

ADVANCED SECURE AND PRIVACY PRESERVING DATA DEDUPLICATION USING EXTREME LEARNING MACHINE

¹Pathella Mahidhar Sai ²Dr. P Harini ³N Lakshmi Narayana

¹M. Tech Scholar, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.

²Professor & HOD, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.

³Assistant Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.

e-mail: drharinicse@gmail.com

ABSTRACT: Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. The centralized DeKey based deduplication system in which user do not need to handle any keys on their own but as a substitute securely distribute the convergent key across multiple servers.. The existing system is limited high overheads and may also suffer from falsified aggregated results returned by a malicious server. Wireless sensor networks (WSN) and Cloud Computing to conduct an analysis of previously published research and provide a summary of the efforts put into researching BC applications for network security. We present the research paper is developed practical privacy security analytics in information systems. Cloud Technology is providing prompt Internet access as well as feasible access to information from any location and platform at any given time. There has been a significant rise in the volume of information produced as well as the take different involved in its data type. We propose a mapping framework to employ a fine-tuned multilayer feed forward artificial neural network (ANN) and extreme learning machine (ELM) for role engineering in the SCADA-enabled IIoT environment to ensure privacy and user access rights to resources. Our proposed system to entire application space for applying reinforcement learning to IIoT systems into four scenarios, namely, non-continuous learning without learning model sharing, non-continuous learning with learning model sharing, continuous learning without learning model sharing, and continuous learning without learning model sharing. Security analysis shows that our scheme protects the privacy of inputs, ML model and prediction results.

INDEX TERMS: Efficiency, Cloud Storage, Data Deduplication, Data Privacy, Encryption, Hash Key, Reliability.**INTRODUCTION**

Cloud providers offer potentially infinite storage space, where users can use as much space as they can and vendors constantly look for techniques which aimed to minimize redundant data (multiple copies) and maximize space savings [1]. The production control system with industrial control and monitoring capabilities that provides multiple enterprise-related services. IoT applications is recently deployed in cross-industry applications based on the principles of public information services [2]. The ML algorithms like regression models have attracted considerable attention over deep learning in application domain [3]. The advantages of collaborative learning they are two major concerns input data security and vulnerability of locally trained models to data leakage [4]. It was developed for monitoring transactions including decentralized digital currency every node

in the P2P network can receive updated data regarding the different transactions validated in a decentralized and distributed database [5]. The selection of data security method is design of security preserving analytics algorithms. The protection techniques only provide room for limited operations on the obfuscated data complex algorithms is disintegrated to these simpler operations [6]. Formed and standardized by the National Institute of Standards and Technology (NIST) the SHA (Secure Hash Algorithm) gives ideal performance in maintaining data integrity process. Taking offline supervised learning is example datasets should be available to the model during the training process. This model is not usable unless the training process is completed [7]. Online continuous machine learning is useful for handling the learning process in large-scale machine learning tasks on dynamic systems such as IIoT systems [8]. Although it supports model training based on different public keys, to extra encryption and decryption operations increase the computational burden is proposed is based naive Bayesian disease risk prediction scheme for online medical treatment [9].

