applied sciences

MDPI

Article

Research on Attack Detection for Traffic Signal Systems Based on Game Theory and Generative Adversarial Networks

Kailong Li [1,*], Ke Pan [2], Weijie Xiu [1], Min Li [1], Zhonghe He [1] and Li Wang [1,*]

[1] Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology, Beijing 100144, China; xiuwj@ncut.edu.cn (W.X.); liminncut@ncut.edu.cn (M.L.); zhonghehe@ncut.edu.cn (Z.H.)

[2] Traffic Information Engineering Institute, Guangxi Transport Vocational and Technical College, Nanning 530004, China; panke@mail.ncut.edu.cn

* Correspondence: leeph.d2023@mail.ncut.edu.cn (K.L.); its-leel@aliyun.com (L.W.)

Abstract: With the rapid development of intelligent transportation systems and information technology, the security of road traffic signal systems has increasingly attracted the attention of managers and researchers. This paper proposes a new method for detecting attacks on traffic signal systems based on game theory and Generative Adversarial Networks (GAN). First, a game theory model was used to analyze the strategic game between the attacker and the defender, revealing the diversity and complexity of potential attacks. A Bayesian game model was employed to calculate and analyze the attacker's choice of position. Then, leveraging the advantages of GAN, an adversarial training framework was designed. This framework can effectively generate attack samples and enhance the robustness of the detection model. Using empirical research, we simulated the mapping of real traffic data, road network data, and network attack data into a simulation environment to validate the effectiveness of this method. In a comparative experiment, we contrasted the method proposed in this paper with the traditional Support Vector Machine (SVM) algorithm, demonstrating that the model presented here can achieve efficient detection and recognition across various attack scenarios, with significantly better recall and F1 scores compared to traditional methods. Finally, this paper also discusses the application prospects of this method and its potential value in future intelligent transportation systems.

Keywords: traffic signal system; Nash game; Bayesian equilibrium; GAN

check for updates

Citation: Li, K.; Pan, K.; Xiu, W.; Li, M.; He, Z.; Wang, L. Research on Attack Detection for Traffic Signal Systems Based on Game Theory and Generative Adversarial Networks. *Appl. Sci.* **2024**, *14*, 9709. https://doi.org/10.3390/app14219709

Academic Editor: David Megías

Received: 14 August 2024
Revised: 3 October 2024
Accepted: 9 October 2024
Published: 24 October 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Traffic signal control systems are one of the key measures to regulate traffic flow and address urban traffic congestion. In the current context of urban smart development and the advancement of traffic information technology [1], the efficient operation of these systems will require more detection devices and computational resources [2]. However, information security prevention is often overlooked; in other words, future issues related to information security could increasingly jeopardize the safety of traffic signal control systems. Li et al. [3,4] discovered and demonstrated that there are many security flaws in the networked traffic signal systems currently deployed in the United States due to system failures by designers. They proposed a game-theoretic framework that can promote cybersecurity while facilitating efficient traffic management. Haddad et al. [5,6] proposed a boundary control algorithm, based on MFD, that can continuously and adaptively supplement the effects of network attacks. Khattak Z H et al. [7] analyzed the security vulnerabilities of Active Traffic Management (ATM) systems in Virginia, offering methods to determine whether the real-time monitoring systems are under attack. Other security researchers have also identified various vulnerabilities in traffic light systems [8–10]. Ahmad F et al. [11] categorized years of research related to traffic games into several directions, including flow management, behavioral interactions, routing operations, and transportation safety,

noting that substantial research opportunities still exist in the area of information security for urban traffic control systems. Cheng P et al. [12] proposed an intelligent traffic detection and recognition method based on an adaptive unknown input observer. Ruixiao Sun et al. [13] developed an anomaly detection model for traffic networks that employs non-parametric Long Short-Term Memory (LSTM) and parametric Gaussian processes for online data-driven analysis. C. Zhang et al. [14] proposed a network intrusion detection method based on deep learning, which uses circumvention technology to enhance resistance to attacks. At the same time, it is clear that in order to reduce the damage caused by malicious attacks on the Internet-connected traffic flow detector equipment at intersections, the traffic flow of malicious nodes in the Internet of Things can be identified. Yunyun Hou et al. [15] proposed a method for detecting malicious nodes in the Internet of Things based on autoencoders. Compared to traditional K-means and Gaussian mixture model approaches, the proposed method demonstrated improved model detection performance. Xin Huang et al. [16] designed a novel adaptive attack compensator that allows a fuzzy controller equipped with a compensator to ensure the stability of the attacked system and the required control performance. The main problems and research results are summarized in Table 1.

Table 1. Summary of existing research results.

Problem Summary	Research Basis
System defect	System security vulnerabilities; issues with insufficient network security facilities. For example, there is a need to enhance the construction of network security facilities.
Hardware is attacked	System security vulnerabilities; issues with insufficient network security facilities. For example, there is a need to enhance the construction of network security facilities.
The system is attacked	System security vulnerabilities; issues with insufficient network security facilities. For example, there is a need to enhance the construction of network security facilities.

