University of Missouri

VIMAN Lab
VIRTUALIZATION, MULTIMEDIA AND NETWORKING LAB

ClaimChain: Secure Blockchain Platform for handling Insurance Claims Processing

Presented by: Masrik Dahir, George Stafford

Final REU 2021 Presentation

Faculty Advisor: Dr. Prasad Calyam

Graduate Mentors: Ramya Bhamidipati, Roshan Neupane, Varsha Vakkavanthula

Agenda

Introduction	• Background on Insurance Industry and related Challenges • Problem Statement
Solution Approach	• Solution Approach • Blockchain Vs. Database • Related Works
Infrastructure-level Threat Modeling	• ClaimChain System Architecture • Hyperledger Platform • Evaluation Results • Threat Modeling using Attack Tree Formalism
Application-level Fraudulent Claims Analysis and Detection	• Fraud Detection Method • Dataset Collection • Anomaly Detection Algorithm • Evaluation Results
Conclusion	• Concluding Remarks • Future Work

Agenda

Introduction	• **Background on Insurance Industry and related Challenges** • **Problem Statement**
Solution Approach	• Solution Approach • Blockchain Vs. Database • Related Works
Infrastructure-level Threat Modeling	• ClaimChain System Architecture • Hyperledger platform • Evaluation Results • Threat Modeling using attack tree formalism
Application-level Fraudulent Claims Analysis and Detection	• Fraud Detection Method • Dataset Collection • Anomaly Detection Algorithm • Evaluation Results
Conclusion	• Concluding Remarks • Future Work

Background on Insurance Industry

- Insurance: Fast growing industry requiring cutting edge technology support for background processing and handling of accident or property loss claims
  - More than $170 billion in car insurance claim payments are made by US insurance companies each year
  - Some claims are fraudulent, and cause huge capital losses for the companies
- Need to intelligently automate claim handling process to reduce the involvement of multi-domain entities to improve:
  - System Trust, Claim Processing Time, Avoid Litigation

Overview of Challenges & Threats

Challenges

- **Fraud Detection**
- **Trust Establishment**
- **Security Issues**

Threats

- **DDoS**
- **Phishing**
- **Malware**
- **Cryptojacking**
- **Timestamp Manipulation**

General Issues

- Duplicate claims
- Multiple domain entities such as police, law, insurance agents are involved in insurance claim
- Threat Scenarios

How to address these challenges and improve the process in an effective manner?

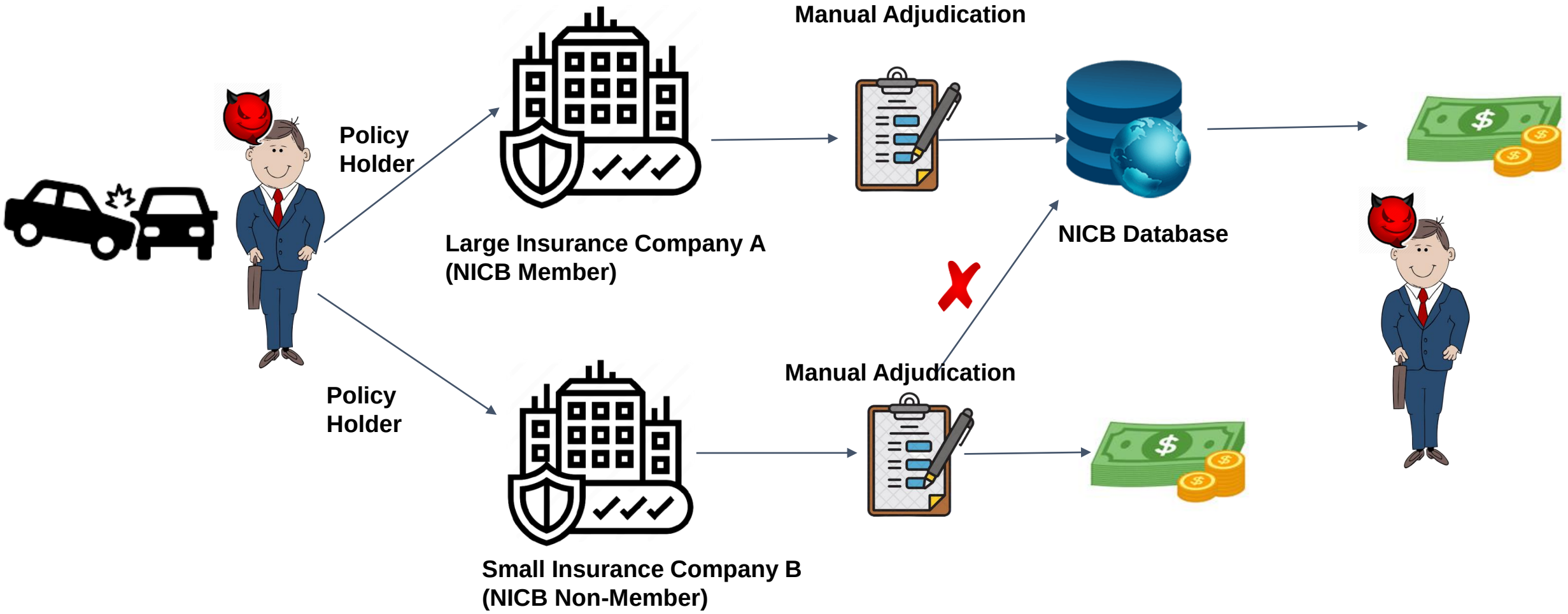
Existing System for Cooperation of Insurance Companies

In current system, National Insurance Crime Bureau (NICB) maintains a database that is accessible by participating insurance companies – Vulnerable to Loss of Integrity issues

The member companies pay large fees $ to be a part of NICB; however, not all companies (especially small-sized companies) can afford to be part of NICB

Critical issue is increased fraud in cases where participating and non-participating companies in NICB do not have a way to share their intelligence on claims - Leads to duplicate claims approval and double-spending!

Duplicate Claim Approval Scenario



Problem Statement

- Existing manual Insurance Claim processes are slow, and the cooperative mechanism for monitoring the total population of claims has several key drawbacks
- Insurance Industry is prone to cyber-attacks at both **application-level** and **infrastructure-level** which **results in loss of system integrity and/or exposure of sensitive customer data**

Research Question -1) How to utilize Blockchain Technology to create a common platform for automated insurance processing that achieves greater trust and participation?

Research Question -2) How to defend new Blockchain-based platform against fraud at the application-level and prominent cyber-threats to infrastructure-level?

Agenda

Introduction	• Background on Insurance Industry and related Challenges • Problem Statement
Solution Approach	• **Solution Approach** • **Blockchain Vs. Database** • **Related Works**
Infrastructure-level Threat Modeling	• ClaimChain System Architecture • Hyperledger Platform • Evaluation Results • Threat Modeling using attack tree formalism
Application-level Fraudulent Claims Analysis and Detection	• Consortium Blockchain Platform • Fraud Analysis Method • Anomaly Detection Algorithm • Evaluation Results
Conclusion	• Concluding Remarks • Future Work

