Website:https://www.bgpulseusa.com/index.php/ICAIA Volume 7,Issue 01,November-2026

DEVELOPMENT OF A MOBILE ANTI-FRAUD MODEL BASED ON SESSION AND TRANSACTION ANALYSIS

Odilov O.R.
Ashirmurotov M.N.
Nuraliyev Sh.Sh.
Bobonazarova M.Sh.
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,Tashkent, Uzbekistan
DOI:https://doi.org/10.5281/zenodo.24050234

Abstract. This article addresses the detection of fraud in mobile banking at the point at which a payment is authorised. The prevailing practice evaluates the payment order in isolation, by rules applied to its amount, time and beneficiary. That practice fails against the two forms of attack that now dominate: remote control of the customer device, in which the payment originates from the genuine device and the genuine credentials, and social engineering, in which the customer authorises the payment himself. In both the transaction is formally legitimate and the evidence of fraud lies not in the order but in the session that produced it. A four-layer reference model is constructed in which device, behavioural and transactional signals are assembled into a session-context feature vector, a risk score is computed before authorisation, and the score is mapped to a graduated decision policy whose thresholds are set from the cost of each class of error rather than from a single accuracy figure. A taxonomy of session signals is given with the attack class each signal detects; the construction of the model under extreme class imbalance is described; and an evaluation methodology is stated in which recall at a fixed review capacity, rather than accuracy or the area under the ROC curve, is the operative metric. Adversarial adaptation, concept drift and the requirements of the legislation of the Republic of Uzbekistan are examined.

Keywords: anti-fraud, mobile banking, session analysis, transaction monitoring, behavioural biometrics, device fingerprinting, account takeover, authorised push payment fraud, remote access trojan, class imbalance, cost-sensitive learning, concept drift.

Annotatsiya. Ushbu maqolada mobil bankingda to‘lov tasdiqlanadigan nuqtada firibgarlikni aniqlash masalasi ko‘rib chiqiladi. Amaldagi yondashuv to‘lov topshirig‘ini alohida, uning summasi, vaqti va qabul qiluvchisiga qo‘llanilgan qoidalar orqali baholaydi. Bu yondashuv bugungi ikki asosiy hujum turiga qarshi samarasiz: mijoz qurilmasini masofadan boshqarish — bunda to‘lov haqiqiy qurilma va haqiqiy hisob ma’lumotlari bilan amalga oshiriladi — hamda ijtimoiy muhandislik, bunda to‘lovni mijozning o‘zi tasdiqlaydi. Har ikki holatda tranzaksiya rasman qonuniy bo‘lib, firibgarlik dalili topshiriqda emas, uni keltirib chiqargan sessiyada bo‘ladi. Maqolada to‘rt qatlamli etalon model quriladi: qurilma, xulq-atvor va tranzaksiya signallari sessiya kontekstidagi belgi vektoriga birlashtiriladi, xavf bahosi tasdiqlashdan oldin hisoblanadi va u bosqichma-bosqich qaror siyosatiga akslantiriladi; chegaralar yagona aniqlik ko‘rsatkichidan emas, har bir xato turining narxidan kelib chiqib belgilanadi. Sessiya signallarining tasnifi, sinflar nomutanosibligi sharoitida model qurish va baholash metodologiyasi bayon etiladi. Moslashuvchan hujumlar, kontseptual siljish hamda O‘zbekiston Respublikasi qonunchiligi talablari tahlil qilinadi.

Kalit so‘zlar: antifrod, mobil banking, sessiya tahlili, tranzaksiya monitoringi, xulq-atvor biometriyasi, qurilma barmoq izi, hisobni egallash, tasdiqlangan to‘lov firibgarligi, masofaviy kirish troyani, sinflar nomutanosibligi, narxga sezgir o‘qitish, kontseptual siljish.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

Introduction. The migration of retail payments to the mobile channel has moved the point of attack. Where the perimeter was once the bank, it is now the customer's handset, and the instruments of fraud have adapted accordingly. Two classes of attack now predominate, and both share a property that defeats conventional transaction monitoring: the payment order they produce is, on its face, entirely legitimate.

