking. These three techniques have their own applicability and limitations.

a) Cryptography

Cryptography is an art to convert readable data into an unreadable format. It totally hides the meaning of the original data. Process of converting readable data into unreadable format is called as encryption whereas converting unreadable data into a readable format is called decryption. To perform encryption and decryption, sender and receiver use either the same or different key. Based on the similarity of the key used at sender and receiver side, cryptography is classified as 1. Private key cryptography 2. Public key cryptography. Public key Cryptography uses to separate keys for encryption and decryption. The algorithms like RC2, data Encryption standard, triple Data Encryption standard, advanced encryption standard comes under the category of private key cryptography whereas algorithm likes RSA comes under public-key cryptography.

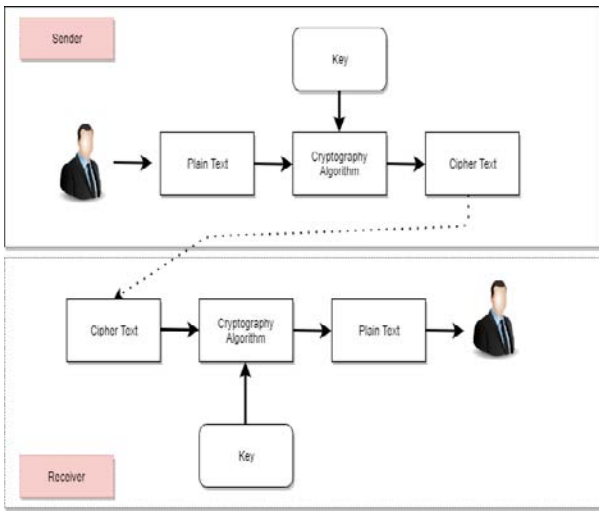

Fig. 1: Overview of Cryptography

Cryptography convert readable data set *fp1; p2; p3; p4; p5:::::; pn g* into unreadable(encrypted) data set *fe1; e2; e3; e4; e5:::::; en g* Cryptography can be defined

Author α: Virendra Nikam is a Ph.D. student who is working on secure data concealing and transmission for wireless networks using steganographic techniques. Data security, information management, and other topics are among his main research interests.
e-mail: virendranikamphd@gmail.com
Author σ: Dr. Shital S. Dhande is a professor at SIPNA College of Engineering and Technology in Amravati, where she teaches computer science and engineering. She's worked in data and information management, security applications, and data query processing for over 20 years. She mentored over ten Ph.D. students and was a member of several important o rganizations such as IEEE, CSI, and others.

as $f(x) = f(p; k)$ for encryption and $f(x) = f(f0(x); k)$ for decryption.

b) Steganography

It's the practice of concealing or hiding hidden information behind a carrier item. With the use of carrier medium, it entirely conceals the existence of data. There is a considerable similarity among Cryptography and steganography that, both hide the real presence of data to an unauthorized user. Result of steganography is exactly similar in appearance with the original carrier object. This is the one mandatory feature of steganography which not allows any change occur in carrier object. Steganography take one carrier object *fc* & secret data *fd* and hide it behind carrier object $f(H) = f(f(c); f(d))$ The similarity of result stego object can be

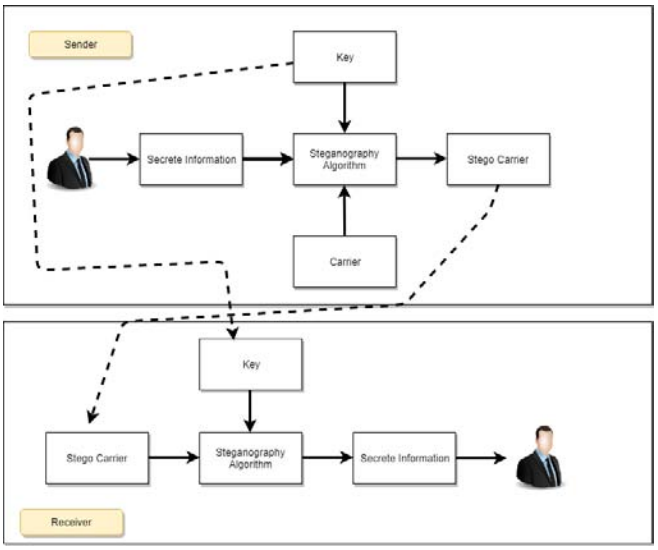

Fig. 2: Overview of Steganography

measured with parameters like peak signal to noise ratio (PSNR), mean square error (MSE), absolute difference, structural content, mean difference, normalized absolute error (NAE) and others. If the result object finds dissimilar visual perceptual quality with original carrier object $f(c)$ $6=$ $f(H)$, the process is not called steganography.

c) Watermarking

It is generally used for copyright protection. Few authors also called watermarking as steganography. Watermarking is of two types 1. Visible watermarking and 2. Invisible watermarking. In watermarking, original carrier replaced with watermark data bits. It's completely changed carrier object especially in visible watermarking. The concept of watermarking was originally designed for copyright protection later on it is expanded for secret data hiding and transmission. This paper focuses on data Hiding and extraction mechanism by optimizing sample selections approach for minimizing quantization error. Quantization error is a difference between result stego object and original carrier object *Qe = jCo -Crj*. Many existing steganography techniques does not focus on minimizing quantization error. Existing technique directly select sample irrespective of its value that results in large quantization error. Maximum quantization error can create difference among original and result carrier object. This paper has a primary focus on effective sample selections with minimum difference.

Basic questions come into mind "How to select optimize sample during Data Hiding process?". This paper proposes an algorithm that finds the best matching carrier sample using best fit strategic approach. The best fit strategic approach is one that generally used in memory allocation. While allocating memory, a primary focus is given on memory fragmentation. Memory gets fragment when memory block of either larger or smaller size get allocated to the required content. Best fit strategic approach reduce memory fragmentation and hence it's a good choice by memory allocator. Best bit strategic approach is chosen by many programmers due to its effective selection of required memory block from available blocks of memory. Consider an item set *fi1; i2; i3; i4; i5:::::; ing* and the required item to search is *fsig* At very first stage, difference among *fsig* and item set please find out *fji1 - sij; ji2 - sij; ji3 - sij; :::jin - sij* An item from item Set with minimum difference *Best match = min jii -sij* Can be chosen as best fit or maximum matched for further processing.

II. Background History

Steganography is the method of concealing hidden information behind a carrier. In its true meaning, steganography is a centuries-old notion that was first executed 300 years ago. People utilized this approach to manually transfer a message from one location to another using a fly in ancient times. The message is coded on the fly's neck after the hair on its back neck is removed. They wait for their neck hair to develop before sending this message to their target. Nowadays, "digital steganography," in which digital data is hidden behind a digital carrier object, is a whole distinct type of steganography. A photo, music, or video might be used as the carrier item. In recent years, many steganography techniques have been invented that make data transfer more secure by preventing an intruder from inferring data.

Least Significant Bit Substitution (LSB) is the original steganography technique, and it includes hiding a secret information bit at the first (from right to left) position of the carrier sample. Let's look at a carrier sample in binary format (0001010*textbf*1) with the secret bit set to 0. After concealing secret bit 0, the output carrier sample is (0001010*textbf*0). This approach is straightforward to use and retains the carrier's audio-

visual perceptual quality. However, because of its simplicity, an attacker may be able to readily find hidden bit locations, thus allowing unlawful data extraction.

By concealing hidden bits at higher and higher LSB places, the problems of the least significant bit replacement approach were eliminated. Moving from the least significant bit (LSB) to the most significant bit (MSB) increases quantization error *Qe*. An intruder can easily detect the existence of secret information bits behind the carrier owing to a discrepancy between the result and the original carrier object caused by a quantization mistake. A quantization error is a number that ranges from 0 to 255.

Spread spectrum analysis, wavelet analysis, and sample selection procedures were offered as ways to reduce quantization error. All of these methods choose the optimal sample carrier and then hide hidden data behind it. These techniques bring quantization error down to an acceptable level. However, each strategy has its own set of benefits and drawbacks. These methods take a long time to implement and are difficult to master. In terms of data concealing capacity, quantization error, time processing, and so on, none of the solutions are perfect.

III. Literature Survey

M. Nosrati, R. Karimi, H. Nosrati, and A. Nosrati [1] published a paper in 2011 that described a new approach for detecting hidden information in a 24 bit RGB color picture. After the data has been hidden, carrier samples are connected together to remember where the secret information bits are kept. They're disguising data with a randomized sample selection strategy, which might make the data extraction procedure more complicated and challenging. It connects data samples with each other via a directed data connection list. It also sets aside portion of the carrier sample in order to check for the presence of secret information bits.

Wen-Chung Kuo, Dong-Jin Jiang, and Yu-Chih Huang [2] introduced a data concealing and extraction approach based on block division in 2008. This approach separates the data into numerous blocks, then generates a histogram for each of them. The histogram's lowest and maximum points are found, allowing hidden data to be stored in each block and enhancing its data concealing capability.

Steel hypothetical, audio/video steganography, IP datagram steganography, and other data concealment techniques were introduced by Soumyendu Das, Subhendu Das, Bijoy Bandy opadhyay, and Sugata Sanyal [3] in 2008. The focus is on creating a carrier histogram to determine the amount of space available for data concealing and encryption to increase the security of sensitive data transmitted across an unsecured wireless network.

In 2002, Ming Sun Fu and O.C. Au [4] introduced a halftone picture data-hiding approach. When original multitone photos are unavailable due to force pair toggling, this approach offers a high data concealing capacity. The visual perception of the outcome image's quality is so similar to that of the original carrier that it's impossible to tell them apart.

H. B. Kekre, Archana Athawale, Archana Athawale, Archana Athawale, and Uttara Athawale [6] suggested a Data Concealing technique for hiding data in audio by producing stego audio carriers in 2010. This recommends that instead than hiding data directly in the LSB of an audio sample, the parity of the sample should be calculated first, and then the choice to hide secret sample information be made. An intruder will find it incredibly difficult to estimate where secret information bits are concealed in a carrier using this method.

Data Hiding utilizing video as a carrier was proposed by Xiaoyin Qi, Xiaoni Li, Mianshu Chen, and Hexin Chen [7] in 2011. This method use an add-up strategy to choose the best sample from a carrier that conceals hidden information. In this approach, a 4X4 DCT block is scanned in a zigzag pattern to obtain the best matching sample.

By breaking a picture into numerous blocks, T. Hong, W. Chen, and H. Wu [8] describe a Data Hiding technique in 2012. Finding the smoothness of a carrier sample is used to extract data. The data extraction technique does not include the four edges of each block. When the size of a block is smaller, the extraction process slows down. On the receiver side, the extraction process is completed by measuring the perceptible smoothness to at least one.

In 2013, C. Qin, C.-C. Chang, Y.-H. Huang, and L.-T. Liao introduced a prediction-based reversible data concealment strategy that chooses image carrier samples based on their distribution features. To create a predicted picture with similar structural content to the original carrier image, an image in-painting approach is applied. To incorporate hidden information bits, the histogram of the difference is altered. Secret information bits can be retrieved precisely and without alteration on the receiver side. The prediction technique is used to estimate the covered picture pixels and quantization error before moving on to Data Hiding.

In 2014, X. Zhang, Z. Qian, G. Feng, and Y. Ren [10] suggested using lossless compression to hide data in encrypted photos. The encrypted image's route is compressed with LDPC coding before being implanted with extra secret info. Encrypted data has a high level of quality. Hidden data is successfully retrieved on the receiver side. LDPC coding is used to hide data at the 4*th* LSB of an image.

In 2015, Jiantao Zhou, Weiwei Sun, Li Dong, Xianming Liu, Oscar C. Au, and Yuan Yan Tang suggested a datahiding technique in the encrypted domain [11]. There is no need to disclose a private

encryption key because images are encrypted and decrypted with a public key. This method encrypts sensitive information using public-key cryptography once it is hidden in a carrier. This method eliminates the need for additional transmission expenses to send secret encryption techniques. The picture is decrypted with a separate key on the receiver's end, and then the concealed bits are retrieved.

IV. Proposed Methodology

Proposed methodology concerned with an actual implementation of suggested work. It includes a data flow diagram, algorithms, system model, etc. Searching and locating an optimize sample is not a simple task. It needs lots of comparison and sorting of intermediate results. The overall process is divided into a total four states

a) Carrier Selection.
b) Carrier Classification.
c) Searching Best Fit Sample from Class.
d) Updating Result Carrier with Newly Find Best-Fit Sample.

a) Carrier Selection

A carrier is an object that has the capability to carry sensitive data. In the proposed concept, carrier may be an image, audio or video. Carrier is an integration of samples *cr* = R*in*=1 *fcig* Which can be consist of 8 bits only. If a carrier is an image, it consists of RGB colour combination and represented each colour (RGB) with 8 bits. If a carrier is an audio, it consists of left and right stereo channels. Each carrier sample *ci* have a minimum value 0 and maximum value 255. For an image carrier sample *ci* is an integration of RGB values R (*R; G; B*) . While processing image as a carrier, colour channels are considered and manipulated separately.

b) Carrier Classification

Carriers are classified based on hiding position of the sensitive information bit. If sensitive information is decided to hide at 3*rd* position of carrier binary, then samples are classified accordingly. An effectiveness of proposed algorithm entirely depends on hiding the position of the secret information bit. Security is a major issue that is achieved with high complexity. Security level gets increase from the Least Significant Bit (LSB) to Most Significant Bit (MSB).The number of item in each class gets affected by the position of secret information bit. It is always better to classify carrier samples as per the MSB position. Security of data can be increased with higher MSB position. However, all samples of carrier are not sure in table 1 due to its large size. Samples are equally divided that is 128 in class 0 and class 1 each.

Table I: Carrier Samples Classification

Class 0	Binary	Class 1	Binary
0	00000000	4	00000100
1	00000001	5	00000101
2	00000010	6	00000110
3	00000011	7	00000111
8	00001000	12	00001100
9	00001001	13	00001101
10	00001010	14	00001110
11	00001011	15	00001111
16	00010000	20	00010100
17	00010001	21	00010101
18	00010010	22	00010110
19	00010011	23	00010111

c) Searching Best Fit Sample from Class

The best fit sample searches according to the secret information bit. If secret information bit is 0, best fit sample search into Class 0 else it is searched in class 1. The best fit sample is a sample with a minimum difference with the original sample. Suppose secret bit is 0 and sample taken for hiding it is equal to (7)10 = (00000111)2. At position 3, bit 1 is present. After replacing 3*rd* bit with 0, the final value of the sample becomes (00000011)2 = (3)10. In this case, Quantization error *Qe* = *j*3 - 7*j* = 4 . Propose approach help to focus on minimizing this condition error. Find out the best-fitted sample in class 0 which is sample (8)10 = (00001000)2. After finding best sample, Quantization error becomes *Qe* = *j*8 - 1*j* = 1.

d) Updating Result Carrier with Newly Find Best-Fit Sample

While doing replacement of original carrier sample with resultant sample, a proposed algorithm takes care for not generating quantization error more than ±16. Effectiveness of any steganography algorithm depends on the difference between result and original carrier object. The audiovisual perceptual quality of carrier gets affected with large quantization value which may violate a definition of steganography. Result sample after setting it to carrier will change its originality which is measured by parameters like peak signal to noise ratio, mean square error, absolute difference, minimum difference, structural content, cross correlation etc.

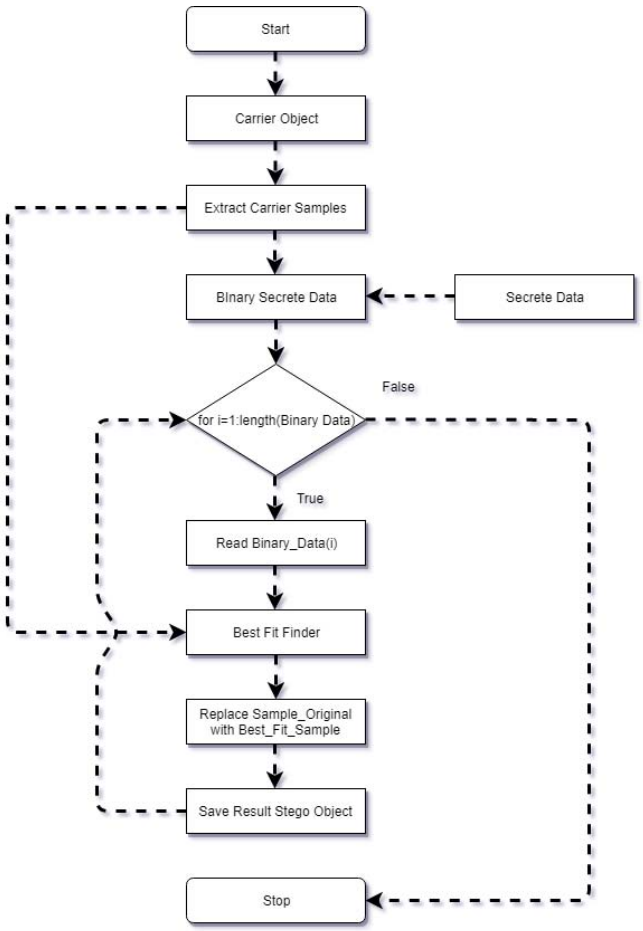

Fig. 3: Data Flow Diagram

As shown in the data flow diagram, user first select carrier object which may be anything like a picture, audio or video. Samples are extracted from this carrier which is given to best fit finder. Sensitive information given by the user is converted into its equivalent binary format. For each bit of binary sensitive information, best fit finder locates best carrier samples with minimum quantization error *Qe*. Loop continues its execution until all binary bits of secret pieces of information get completely hidden behind carrier samples. Below algorithm is used for Data Hiding process which is effectively implemented and executed by proposed system. Effectiveness of Data Hiding algorithm is based on at what position secret information bit get hidden?. The algorithm does not focus on the position of secret bits in the carrier object. It is the generalized algorithm which can be fit for any carrier with any position.