In summary, during the anomaly detection process, most researchers utilizing autoencoder-based methods through dimensionality reduction and clustering techniques are able to effectively distinguish between normal and abnormal traffic. However, there remain challenges in enhancing real-time capabilities and adapting to more complex attack scenarios. Issues such as how to respond to more hidden and complex attack methods, how to improve the system's robustness and adaptability, and how to better integrate IoT technology with security control strategies need to be addressed. Future research should focus on further improving and advancing these technologies to ensure the security and reliability of road traffic signal systems.

This article focuses on the scientific research on attack detection in urban traffic signal control systems, with the main goals as follows:

(1) Since defenders cannot accurately understand the attack locations, this article transforms the game into a complete yet imperfect information game for analysis. It is necessary to calculate the probability of the TSM and UC being attacked, solve for the pure-strategy Bayesian–Nash equilibrium and mixed-strategy Bayesian–Nash equilibrium points of the attacker, clarify the true attack locations, and further select defense measures.

(2) A generative adversarial network model is used to learn traffic flow data and current detection data to achieve anomaly detection and data recovery, generating three common types of network attack behaviors: denial of service, tampering attacks, and replay attacks. This approach aims to generate key target data attackers use to influence traffic control for their own purposes, while also achieving comparative effectiveness.

2. Attack Position Assessment

2.1. Bayesian Game Model

The process of offense and defense in traffic signal systems involves incomplete information and uncertainty. Bayesian games can transform this uncertainty into probability distributions. This allows each decision-maker to choose the optimal decision strategy based on probabilistic judgments of the behaviors of other decision-makers. Bayesian games can help analyze the interactions and influences among various decision-makers under conditions of incomplete information and uncertainty, leading to the identification of optimal strategies. Additionally, Bayesian games can provide a quantitative evaluation of the likelihood of survey results, rather than simply accepting them completely or rejecting them entirely, as is common in traditional decision-making methods.

In urban traffic control, traffic signal controllers are positioned at intersections and are primarily responsible for receiving data from upper management systems and various traffic detectors. They control the operation of traffic lights based on these instructions. Generally, the upper management system is installed at traffic management agencies, coordinating with traffic signal controllers through dedicated network resources.

Assumption 1: Both the attackers and defenders participating in the game are purely rational. Therefore, the traffic network security value S must exceed any costs; otherwise, there would be no incentive for either party to adopt strategies aimed at enhancing security. Thus,

$$S > max(Cau, Cas, Cdc, Cdp, Cdcp) \tag{1}$$

where Cau denotes the cost of attacking the UC; Cas denotes the cost of attacking the TSM; Cdx represents the cost corresponding to the defense strategy of $dx \in \{dcp, dc, dp\}$, dcp means using information and control strategy defense at the same time, dc means using only information strategy defense, and dp means using information and control strategy defense.

Next, we will discuss the benefits derived from the attack positions under various strategies of both parties in the game.

When both parties in the game choose strategy (a, nd), the attacker benefits $S - Cau$, while the defender also benefits $-S$. When the strategy is (a, dx), the defender benefits $\alpha S - (1-\alpha)S - Cdx = (2\alpha - 1)S - Cdx$, and the attacker's benefit is the defender's benefit minus the cost of the attack: $(1-\alpha)S - \alpha S - Cau = (1-2\alpha)S - Cau$. In the case of (na, dx), the detection system has an error, resulting in the defender receiving benefit: $-Cdx$. The strategy benefits of attacking the UC or TSM can be obtained, as shown in Table 2.

Table 2. Payoff matrix.

	dcp	***dc***	***dp***	***nd***
a	$(1-2\alpha)\ S - Cx,$ $(2\alpha - 1)\ S - Cdcp$	$(1-2\alpha)\ S - Cx,$ $(2\alpha - 1)\ S - Cdc$	$(1-2\alpha)\ S - Cx,$ $(2\alpha - 1)\ S - Cdp$	$S - Cx,$ $-S$
na	$0, -\beta Cdcp$	$0, -\beta Cdc$	$0, -\beta Cdp$	$0, 0$

In the payoff matrix, Cx = Cau in case UC, and Cx = Cas in case TSM. N_a denotes the attacker, the strategy includes $\{a, na\}$, a means attacking, and na means not attacking; N_d denotes the defender of the system, including UC and TSM, the defense strategy includes $\{dcp, dc, dp, nd\}$, and nd means not defending; S denotes the value of the security of the transportation network; α represents the detection rate; β represents the false alarm rate.

2.2. Calculation of Attack Position

The defender cannot accurately understand the attacker's position, which constitutes an incomplete information game. This paper employs the Harsanyi transformation method to convert the game into a complete but imperfect information game for analysis [17]. The defender is aware that the attacker has a probability p of targeting the UC and a probability

$1-p$ of targeting the TSM. Based on Tables 1 and 2, the extended form of the Bayesian game is illustrated in Figure 1.