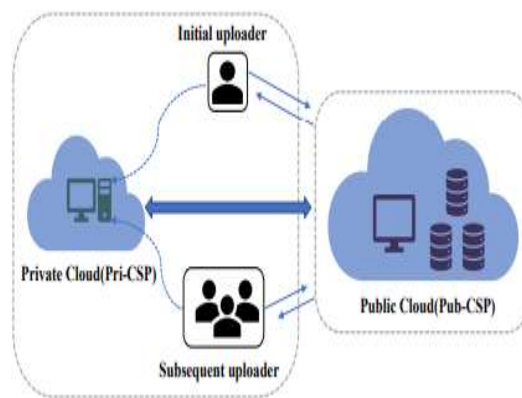


Figure 1. Architecture of a data deduplication system

1. RELATED WORK

Server-aided encryption scheme (DupLESS) for data deduplication was proposed independent key server generates the key which suffers from large computation time in block-level deduplication since it takes long time to generate keys proposed a client-side encryption that requests the data owner to do ownership check and deduplication which is impractical proposed a deduplication scheme by using attribute-based access control technique under hybrid cloud environment [10]. The interconnected nature of multiple nodes systems forming a chain and every node stores a duplicate of the primary chain hackers is quickly access the information [11]. We focus on the building of security preserving data mining algorithms relevant to informatics and then analyse the candidate process. While discussing these process we will try to understand their intrinsic trade-offs many security cost and utility [12]. The recommended the model of security-preserving collaborative model learning using skyline computation referred to as PCML which relies upon papillae cryptosystem take threshold decryption and distributed skyline computation [13]. The design of IIoT architectures and optimization of IIoT results. The application of IIoT,

number of research efforts have leveraged the data generated by sensors to assist in the operation of the industrial manufacturing processes [14].

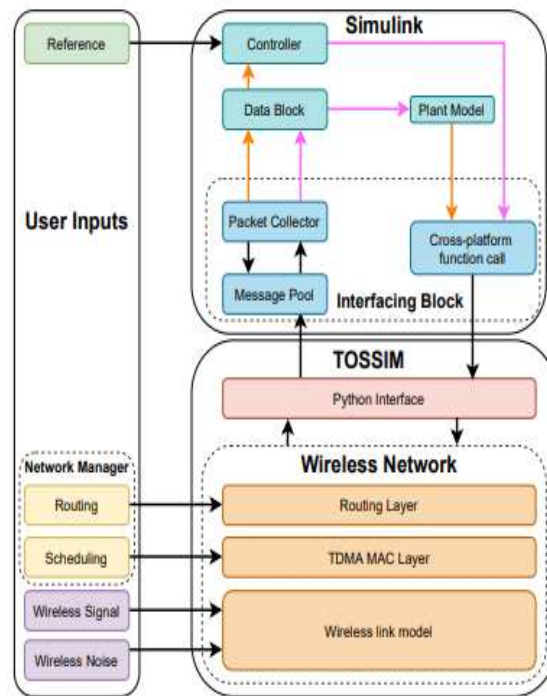


Figure. 2. Architecture of WCPS.

2. SYSTEM DESING

This construction can identify deduplication at a client to save network bandwidth and guarantee that the client holds a file rather than some part chose the projection of a file onto some randomly selected bit-positions as proof to realize the PoW. But both schemes above do not pay attention to data privacy and the server for data storage could be aware of the file content [15]. The ML prediction in HCPSs can provide customers with high-accurate prediction services in trained models and their owned data. we introduce the system and security models of VPMLP for edge enhanced HCPSs [16]. In all four scenarios we use a representative process control system called the Continuous Stirred Tank Reactor (CSTR) system is fluid temperature control system. The objective is to control the temperature of the liquid in the tank by modifying the steam flow rate [17]. The latest SCADA system consists of a central controller and a number of devices including sensors and actuators. They are widely used in industrial areas for controlling the process of the systems [18].

