Security Design for Natural Language Interaction Protocol

Elisa Bertino, Jan Bieniek, Yan-Ming Chiou, Raj Dodhiawala, Sanjay Aiyagari, Sugih Jamin, Ashish Kundu, Jonathan Lenchner, Matthew Louis Mauriello, Abhay Ratnaparkhi, Mohamed Rahouti, Tom Sheffler, Chien-Chung Shen, Dinesh Verma, Jinjun Xiong, Luyi Xing, Wenpeng Yin, Hasan Zengin

Purdue University, Fordham University, SRI International, Independent, RedHat, University of Michigan, Cisco, IBM, University of Delaware, IBM, Fordham University, Independent, University of Delaware, IBM, University at Buffalo, Indiana University, Penn State University, University of Michigan

INTRODUCTION

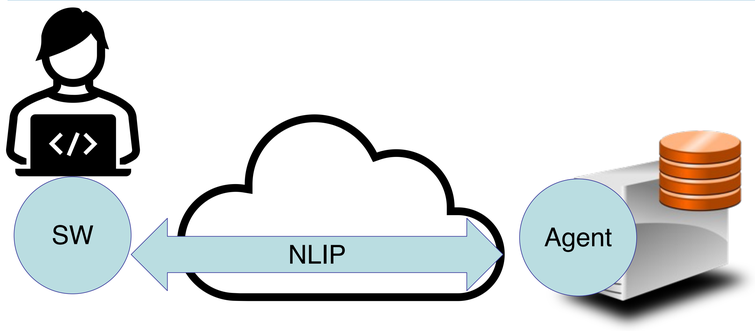
NLIP - Natural Language Interaction Protocol

- A protocol for communication among AI Agents
- A Human user considered equivalent to an AI Agent
- Allows a Universal Application-Level Transport Protocol
- Targeted for inter-organization Agent communication

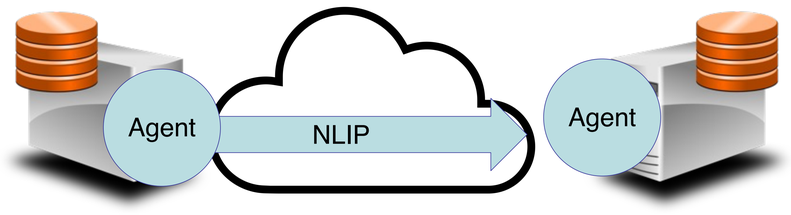
NLIP Objectives

- Unlock value of generative AI to society at large
- Inspired by HTTP unlocking value of Interent technology

Human Interacting with an Agent



Inter Agent Communication



DESIGN PHILOSOPHY

Leverage existing mechanisms

Maintain Simplicity in Protocol Design

Open Source and Open Standards

Royalty Free Protocol Design

Standardize

- Standardization being done as an ECMA Technical Committee

SECURITY CHALLENGES

As an inter-organization protocol, NLIP Security needs to address:

- **Authentication**: Prevent unauthorized access to services
- **Denial of Service**: Agent operations using LLMs are slow, exposing new ways to launch denial of service attacks.
- **Open Access**: Enable unauthenticated users to access some services (e.g. browsing at an eCommerce Agent)
- **Third Party Support**: Leverage existing mechanisms for Single Sign On/Credentials available from 3rd parties.

Security Mechanism encoded in NLIP

NLIP Supports Carrying of Authentication Tokens:

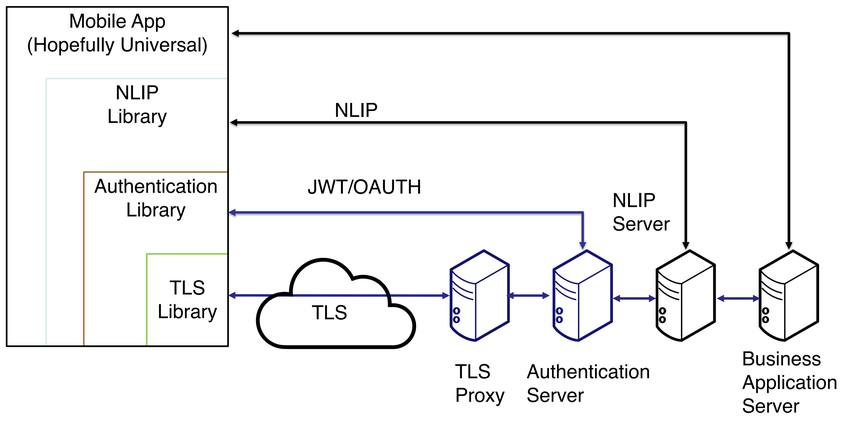
- **Format**: A special message to carry tokens
- **Subformat**: Authentication enables transfer of opaque tokens

NLIP messages consist of submessages with format and subformat

NLIP Approach for reusing security mechanism

- Leverage TLS to encrypt data in motion
- Leverage existing Authentication Services

NLIP Notional Architecture

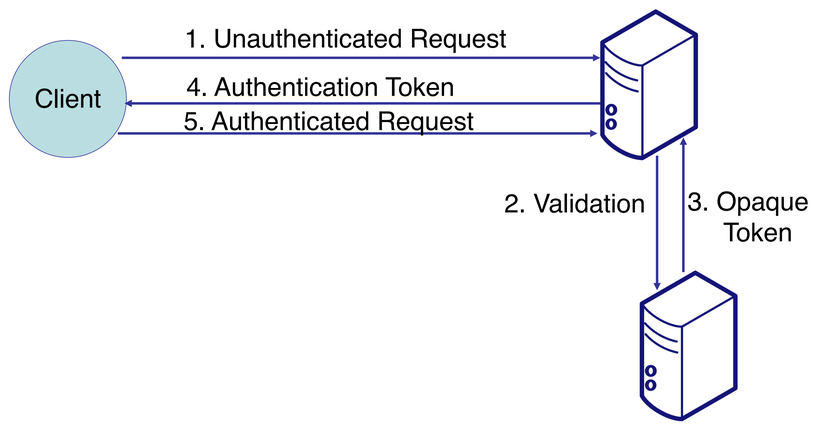


REFERENCES

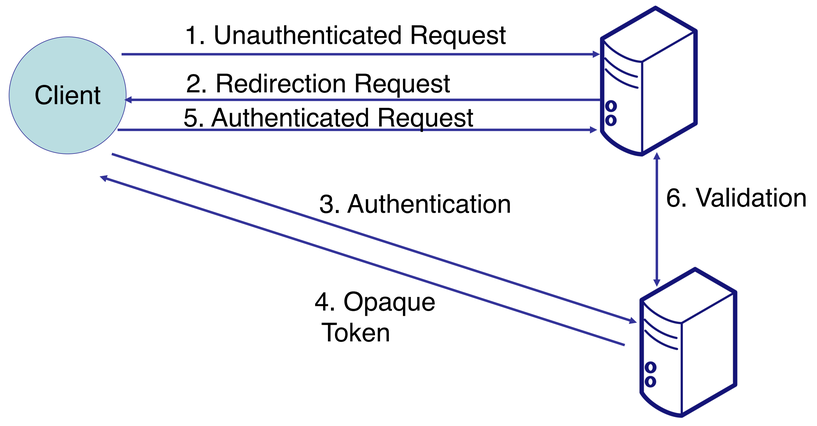
1. NLIP Specifications & Open Source: `github.com/nlip-project`
2. ECMA Technical Committee 56 on NLIP

SECURITY MECHANISMS

NLIP Authentication



NLIP Authentication (Alternative)



DDOS Prevention

NLIP prevents DDOS attacks using multiple approaches

- A control mechanism to define limits on interactions
  - the amount of interaction history at each side
  - rates of requests
- Performance Optimizations
  - Special messages for error controls
  - Special mechanisms for large file uploads