IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Identyfying Picpoket Suspects By Analysing Large Of Public Transit Records

[1]Dr. U. Prasad, [2]M. Ganesh, [3]S. Uttam Kumar, [4]K. Dileep Kumar, [5]D. Kalyan Chakravarthy

[1]Professor, Department of Information Technology, QIS college of Engineering and Technology, Ongole, Andhra Pradesh, India
[2,3,4,5]Student, Department of Information Technology, QIS college of Engineering and Technology , Ongole, Andhra Pradesh, India

***Abstract:*** Massive amounts of information about commuter habits and public transportation systems as a whole are collected by automated fare collecting (AFC) systems. Traditional AFC data analyses have centred on identifying passenger behaviour patterns. But we came up with an original use for this data: monitoring the area for pickpockets. Pickpocketing on public transportation is down, which is great news for passengers and for society as a whole. However, in practise, it is challenging to distinguish criminals from law-abiding commuters. In this article, we look at the process of developing a surveillance and detection system that can identify potential pickpockets by analysing a person's daily transportation logs. In order to be more specific, we identified a set of modifiable features common to the daily commutes of all riders. Then, because criminals move around in unusual ways, we used a two-pronged approach that involved both unsupervised outlier identification and supervised classification models to spot them. Our method was proven effective in experiments. We also developed a prototype system that security personnel could use.

***Index Terms -*** fare, pickpocket, thief, transport data, passenger, identification, offender.

## I. INTRODUCTION

Many riders of public transportation are in a hurry and easily distracted, especially when the train or bus is overflowing with passengers. Their attention wanders from their possessions, making them easy targets for thieves [1, 2]. In the first nine months of 2014, there were reports of 350 pickpockets being caught in Beijing's subways and 490 being caught on buses. There is a serious issue with pickpocketing in many European cities, including Barcelona, Rome, and Paris. Smart, crafty criminals who know how to avoid being caught provide a big challenge to investigators looking for stolen items. It is essential to provide security officers in public transportation with a state-of-the-art surveillance and tracking device. Automated fare collection (AFC) systems collect useful information about passenger mobility patterns and the dynamics of urban areas [3, 4, 5, 6, 7]. Current efforts in the field are focused on elucidating common, communal mobility behaviours including commute flows and transportation networks. To the best of our knowledge, our study is the first to employ AFC data for the sole purpose of identifying criminals. Therefore, AFC data can be used in the fight against theft by utilising the mobility footprints to identify possible suspects based on their unique behaviours. These tendencies include doing things like taking longer routes, taking more transfers than necessary, or making unscheduled stops along your usual route. Intelligent systems can now be built to automatically extract predetermined behavioural features for the purpose of dynamically detecting and tracking potential pickpocketing suspects. Finding prospective thieves in AFC data is more complicated than just checking for anomalies. FIGURE 1: Known thieves; FIGURE 2: Atypical individuals. There are multiple routes that can be inferred from the apparent separation of the two hottest regions (A and B). From our statistical analysis, we know that the vast majority of travellers choose for the path that involves the fewest number of transfers or the lowest total trip time and distance. A traveller (a known suspect) who follows the route A -> C -> D ->! B stands out since it is not required to transfer at C and D to go to B. Customers whose actions reflect this profile will be picked out for further scrutiny. You're a rare bird if you choose the path from A to B that goes via E. And yet, this visitor is probably no different from any other tourist from a less populous area. The fact that even routine passenger excursions might be disguised as something more nefarious adds to the difficulty of catching criminals. Even while it's not out of the ordinary for frequent commuters to take outings to see friends or places of interest, the fact that these excursions are so obviously out of character raises suspicion. Collecting AFC information from millions of passengers complicates an already complex situation when just a tiny fraction of travellers are genuine pickpockets. It's like trying to find a needle in a haystack to track down a certain subset of individuals in a huge database. Our model-based information must be quickly adapted into a decision-making tool while this is occurring. This system has to provide timely decision suggestions to security officers for them to conduct their responsibilities more efficiently.