Propose concept deal with information hiding in a carrier by minimizing its quantization error so that the difference between resultant and original carrier should be minimum as possible as.

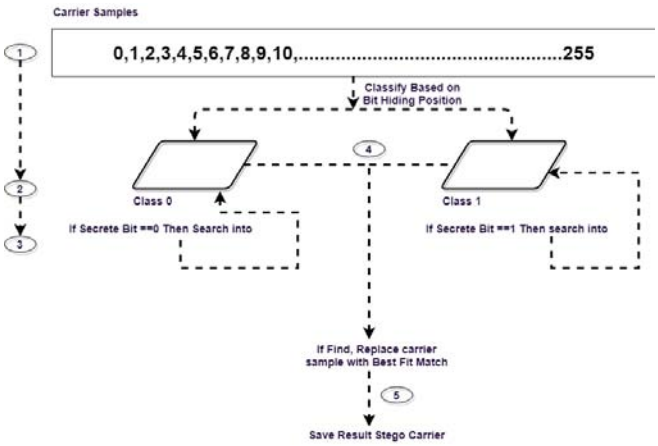

Fig. 4: Detailed Process Data Flow Diagram

Figure 4 demonstrates the proposed Data Hiding approach with detailed process data flow diagram. Carrier samples consist of 8 bits which have minimum value 0 and maximum value 255. These samples classified into class 0 and class 1 based on at what position secret information bit get it done? Suppose sender decides to hide secret information bit at third position, class 0 contains all the values which have 0 value at their third position of binary and class 1 contain all sample having 1 at 3rd position. If secret information bit equal to 0, optimized samples will be searched in Class 0 else search into class 1.

Algorithm

1 Start
2 Input Carrier c_r
3 Extract Carrier Sample c_s
4 Input secrete information s_i
5 Convert s_i to Binary B_{s_i}
6 For i 1 to count(B_{s_i})
Find Best Fit Sample c'_s from c_s
c_s c'_s
c_r c_s
End
7 Save c_r
8 Stop

Let's illustrate the algorithm with a specified example. Consider a set of carrier samples 10; 24; 30;

50:::::*n* and secret binary information 00101100. Read the first bit of secret binary information from Most Significant Bit to Least Significant Bit. The first bit is 0 and supposes carrier sample is (23)10 = (00010111)2 . An algorithm decided to hide secret information bit at 3*rd* position from LSB side. An algorithm finds the best matching carrier sample from a set of carrier samples i.e (24)10 = (00011000)2 With selected best fit sample, contribution error *Qe* = *j*23 - 24*j* = 1.

V. Result Analysis

Result analysis is used to compare the proposed concept output with existing results. An outcome of result analysis must be the final decision that it shows its applicability, acceptability and it's implications.

Fig. 5: 1, 2

From figure 5, 1 represent to original image whereas 2 represents result carrier image. From above images, one can conclude that propose best fit sample selection strategy for Data Hiding preserve the visual perceptual quality of carrier. The similarity between these two images can be measured with the other parameters shown in below Table II and Table III. Propose concept have been tested on more than 50 carrier images of uncompressed types. We intentionally proceed with uncompressed carrier to successfully extract hidden data.

Propose concept compared with existing methods like least significant bit substitution (LSB), higher LSB and many more. From Table II, quantization error occurs is very from 0 to 8. It implies that, propose a steganographic technique generates very less noise/ quantization error which best suited for an image as a carrier. For audio, quantization error *Qe* is tolerable up to ±16 . Hence it is also fit for audio-based steganography. Parameters mentioned in table II are used to compare the original and resultant carrier object. From table II, it is concluded that with the proposed best fit strategy, it is possible to maintain too much similarity between original and result carrier. Structural content, cross-correlation and normalized absolute error(NAE) remain untouched which indicates its effectiveness and acceptability of the proposed algorithm.

Table II: Parameter Comparison

Average Red	Average Green	Average Blue	MSE	PSNR	Max Diff	Min Diff	Average Diff	NAE	Cross Correlation	Structural Content
0.39	0.27	0.37	2.72	43.79	71	0	0.34	0	1	1
0.36	0.25	0.32	1.21	12.52	25	0	0.21	0	1	1
0.34	0.21	0.29	1.89	10.58	10	0	0.25	0	1	1
0.41	0.22	0.40	0.52	13.56	14	0	0.26	0	1	1

means square error (MSE) = *n*1 P*n i*=1 (*ci - c0 i*)2, peak signal to noise ratio = 10 ∗ log10 *MAX MSE*2 , Min difference = *Min jco - crj* , Max difference = *Max jco – crj*, nor *n i*=1 P*m j*=1*i*=1correlation (CC) = (*f* ∗ *g*) (*τ*) Δ *1 f* ¯(*t*)*g* (*T* + *τ*) *dt* and structural content P*n i*=1 P*m j*=1(*coij*)2 P*n i*=1 P*m j*=1(*crij*)2 is a measure of the difference between original and result carrier. Table II imply that difference occurs at a minimum level.

Table III: Parameter Comparison Result Analysis

Image Type	Red Mean	Green Mean	Blue Mean	Mean	Pure Height	Pure Width	Entropy
Original Image	105	108	81	221598	50	45	449.33
Result Image	104	108	81	221395	50	45	449.33

Entropy is a measure of information represented by the carrier object. Samples which are closer to 0 and 255 does not contributes to entropy. From table III, one can conclude that with proposing a best-fit strategic approach for sample selections of carrier will reduce the quantization error at minimum level. Changes occur in the image as a carrier shown with the parameters like average red $Avg_{red} = (\sum_{ri=1}^{n} P_{ri})/n$, average green $Avg_{green} = (\sum_{gi=1}^{n} P_{gi})/n$, , average blue $Avg_{blue} = \sum_{bi=1}^{n} P_{bi})/n$ and $Entropy = \sum kP_k \log_2(P_k)$ etc.

Table IV: Difference Between Original and Result Carrier Sample

Red Original	Red Result	*Qe*	Secrete Bit	Green Original	Green Result	*Qe*	Secrete Bit	Blue Original	Blue Result	*Qe*	Secrete Bit
34	34	0	0	25	23	2	0	16	15	1	1
51	51	0	0	42	39	3	0	35	35	0	0
48	48	0	0	40	40	0	1	37	40	3	1
166	166	0	0	127	128	1	0	132	132	0	0
181	181	0	0	138	138	0	1	147	143	4	1
155	151	4	0	113	111	2	1	127	127	0	1
173	176	3	0	140	140	0	1	157	160	3	0
119	119	0	0	115	115	0	0	138	135	3	0
187	183	4	0	167	167	0	0	192	192	0	0
200	199	1	0	172	172	0	1	197	200	3	1
108	112	4	0	93	93	0	1	112	112	0	0
26	23	3	0	20	24	4	1	24	23	1	0
20	20	0	0	14	14	0	1	16	15	1	1
255	247	8	0	210	207	3	1	220	220	0	1
227	227	0	0	176	175	1	1	191	191	0	1
212	212	0	0	174	174	0	1	197	200	3	1
68	68	0	0	59	59	0	1	90	87	3	0
161	161	0	0	140	144	4	0	173	173	0	1
221	224	3	0	185	185	0	1	221	224	3	0
194	194	0	0	158	158	0	1	194	194	0	0
146	146	0	0	123	119	4	0	151	151	0	0
113	113	0	0	103	103	0	0	114	114	0	0
101	101	0	0	94	96	2	0	102	104	2	1
87	87	0	0	80	80	0	0	87	87	0	0
74	71	3	0	57	57	0	1	37	37	0	0
55	55	0	0	36	36	0	0	22	22	0	0
71	71	0	0	47	48	1	0	45	45	0	1

VI. Conclusion

Table II, III, IV shows that quantization error occurs with the proposed approach is at its minimum level. This concept also preserves the audiovisual perceptual quality of carrier due to which it will be a future choice by many security application developers for transmitting sensitive data over insecure wireless network. The best fit strategic approach is applicable to all types of carrier media including an image, audio, and video. It does not affect size, length at other parameters of carrier.

VII. Future Scope

Future scope concentrates on limitation of proposed work. Even the proposed system reduces quantization error at the minimum level, searching for the best fit sample is time-consuming and needs to have a precise and accurate approach. The number of iterations required to find out the best fit sample is directly proportional to number of samples present in carrier. If propose work requires to avoid searching iterations for the same samples which reduce its overall implementation time.

References Références Referencias

1. M. Nosrati, R. Karimi, H. Nosrati, and A. Nosrati, Embedding stegotext in cover images using linked list concepts and LSB technique, Journal of American Science, Vol. 7, No. 6, 2011, pp. 97-100.
2. Wen-Chung Kuo, Dong-Jin Jiang, Yu-Chih Huang, A Reversible Data Hiding Scheme Based on Block Division, Congress on Image and Signal Processing, Vol. 1, 27-30 May 2008, pp. 365-369
3. Soumyendu Das, Subhendu Das, Bijoy Bandyopadhyay, Sugata Sanyal, Steganography and Steganalysis: Different Approaches, International Journal of Computers, Information Technology and Engineering (IJCITAE), Vol. 2, No 1, June, 2008, Serial Publications, pp. 1-11.
4. Ming Sun Fu and O.C. Au, Data hiding watermarking for halftone images", IEEE

Transactions on Image Processing, Vol.11, No. 4, Apr. 2002, pp.477-484.

5. Sandipan Dey, Ajith Abraham, Sugata Sanyal, An LSB Data Hiding Technique Using Prime Numbers, IEEE Third International Symposium on Information Assurance and Security, Manchester, United Kingdom, IEEE Computer Society press, USA, 29 31 Aug. 2007, pp.101-106.
6. H. B. Kekre, Archana Athawale, Archana Athawale, Uttara Athawale, Information Hiding in Audio Signals, International Journal of Computer Applications IJCA, Vol. 7, No. 9, Foundation of Computer Science, New York, USA, pp. 14-19 2010.
7. Xiaoyin Qi, Xiaoni Li, Mianshu Chen, Hexin Chen, Research on CAVLC audiovideo synchronization coding approach based on H.264, IEEE International Conference on Uncertainty Reasoning and Knowledge Engineering (URKE), Vol. 2, 4- 7 Aug. 2011, pp.123-126.
8. T. Hong, W. Chen and H. Wu, An improved reversible data hiding in encrypted images using side match, IEEE Signal Processing Lett., vol.19, no. 4, pp. 199-202, 2012.
9. C. Qin, C. -C. Chang, Y.-H. Huang, and L.-T. Liao, An in painting Assisted reversible steganographic scheme using a histogram shifting mechanism, IEEE Trans. Circuits Syst . Video Technol., vol. 23, no. 7, pp. 1109-1118, 2013.
10. X. Zhang, Z. Qian, G. Feng, and Y. Ren,Efficient reversible data hiding in encrypted images, J. Vis. Commun. Image R., vol. 25, no. 2, pp. 322- 328, 2014.
11. Jiantao Zhou, Weiwei Sun, Li Dong, Xianming Liu, Oscar C. Au Yuan Yan Tang,Secure Reversible Image Data Hiding over Encrypted Domain via Key Modulation, IEEE Trans. On Circuits and Systems for Video Technology, vol. 26, issue 3, pp.441 452,2015.
12. Po-Chun Chang, Kuo-Liang Chung, Jiann-Jone Chen, Chien-Hsiung Lin ,etc. A DCT/DST-based error propagation-free data hiding algorithm for HEVC intracoded frames [J]. Journal of Visual Communication & Image Representation. 2013, 25(2):239253
13. A Piva, R Caldelli, F Filippini. Data hiding for error concealment in H.264/AVC[C]. IEEE Workshop on Multimedia Signal Processing, 2004:199-202
14. Lin T, Lie W, Tsai D, Lin G. Error Resilient Coding Based on Reversible Data Embedding Technique for H. 264/AVC Video[C]. IEEE International Confer ence on Multimedia and Expo, (2005) :1174-1177
15. SD Lin, HC Meng, YL Su. A novel error resilience using reversible data embedding in H.264/AVC[C]. International Conference on Information, 2008, 7(5):1-5
16. X. J. Ma, Z. T. Li, J. L and W. D. Wang. Data Hiding in H.264/AVC Streams with Limited Intra-Frame Distortion Drift [C]. Computer network and Multi media Technology, CNMT 2009.
17. X. J. Ma, Z. T. Li, H. Tu, B. Zhang. A data hiding algorithm for H. 264/AVC video streams without Intra frame Distortion Drift [J]. IEEE Trans. Circuits Syst. Video Technol. 2010, 20(10):13201330.
18. Y.X. Liu, M.S. Hu, X.J. Ma, H.G. Zhao. A new robust data hiding method for H.264/AVC without intra-frame distortion drift [J]. Neu rocomputing. 201 5,1076-1085

Global Journal of Computer Science and Technology: E Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Improved Energy-Aware Distributed Unequal Clustering Protocol using BBO Algorithm for Heterogeneous Load Balancing

By Maryam Maleki & Amir Massoud Bidgoli (CA)
Islamic Azad University

Abstract- With the rapid extension of IoT-based applications, various distinct challenges are emerging in this area. Among these concerns, the node's energy efficiency has a special importance; since it can directly affect the functionality of IoT-Based applications. By considering data transmission as the most energy-consuming task in IoT networks; clustering has been proposed to reduce the communication distance and ultimately overcome node energy wastage. However, cluster head selection as a non-deterministic polynomial-time hard problem will be challenging notably by considering node's heterogeneity and real-world IoT network constraints which usually have conflicts with each other. Due to the existence of conflict among the main system parameters, various solutions have been proposed in recent years that each of which only considered a few real-world limitations and parameters.

Index Terms: internet of things, optimization method, energy consumption optimization, clustering algorithms, heterogeneity, load-balancing, swarm intelligence, distributed artificial intelligence, node heterogeneity, packet size heterogeneity, buffer size heterogeneity, energy heterogeneity, distributed clustering, wireless sensor networks, distributed computing, biogeography-based optimizationi.

GJCST-E Classification: C.2.2

Strictly as per the compliance and regulations of:

An Improved Energy-Aware Distributed Unequal Clustering Protocol using BBO Algorithm for Heterogeneous Load Balancing

Maryam Maleki [α] & Amir Massoud Bidgoli (CA) [σ]

Abstract- With the rapid extension of IoT-based applications, various distinct challenges are emerging in this area. Among these concerns, the node's energy efficiency has a special importance; since it can directly affect the functionality of IoT-Based applications. By considering data transmission as the most energy-consuming task in IoT networks; clustering has been proposed to reduce the communication distance and ultimately overcome node energy wastage. However, cluster head selection as a non-deterministic polynomial-time hard problem will be challenging notably by considering node's heterogeneity and real-world IoT network constraints which usually have conflicts with each other. Due to the existence of conflict among the main system parameters, various solutions have been proposed in recent years that each of which only considered a few real-world limitations and parameters. Here we present an effective improvement on the existing energy-aware distributed unequal clustering protocol (EADUC). Our recommended improvement attempts to decrease the number of dead nodes and energy wastage by utilizing a load-balancing technique which is beneficial for heterogeneous node networks. This solution employs a well-known swarm intelligence algorithm (SI), named Biogeography-based optimization (BBO) in a distributed manner. Our proposed work considered a variety of real-world limitations such as energy, time, communication radius, and buffer size which have not been reflected in many previous works simultaneously. Our simulations show approximately a 6% improvement in the total number of alive nodes and a 1.59 % drop in energy consumption in comparison with the existing EADUC algorithm.

Index Terms: internet of things, optimization method, energy consumption optimization, clustering algorithms, heterogeneity, load-balancing, swarm intelligence, distributed artificial intelligence, node heterogeneity, packet size heterogeneity, buffer size heterogeneity, energy heterogeneity, distributed clustering, wireless sensor networks, distributed computing, biogeography-based optimizationi.