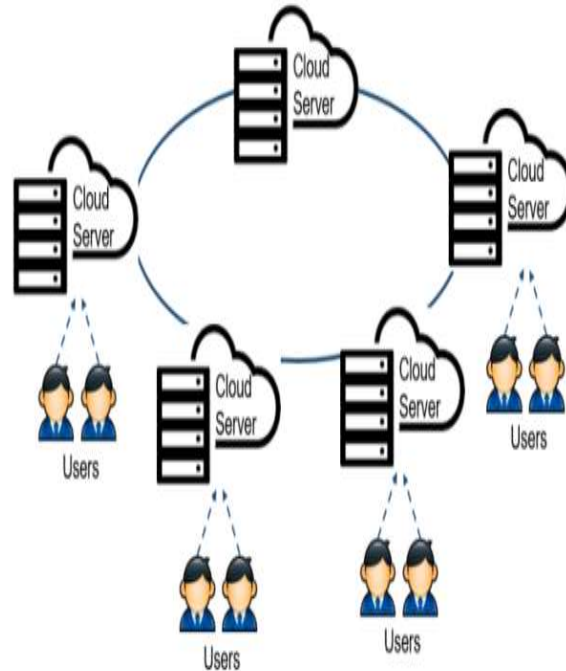


Figure. 3. Architecture of the CSTR System

3. PROPOSED MODEL

The criteria or framework for the process provide an overview of the activities that are carried organisation to expanding. In the Primo Chain framework the smart contract, DISTEN, CENTAUTH, IPFS, organised the framework is developed, dubbed PriModChain (security preserving trustworthy machine learning model training and distubuted framework based on block chain), tackles the security and trust concerns with machine learning in IIoT systems[19].The proposed system outlines the key components and functionalities of the project, focusing on distributed deduplication, privacy-preserving encryption, dynamic key generation, verification mechanisms, data retrieval, reliability, fault tolerance, and performance optimization. These aspects collectively contribute to the development of an advanced and privacy-conscious data deduplication system tailored for joint cloud storage environments. The proposed system entails the development of a distributed deduplication framework designed to enhance reliability and efficiency in joint cloud storage environments. [20]. integrating the machine-learning-based automated role assignment is provide accurate modelling of user–role relationships making the system efficient and effective in terms of time and cost [21]. The Block chain technique in association with SHA-256 enables and accelerates security preserving patient centric cryptographic hash algorithms like SHA assures trustworthy transactions cryptographic hash algorithms [22].

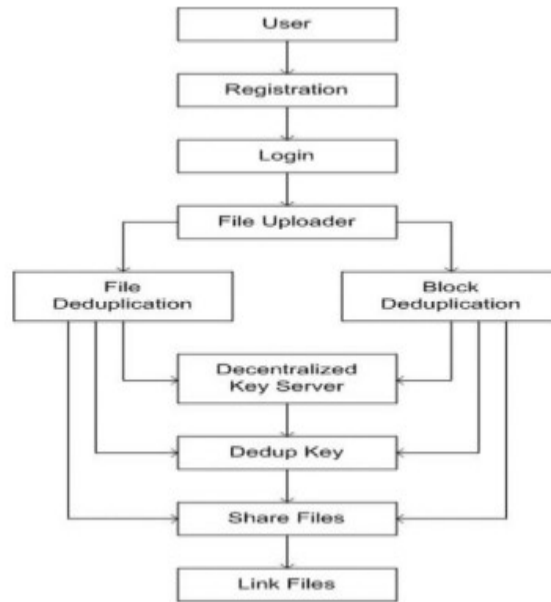


Figure 4: The elements of the proposed framework

1. Implementation of Block chain Technology in Network Applications

The implementation of BC technology utilized number of network applications in recent years and the domains in which BC technology could be applied along with image representation. CPS developed a protection model for its operational and data security based on BC technology [23]. The block chain was utilized to find a solution to the problem of information security. Present advancements in IoT and fifth-generation mobile networks (5G) is substantially increase the amount of big data collected by 5G-enabled industrial automation[22]. The building an efficient deep learning paradigm for IoT has several including a single point of failure the potential for IoT devices to leak personal information [24].