## I. RELATEDWORKS

The current body of research places a heavy emphasis on methods for identifying patterns in data gathered from passenger traffic. Multiple applications exist for this data, and it is essential for determining and satisfying passenger needs. Among them include assessing the overall health of the system, identifying and repairing ineffective bus routes, improving the accuracy with which passenger flow is projected between two locations, and modifying services to reflect daily and seasonal variations in demand. For instance, in [4] the authors estimate the activity levels of important transportation hubs using AFC data. The cyclical nature of public transportation usage was measured by academics in [9].

Existing studies that investigate urban sensing data for anomalies may be divided into two categories: those that look at particular places and those that look at specific pathways. In [15], the authors provided a framework for location-based anomaly identification via the study of the context of several operational urban areas; this is the basis for our feature extraction technique. In addition, [16] sought to find temporal and geographical anomalies that were coincidentally linked. [17] gathered key terms from social media posts made after city-specific situations like accidents and protests. The black-hole and volcano patterns discovered in [18]'s analysis of people's movements in a city might be utilised to quickly identify large-scale events like football games and concerts. When anomalies are spotted, alerts may be generated and data utilised for intelligent decision support, such enhancing traffic flow.

**Disadvantages**

Activity patterns and transit data for passengers are now recorded manually, therefore using a smart card is no longer an option. Due to the no-checkout policy, the system was unable to track down the actual perpetrators of the crimes that were reported by victims. As an alternative, the algorithm labelled people as possible thieves based on their behaviour. More specifically, the computer first compiled a whole set of riders for a certain time period, and then analysed a visualisation of their itineraries to assess how typical they were.

## II. PROPOSED SYSTEM ARCHITECTURE

To solve the problem of identifying pickpockets, the proposed technology considered the problem from every possible angle. This diagram represents the basic outline of our response. The city was originally divided into zones according to its function under the scheme. Then, using the collected transit information, the mobility characteristics of the passengers were calculated in real time. A two-step passenger classification approach was crucial to the strategy, with the first stage focusing on weeding out "regular" passengers and the second stage locating "suspect" people. At last, user feedback information, such as newly confirmed thieves, was entered into the utility function that strikes a balance between performance and relevance for use in training subsequent models (i.e., recency). This system might provide a more in-depth analysis of the system. The main takeaway from our research is outlined here. We started by isolating certain characteristics of AFC data that may be used to distinguish between criminals and law-abiding tourists. Since the number of "good" and "bad" samples may be so drastically different in a large-scale data setting, a two-step approach was proposed to handle the suspect detection problem. Finally, our advancement in dynamic filtering greatly reduced the typical computation costs while still retaining leading accuracy in the market. Moreover, real-world, large-scale data was utilised to develop and test a system for the end user. In this situation, the detection of pickpockets, our approach is the first to employ big data to address a critical social issue. The publication The Economist has recognised the significance of this issue by devoting an entire article to it.

**Advantages**

Smart card-based public transportation record-keeping; effective techniques for spotting irregular patterns of use. **Modules**

If the Web Server wants to see this content, it will need to provide a username and password. Once he logs in, he'll be able to use options like the User List and the Authorize button. Existing routes may be seen and edited, smart card information can be viewed, facts about all passenger journeys can be viewed, potential pickpockets can be viewed, and the results of passenger trips and transit records can be examined. **User**

Currently, n users are logged into this section. A user must sign up before they may access some features. After completing registration, he will be asked to log in using his personal ID and password. When he logs in, he can do things like see his profile, add a smart card, view the data of that card, add information about a boarding station, view and add information about an exit station, and view information about his trips. People using crowded public transit systems are more likely to be in a hurry and less attentive. They are common targets for pickpockets because they often misplace their belongings [1, 2]. On the first nine months of 2014, there were allegedly 350 pickpockets nabbed in the Beijing subway and 490 on Beijing buses. Many other large cities throughout the world have the same problems with pickpocketing as do Barcelona, Rome, and Paris. Criminals who are cunning enough to evade capture are rare but not impossible to apprehend. For the safety of the public, it is essential that transit authorities have access to sophisticated monitoring and location-stability tools. Data collected by automated fare collection (AFC) systems has allowed for an examination of passenger movement patterns and urban dynamics thanks to the rapid growth of IT and associated infrastructure.