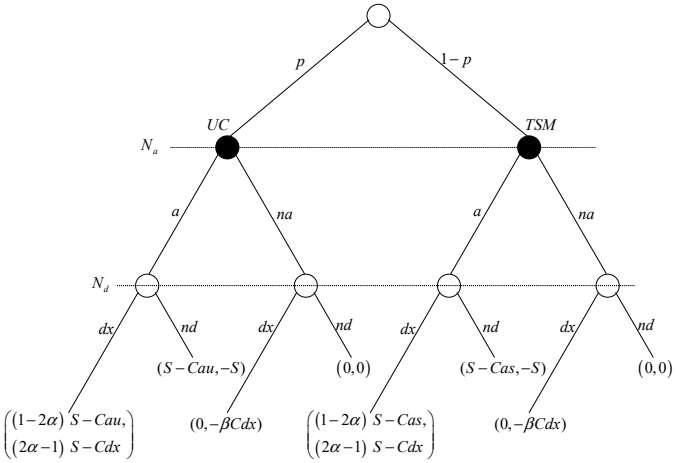


Figure 1. The Bayesian game in an extensive form.

(1) When the attacker chooses the pure strategies of (a, a) and (na, na), there is no Bayesian–Nash equilibrium for pure strategies and no Bayesian–Nash equilibrium for mixed strategies.

Proof. When attacker N_a adopts a pure strategy against dx, the expected payoff for defender N_a when implementing strategy (a, a) is as follows:

$$U(dx) = (2a-1)S - Cdx \tag{2}$$

$$U(nd) = -S \tag{3}$$

In this case, the expected returns $U(dx)$ and $U(nd)$ for system defender N_d are independent of the probability p of attack locations, regardless of whether defensive strategies are adopted. Therefore, both $((a, a), dx)$ and $((a, a), nd)$ are not Bayesian–Nash equilibria.

When attacker N_a uses a pure strategy against (na, na), the expected payoff for defender N_d when adopting that strategy dx is as follows:

$$U(dx) = -\beta Cdx \tag{4}$$

$$U(nd) = 0 \tag{5}$$

Similarly, $((na, na), dx)$ and $((na, na), nd)$ are not Bayesian–Nash equilibria. □

(2) When $p > \frac{\beta C dx_m}{2\alpha S - (1-\beta) C dx_m}$ and $Cas > (1-2\alpha)S > Cau$, there exists a pure strategy $((a, na), dx_m, p)$ as a Bayesian–Nash equilibrium; when $p \leq \frac{\beta C dx_m}{2\alpha S - (1-\beta) C dx_m}$, there exists a mixed-strategy Bayesian–Nash equilibrium, with the attack position being the UC.

Proof. When attacker N_a employs pure strategy (a, na), then

$$U(dx) = (2\alpha - 1)pS - (1-\beta)pCdx - \beta Cdx \tag{6}$$

$$U(nd) = -pS. \tag{7}$$

Let $\max(U(dx_m)) > U(nd)$, brought into Equations (6) and (7), have

$$p > \frac{\beta C dx_m}{2\alpha S - (1-\beta) C dx_m}. \tag{8}$$

The defender's dominant strategy is dx_m. According to Tables 1 and 2, when $Cas > (1-2\alpha)S > Cau$, the attacker's dominant strategy is (a, na). So, $((a, na), dx_m, p)$ is a pure-strategy Bayesian–Nash equilibrium.

Let $\max(U(dx_m)) \leq U(nd)$, brought into Equations (6) and (7), have

$$p \leq \frac{\beta C dx_m}{2\alpha S - (1-\beta) C dx_m} \tag{9}$$

The defender's dominant strategy is nd. Due to $\alpha > 0$, according to Tables 1 and 2, the attacker's dominant strategy is (a, a). So $((a, na), dx_m, p)$ is not a pure-strategy Bayesian–Nash equilibrium.

Next, we will discuss mixed-strategy games. Let the attacker adopt attack strategies a and na with probabilities p_a and $(1-p_a)$, respectively, while the defender employs defense strategies dx and nd with probabilities p_{dx} and $(1-p_{dx})$, respectively; $p_{dx} + p_{nd} = 1$.

When attacker N_a uses strategy (a, na), the expected return for each attack strategy is as follows:

$$U(a) = pS(1 - 2\alpha p_{dx}) - pCau \tag{10}$$

$$U(na) = 0 \tag{11}$$

Let $U(na) = U(a)$ be brought into Equations (10) and (11), as follows:

$$p^*_{dx} = \frac{S - Cau}{2\alpha S} \tag{12}$$

When attacker N_a uses strategy (a, na), the expected returns for each defense strategy are as follows:

$$U(dx) = (2\alpha - 1) p p_a S - (1-\beta) p p_a C dx - \beta C dx \tag{13}$$

$$U(nd) = -p p_a S \tag{14}$$

Let $U(dx) = U(nd)$ be brought into Equations (13) and (14), as follows:

$$p^*_a = \frac{\beta C dx}{2\alpha p S - (1-\beta) p C dx} \tag{15}$$

According to the definition of the mixed-strategy Nash equilibrium, the attacker targets the UC with probability p^*_a, while the defender responds with probability p^*_{dx}. The mixed-strategy Nash equilibrium occurs when the chosen defensive strategy dx_m has the lowest cost. □

(3) When $p < \frac{2\alpha S - C dx_m}{2\alpha S - (1-\beta) C dx_m}$ and $Cas < (1-2\alpha)S < Cau$, there exists a pure strategy $((na, a), dx_m, p)$ as a Bayesian–Nash equilibrium; when $p \geq \frac{2\alpha S - C dx_m}{2\alpha S - (1-\beta) C dx_m}$, there exists a mixed-strategy Bayesian–Nash equilibrium, with the attack position being the TSM.