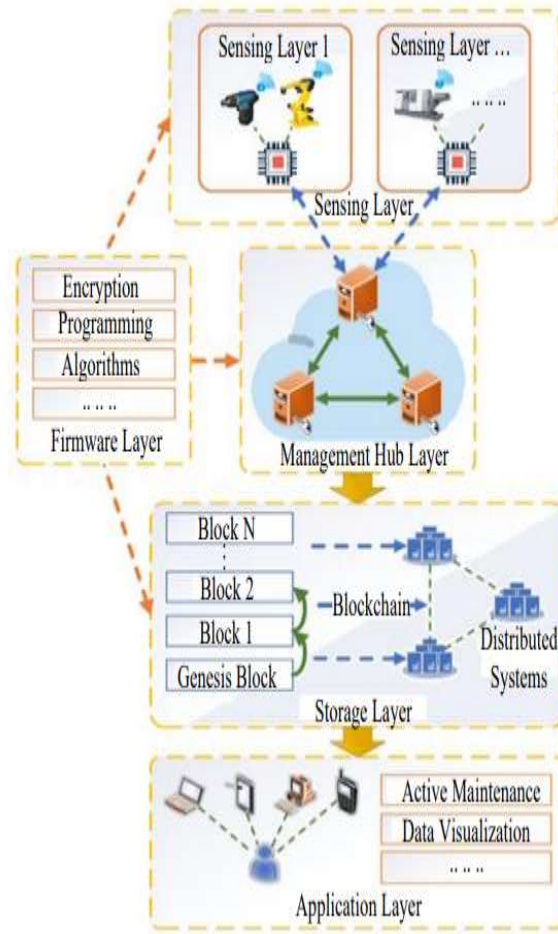


Figure. 5 Architecture of BC-based IIoT for a Smart Factory

4. SERVICE PROVIDER

The extreme scenario of users does not trust the SP itself the SP might be commercially motivated or not credible enough to win users' faith. The users is allured to the services the SP offers [25]. A possible remedy is introduce another honest-but-curious party, called crypto service provider (CSP), is manage secret keys decrypt intermediate results, and assist SP to finish the modelling task to framework in SP and CSP learn models over encrypted/masked data and the generated models are only decidable by the individual users [26].

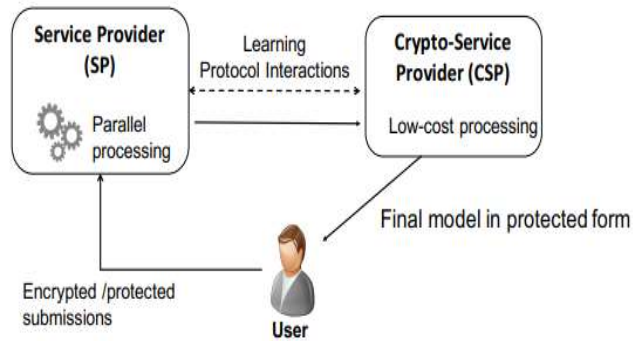


Figure. 6 Cryptographic Service Provider (CSP) and SP with preserved security

1. Structure of SHA-256 Algorithm SHA-256

Algorithms:

Input: Block of Message

Output: Fixed Size bits

Step 1: Pre-Processing

i). Indexing and Padding with 0's until data is a multiple of 512, less 64 bits.

Step 2: Initialize Hash Values

ii). Now create hash values

Step 3: Initialize Round Constants

iii). Similar to step 2, we are creating constants. This time, there are 64 of them.

Step 4: Chunk Loop

IV). The following steps for each 512-bit "chunk" of data from our input.

Step 5: Create Message Schedule

V). Copy the input data entry is a 32-bit word

Step 6: Compression

vi). Initialize variables hash values respectively

Step 7: Modify Final Values

vii). after the compression loop and change variables to them.

Step 8: Concatenate Final Hash

Viii). Combine them all together to get fixed length bit size

SHA-256 hashing calculation elaborated the official NIST standard there are main two steps in the SHA-256 calculation. Pre-measure of the first messages by message cushioning to extending the directive for the round calculation [27]

5. EXPERIMENTAL RESULTS

The results section of this project presents a thorough analysis of the implemented secure data deduplication system in Joint Cloud Storage environments. It covers the system's deployment outcomes, functionality, reduction in storage overhead, security measures, and overall efficiency

Data security methods are used for micro- data publishing without rigorous theoretical foundation they suffer from various background-knowledge based attacks. The simulations are executed at the different security levels with different sizes of the query vectors and the result some of the schemes is designed for neural network services for the sake of fairness these schemes are adjusted to the same scale of LR in the same reflect the similar tendencies on the execution time of these schemes under different security levels.

