An Effective Study on Data Science Approach to Cybercrime Underground Economy Data

M. Kiran Kumar[1] & Dr. K. Bhargavi[2]
[1]*Assistant Professor, Department of Information Technology, Guru Nanak Institutions Technical Campus, Hyderabad-Telangana.*
[2]*Associate Professor, Department of Computer Science and Engineering, PVKK Institute of Technology, Anantapur-Andhra Pradesh.*

ABSTRACT:

In spite of the rapid escalation of digital threats, there has still been little research into the foundations of the subject or philosophies that could serve to manage Information Systems researchers and practitioners who deal with cyber security. In addition, little is alluded to about Crime-as-a-Service (CaaS), a criminal plan of action that bolsters the cybercrime underground. This research gap and the practical cybercrime issues we face have motivated us to investigate the cybercrime underground economy by taking a data analytics approach from a design science perspective. To achieve this goal, we propose (1) a DATA SCIENCE framework for analyzing the cybercrime underground, (2) CaaS and crime ware definitions, and (1) an associated classification demonstrate. In addition, we (1) develop an example application to demonstrate how the proposed framework and classification model could be actualized in practice. We at that point use this application to investigate the cybercrime underground economy by analyzing a large dataset obtained from the web hacking community. By taking a design science research approach, this examination adds to the design of artifacts, foundations, and strategies right now. Additionally, it gives supportive practical bits of information to practitioners by proposing rules as to how governments and organizations in all organizations can prepare for attacks by the cybercrime underground.

Keywords: *Crimeware-as-a-Service, crimeware, underground economy, hacking community, machine learning, design science research.*

I. INTRODUCTION

As the threat displayed by massive digital attacks (e.g., ransomware and passed on denial of service attacks (DDoS)) and cybercrimes has created, individuals, organizations, and governments have attempted to find ways to shield against them. In 2017, ransomware known as WannaCry was in charge of nearly 15,000 attacks in almost 100 nations [1]. The hazardous impact of cybercrime has put governments constrained to increase their cybersecurity spending plans. Joined States President Barack Obama proposed spending over $19 billion on cybersecurity as part of his fiscal year 2017 spending plan, an increase of over 15% since 2016 [2].

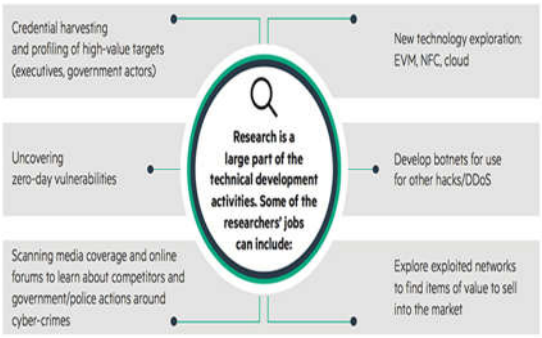


Fig 1 : Cybercrime Economy Solution

The attackers' worth chain

The present enemies regularly make a formalized working model and 'worth chain' that is fundamentally the same as authentic organizations in structure and passes on progressively unmistakable ROI for the cybercriminal association all through the assault lifecycle. In case adventure level security pioneers, controllers and law implementation are to vexed the attackers' association, they ought to initially observe each movement in the value chain of this cybercrime economy.

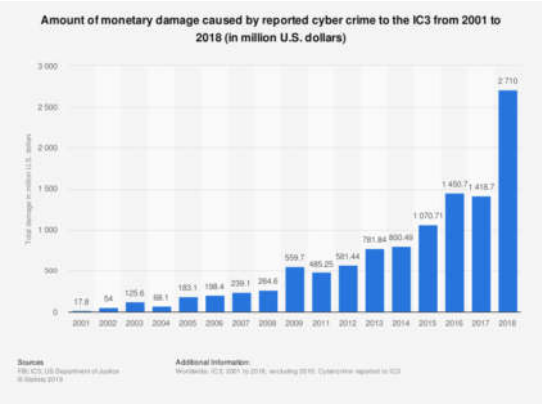


FIG 2 The impact of cybercrime has soared over two decades,

Basic components to the attackers' worth chain models regularly include:

HR the officials – Includes selecting, screening and paying the supporting 'staff' expected to pass on express assault necessities; the aptitudes based preparing and guidance of attackers in like manner fall inside this classification.

