Formally Verifying Flow Properties in Industrial Systems
Jannik Dreier, Maxime Puys, Marie-Laure Potet, Pascal Lafourcade, Jean-Louis Roch

To cite this version:

HAL Id: hal-01527913
http://hal.univ-grenoble-alpes.fr/hal-01527913
Submitted on 26 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Formally Verifying Flow Integrity Properties in Industrial Systems

Jannik Dreier\(^3\), Maxime Puys\(^1\), Marie-Laure Potet\(^1\), Pascal Lafourcade\(^2\) and Jean-Louis Roch\(^1\)

(1) Verimag, University Grenoble Alpes, Saint-Martin-d’Hères, France
firstname.lastname@univ-grenoble-alpes.fr

(2) LIMOS, University Clermont Auvergne, Campus des Cézeaux, Aubière, France
pascal.lafourcade@uca.fr

(3) LORIA, University of Lorraine, INRIA, CNRS, France
jannik.dreier@loria.fr

Keywords: Security protocols, industrial systems, SCADA, symbolic model, automated verification, flow integrity.

Abstract: In contrast to other IT systems, industrial systems often do not only require classical properties like data confidentiality or authentication of the communication, but have special needs due to their interaction with the physical world. For example, the reordering or deletion of some commands sent to a machine can cause the system to enter an unsafe state with potentially catastrophic effects. To prevent such attacks, the integrity of the message flow is necessary. We provide a formal definition of Flow Integrity. We apply our framework to two well-known industrial protocols: OPC-UA and MODBUS. Using TAMARIN, a cryptographic protocol verification tool, we confirm that most of the secure modes of these protocols ensure Flow Integrity given a resilient network. However, we also identify a weakness in a supposedly secure version of MODBUS.

1 Introduction

Industrial systems are often used to monitor and control a physical process such as energy production and distribution, water cleaning or transport systems. They are often simply called Supervisory Control And Data Acquisition (SCADA) systems. Due to their interaction with the real world, the safety of these systems is critical and any incident can potentially harm humans and the environment. Since the Stuxnet worm in 2010 (Langner, 2011), such systems increasingly face cyberattacks caused by various intruders, including terrorists or enemy governments. As the frequency of such attacks is increasing, the security of SCADA systems becomes a priority for governmental agencies, e.g. (Stouffer et al., 2011) for the NIST or (ANSSI, 2012) for the ANSSI.

While security objectives for IT systems are usually confidentiality, integrity and availability (CIA), industrial systems put a particular emphasis on integrity and availability. One property required by such systems is that all sent commands are received in the same order by the industrial machine, which is part of what we call Flow Integrity. This property is crucial in industrial systems since most of commands require the system to be in a specific state when they are launched. For instance, if an electric device requires to be unplugged to be manipulated, the shutdown command must arrive before any manipulation command. Inverting them could cause the device, along with its environment, to be damaged.

Automated protocol verification has been performed during the past twenty years and multiple efficient tools such as ProVerif (Blanchet, 2001), AVISPA (Armando et al., 2005), Scyther (Cremers, 2008) or TAMARIN (Meier et al., 2013) have been developed. However, they focused on cryptographic protocols for Internet such as TLS (Cremers et al., 2016) or special applications such as electronic voting (Kremer and Ryan, 2005) or auctions (Dreier et al., 2013).

The Flow Integrity property differs from the properties usually verified in these classical protocols. For example, we want to ensure that messages are delivered (a liveness property), which requires a resilient channel. As Internet is not resilient, resilient channels are difficult to model in most tools that were designed to verify Internet protocols. Moreover, the order of messages is ensured in most Internet protocols as the messages...
have different formats, so reordering the messages simply aborts the protocol. In the context of industrial systems most of the protocols are used to transport commands, meaning that the messages always have the same format, rendering the ordering crucial. In order to ensure the correct ordering of the messages, most of the transport protocols including industrial ones use timestamps, counters and sequence numbers. These solutions are notoriously difficult to model and verify using actual tools due to some theoretical limitations of the tools that often lead to non-termination. In order to face these limitations, we use the verification tool TAMARIN (Meier et al., 2013), that allows us to model counters and resilient channels that can build on previous work concerning the verification of liveness properties (Backes et al., 2017).

