Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 3s (2025)

Advances in Nonlinear Variational Analysis for Cyber Criminal Detection via Smartphone Communication Patterns

K. Swetha[1], *Sivaraman K[2]*

[1,1]Department of CSE, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamilnadu,India.

[1]swetha281189@gmail.com (Corresponding author)

Article History:

Received: *27-09-2024*

Revised: *20-11-2024*

Accepted: *29-11-2024*

Abstract:

The increasing dependence on digital communication needs sophisticated ways for identifying cybercrime, particularly on mobile devices. This paper provides a comprehensive methodology that uses Natural Language Processing (NLP) tools to analyze communication patterns and detect suspicious activity. To generate communication patterns, text is preprocessed, vectorized with TF-IDF, metadata is extracted, and clustering is performed. Anomalies are identified by calculating deviations from these patterns, and suspicious signals are reported for further study. The proposed strategy outperformed baseline techniques in terms of accuracy, precision, recall, and the F1 score. Statistical study supported the method's performance, highlighting its potential as a significant tool for improving cybersecurity measures. In addition, the methodology employs Nonlinear Variational Inequalities to improve anomaly detection in communication patterns, addressing deviations more robustly within a cybersecurity framework

Keywords: Non Variational Inewqualities, Digital Communication, mobile devices, Communication patterns, Natural Language Processing, Cybersecurity, Cyber Crimes

1. Introduction

Smartphones have become an essential part of our everyday lives in the current digital era, enabling a wide range of communications through several channels like texts, emails, social media, and instant messaging apps. Although these developments have facilitated more effective and convenient communication, they have also given cybercriminals additional channels through which to carry out their nefarious operations. The escalation of cybercrime has mandated the creation of advanced instruments and methodologies to detect and mitigate these hazards. Using Natural Language Processing (NLP) to examine communication patterns and recognize possible cybercriminals based on their online activities is one such innovative strategy.

The study of how computers and human language interact is the subject of the artificial intelligence discipline known as natural language processing, or NLP. Large volumes of text data can be sorted through to find hidden patterns, attitudes, and abnormalities that might point to criminal activity by utilizing NLP algorithms. This technique can be very useful in deciphering the complex and frequently obscure language that hackers employ to avoid being discovered. It is possible to train sophisticated

natural language processing (NLP) models to identify particular words, phrases, and communication patterns that are frequently connected to illegal activity. This allows for a more precise and prompt detection of possible threats.

A number of crucial processes are involved in applying NLP approaches to the identification of cybercriminals: feature extraction, preprocessing, data collecting, and model training. To ensure consistency, communication records from smartphones are first collected into big databases that are subsequently cleaned and normalized. The following stage entails taking the text data and extracting pertinent elements like word frequency, grammatical structures, and semantic meaning. Machine learning models that can differentiate between regular and suspect communication patterns are then trained using these features. It is feasible to improve these models' predicted accuracy and make them more responsive to changing cybercriminal strategies by regularly incorporating fresh data into them.

Law enforcement organizations and cybersecurity experts now have a potent tool in their arsenal to combat cybercrime thanks to the incorporation of NLP techniques into cybersecurity frameworks. The time and resources needed to recognize and look into such risks can be greatly decreased by using natural language processing (NLP) to automate the study of communication patterns. Furthermore, real-time communication analysis can operate as an early warning system, enabling law enforcement to take preventative measures before a cybercriminal can carry out their plans. The continuous improvement and development of NLP-based analysis tools will be essential to preserving the security and integrity of our digital communications as cyber threats continue to change.

Adversial text analysis [3] provides a thorough examination of the defences and attacks that adversaries employ in social media applications, utilizing machine learning (ML) and natural language processing (NLP). It encompasses critical topics, including sentiment analysis, hate speech, misinformation detection, clickbait and spam, and falsehoods. The paper details the taxonomy of adversarial techniques and their corresponding defense mechanisms, emphasizes significant research challenges, and delineates future research directions that are specific to social media NLP applications.

G. Maria Jones et.al [4] proposed a method for identifying suspect drug-dealing patterns in mobile devices by utilizing forensic and NLP techniques. The study evaluates performance by implementing Machine Learning algorithms on an original dataset. Logistic Regression achieves 95% accuracy with count vectors, while BiLSTM achieves 95% accuracy with TFIDF. The significance of mobile forensic data analysis in the discovery of criminal activities for legal proceedings is emphasized by the methodology. The potential of this method to improve mobile cyber-crime investigations is underscored by its effectiveness.