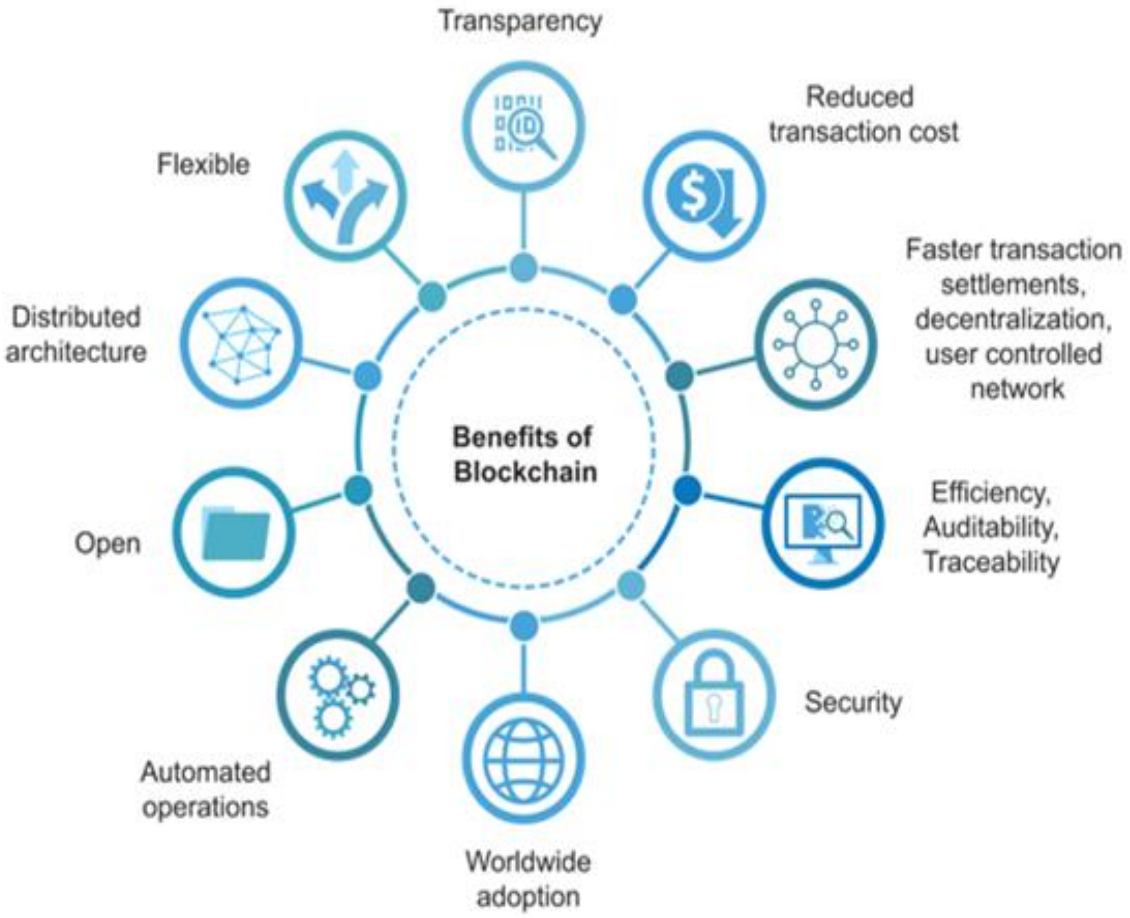
Our Solution Approach

- We propose a novel Blockchain-based solution viz., "ClaimChain" to replace existing NICB database system and manual processing
- We Identify all possible **threats** to infrastructure in ClaimChain Threat Modelling through attack tree formalism and prescribe Security Design Principles to mitigate attack risk
- Implement **machine learning** techniques via a ClaimChain Fraud Model to detect fraudulent claims at the application level

Hypothesis: The use of Blockchain technology and data-driven algorithms will achieve greater participation, processing efficiency, and trust at the application level; the implementation of Security Design Principles will mitigate attack risk while yielding greater security at the system level

Blockchain

- Blockchain is a digital ledger of transactions maintained over a peer-to-peer network
- Once a block is inserted into the chain, it cannot be removed without rewriting the entire chain
- Each peer on the network maintains a copy of the ledger
- The smart contracts that dictate transactions over tracked assets must be approved by each peer on the network
- Improves integrity of data and information
- Decentralized in nature



NICB Database vs ClaimChain

Attributes	NICB Database	ClaimChain
Authority	The database is centralized in nature	ClaimChain uses blockchain which is decentralized
Transparency	NICB administrators only decide what data to be made public	ClaimChain offers transparency
Integrity	NICB uses database that can be altered by malicious actors and can lose data integrity	ClaimChain supports integrity in data as any update made is validated through consensus algorithm
Data Handling	The data can be erased or replaced as databases utilize CRUD (Create, Read, Update, Delete)	ClaimChain offers immutability meaning no data tampering is possible within the network
Architecture	Client/server architecture	Peer-to-peer architecture
Privacy	Require administrator permission	Permissionless ledger

Related Work

Related Work	Focus	Our Focus/ Novelty
[1] Economic evaluation and optimization of the degree of automation in insurance processes.	• This paper catalogs the need for automation in the handling of insurance processes • The authors discuss the diminishing returns tied to automating business processes and attempt to prescribe a degree of automation that maximizes cost/benefit	The novelty is in layering our system with Attack tree threat modeling to categorize various attack scenarios. This system also has the feature of detecting fraudulent claims through different Machine Learning techniques and fraud detection module.
[2] Avoiding Insurance Fraud: A B lockchain based Solution for the Vehicle Sector	• This paper describe a blockchain-based solution that demonstrates the potential use of this emergent technology in the specific case of insurance fraud avoidance	
[3] A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement	• This paper explains about advantages and ways to use Machine Learning models to detect fraudulent claims and to categorize the claims into different types	

[1] Braunwarth, K.S., Kaiser, M. & Müller, AL. Economic Evaluation and Optimization of the Degree of Automation in Insurance Processes. *Bus Inf Syst Eng* 2, 29–39 (2010). https://doi.org/10.1007/s12599-009-0088-6
[2] Rui Roriz, José Luis Pereira, "Avoiding Insurance Fraud: A Blockchain-based Solution for the Vehicle Sector," Procedia Computer Science, Volume 164, 2019, Pages 211-218, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.12.174
[3] Dhieb, Najmeddine & Ghazzai, Hakim & Besbes, Hichem & Massoud, Yehia. (2020). A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2983300.

Related Work (2)

Blockchain based Frameworks	Method	Tools	Scalability	Transparency	Fraud Detection	Advantage
Inspeer	Proprietary	AI and ML predictive models	Medium	High	No	Increase and insure deductible
Friendsurance	Digital Brokerage and Bancassurance	Modular platform	High	Low	No	Reward for staying claims-free
Etherisc	Application specific smart contracts	Decentralized Insura nce Platform	High	High	No	Earn interest in cryptocurrency
B-FICA	Dynamic block	Blockchain Partition	Medium	Low	No	Resistant to Sybil attack
WISChain	Smart contract	DengLu	Low	Low	No	Reward Insurers for data packing
ClaimChain	Smart contract	Tableau, CouchDB, Wireshark	High	Low	Yes	Threat model and Protection against LOI attacks