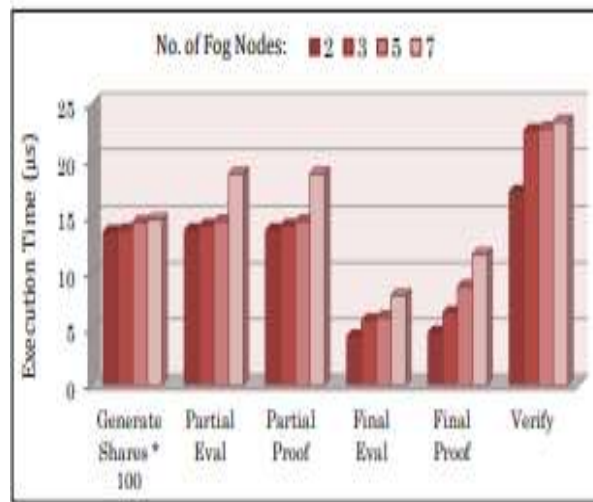


Figure. 7: Execution Time of each sub-task in Verifiability with increasing Fog nodes

6. CONCLUSION AND FUTURE WORK

The system adopts distributed deduplication frameworks with increased reliability, ensuring that data chunks are distributed across multiple cloud servers. Through the use of traditional De Key-based deduplication detection and Big Data processing, the system effectively avoids unnecessary traffic and storage, optimizing resource utilization in the cloud environment the security model is risk the desired algorithms quality and audience of the models must be conducted. Depending on the desired analytics and privacy level an additional party such as a cryptographic service provider might need to introduce to a framework. The security of SPRITE is analysed under an honest-but-curious setting where the cloud is untrustworthy. In our experiments MLP outperformed is more security results but the convergence efficiency with respect to time was better in any other application. In feature role, tuples in real-life use cases such as smart transportation, smart data,. Customized data is further investigated using advanced tailored machine learning algorithms and multilayer extreme learning machines and hybrid deep models to achieve robust security for role engineering and propagation in fine-grained access control. The successful implementation and validation of the system underscore its potential for widespread adoption, offering a reliable and secure approach to data management in collaborative cloud storage settings.

8. REFERENCES

- [1] Wang, G., Xu, L., Long, F., Li, L., & Fan, J. (2022). Privacy-Preserving Deduplication and Dynamic Data Sharing in Cloud Storage. *IEEE Transactions on Dependable and Secure Computing*, 19(1), 110-123.
- [2] Kumar, A., Varun, & Kumar, S. (2021). A Novel Approach of Data Deduplication and Privacy-Preserving in Cloud Computing. *Wireless Personal Communications*, 120(4), 3025-3041.
- [3] Ma, C., Hu, Y., Liu, J., & Zhao, D. (2020). A Privacy-Preserving Data Deduplication Scheme in Cloud Storage. *Journal of Computer and System Sciences*, 109, 143-152.
- [4] Singh, A., & Kumar, S. (2020). Secure and Privacy-Preserving Data Deduplication in Cloud Storage. *Future Generation Computer Systems*, 102, 815-825.
- [5] Ratna Raju Mukiri, B.V.V.S. Prasad, "Developing Secure Storage of cloud with IoT Gateway" *International Conference on Advancements in Computing & Management (ICACM-2019)*
- [6] Matt Fredrikson, SomeshJha, and Thomas Ristenpart. Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, page 1322–1333, 2015.
- [7] Luca Melis, Congzheng Song, Emiliano De Cristofaro, and VitalyShmatikov. Exploiting Unintended Feature Leakage in Collaborative Learning. In *2019 IEEE Symposium on Security and Privacy*, pages 691–706, 2019.
- [8] Zhibo Wang, Mengkai Song, Zhifei Zhang, Yang Song, Qian Wang, and Hairong Qi. Beyond Inferring Class Representatives: User-Level Privacy Leakage From Federated Learning. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 2512–2520, 2019.
- [9] Huadi Zheng, Qingqing Ye, Haibo Hu, Chengfang Fang, and Jie Shi. BDPL: A Boundary Differentially Private Layer Against Machine Learning Model Extraction Attacks. In *Computer Security – ESORICS 2019*, pages 66–83, 2019.
- [10] Yaochen Hu, Di Niu, Jianming Yang, and Shengping Zhou. FDML: A Collaborative Machine Learning Framework for Distributed Features. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2232–2240, 2019.
- [11] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, FarhadFarokhi, Shi Jin, Tony Q. S. Quek, and H. Vincent Poor. Federated Learning With Differential Privacy: Algorithms and Performance Analysis. *IEEE Transactions on Information Forensics and Security*, 15:3454– 3469, 2020.
- [12] Yunlong Lu, Xiaohong Huang, Yueyue Dai, SabitaMaharjan, and Yan Zhang. Differentially Private Asynchronous Federated Learning for Mobile Edge Computing in Urban Informatics. *IEEE Transactions on Industrial Informatics*, 16(3):2134–2143, 2020.
- [13] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, page 1175–1191, 2017.