I. Introduction

IoT has enabled us to develop a new generation of intelligent systems which had not been possible previously. The ability to connect every single entity and object in the physical world to the internet, through a mixture of communication protocols and techniques has applications in several domains. From industrial usage to smart city management systems [15] [16], IoT has infiltrated all aspects of our lives. Since IoT-based infiltrated all aspects of our lives. Since IoT-based applications demonstrated their capacity to enhance the quality of our lives, their pace of evolution has increased. Meanwhile, several concerns and challenges have been raised; here a primary problem is how to provide the required energy for these flourishing applications?

Consider a smart city management system that comprises an enormous number of heterogeneous nodes that can sense and measure a mixture of vital system parameters. IoT nodes are mostly equipped with inexpensive and limited batteries which recharging and replacing them would not be practical or even possible. These Nodes are randomly placed in a large-scale context of a city and can have a distance of more than 2000 meters from the base station. Collecting information from these nodes will be costly in terms of energy, especially by increasing the required communication distance [17]. Additionally, a variety of other environmental constraints and limitation exists which must be considered.

Clustering has been identified as a suitable solution for solving the challenge of long-distance data transmission [17] [18]. Unfortunately, this solution is complicated due to several parameters that should be considered. Most of these parameters are in contrast with each other. Hence, their simultaneous optimization would be difficult or sometimes impossible. In recent years, many valuable works have been done in this region that each of them considers a combination of parameters in different ways. In the following, we present five categories to provide a better understanding of reviewed solutions.

a) Centralized or Distributed

Centralized clustering algorithms are solutions in which the base station collects general network information, performs clustering, and finally informs the nodes. The 2 base station needs to collect general network information consists of different node's properties such as id, position, energy level, buffer size, data packet length, and so on; This information is applied for clustering the network. Although the

Author α σ: Islamic Azad University-Tehran North Branch Tehran, Iran. e-mail: am_bidgoli@iau-tnb.ac.ir

effectiveness of this type of solution has been proven; using these approaches in heterogeneous crowded networks would not be practical since the base station have to gather and process large volumes of data which would be time-consuming and inefficient. Several protocols are located in this category, for instance, Stable-Aware Evolutionary Routing Protocol (SAREP) [1], at the beginning of this approach nodes are required to inform the base station of their id, position, and energy level. Then, N solution will be generated by the base station in each of which nodes with a high amount of residual energy have a higher chance to be picked as a cluster head. Each of these solutions presents a different network clustering strategy. An objective function will be utilized to evaluate a produced solution. After assessing each solution, the base station collects the best ones as next-generation parents. This process will continue until achieving a certain value of the objective function or till we produce and evaluate a certain number of generations or by reaching a defied limited execution time.

[2] presented a DPSO-Based clustering routing algorithm that utilizes a two-field structure to produce experimental sequences. In other words, in this DPSO-Based solution, a set of random structures of the network (particles) is generated as the solution that each of which represents a clustering strategy. When the network information-gathering phase is completed, the base station collects h*2 nodes with the highest energy level as the cluster head candidates. Then each particle picks its cluster heads from this collection and randomly assigns them member nodes. Produced particles will be evaluated; consequently, the acceleration and success of the solutions will be updated. In this approach, the best solution is applied to orient new solutions.

UMBIC [3] is an unequal multi-hop balanced immune clustering protocol that applies the MOIA algorithm, to collect an optimal number of nodes with the highest energy as the cluster head in a centralized manner. Then, the rest of the cluster formation process and building routing tree will be done by nodes cooperation; the main goal of this phase is to minimize communication costs and discard duplicate data overhead.

As a powerful optimization tool, swarm intelligence has been employed in a variety of useful approaches such as SMO [4]. In this work, the problem of optimal clustering is considered as a Boolean problem so that a binary spider monkey optimization solution is presented to find the optimal solution. Briefly, SMO attempts to simulate search-and-find behavior among a group of spider monkeys. In this strategy, each solution considers as a spider monkey, and evaluation will be performed based on four objectives. Ultimately, each of these four objectives has a specific pre-defined weight that explicates the importance of each objective.

LEACH [5] has been known as a fundamental algorithm for clustering IoT networks; several editions and versions have been proposed on this approach, WHCBA [6] is one of these editions. Operations of this approach are inspired by the actions of a group of bats during hunting. In this population-based solution, parameters such as the position of the bats, acceleration, frequency, propagation rate, and intensity are effective in the exploration of the optimal solution. Besides, after each solution generation, the best one will be applied to calculate the acceleration for future stages. In this approach, the objective function is a function of the standard deviation of the distance between the member nodes and the cluster head. The ultimate goal of such a consideration is to achieve the optimal solution with the minimum inter-cluster distance. Central-Id strategy is also adopted for producing new solutions. In this strategy, the farthest and closest nodes to the cluster head are picked and a search is performed to find the hidden nodes which are located within the aforementioned distance.

BBO: EEEA [7] is another Si-based work in this context. After network information gathering, the base station produces N random solutions named habitant, each of which will be evaluated to calculate Habitat Suitability Index (HSI). Produced solutions will be prioritized based on their HSI and the emigration and immigration rate of each habitat is measured to be used in the production of next-generation solutions. In the biogeography optimization algorithm, migration is considered the main activity to produce the optimal solution. In this approach, the roulette wheel selection mechanism has been adopted to choose the habitat and migrate to it; Migration is used to generate new solutions. To produce more diversity, the mutation rates will adjust dynamically. Ultimately, after the production of new solutions, the immigration and emigration rate of each habitat are recalculated.

To summarize, we can say most of the centralized Si-based approaches in this context, have a genetic-like functionality in producing N different solutions, evaluating them, selecting parents, and finally producing new generations. Although centralized approaches have demonstrated their skill in finding the optimal solution, they aren't suitable for large-scale IoT networks such as smart cities; which would consist of more than 5000 nodes. Due to their dependency on accessing global network information and its time-consuming mechanism for generation, composition, and evaluation of new solutions. This is where distributed solutions come in handy with large, scalable networks which have to achieve an optimized solution by considering time and energy constraints. 3 In Distributed approaches, IoT nodes are responsible to shape the clusters. In these methods, nodes are interacting and collaborating to produce a single optimal

global answer. In this way, they do not require to send their local information fully to the base station and wait for the clustering result. The common skeleton of these approaches is consisting of these sub-phases; 1) Neighbor node discovery phase. 2) Cluster head competition phase. 3) Shaping clusters phase. 4) Cluster optimization phase. 5) Building a data transportation path phase. HUCL [8], EADUC [9], HEEC [10], HCD [11] are some of the proposed methods in this category. Although the above algorithms follow a similar structure they proposed diverse formula for node radius, waiting time (TW), and relay node evaluation calculation by considering different parameters which brought diversity in the final achieved answer.

b) Equal or Unequal Clusters

Regularly, Cluster radius can be categorized into three groups; 1) non-defined radius. 2) Defined and equal radius. 3) Defined and unequal radius.

Non-defined radius: BBO: EEEA [7] is one of the strategies that do not consider a certain limited radius for cluster head. In general, distributed approaches can't be placed in this category. In other words, the neighbor node discovery phase has required distributed methods to define a specific radius for nodes to set a border for node communication and cooperation.

Defined and Equal Radius: Although fixing an equal radius for clusters seems to be a practical solution for load balancing in homogenous IoT networks; it still has a negative aspect when we utilize multi-hop data transferring mechanism. In such situation, cluster nodes that are located next to the base station have to carry a lot of loads; which eventually leads to the energy hole problem near the base station and consequently reduces network coverage. To deal with this problem, the unequal cluster radius has been proposed.

Table I: Categorized Reviewed IoT Clustering Algorithms

Approach	Inter Cluster Layering	Inner Network Layering	Assistance Node Usage	Equal Or Unequal	Centralized Or Distributed
SAREP[1]	-	-	-	Unequal	Centralized
DPSO[2]	-	-	-	Unequal	Centralized
UMBIC[3]	-	-	VCH	Unequal	Centralized
SMO[4]	-	+	-	Unequal	Centralized
WHCBA[6]	-	-	-	Unequal	Centralized
BBO:EEEA[7]	-	-	-	Unequal	Centralized
HUCL[8]	-	+	-	Unequal	Distributed
EADUC[9]	-	+	-	Unequal	Distributed
HEEC[10]	+	+	ACH	Unequal	Distributed
HCD[11]	+	+	ACH	Unequal	Distributed

Defined and Unequal Radius: The unequal cluster radius has been proposed to tackle the energy hole problem. In this solution to overcome high load near the base station, the nodes which are positioned nearby the base station will have a smaller radius. HUCL [8] is a sample of this class.

c) Usage of Assistant Nodes

Assistant nodes selection can be employed to minimize the energy depletion of cluster heads and also member nodes. The assistant nodes can be classified into two categories; 1) ACH. 2) VCH.

ACH: These nodes are placed within a limited radius of the cluster head and are required to collect and aggregate data from members; they ultimately forward aggregated data to the related cluster head. Thus, not only inter-cluster communication distance will be minimized but also cluster heads need to waste less energy for data receiving and aggregation. HCD [11] is a sample of this category.

VCH: UMBIC [3] is an example that has employed VCH nodes. These nodes are selected as substitutes for the cluster heads so that if the cluster head energy is dropped under a defined threshold; it could be replaced immediately without any additional processing. Employing this kind of assistance nodes will be essential especially for long-distance networks that when a cluster head dead occurs, member nodes are required to transfer their data directly to the base station which causes quick energy depletion.

d) Utilizing Inner-network Layering

In general, we can categorize mentioned algorithms into these groups; 1) Two-layered networks. 2) Four-layered networks.

Two-Layered Networks: EADUC [9] simply defined a predefined threshold. So that, we can divide the network into two layers. The first layer comprises nodes that are placed near the base station. To overcome the energy hole problem these nodes are not permitted to pick a relay node.

Four-Layered Networks: Nodes that are located in the farther layers have a greater radius in comparison with the ndes that are placed in lower layers. On the other hand, this layering makes it possible to construct a data

transmission path by avoiding the loop. HCD [11] is a sample of this category.

e) Utilizing Intra-Cluster Layering

Due to the remarkable performance of inner-network layering, some solutions decided to apply this model for layering clusters as well as networks. Based on this idea, the cluster heads can share their data load among nodes that are located in the first layer. This mechanism is adopted for finding node assistants. HEEC [10] is an example of these algorithms.

II. System Model

In this section, we present a real word smart city IoT network platform named SmartSantander; which inspired our performed simulation. SmartSantander [12] is a city scaled IoT project that has followed two principal aims: First, With the expansion of urbanization, city management is no longer as simple as it uses to be, and the former conventional systems are not proper for assisting city administrators; to make right decisions and perform timely and decent actions. By extending an efficient IoT platform, now it is possible to develop and present a varied range of services for citizens to improve the quality of resident's life.

Second, to create a platform for researchers and developers for testing their IoT applications and services, in a real-world IoT platform. Smart Santander platform is composed of more than 20.000 indoor and outdoor IoT heterogeneous nodes which have been distributed on a city scale. These nodes can be utilized for multiple purposes; 1) Environmental monitoring. 2) Outdoor parking area management. 3) Parks and gardens irrigation. Three proposed applications Are some of the applications that can be implemented by applying this platform and has adopted a 3-tiered architecture, contains:

1) IOT Node

Which are typical IoT nodes.

2) Repeaters

Are IoT nodes that not only sense environmental parameters and send their prepared data to the base station but also forward other node's data. The communication between repeaters and IoT nodes performs through the 802.15.4 protocol which its range is limited between 10 to 100 meters [13]. So that we have employed this restriction in our simulations to achieve an estimation near the real world.

3) Gateways

Both regular IoT nodes and repeaters are configured to send their data to gateways. finally, gateways send data to other machines or data centers. In this architecture, several gateways have been used in the environment to reduce data transformation distance.

The following assumptions are made for our system model:

- Nodes are distributed randomly and are stationary.
- Nodes are configured to send their data to the nearest gateway.
- Nodes are heterogeneous in terms of energy, produced data packet length, and buffer size.
- Buffer limitation forces to forward data to the gateway by buffer filling event and ultimately, clear buffer to be able to accept new data from cluster members.
- The node buffer's size is assumed to be 3 times more than each node's data packet length.
- The maximum radius node is assumed to be 100 meters.

We utilize a radio model same as [11], [14]. Here, consumed energy for transferring K bits data through dij distance can be calculated using the following equation.

$$E_{TX}(i,K,d_{ij}) = \begin{cases} E_{elec}\,K + E_{fs}Kd_{ij}^2 & d_{ij} \le d_0 \\ E_{elec}K + E_{mp}Kd_{ij}^4 & d_{ij} > d_0 \end{cases} \quad (1)$$

dij is the Euclidean distance between two nodes; d0 is a threshold distance that can be calculated by employing the following equation.

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (2)$$

Finally, consumed energy for receiving K bits data can be calculated using (3).

$$E_{RX}(i,K) = E_{elec}K \quad (3)$$

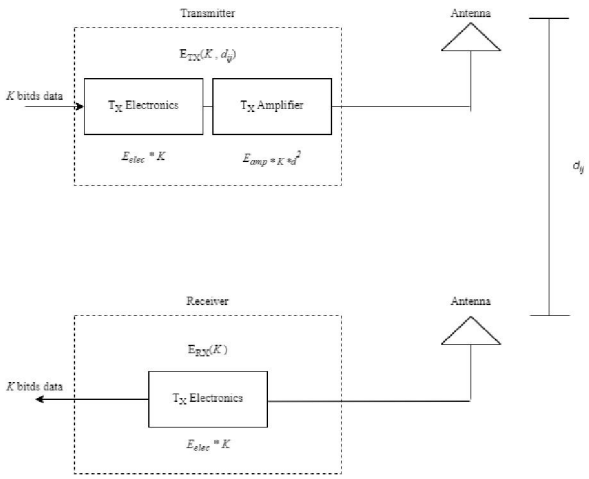

Fig. 2: IoT Radio Model.

III. Performed Improvement

With the rapid expansion of IoT applications, various challenges such as energy optimization and scalability have been considered. These problems are further complicated by considering several real-world constraints, such as energy, time, communication

radius, and buffer size restrictions which are often in conflict with each other.

During recent years several notable solutions have been introduced to face these concerns; each of them is different in terms of the combination of considered parameters and limitations.

Here we present an effective improvement on the existing energy-aware distributed unequal clustering protocol (EADUC) algorithm. Our presented method is appropriate for clustering heterogeneous, large-scale IoT networks. Our recommended improvement attempts to enhance network lifetime by minimizing the number of dead nodes and consumed energy through utilizing three main enhancements:

a) *Neighbor Node Discovery Phase.*
b) *Optimizing Clusters Phase.*
c) *Assistance Node Selection Phase.*

Considering a smart city network that consists of more than 2000 heterogeneous nodes; gathering global network information, constructing and evaluating solutions, reproduction of new solutions generations, and ultimately providing an optimal answer would be time-consuming especially when a mixture of parameters should be considered. Thus, centralized strategies are not appropriate due to their time-consuming processes. However, distributed IoT clustering approaches are suited for networks with a considerable number of nodes, particularly under time restraints.

Our distributed approach is consisting of six sub-phases: 1) Neighbor node discovery phase. 2) Cluster head competition phase. 3) Shaping clusters phase. 4) Cluster optimization phase. 5) Assistance node selection phase. 6) Building a data transportation path phase. We further describe these sub-phases in detail.

a) *Neighbor Node Discovery Phase*

In this sub-phase, each node gathers local network information by interacting with the neighbors; this information will be required for future decisions. Here, three significant questions must be answered:

- What effects does radius computation have on energy efficiency?
- How to calculate the optimal radius for a node?
- How to lessen the energy consumption in the neighbor node discovery sub-phase?

Above mentioned questions will be answered in this section one by one. As mentioned earlier, to tackle the energy hole challenge near the gateway, we produce clusters with unequal radius. Since by lessening the node's radius the number of neighbor nodes or future cluster members will be dropped. Hence, the radius of a node directly affects the load of the cluster, this fact has been used to coordinate the node's load by considering a variety of properties and restrictions such as 1) node's current energy. 2) node's network position. 3) node's buffer size. 4) node's data packet size.

Moreover, cluster radius directly influences the number of clusters in the network. Since as the node's radius decreases more clusters will be needed to cover the same network area. In this situation, although the cluster head's load decreased; increasing the number of clusters will cause grow the total transmission distance and subsequently, total network consumed energy.

In distributed clustering algorithms, optimal radius determination has been considered as a multi-objective optimization problem since there exists more than one parameter or objective for minimization or maximization; which some of them may have conflicts with each other.