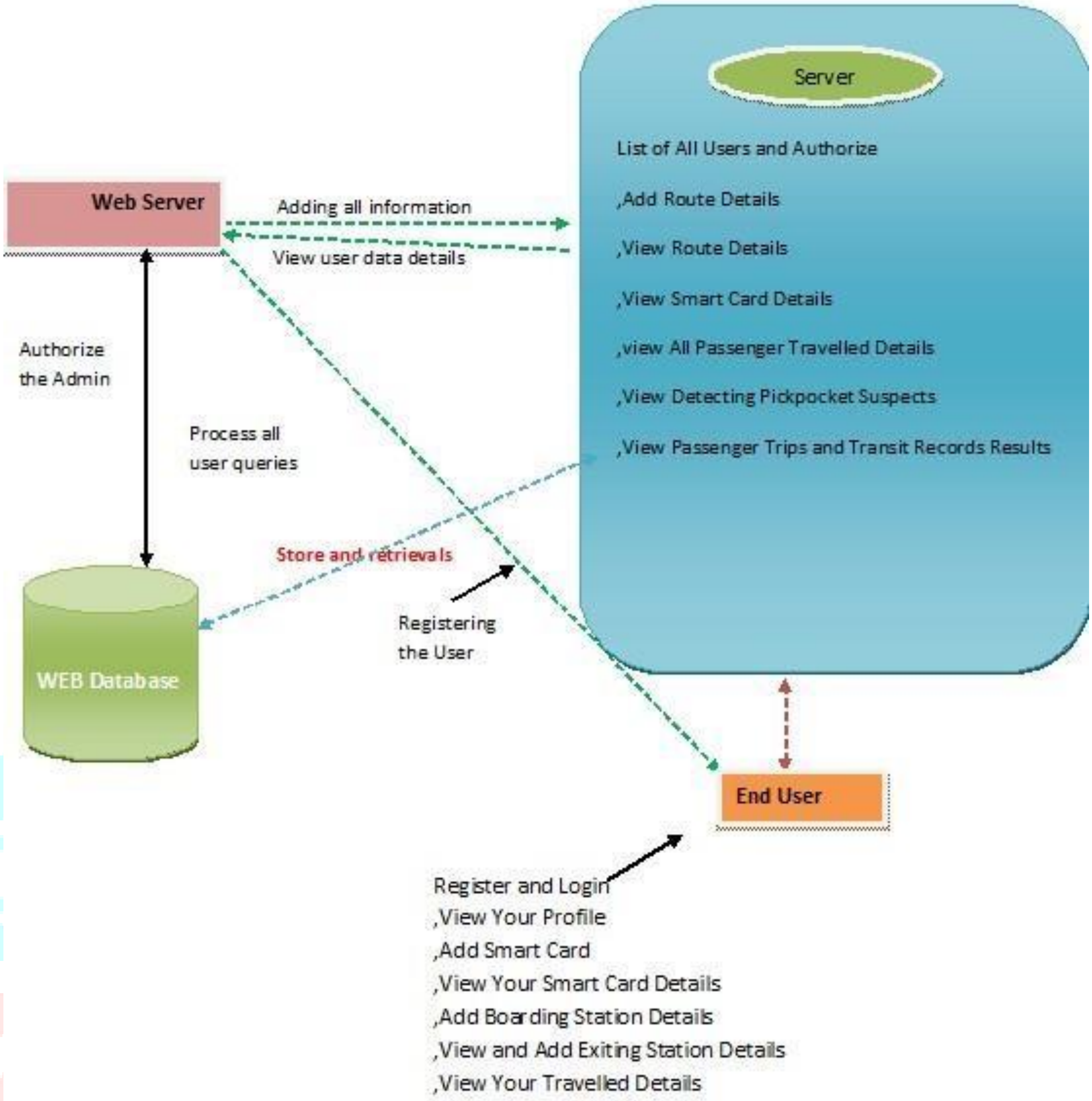


Fig.1 Proposed system architecture

## III. RESULTS AND DISCUSSION

The following is presented as reasoning for a two-stage framework to be used in identifying possible pickpockets: Since the great majority of passengers are law-abiding citizens, it would be hard to implement a classification system. Particularly, the number of confirmed pickpockets in the tourist industry is vanishingly tiny. Simple heuristics like oversampling and under sampling wouldn't help with our country's huge class divide. When it comes to building reliable machine learning models for such unbalanced data, there is still a lot of uncharted territory to be discovered in the literature. Contrarily, using anomaly detection algorithms, which are often unsupervised, cannot scale effectively and may potentially lead to considerable false positives, as many regular travellers who occasionally undertake unusual actions may be wrongly labeled as suspects. Fig.2 depicts the final displays seen after the system has been performed and executed. In this paper, we provide the experimental results of using our proposed method. We start with a description of the test environments and the technical details of the implementation. We next demonstrate the value of our approach by contrasting it to a wide range of established benchmarks. The procedures we followed and the apparatus we used are outlined here. We give some quick platform, baseline, and performance data.