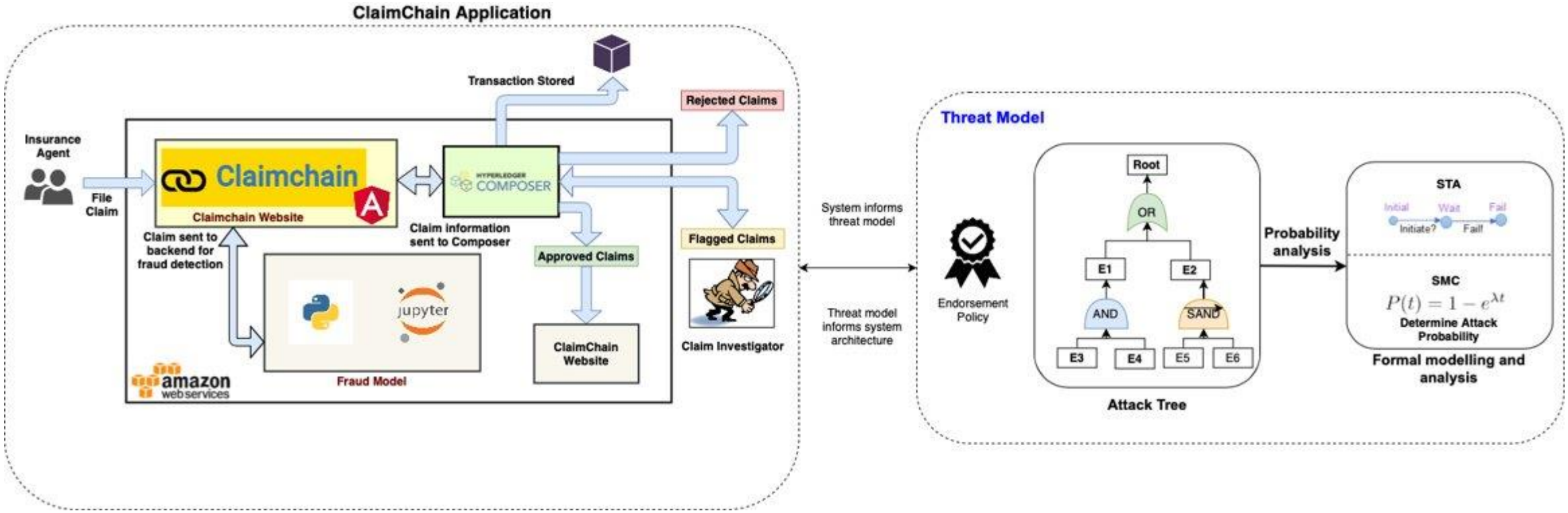
Agenda

Introduction	• Background on insurance Industry and related Challenges • Problem Statement
Solution Approach	• Solution Approach • Blockchain Vs. Database • Related Work
Infrastructure-level Threat Modelling	• **ClaimChain System Architecture** • **Hyperledger Platform** • **Evaluation Results** • **Threat Modeling using attack tree formalism**
Application-level Fraudulent Claims analysis and Detection	• Fraud Detection Method • Dataset Collection • Anomaly Detection Algorithm • Evaluation Results
Conclusion	• Concluding Remarks • Future Work

ClaimChain System Architecture

Goal of the system:

- Threat modeling to identify possible attacks
- Quantitative analysis of attacks
- To assess the risk of fraud claims on the system



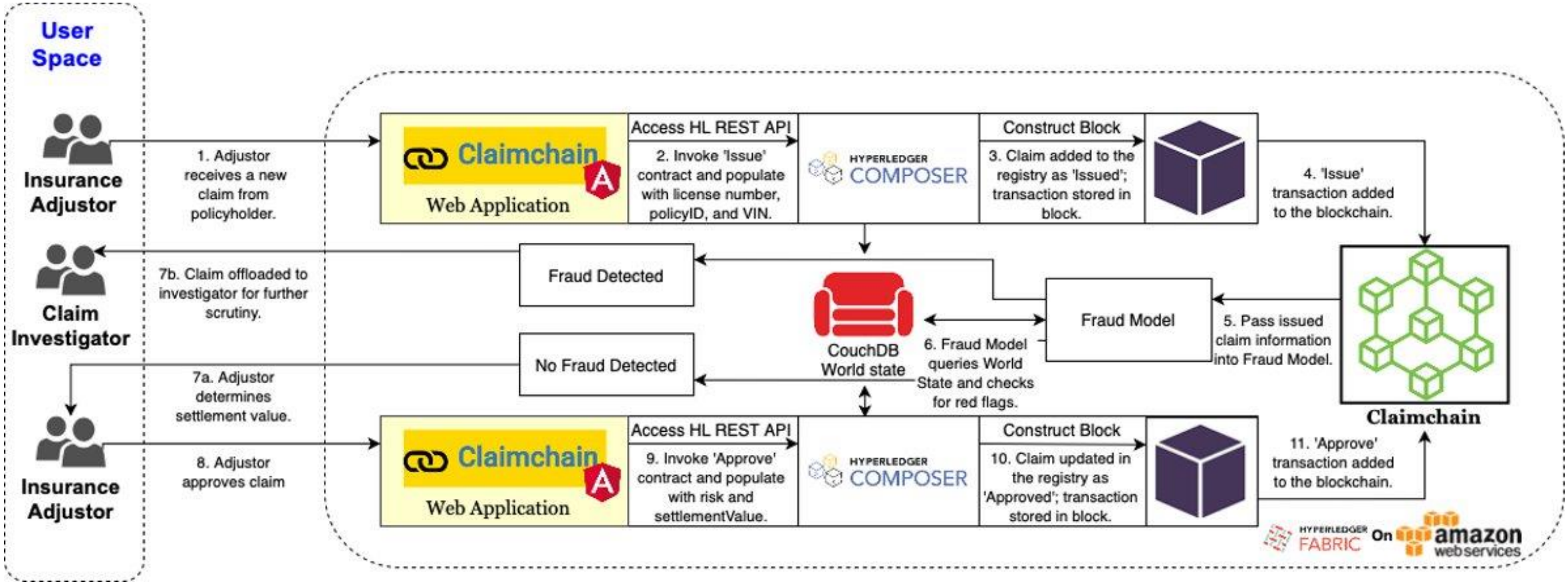
Hyperledger Composer

- Hyperledger Composer is a set of open-source tools that allows business owners, operators, and developers a way to create blockchain applications and smart contracts
- Useful for coordinating a single business' blockchain operations

ChainCode - Claim Operations

- Issue (issue)
  - Takes driver's license number, policy id, and vin stored in 'issue' object as input
  - Creates a new claim and claim id, adds it to the registry
- Approve (approve)
  - Takes a claim id, the settlement value, and risk stored in 'approve' object as input
  - Retrieves a claim from the registry and declares it approved by the agency
- ClaimChain Chaincode implements Insurance Claim Processes as Blockchain transactions

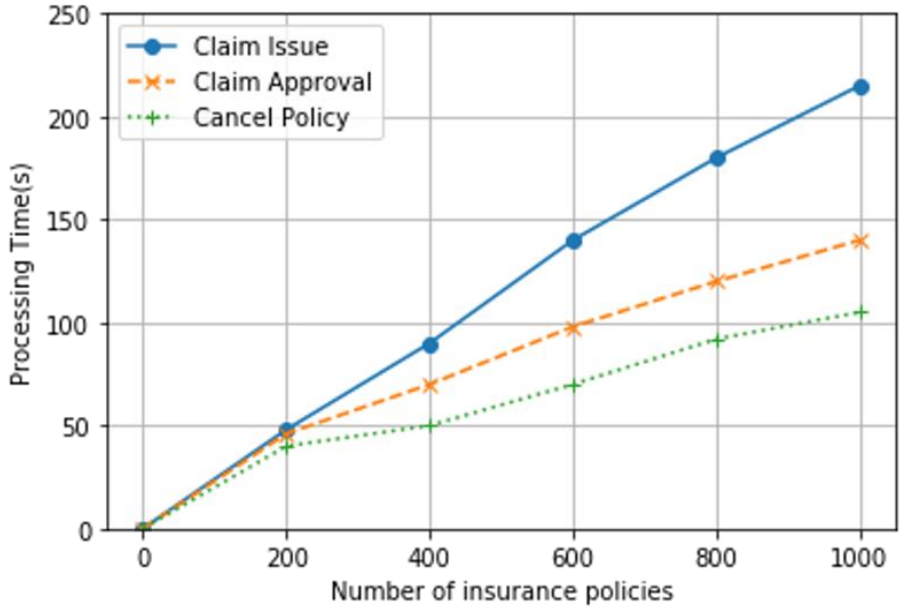
Lifecycle of Claim Asset



Hyperledger Composer REST Server

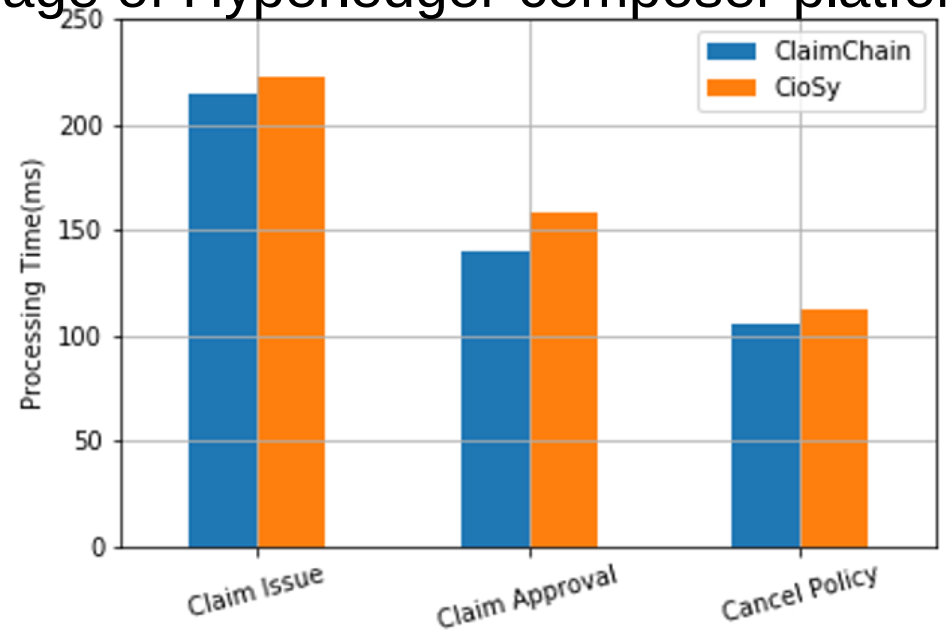
ClaimChain Scalability

- Processing time needed to validate different claim transactions
- Smart contracts yield greater scalability than manual processing