Tasks – The 'supervisory group' that guarantees the smooth movement of data and assets all through the assault lifecycle; this gathering will viably attempt to decrease expenses and lift ROI at each movement.

Specialized advancement – The cutting edge 'laborers' giving the specialized aptitude required to play out some random assault, including research, vulnerability abuse, computerization, and that's simply the start.

Advertising and deals – These gatherings guarantee that the assault gathering's reputation in the underground commercial place is strong and the illegal things are both known and trusted among the proposed intrigue gathering of potential purchasers.

Outbound coordinations – This incorporates both the individuals and frameworks liable for passing on purchased items to a purchaser, be it tremendous lots of taken Visa information, clinical records, authorized innovation or something different.

"Cybercriminals are exceptionally capable, have incredible financing, and are cooperating to dispatch concentrated assaults," said Chris Christiansen, Program Vice President, Security Products and Services, IDC.

1.1 Cybercrime economy

Upsetting the chain and moving undertaking security

HPE endorses various approachs for huge business security specialists to all the almost certain shield against these sifted through attackers:

Decrease the advantages – Limit the cash related prizes adversaries can understand from an assault on the undertaking by actualizing start to complete encryption arrangements. By scrambling information despite everything, moving and being utilized, the data is rendered inconsequential to the attackers, restricting their capacity to sell and decreasing advantages.

Lessen the goal pool – The augmentation of versatile and IoT has significantly expanded the conceivable assault surface for all endeavors. Associations must incorporate security with their improvement methods, and focus on guaranteeing the communications between information, applications, and customers paying little regard to the gadget to all the almost certain alleviate and upset adversary assaults.

Gain from the foes – New advancements, for example, 'misleading matrices' give methods for catching, checking and gaining from attackers as they investigate their way through a practical duplication of the framework. Adventures can use this data to all the more readily secure their certifiable framework, upset comparative assaults before they start, and thwart the advancement of attackers.

This research gap and the practical issues faced by cybercriminals motivate our investigation. We take a data analytics approach and investigate the cybercrime economy from a design science perspective. To achieve this goal, we (1) propose a DATA SCIENCE framework for analyzing the cybercrime underground to manage researchers and practitioners; (2) characterize CaaS and crimeware to all the more readily reflect their features from both academic research and business practice perspectives; (1) use this to manufacture a classification appear for CaaS and crimeware; and (1) assemble an application to demonstrate how the proposed framework and classification model could be actualized in practice. We at that point evaluate this application by applying it in a case consider, namely investigating the cybercrime economy by analyzing a large dataset from the electronic hacking community.

This investigation takes design science research (DSR) approach. Design science "creates and evaluates information innovation artifacts planned to tackle perceived problems"[1]. DSR incorporates developing a range of IT artifacts, for example, decision emotionally steady systems, models, frameworks, gadgets, strategies, and applications [1]. Where behavioral science research attempts to create and legitimize speculations that explain or anticipate human

or organizational phenomena, DSR hopes to expand the boundaries of human and organizational capabilities by creating new and innovative artifacts [1]–[8]. DSR's responsibility is to add value to the literature and practice as far as "design artifacts, design advancement information (e.g., foundations), and/or design evaluation learning (e.g., strategies)," [1].

This examination seeks after these DSR manages and contributes design artifacts, foundations, and strategies [1]. In particular, DSR must demonstrate that design artifacts are "implementable" in the business condition to take care of an important issue [1], so we give an implementable framework rather than a conceptual one. We also create a front-end application as a case example to demonstrate how the proposed framework and classification model could be actualized in practice. In addition, this examination adds to the design theory [9], [10].

In spite of the fact that Information Systems (IS) researchers and practitioners are taking an increasing enthusiasm for cybercrime, because of the critical issues arising from the rapid increase in digital threats, few have attempted to put this new enthusiasm on a solid foundation or create suitable strategies. Past investigations have not analyzed the underground economy behind cybercrime start to finish. Additionally, little is considered CaaS, one of the primary plans of action behind the cybercrime underground. There is an overall lack of understanding, both in research and practice, of the nature of this underground and the mechanisms fundamental it. Financial impact of cybercrime

The quantity of attacks obstructed by Symantec is only a glimpse of something larger and cybercrime is without a doubt answerable for the majority of them. The advantaged weapons utilized are the malware, malicious code of increasing unpredictability that much of the time have escaped security mechanisms, principal vectors of attacks abused by cybercriminals are web attack packs and socially built phishing procedures.