**Contributions:** To the best of our knowledge, the Flow Integrity property has not yet been formalized in the context of industrial protocols. Hence, we have two main contributions:

- We provide a formal definition of Flow Integrity in industrial control systems; a property that ensures that all messages are received without alteration, and in the same order as they were sent. We also define weaker properties, including Non-injective and Injective Message Authenticity, which ignore the ordering of messages but ensure that all received messages are unmodified (and cannot be duplicated in the injective case). We also define the corresponding Non-injective and Injective Message Delivery properties, making sure that all messages are delivered (and in the injective case the correct number of times).
- We study Flow Integrity for two real industrial protocols: MODBUS and OPC-UA. Using TAMARIN, we apply our framework to multiple versions of these protocols and discover a weakness in a version of MODBUS.

**Outline:** In Section 2, we discuss related work. Then in Section 3, we explain our definitions of the different properties, and in Section 4 how we modeled these properties with the TAMARIN prover. In Section 5, we apply the verification of our property to the MODBUS and OPC-UA industrial protocols. Finally, we conclude in Section 6.

## 2 Related Work

The notion of integrity can vary a lot depending on the context. A generic definition could state that integrity is the maintenance and assurance of the accuracy and consistency of some data over its life-cycle. For instance, this notion has been applied in 1987, by Clark and Wilson in (Clark and Wilson, 1987). They proposed an access control model able to specify and analyze integrity policies. In such model, data alteration is restricted to those authorized. In 1998 in a different field, Heintze et al. (Heintze and Riecke, 1998) analyzed the consistency of the values of variables during a program execution. Within their framework, they are able to ensure properties relying on integrity such as non-interference (i.e. the modification of a variable should not affect another). Again in a different field, in 2005, Umezawa et al. (Umezawa and Shimizu, 2005) proposed a methodology to ensure that the description of hardware components (such as VHDL code) respects some temporal logic properties such as invalid states for state machines or invalid values for counters. Their approach relies both on model-checking and simulation.

In this paper, we studied the integrity of messages exchanged over a potentially insecure network. Traditionally, message integrity is used to detect accidental changes using error detection codes such as Cyclic Redundancy Checks (CRC). However, such detection codes do not protect against a malicious intruder since he can easily recalculate CRCs of the messages he changes. Similarly the TCP protocol protects against an accidental reordering of messages, but not against a malicious intruder that also modifies the sequence numbers used for this purpose. To guarantee message integrity in presence of malicious intruders, cryptographic primitives are needed, such as digital signatures or Message Authentication Codes (MAC).

Early works concerning the security of industrial protocols focused on discussing the security properties supported or not by protocols. In 2004, Clarke et al. (Clarke et al., 2004) studied the security of DNP3 (Distributed Network Protocol) and ICCP (Inter-Control Center Communications Protocol). In 2005, Dzung et al. (Dzung et al., 2005) surveyed the security in SCADA systems including informal analysis on the security properties offered by various industrial protocols: OPC (Open Platform Communications), MMS (Manufacturing Message Specification), IEC 61850, ICCP and EtherNet/IP. In 2006, authors of the technical documentation of OPC-UA (OPC Unified Architecture) detailed the security measures of the protocol. In 2015, Wanying et al. (Wanying et al., 2015) summarized the security offered by MODBUS, DNP3 and OPC-UA. None of these works give any formal proof of security properties on the protocols.

In more recent works, formal analyses started to appear for industrial protocols. In (Patel and Yu, 2007) the authors proposed a formal verification of DNP3 using OFMC (Basin et al., 2003, Open-Source Fixed-Point Model-Checker) and SPEAR II (Saul and
Hutchinson, 1999, Security Protocol Engineering and Analysis Resource). In (Dutertre, 2007), they detailed formal specifications of MODBUS developed using PVS, a generic theorem prover in order to help proving the consistency of an implementation with the standards. In (Fovino et al., 2009), the authors proposed a secure version of MODBUS relying on well-known cryptographic primitives such as RSA and SHA2. In (Hayes and El-Khatib, 2013), they designed another secure version of the MODBUS protocol using hash-based message authentication codes and built on SCTP (Stream Transmission Control Protocol). In (Bratus et al., 2016), authors provided a Deep-Packet Inspection tool to verify syntactic correctness of DNP3 packets using the Hammer tool (Paterson and Hirsch, 2014). In (Puys et al., 2016), the authors formally verified secrecy and authentication properties of OPC-UA handshake protocols using the ProVerif tool (Blanchet, 2001). However, none of these works formally define or verify Flow Integrity.