Comparison with State-of-the-art System

- CioSy[4] system has similar smart contract methods required for insurance claim processing
- As we can observe, processing time for ClaimChain system is less compared to CioSy due to the usage of Hyperledger composer platform



[4] Loukil, F., Boukadi, K., Hussain, R. and Abed, M., 2021. CioSy: A Collaborative Blockchain-Based Insurance System. *Electronics*, *10*(11), p.1343.

Threat Model

- We use Attack trees for formalizing our threat model based on threats characterized
- We further analyze the probability of occurrence of different threats which is achieved through statistical model checking using UPPAAL[5] tool
- Application layer and the infrastructure layer are the set of points for our attack surface

[5] G. Norman, D. Parker, J. Sproston, “Model checking for probabilistic timed automata”,Formal Methods in System Design, 2013

Threat Modeling via Attack Tree Formalism

- Attack trees[6] are hierarchical models that show attacker goal that can be further divided into smaller goals making use of gates such as AND, OR, and SAND
  - AND gate is activated when all of the child nodes are activated
  - OR gate is activated when at least one child node is activated
  - SAND gates are activated as the child nodes are activated from left to right based on the success of previous stage and later determines the activation of the next child node
- Application and infrastructure issues in the system are built using the attack trees

[6] R. Kumar, M. Stoelinga, ``Quantitative Security and Safety Analysis with Attack-Fault Trees'', \textit{IEEE 18th Int. Symposium on HASE}, 2017.

Attack Characterization

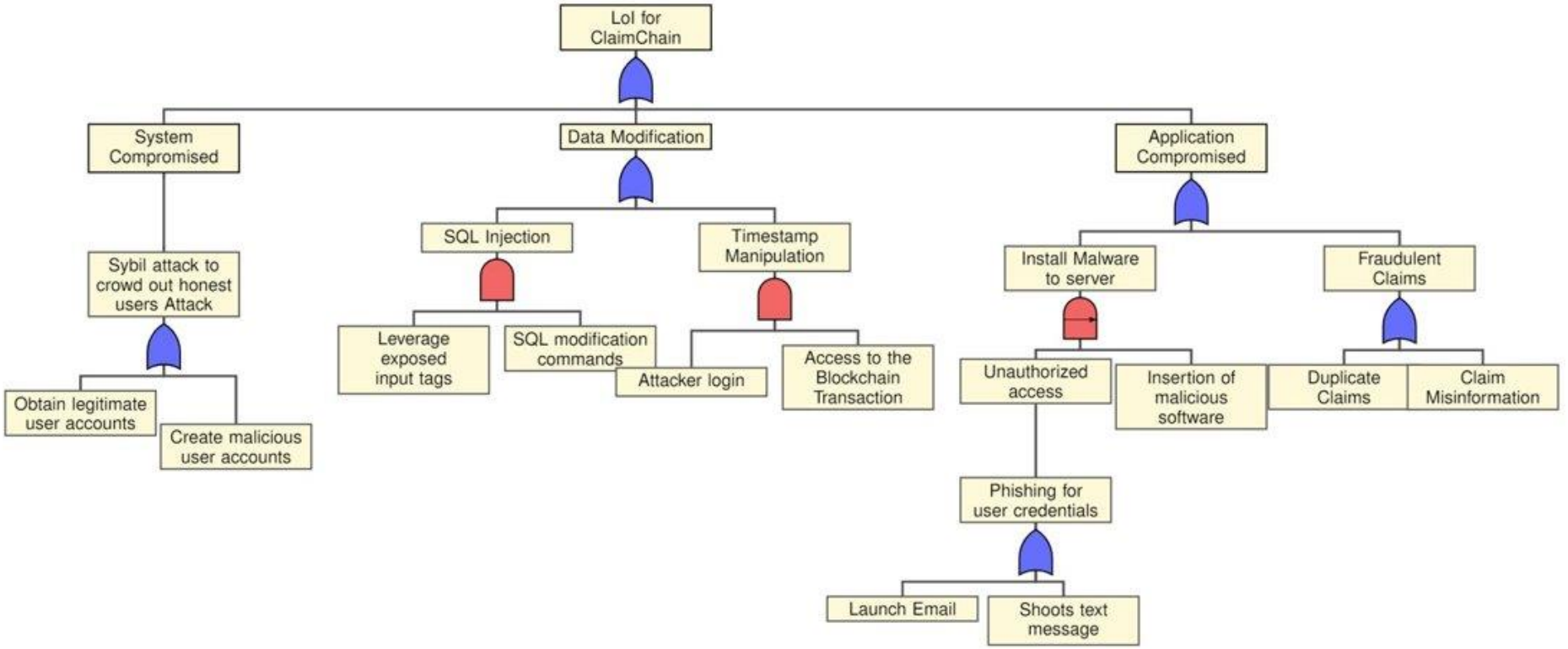
- ClaimChain platform is prone to various attacks that we categorize based on the following three types in CIA Triad
  - Loss of Integrity (LoI)
  - Loss of Confidentiality (LoC)
  - Loss of Availability (LoA)
- These attacks make ClaimChain's application layer and the blockchain infrastructure vulnerable. Such attacks need to be considered in securing our platform.

LoI Attacks

- **Sybil Attack**: Create's huge amount of fake accounts within a block to subsequently gain control over the block
- **Injection Attack**: Use malicious data to attack software systems
- **Fraudulent claims**: Duplicate claims or provide false information in the claim form
- **Malware**: Modification of system parameters toward non system-critical functions
- **Timestamp Manipulation**: Stall efforts to identify fraudulent claims by modifying transaction timestamps

Attack Tree for LoI Attacks

- Attack Tree formalization of attacks resulting in Loss of ClaimChain System Integrity

