산업 제어 시스템 프로토콜 분석을 위한 메시지 타입 정의 방법

심규석, 구영훈, 신무곤, 김명섭
고려대학교

{kusuk007, gyh0808, tm0309, tmskim} @korea.ac.kr

The Method of Define Message Types for Analysis of Industrial Control System Protocol

Kyu-Seok Shim, Young-Hoon Goo, Mu-Gon Shin and Myung-Sup Kim
Korea Univ.

요 약

본 논문은 대부분 비공개 되어 있는 산업 제어 시스템 프로토콜(Industrial Control System Protocol)의 구조를 분석하기 위해 프로토콜 내에서 발생하는 메시지들의 타입을 분류하는 방법을 제안한다. 메시지들의 타입을 분류하는 것은 프로토콜이 발생하는 메시지들의 종류를 구분하는 것으로 프로토콜 구조 분석을 위한 가장 핵심적인 단계이다. 기존에는 단순히 클러스터링 알고리즘을 통해 메시지 타입을 구분하였지만, 분류 시간이 많이 소요되고, 계산 처리량이 많아서 많은 양의 데이터를 처리하지 못하는 단점이 존재하였다. 따라서 본 논문에서는 산업 제어 시스템 프로토콜의 특징을 이용하여 많은 양의 데이터를 처리할 수 있는 방법을 제안한다. 제안하는 방법은 총 4 단계로 메시지 사이즈 그룹핑 단계, 메시지 유사도 클러스터링 단계, 고정 필드 추출 단계, 고정 필드를 이용한 메시지 합병 단계로 구성된다. 본 방법을 통해 상대적으로 많은 데이터 양을 같은 메시지 타입끼리 분류하는 것을 확인할 수 있다.

Ⅰ. 서 론

산업 제어 시스템 프로토콜은 산업 현장에서 제조, 생산, 발전, 가공, 제련, 기반 시설, 설비를 바탕으로 하는 작업 공정에 필요한 데이터를 전송하고, 명령어를 송, 수신하며 이러한 작업들이 자동으로 이루어질 수 있게 만들어 주는 역할을 한다[1].

오늘날 산업 현장에는 자동화가 도입되면서 네트워크는 필수가 되고 있다. 또한 산업 현장의 효율적인 데이터 전송을 위해 대부분의 산업용 프로토콜은 자체적으로 개발된 프로토콜을 사용한다. 자체적으로 개발된 프로토콜은 기술 유출 및 보안상의 이유로 대부분 비공개한다. 하지만, 네트워크 관리 입장이나 산업 현장 관리 입장에서는 프로토콜 분석으로부터 얻을 수 있는 정보들이 있기 때문에 프로토콜의 구조를 파악하는 것은 많은 비용을 줄일 수 있다.

기존 프로토콜의 구조를 추론하기 위해 수동적으로 분석하였다. 그러나 산업용 프로토콜의 종류가 다양하고, 환경적 상황에 따라 프로토콜의 규격은 변화되기 때문에 많은 어려움이 있다. 따라서 최근에는 자동으로 프로토콜의 구조를 추론할 수 있는 방법들이 연구되고 있다. 기존 자동화된 프로토콜 역공학 기술들이 많이 발표되었지만, 대부분 인터넷 프로토콜을 대상으로 개발되었고 이러한 기술들은 산업 제어 시스템 프로토콜의 구조를 분석하기에는 부족한 부분이 있다[2].

따라서 산업 제어 시스템 프로토콜을 분석할 수 있는 방법이 필요하다. 본 논문에서는 산업 제어 시스템 프로토콜 분석을 위해 산업 제어 시스템 프로토콜의 특징을 이용하여 많은 양의 데이터를 신속하고 정확하게 처리할 수 있도록 프로토콜 메시지들을 분류하는 방법을 제안한다. 제안하는 방법은 총 4 단계로 메시지 사이즈 그룹핑 단계, 메시지 유사도 클러스터링 단계, 고정 필드 추출 단계, 고정 필드를 이용한 메시지 합병 단계로 구성된다.