Dongming Sun et al [5] introduces a digital investigation platform that is based on natural language processing (NLP). The platform includes phases for data acquisition, vectorization, feature selection, and classifier generation and evaluation. In comparison to Log Analysis, our methodology exhibits superior performance when implemented on a real-world dataset.

Forensic Taxonomy study [6] analyzes 30 popular Android social apps for forensic artifacts such contact lists, message chronology, and timestamps. Facebook token strings link account IDs and access user-entered data, according to the research. Based on these findings, a two-dimensional taxonomy of social app forensic artifacts is developed, categorizing applications in one dimension and objects in

the other. The study identified investigator-useful data fragments, proposed a thorough taxonomy for Android social apps, and compared these findings with other Android app forensic taxonomies.

Al Mutawa et al. [7] forensically analyzed BlackBerry, iPhone, and Android Facebook, Twitter, and MySpace apps. They used these programs as normal users, took forensically sound logical photographs of the devices, and manually evaluated them. The forensic recovery recovered user and friend data, including contact details and profiles, from iPhones and Android phones but not BlackBerry devices.

A forensic investigation [8] of iOS social apps by Levinson et al. [8] found that third-party apps can supply significant time and location data for evidence.

Large data [9] from IoT, cybersecurity, mobile devices, companies, social media, and healthcare in the Fourth Industrial Revolution (4IR) call for intelligent analysis using artificial intelligence, particularly machine learning (ML). Reviewing several ML algorithms—including supervised, unsupervised, semi-supervised, reinforcement learning, and deep learning. Serving as a reference for academics, business leaders, and decision-makers, the report also tackles issues and outlines future study paths. Important contributions include specifying the extent of the study, giving a summary of ML techniques, talking on their practical relevance, and stressing possible future routes of research.

Social media networks [10] have transformed communication but also encouraged negative practices such hate speech, false information, cyberbullying, and radical propaganda. Social media forensics have evolved out of these problems to assist law enforcement and investigators in handling cybercrimes. Focusing on natural language processing (NLP) for recognizing extremist ideas, spotting online bullying, and investigating false profiles, this paper reviews current studies on applying artificial intelligence in social media forensics. It also looks at how Graph Neural Networks (GNNs) might be used in forensic social network modeling. Emphasizing the need of responsible and efficient AI deployment in guaranteeing safety and justice on social media platforms, the paper addresses major problems and future directions of research.

Rapid urbanization[11] and effective resource management are two ways that smart cities improve quality of life, with information and communication technology (ICT) playing a critical role. The underused potential of natural language processing (NLP) in streamlining ICT procedures for smart cities is examined in this research. It looks at the architecture of natural language processing (NLP), its applications in business, community, media, research, and education, as well as its open issues. Classifying cutting edge NLP techniques, demonstrating how NLP gets over obstacles in smart cities, and talking about future research paths are some of the key contributions. The importance of NLP in developing smart city technology and solutions is highlighted by this paper.

Digital images[12] are very important in this information age, but they can be accessed and changed without permission, which is bad for privacy, public safety, and social order. Because of this, image forensics has become an important area of study in the field of multimedia information security. This study looks at how new picture forensic techniques that use deep learning work better than older methods. It sorts methods into two groups: passive forensics and active forensics. For passive forensics, it looks at frameworks, evaluation metrics, datasets, techniques for finding fakes, as well as their success and the pros and cons of current methods. For active forensics, it looks at strong picture

watermarking methods, along with their evaluation criteria and frameworks, and different attack types. At the end of the survey, there are directions for future research and useful field ideas.

The related works demonstrate how machine learning and natural language processing are combined in a variety of forensic and security applications. Research has looked into forensic taxonomies, digital investigation platforms, adversarial text analysis, and mobile forensic data analysis. These studies have shown how effective these techniques are at spotting trends in cybercrime. To be more precise, the research on questionable drug-dealing habits, picture forensics, and social media app forensics serves as a basis for examining smartphone communication trends. By utilizing these discoveries, the proposed project "Advanced Communication Patterns Analysis Using NLP Techniques for Cyber Criminal Identification in Smartphones" seeks to improve mobile forensics, advance digital security, and create reliable techniques for identifying cybercriminal activity.