The first is device takeover. Contemporary Android banking trojans obtain Accessibility Service permissions and thereby acquire full control of the handset: they map every visible interface element with its coordinates and interactive state, log keystrokes and selection changes, capture the unlock pattern, and present a full-screen overlay in place of the genuine banking screen in order to harvest PIN codes and one-time passwords. Evasion is correspondingly developed — one family routes Google Play Protect traffic through a null VPN channel, signs its payload with a freshly generated key to defeat hash-based detection, and blocks the removal and factory-reset screens. A single such family was confirmed against more than thirty financial institutions in six countries within months of first observation. A payment initiated under this control originates from the registered device, through the genuine application, with correct credentials and a correctly entered second factor.

The second is social engineering, in which the customer is persuaded to authorise the payment himself. The scale is no longer marginal. In the United Kingdom, where the statistics are published in a consistent series, total payment fraud losses reached £1.28 billion in 2025, an increase of four per cent, across more than four million confirmed cases. Within that total, authorised push payment losses rose by nineteen per cent to £576.4 million over 248,070 cases, while unauthorised losses fell by five per cent to £703.4 million. The divergence of the two series is the central fact for the design of a detection system: the category that controls can suppress is in retreat, and the category in which the customer himself issues the instruction is advancing. It is also notable that sixty-six per cent of authorised push payment cases began online and seventeen per cent through telecommunications, so that the origin of the deception lies outside the bank entirely.

In both classes the transaction record contains no anomaly. The device is known, the credentials are correct, the authentication succeeded, and in the second class the customer's intention to pay is genuine. A rule engine operating on amount, hour and beneficiary will either pass the payment or, if its thresholds are lowered far enough to catch it, generate a volume of false positives that no review function can absorb and that drives legitimate customers out of the channel.

The evidence that distinguishes these payments from genuine ones exists, but it lies in the session rather than in the order. Under remote control the interaction carries the signature of the machine: navigation without hesitation, input timing without human variance, an accessibility service active where none is required, a screen-sharing or remote-control channel open. Under social engineering the interaction carries the signature of distress: an unusually long dwell on the transfer screen, the hallmarks of a customer reading instructions dictated to him, a concurrent telephone call, hesitation and correction at the beneficiary field. Neither appears in the payment record, and both are observable before the payment is authorised.

The contribution of this article is fourfold. First, a four-layer reference model is constructed in which session and transaction evidence are combined into a single risk score computed before authorisation. Second, a taxonomy of session signals is given in which each signal is matched to the attack class it detects and to its resistance to falsification. Third, the decision stage is formulated as a cost problem rather than a classification problem, with thresholds derived from the cost of each class of error and from the capacity of the review function. Fourth, an evaluation methodology is stated that is appropriate to extreme class imbalance, together with the requirements imposed by the legislation of the Republic of Uzbekistan on the collection of the signals the model requires.

Literature review and methodology. The detection of payment fraud was first treated as a supervised classification problem on transaction attributes, and the methodological difficulties of that formulation were established early: extreme class imbalance, a delay of days or weeks between a transaction and its confirmed label, and a cost asymmetry in which a missed fraud and a false alarm are

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

not commensurable. Dal Pozzolo and colleagues demonstrated that under these conditions conventional accuracy measures are uninformative and that the operative question is the precision of the alerts a bounded review function can actually process.

Behavioural biometrics developed as a separate line. The proposition is that the manner of interaction — the dynamics of typing, the geometry of swipes, the pressure and orientation of the handset, the path taken through the application — identifies the user continuously and passively, without an explicit authentication event. Commercial deployments report that this evidence detects precisely the two classes above, because automated interaction and coerced interaction both diverge from the customer's own established pattern even when the credentials are correct.

Device intelligence constitutes a third line, concerned with the integrity and identity of the endpoint: fingerprinting, the detection of rooting, emulation and instrumentation frameworks, the enumeration of active accessibility services, and the detection of screen capture or remote-control channels. The threat intelligence literature on mobile banking trojans supplies the adversary model against which these detections must be designed.

Two gaps are apparent. The first is that the three lines are generally treated as separate products and evaluated separately, whereas the attacks in question are detected by their combination: a device signal alone yields false positives on legitimate accessibility users, and a transaction signal alone yields nothing at all on an authorised payment. The second is that the decision stage is usually reduced to a single threshold on a score, whereas in operation the response is graduated — step-up authentication, delay, warning, refusal — and the thresholds that separate those responses are determined by costs that the classification literature does not model.