Proof. When attacker N_a uses pure strategy (a, na),

$$U(dx) = p(-\beta C dx) + (1-p)((2\alpha - 1)S - C dx) \tag{16}$$

$$U(nd) = (p-1)S \tag{17}$$

Let $\max(U(dx_m)) > U(nd)$ be brought into Equations (16) and (17), as follows:

$$p < \frac{2\alpha S - Cdx_m}{2\alpha S - (1-\beta)Cdx_m} \tag{18}$$

The defender's dominant strategy is dx_m. According to Tables 1 and 2, when $Cas < (1-2\alpha)S < Cau$, the attacker's dominant strategy is (na, a). So, $((na, a), dx_m, p)$ is a pure-strategy Bayesian–Nash equilibrium.

Let $\max(U(dx_m)) \leq U(nd)$, brought into Equations (16) and (17), have

$$p \geq \frac{2\alpha S - Cdx_m}{2\alpha S - (1-\beta)Cdx_m} \tag{19}$$

The defender's dominant strategy is nd. Due to $\alpha > 0$, according to Tables 1 and 2, the attacker's dominant strategy is (a, a). So $((na, a), dx_m, p)$ is not a pure-strategy Bayesian–Nash equilibrium.

Similarly, discuss mixed strategy games.

When attacker N_a uses strategy (na, a), the expected return for each attack strategy is as follows:

$$U(a) = (1-p)(-2\alpha p_{dx})S - (1-p)Cau \tag{20}$$

$$U(na) = 0 \tag{21}$$

Let $U(na) = U(a)$ be brought into Equations (20) and (21), as follows:

$$p^*_{dx} = \frac{-Cau}{2\alpha S} \tag{22}$$

When attacker N_a uses strategy (na, a), the expected returns for each defense strategy are as follows:

$$U(dx) = (1-p)(2\alpha - 1)Sp_a - \beta Cdx - (1-\beta)p_a Cdx \tag{23}$$

$$U(nd) = (p-1)p_a S \tag{24}$$

Let $U(dx) = U(nd)$ be brought into Equations (23) and (24), as follows:

$$p^*_a = \frac{\beta Cdx}{(1-p)2\alpha S + (1-\beta)Cdx} \tag{25}$$

According to the definition of the mixed-strategy Nash equilibrium, the attacker targets the TSM with probability p^*_a, while the defender responds with probability p^*_{dx}. The mixed-strategy Nash equilibrium occurs when the chosen defensive strategy dx_m has the lowest cost. □

3. Modeling and Dataset Building

3.1. Traffic GAN Model

In information security, attackers can bypass security systems and obtain legitimate identities through methods such as system vulnerabilities and backdoors to maliciously alter data. Therefore, it is essential to develop a method that treats the entire system as a black box, where it is possible to determine if the system has been attacked by analyzing the input and output data from the traffic control system. This method should be able to identify the locations of attacks within the system and restore any abnormal data. Utilizing Generative Adversarial Network (GAN) models can facilitate the learning of relationships between control signal data and current detection data for anomaly detection and data restoration, as illustrated in Figure 2.

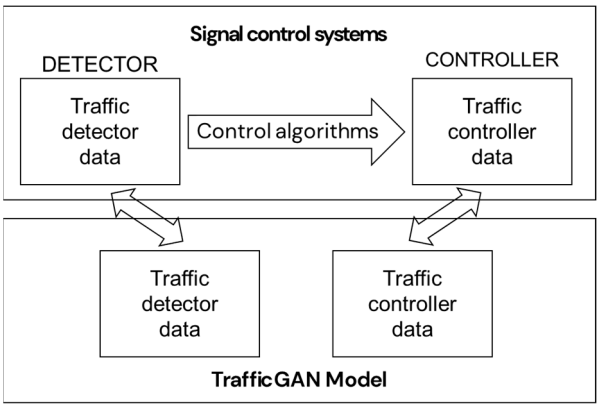


Figure 2. Sample Data Formation.

We combined traffic detection and traffic control data. Traffic detection data refers to the basic traffic parameters collected by traffic detection devices during normal operation, including the volume of traffic flow, roadway occupancy rate, queue length, travel speed, and vehicle spacing. Traffic control data consists of the commands used by signal control devices to adjust vehicle movement during normal operation, including signal cycles, green light duration, and phase differences. These two sets of data have a temporal and spatial correlation, and the size of the dataset is related to the physical boundaries of the selected road network. The signal control data is calculated by the control algorithm according to the current traffic condition, so there is a certain relationship between the control signal data and the current detection data. Therefore, it is possible to select associated data as training samples to train the GAN model. This approach has several advantages:

Data Completion: If certain data in the system (such as detectors or control signals) are abnormal, we can restore them using the trained generator in the GAN model.

Data Prediction: Given that traffic flow exhibits spatiotemporal correlations, the GAN model can also be employed for data prediction.

Anomaly Detection: If certain data in the system (including detection data, control signals, and system time) becomes abnormal due to equipment failure or malicious attacks, we can identify these issues using the trained discriminator in the GAN model.