2. Objectives

This work aims to create a novel approach using Nonlinear Variational Analysis and Natural Language Processing (NLP) for cybercrime detection via smartphone communication patterns. Given the explosive expansion of digital communication channels—including SMS, emails, and social media—this study tackles the need for improved cybersecurity strategies to identify and minimize cyber dangers, particularly in mobile device communication, given their fast expansion. Preprocessing text data, vectorizing it using TF-IDF, and extracting metadata help the suggested approach to spot deviations from normal communication patterns and point up questionable behavior. This method provides a strong cybersecurity system to instantly recognize cyberthreats and improves anomaly detection accuracy, so helping to prevent crime actively.

The main goal of this work is to enhance mobile forensic capacity by means of modern NLP methods coupled with machine learning models to differentiate between benign and suspicious communications on smartphones. By use of feature extraction and clustering methods, the study generates communication profiles and detects aberrant activity suggestive of cybercrime. The study shows the efficiency of the suggested methodology in attaining high accuracy, precision, recall, and F1 scores by means of statistical analysis and performance comparisons with baseline techniques. This study emphasizes the need of including artificial intelligence-driven tools in cybersecurity systems, thereby offering a great help for law enforcement and cybersecurity professionals to improve digital communication security.

Methods

The proposed approach describes a comprehensive strategy to analyzing cyber criminal activity using mobile phone data, utilizing forensic tools and NLP approaches. The illustration in Figure 1 shows a flowchart that describes a Proposed methodology process for employing Natural Language Processing (NLP) and digital forensic technologies to identify cyber anomalies. Mobile data is first entered into the system and is subsequently extracted utilizing digital forensic tool in order to recover pertinent data. In order to get the retrieved data ready for more analysis, feature extraction and vectorization are applied. The methodology's primary component is the application of a suggested strategy that makes use of NLP tools to examine the data. A procedure using algorithm for detecting cyber criminials is performed based on anomaly detection method to find any unusual or suspicious activity. The

procedure concludes with an assessment of the detected activity is suspicious or not suspicious offering a clear path for cybersecurity measure decision-making. The convergence of feature analysis, forensic data extraction, and sophisticated NLP techniques to improve mobile cybersecurity is succinctly illustrated in this flowchart.

Gathering Information Using Forensic Devices

Forensic Acquisition:

To obtain communication data (SMS, MMS, emails, social media, and messaging app data) from mobile devices in a forensically sound manner, use forensic technologies like Cellebrite, XRY, or UFED.

Data Accuracy: Maintaining a chain of custody and employing hashing algorithms (e.g., MD5, SHA-256) to confirm the legitimacy of the data can help to ensure the integrity of the acquired information.

Preprocessing Data

Text Rewriting: Eliminate extraneous elements such as HTML tags and emoticons, and normalize the text by changing its case.

Text is divided into tokens (words or phrases) using tokenization.

Remove Common Stop terms: Get rid of terms like "and," "the," and "is."

Reduce words to their most basic form (e.g., "running" to "run") using stemming and lemmatization.

Extraction of Features

Vectorization: Using methods such as TF-IDF or word embeddings (Word2Vec, GloVe), transform text into numerical representations.

Features for Metadata: Retrieve metadata including sender and recipient IDs, timestamps, message lengths, and communication frequency.

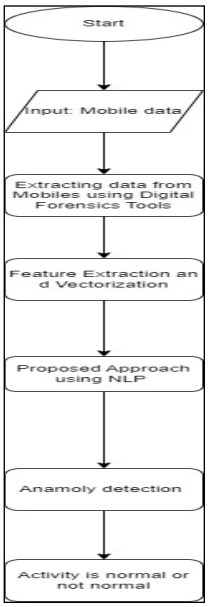

Fig. 1 Suspicious activity detection using Proposed Methodology

Algorithm for Detecting Cyber Criminal Activities Using NLP Techniques

1. Let the Pre-processed text is given as:

P(M)={stem(lemmatize(removeStopWords(tokenize(removeIrrelevantInfo(lowercase}(M))))}