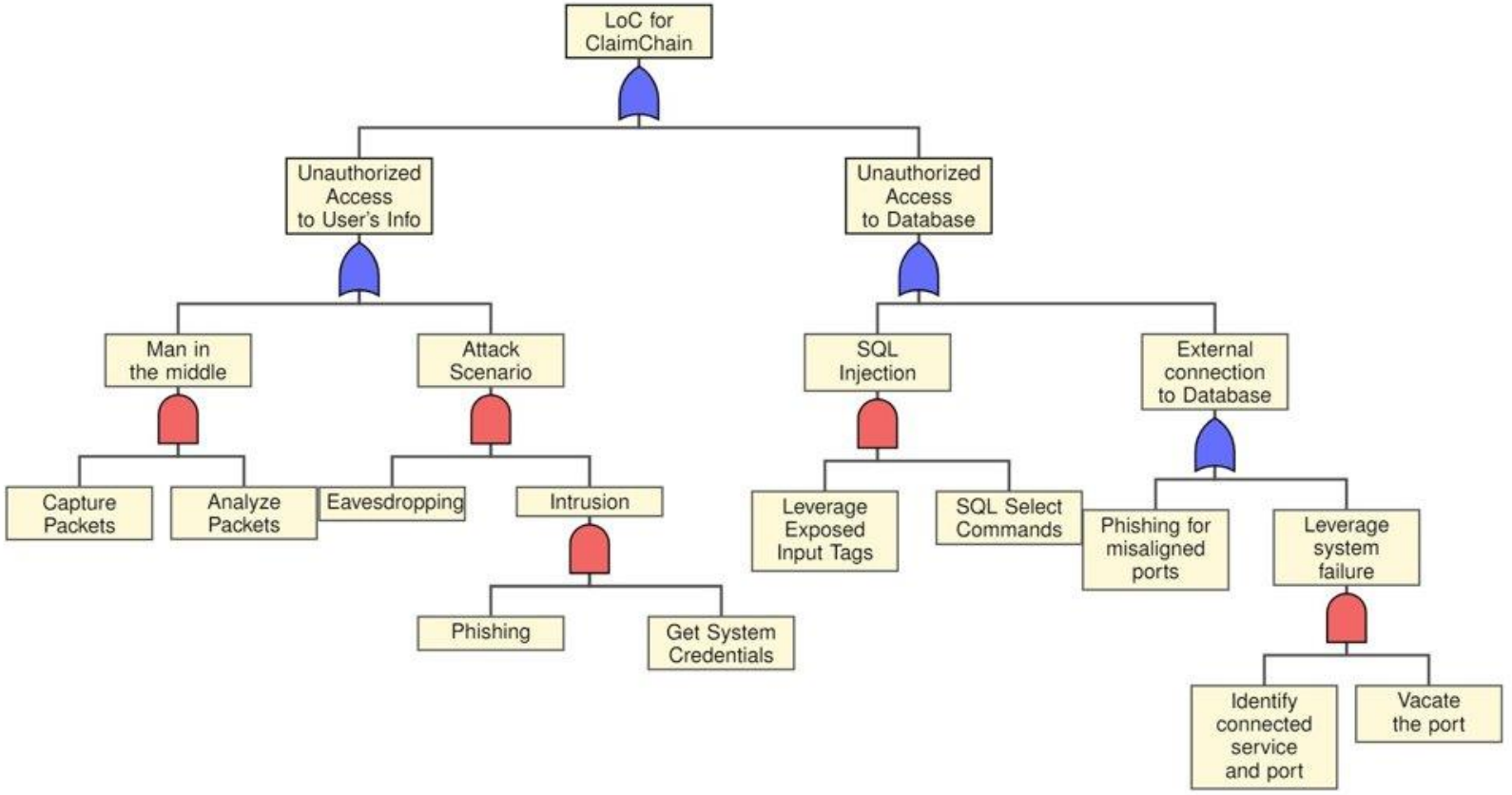


LoC Attacks

- **Man in the Middle**: Intercept packets of information in transit
- **SQL Injection**: Inserting or "injecting" of a SQL query into the input fields/forms of the application at the client side
- **Phishing**: Send a fraudulent message designed to trick the users into revealing sensitive information such as user credentials, credit card details etc.

Attack Tree for LoC Attacks

- Attack Tree formalization of attacks resulting in Loss of ClaimChain System Confidentiality

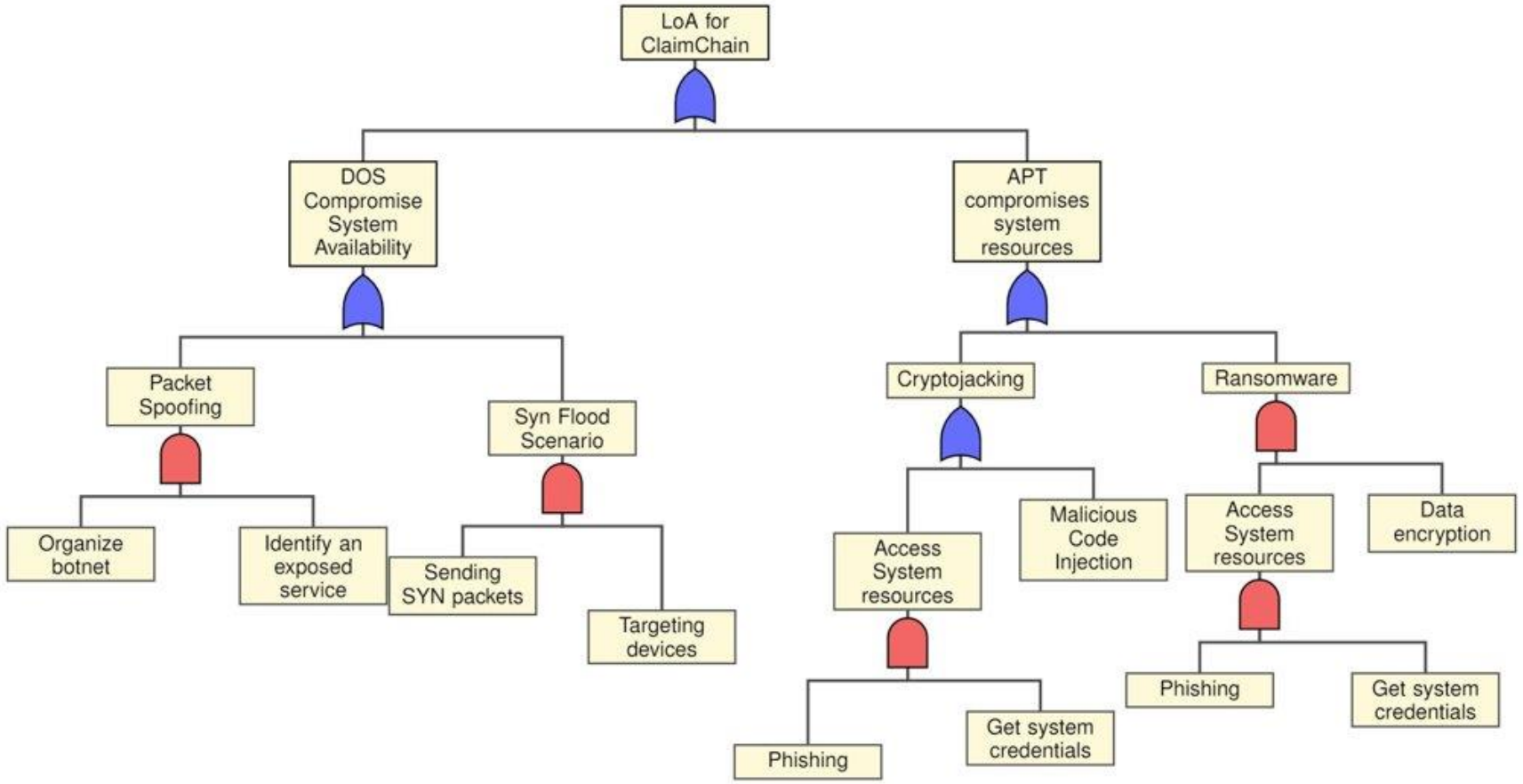


LoA Attacks

- **Packet Spoofing**: Falsifying content of a Source IP header
- **Syn Flood**: Overloads available ports to prevent legitimate traffic from reaching a server
- **Cryptojacking**: Attempt to insert crypto-mining scripts into critical computer resources
- **Ransomware**: Threatens to publish the victim's files and in turn their data or perpetually block access to these files unless a ransom is paid

Attack Tree for LoA Attacks

- Attack Tree formalization of attacks resulting in Loss of ClaimChain System Availability