As for techniques, the creative advancement and utilization of evaluation strategies give DSR duties [1]. Accordingly, this investigation uses a dynamic analysis to lead an ex-ante evaluation of the classification to demonstrate. It also coordinates an ex-post evaluation of a front-end application using observational strategies (case examples). From a practical perspective, this investigation also outfits practitioners with valuable bits of information by making recommendations to coordinate governments and organizations in all endeavors in taking care of the issues they face while preparing for attacks from the cybercrime underground

.II. RELATED WORK

The cybercriminal underground as an overall commercial community overflowing with obscure purchasers and dealers. It's made out of a couple of smaller markets special to each region. As of now, the most noticeable are those in Russia, China, and Latin America. These virtual underground markets generally thrive in conversations or visit rooms where various cybercriminals go about as strange agents who exchange items and dares to make advantage. What's progressively, much like average representatives, these cybercriminals hold fast to express plans of action:

Plan of action: It's the most clear approach. Cybercriminals hawk items like taken customer certifications, malware, abuses, and so forward, and offer them to any individual who's fascinated.

Created bad behavior model: This set-up incorporates a couple of individuals or gatherings who cooperate to accomplish certain destinations. Each individual in the chain serves their own unique capacity essential to the whole gathering's tasks.

Redistributing model: This plan of action requires cybercriminals to band together with outside PC owners, whose arrangement of machines they can lease as botnets for malicious plans.

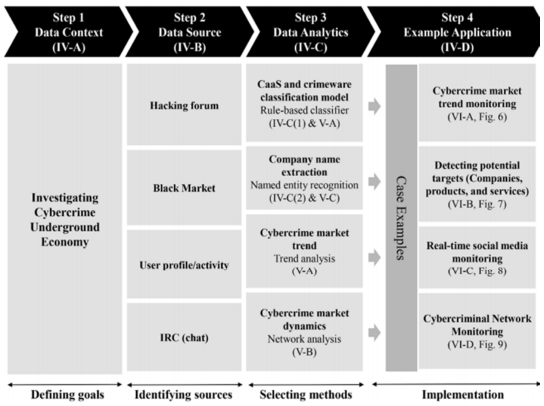
Guide follower model: Those fascinated to drive their skill in hacking or making noxious applications can get dynamically talented cybercriminals to pass on what they know.

Whatever model cybercriminals choose to follow, one thing's point of fact, in case you're using the Internet, you, including your own data and cash, are slanted to their plans. Regardless of whether you're perusing email, messing around, or interfacing with companions via electronic systems administration media, cybercriminals are continuing to create items expected to piggyback on all of your activities. For whatever period of time that cybercriminals have something to profit by, the cybercriminal underground is constantly open for business.

Because CaaS is designed for juveniles, its customers don't have to run a hacking server or have abnormal state hacking aptitudes. In this manner, the CaaS plan of action can incorporate the accompanying employments: forming a hacking program, playing out an attack, appointing an attack, giving an attack server (infrastructure), and laundering the profits. Sood and Enbody [1] have suggested that crimeware marketplaces have three key segments, namely actors (e.g., coders, operators, or purchasers), value chains, and strategies for operation (e.g., CaaS, pay-per-install, crimeware tool kits, brokerage, or giving data). Irregular checking and analysis of the substance of cybercrime marketplaces could help anticipate future digital threats [1].

IV. ANALYTICAL FRAMEWORK AND METHODS

The assembles used in DSR are component representations [10] that give the vocabulary and images expected to characterize issues and arrangements [1]. Accordingly, the design segments used right now the cybercrime underground, criminal things (CaaS and crimeware), classifications, and front-end framework applications, and the artifacts are based on these fabricates. These artifacts are evaluated in two stages [19]: ex-ante (classification evaluation) and ex-post (case example). Because DSR should be tentative, this ex-present evaluation is essential on the search method used by iterative DSR, which includes search, design, ex-ante evaluation, advancement, artifact, ex-post evaluation, and research [19]. Based on this, we propose the DATA SCIENCE framework appeared in Fig. 1.



Proposed data analytical framework. Sections are in parentheses.