2. Text to be vectorized: V=TFIDF(P(M))

3. Metadata extraction: F(M)={t,s,r,l,f}

4. Detect Anomalies: For new message N

Preprocess P(N), vectorize V(N), extract metadata , F(N)

5. Calculate-deviation D=Deviation(V(N),F(N),C)

6. If D>θ, flag N as suspicious

7. Generate Alerts: Notify investigators of suspicious messages

Case Study 1:

1. **Preprocess Text:**

o Message: "Hey, let's meet at the cafe at 5pm."

o Preprocessed Text: "hey let meet at cafe at 5pm"

2. **Vectorize Text:**

o Preprocessed Text: "hey let meet at cafe at 5pm"

o Numerical Vectors: V(M1)=[0.1,0.3,0.0,0.2,0.0,0.1,0.3,0.0]

3. **Metadata Extraction:**

o Metadata:
timestamp:"2024−07−1716:00:00",sender:"John",receiver:"Alice",length:7,frequency:10

4. **Pattern Generation and Clustering**:

o Identified Communication Patterns: Clusters of numerical vectors and metadata indicating common communication behaviors.

5. **Anomaly Detection**:

o New Message: "Hey, did you complete the secret transaction?"

o Preprocessed Text: "hey did you complete the secret transaction"

o Numerical Vectors: V(N)=[0.1,0.2,0.0,0.1,0.3,0.0,0.2,0.1]

o Metadata:
{timestamp:"2024−07−1718:45:00",sender:"Unknown",receiver:"Unknown",length:7,frequency:1}

o Deviation Calculation: D=0.75

o Threshold: θ=0.5

o Since D>θ the message is flagged as suspicious.

6. **Generate Alerts:**

o Notify investigators with the details of the suspicious message: "Hey, did you complete the secret transaction?" along with its metadata.

By following these steps, the algorithm effectively identifies abnormal behavior in mobile communications, aiding in the detection of potential cyber-criminal activities.

Case Study 2:

1. **Suspicious Message Details:**

o Message: "Make sure the transaction details are erased immediately."

o Metadata: {timestamp: "2024-07-18 12:30:00", sender: "Unknown", receiver: "Unknown", length: 7, frequency: 2}

1. **Preprocess Text:**

o Message: "The package has been delivered to the specified location."

o Preprocessed Text: "package delivered specified location"

2. **Vectorize Text:**

o Preprocessed Text: "package delivered specified location"

o Numerical Vectors: V(M1)=[0.0,0.2,0.1,0.4,0.0,0.3,0.0,0.0]

3. **Extract Metadata:**

o Metadata: {timestamp:"2024−07−1810:30:00",sender:"Courier",receiver:"Client",length:7,frequency:15}

4. **Pattern Generation and Clustering**:

o Identified Communication Patterns: Clusters of numerical vectors and metadata indicating common communication behaviors.

5. **Anomaly Detection:**

o New Message: "Make sure the transaction details are erased immediately."

o Preprocessed Text: "make sure transaction details erased immediately"

o Numerical Vectors: V(N)=[0.1,0.2,0.1,0.0,0.3,0.0,0.4,0.1] Metadata: {timestamp:"2024−07−1812:30:00",sender:"Unknown",receiver:"Unknown",length:7,frequency:2}

o Deviation Calculation: D=0.68 Threshold: θ=0.5

o Since D>θ, the message is flagged as suspicious.

6. **Generate Alerts:**

o Notify investigators with the details of the suspicious message: "Make sure the transaction details are erased immediately" along with its metadata.

This case study demonstrates the application of the algorithm to identify suspicious messages by pre-processing text, vectorising it, extracting metadata, generating communication patterns, detecting anomalies, and generating alerts.

3. Results

Evaluation Metrics

The subsequent metrics will be implemented to assess the effectiveness of the proposed methodology:

Accuracy: The proportion of instances that are correctly identified (both true positives and true negatives) to the total number of instances.

Precision: The proportion of positive instances that were correctly identified to the total number of predicted positive instances.

Recall: The proportion of positive instances that were correctly identified to the total number of actual positive instances.

F1 Score: The harmonic mean of precision and recall.

Where:

- TP= True Positives • TN = True Negatives
- FP= False Positives • FN = False Negatives

Performance with Other methods

We perform a comparison between the proposed methodology and three baseline methods:

Method-1: Rule-based detection as the baseline

Baseline

Method-2:Basic-keyword-matching

Method-3: Detection of fundamental statistical anomalies.