The methodology of this article is analytical and constructive. A reference model is built and the risk function is formalised. The signals are classified by attack coverage and by resistance to falsification. The decision stage is formulated as the minimisation of expected cost subject to a capacity constraint. An evaluation protocol appropriate to the imbalance and to the labelling delay is specified. Published statistics are cited as of the date of writing.

Reference model. The process is modelled as four layers, shown in Figure 1. The model differs from the classical transaction-monitoring pipeline in that the unit of analysis is the session rather than the payment, and in that the score is computed before authorisation rather than after settlement.

At the first layer signals are acquired continuously from the moment the application is opened: device and session attributes, behavioural telemetry, and the attributes of the payment under construction. At the second layer these are assembled into a single feature vector in which each value is expressed, wherever possible, as a deviation from the customer's own established profile rather than as an absolute quantity. At the third layer a risk score is computed by a learned model, retaining a rule layer for hard constraints that must not be subject to a probabilistic judgement — sanctions screening, accounts already confirmed as mule accounts, statutory limits. At the fourth layer the score is mapped to one of four graduated responses.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

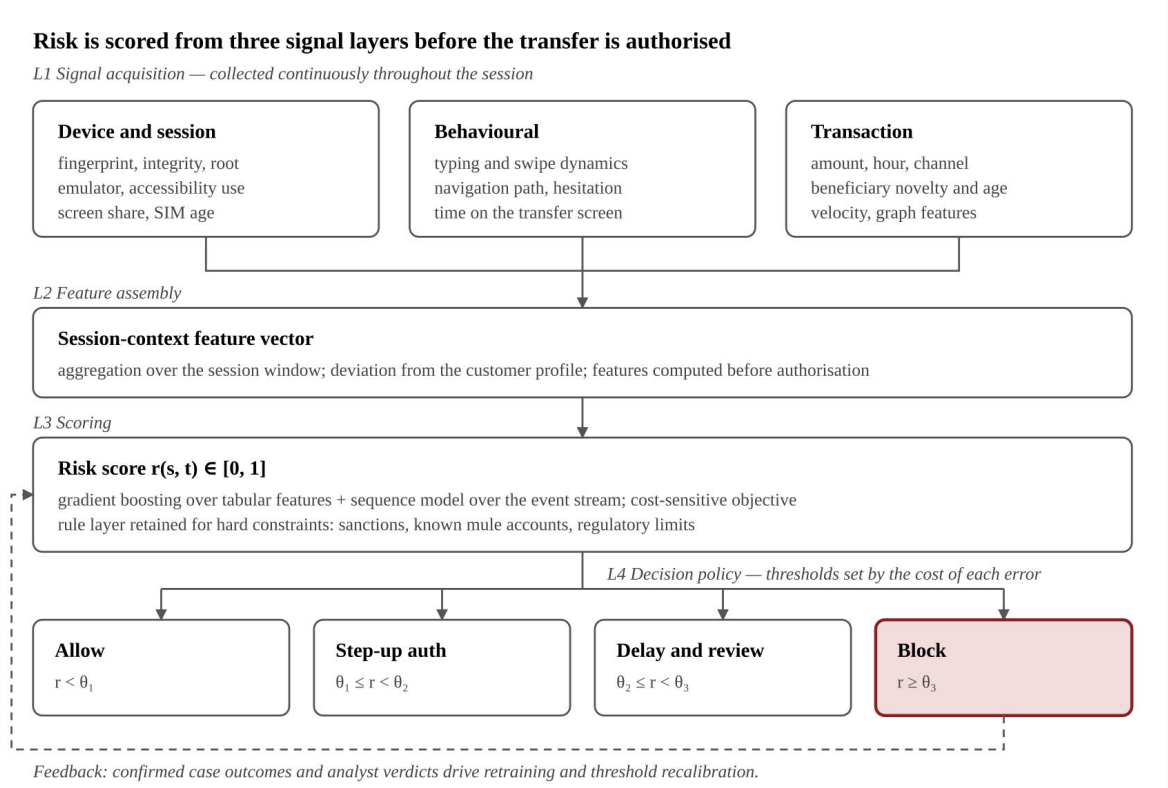


Figure 1. Four-layer reference model of the mobile anti-fraud system with the decision policy and the feedback loop from confirmed outcomes

Two properties of this construction deserve emphasis. The first is that the score is a function of the session and the transaction jointly, r(s, t), and not of the transaction alone; the same payment order carries a different risk depending on how the session that produced it behaved. The second is that the feedback path is part of the model and not an operational afterthought: confirmed case outcomes and analyst verdicts are the only source of labels, they arrive with a delay, and the thresholds must be recalibrated as the attack mix changes.