- [14] Stacey Truex, Nathalie Baracaldo, Ali Anwar, Thomas Steinke, Heiko Ludwig, Rui Zhang, and Yi Zhou. A Hybrid Approach to PrivacyPreserving Federated Learning. In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, page 1–11, 2019.
- [15] Yong Yu et al., “Blockchain-Based Solutions to Security and Privacy Issues in the Internet of Things,” *EEE Wireless Communications*, vol. 25, no. 16, pp. 12-18, 2018. [CrossRef] [Google Scholar] [Publisher Link]
- [16] Minhaj Ahmad Khan, and Khaled Salah, “IoT Security: Review, Blockchain Solutions, and Open Challenges,” *Future Generation Computer Systems*, vol. 82, pp. 395-411, 2018. [CrossRef] [Google Scholar] [Publisher Link]
- [17] Sophocles Theodorou, and Nicolas Sklavos, *Blockchain-Based Security and Privacy in Smart Cities*, Smart Cities Cybersecurity and Privacy, Elsevier, Chapter 3, pp. 21-37, 2019. [CrossRef] [Google Scholar] [Publisher Link]
- [18] Lakshmana Kumar Ramasamy et al., “Blockchain-Based Wireless Sensor Networks for Malicious Node Detection: A Survey,” *IEEE Access*, vol. 9, pp. 128765-128785, 2021. [CrossRef] [Google Scholar] [Publisher Link]
- [19] Bandar Alamri, Katie Crowley, and Ita Richardson, “Blockchain-Based Identity Management Systems in Health IoT: A Systematic Review,” *IEEE Access*, vol. 10, pp. 59612-59629, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [20] Abdullah Al Mamun, Sami Azam, and Clementine Gritti, “Blockchain-Based Electronic Health Records Management: A Comprehensive Review and Future Research Direction,” *IEEE Access*, vol. 10, pp. 5768-5789, 2022
- [21] G. Shah and A. Tiwari, “Anomaly detection in iiot: A case study using machine learning,” in Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, 2018, pp. 295– 300.
- [22] B. Yang, X. Cao, X. Li, Q. Zhang, and L. Qian, “Mobile-edgecomputing-based hierarchical machine learning tasks distribution for iiot,” *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2169–2180, 2020.
- [23] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, “A trustworthy privacy preserving framework for machine learning in industrial iot systems,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6092–6102, 2020.
- [24] M. Liu, F. R. Yu, Y. Teng, V. C. Leung, and M. Song, “Performance optimization for blockchain-enabled industrial internet of things (iiot) systems: A deep reinforcement learning approach,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3559–3570, 2019.
- [25] B. Chen, J. Wan, Y. Lan, M. Imran, D. Li, and N. Guizani, “Improving cognitive ability of edge intelligent iiot through machine learning,” *IEEE Network*, vol. 33, no. 5, pp. 61–67, 2019.
- [26] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.
- [27] M. O. Duff, “Q-learning for bandit problems,” in *Machine Learning Proceedings 1995*. Elsevier, 1995, pp. 209–217.