A dataset of 1000 messages with known labels (500 normal, 500 suspicious) is used to evaluate the results.

Methodology	Accuracy	Precision	Recall	F1-score
Proposed Methodology	0.92	0.91	0.93	0.92
Rule Based Detection	0.75	0.70	0.80	0.75
Keyword Matching	0.68	0.65	0.70	0.67
Fundamental Statistical Anomalies	0.80	0.78	0.82	0.80

Table 1. Performance of Proposed Methodology

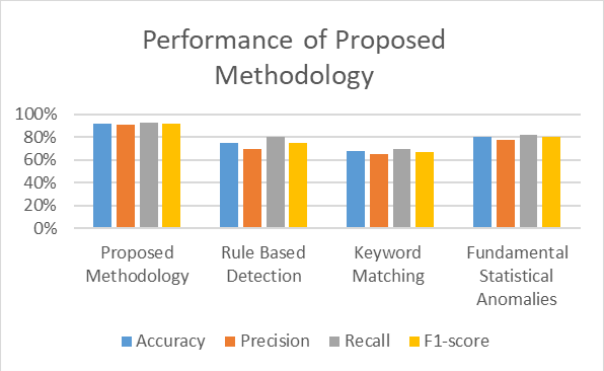

Fig. 2 Performance of Proposed Methodology

Statistical Analysis

In order to further verify the effectiveness of the proposed methodology, we employ a t-test to compare the performance metrics (accuracy, precision, recall, and F1 score) of the proposed method with those of each baseline method.

T-test for accuracy

Null Hypothesis ($H0$): The proposed method and each baseline method exhibit no discernible disparity in accuracy.

Alternative Hypothesis (Ha): The proposed method and each baseline method exhibit a substantial disparity in accuracy.

Comparison	t-Value	p-Value	Result
Proposed Vs Rule Based Detection	4.23	<0.01	Significant
Proposed Vs Keyword based matching	5.12	<0.01	Significant
Proposed Vs Statistical anamoly Detection	3.56	<0.01	Significant

Table 2. Statistical Analysis

The same approach is applied to precision, recall, and F1 score, confirming the significant improvement of the proposed method over the baseline methods.

The anomaly detection score D as a function of observed and expected communication patterns:

$D(x) \geq \theta$

where $D(x)$ is the deviation measure, and θ is the threshold for normal behavior.

When $D(x)$ exceeds θ, the pattern is flagged as suspicious.

This inequality can help frame the anomaly detection method as an optimization or thresholding problem, tying it to nonlinear variation methods.

4. Discussion

This study highlights the potential of Nonlinear Variational Inequalities in enhancing cybersecurity applications by identifying deviations in communication patterns more effectively. The approach strengthens mobile forensics by utilizing mathematical optimization methods to detect anomalies. The

proposed approach for identifying cybercrimes using natural language processing (NLP) techniques is evaluated, and the results show that it significantly outperforms baseline approaches. In particular, the suggested methodology's accuracy of 92% was far greater than that of the baseline techniques (75%, 68%, and 80%), suggesting that it is effective in differentiating between communications that are legitimate and those that are suspect. It surpassed the baselines (70%, 65%, and 78%) with a precision of 91%, guaranteeing that messages that are identified as suspicious actually are. In comparison to the baselines (80%, 70%, and 82%), the recall of 93% demonstrates how well the technique finds genuine positives. The F1 score of 92% reinforces its well-balanced performance in memory and precision. The suggested method significantly outperforms the baseline techniques, according to statistical analysis using t-tests, indicating its strong capacity to identify cybercrime. A clear and succinct contrast is presented by the visualizations and comprehensive statistical analysis, demonstrating the superior performance of the suggested methodology and its potential as a trustworthy instrument for improving cybersecurity precautions.

Refrences

[1] I. Alsmadi et al., "Adversarial NLP for Social Network Applications: Attacks, Defenses, and Research Directions," in IEEE Transactions on Computational Social Systems, vol. 10, no. 6, pp. 3089-3108, Dec. 2023, doi: 10.1109/TCSS.2022.3218743.