The existing sample data is x and the generated sample data is $G(z)$, with the generation model being G and the discriminative model being D. The objective function of the GAN is as shown in Equation (26):

$$\min_{G}\max_{D}\left\{\underset{x\sim P_r}{E}[\log(D(x))]+\underset{z\sim P_z}{E}[\log(1-D(G(z)))]\right\} \tag{26}$$

The optimization process of the GAN algorithm is a minimax game involving two neural networks: the discriminator and the generator. When training the discriminator, the first term indicates that a higher output is better when the input is real, while the second term suggests that a lower output is preferable when the input is fake. In contrast, when training the generator, the second term signifies that the generator aims to maximize the output of the discriminator. To achieve victory, these two participants in the game must continually enhance their generative and discriminative capabilities until they reach a state of Nash equilibrium.

3.2. Traffic Flow Sample Data Collection

Before collecting samples, it is necessary to establish a mapping relationship between the numbered sections of the actual city traffic road network and the positions in the sample matrix. As shown in Figure 3, if a situation arises during the mapping process where a road segment does not exist in reality, zero padding can be applied.

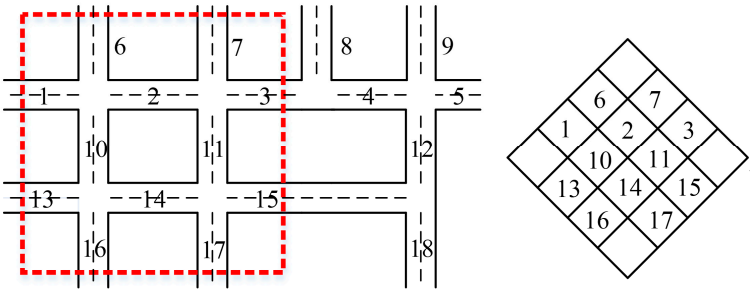
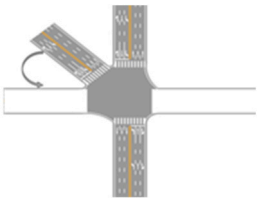
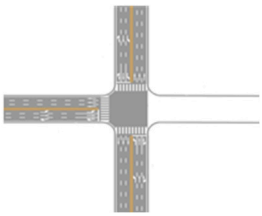
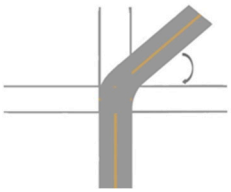


Figure 3. Link Number Mapping Relationship.

Due to the complex terrain of cities and the varied traffic demands, urban traffic networks often involve many uniquely shaped intersections during their design and construction. The road segments associated with these intersections require specific handling when establishing mapping relationships, with the processing rules outlined in Table 3.

Table 3. Image Processing of Special-shaped Links.

Type	Prototype	Optimized	Remarks
Y-shaped intersection			Move inclined sections to horizontal
T-shaped intersection			Perform zero-padding for traffic data where sections do not exist
Boundary segment			Move inclined sections to horizontal and perform zero-padding for non-existent sections

This article utilizes VISSIM's secondary development technology to obtain the number of queued vehicles at key phases for each road segment within any given simulation cycle. This data is used to create a detection data matrix, which serves as input to calculate the corresponding green light time data matrix for each road segment. The dimensions of the detection data matrix match those of the control data matrix, allowing the two matrices to be combined into a single sample matrix, as illustrated in Figure 4.

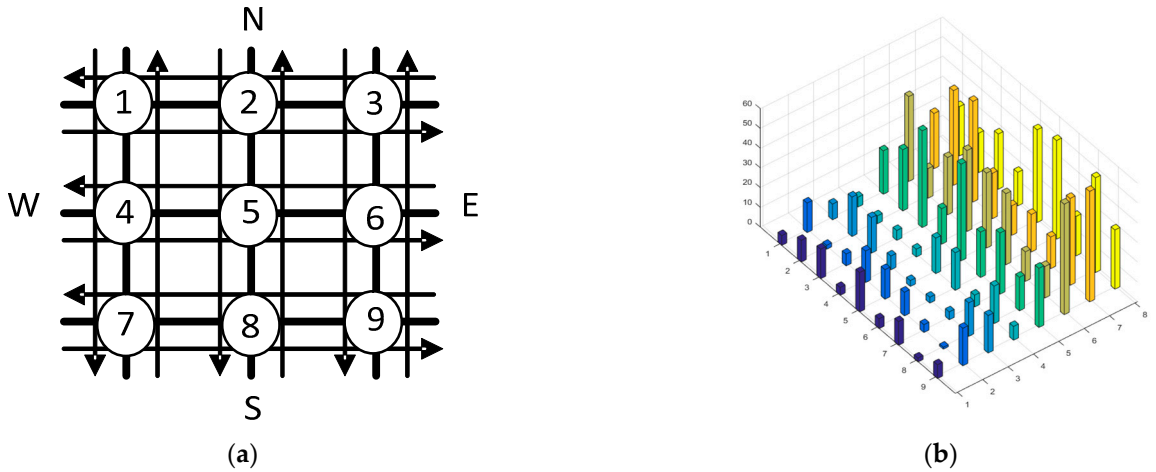


Figure 4. Schematic diagram of sample data. (**a**) Simulation Road Network Topology Diagram, the number represents the simulation intersection number; (**b**) Detection Data Matrix and Control Data Matrix Graph, the bar in blue in the front is the input detection data, and the bar in yellow is the output control data.