Formal statement. Let s denote a session and t a payment order constructed within it. The session is an ordered sequence of events

$$s = (e_1, e_2, \ldots, e_n), \quad e_i = (\tau_i, \text{type}_i, \text{attr}_i),$$

where τ_i is the timestamp of the event, type_i its kind — screen transition, field entry, authentication, permission query — and attr_i its attributes. The feature assembly layer computes

$$x = \varphi(s, t) \in \mathbb{R}^d,$$

in which the components fall into the three groups set out in Table 1. The scoring layer computes

$$r(s, t) = f(x) \in [0, 1],$$

the estimated probability that the payment is fraudulent, and the decision layer applies the policy

$$d(r) = \text{allow if } r < \theta_1; \text{ step-up if } \theta_1 \le r < \theta_2; \text{ delay if } \theta_2 \le r < \theta_3; \text{ block if } r \ge \theta_3.$$

The thresholds are not free parameters to be tuned for accuracy. Let c_h be the expected cost of a fraudulent payment that is allowed, c_e the cost of friction imposed on a legitimate customer at each level of response, and let the review function have capacity m cases per period. The thresholds solve

$$\min \Sigma \left[c_h \cdot P(\text{fraud, allowed}) + \Sigma_k c_e(k) \cdot P(\text{legitimate, response } k) \right] \quad \text{subject to} \quad E[\text{cases referred}] \le m.$$

The capacity constraint is the element most often omitted and the one that governs the design in practice. A model that refers more cases than the review function can process does not detect more fraud; it detects less, because the queue ages and the alerts are cleared without examination.

The costs are not symmetric and not constant. A fraudulent payment allowed carries the loss, the reimbursement where the regime requires it, and the investigative cost. Friction carries an abandonment

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

probability that rises sharply with the severity of the response: a step-up challenge is tolerated, a refusal is not, and a legitimate customer refused twice is substantially likely to leave the channel. It follows that the graduated policy is not a convenience but a requirement, since collapsing four responses into a binary one forces the choice between accepting the loss and imposing the maximum friction on the whole of the affected population.

Session signals and their coverage. The signals available within a session are heterogeneous in what they detect and in how readily an adversary can falsify them. Table 1 classifies them on both axes. The classification is the basis of the fusion rule: a signal of low resistance may raise suspicion but may not alone determine a response, whereas the coincidence of signals from different groups is what carries weight.

Table 1. Taxonomy of session signals, the attack class detected and resistance to falsification

Group	Signal	Attack class detected	Resistance
D1 Device identity	Device fingerprint, binding to the registered handset, SIM identifier and age of the current SIM	Takeover from an unregistered device; SIM-swap interception of one-time codes	Medium: fingerprints can be cloned, but a change is reliably observed
D2 Device integrity	Rooting, unlocked bootloader, emulation and instrumentation frameworks, repackaged application, debugger attached	Automated fraud at scale; farms of emulated devices	Medium to high: concealment is possible but costly and leaves artefacts
D3 Remote control	Accessibility service active and not required by the user, screen capture or casting, remote-administration channel, overlay window detected	Device takeover by a banking trojan; fraudulent "technical support"	High: the capability must be present for the attack to proceed
B1 Input dynamics	Typing rhythm, swipe geometry, pressure and handset orientation, touch area	Automated interaction; an operator other than the customer	High: reproducing an individual pattern is substantially harder than stealing a credential
B2 Navigation	Path through the application, absence of exploratory movement, direct entry to the transfer screen, use of paste where typing is habitual	Scripted interaction; a party familiar with the target field but not with the customer's habits	Medium to high
B3	Unusual dwell on	Social	High for the

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

Distress markers	the transfer screen, hesitation and correction at the beneficiary field, concurrent call, reading cadence in field entry	engineering in progress; coerced authorisation	attacker, since the behaviour is the customer's own
T1 Transaction	Amount relative to the customer's distribution, hour, channel, novelty and account age of the beneficiary, velocity, position in the payment graph	Established fraud patterns; mule account networks	Low to medium: amounts and timing are freely chosen by the attacker

Note. No signal in group T1 is sufficient on its own against the two attack classes considered here, since the payment order is legitimate in both. The detection rests on the coincidence of signals from groups D3, B1 and B3 with an ordinary-looking transaction.