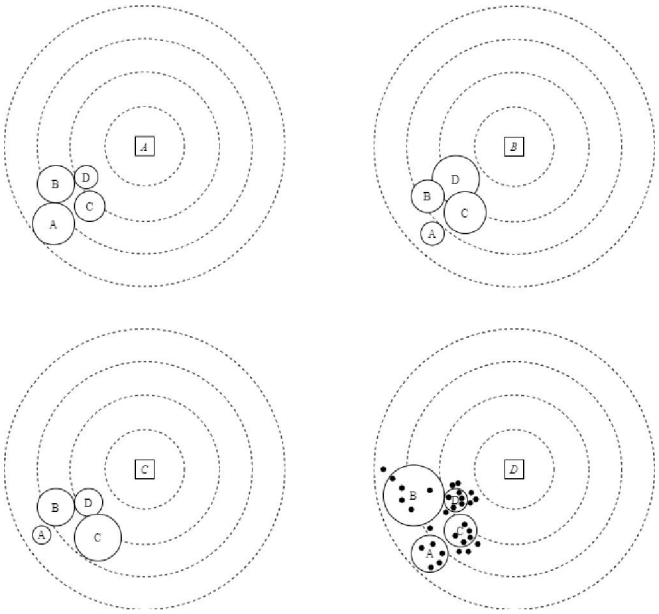

Fig. 3: Multi-Objective Radius Optimization.

Assume four nodes: A, B, C, D; which are varied in terms of 1) Distance to the gateway. 2) Energy level. 3) Buffer size. 4) Area density. Suppose that We have considered four goals in node's radius optimization:

A: To minimize the radius of nodes that have less distance to the gateway. Consequently, we would be able to reduce the load of nodes near the gateway and prevent the energy hole problem.

B: To maximize the radius of nodes that have more residual energy or in other words have consumed less energy during previous rounds. Hence, we could lessen the load of nodes that have less remaining energy. 6

C: To minimize the radius of nodes that have a smaller buffer size.

D: To minimize the radius of the nodes which are located in a dense area.

Figure 4 is presenting our four radius optimization strategies and conflicts among them. For instance, in the first strategy A has the maximum radius since it has more distance to the gateway. However, in the second strategy A has the least radius since it has a lower energy level.

As we stated node radius optimization has been considered as a multi-objective optimization problem in IoT distributed clustering approaches; due to its intent to minimize or maximize multiple parameters and achieving multiple conflict goals. Here, we utilized a decomposition multi-criteria optimization solution named weighted sum same as radius computation formula which is employed in EADUC [9]. These objective functions can be considered:

To maximize the radius of nodes that have more distance to the gateway. Thus we would be able to lessen the load of nodes that are located near the gateway and prevent the energy hole problem. To this end, we used the (4) as an objective function that encounters the distance from the maximum value. (dmax - dmin) is the network length of the current gateway. The resulted value will be bounded between 0 and 1. For nodes that are located near the gateway this value is smaller.

$$Z_5 = \frac{X - \max(X)}{\max(X) - \min(X)} = \frac{d_{max} - d(s_j, BS)}{d_{max} - d_{min}} \quad (4)$$

To maximize the radius of nodes that have consumed less energy during previous rounds. Consequently, we could lower the load of nodes that have consumed more energy previously. Thus, we applied (5) as an objective function that expresses the relative proximity of each node energy value to its maximum energy. The resulted value will be bounded between 0 and 1. For nodes that have been wasted more energy this value is smaller.

$$Z_3 = \frac{X}{\max(X)} = \frac{E_r}{E_{max}} \quad (5)$$

Ultimately, by multiplying each objective function by -1 or by subtracting the limit value (between 0 and 1) by 1; we can convert the minimization objective function to maximization or vice versa. The Rlmax is a predefined coefficient that directly affects the radius of nodes and maximizes our control on radius computation.

When it comes to reducing energy consumption in the neighbor node discovery sub-phase; Restricting the number and the size of broadcasted packages will be profitable. As the result, here we have applied an improvement in the original EADUC approach.

In regular neighbor node discovery strategies, each node broadcasts a request for information message in their restricted radius. Then, each node that hears this message will respond back by sending a packet containing its properties. consequently, if a node hears n requests for information message; it will send n packages in reply. Our applied strategy required a node to broadcast a replay packet only once and reduces the consumed energy from nETX to maximum ETX. Fig 4 is representing mentioned strategy. Fig 6 displays, our presented method, which can be split into two sub-steps:

T1: This is the duration in which nodes broadcast the request for the information message. Each neighbor node will receive the message and saves the id and position of each received request in a list. So that in the next step it can increase its radius just in case to obtain the maximum required radius to answer the message once; temporary. Since our assumption considered the 802.15.4 protocol range limitation, we can be assured that this radius will never exceed 100 meters.

T2: This is the duration in which each node checks its list to find the maximum necessitated radius to broadcast its information once.

b) Cluster Head Competition Phase

The primary purpose of this sub-phase is to collect optimal cluster heads by maximizing their opportunity to be selected. In this strategy, a waiting time will be set by each node. Nodes that have not received the head message till the end of their waiting time; will consider themselves as cluster heads and broadcast a head message in their radius to inform their neighbors. For maximizing the chance of optimal nodes to become a cluster head or minimize the waiting time of optimal nodes; this problem can be considered as an optimization problem. To this end, a variety of node's properties can be considered:

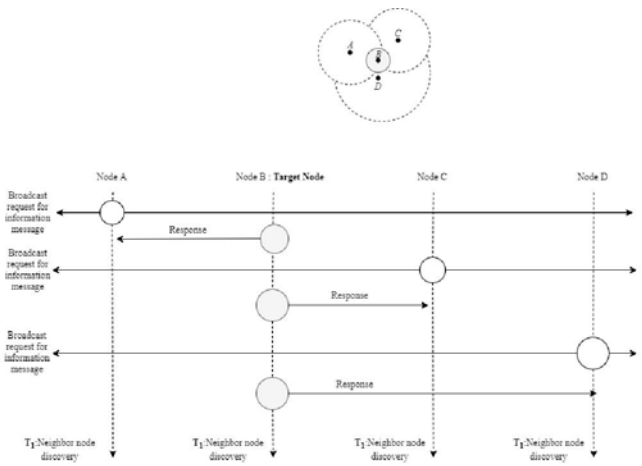

Fig. 4: Regular Neighbor Node Discovery.

Node energy state, the proportion of average energy of the neighbors (6) and current node residual energy (7); can be considered as an objective function. consequently, nodes that have more energy in comparison to their neighbor average energy; have more chance to become a cluster head.

$$E_{avg\_res} = \frac{\left(\sum_{j=1}^{m} s_j E_r\right)}{nb} \quad (6)$$

$$Z_3 = \frac{X}{\max(X)} = \frac{E_{avg-res}}{E_r} \qquad E_r \geq E_{avg-res} \quad (7)$$

Or the amount of consumed energy (8), in this way we could maximize the chance of becoming cluster heads for nodes which, have lost less energy.

$$Z_3 = \frac{X}{\max(X)} = \frac{E_{rem}}{E_{int}} \quad (8)$$

The number of times that a node acts as a cluster head, by applying this parameter, we can improve the chance of becoming a cluster head for nodes that have not spent much energy as a cluster head. 8 The number of neighbor nodes, this parameter can be employed to coordinate the load balancing among the nodes.

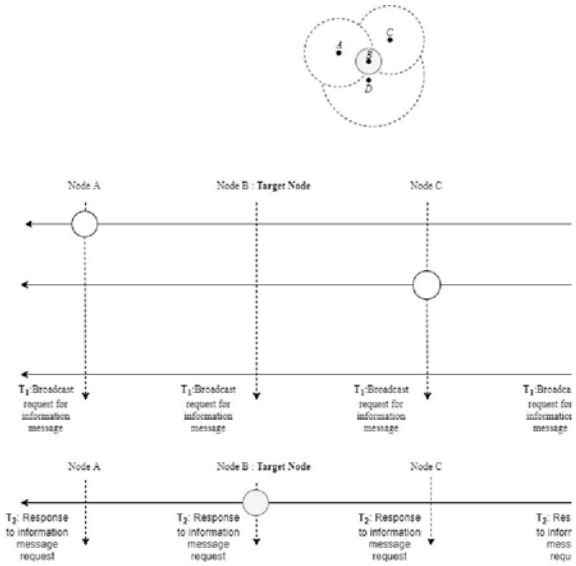

Fig. 5: Proposed Neighbor Node Discovery.

c) Shaping Clusters Phase

In distributed approaches after completing the cluster head competition phase; entire the network is covered by clusters. So that each alive node has the rule of cluster head or member node. In the cluster formation phase, member nodes are required to select an appropriate cluster head from the developed cluster head list.

To reduce the intra-cluster distance and consequently consumed intra-cluster communication energy. Regularly, member nodes are expected to pick the cluster head with the least distance. Hence, a cluster head's load highly depends on the density of the located area, radius, and Tw of its surrounded nodes. To enforce more control on load balancing here we have applied an additional phase named the cluster optimization phase.

d) Clustering Optimization Phase

To enhance better load balancing, here we have implemented our improvement by utilizing a well-known swarm intelligence algorithm (SI), named Biogeography-based optimization (BBO) [19] in a distributed manner.

BBO is an intelligence optimization algorithm introduced in 2008 by Dan Simon. The basic idea is based on this central question, how various plant and animal species are distributed in geographical areas. As a fact, we know that animals tend to monopolize resources, so the ideal locality for them is a place with the lowest possible population.

This performance of animals in emigrating from densely populated areas and immigration to sparsely populated areas is the basic idea of the BBO optimization algorithm; which can also assist us to achieve better load balancing in IoT clustering. To provide a better perception of our implementation, consider four nodes A, B, C, and D.D selected C as cluster head among its head list since it has the least distance. C has got the highest population among two other possible habitats so its emigration rate is higher than two other options. As a result, D has a high intend to leave C and find a new cluster or habitant that has a less population.

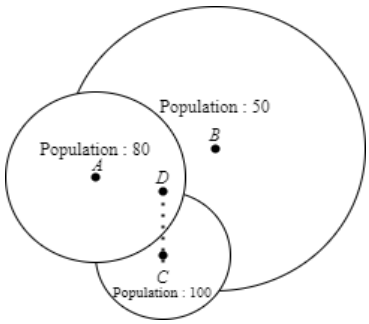

Fig. 6: Node possible habitants.

This is our viewpoint in employing the BBO algorithm; in other words, by utilizing head list as a possible habitat list and BBO, each member node does immigration and 9 emigration to distribute load among possible habitats and achieve better load balancing.

To this end after regular cluster shaping; each cluster head as a habitant calculates its current load and informs its member by broadcasting a small 32 bits message. Thus, each member node can compute the habitat suitability index (HSI) for each of the possible habitats in the head list. This HSI is an objective function that can improve cluster head load balancing. Here we considered three habitant suitability index (SIV) that forms our final HIS:

The required energy to transfer the node's data packet. This suitability index has a direct relation to the distance. So, the nearest cluster head is the most proper option. Here we considered the distance to the minimum Etx needed as an SIV.

$$Z_5 = \frac{X - \min(X)}{\max(X) - \min(X)} = \frac{E_{tx} - \min(E_{tx})}{\max(E_{tx}) - \min(E_{tx})} \quad (9)$$

Our second suitability index can be considered as a constraint. Hence, we can convert our HSI to a constrained optimization problem. To this end here we defined load violation as a parameter that affects our final objective function. Our main goal is to keep the habitant's load less or equal to its buffer size.

$$V\{g(x) < g_0\} = \begin{cases} 0 & g(x) \le g_0 \\ \frac{g(x)}{g_0} - 1 & g(x) > g_0 \end{cases} \quad (10)$$

Here g0 is the buffer size constraint of habitant x and g(x) is the current load of habitant x. This load violation value will use as a penalty value which can be appended to the objective function and increased the value of the HSI in the minimization problem. Our third SIV attempts to prevent the accumulation of load in a habitat. Worst load violation is a parameter that shows load violation when the habitant accepts the maximum possible load. Here we apply (10) for violation computation but we considered g(x) as the current habitant worst load. Ultimately, our HSI will be presented as follow:

$$HSI = w_1 \left(\frac{E_{tx} - \min(E_{tx})}{\max(E_{tx}) - \min(E_{tx})} \right) + w_2(Current\ Load\ Violation) + w_3(Worst\ load\ Violation) \quad (11)$$

w1, w2, and w3 are weights that reveal the parameter importance. After calculating HSI, evaluated habitats will be sort in descending order. Ultimately, the immigration rate and emigration rate can be computed based on [19].

Although the BBO algorithm has been utilized in other centralized IoT clustering approaches previously; we were able to present a distributed usage of this algorithm which helps us not only get profits of its capabilities but also consider the time limitation in our network.

In the other words, to present more controllable load balancing we permit our member nodes to reshape the clusters; utilizing their local information and BBO as a practical load balancing tool. On the other hand, we have been able to enforce buffer size constraints by employing a constrained objective function as HSI.

e) Assistance Node Selection Phase

To reduce energy dissipated in the urgent conditions that a cluster head dies or lost its energy more than the expected defined value. we have employed VCH assistant nodes selection. These nodes are selected as substitutes for the cluster heads so that if the cluster head died; it could be replaced immediately without any additional processing.

Employing this kind of assistance nodes will be essential especially for long-distance networks when a cluster head dead happens, member nodes are required to transfer their data directly to the base station which causes quick energy depletion.

To pick an appropriate VCH, each cluster head does member evaluation to find a VCH that has residual energy more than average member energy and has the least distance to the current cluster head

f) Building A Data Transportation Path Phase

Here a multi-hop strategy is adopted, to reduce communication distance between the cluster heads and the gateway. Same as the EADUC algorithm, a threshold distance named dist_th is defined. If a cluster head is placed in a distance less than dist_th; it is not allowed to pick a relay node; so that we can face the energy hole problem.

To build a data transmission path, each cluster head broadcasts its information in its radius to inform other cluster heads. Each cluster head that receives this information; will save this data in a list so that it can evaluate each cluster head based on required relay energy and ultimately picks the one which needs less energy waste. 10 Fig. 8 A comparison of Average Cluster Head's Load Standard Deviation before and after improvement.

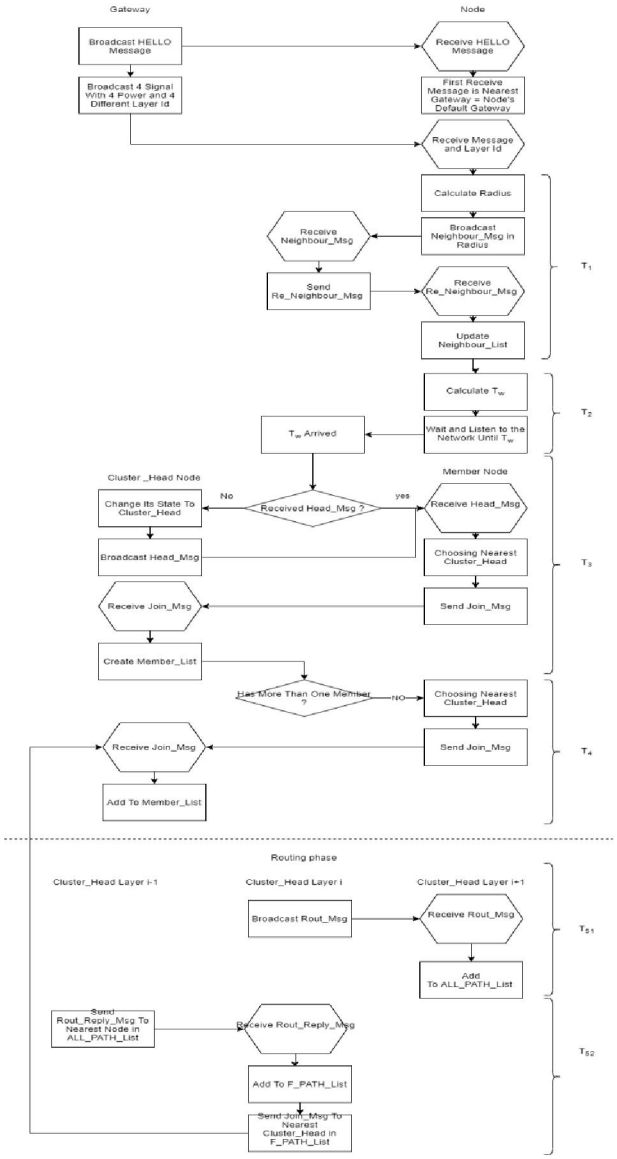

Fig.7: Flowchart of Presented Improved EADUC Algorithm.

IV. Experimental and Performance Analysis

To carry out our simulation MATLAB R2019b has been employed. To present a more explicit analysis of our presented improvement. First, we consider a simple 200×200 square-shaped environment that includes 3,401 randomly distributed heterogeneous nodes; each of which is modified to interact with the nearest gateway. Table II and III demonstrates the network composition which has considered as our first scenario.