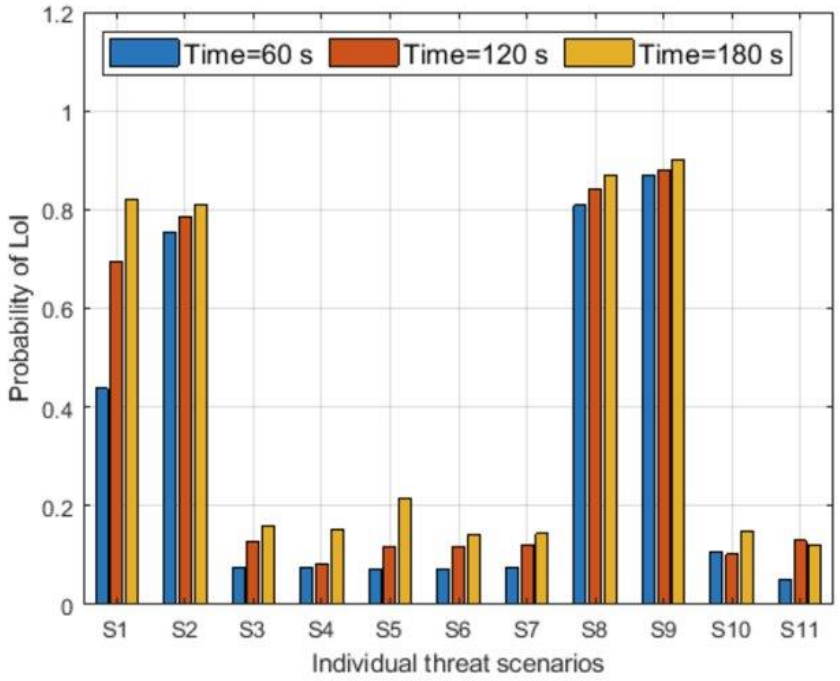
Probability Analysis

- Probability analysis uses a λ value to represent exponential rate for an attack
- $P(t) = 1 - e^{\lambda t}$
- Based on the λ values, Probability is calculated over Time (i.e., 60 sec, 120 sec, 180 sec)
- Considered different kinds of attack scenarios based on CIA Triad
- Calculated our results of probability of attacks using our ClaimChain architecture
- Assigned a λ value to a leaf node and utilized a small positive constant (K) of ~0.0002 for all the remaining nodes while calculating the likelihood of particular attack

Probability Result – Loss of Integrity

- Based on λ values, probability of different LoI attacks is calculated using UPAAL tool

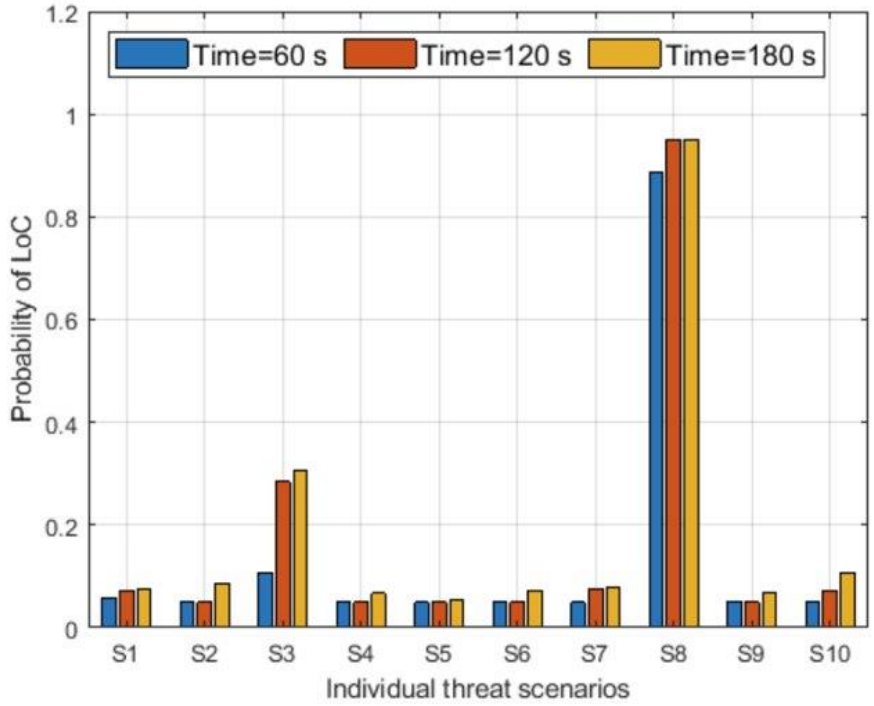
Type	Threat Events	λ
S1	Obtain legitimate user accounts	0.01
S2	Create malicious user accounts	0.03
S3	Leverage exposed input tags	0.03
S4	SQL modification commands	0.01
S5	Attacker login	0.03
S6	Access to the Blockchain Transaction	0.007
S7	Malicious software	0.03
S8	Duplicate Claims	0.04
S9	Claim Misinformation	0.05
S10	Email	0.03
S11	Instant message	0.02



Probability Result – Loss of Confidentiality

- Based on λ values, probability of different LoC attacks is calculated using UPAAL tool

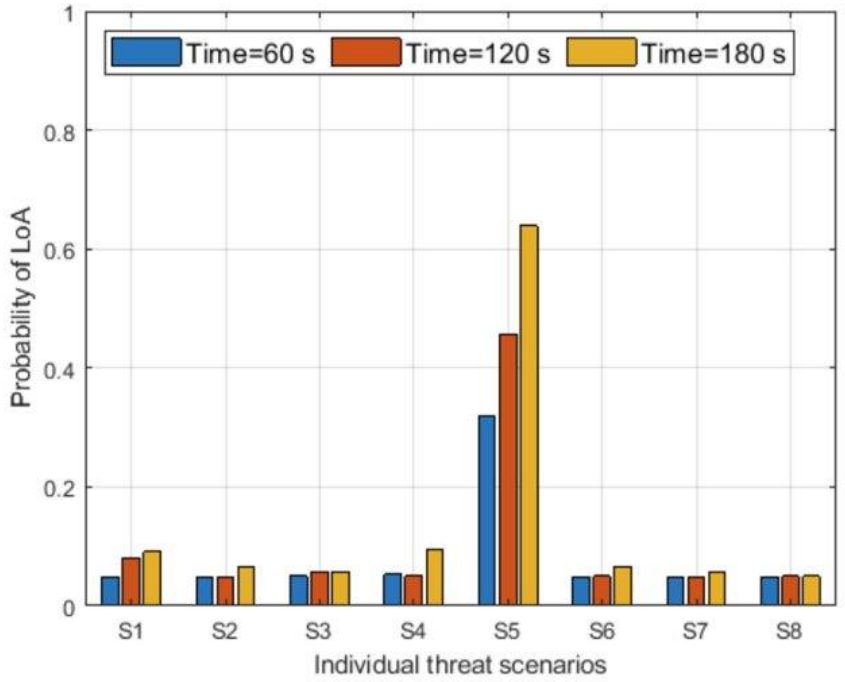
Type	Threat Events	λ
S1	Capture Packets	0.04
S2	Analyze Packets	0.007
S3	Application Eavesdropping	0.04
S4	Phishing	0.05
S5	Get System Credentials	0.007
S6	Leverage Exposed Input Tags	0.03
S7	SQL SELECT commands	0.03
S8	Phishing for misaligned ports	0.04
S9	Identify connected service and ports	0.007
S10	Vacate the ports	0.02



Probability Result – Loss of Availability

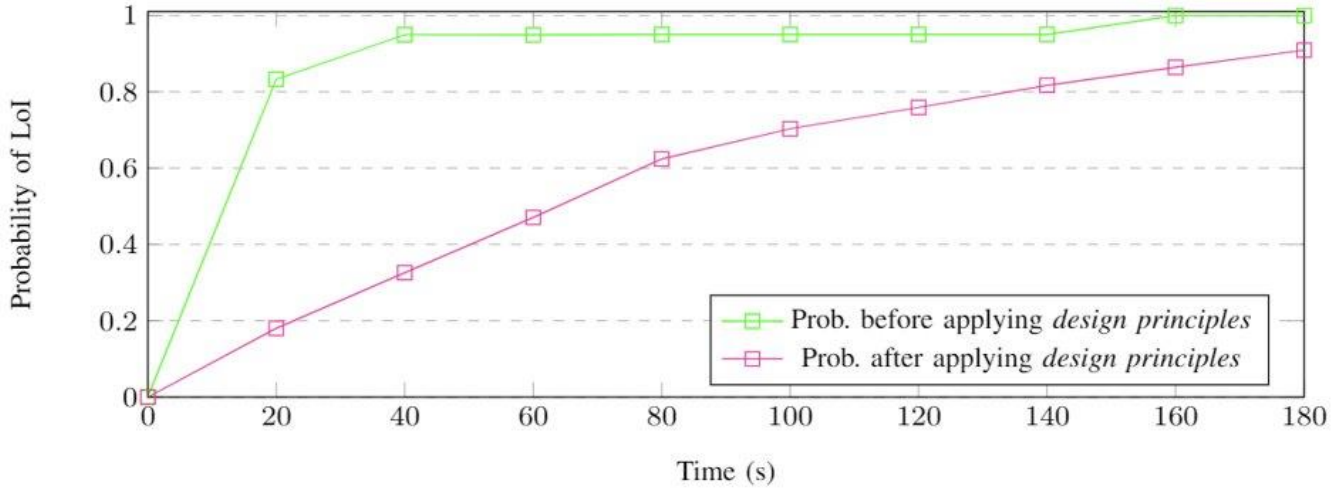
- Based on λ values, probability of different LoA attacks is calculated using UPAAL tool