Group D3 requires particular attention in implementation. An accessibility service is a legitimate facility used by customers with disabilities, and its presence may not be treated as evidence of fraud without qualification. The discriminating observation is not that the service is active but that it is active in a session whose behavioural profile has simultaneously changed, or that it appeared for the first time shortly before a high-value transfer to a new beneficiary. This is an instance of the general rule that in this problem the conjunction carries the information and the individual signal does not.

Group B3 is the only group that addresses authorised push payment fraud directly, and it is the group least represented in deployed systems. The customer under instruction behaves measurably differently from the customer paying a familiar counterparty: the session is longer, the entry of the beneficiary details is slower and more frequently corrected, and the pattern of attention alternates between the handset and an external source. These are the markers that the published case studies of behavioural deployment report as the basis of detection in this category.

Construction of the scoring model. The feature vector mixes tabular quantities with an event sequence, and the two are best treated by different means. Gradient boosting over decision trees remains the strongest and most practical estimator for the tabular part: it accommodates mixed types and missing values, requires no scaling, trains in minutes on the volumes in question, and yields feature attributions that an analyst and a regulator can inspect. The event sequence — the ordered stream of screens, fields and timings — carries information that aggregation destroys, and is better handled by a sequence model whose output is supplied to the boosted ensemble as an additional feature.

The dominant statistical difficulty is class imbalance. Fraudulent payments constitute a small fraction of one per cent of volume, so that a model predicting the majority class universally attains an accuracy indistinguishable from unity while detecting nothing. Three measures are required jointly. The objective must be cost-sensitive, weighting the classes by the costs stated above rather than by their frequencies. The sampling scheme must preserve the temporal structure: synthetic oversampling applied across a time boundary manufactures cases that could not have existed and inflates the measured result. And the validation must be temporal — trained on an earlier period and tested on a later one — because a random split allows the model to learn from the future and produces an estimate that deployment will not reproduce.

Label delay compounds the difficulty. The confirmation that a payment was fraudulent arrives days or weeks later, when the customer reports it or the receiving institution responds, and cases that are never reported remain labelled as legitimate. The consequence is that the most recent data are the least reliably labelled, precisely when they are the most relevant to the current attack mix. The practical response is to train on a window closed far enough in the past for labels to have matured, to treat the

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

open window as unlabelled and use it for drift monitoring only, and to accept that the negative class is contaminated and to weight it accordingly.

A further asymmetry is that blocked transactions have no outcome. A payment refused by the model never produces a confirmed label, so the training set contains no record of what would have happened, and the model progressively loses the ability to evaluate the region of the space it already acts upon. The standard remedy is a small randomised control — a fraction of high-score cases allowed through under close monitoring, or referred to manual review rather than refused — so that the region continues to generate evidence.

Evaluation methodology. Accuracy is uninformative under this imbalance, and the area under the ROC curve is only slightly less so, since it is dominated by the majority class and a model may improve it substantially without improving any operationally relevant quantity. Table 2 sets out the metrics that bear on the decision, each with the question it answers.

Table 2. Evaluation metrics and the operational question each answers

Metric	Operational question	Remark
Recall at fixed review capacity	Of the fraud that occurred, what proportion is caught if the review function processes m cases per day?	The primary metric; capacity is a hard constraint, not a parameter
Alert precision	Of the cases referred, what proportion are genuinely fraudulent?	Governs analyst workload and the credibility of the system with the review team
PR-AUC	How does the model perform across the operating region as a whole?	Preferred to ROC-AUC under imbalance; insensitive to the majority class
Value detection rate	What proportion of the monetary exposure is prevented, as distinct from the count of cases?	Diverges sharply from case recall when losses are concentrated in few payments
False positive rate on legitimate volume	How many genuine customers are subjected to friction per thousand payments?	Must be reported separately for each response tier, since costs differ by an order of magnitude
Time to decision	Is the score available before authorisation, within the latency budget of the payment flow?	A model that answers after settlement converts prevention into recovery
Stability over time	How far does performance degrade between retraining cycles?	Measured on a rolling temporal split; the basis of the retraining schedule

Note. The metrics are reported jointly and not reduced to a single figure. A model superior on PR-AUC but inferior on recall at the actual review capacity is the inferior model in deployment.