Because cybercrime varies from general bad behavior from various perspectives, we have to coordinate a variety of analyses using a large dataset. A past report [10] proposed a data burrowing framework for bad behavior, separating violations harmful to the general open into eight categories: traffic violations, sex bad behavior, burglary, fraud, arson, gang/sedate offenses, awful bad behavior, and cybercrime.

Although the past investigation explained how data mining strategies could be applied to bad behavior analysis, it didn't consider the particular features of cybercrime. Additionally, it just explained the data mining strategies rapidly, rather than displaying a broad audit of the framework [10]. In contrast, the goal of our DATA SCIENCE framework is to lead a major picture investigation of the cybercrime underground by covering all phases of DATA SCIENCE from the earliest starting point as far as conceivable (see Fig. 1). This framework incorporates four stages: (1) characterizing goals; (2) recognizing sources; (1) picking analytical strategies; and (1) actualizing an application.

Because this investigation emphasizes the importance of RAT for analyzing the cybercrime underground, the proposed RAT-based definitions are critical to this framework: Steps 1–1 all contain the RAT parts, as Fig. 1 appears.

A. Stage 1: Defining Goals

The initial advance is to recognize the conceptual degree of the analysis. Specifically, this movement perceives the analysis setting, namely the targets and goals. To gain a through and through understanding of the force CaaS research, we investigated the cybercrime underground, which operates as a shut community. Right now, goal of the proposed framework is to "investigate the cybercrime underground economy."

B. Stage 2: Identifying Sources

The subsequent advance is to perceive the data sources, based on the goals characterized by Step 1. This movement should consider what data is required and where it may be obtained. Since the goal of this examination is to investigate the cybercrime underground, we consider data on the cybercrime underground community. We, therefore, gathered such data from the community itself and obtained a malware database from a leading global cybersecurity research firm.

Because cybercriminals regularly change their IP addresses and use anti-crawling substance to conceal their communications, we used a self-created crawler that can resolve captchas and anti-crawling substance to gather the necessary data. We gathered a total of 2,172,091 posts moving CaaS or crimeware, made between August 2008 and October 2017, from a large hacking community site (www.hackforums.net) with in excess of 178,000 individuals and more than 10 million posts. We also gathered 16,172 customer profiles of merchants and potential purchasers, based on their communication narratives, as well as expenses and questions and answers about the transactions.

The black market uses traditional conversation threads (e.g., release boards) instead of typical online business platforms (e.g., eBay, and Amazon). For example, merchants create threads in marketplace gatherings to move things, and potential purchasers remark on these threads. A standout amongst the most significant challenges was, along these lines, changing over this unstructured data into organized data. Since the thing features, expenses, and portrayals were explained inside longer messages, we used a variety of substance mining methods to extract the important features: for example, we used named component acknowledgment to extract company names (see Section IV-C(2)). Since these works included many typographic mistakes and jargon terms, we had to create a dictionary for use amid a preprocessing step.

In addition, we obtained a malware database from a cybersecurity firm containing in excess

of 13,815 segments covering cybercrimes between May 11, 2010, and January 13, 2014. This exceptional dataset strengthened our investigation by giving real-world verification from an alternate viewpoint.

C. Stage 1: Selecting Analytical Methods

1) CaaS AND CRIMEWARE CLASSIFICATION MODEL:

A various range of things is sold in the cybercrime underground, with various degrees of associated hazard. For this examination, we concentrated mainly on things critical to hacking. We originally filtered the messages to pick only those that carried significant dangers and then isolated them into the categories.

To choose whether a given message is dangerous, our classification demonstrates registers whether it falls with one of the accompanying five categories: Threat, Product/Service, File Extension, Market, and Exclusion. Fig. 2 demonstrates an improved example to clarify this standards-based approach. We used a dictionary containing 1,191 watchwords spread across five categories, fabricated using data obtained from the cybersecurity research firm, anti-contamination merchants, Wikipedia, and gatherings.

To be classified as a dangerous Threat, for example, a message ought to also contain Market-related catchphrases. Messages containing the two Threats-and Market-related watchwords are seen as increasingly dangerous (e.g., "Moving calm Microsoft Office abuse") at that point messages with just Threat-related catchphrases (e.g., "Can I conceal a record inside a word doc?"). In like manner, messages related to the Product/Service, Market, and File Extension categories are not perceived as dangerous in case they simply contain catchphrases related to one category. In addition, messages containing Exclusion-related catchphrases (e.g., "tutorials" or "tips") are not perceived as dangerous.