In general – outside industrial systems – formal verification of authentication properties (Lowe, 1997) is common. As shown by (Lafourcade and Barthe, 2007), they de-
Definition 1. Let $S_{A,B,D}$ be the sequence that contains the subterms of type $D$ of all messages sent by agent $A$ to agent $B$, and the sequence $R_{A,B,D}$ contains the subterms of type $D$ of all messages received by agent $B$ from $A$.

For example, given a protocol that sends the message $m$ of type $D$ together with its hash $h(m)$ of type $H$, $S_{A,B,D}$ only contains the messages, and not the hashes. Since $A$ might not only send messages to $B$ but also to another agent $C$, and $B$ might receive messages from $A$ and $E$, we define the ordered sequence of messages that $A$ sends to $B$, and the ordered sequence of messages that $B$ received from $A$.

Note that we understand the notions of origin and destination from the agents perspective, i.e., a message $m$ is in $R_{A,B,D}$ if $B$ believes that it came from $A$. Similarly, a message $m$ that $A$ wanted to send to $B$, but was received by $C$, is still in $S_{A,B,D}$.

We now define several notions of integrity, authenticity, and delivery. We have three levels of integrity, authenticity and delivery, where at each level integrity is defined as the conjunction of the corresponding authenticity and delivery properties. Intuitively, the authenticity properties ensure that message have not be altered during transmission, and the delivery properties ensure that messages are not lost.

The first notion of authenticity requires that all received messages were sent by the sender, but messages can be lost or duplicated.

Property 1. A protocol ensures Non-Injective Message Authenticity (NIMA) between sender $A$ and receiver $B$ for data $D$ if $\text{set}(R_{A,B,D}) \subseteq \text{set}(S_{A,B,D})$.

It does only ensure that messages are unmodified, but not that they are actually delivered. For this, we define the corresponding delivery property.

Property 2. A protocol ensures Non-Injective Message Delivery (NIMD) between sender $A$ and receiver $B$ for data $D$ if $\text{set}(R_{A,B,D}) \supseteq \text{set}(S_{A,B,D})$.

Note that message delivery is difficult to achieve using an insecure asynchronous network such as Internet, but industrial systems often use special (real-time) networks with stronger channel guarantees such as Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR) (IEC-62439, 2016). Taking the above properties together, we obtain Non-Injective Message Integrity.

Property 3. A protocol ensures Non-Injective Message Integrity (NIMI) between sender $A$ and receiver $B$ for data $D$ if $\text{set}(R_{A,B,D}) = \text{set}(S_{A,B,D})$.

To ensure that messages cannot be duplicated, we have Injective Message Authenticity and Injective Message Delivery.

Property 4. A protocol ensures Injective Message Authenticity (IMA) between sender $A$ and receiver $B$ for data $D$ if $\text{multiset}(R_{A,B,D}) \subseteq \text{multiset}(S_{A,B,D})$.

Property 5. A protocol ensures Injective Message Delivery (IMD) between sender $A$ and receiver $B$ for data $D$ if $\text{multiset}(R_{A,B,D}) \supseteq \text{multiset}(S_{A,B,D})$.

Both properties can be verified at the same time by checking Injective Message Integrity.

Property 6. A protocol ensures Injective Message Integrity (IMI) between sender $A$ and receiver $B$ for data $D$ if $\text{multiset}(R_{A,B,D}) = \text{multiset}(S_{A,B,D})$.

Again it is easy to see that a protocol ensuring Injective Message Integrity also ensures Injective Message Delivery and Injective Message Authenticity, and that vice versa a protocol ensuring Injective Message Delivery and Injective Message Authenticity also ensures Injective Message Integrity.

Injective Message Integrity ensures that all messages are delivered, and not duplicated, but they can still be reordered. This is prevented by Flow Authenticity and Flow Delivery.

Property 7. A protocol ensures Flow Authenticity (FA) between sender $A$ and receiver $B$ for data $D$ if $R_{A,B,D}$ is a subchain of $S_{A,B,D}$.