Here, (a) is a digital schematic diagram of the simulated road network, primarily reflecting the physical relationships between nine intersections within the network. The control methods include single-intersection signal control, arterial coordinated signal control, and multi-intersection sub-region coordinated control, fully illustrating the variations in traffic flow at actual intersections. (b) is a concatenated matrix of the queue lengths at the inbound lanes of each intersection and the signal timing for various phases in the simulated road network. To facilitate a unified feedback loop for the GAN model's data input and output, longer queue phases should correspond to longer passage durations.

By modifying the random number seed in VISSIM and simulating traffic conditions during different periods, such as peak and low times, the above steps were repeated to collect samples. Utilizing data augmentation techniques, a total of 3000 sample sets were generated to prepare for the training of the GAN model.

3.3. Attack Sample Data Collection

The attack data primarily comprised three common types of network attack behaviors: denial of service, tampering attacks, and replay attacks. Analyzing the control process of traffic signal systems revealed that the main targets for attackers, who aim to influence traffic control for their own purposes, are timing $g(k+1)$, phase differences $offset(k+1)$, and detection of queue lengths $l(k+1)$, where k is the current moment. By altering these factors, attackers can quickly and directly impact intersection control effectiveness. The data types are shown in Table 4.

Table 4. Illustration of Attack Data.

Attack Behaviors	Data Set
Denial of Service	$g(k+1) = null$ $l(k+1) = null$ $offset(k+1) = null$
Tampering attacks	$g(k+1) = false(g(k+1))$ $l(k+1) = false(l(k=1))$ $offset(k+1) = false(offset(k+1))$
Replay attacks	$g(k+1) = g(k-t), t \in [0,k]$ $l(k+1) = l(k-t), t \in [0,k]$ $offset(k+1) = offset(k-t), t \in [0,k]$

Denial of service (DoS) attacks disrupt and block normal communication between system components by flooding the communication channel with a large volume of meaningless access requests [18,19].

Tampering (or spoofing) attacks involve an attacker eavesdropping on and recording communication data over a period to avoid detection when injecting external inputs. Later, they resend the recorded data to deceive the system to achieve their goals, as seen with the Stuxnet virus [20,21].

Replay attacks involve malicious actions that disrupt or damage the system's normal operation by fully injecting or propagating data from a prior time period.

The simulation environment randomly utilized two types of attack models, centralized and distributed, to analyze the relationship between data changes and to collect relevant data, thereby validating the effectiveness of the proposed algorithm in attack detection and data restoration. We focused on the detection of anomalous data, so we chose the three most common types of attacks: denial of service, tampering, and replay attacks. The system we built randomly generates attack types and frequencies to validate the detection capability of the trained system.

4. Analysis of Experimental Results

4.1. Determining the Detection Threshold

Since the output of the discriminator ranges from 0 to 1, with values closer to 1 indicating that the input data is more likely to be genuine, selecting an appropriate threshold for determining the authenticity of the input is crucial. The choice of the detection threshold can refer to the general methods for selecting classifier thresholds in machine learning, which involves introducing four parameters:

False Positive (FP), the number of abnormal samples incorrectly identified as normal;
True Positive (TP), the number of normal samples correctly identified as normal;
False Negative (FN), the number of normal samples incorrectly identified as abnormal;
True Negative (TN), the number of abnormal samples correctly identified as abnormal.

For a classifier, different thresholds applied to the same test sample can result in a confusion matrix, can calculate the corresponding true positive rate and false positive rate.

True positive rate (TPR) refers to the proportion of normal samples that are correctly identified as normal.

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \tag{27}$$

Here, P represents the number of normal samples in the test set.

The false positive rate (FPR) refers to the proportion of abnormal samples that are incorrectly identified as normal.

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} \tag{28}$$

Here, N represents the number of anomalous samples in the testing set. According to the above formulas, it is evident that both True Positive Rate (TPR) and False Positive Rate (FPR) are fixed for the same test set. Naturally, we hope to maximize TPR and minimize FPR. By plotting the error rate and detection rate corresponding to each threshold as the x and y coordinates, we can obtain the ROC curve, as shown in Figure 5. From the figure, we can see that the closer the ROC curve is to the top left corner, the higher the TPR and the lower the FPR. Therefore, the threshold corresponding to this balance point is what we are looking for, and this balance point selection is shown in Figure 5.

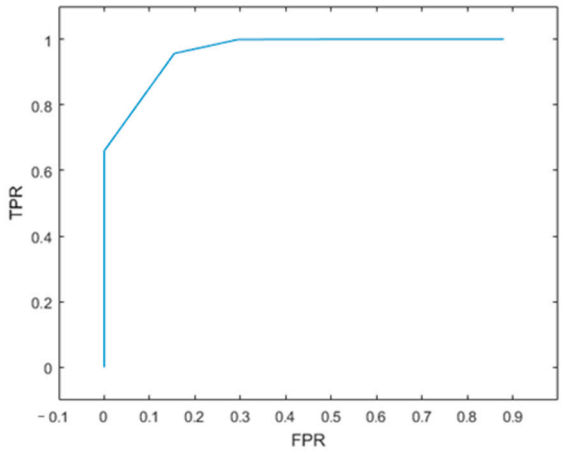


Figure 5. ROC curve.