Type	Threat Events	λ
S1	Organize botnet	0.02
S2	Identify an exposed service	0.04
S3	Sending SYN packets	0.005
S4	Targeting devices	0.04
S5	Malicious Code Injection	0.0055
S6	Data encryption	0.0055
S7	Phishing	0.05
S8	Get system credentials	0.007



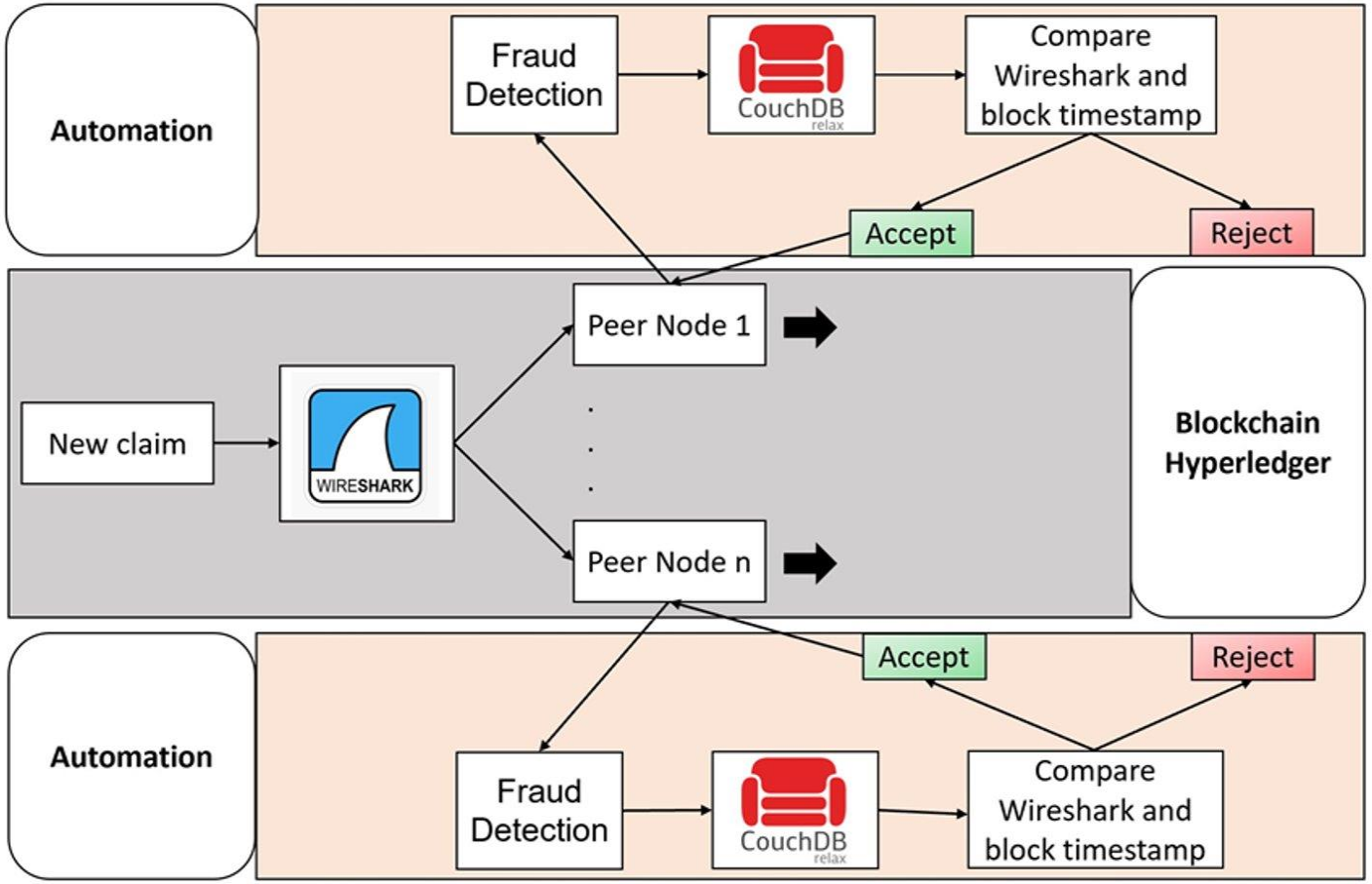
Probability Reduction

- To address the high probability attacks and to make the system better, certain design principles can be considered
- Existing works such as NIST SP800-160 explains different security design principles which helps in improving efficiency and secure system architecture



Timestamp Manipulation

- Wireshark record the time for captured packet
- Recorded time is sent to the peers
- Peers compare Wireshark and block time stamp
- If there is a significant difference in time, the claim is rejected by the peers

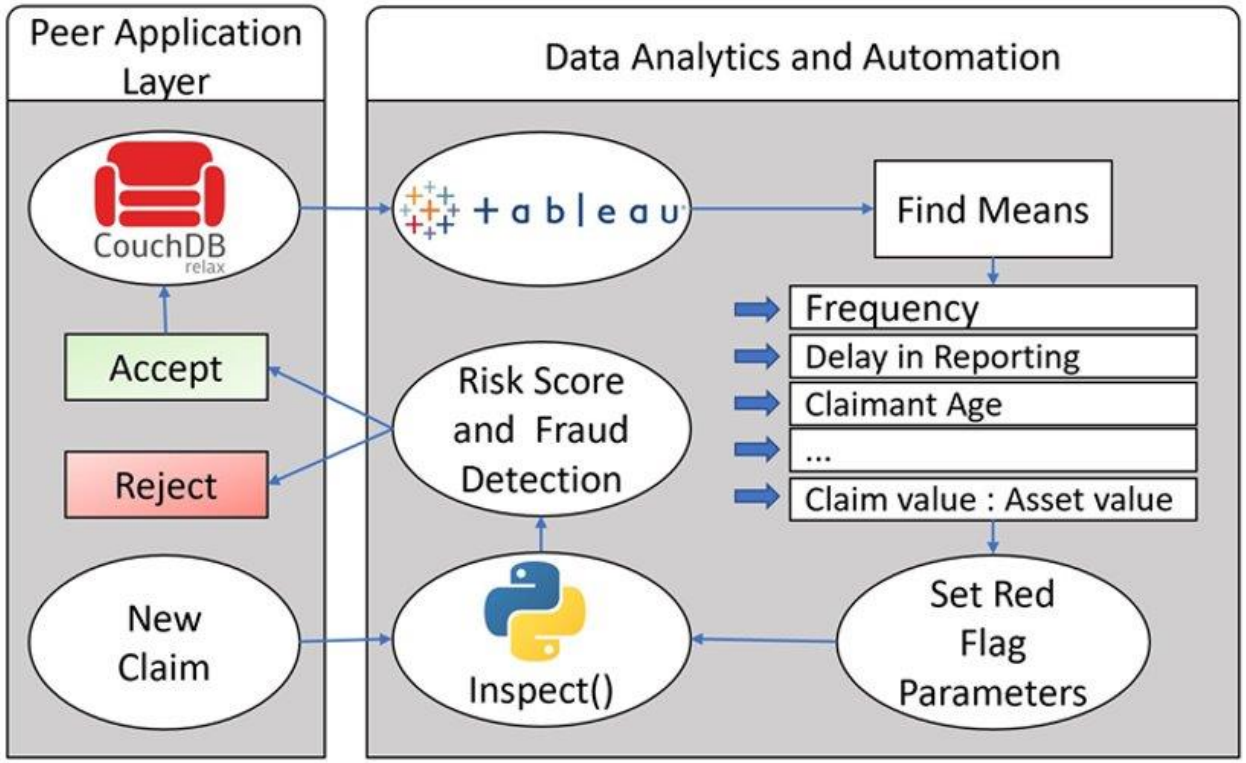


Agenda

Introduction	• Background on Insurance Industry and related Challenges • Problem Statement
Solution Approach	• Solution Approach • Blockchain Vs. Database • Related Works
Infrastructure-level Threat Modeling	• ClaimChain System Architecture • Hyperledger Platform • Evaluation Results • Threat Modeling using attack tree formalism
Application-level Fraudulent Claims Analysis and Detection	• **Fraud Detection Method** • **Dataset Collection** • **Anomaly Detection Algorithm** • **Evaluation Results**
Conclusion	• Concluding Remarks • Future Work