Table II: Scenariosimulation Parameters

Parameter	Value
Total Temperature Repeater Nodes	920
Total Light Repeater Nodes	353
Total Noise Repeater Nodes	558
Total Gases Repeater Nodes	443
Total Parking Sensor Nodes	337
Total Multimedia Nodes	750
Rlmax	10

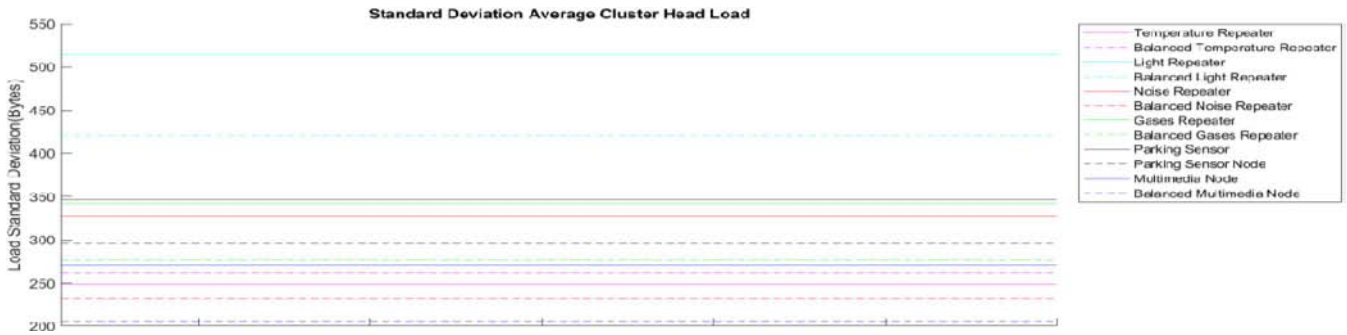

Fig. 8: A comparison of Average Cluster Head's Load Standard Deviation before and after improvement.

In the homogeneous networks, load balancing can be simply measured by calculating the standard division of the cluster head's load. To this end, standard division can be calculated in a single round. Since the goal is to build clusters with the nearly same amount of load. However, in heterogeneous networks, each cluster load depends on its constraints, and their capacity is not comparable. To achieve this comparison possibility, here we present a measurement named standard division of average cluster head's load. To measure this parameter, in each round we calculate the average load of each cluster head types separately ultimately, we can compare this average for each type using standard division considering total rounds.

Fig 9 illustrates the ability of our improvement in achieving load balancing in heterogeneous networks. Since the standard division of each cluster head type reduced successfully after applying our presented improvement.

$$\sigma i = \sqrt{\Sigma} |\mu i - \bar{x} i| 2Nj = 1N \qquad (12)$$

Formula (12) represents the standard division of cluster head type i, during N rounds of clustering and data transition. μj is the average load of node type i in j-th round and $\bar{x}i$ is the average of average load for type i during all N rounds.

Our simulation with examining scenario 1, shows a reduction in total load standard division from 2.0514e+03 to 1.6926e+03 after implementing BBO load balancing improvement; which reveals a decline near 17.49 percentage.

Table III: Simulation Parameters Parameter

Parameter	Value
Temperature Repeater Initial Energy	5 J
Temperature Repeater Packet Size	300 bites
Temperature Repeater Buffer Size	300×3 bites
Light Repeater Initial Energy	6 J
Light Repeater Packet Size	500 bites
Light Repeater Buffer Size	500×3 bites
Noise Repeater Initialize Energy	7 J
Noise Repeater Packet Size	400 bites
Noise Repeater Buffer Size	400×3 bites
Gases Repeater Initial Energy	8 J
Gases Repeater Packet Size	400 bites
Gases Repeater Buffer Size	400×3 bites
Parking Sensor Initial Energy	9 J
Parking Sensor Packet Size	200 bites
Parking Sensor Buffer Size	200×3 bites
Multimedia Node Initial Energy	10 J
Multimedia Node Packet Size	400 bites
Multimedia Node Buffer Size	400×3 bites
Eelec	50 nJ/bit
Efs	10 pJ/bit/m2
Emp	0.0013 pJ/bit/m4
EDA	5 nJ/bit/message
MiniSlot	20
MajorSlot	2
Rlmax	-
Packet Header Size	50 bites
I,E	1
α,β,	0.333
Total Gateways	25
W1	0.2
W2	0.4
W3	0.4

To confirm the efficiency of our presented improvement, in a near-real word environment the second scenario is considered. Moreover, four other existing distributed clustering algorithms named EADUC, HCD, HUCL, and HEEC have been simulated using the same scenario parameters to perform a satisfactory comparison. To measure the performance of each approach four different parameters has been examined:

1) Total Network Remaining Energy

Energy consumption is identified as a key restriction in prolonging network lifetime. This parameter has close dependencies with other network measures such as the number of cluster heads and cluster load that has made the optimization more complicated. However, by expanding cluster heads load will distribute more and reduced. This relation can be demonstrated in

fig 8 and fig 11. Here we have selected EADUC as the base algorithm to implement our performed improvement; since it has the least amount of cluster heads among the three other approaches.

The following chart displays the performance of our improvement in energy efficiency; Our improved EADUC approaches were able to waste the least energy in contrast to the other four approaches. In comparison to the original EADUC by remaining network energy 56728.83, our improved approaches were able to keep 57631.79 which shows a 1.59 percent drop in energy consumption.

2) Total Number of Cluster Heads in Each Round

3) Total Number of Alive/Dead Nodes in Each Round

The number of alive/dead nodes is the next essential criterion for prolonging network lifetime and preserving network coverage. The two following charts display the ability of our offered approach to decrease the ratio of losing nodes.

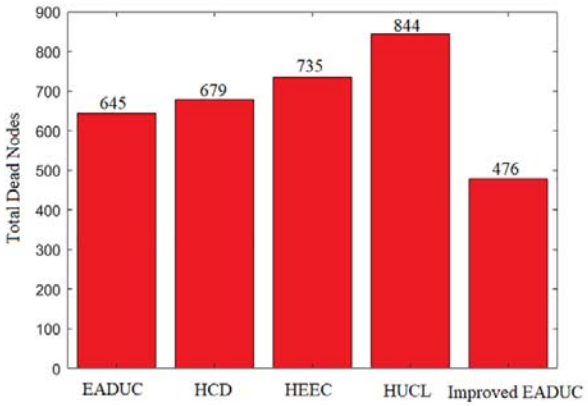

Fig. 13: Total Dead Nodes Scenario 2.

Table III: Scenario 2 Simulation Parameters

Parameter	Value
M×M	2000×2000
Total Temperature Repeater Nodes	1920
Total Light Repeater Nodes	1353
Total Noise Repeater Nodes	1558
Total Gases Repeater Nodes	1443
Total Parking Sensor Nodes	1377
Total Multimedia Nodes	1750
RImax	100

V. Data Transmission Simulation and Heterogeneous Buffer Size

With no doubt, heterogeneity is the primary property of an IoT real-world network. Since a typical IoT network may comprise hundreds of nodes, produced by different companies and with a variety of capabilities and constraints. Moreover, these heterogeneous nods may employ diverse communication protocols. This fact has made clustering optimization more complicated. So that, although several practical approaches have been introduced during recent years, not all of the heterogeneous properties and real-world network conditions have been considered by them.

In this paper, we have considered a group of real-world attributes in our presented work and simulations:

1) Node's Energy Heterogeneity.
2) Node's Produced Packet Size Heterogeneity.
3) Node's Buffer Size Heterogeneity.
4) Limited Node Radius (Zigbee Communication Radius Constraint).
5) Clustering Execution Time Limitation.

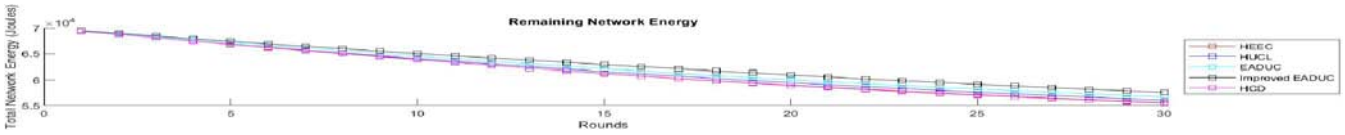

Fig. 9: Network Energy in Scenario 2.

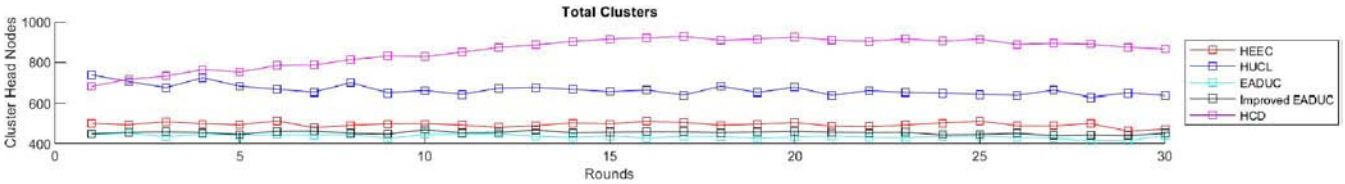

Fig. 10: Total Cluster Heads Scenario 2.

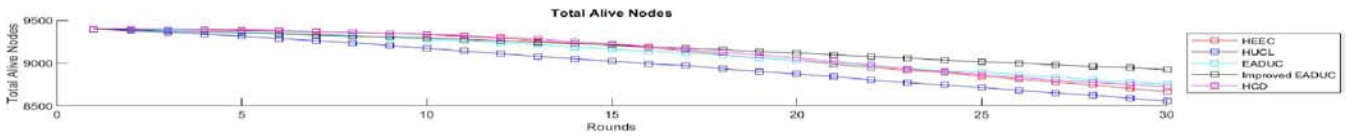

Fig. 11: Total Alive Nodes Scenario 2.

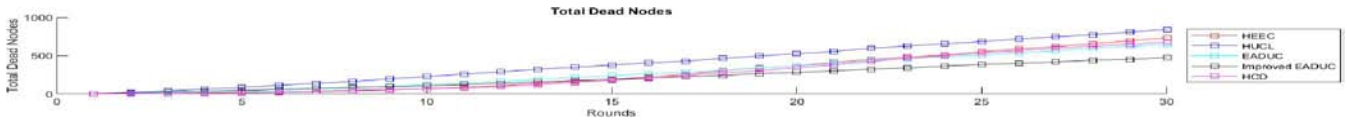

Fig. 12: Total Dead Nodes Scenario 2.

In this section, we present our data transmission phase by considering buffer size constraint and packet size heterogeneity. In most of the data gathering mechanisms, TDMA be instructed to gather data from the farthest nodes first. Hence, data can properly be aggregated by intermediate nodes before forwarding to the gateway. In the other words, the node's defined slot time is related to its distance to the next-hop node.

```
while Remain packets > 0 and Cluster Head is alive do
    if Cluster head's buffer is full then
        Forward data;
        Clear buffer;
    else
        forall members do
            if member is alive and Remaining Head's buffer > 0 and
             member's data has not been forwarded totally then
                Adds member's data to head's buffer;
                if member's data placed in head's buffer completely then
                else
                    Head's buffer is full;
                    Not all member's data has been forwarded completely;
                end
            end
        end
    end
end
```

Fig. 14: Data Transition.

Here to consider buffer limitation, we start data gathering from the farthest nodes. Consider node A in fig 14 as a cluster head; TDMA is set to collect data from nodes based on their distance to A. This data accepting by node A will continue until it reaches buffer size limitation. Then, to resume its process, node A is required to unload its buffer by transferring its collected data to the 13 next hop. consequently, the buffer will be reset to accept the rest of the data.

This procedure will be continuing as mentioned in pscode, until the rest of the data will be gathered or its energy be consumed completely.

VI. Conclusion

In this paper, first, we categorized our studied approaches based on a variety of properties; then based on the presented classification we selected distributed approaches as an appropriate clustering method for large-scale heterogeneous IoT networks. Due to their ability to present optimized solutions by utilizing local limited information promptly. Eventually, to perform our enhancement we selected EADUC approaches among three other existing distributed strategies. The main reason for our selection is the capability of EADUC to present the most energy conservation clustering solutions by employing the fewest amounts of cluster heads. Finally, we implement three main improvements to the selected strategy:

a) Neighbor Node Discovery Phase

Our performed enhancement can lessen the number of required messages and consequently wasted energy in this phase.

b) Optimizing Clusters Phase

Here we employed BBO as a practical tool for distributing loads among heterogeneous neighbor clusters. To orient the migrations based on heterogeneous limitations we have implemented a constrained optimization objective function as a habitat suitability index.

c) Assistance Node Selection Phase

We recognized two types of assistance nodes in the current presented approaches, as we have considered large-scaled long-distance IoT networks; missing cluster heads during the data transmission phase will cost a lot. Since by losing the cluster head, all members are required to send their data directly to the gateway; This will cause energy wastage. To assure the cluster structure we have employed VCH assistance node selection. So that by missing a cluster head it can immediately be replaced by other members with no further computation; hence the cluster structure will be maintained.

Three cited advancements are presented in the EADUC algorithm to provide an optimized distributed clustering approach for heterogeneous IoT networks.

References Références Referencias

1. Khalil, E.A. and Bara'a, A.A., 2013. Stable-aware evolutionary routing protocol for wireless sensor networks. Wireless Personal Communications, 69(4), pp.1799-1817.

2. Li, J. and Liu, D., 2015, October. DPSO-based clustering routing algorithm for energy harvesting wireless sensor networks. In 2015 International Conference on Wireless Communications & Signal Processing (WCSP) (pp. 1-5). IEEE.
3. Sabor, N., Abo-Zahhad, M., Sasaki, S. and Ahmed, S.M., 2016. An unequal multi-hop balanced immune clustering protocol for wireless sensor networks. Applied Soft Computing, 43, pp.372-389.
4. Sabor, N., Abo-Zahhad, M., Sasaki, S. and Ahmed, S.M., 2016. An unequal multi-hop balanced immune clustering protocol for wireless sensor networks. Applied Soft Computing, 43, pp.372-389.
5. Handy, M.J., Haase, M. and Timmermann, D., 2002, September. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In 4th international workshop on mobile and wireless communications network (pp. 368-372). IEEE.
6. Cui, Z., Cao, Y., Cai, X., Cai, J. and Chen, J., 2019. Optimal LEACH protocol with modified bat algorithm for big data sensing systems in Internet of Things. Journal of Parallel and Distributed Computing, 132, pp.217-229.
7. Mehta, K. and Pal, R., 2017, October. Biogeography based optimization protocol for energy efficient evolutionary algorithm:(BBO: EEEA). In 2017 international conference on computing and communication technologies for smart nation (IC3TSN) (pp. 281-286). IEEE.
8. Malathi, L., Gnanamurthy, R.K. and Chandrasekaran, K., 2015. Energy efficient data collection through hybrid unequal clustering for wireless sensor networks. Computers & Electrical Engineering, 48, pp.358-370.
9. Gupta, V. and Pandey, R., 2016. An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks. Engineering Science and Technology, an International Journal, 19(2), pp.1050-1058.
10. Bozorgi, S.M. and Bidgoli, A.M., 2019. HEEC: a hybrid unequal energy efficient clustering for wireless sensor networks. Wireless Networks, 25(8), pp.4751-4772.
11. Han, T., Bozorgi, S.M., Orang, A.V., Hosseinabadi, A.A.R., Sangaiah, A.K. and Chen, M.Y., 2019. A hybrid unequal clustering based on density with energy conservation in wireless nodes. Sustainability, 11(3), p.746.
12. Sanchez, L., Galache, J.A., Gutierrez, V., Hernandez, J.M., Bernat, J., Gluhak, A. and Garcia, T., 2011, June. Smartsantander: The meeting point between future internet research and experimentation and the smart cities. In 2011 Future Network & Mobile Summit (pp. 1-8). IEEE.
13.
14. Jara, A., Genoud, D. and Bocchi, Y., 2014. Sensors data fusion for Smart Cities with KNIME. In Proceedings of the 1st IEEE World Forum on Internet of Things 2014 (No. CONFERENCE). 06-08 March 2014.
15. Shaikh, F.K., Zeadally, S. and Exposito, E., 2015. Enabling technologies for green internet of things. IEEE Systems Journal, 11(2), pp.983-994.
16. Longhi, S., Marzioni, D., Alidori, E., Di Buo, G., Prist, M., Grisostomi, M. and Pirro, M., 2012, May. Solid waste management architecture using wireless sensor network technology. In 2012 5th international conference on new technologies, mobility and security (NTMS) (pp. 1-5). IEEE.
17. Parwekar, P. and Nagireddy, V., 2018, August. Comparative Analysis of SGO and PSO for Clustering in WSN. In 2018 International Conference on Computing and Network Communications (CoCoNet) (pp. 78-82). IEEE.
18. Mazumdar, N., Roy, S. and Nayak, S., 2018, September. A survey on clustering approaches for wireless sensor networks. In 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA) (pp. 236-240). IEEE.
19. Ergezer, M.E.H.M.E.T., Simon, D. and Du, D., 2008. Biogeography-based optimization. IEEE T Evolut Comput, 12, pp.702-713.
20. Deshpande, V.V. and Patil, A.R.B., 2013, July. Energy efficient clustering in wireless sensor network using cluster of cluster heads. In 2013 tenth international conference on wireless and optical communications networks (WOCN) (pp. 1-5). IEEE.
21. Younis, O. and Fahmy, S., 2004. HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Transactions on mobile computing, 3(4), pp.366-379. 14
22. Sotres, P., Santana, J.R., Sánchez, L., Lanza, J. and Muñoz, L., 2017. Practical lessons from the deployment and management of a smart city internet-of-things infrastructure: The smartsantander testbed case. IEEE Access, 5, pp.14309-14322.
23. Jara, A., Genoud, D. and Bocchi, Y., 2014. Sensors data fusion for Smart Cities with KNIME. In Proceedings of the 1st IEEE World Forum on Internet of Things 2014 (No. CONFERENCE). 06-08 March 2014.