To classify messages accurately, we also use catchphrases related to CaaS and crimeware. This classification step is applied after the messages have been separated as above, so many watchwords are not required and the criteria are less intricate. In any case, when a message fits into various categories, this overlap is recorded so as to get additional encounters from the later analysis and applications. The sorts of the catchphrase used for the proposed classification show are according to the accompanying.

- Threat: watchwords specifically related to threats or digital attacks (e.g., "abuse" or "botnet").

- Product/Service: watchwords related to things or services (e.g., "Facebook" or "Skype").
- File Extension: watchwords related to software or add-ons (e.g., "doc" or "ppt").
- Market: watchwords related to markets or transactions (e.g., "moving" or "$").
- Exclusion: watchwords that are not related to malware (e.g., "tutorial" or "tips").

To enhance the quality of the training data, we alluded to the malware database obtained from the cybersecurity research firm. Since this database contained labeled black market communications by cybersecurity professionals, it gave an appropriate manual for building the training dataset.

Notwithstanding, the database was somewhat outdated (May 11, 2016, to January 13, 2018), so we also alluded to later data from anti-contamination dealers' destinations. Four undergraduate understudies (two gatherings of two) with cybersecurity backgrounds assisted in validating this data. Before creating the training dataset, we gave the participants a ton of rules and methods based on the malware dataset. After they had totally grasped and talked about these, we used them to create the training data. At the moment that two understudies disagreed, someone from the other gathering talked about the matter with them to help accommodate the disagreement. The between rater reliability score was 82%. This is above the proposed reliability least (80%), and so was seen as adequate [11].

We use the naïve Bayes algorithm, a probabilistic classification algorithm [12], [13] that addresses probabilistic reasoning under uncertainty because it is the least troublesome approach for content classification [14]. Its forecasts self-directly as new information is experienced, so they end up being dynamically accurate with more data. The conditional probability is given by Bayes' speculation:

$$P(C_i \mid d) = \frac{P(d \mid C_j)P(C_j)}{P(d)} . (1)$$

Here, P(Cj) and P(Ci|d) is the earlier and back probabilities of class Ci, while P(d) and P(d|Ci) is the earlier and back probabilities of the indicator d. The reliant feature vector is x = (x1, x2,… , Xn) and Bayes' speculation gives us the accompanying.

$$C_i = argmax\, P(x_1, x_2, x_3, x_4, \ldots, x_n \mid C)\, P(C_i) (2)$$

$$C_i = \arg\max \prod_{i=1}^{n} P(x_n \mid C_i) P(C_i) (3)$$

$$P(x_i) = \frac{Number\ of\ x_i\ in\ documents\ of\ class\ C}{Number\ of\ words\ in\ documents\ of\ class\ C} (4)$$

Basing the probabilistic classifier on the naïve Bayes display streamlines the conditional opportunity assumptions for the CaaS and crimeware classes. The sentences in a record are tokenized into words, which are classified

as relating to either CaaS or crimeware. The probability of the report having feature xi can then be figured by partitioning "the quantity of features xi in archives of class C" by "the quantity of words in records of class C".

2) COMPANY NAME EXTRACTION

Named component acknowledgment is an information extraction framework that classifies named substances based on a predefined dictionary. We used the Open Calais API to see the company and personal names. For example, Fig. 1 demonstrates that "Apple" is seen as alluding to the company rather than the natural item. We use named substance acknowledgment to perceive the company names referenced in the cybercrime underground, which we consider as potential targets (e.g., RAT suitable targets) [13], [14].

D. Stage 1: Implementing an Application

Although organizations emphasize the measures they take to avert cybercrime, their overall viability has yet to be empirically demonstrated in practice. In the last advance of our framework, we demonstrate the utilization of the proposed CaaS and crimeware definitions, classification model, and analysis framework.