Fraud Detection

- Duplicate claim detection
- Give risk score to individual claim
- Reject the high-risk claims
- Legit claim integrates into the Apache Couch database



Red Flag Conditions

- A Red Flag is a **suspicious** circumstance, pattern, practice, or specific activity that indicates the possibility of identity theft
- In our ClaimChain system, we considered the most frequently occurring red flag conditions.
  - Difference in policy issue date and policy report date
  - If policy is taken within 10 days of incident, then it has risk score
  - If someone tries to claim for fraud claims again and again then we will update all the insurance agents to verify it multiple times before we confirm it and approve the request
  - You participate in sports outside of work
  - If the claim is more than the insured asset
  - When there is a delay in reporting the injury

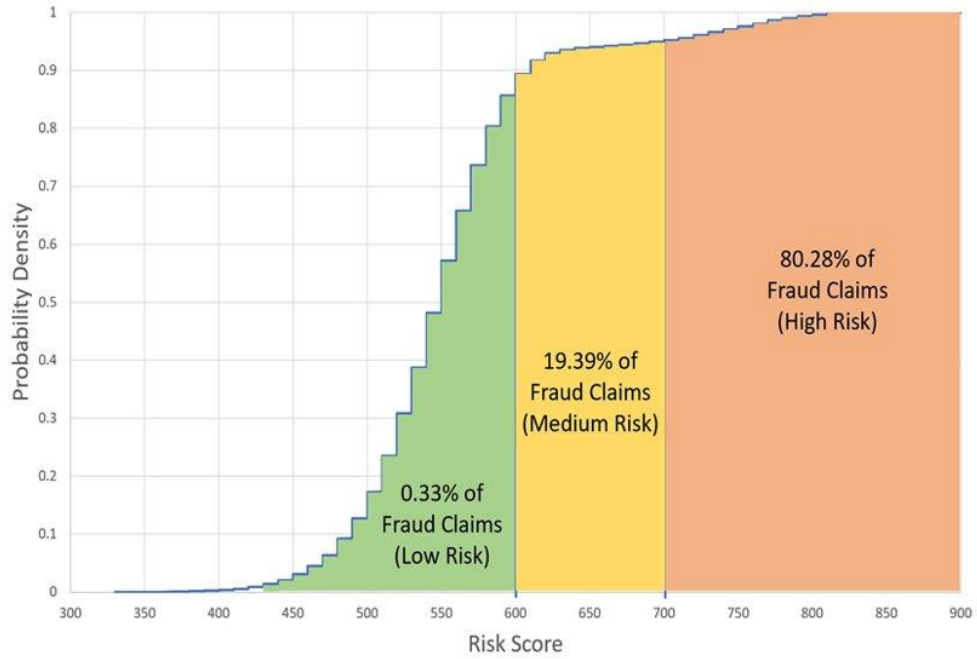
Dataset Collection

- The dataset [7] we are using is publicly available dataset on a blog by Charlie Berger to describe insurance transactions in 2015
- The dataset describes insurance vehicle accident claims for an undisclosed insurance company. It contains 15,430 claims; each claim comprises 33 attributes describing the following components:
  - Customer demographic details (Age, Sex, Marital Status)
  - Purchased policy (Policy Type, Vehicle Category, No: of supplements, Agent Type)
  - Claim circumstances (day/month/week claimed, policy report filed, witness present, past days between incident-policy report and incident-claim)
  - Other customer data (number of cars, previous claims, Driver Rating)

[7] Dataset for Fraudulent Claims: https://blogs.oracle.com/machinelearning/fraud-and-anomaly-detection-made-simple

Fraud Model Evaluation

- We have evaluated 15,430 claims out of which 924 are fraudulent
  - 0.33% fraud claims on the **low-risk** score area
  - 19.39% fraud claims on the **medium-risk** score area
  - 80.28% fraud claims on the **high-risk** score area



Anomaly Detection

- Anomaly Identification (also known as outlier analysis) is a data mining step that detects data points, events, and/or observations that differ from the expected behavior of a dataset
- A typical data might reveal significant situations, such as a technical fault, or prospective possibilities, such as a shift in consumer behavior
- Anomaly detection is increasingly automated thanks to machine learning

ML Models Performance

- Various Machine Learning Models are used based on their unique significance Among all XGBoost has highest attack detection rate i.e., it identifies the fraud claim anomaly pattern more accurately
- The data is divided in the ratio of 80 and 20 to train and test, respectively

ML Model	Accuracy	Precision	Recall	F-Score
Random Cut Forest	0.96	0.96	0.96	0.96
K-Nearest Neighbors	0.82	0.96	0.67	0.78
Logistic Regression	0.76	0.70	0.74	0.72
XGBoost	0.98	0.98	0.98	0.98

Fraud Data Analysis

Performance Metrics & Statistical Significance

At a descriptive level, we first summarized a macro-profile for 924 cases of fraud from our dataset:

- 88.6% of the fraudsters were male;
- 67.2% of the fraudsters were married;
- The average age of fraudsters was 38.2 years;
- 51.7% of the fraudsters has rating greater than 2 i.e., 3 & 4
- 98.2% of fraudsters doesn't have police reports
- 99% of fraudsters doesn't have a witness

[8] Matheus Kempa Severino, Yaohao Peng, "Machine learning algorithms for fraud prediction in property insurance: Empirical evidence using real-world microdata," Machine Learning with Applications, Volume 5, 2021, 100074, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2021.100074.

Key Results

Architecture Merits:

- *ClaimChain* consortium Blockchain is a futuristic alternative to the NICB's ISO database achieving greater participation, processing efficiency, and trust

Threat Modeling Merits:

- *ClaimChain* fraud model is effective at detecting known red-flags with up to a 98% detection accuracy and effective at combatting duplicate claims among participating agencies

Fraud Detection Merits:

- *ClaimChain* security design principles are effective in protecting insurance claims processing system as seen in reduction of probability of Loss of Integrity before and after application

Agenda

Introduction	• Background on Insurance Industry and related Challenges • Problem Statement
Solution Approach	• Solution Approach • Blockchain Vs. Database • Related Works
Infrastructure-level Threat Modeling	• ClaimChain System Architecture • Hyperledger Platform • Evaluation Results • Threat Modeling using attack tree formalism
Application-level Fraudulent Claims Analysis and Detection	• Fraud Detection Method • Dataset Collection • Anomaly Detection Algorithm • Evaluation Results
Conclusion	• **Concluding Remarks** • **Future Work**

Conclusion

- We proposed a Blockchain-based solution for Insurance Claims Processing viz., ClaimChain
- We showcased different claim transactions implemented in Claimchain's Blockchain Network
- We utilized attack trees for the threat analysis and probability determination
- We examined different attacks based on scenarios and focused on Loss of Integrity attacks to secure the ClaimChain
- We presented the key results of our solution that include:
  - Key Result 1—Architecture Merits
  - Key Result 2—Threat Modelling Merits
  - Key Result 3—Fraud Detection Merits

Future Work

- Expand Fraud detection module to detect anomalous claim data and help the insurance company to conduct fraud analytics at large-scale
- Expand the security suite of our Blockchain-based application to support detection and mitigation of prominent threats e.g., Timestamp Manipulation, Sybil

Thank you for your attention!

Any questions?