This page is intentionally left blank

Global Journal of Computer Science and Technology: E Network, Web & Security
Volume 22 Issue 2 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A 3 Tier Open Ict Ecosystem Network Design (3TOIEND) and Implementation to Enhance Utilization of Local Ict Infrastructure Via Campus Wlan

By Ikechukwu Innocent Umeh

Nnamdi Azikiwe University

Abstract- This research focused on developing of an open ICT ecosystem network design to enhance the utilization of local ICT infrastructure via network, and reduce overdependence on the Internet in the Nigerian campus network environments. A network is used to intelligently share computer infrastructure, and resources by the users involved. The research first surveyed selected tertiary institutions in Nigeria to ascertain the level of utilization, of expensive ICT facilities including, Internet bandwidth via campus LANs, and to discover the causative and prevailing bottlenecks, which hindered optimal utilization of ICT resources via campus networks. The analysis of the findings, alongside the literatures reviewed, revealed that poor network planning and design were the most prominent factors that cause underutilization of ICT resources via a WLAN, in the Nigerian University campus environments.

Keywords: ecosystem, open ICT, network, design, bandwidth, resource, utilization, wlan, campus, and infrastructure.

GJCST-E Classification: C.2.1

Strictly as per the compliance and regulations of:

A 3 Tier Open Ict Ecosystem Network Design (3TOIEND) and Implementation to Enhance Utilization of Local Ict Infrastructure Via Campus Wlan

Ikechukwu Innocent Umeh

Abstract- This research focused on developing of an open ICT ecosystem network design to enhance the utilization of local ICT infrastructure via network, and reduce overdependence on the Internet in the Nigerian campus network environments. A network is used to intelligently share computer infrastructure, and resources by the users involved. The research first surveyed selected tertiary institutions in Nigeria to ascertain the level of utilization, of expensive ICT facilities including, Internet bandwidth via campus LANs, and to discover the causative and prevailing bottlenecks, which hindered optimal utilization of ICT resources via campus networks. The analysis of the findings, alongside the literatures reviewed, revealed that poor network planning and design were the most prominent factors that cause underutilization of ICT resources via a WLAN, in the Nigerian University campus environments. Sequel to the above findings, an intelligent design of an open ICT ecosystem network that will effectively proffer high utilization of ICT infrastructure and resources via a campus network, was developed to improve the overall network performance throughput. In the end, a new network design tagged 3TOIEND (3 TIER OPEN ICT ECOSYSTEM NETWORK DESIGN) grew. The implementation of the new design yielded the integration of bandwidths sources and available ICT infrastructure in a campus environment into a 3-layer/tier open ICT ecosystem to achieve more than 80% improvement in ICT utilization via the campus network.

Keywords: ecosystem, open ict, network, design, bandwidth, resource, utilization, wlan, campus, and infrastructure.

I. Introduction

The current era of the COVID-19 pandemic has accelerated dependence on ICT for solving communication problems through virtual means. Virtual activities are computer networks-based, and, often Internet dependent. The 2020 COVID-19 pandemic lockdowns, escalated the awareness in virtual activities such as; eLearning, E-commerce, E-entertainment, E governance and etc. The rise can be attributed to the directive by the World Health Organization to reduce physical contact between persons to avert contracting the pandemic (World Health Organization, 2020.). The notable upsurge in virtual activities, triggered the emergence, and the implementation of several new trends in network technology advancements to balance the increasing number of users and network-based facilities around the globe.

Recent trends in computer techno fad such as, the Internet of things (IoT), and other similar technologies, have continued to stimulate the growth in intelligent ICT infrastructures. Modern devices and solutions are intelligently designed to be able to meet up with the dynamic network trend in ICT resource needs, and for resources utilization. The new trends bothers on users, manufacturers, and service providers. Unfortunately, the massive improvements on ICT infrastructure notwithstanding, the quest for the optimal performance and utilization of computer networks remains a severe issue. Since the advent of wireless LANs, which use dual-band routers for the delivery of Internet access and ICT resources via a campus LAN, quite unlike the wired LAN, ICT resource underutilization in educational network environments has escalated, and remained a reoccurring issue (DQE Communications, 2022.) and (Project Store, 2021).

At the moment, ICT and networks are inevitable tools for achieving organizations' set out goals in an educational environment. But, in many developing countries, campus wide area network deployments by the vendors of different ISPs (Internet Service Providers), are solely used for Internet access provision. Most vendors' sole interest in the deployment, is to sell Internet bandwidth, but not concerned with the utilization of the sold bandwidth, talk less of other numerous possible ICT services available for use in a client's network environment (Guimaraes, 2007).

Therefore, the scenario of over-prioritizing bandwidth as the sole resource component of a campus WLAN, causes organizations to spend huge amount of funds for only Internet access provision. Thus, lowering the quest for the utilization of a network for other possible benefits. Also, the problem of high cost disparity in designs, and the modes of internet access bandwidth delivery pose hitches for optimal utilization of ICT infrastructure in many of institution' WLANs. This is because, poor planning, and poor

Author: PhD, Computer Science Department Nnamdi Azikiwe University Awka, Anambra State Nigeria. e-mail: Ik.umeh@unizik.edu.ng

equipment choice, coupled with the unintelligent network design models also contribute to the low utilization of resourceful ICT infrastructures in campus WLANs. (Mahajan et al., 2005). The ISPs often deliver Internet access bandwidth to their clients via fiber-optics, or last-mile microwave radios, which are differentiated by their designs. The difference in the designs pose hiccups to the proper utilization of ICT facilities in the clients' campus WLANs. Furthermore, Interoperability issues between ISPs and clients, expose clients' WLANs to poor susceptible design that cause low utilization of networks (Hassidim et al., 2020).

Finally, bandwidth is an expensive, and difficult to save against downtimes, a feature which pose a serious concern to network administrators as they battle to optimize bandwidth usage for optimal network performance (Dajvit et al., 2015). The poor design at ISPs backbone and clients' network, facilitate network underutilization which promotes resilience to network failure (Richard, 2014). In the same manner, low utilization of bandwidth, or ICT resources available in a network spells inefficiency of administrators, in the administration of their networks. Therefore, securing optimal utilization of ICT resources via a network to attain high throughput, is among the paramount desire of every organization. This can only be achieved through a good network design (Kassim, 2013).

II. Problem Statement

The daily increase in the pivotal role played the Internet and ICT in providing solutions for tertiary educational institutions cannot be overemphasized. But, the inadequacy in using ICT facilities via tertiary institutions' network of some developing countries poses serious challenges to them. Some of the major contributory problems besides the high cost of network deployment, are stated below:

1. The poor utilization of ICT infrastructure in WLAN sequel to design.
2. The failure of different Internet Service Provisions to establish a smooth handshake that will proffer optimal bandwidth utilization.
3. The poor design that can enable WLAN to deliver ICT infrastructures via campus networks
4. The low utilization of ICT infrastructure via networks resulting from heterogeneous designs.
5. The poor design framework to accommodate network or ICT infrastructural growth via campus WLAN.

III. Aim of Study

The quest to ameliorate the low utilization of ICT infrastructure via campus WLANs motivated this study, which is, the development of a 3-tier open ICT ecosystem network design (3TOIEND) to enhance the utilization of ICT infrastructures via campus WLANs and reduce the users' overdependence on bandwidth in a campus WLAN.

IV. Objective of Study

The specific objectives set out to achieve the aim of this research are as follows:

1. The development of an intelligent network design plan for a campus area networks.
2. The development of an architecture that can integrate heterogeneous ICT components and services into an open ICT ecosystem.
3. The development of an intelligent hybrid network core layer comprising of different ISPs integrated with local ICT infrastructures.
4. The design a mesh of network aggregation and access layers using a hybrid of fiber optic and point-to- point radio links.
5. The formulation of policies that will proffer good network maintenance.

V. Literature Review

a) Network and Bandwidth Utilization

According to (Bohge et al., 2007), the advent, and choice of WAN and WLAN for economic, academic, governance, and social purposes aggravated research in ICT resource management, with particular interest in bandwidth. During the early 2000s, the said research, in both the private and public sector, yielded improvement in end-to-end throughput of TCP-based networks. But, in the recent years, the rapid growth in digital data, and the concomitant need to store and protect bandwidth, for big data has presented the need for data center-to-data-center (DC2DC) WLAN optimization. Therefore, (Lougee-Heimer, 2003) recommended the two common business WAN topologies of optimization, which include; Branch-to-Headquarters and Data-Center-to-Data-Center (DC2DC) topologies. Also, (Lougee-Heimer, 2003) noted that branch WAN also support more simultaneous connections, smaller connections, more short-lived connections, yet handle a greater variety of protocols, which are used for business applications such as, email, content management systems, database applications, and web delivery. By comparison, campus DC2DC WAN links tend to require more bandwidth, and are more distant, involve fewer connections, with connections bigger than 100 Mbit/s to 1Gbit/s flows for longer duration, unlike headquarters' DCDC, and thus, branch WANlinks use less bandwidth.

b) Effect of Network Planning on Network Utilization

(Devajit et al., 2013) Showed that the inclusion of network topology in the design of a network proffer effective bandwidth management and overall throughput of a network, while having good network design will give a good understanding of a network, and the location of the resources, which are embedded in a network. Also,

(Scte.org, 2013) maintained that, successful management of a network and the bandwidth therein, begin with good planning and a sound design for the implementation and maintenance of a network to enable it, meet the constant dynamics experienced with modern ICT infrastructures and solutions. Consequently, (John et al., 2011) also revealed that most of the network failures in campus environments emanate from poor designs and poor utilization of the network infrastructures.

c) Network Design and Network Utilization

Many researchers agree that network deployment must begin with an excellent network design, but none of the researchers suggested how to include available ICT infrastructures for utilization via a network. (Kassim et al., 2012) Survey on bandwidth management in LANs, WLANs, and CAMs showed that of the most networks studied were underutilized, and short-lived because of poor design and lousy bandwidth management plan. At present, network devices are continually being reengineered to become more intelligent, to improve network resource utilization for better efficiency, and performance. As such, most network designs have been improved, yet many ICT infrastructures are still underutilized via the networks. Having a good network design before deployment is necessary, and will help to identify where bottlenecks may exist, and the locations from where to collect the best data to help troubleshoot, or monitor the networks better resource utilization (Robert, 2002).

d) Bandwidth Optimization and Network Utilization

Although, bandwidth optimization is significant when considering network utilization. Still, the components which cause over or under-dependence of a network on bandwidth in place of other ICT infrastructure, have to be considered too. As an example, (Ketan, 2002) noted that audio streaming is less concerned when considering network utilization concerning bandwidth. This is because, an audio streaming consumes less bandwidth when compared to a video streaming concerning bandwidth. (Ketan, 2002) Did not state how to improve this condition in a network design, but (Lian, 2001) and (Koller et al., 2003) agreed that high network utilization often causes serious message delays, jitters, or losses, which are caused by the large number of communicating devices in a network. Such condition, leads to degradation in performance or instability of the control loops in a network.

Also, network bandwidth utilization problems are aggravated by the popularity of wireless sensor networks, and the increase of power line transmission. Wireless sensor networks use smaller bandwidths which has higher error rates resulting from low bandwidth utilization (Lian, 2001) and (Koller et al., 2003).

It is obvious that moderate network utilization are more realisable in developed regions because of better bandwidth consumption rates. But, such moderate utilization may not be constant if the design is not taken into consideration at beginning of the network life (Xipeng, 2008). This position buttresses the importance of network design to a network's utilization. As an example, the typical utilization of a last mile broadband link is often fairly low, and less than 10% efficient, because the design offer multiple tiers of broadband access speed, yet many subscribers do not sign up for the highest tier. Therefore, while it is generally thought that having more bandwidth will cause a better link utilization, better design instead does (Xipeng, 2008). Bandwidth availability can be likened to food because, having excess food does not guarantee satisfaction; likewise, a network user may consume excessive bandwidth, without achieving the required network utilization, or the network achieving optimal throughput.

e) The ICT Ecosystem

The ICT ecosystem is a new concept that suggests ICT as a part of an expanded network that recognizes the non-technical dynamics of socioeconomic, political, and spatial concerns (Simon et al., 2018). The ecosystem concept was borrowed from biology to describe the interaction of various actors in a network environment. Accordingly, several studies within the ICT research sphere have applied the ecosystem metaphor (Xipeng, 2008). The Open ePolicy Group, also described how open ICT ecosystem have been carried out in many information technology research (Open ePolicy Group, 2005). Therefore, the thought of applying open ecosystem concept in a campus wired local area network (WLAN) environment could be rewarding, because all homogeneous and heterogeneous components within the ecosystem will handshake every network component to yield, ICT resources, and network utilization, by turning the value chain to a value ecosystem (Rafigue et al., 2012).

f) Students Presence in Campus WLANs

A case study of selected Nigerian university students' application of the Internet for studies revealed that Nigerian undergraduate students make extensive use of the internet by relying on private, or commercial cyber cafes, which are grossly poor for studies (Okon, 2010). The students are often, not considered in the institutions' Internet infrastructural design (Okon, 2010). On the other hand, most postgraduate students who can access the Internet from campus WANs do not possess the necessary skills to utilize available ICT infrastructure via the networks for studies. In some cases, the ICT infrastructures are basically not utilized via a campus WANS (Adeleke & Emeaghara, 2016). A study on the usage of electronic resources in Nigerian

university libraries showed that the ICT infrastructures, and resources are grossly underutilized (Obidike & Mole, 2015). Students are often not aware of available ICT resources, and as such, do not utilize same for studies (Mohamed, 2008). Findings from oral interviews with students in different tertiary institutions in Nigeria showed a few concerns that often compel the management, and administrators of campus networks from fully including students' presence in the institutions' networks. (Siddiquah et al., 2017) Itemized speed of computers, the Internet signal problems, virus threat issues, poor working condition of computers, load shedding, and lack of access to the Internet as the factors that deter students from utilizing campus ICT infrastructures via campus networks. (Valerie, 2018) Attributed poor wireless connectivity problems as one reasons why the students in higher education do not use campus WLAN. The poor budgetary for student presence in the campus WAN designs is also a serious contributor to ICT underutilization via the campus networks of the tertiary institutions.

g) Conclusion on Literature Review

From the literature review, it was evident that good network design is a major prerequisite process in the life of any performance-oriented computer network, and building an ICT ecosystem which comprises of bandwidth, and ICT infrastructure in a tertiary institution's network environment will harness the accommodation, and utilization of the necessary components required by users in a campus network. Such integration, will yield optimal network performance through high ICT resource utilization. In Conclusion, no existing documented algorithm, for the design of an open ICT ecosystem integration or utilization via a campus WLAN was found in the literatures reviewed.

VI. Design Methodology

The Cisco Service-Oriented Architecture (SONA) that adopts architectural modularization, with emphasis on infrastructural considerations, and the Top-Down modular method of splitting a proposed network design into three modular tiers of core, distribution, and access layers was adopted to aid flexibility, facilitation of implementation, ease of testing, and troubleshooting of a design was adopted for the proposed design.

VII. The 3 Tier Open ICT Ecosystem Network Design (3TOIEND)

Sequel to the findings and conclusion drawn from the literature review, a 3 (three) Tier Open ICT Ecosystem Network Design (3TOIEND) would be appropriate to curb the menace of ICT infrastructure/ resources underutilization, and overdependence of only the Internet access provision in a campus network. The 3TOIEND is an intelligent network design that will provide a campus networks with the capability of adequately utilizing every available ICT resources in a network, and depending less on the Internet as the sole source for data and information in a campus WLAN and proffer optimal performance.

a) 3TOIEND Design Algorithm

The 3TOIEND shall be achieved through the following algorithm and flowchart in Figure 1:

Step 1: Planning Campus WLAN design

Step 2: Choosing WLAN design structure and components

Step 3: Provide an Open ICT Ecosystem architecture

Step 4: Design a 3 Tier open ICT ecosystem Network

Step 5.1: Create a robust hybrid ISP open ecosystem core, test layer design

Step 5.2: Create Aggregation Layer, test layer design

Step 5.3: Create Access layer, test layer design

Step 5.4: Network Security Design

Step 6: Test the 3TOIEND design

Step 7: Implement the 3TOIEND design

Step 8: Refine design for future upgrade

b) The 3TOIEND Algorithm Implementation

Step 1: Planning a Campus-Wide Local Area Network Design

In computer science, planning is usually the first step in designing a system. Networks are expensive; therefore proper planning of a network for future developments, or upgrades is required before proceeding on the design and execution. A proper planning will enable the design to accommodate the inclusion of all the necessary requirements. In this case, the requirements included, every available ICT infrastructure and network-oriented services in a campus environment. The planning process began with gaining sufficient in-depth knowledge of the existing ICT facilities, needs, users, and network problems, which are required to be solved by the new design. Also, the planning considered the choice and viability of the platform that will be used, as well as, the available sources of Internet bandwidth subscriptions. Afterwards, the plan was tested for viability, and the test results were considered for every stage of the design process.