CONCLUSION:

This examination adds to the DSR literature in a broader IS setting in several ways. Because it takes a DSR approach, it adds to the design artifacts, foundations, and techniques right now [8]. In any case, by creating example front-end applications, we have demonstrated how our design artifacts (the proposed framework and classification display) can be actualized in practice. Despite the rapidly creating threat from cybercrime, there has been little research into practical frameworks for future cybersecurity researchers: the past investigations have not attempted to analyze the data or take a systematic displaying approach [1], [18]–[12]. In DSR, we ought to demonstrate that the artifacts can be actualized in a business domain for them to qualify as taking care of an important unsolved issue [1]. We have right now an implementable framework, not just a conceptual one. Second, this investigation adds to the creating cybersecurity literature by giving a foundation on which to manufacture [1].In addition, there is at present a lack of good CaaS and crimeware definitions and classification models. Third, this examination adds to the gathering of information by demonstrating new approaches to the issues digital wrongdoing and social media researchers face [1], [13]. Finally, this examination adds to RAT [13], [14] by applying it to the cybercrime underground. The same three factors can be applied to cybercrime and general violations, so we have classified CaaS and crimeware with regards to the cybercrime underground and analyzed them accordingly.

REFERENCES

1. An, J., & Kim, H. W. (2018). A data analytics approach to the cybercrime underground economy. Ieee Access, 6, 26636-26652.
2. Yip, M., Shadbolt, N., Tiropanis, T., & Webber, C. (2012). The digital underground economy: A social network approach to understanding cybercrime.
3. Yip, M., Shadbolt, N., & Webber, C. (2012, June). Structural analysis of online criminal social networks. In 2012 IEEE International Conference on Intelligence and Security Informatics (pp. 60-65). IEEE.
4. Pastrana, S., Hutchings, A., Caines, A., & Buttery, P. (2018, September). Characterizing eve: Analysing cybercrime actors in a large underground forum. In International Symposium on Research in Attacks, Intrusions, and Defenses (pp. 207-227). Springer, Cham.
5. Rani, A. S., & Manasa, G. (2019). An Effective Data Analytics Approach to Cybercrime Underground Economy Using Ml Methodologies. International Journal of Research, 6(3), 1022-1036.
6. Sood, A. K., & Enbody, R. J. (2013). Crimeware-as-a-service—a survey of commoditized crimeware in the underground market. International Journal of Critical Infrastructure Protection, 6(1), 28-38.
7. J. C. Wong and O. Solon. (2017, May 12). Massive ransomware cyber-attack hits nearly 100 countries around the world. [Online]. Available: https://www.theguardian.com/technology/2017/may/12/global-cyber-attack-ransomware-nsa-uk-nhs
8. "FACT SHEET: Cybersecurity National Action Plan," ed: The White House, 2016.
9. K. Sood and R. J. Enbody, "Crimeware-as-a-service—A survey of commoditized crimeware in the underground market," Int. J. Crit. Info. Prot., vol. 1, no. 1, pp. –18, 2013.
10. S. W. Brenner, "Organized Cybercrime? How Cyberspace May Affect the Structure of Criminal Relationships," N. C. J. Law & Technol., vol. 1, no. 1, pp. 1-10, 2002.
11. K. Hughes, "Entering the world-wide web," ACM SIGWEB Newsl., vol. 1, no. 1, pp. 1–8, 1994.
12. S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact," MIS Quart., vol. 17, no. 2, pp. 17-156, 2013.
13. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," MIS Quart., vol. , no. 1, pp. 15-105, 2004.
14. [8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," J. Manag. Inf. Syst., vol. , no. 1, pp. 15–17, 2007.
15. S. Gregor, "Design theory in information systems," Aust. J. Inf. Syst., vol. 10, no. 1, pp. 14–, 2002.
16. S. Gregor and D. Jones, "The Anatomy of a Design Theory," J. the Assoc. Inf. Syst., vol. 8, no. 1, pp. 1–15, 2007.
17. M. Yar, "The Novelty of 'Cybercrime': An Assessment in Light of Routine Activity Theory," Eur. J. Criminol., vol. 2, no. 1, pp. 107– 1, 2005.
18. K.-K. R. Choo, "Organised Crime Groups in Cyberspace: a Typology," Trends in Organized Crime, vol. 11, no. 1, pp. 0–5, 2008.
19. L. E. Cohen and M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach," Am. Social. Rev., vol. 14, pp. 188–108, 1979.
20. M. Felson, "Routine Activities and Crime Prevention in the Developing Metropolis," Criminol., vol. , no. 1, pp. 911–92, 1987.