Property 8. A protocol ensures Flow Delivery (FD) between sender $A$ and receiver $B$ for data $D$ if $S_{A,B,D}$ is a subchain of $R_{A,B,D}$.

Both properties can be verified at the same time by checking Flow Integrity, which corresponds to the property one would like to achieve in real systems.

Property 9. A protocol ensures Flow Integrity (FI) between sender $A$ and receiver $B$ for data $D$ if $S_{A,B,D} = R_{A,B,D}$.

Again it is easy to see that a protocol ensuring Flow Integrity also ensures Flow Delivery and Flow Authenticity, and that vice versa a protocol ensuring Flow Delivery and Flow Authenticity also ensures Flow Integrity.

Note that a protocol ensuring Flow Integrity also ensures Injective Message Integrity, and that a protocol ensuring Injective Message Integrity also ensures Non-Injective Message Integrity (and analogously for the authenticity and delivery properties). This is summed up in Figure 1.

Moreover, if a protocol ensures either Flow Authenticity and Injective Message Delivery, or Flow Delivery and Injective Message Authenticity, this is sufficient to ensure Flow Integrity, as the following Theorem 1 shows.

Theorem 1. A protocol that ensures Flow Delivery and Injective Message Authenticity also ensures Flow...
Integrity ($FD \land IMA \Rightarrow FI$). Similarly, a protocol that ensures Flow Authenticity and Injective Message Delivery, also ensures Flow Integrity ($FA \land IMD \Rightarrow FI$).

Proof. Let $[s_1,\ldots,s_n] = S_{A,B,D}$ and $[r_1,\ldots,r_m] = R_{A,B,D}$. Suppose that a protocol ensures Flow Delivery and Injective Message Authenticity, i.e. we have that $S_{A,B,D}$ is a subchain of $R_{A,B,D}$, and $\text{multiset}(R_{A,B,D}) \subseteq \text{multiset}(S_{A,B,D})$. Moreover, as any protocol ensuring Flow Delivery also ensures Injective Message Delivery, we have $\text{multiset}(R_{A,B,D}) \supseteq \text{multiset}(S_{A,B,D})$. Thus, $\text{multiset}(R_{A,B,D}) = \text{multiset}(S_{A,B,D})$.

This means that $n = m$, i.e. both sequences have the same length. By the definition of subchains we have that there exist sequences $z_0,\ldots,z_n$ such that $z_0 \cdot [s_1] \cdot z_1 \cdot \ldots \cdot z_{n-1} \cdot [s_n] \cdot z_n = [r_1 \ldots r_m]$. As $n = m$, we have that $[s_1,\ldots,s_n] = [r_1,\ldots,r_n]$, which is what we wanted to show. The second proof is similar.

4 The TAMARIN prover

We now recall the syntax and semantics of labeled multiset rewriting rules, which constitute the input language of the TAMARIN prover (Schmidt et al., 2012).

4.1 Introducing the TAMARIN prover

In TAMARIN, equations are used to specify properties of functions, where an equation over the signature $\Sigma_{Fun}$ is an unordered pair of terms $s,t \in T_{Fun}(V)$, written $s \simeq t$. An equational presentation is a pair $E = (\Sigma_{Fun},E)$ of a signature $\Sigma_{Fun}$ and a set of equations $E$. The corresponding equational theory $=_{E}$ is the smallest $\Sigma_{Fun}$-congruence containing all instances of the equations $E$. We often leave the signature $\Sigma_{Fun}$ implicit and identify the equations $E$ with the equational presentation $E$. Similarly, we use $=_{E}$ for the equational theory $=_{E}$. We say that two terms $s$ and $t$ are equal modulo $E$ iff $s =_{E} t$. We use the subscript $E$ to denote the usual operations on sets, sequences, and multisets where equality is modulo $E$ instead of syntactic equality. For example, we write $\in E$ for set membership modulo $E$.