The main basis for classifying abnormal data is the spatiotemporal correlation of traffic flow data and signal control data. Therefore, when using the trained model to identify abnormal data during different time periods in the same road environment, the classifier leverages the P and N in the test data set to identify anomalies, which is beneficial for quickly finding a balance point.

4.2. Comparison of Experimental Results

The Support Vector Machine (SVM) model [22] was compared with the GAN method presented in this paper. A total of 2000 abnormal samples (where data for a random sample varies within interval $\pm[30\%, 50\%]$) and 3000 normal samples were generated using VISSIM. Precision, recall, and F1 score are commonly used metrics to evaluate the performance of an anomaly detection model.

Precision refers to the probability that a sample predicted as correct is actually correct, and it differs from the concept of accuracy. It mainly indicates the degree of correctness in the results of positive samples. Recall, on the other hand, refers to the probability that a sample is predicted as positive among those that are actually positive, primarily reflecting the extent of prediction for positive samples. The F1 score considers both precision and recall, aiming to maximize both simultaneously. The specific experimental results are shown in Table 5.

Table 5. Comparison of Evaluation Indicators.

Norm	SVM/%	GAN Model/%
precision	100	86.7
recall	78.4	95.7
F1 score	87.9	90.1

Through comparative experiments, it was found that the classical SVM can effectively identify all positive sample data during the judgment process, with a precision higher than that of the method proposed in this paper. However, classical SVM tends to misclassify normal samples as anomalies quite frequently, which is precisely the focus of this paper. This is because, when a signaling system is attacked, hackers can generate some anomalous data that disrupts the normal operation of signals and control systems. Therefore, we need to accurately identify all anomalous data; thus, the model design places emphasis on the authenticity of the identification results themselves. The algorithm presented in this paper exhibits a higher recall rate. Ultimately, the F1 score of this method surpasses that of classical SVM, reflecting the algorithm's consideration of the overall relevance of the problem. Its primary goal is to quickly identify anomalous data to assist the signal

control system in restoring normalcy. Therefore, the detection performance of the proposed method is superior to that of the SVM algorithm.

GAN model's discriminator can be used to determine if there are any anomalies in traffic information. If anomalies are detected, the generator of the GAN model can be used to restore the data, allowing for a comparison with the anomalous data to identify the specific locations of the anomalies. The algorithm's accuracy and restoration effectiveness were validated using 1000 anomalous samples, as shown in Table 6.

Table 6. Evaluation Indicators.

Norm	GAN/%
determine Accuracy Rate	97.6
RMSE	13.9

Those interested in using this method can select traffic flow data and signal control data from a target area to train a model on a dataset. Then, they can utilize a simulation environment to randomly generate anomalous data. Through our validation, we found that the accuracy in identifying anomalous data can reach over 97.6%. The deviation in repairing anomalous data using the GAN model's generation network can be kept within 13.9%, making it superior compared to similar methods.

5. Conclusions

This paper first analyzes the selection criteria for an attacker's position from a theoretical perspective, providing a foundation for rational attackers when choosing between attacking the superior UC or the TSM. This lays the theoretical groundwork for developing defense strategies. From a data perspective, it proposes an attack detection algorithm based on a GAN model, establishing a Bayesian game model with the attacker, UC, and sTSM as participants. The paper examines the conditions for Nash Equilibrium under various strategic combinations. Traffic flow sample data and attack sample data were collected, and a joint simulation of information attacks on road network data was conducted using MATLAB–VISSIM to simulate abnormal situations, which validated the effectiveness of the algorithm. Additionally, a comparison was made between the SVM model and the proposed traffic GAN model. It was found that SVM did not misclassify any abnormal samples as normal, achieving a higher precision than the method proposed in this paper. However, SVM misclassified more normal samples as abnormal, while the recall rate of the proposed algorithm was higher. Ultimately, the F1 score of the proposed method exceeded that of SVM, indicating that the detection performance of the proposed method is superior to that of the SVM algorithm. In the future, we will further explore the practical problems of urban road traffic control and continue to revise the model and database to prepare for application in practical projects.