Evaluation must also be stratified by attack class. A model that performs well in aggregate may owe that performance entirely to the detection of unsophisticated unauthorised fraud while detecting nothing in the authorised category, which is the category that is growing. Reporting the three classes separately — unauthorised remote fraud, device takeover, and authorised push payment — is the minimum necessary to see whether the system addresses the threat that is actually present.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

Adversarial adaptation and concept drift. The data in this problem are generated by an adversary who observes the response of the system and adapts, so the stationarity assumed by standard learning theory does not hold. Drift arises from two distinct sources, and they require different responses.

Natural drift reflects change in legitimate behaviour: a new release of the application alters navigation paths, a seasonal pattern shifts the distribution of amounts, a new payment instrument changes the mix of channels. It is gradual, it is observable in the feature distributions without reference to labels, and it is addressed by scheduled retraining and by expressing features as deviations from the customer's own profile rather than as absolute values, which absorbs much of the population-level shift.

Adversarial drift reflects deliberate evasion. When a detection becomes effective, the attacking family is modified to defeat it: input timing is randomised to imitate human variance, the accessibility service is enabled only at the moment it is needed and disabled immediately afterwards, amounts are fragmented below the thresholds the attacker has inferred by probing. This drift is abrupt, it is not visible in aggregate feature distributions because it is confined to the fraudulent subpopulation, and it is detected by monitoring the score distribution of confirmed cases rather than of all traffic. A sudden fall in the mean score of confirmed fraud is the signature of a successful evasion and is the operative alarm.

Three design consequences follow. Thresholds must not be published or inferable: a constant threshold is discovered by probing within days, and a degree of randomisation in the decision boundary raises the cost of that reconnaissance considerably. Detection must not rest on a single signal, since a single signal defines a single evasion target, whereas a conjunction across groups requires the adversary to defeat several mechanisms simultaneously. And the signals of group D3 are comparatively durable precisely because the capability they detect is necessary to the attack: the trojan cannot conceal the accessibility permission and still use it.

Requirements of the legislation of the Republic of Uzbekistan. The signals on which the model depends are personal data, and several of them — behavioural dynamics in particular — are capable of identifying an individual. Their collection and processing are governed in the Republic of Uzbekistan by the Law on Personal Data, which establishes the requirements of lawful basis, purpose limitation, proportionality of the data collected to the stated purpose, and the localisation of processing. The Law on Cybersecurity establishes the system of cybersecurity subjects and the obligations of operators of critical information infrastructure, within which the banking system falls, and the Law on Payments and the Payment System governs the duties of payment service providers in the execution and suspension of payment orders.

Three requirements follow for the design. The first is that the lawful basis for behavioural telemetry must be established explicitly, and the purpose stated narrowly as fraud prevention; data collected for that purpose may not be repurposed for marketing or scoring without a separate basis. The second is proportionality: the model should collect the derived features it requires — timing statistics, deviation measures — rather than raw keystroke content, which is both more intrusive and no more informative for the purpose. The third is that an automated decision with legal or similarly significant effect on the customer, which the refusal of a payment plainly is, requires a route to human review and an explanation; this is the operational justification, independent of any statistical one, for retaining an interpretable model class and for the graduated policy in which refusal is the exception rather than the standard response.

Retention is the remaining question. Session telemetry is voluminous and its evidentiary value decays, but the investigation of a reported case may require the session record weeks after the event. The resolution adopted in practice is a two-tier schedule: full event detail retained for a short defined period sufficient to cover the reporting window, and derived aggregate features retained longer for model training, with the identifiers separated from the behavioural record wherever the training does not require them.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

Discussion. Three conclusions of practical consequence follow from the analysis.

The first is that the unit of analysis must change. So long as the payment order is the object of scrutiny, the two dominant attack classes are undetectable in principle, because in both the order is genuine. Moving the unit of analysis to the session does not merely improve detection at the margin; it is what makes detection possible at all in these categories. The corollary is that a bank whose anti-fraud capability consists of transaction rules is not weakly protected against these attacks but unprotected.