가변길이 필드가 많이 존재하지 않는 산업 제어 시스템 프로토콜의 특징을 이용하여 메시지 사이즈가 같으면 같은 타입의 메시지로 판단할 수 있기 때문에 메시지 사이즈를 기반으로 그룹핑 할 수 있고, 사이즈는 같지만 다른 타입의 메시지들이 존재할 수 있기 때문에 그룹핑 된 메시지들끼리 유사도를 판별하여 다른 타입의 메시지들을 분류할 수 있다. 분류된 메시지들 내에서 고정 필드를 추출하고, 추출된 고정 필드를 이용하여 사이즈는 다르지만 같은 타입의 메시지들을 같은 타입으로 합병할 수 있다.

본 논문의 구성은 본장 서론에 이어, 본론에서 산업 제어 시스템 프로토콜 구조 추론을 위한 메시지 타입 분류 방법에 대해 설명하고, 마지막으로 결론 및 향후 연구를 언급한다.

이 논문은 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2018R1D1A1B07045742)과 2018 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2018-0-00539-002, 블록체인의 트랜잭션 모니터링 및 분석 기술개발)

Ⅱ. 본론

본 논문에서 제안하는 산업 제어 시스템 프로토콜 메시지 타입 분류 방법은 다음 그림 1 과 같다. 메시지 그룹핑 단계에서는 단일 프로토콜에서 발생하는 메시지를 입력 받아서, 메시지의 사이즈별로 분류하는 단계이다. 산업 제어 시스템 프로토콜에서 발생하는 메시지에는 가변길이를 갖는 필드들이 많이 존재하지 않기 때문에 다음의 단계를 포함하여, 다음 단계인 유사도를 기준으로 메시지 타입 분류과정에서 시스템 과부화를 감소시킬 수 있다. 다음단계인 유사도를 기준으로 메시지 타입을 분별하는 메시지 타입 클러스터링 단계에서는 클러스터링 알고리즘인 K-means 알고리즘을 이용하여 메시지 타입을 분류한다. 이때, 사이즈로 분류된 메시지 타입마다 몇 개의 군집으로 분류되는지 알 수 없기 때문에 우리는 Elbow-method 를 통해 최적의 K 값을 찾아 K-means 입력으로 넣는다.

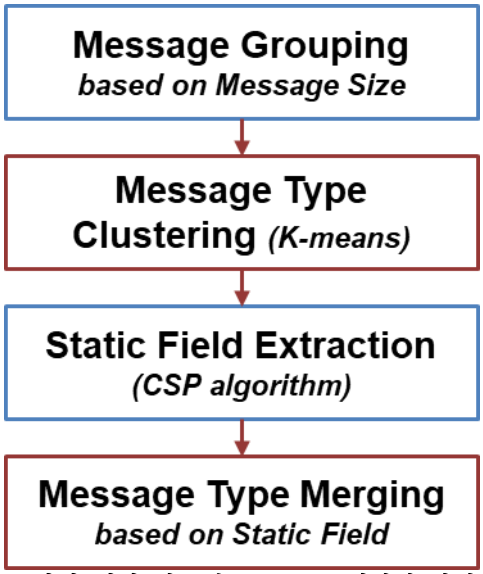

그림 1. 산업 제어 시스템 프로토콜 메시지 타입 분류 방법

메시지 타입 클러스터링 단계까지 최대한 상세하게 메시지 타입이 분류되고, 고정 필드 추출 단계에서는 각 분류된 타입내의 메시지들끼리 CSP(Contiguous Sequence Pattern) 알고리즘을 통해 고정적으로 발생하는 필드를 추출한다[3]. 각 메시지타입에서 고정 필드를 추출하면, 추출된 필드를 이용하여 메시지 타입 합병단계에서 같은 타입이지만, 사이즈가 다른 메시지들을 하나의 타입으로 병합한다. 이전 단계까지는 사이즈가 다르면 다른 타입일 수밖에 없지만, 본 단계를 통해 사이즈가 다르더라도 추출된 고정필드가 타입을 병합하는 조건에 충족한다면 같은 타입으로 병합할 수 있다.