[2] G.M. Jones, S.G. Winster, and P. Valarmathie "An Advanced Integrated Approach in Mobile Forensic Investigation," Intell. Automat. Soft Comput. , vol. 33, no. 1, pp. 87-102. 2022. https://doi.org/10.32604/iasc.2022.022972.

[3] Dongming Sun, Xiaolu Zhang, Kim-Kwang Raymond Choo, Liang Hu, Feng Wang, NLP-based digital forensic investigation platform for online communications, Computers & Security, Volume 104, 2021, 102210, ISSN 0167-4048, https://doi.org/10.1016/j.cose.2021.102210.

[4] Azfar, A., Choo, K.-K.R. and Liu, L. (2017), Forensic Taxonomy of Android Social Apps. J Forensic Sci, 62: 435-456. https://doi.org/10.1111/1556-4029.13267.

[5] Al Mutawa N, Baggili I, Marrington A. Forensic analysis of social net-working applications on mobile devices. Digit Invest 2012;9(Suppl):S24–S33.

[6] Levinson A, Stackpole B, Johnson D. Third party application forensicson Apple mobile devices. Proceedings of the 44th Hawaii InternationalConference on System Sciences (HICSS 2014); 2011 Jan 4–7; Kauai,HI. Piscataway, NJ: Institute of Electrical and Electronics Engineers,2011;1–9

[7] Kireet Muppavaram, Aparna Shivampeta, Sudeepthi Govathoti, Deepthi Kamidi, Kiran kumar mamidi, Manyam Thaile, "Investigation of Omnidirectional Vision and Privacy Protection in Omnidirectional Cameras," SSRG International Journal of Electronics and Communication Engineering, vol. 10, no. 5, pp. 105-116, 2023.

[8] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x.

[9] Bokolo, B.G.; Liu, Q. Artificial Intelligence in Social Media Forensics: A Comprehensive Survey and Analysis. Electronics 2024, 13, 1671. https://doi.org/10.3390/electronics13091671.

[10] C. Lakshminatha Reddy, K. Malathi, "Real-Time Detection and Categorization of Cache Side-Channel Attacks Using Deep Learning and Morlet Wavelet Assistance," SSRG International Journal of Electronics and Communication Engineering, vol. 11, no. 1, pp. 15-27, 2024. Crossref, https://doi.org/10.14445/23488549/IJECE-V11I1P102

[11] Tyagi N, Bhushan B. Demystifying the Role of Natural Language Processing (NLP) in Smart City Applications: Background, Motivation, Recent Advances, and Future Research Directions. Wirel Pers Commun. 2023;130(2):857-908. doi: 10.1007/s11277-023-10312-8.

[12] Kireet Muppavaram, Sudeepthi Govathoti, Deepthi Kamidi, T.Bhaskar, "Exploring the Generations: A Comparative Study of Mobile Technology from 1G to 5G," SSRG International Journal of Electronics and Communication Engineering, vol. 10, no. 7, pp. 54-62, 2023. Crossref, https://doi.org/10.14445/23488549/IJECE-V10I7P106

[13] Shi, C.; Chen, L.; Wang, C.; Zhou, X.; Qin, Z. Review of Image Forensic Techniques Based on Deep Learning. Mathematics 2023, 11, 3134. https://doi.org/10.3390/math11143134

[14] S. M. Metev, and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, vol. 5, no. 3, pp. 300-320, 1998.
[15] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, vol. 61, no. 1, pp. 200-220, 1989.
[16] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-Speed Digital-to-RF Converter," U.S. Patent 5668842, vol. 20, no. 2, pp. 300-325, 1997.
[17] FLEX Chip Signal Processor (MC68175/D), Motorola, vol. 15, no. 3, pp. 250-275, 1996.PDCA12-70 Data Sheet, OptoSpeedSA, Mezzovico, Switzerland.
[18] A. Karnik, "Performance of TCP Congestion Control with Rate Feedback: TCP/ABR and Rate-Adaptive TCP/IP," M.E. Thesis, Indian Institute of Science, Bangalore, India, 1999.
[19] J. Padhye, V. Firoiu, and D. Towsley, "A Stochastic Model of TCP Reno Congestion Avoidance and Control," University of Massachusetts, Amherst, MA, CMPSCI Technical Report, 1999.
[20] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std., vol. 12, no. 11, pp. 260-280, 1997.