The second is that the binding constraint is the capacity of the review function and not the discrimination of the model. The quantity that determines prevented loss is recall at the capacity that exists, and a model that improves PR-AUC while referring more cases than can be examined reduces prevented loss. The graduated policy is the mechanism by which the constraint is respected, since step-up authentication consumes no analyst time and therefore extends the effective reach of the system far beyond the review queue.

The third is that detection in this domain is a conjunction and not a classification. No individual signal in Table 1 distinguishes the attacks considered here with acceptable precision, and systems built on a single family of signals fail in characteristic ways: device intelligence alone penalises legitimate accessibility users, behavioural analysis alone is defeated by a patient operator, and transaction rules alone see nothing. The detection lies in the coincidence of an ordinary transaction with an extraordinary session.

Several limitations should be stated. The cost model is formulated at a level of generality intended to apply across institutions, whereas the actual costs — particularly the abandonment probability associated with each response tier — are specific to a customer population and must be estimated locally before the thresholds can be set. The taxonomy of Table 1 reflects the attack families observed at the time of writing and will require revision as those families adapt. The model has not been validated on an Uzbek banking dataset, for the reason that the labelled session telemetry required for such a validation is not publicly available, and the figures cited are drawn from jurisdictions that publish consistent fraud statistics.

These limitations indicate the direction of further work: the estimation of the cost parameters on a domestic customer population, the construction of a labelled session dataset under the constraints of the Law on Personal Data, and an empirical comparison of the fusion rules proposed here against the transaction-only baseline on that dataset.

Conclusion. This article has examined the detection of fraud in mobile banking and has argued that the two attack classes which now dominate — remote control of the customer device and the social engineering of an authorised payment — produce transaction records that are indistinguishable from legitimate ones, so that the evidence of fraud must be sought in the session rather than in the payment order.

A four-layer reference model was constructed in which device, behavioural and transactional signals are assembled into a session-context feature vector, a risk score r(s, t) is computed before authorisation, and the score is mapped to a graduated policy of four responses. The decision stage was formulated as the minimisation of expected cost subject to the capacity of the review function, and it was shown that this constraint, rather than the discriminative power of the model, governs the loss actually prevented.

A taxonomy of session signals was given in which each signal was matched to the attack class it detects and to its resistance to falsification, and it was shown that no individual signal is sufficient against the attacks in question, so that detection rests on the conjunction of an ordinary transaction with an anomalous session. The construction of the model under extreme class imbalance and delayed, contaminated labelling was described, and an evaluation protocol was specified in which recall at a fixed review capacity, stratified by attack class, is the operative metric. Finally, the requirements imposed by the legislation of the Republic of Uzbekistan on the collection of behavioural telemetry were stated, together with their consequences for model class, retention and the right to human review.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