Author Contributions: Conceptualization, Data curation, Methodology, Writing—review & editing, K.L.; Conceptualization, Investigation, Writing—review & editing, K.P.; Formal analysis, Project administration, Software, W.X.; Investigation, Resources, Visualization, M.L.; Formal analysis, Software, Writing—review & editing, Z.H.; Conceptualization, Methodology, Writing—original draft, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Key R&D Program Project, grant number 2022YFB2601804 and Yuxiu Innovation Project of NCUT (No. 2024NCUTYXCX218). This research was also supported by North China University of Technology.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Elassy, M.; Al-Hattab, M.; Takruri, M.; Badawi, S. Intelligent transportation systems for sustainable smart cities. *Transp. Eng.* **2024**, *16*, 100252. [CrossRef]
2. Abirami, S.; Pethuraj, M.; Uthayakumar, M.; Chitra, P. A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system. *Case Stud. Transp. Policy* **2024**, *17*, 101247. [CrossRef]
3. Li, Z.; Jin, D.; Hannon, C.; Shahidehpour, M.; Wang, J. Assessing and mitigating cybersecurity risks of traffic light systems in smart cities. *IET Cyber-Phys. Syst. Theory Appl.* **2017**, *1*, 60–69. [CrossRef]
4. Li, Z.; Shahidehpour, M. Deployment of cybersecurity for managing traffic efficiency and safety in smart cities. *Electr. J.* **2017**, *30*, 52–61. [CrossRef]
5. Haddad, J.; Mirkin, B. Resilient Perimeter Control of Macroscopic Fundamental Diagram Networks under Cyberattacks. *Transp. Res. Part B Methodol.* **2020**, *132*, 44–59. [CrossRef]
6. Haddad, J.; Mirkin, B. Resilient Perimeter Control of Macroscopic Fundamental Diagram Networks under Cyberattacks—ScienceDirect. *Transp. Res. Procedia* **2019**, *38*, 56–76. [CrossRef]
7. Khattak, Z.H.; Park, H.; Hong, S.; Boateng, R.A.; Smith, B.L. Investigating Cybersecurity Issues In Active Traffic Management Systems. *Transp. Res. Rec.* **2018**, *2672*, 79–90. [CrossRef]
8. Cerrudo, C. Hacking US Traffic Control Systems. *Proc. Defcon.* **2014**, *22*, 1–15.
9. Ghena, B.; Beyer, W.; Hillaker, A.; Pevarnek, J.; Halderman, J.A. Green Lights Forever: Analyzing the Security of Traffic Infrastructure. In Proceedings of the 8th USENIX Conference on Offensive Technologies, San Diego, CA, USA, 19 August 2014.
10. Goldberg, S.; Kenealy, B.; Shepherd, K. Traffic Hackers Get the Green Light. Available online: https://www.businessinsurance.com/off-beat-traffic-hackers-get-the-green-light/ (accessed on 1 April 2023).
11. Ahmad, F.; Shah, Z.; Al-Fagih, L. Applications of evolutionary game theory in urban road transport network: A state of the art review. *Sustain. Cities Soc.* **2023**, *98*, 104791–104825. [CrossRef]
12. Cheng, P.; Pan, J.; Zhang, Y. Adaptive unknown input observer-based detection and identification method for intelligent transportation under malicious attack. *Meas. Control.* **2023**, *56*, 1377–1386. [CrossRef]
13. Sun, R.; Luo, Q.; Chen, Y. Online transportation network cyber-attack detection based on stationary sensor data. *Transp. Res. Part C Emerg. Technol.* **2023**, *149*, 104058. [CrossRef]
14. Zhang, C.; Costa-Pérez, X.; Patras, P. Adversarial Attacks Against Deep Learning-Based Network Intrusion Detection Systems and Defense Mechanisms. *IEEE ACM Trans. Netw.* **2022**, *30*, 1294–1311. [CrossRef]
15. Hou, Y.; He, R.; Dong, J.; Yang, Y.; Ma, W. IoT Anomaly Detection Based on Autoencoder and Bayesian Gaussian Mixture Model. *Electronics* **2022**, *11*, 3287. [CrossRef]
16. Huang, X.; Dong, J. An Adaptive Secure Control Scheme for T–S Fuzzy Systems Against Simultaneous Stealthy Sensor and Actuator Attacks. *IEEE Trans. Fuzzy Syst.* **2021**, *29*, 1978–1991. [CrossRef]
17. Zhang, Y.; Malacaria, P. Bayesian Stackelberg games for cyber-security decision support. *Decis. Support Syst.* **2021**, *148*, 113599. [CrossRef]
18. Pasqualetti, F.; Dorfler, F.; Bullo, F. Cyber-Physical Security via Geometric Control: Distributed Monitoring and Malicious Attacks. In Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), Maui, HI, USA, 10–13 December 2012; pp. 3418–3426.
19. Amin, S.; Cárdenas, A.A.; Sastry, S.S. Safe and Secure Networked Control Systems under Denial-of-Service Attacks. In *International Workshop on Hybrid Systems: Computation and Control*; Springer: Berlin/Heidelberg, Germany, 2009.
20. Mo, Y.; Kim, T.H.; Brancik, K.; Dickinson, D.; Lee, H.; Perrig, A.; Sinopoli, B. Cyber–Physical Security of a Smart Grid Infrastructure. *Proc. IEEE* **2012**, *100*, 195–209.
21. Zhu, M.; Martinez, S. On the Performance Analysis of Resilient Networked Control Systems Under Replay Attacks. *IEEE Trans. Autom. Control* **2014**, *59*, 804–808. [CrossRef]
22. Shombot, E.S.; Dusserre, G.; Bestak, R.; Ahmed, N.B. An application for predicting phishing attacks: A case of implementing a support vector machine learning model. *Cyber Secur. Appl.* **2023**, *2*, 100036. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.