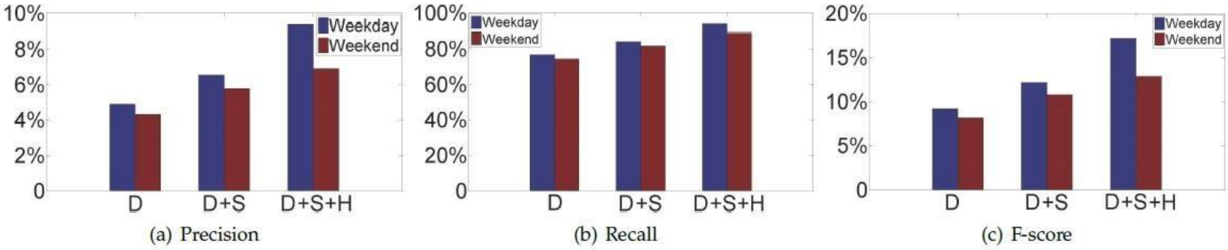


Fig.2 Comparative analysis of existing system with proposed system **Platform.**

We used a Windows Server 2012 64-bit computer with four 2.6GHz Quad-Core CPUs and 128GB of main RAM for all of our offline testing. The real-time solution was deployed using Spark clusters with 10 nodes. Centos 6.5 operating system, 2 terabytes of SATA3.0 hard drive capacity, 8 cores from an Intel i7-4790 CPU, 2 GB of RAM from Kingston, and 3.6 GHz CPU speed are all standard on each node. We built our system and all of its algorithms with the help of the Java and Scala programming languages.

Getting the Numbers Ready. All analyses relied on real-world data, as described in Section 3. About 1.7 billion data points were collected in the period between April and June of 2014. The training set was cleaned of all passengers who had more than three daily recordings. After removing duplicates and extremely infrequent riders, we were left with about 1.6 billion data involving around 6 million passengers throughout the three month period.

We built a 'training set' for the model-building stage, and a 'testing set' for the final checks. The offline experiments use a threemonth (April–June, 2014) training set, with the succeeding two weeks (June–July, 2014) serving as the test set (in July 2014). In order to evaluate the real-time system, we let the models progressively learn and reuse data. Every day, roughly 5 million people's records—totaling about 14 million—are compiled.