다음 그림 2 는 고정필드에 따라 메시지 타입을 합병할 수 있는 조건이다. 첫번째 조건은 각 타입에서 추출된 고정필드가 완전 동일할 때이다. 완전 동일한 고정필드가 추출되었다면, 가변필드로 인해 다른 타입으로 분류되었기 때문이다. 두번째 조건은 추출된 고정 필드가 포함관계일 때이다. 하나의 타입에서 추출된 고정필드가 다른 타입에서 추출된 고정필드와 완전 포함관계에 있다면, 두개의 타입은 하나의 타입으로 병합할 수 있다. 마지막 조건은 각 타입의 고정필드 개수가 다르더라도 포함관계에 있다면 같은 타입으로 병합할 수 있다.

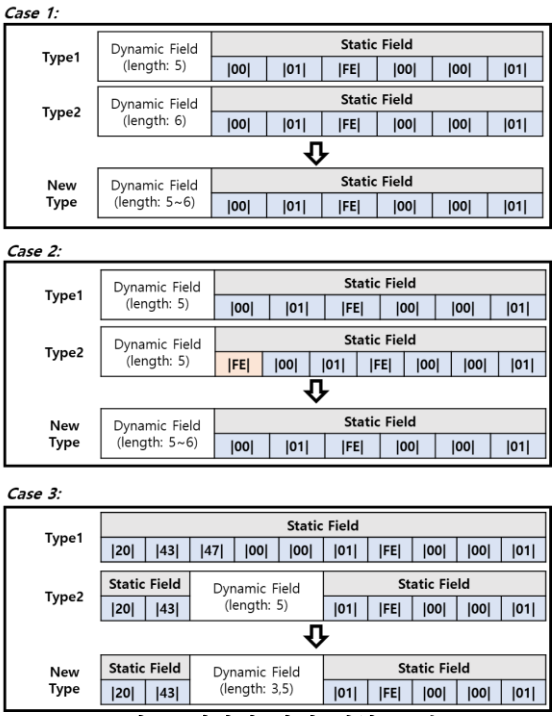

그림 2. 메시지 타입 병합 조건

Ⅲ. 결론

본 논문에서 우리는 산업 제어 시스템 프로토콜 분석을 위한 메시지 타입 정의 방법을 제안하였다. 산업 제어 시스템 프로토콜은 대부분 비공개되어있기 때문에 네트워크 트래픽 분석을 위해서는 프로토콜 구조를 추론해야한다. 따라서 본 논문에서 제안한 산업 제어 시스템 프로토콜 분석을 위한 메시지 타입 정의 방법을 통해 메시지 타입을 분류하고, 분류된 메시지 타입 내에서 필드의 구성을 파악하여 프로토콜의 구조를 분석할 수 있다.

향후연구로는 메시지 타입을 분류하고, 분류된 메시지 타입 내에서 필드를 추출한 상태에서 각 필드의 의미를 파악할 수 있는 Semantics 추론이 이루어져야한다. 예를 들면, 하나의 필드가 메시지 발생 순서에 따라 점차 증가하는 필드이면 해당 필드는 Transaction ID 를 의미하는 필드로 분류할 수 있는 방법을 개발할 예정이다. 또한, 본 시스템을 시스템화 하여 실제 산업 제어 시스템 프로토콜에 적용해야한다.

참 고 문 헌

[1] K. Zetter. Attack code for scada vulnerabilities released online. http://www.wired.com/threatlevel/2011/03/scada-vulnerabilities/, 2011.

[2] G. Bossert, ‘‘Exploiting semantic for the automatic reverse engineering of communication protocols,’’ Ph.D. dissertation, Univ. Gif-sur-Yvette, Rennes, France, Dec. 2014.

[3] Kyu-Seok Shim, Young-Hoon Goo, Min-Seob Lee, Huru Hasanova and Myung-Sup Kim, "Inference of Network Unknown Protocol Structure using CSP(Contiguous Sequence Pattern) Algorithm based on Tree Structure," Proc. of the NOMS 2018 - IEEE/IFIP DISSECT workshop, Taipei, Taiwan, April. 23, 2018, pp.1-4.