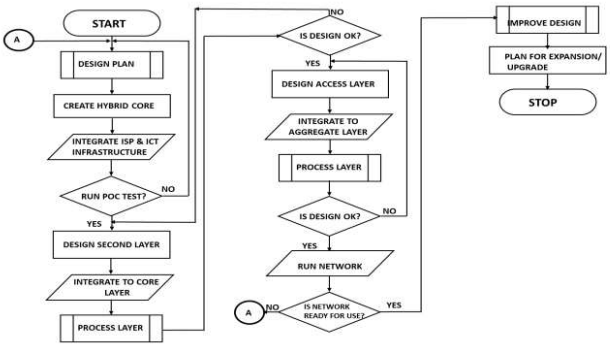

Figure 1: Flowchart for the 3TOIEND

Step 2: Choosing the WLAN Design Structure and Components

A campus network is a portion of an enterprise network infrastructure which enable users to communicate with, and use available ICT resources (CISCO, 2016). The choice of the design architecture adopted for the 3TOIEND was, the layered (hierarchy), modularity design. The choice was based on its resilience and cost effectiveness in guaranteeing ICT infrastructure, resource utilization, ease of management and performance, expansion, equipment integration, and operational compatibility. The choice of the components for the design was made open to nonproprietary vendors' equipment to evade integration-related and operational issues.

Step 3: Development of the Open ICT Ecosystem

The ICT ecosystem is a new concept that suggests ICT resources as a part of a comprehensive network platform (Simon et al., 2018). An Open ICT Ecosystem (OIE) refers to an ICT system architecture that is open to allow for the interaction, operation, and synchronization of both homogenous and heterogeneous ICT devices within a network. The open ICT ecosystem idea enabled the integration of the various hardware and software components of the ecosystem, and allow their inter-operability to create a robust resourceful network design, for optimal resource utilization.

The creation of the open ICT ecosystem began with gathering a list of ICT infrastructure, resource, and services that are required for the network design, and implementation. Future advancements and possible expansion of the network were also considered. The open ICT ecosystem was able to resolve compatibility issues by unifying the operational systems of the different components, by accommodating every available resourceful infrastructure to benefit equipment vendors and the network users. The open ICT ecosystem provided a limitless device-operational environment, and optimal resources utilization via a campus WLAN.

Step 4: Design Requirements for the 3-Tier Open ICT Ecosystem Network Architecture

A 3-Tier (Hierarchical/Tree) Network design was developed according to the flowchart in Figure1. The design comprised of the core, distribution and access layers. The design began with finding the major players for the network ecosystem, which needed to form the layers of the open ICT ecosystem. The components requirement for the open ICT ecosystem included, the Internet service providers, ICT infrastructures viz; servers, routers, converters, printers, and etc.

The following basic hardware were used in Building the 3 Tier Open ICT ecosystem Design:

1. Mikrotik cloud core multilayer router (CCR1036-12G 4s-EM).
2. Smart multilayer cloud router switch (CRS125-24G-1S-IN complete 1sfp 24 port 10/100/1000 layer 3 switch (optional).
3. Mikrotik Firewall OS, Intrusion Detection and Prevention Systems (IDS and IPS).
4. 8 core Fiber optic link for point to points secured links.
5. Netlink Ethernet media converter; 10/100/1000Base-T to 1000 Base-SX/LX.
6. Mikrotik 5.8 SXT GHz distribution radio device for Point to Point link on layer 2.
7. Mikrotik 2.4 GHz router board device for local remote access points.
8. Ubiquitous or TP Link 4G Wireless N router (to serve for access points in the access layer).

VIII. Development of the 3TOIEND

Building the 3TOINED was carried out in three phases viz;

Step 4.1: A. Creation of a Robust Hybrid Core Layer (RFCL) Design

The core layer design started with creating a robust, hybrid, and open ICT ecosystem at the Network Operation Center (NOC). The layer consist of an integrated Internet access service sources using a Mikrotik cloud router *CCR1036-12G-4s-EM*, with inbuilt

firewall OS. The router was connected to a multilayer switch (*Mikrotik CRS125-24G-1S-IN complete 1SFP 24 Port 10/100/1000 layer 3 switch)* to produce one common output of a hybrid internet access backbone. Further integration of more ICT facilities such as, subnetworks, servers, network printer, and etc. followed. The core layer used a hybrid link comprising of a 5.8GHz Mikrotik wireless radio and a fibre optic link, which terminated with two 10/100/1000Base-T to 1000 Base-SX/LX Ethernet fibre media converter at both ends, to provide connections from the core layer to the distribution layer.

Although, both links were merged by the router to produce one output, the links remained independent of one another, and sustained their peculiar features, so that, they can be used independently if the need arises. The design and configuration of the core layer can be likened to a smart phone with multiple SIM cards, which are used for Internet access services for the phone, for seamless switching between alternatives SIM cards to constantly provide a consistent Internet access services for the phone, via any of the SIM cards.

At the end of the core layer design, the *Proof of Concept (POC)* design testing was carried out on the design, to confirm viability of the design before proceeding to the design the next layer.

The figure 2, presents the three layers of the design architecture and the WLAN components.

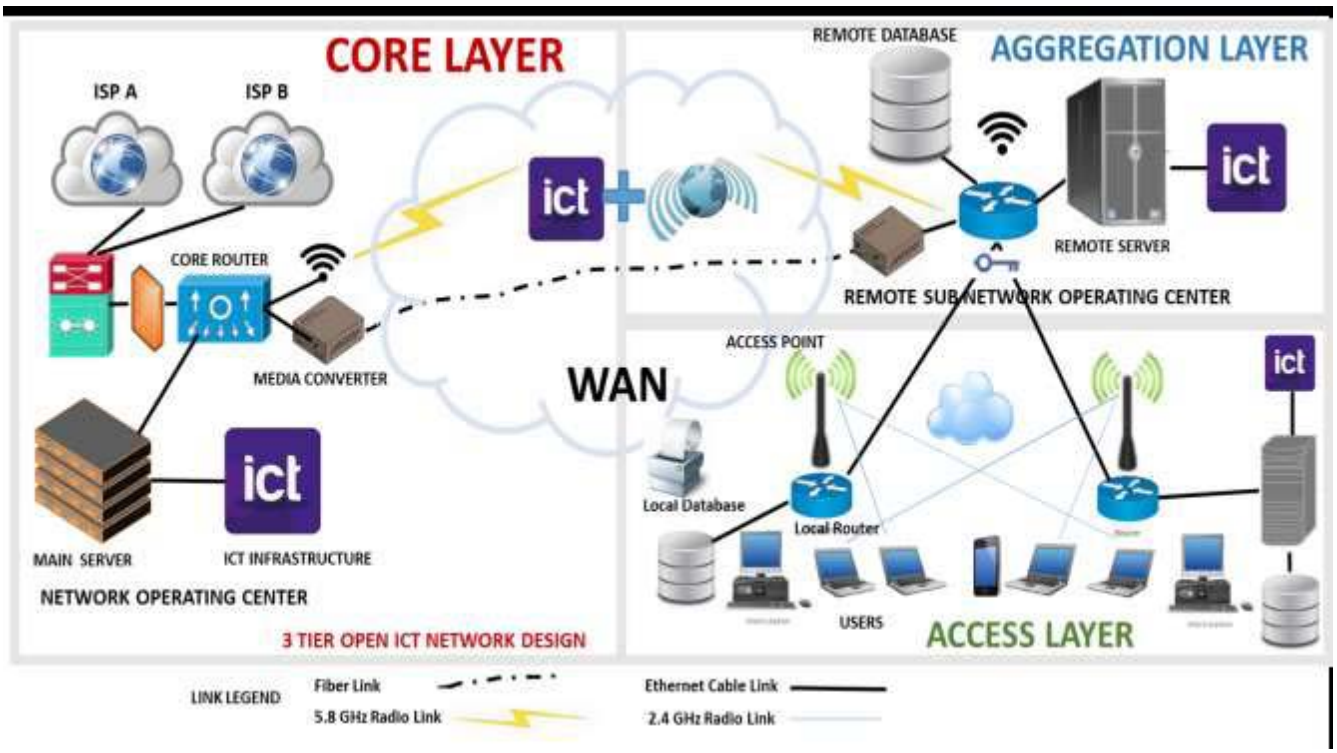

Figure 2: The 3 Tier Open ICT Network Design layout

Step 5.2: The Aggregation/ Distribution Layer Design and Implementation

The aggregation layer, which is also referred to as, the distribution layer, interface between the core layer and the access layer. The aggregation layer serves as the critical point for control and application services of the WLAN. The layer's design began with the creation of Remote Sub Network Operations Centres (RSNOC), and afterwards was linked the core layer via a hybrid link of fiber (Ethernet media converter of 10/100/1000Base-T to 1000 Base-SX/LX at both ends), and a wireless (Mikrotik 5.8GHz SXT) radio device. The duo links terminated at a router (Mikrotik router board) located at the RSNOC.

The Functional features of the Aggregation Layer: This layer was designed to perform among others, the following prominent functions:

1. *The distribution of local campus ICT infrastructure to the access layer:* The RSNOCs links the open ICT ecosystem infrastructure of the core center, to the meshed access layer, for the distribution of the integrated hardware and software contents of the core layer.
2. *The Integration of Remote Local Hadware and Software Resources:* The RSNOC serve as the integration center for ICT infrastructure and resources at remote campus locations. The aim the RSNOCs was to ensure optimal ICT resource integration, and utilization via a campus network, and to evade link redundancy. The Local contents at the RSNOCs included; faculty and departmental e-libraries, expensive hardware; such as network printers and scanners; local databases containing digitized student forms, lecture materials; and other

similar materials. The integration of the ICT resources at the RSNOCs, yielded a robust mesh of resourceful hardware and soft contents that is accessible to users via a campus WLAN, and increased the utilization of the robust open ICT ecosystem shown in figure 2.

3. *The Utilization of the WLAN login page for news and information dissemination:* The design used the local network Security Set Identifier (SSID) of the RSNOC access points, to provide the *local NOC's name alongside a short message.* The network login page layout was designed to use a part of the page layout, for posting institution's news headlines, and other vital information as shown in figure 3. The content of the login page is subject to updates, regularly from the server at the NOC. Network-users will be able to access News, information, and campus updates by merely connecting to the network Wi-Fi's, SSID login page, and get the necessary information, without logging into WLAN, or using the WLAN's bandwidth data.
4. *The Provision Remote Second Security Layer (RSSL):* The distribution layer was designed to provide a second layer security at the RSNOCs, by using the Mikrotik router board's operating system. The Mikrotik router board's configuration, provided monitoring, filtering, and firewall services as shown Figure 2.
5. *The Inclusion of Student-Users' Presence on the Network:* Students comprises a larger population of any tertiary institution, yet their participation is often, not properly captured, or accommodated in most campus network designs. Thus, denying students from extensive use of campus networks, and the underlying ICT infrastructures which are leveraged therein. In order to improve student-users' presence in a campus WLAN, and to improve resource and network utilization, the design included the following features at the distribution layer:

i. *The Creation of Students' Layer in the Network Design:* The student layer of the design made it possible, for the integration of student-related services such as; link for uploaded and download of local forms for students' using Wi-Fi-enabled devices, CBT solutions facilities, and games into the ICT ecosystem, which is accessible via the campus network. Adding students' presence-oriented services, and needs in the design caused more local network traffic, and higher utilization of the robust ICT infrastructure available on the network.

ii. *The Integration of EGranary/ Digital Library Server(s) to the Design:* The design also gave room for the integration of robust local research infrastructural tools, like the EGranary server (Internet in a Box), digital library servers, and database information systems servers, to further enrich the open ICT ecosystem design. The provision of these robust academic materials, which are needed, and accessible by students via a local campus WLAN in the design, grossly reduced the overdependence on the Internet access service of a network. Also, the integration of aforementioned resources to the ecosystem, caused notable reduction of Internet traffic, but increased the local WLAN traffic, because the resources provided alternatives to the Internet resources needed. Furthermore, the integrations grossly reduced bandwidth usage and wastage, to proffer more utilization of the local resources present in a local campus network.

Figure 3: The Wi-Fi Login Page Designed to Hold News Headlines

a) Implementation of the Mesh Access Layer Design

At the periphery of every network lies the access layer. The access layer provide the initial connectivity between the user devices and, a network, through the aggregation layer. The designof the access layer comprise of a mesh of Mikrotik router boards, wireless routers, and other ubiquity devices for delivering access to local open ICT ecosystem contents to users, via a local campus network. The meshed access layer serve the critical-mission clients, general-purpose clients, and the mobile clients, by cable, or wireless connections. The design made provision for the further integration of

available services, and ICT infrastructure to a network. Also, the design made provisions for free Wi-Fi clusters at different locations within a campus. The students can utilize a filtered Internet access remotely via a campus WLAN, at the designate free-Internet access locations. The access layer design and implementation were successfully tested using the *network ready for use* (NRFU) testing tools to mark the end of the 3 Tier Open ICT Ecosystem Network Design (3TOIEND), and implementation.

b) Network Security Design

A network design will be incomplete without the development of basic security measures put in place to prevent intrusion and consequent failure of the design. The OSI layer 3 security devices, and software were used to checkmate network intrusion and detection, and to enhance intrusion prevention for the defence, and optimization of the 3TOIEND. The Mikrotik hardware and software, were properly configure to monitor all connections and their content. Also, operational and maintenance policies for the users in a campus network environment were listed, and formulated. Institution's management are expected to provide the access rights and privileges applicable to their network to curb unauthorised access and usage.

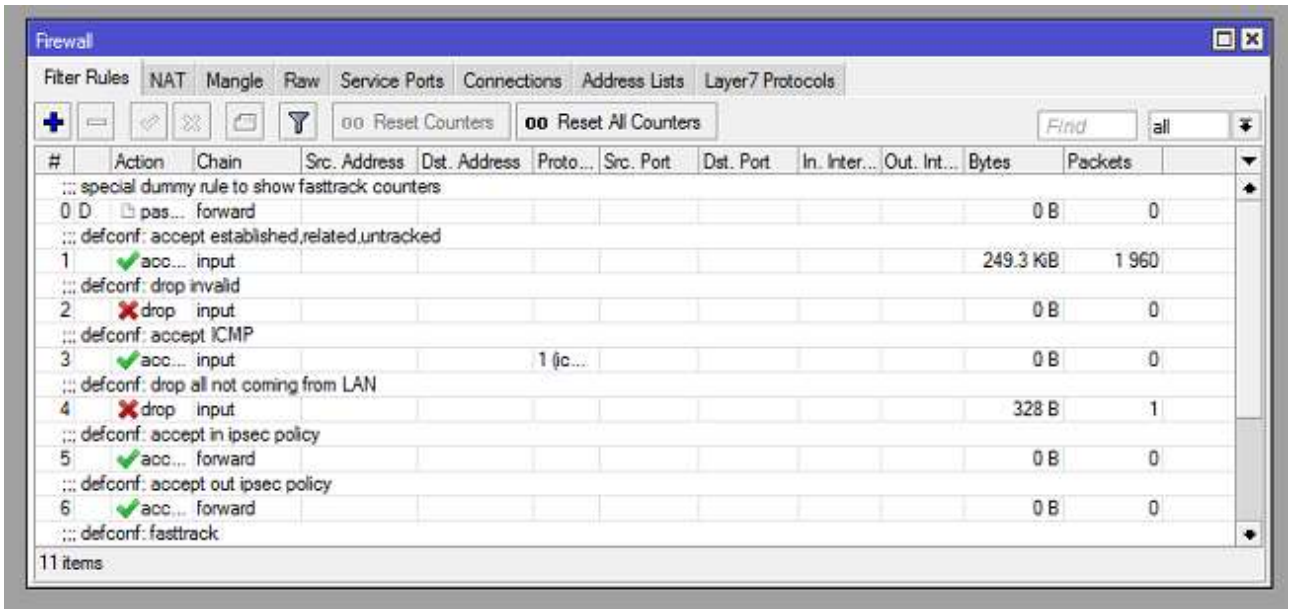

Figure 4: Mikrotik firewall setup

Step 6: Campus Wlan Management Policies Development

Efficient utilization and optimal throughput are the required goals of every network success. Successfully managing any network to deliver, will the required design goals to yield optimum throughput implies setting good management policies as stated in network design algorithm.

c) Curbing Security and Safety Concerns at the Access Layer

Network security architects (NSA) are responsible for analysing network data and systems to select the most appropriate control mechanism for the security required in a network. A security architect will also have a hand in selecting software and hardware usage in a network to control system. Finally, for adequate maintenance of a network performance and utilization, the following concerns must be considered for address while formulating the network design policy:

- ✓ Access point levels security
- ✓ Network usage monitoring
- ✓ Scale usage of a network
- ✓ Checkmate authentic user qualified to access a network
- ✓ Checkmate intrusion into to a network
- ✓ Monitor abuse of bandwidth
- ✓ Monitor abuse of privileges
- ✓ ICT resource access control

IX. Benefits of the 3TIER Open ICT Ecosystem Network Design (3TOIEND)

After the design and implementation, the 3TOIEND design was able to render the following services among others:

a. Provision VOIP (Voice over Internet Protocol) – VOIP service can be configured to run on the network, so that users can leverage the WLAN to make free calls through a network with, or without connecting to the Internet. Optimal performance of this feature require some additional devices and configurations.

b. The integration of local resources - The 3TOIEND WLAN design offer very robust services beyond the

distribution of Internet access. The Implementation an open ICT ecosystem design, positions a network to integrate every local ICT infrastructure in a campus WLAN for adequate utilization.

c. Ease of network expansion- The design is intelligent enough to accommodate future network expansion of any extent, and at any required time. An expansion may include but not be limited to;

1. An Indoor and outdoor equipment expansions
2. The introduction of a new ICT infrastructure
3. A User-based expansion
4. A network device expansion
5. The Internet bandwidth increments
6. A general network expansion

d. On the Go News flash Service on the Login Page: This peculiar feature of having a campus news update on login page, showcases the robust features of the 3TOIEND. Connecting to a Wi-Fi is usually automatic on every Wi-Fi-enabled device. Thus, using a Wi-Fi login page to provide a campus' news updates, means that, users can get a campus news, and information, without having to waste a ampus' Internet bandwidth, but by just securing a connection to a campus WLAN, .

