**Initial Project Proposal Report**

**Project Summary**

The project's primary objective is to implement the Kafka component of a network telemetry system with the purpose of reconstructing traffic routes in a Software-Defined Networking (SDN) environment. Building on the existing Kafka-based event streaming framework, the system will handle flow records, process Open vSwitch data, and incrementally reconstruct end-to-end network pathways. Additionally, there is expected integration with GoFlow2 for initial data capture and testing of the Kafka infrastructure, though its implementation status is uncertain as of this initial proposal. If time permits, the project will explore adding other system components, paving the way for future teams to expand.

**Background**

The project builds upon previous research that demonstrated the potential of leveraging tools such as GoFlow2 and Kafka for network path reconstruction and monitoring. GoFlow2 is used to capture flow records from switches and convert them into a format (such as JSON) suitable for streaming. Our specific focus is on implementing the Kafka component, which processes and streams these flow records for real-time path reconstruction and network telemetry. While GoFlow2 collects the flow data, Kafka handles the distributed data processing to manage large volumes, improve path reconstruction accuracy, and provide insights into network performance.

**Potential Contributions and Advantages**

This system attempts to address accurate and real-time path reconstruction in SDN systems, which can be a major difficulty in network monitoring. Complex network topologies may demand off-the-shelf solutions that aren't scalable or have real-time processing capabilities. Using the distributed streaming infrastructure provided by Kafka, this project offers the following benefits:

**Scalability:** Handling large volumes of network telemetry data requires a highly scalable and fault-tolerant data streaming solution, which Kafka’s design effectively provides.

**Real-Time Processing:** Kafka enables rapid flow record processing, allowing real-time path reconstruction.

**Increased Accuracy:** Network path representations are more accurate when partial path updates and flow records are used to reconstruct data flow incrementally.

**Broader Impacts**

Several benefits could result from this project's successful implementation.

**Operational Benefits**: Enhances network performance and reliability by providing network managers with advanced tools for optimizing and managing SDN networks.

**Sustainability Benefits**: More efficient resource utilization and reduced energy consumption in data centers through effective network management.

**Infrastructure Benefits**: Improved network performance and visibility helps ensure the stability and efficiency of internet infrastructure, benefiting both users and businesses.

**Approach and Methodology**

**Strategy**

Integration of Kafka and GoFlow2**:** Configure Kafka to handle flow data from OVS switches.

Kafka Consumer Agents**:** Implement Kafka consumer agents in Go for processing flow records and reconstructing network paths.

Testing and Validation**:** Deploy the system in a cloud environment to test and validate its performance in reconstructing network paths and handling high data volumes.

**Methodology**

Configuration**:** Set up Kafka brokers, ZooKeeper clusters, and topic queues for data streaming.

Development**:** Code the Kafka consumer agents and integrate them with GoFlow2.

Deployment: Use Mininet for network simulation and ONOS for SDN control to emulate the network environment.

Evaluation**:** Measure the system's performance in terms of path reconstruction accuracy and latency.

**Tools and Resources**

Software**:** Kafka, GoFlow2, Mininet, ONOS, Docker.

Hardware**:** Standard cloud resources for deployment and testing.