## IV. FUTURE SCOPE AND CONCLUSION

For this research, we analysed a large database of transit data to develop a method for keeping tabs on questionable people. The gadget may be used to proactively monitor high-risk areas and to detect possible pickpocketing suspects. A feature representation was constructed before any passenger profiling could commence. Then we came up with a ground-breaking two-stage process to spot possible pickpockets in the throng. Finally, by using real-world information from a number of sources for model training and validation, we were able to construct a prototype system for end users. Experimental results using real-world data confirmed the efficacy of our proposed approach.

## REFERENCES


[1] G. R. Newman and M. M. McNally, "Identity theft literature review," United States Department of Justice, Report 210459, July 2005.

[2] M. Felson and R. V. Clarke, "Opportunity makes the thief: Practical theory for crime prevention," Policing and Reducing Crime Unit: Police Research Series, Report 98, 1998.

[3] T. L. C. da Silva, J. A. F. de Macˆedo, and M. A. Casanova, "Discovering frequent mobility patterns on moving object data," in MobiGIS, 2014, pp. 60–67.

[4] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: understanding tube station congestion patterns from trip data," in UrbComp, 2012, pp. 134–141.

[5] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders travel patterns," Transportation Research Part C, vol. 36, pp. 1–12, 2013.

[6] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, "Online discovery of gathering patterns over trajectories," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 8, pp. 1974–1988, 2014.

[7] Y. Liu, C. Liu, J. Yuan, L. Duan, Y. Fu, H. Xiong, S. Xu, and J. Wu, "Intelligent bus routing with heterogeneous human mobility patterns," Knowledge and Information Systems, 2016, forthcoming, DOI: 10.1007/s10115-016- 0948-6.

[8] B. Du, C. Liu, W. Zhou, Z. Hou, and H. Xiong, "Catch me if you can: Detecting pickpocket suspects from large-scale transit records," in KDD, 2016, pp. 87–96.

[9] C. Morency, M. Tr´epanier, and B. Agard, "Analysing the variability of transit users behaviour with smart card data," in ITSC, 2006, pp. 44–49.

[10] M. J. Sung, "Analysis of travel patterns of the elderly using transit smart card data," in TRB, no. 11-2357, 2011.

[11] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barab´asi, "Understanding individual human mobility patterns," Nature, vol. 453, no. 7196, pp. 779–782, 2008.

[12] C. Song, Z. Qu, N. Blumm, and A.-L. Barab´asi, "Limits of predictability in human mobility," Science, vol. 327, no. 5968, pp. 1018–1021, 2010.

[13] C. Chen, D. Zhang, Z.-H. Zhou, N. Li, T. Atmaca, and S. Li, "B-planner: Night bus route planning using largescale taxi gps traces," in PerCom, 2013, pp. 225–233.

[14] W. Luo, H. Tan, L. Chen, and L. M. Ni, "Finding time period-based most frequent path in big trajectory data," in SIGMOD, 2013, pp. 713–724. [15] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in KDD, 2012, pp. 186–194.

[16] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in KDD, 2011, pp. 1010–1018.

[17] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in GIS, 2013, pp. 344–353.

[18] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou, "Detecting urban black holes based on human mobility data," in GIS, 2015, pp. 35:1–35:10.

[19] P. Bouman, E. Van der Hurk, L. Kroon, T. Li, and P. Vervest, "Detecting activity patterns from smart card data," in BNAIC, 2013.

[20] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," IGIEEE MOD Rec., vol. 29, no. 2, pp. 93–104, May 2000.

[21] C. Liu, H. Xiong, Y. Ge, W. Geng, and M. Perkins, "A stochastic model for context-aware anomaly detection in indoor location traces." in ICDM, 2012, pp. 449–458.

[22] S. Jiang, J. Ferreira Jr, and M. C. Gonzalez, "Discovering urban spatial-temporal structure from human activity patterns," in UrbComp, 2012, pp. 95–102.

[23] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in ICDM, 2011, pp. 181– 90.