Step 7: Testing the Design and Results Discussion

Network design testing is essential, because networks are hard to build correctly. A network design may appear subtle, whereas, there may be bugs therein that may require intermittent attention. A network design testing as an investigation, conducted to provide network Stakeholder (s) with information about the quality of the product, service, or design of a proposed network. A network design testing is similar to a software design testing and aims to ensure that a network's implementation, works as designed. The flowchart in figure 1, included design testing after each step of the design. (CISCO, 2012) Recommended two types of network design tests viz; firstly, a proof of concept testing (POC) is required, to ensure that a newly invented design, or to test if a design technology will work as expected in the context of the design. Afterwards, the Network ready for use testing (NRFU) method follows immediately. The NRFU is typically executed at the last step of a design, to certify that the design is ready to carry expected traffic.

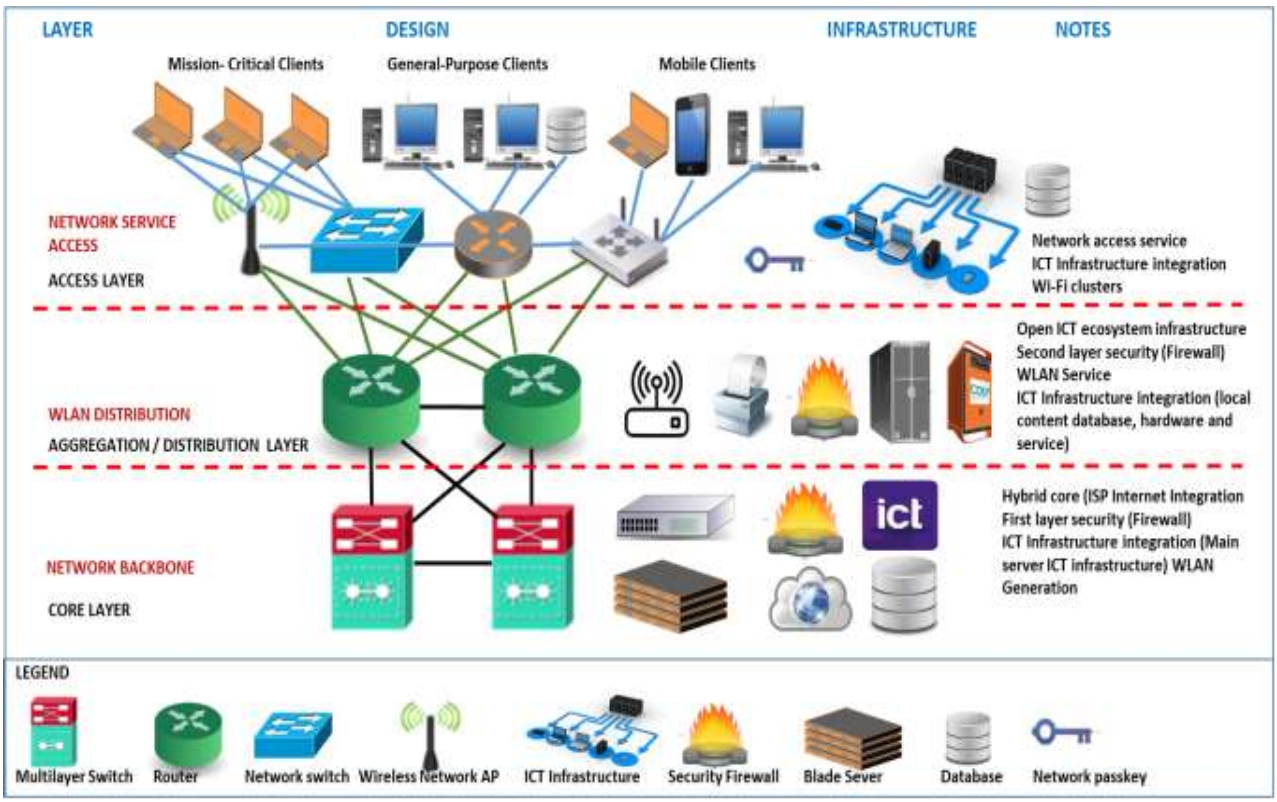

Figure 5: The 3-Tier Open ICT Ecosystem Network Design (3TOIEND) Model

a) Design Testing Results

The design testing results are as follows:

i. A verification that the design meets desired technical goals.

ii. The validation of the WAN and LAN technology, and the device selection for the design.

iii. The identification of any possible problem area or connectivity issues.

Table 1: 3TOIEND testing parameters

Comparison tests were carried out using various parameters to compare the 3TOIEND, with the previously existing campus WLAN design at Nnamdi Azikiwe University Awka campus. The table below, shows the parametric data used to test the new 3TOIEND against the previously existing WLAN design at the Nnamdi Azikiwe University Awka. The results of the tests are presented in the column chart of figure 6.

Table 1

Design Name	Resource Coverage	Resource Utilization	Network Stability	Student's Presence	Network Redundancy
Other Designs	85	70	65	83	25
3TOIEND	40	45	50	38	55

X. 4.0 Discussion of Test Results and Design Validation

As shown in table 1, different types of tests were carried out to test the efficacy of the 3TOIEND. A POC testing was carried out, to certify the design plan, at the same time, the NRFU test was performed to test the bandwidth, and the open ICT ecosystem's infrastructures utilization via a network. The 3TOIEND model yielded amazing test results. Statistical tools were further used to carry out comparison tests, between the new model's performance, and an existing model's performance. The test applied the following parameters; resource coverage, resources utilization, network stability, student's presence, and network redundancy. Eventually, the comparison test results, which are summarized in the bar chart of figure 6 validated the 3TOIEND model.

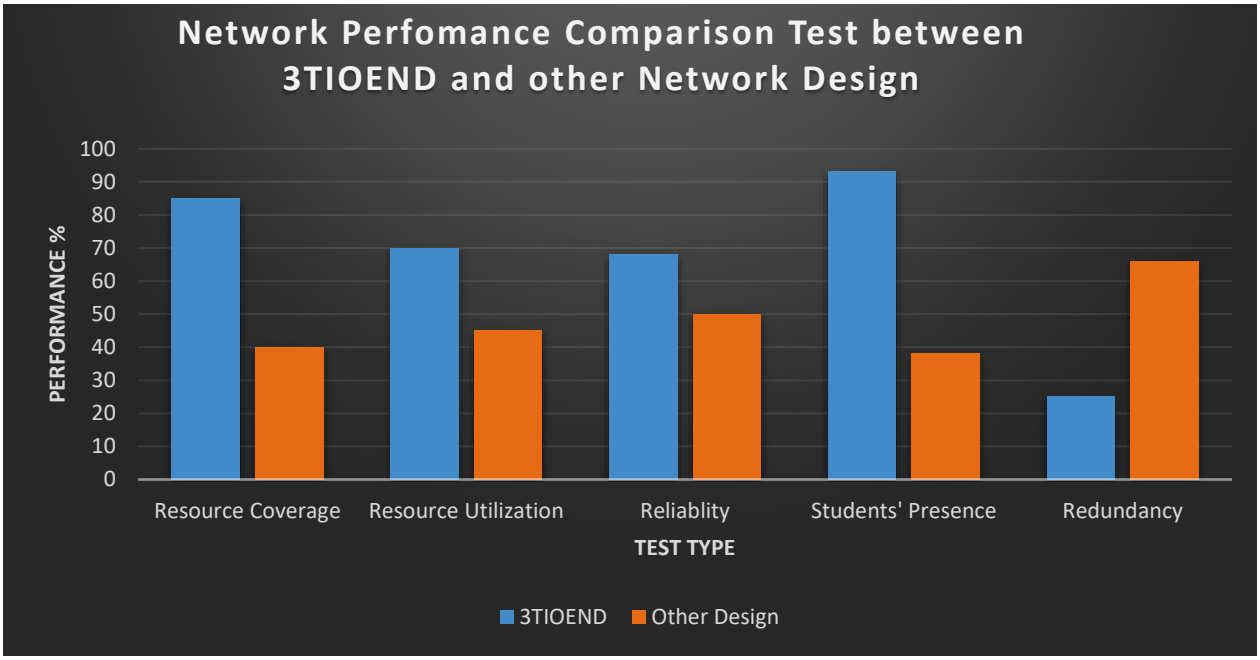

Figure 6: Network Performance Comparison Test between 3TOIEND and other Network Design

XI. Conclusion

In conclusion, the 3-Tier Open ICT Ecosystem Network Design (3TOIEND) is intelligent, robust, and easy to implement. The 3TOIEND was implemented at Nnamdi Azikiwe University Awka, Anambra State, Nigeria, and the intelligent design proffered an enhanced, value-added, and sustainable network, which optimally, utilized the ICT resources available in the Open ICT Ecosystem, via the Network. The implementation of the design at Nnamdi Azikiwe University Awka, was financed, through a grant awarded, by the *TETFund INSTITUTION BASED RESEARCH (IBR) GRANT,* approved with the reference number, *(TETF/DR& D/CE/UNI/AWKA/RG/2022/VOL.1 on July 8th 2022,* in favour of the author.

Also, the implementation of the design at Nnamdi Azikiwe University, enable the campus WLAN, to satisfactorily utilized the available bandwidth

subscription in the campus ICT ecosystem, and created a very noticeable network optimization; reduced bandwidth wastage; evaded bandwidth abuse; drastic reduction in network and resource redundancy; optimal utilization of available local ICT resources via the campus WLAN; a drastic reduction in the overdependence on Internet; and increased network reliability.

References Références Referencias

1. Adeleke Dare Samuel and Emeaghara Evelyn Nkechi, (2016). Relationship between information Literacy and Use of Electronic Information Resources by Post graduate Students of the University of Ibadan, DigitalCommons @University of Nebraska – Lincoln, ResearchGate.
2. Ananata Hassidim, D. Raz, M Segalov, A Shqed (2013). Network Utilization - an overview | ScienceDirect Topics: The flow view proceedings IEEE INFOCOM, pg 1429 1437
3. Aisha Siddquah, Zeema Salim, (2017). The ICT Facilities, Skills, Usage, and the Problems Faced by the Student of Higher Education, EURASIA Journal of Mathematics Science and Technology Education. ISSN: 1305-8223 (online) 1305-8215 (print). DOI: 10.12973/Eurasia.2017.00977a. pp 4921 4931
4. Cisco (2012). Cisco Unified Communications System 8.x SRND. Protocol, Retrieved from http://www.Cisco.com/c/en/us/td/docs/bandwidthmanagement/www.Cisco.com/web/bandwidth/ac143/achived s_issue/ipj.12html
5. Cisco (2016). Cisco Unified Communications System 8.x SRND. Call Admission Protocol, retrieved from http://www.Cisco.com/c/en/us/t/docs/voice_ip_comm/cucm/srn/8x/uc8x/cac.html.
6. Cisco Press (2020). Network ready for use test, the Cisco learning Network, www. Learning-networkcisco.com retrieved on 20th January, 2022
7. Kassim M., Ismail M., Jumari K., and Yusof J. (2013). International Scholarly and scientific Research & Innovation 6(2) World Academy of Science, Engineering and Technology 6, 02-22, 332 -339
8. Ketan Mayer-Patel (2002). Readings in Multimedia Computing and Networking, Network Utilization - an overview | ScienceDirect Topics
9. Lougee_heimer (2003). The Common Optimization Interface for Operations Research: Prmoting open-source software in the operation community, IBM Journal of research and development, Vol 47,
10. Mahajan R, David W. and Thomas A., (2005).Negotiation-based routing between neighboring ISPs, proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation–Volume 2. 2005.
11. Obidike, N, A. & Mole A.J.C., (2015). Assessment of Electronic Collection Development in Nigeria University Library, http://digital-commons.unl.edu/libphilprac/1326.
12. Okon Ani (2010). Internet Access and Use: A Study of Undergraduate Students in Three Nigerian Universities, The Electronic Library. 28 (4): 555-567. DOI 10.1108/0264047101165373
13. Projectstore (2021) Computer Engineering project topics and materials. pg 449 www.project-store.com.ng retrieved on 15th January, 2021
14. Robert J. Shimonski, Yuri Gordienko (2002). Sniffer Pro Network Optimization and Troubleshooting Handbook. Pg 20
15. Simon Petterson Nguyen, Masoud Hussein Mahundi. (2018). The dynamics of national ICT ecosystems. The electronic journal of information systems in developing countries, vol 85, Issue 1e12058
16. Stefano Gridlli (2016). Network Engineering, Why network testing is important. http://net-beez.netretrieved on 1st February, 2022
17. Suresh (2016). Wireless Network testing, Cisco Press, Network Design Testing, https://www.Coursegero.com/file/16307890/Network-Design Testing/ retrieved on 27th January, 2022
18. Valerie Burgess, (2018). Students face issues with WiFi connection, Daily Trojan. Retrieved from www.dailytrojan.com on December 12 2021.
19. XiPeng Xiao. (2008). Network Utilization. Technical, Commercial and Regulatory Challenges of QoS.
20. Network Utilization - an overview | ScienceDirect Topics. Retrieved on 22nd January, 2022

Global Journals Guidelines Handbook 2022

www.GlobalJournals.org

Memberships

Fellows/Associates of Computer Science Research Council

FCSRC/ACSRC Memberships

Introduction

FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

Fellow of Computer Science Research Council

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.

Benefit

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Fellows can use the honored title of membership. The “FCSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 60% of sales proceeds

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

Access to Editorial Board

Become a member of the Editorial Board

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career | Credibility | Exclusive | Reputation

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LOR and Laser-Momento

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Associates can use the honored title of membership. The “ACSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 30-40% of sales proceeds

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.

Associate	Fellow	Research Group	Basic
$4800 **lifetime designation**	$6800 **lifetime designation**	$12500.00 **organizational**	APC **per article**
Certificate, LoR and Momento **2** discounted publishing/year **Gradation** of Research **10** research contacts/day **1 GB** Cloud Storage **GJ** Community Access	**Certificate,** LoR and Momento **Unlimited** discounted publishing/year **Gradation** of Research **Unlimited** research contacts/day **5 GB** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**Certificates,** LoRs and Momentos **Unlimited** free publishing/year **Gradation** of Research **Unlimited** research contacts/day **Unlimited** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**GJ** Community Access

Preferred Author Guidelines

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from https://globaljournals.org/Template.zip

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

Before and during Submission

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct,* along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s') names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

Policy on Plagiarism

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures

- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

Authorship Policies

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

Preparing your Manuscript

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.

Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11'", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

a) A title which should be relevant to the theme of the paper.
b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
d) An introduction, giving fundamental background objectives.
e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
f) Results which should be presented concisely by well-designed tables and figures.
g) Suitable statistical data should also be given.
h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
j) There should be brief acknowledgments.
k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.

Format Structure

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, “What words would a source have to include to be truly valuable in a research paper?” Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.

Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

Preparation of Eletronic Figures for Publication

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

Tips for writing a good quality Computer Science Research Paper

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.

6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10.Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.

20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.

Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.

The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."

Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

The Administration Rules

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350