References

1. Annual Fraud Report 2026: the definitive overview of payment industry fraud in 2025. – London: UK Finance, 2026.
2. UK payment fraud losses topped £1.28bn in 2025, new UK Finance report finds. – URL: https://www.fintechfutures.com/financial-crime-fraud-prevention/uk-payment-fraud-losses-topped-1-28bn-in-2025-new-uk-finance-report-finds (accessed 4 November 2026).
3. UK Finance Fraud Report 2026: payment fraud losses reach £1.28 billion. – URL: https://www.neopay.co.uk/uk-finance-fraud-report-2026-payment-fraud-losses-reach-1-28-billion/ (accessed 4 November 2026).
4. RemControl Banking Trojan Gives Attackers Remote Control of Android Devices. – Infosecurity Magazine, 2026. – URL: https://www.infosecurity-magazine.com/news/banking-trojan-remote-control/ (accessed 4 November 2026).
5. BankBot YNRK Trojan Reinforces Mobile Banking Threats. – Zimperium Mobile Threat Watch, 2026. – URL: https://zimperium.com/blog/mobile-threat-watch/new-android/bankbot-ynrk-trojan-reinforces-mobile-banking-threats (accessed 4 November 2026).
6. Anatsa Banking Trojan Targeting Android Devices: advisory. – Abuja: Nigeria Computer Emergency Response Team, 2026. – URL: https://cert.gov.ng/advisories/anatsa-banking-trojan-targeting-android-devices (accessed 4 November 2026).
7. Preventing Fraud on Mobile Devices Using Behavioral Biometrics: white paper. – Tel Aviv: BioCatch, 2025. – URL: https://www.biocatch.com/resources/white-paper/prevent-fraud-mobile-devices-using-biometric-authentication (accessed 4 November 2026).
8. Top LATAM bank reduces social engineering fraud targeting mobile users with behavioral biometrics: case study. – BioCatch, 2025. – URL: https://www.casestudies.com/company/biocatch/case-study/top-latam-bank-reduces-social-engineering-fraud-targeting-mobile-users-by-67-with-behavioral-biometrics (accessed 4 November 2026).
9. Real-Time Fraud Detection Using AI-Powered Behavioral Biometrics. – GeekyAnts, 2025. – URL: https://geekyants.com/blog/real-time-fraud-detection-using-ai-powered-behavioral-biometrics (accessed 4 November 2026).
10. Dal Pozzolo A., Boracchi G., Caelen O., Alippi C., Bontempi G. Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy // IEEE Transactions on Neural Networks and Learning Systems. – 2018. – Vol. 29, No. 8. – P. 3784–3797.
11. Dal Pozzolo A., Caelen O., Johnson R.A., Bontempi G. Calibrating Probability with Undersampling for Unbalanced Classification // Proceedings of the IEEE Symposium Series on Computational Intelligence. – 2015. – P. 159–166.
12. Bhattacharyya S., Jha S., Tharakunnel K., Westland J.C. Data mining for credit card fraud: A comparative study // Decision Support Systems. – 2011. – Vol. 50, No. 3. – P. 602–613.
13. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2016. – P. 785–794.
14. Ke G., Meng Q., Finley T. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Advances in NeurIPS. – 2017. – P. 3146–3154.
15. Prokhorenkova L., Gusev G., Vorobev A. et al. CatBoost: unbiased boosting with categorical features // Advances in NeurIPS. – 2018. – P. 6638–6648.
16. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic Minority Over-sampling Technique // Journal of Artificial Intelligence Research. – 2002. – Vol. 16. – P. 321–357.
17. Elkan C. The Foundations of Cost-Sensitive Learning // Proceedings of IJCAI. – 2001. – P. 973–978.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

18. Saito T., Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets // PLoS ONE. – 2015. – Vol. 10, No. 3. – e0118432.
19. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation // ACM Computing Surveys. – 2014. – Vol. 46, No. 4. – Article 44.
20. Frank M., Biedert R., Ma E., Martinovic I., Song D. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication // IEEE Transactions on Information Forensics and Security. – 2013. – Vol. 8, No. 1. – P. 136–148.
21. Sitová Z., Šeděnka J., Yang Q. et al. HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users // IEEE Transactions on Information Forensics and Security. – 2016. – Vol. 11, No. 5. – P. 877–892.
22. Lundberg S.M., Lee S.-I. A Unified Approach to Interpreting Model Predictions // Advances in NeurIPS. – 2017. – P. 4765–4774.
23. Weber M., Domeniconi G., Chen J. et al. Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics // arXiv:1908.02591. – 2019.
24. Guidelines on the security measures for operational and security risks of payment services under PSD2, EBA/GL/2017/17. – Paris: European Banking Authority, 2017.
25. Payments Fraud Report 2025. – Frankfurt: European Banking Authority and European Central Bank, 2025.
26. Digital Identity Guidelines: NIST Special Publication 800-63-3. – Gaithersburg: National Institute of Standards and Technology, 2017.
27. OWASP Mobile Application Security Verification Standard, version 2.0. – Open Worldwide Application Security Project, 2024.
28. On Personal Data: Law of the Republic of Uzbekistan of 2 July 2019, No. ZRU-547 (as amended). – URL: https://lex.uz/ru/docs/4396428
29. On Cybersecurity: Law of the Republic of Uzbekistan of 15 April 2022, No. ZRU-764. – National Database of Legislation of the Republic of Uzbekistan.
30. On Payments and the Payment System: Law of the Republic of Uzbekistan of 1 November 2019, No. ZRU-578 (as amended). – URL: https://lex.uz/docs/4581969
31. On Informatisation: Law of the Republic of Uzbekistan of 11 December 2003, No. 560-II (as amended). – URL: https://lex.uz/ru/docs/82956
32. On the Strategy of Development of New Uzbekistan for 2022–2026: Decree of the President of the Republic of Uzbekistan of 28 January 2022, No. UP-60.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.