Example 1. To model MACs, let $\Sigma_{Fun}$ be the signature consisting of the functions $\text{mac}(\cdot,\cdot)$ and $\text{verify}(\cdot,\cdot)$ together with the equation

$$\text{verify}(\text{mac}(x,k),x,k) \simeq true.$$
for the rule instance. As linear facts are consumed upon rewriting, we use multiset inclusion, written \( \subseteq \), to check that all facts in \( \text{facts}(l) \) occur sufficiently often in \( S \). For persistent facts, we only check that each fact in \( \text{pfacts}(l) \) occurs in \( S \). To obtain the successor state, we remove the consumed linear facts and add the generated facts. The actions associated to the transition contain the set of actions of the rule instance, the identifier of the rule, and the newly introduced variables.

**Example 2.** The following multiset rewrite rules describe a simple protocol that sends messages together with a hash of the message. The first rule describes the agent \( A \): he uses the key shared with \( B \) to send a fresh message \( m \) to \( B \). The second rule describes \( B \): he receives a message together with its hash. Note that the second rule can only be triggered if the input matches the premise, i.e., if the hash is correctly computed.

\[
\begin{align*}
\text{Send Message}_A &: \quad [\text{Fr}(m)] \rightarrow [\text{Sent}(m) \rightarrow \text{Out}((m,h(m)))], \\
\text{Receive Message}_B &: \quad [\text{In}((m,h(m)))] \rightarrow [\text{Received}(m)]
\end{align*}
\]

**Example 3.** Consider the multiset rewrite rules given in Example 2. The following property specifies that any message received by \( B \) was previously sent by \( A \):

\[
\forall i : \text{time}, m : \text{msg.} \quad \text{Received}(m) @ i \Rightarrow (\exists j. \text{Sent}(m) @ j \land j < i)
\]

For the formal definition of the semantics, see (Schmidt et al., 2012).

### 4.2 Defining our Security Properties

Using trace formulas we can specify all our properties in TAMARIN as follows. To make messages visible on the trace, we instrument the protocol rules in TAMARIN with two actions, \( \text{Sent}(A,B,m) \) and \( \text{Received}(A,B,m) \), where the first one denotes that the message \( m \) was sent by \( A \) to \( B \), and \( \text{Received}(A,B,m) \) denotes that \( B \) received message \( m \) from \( A \). Note that here we only use the message payload, i.e. \( m \) is the part of the protocol message that is of type \( D \). Using these actions, we can define Non-Injective Message Authenticity in TAMARIN as follows.

**Property 10.** A TAMARIN protocol model ensures Non-Injective Message Authenticity (NIMA) between sender \( A \) and receiver \( B \) for data \( D \) if the following formula is satisfied on all traces:

\[
\forall i : \text{time}, A, B, m : \text{msg.} \quad \text{Received}(A,B,m) @ i \Rightarrow (\exists j. \text{Sent}(A,B,m) @ j \land j < i)
\]
This definition captures precisely the definition from Section 3: we require that any message \( m \) received by \( B \) from \( A \), i.e. \( m \in \text{set}(R_{A,B,D}) \), is included in \( \text{set}(S_{A,B,D}) \), i.e. was sent by \( A \) to \( B \). We can define Non-Injective Message Delivery analogously by interchanging the Sent and Received actions.

**Property 11.** A TAMARIN protocol model ensures Non-Injective Message Delivery (NIMD) between sender \( A \) and receiver \( B \) for data \( D \) if the following formula is satisfied on all traces:

\[
\forall i : \text{time}, A, B, m : \text{msg.Sent}(A, B, m)@i \\
\Rightarrow (\exists j.\text{Received}(A, B, m)@j \land i < j)
\]

To verify Non-Injective Message Integrity we can simply check whether both Non-Injective Message Authenticity and Non-Injective Message Delivery hold.

To ensure injectivity, we have to ensure that a message cannot be duplicated, which we express as follows.

**Property 12.** A TAMARIN protocol model ensures Injective Message Authenticity (IMA) between sender \( A \) and receiver \( B \) for data \( D \) if the following formula is satisfied on all traces:

\[
\forall i : \text{time}, A, B, m : \text{msg.Received}(A, B, m)@i \\
\Rightarrow (\exists j.\text{Sent}(A, B, m)@j \land j < i \land (\exists i_2 : \text{time}, A_2, B_2 : \text{msg.Sent}(A_2, B_2, m)@i_2 \land i_2 = i))
\]

This ensures that any received message was previously sent, and that there is not other time point where the same message is received, thus capturing the injectivity requirement.\(^4\) The corresponding delivery property definition is obtained easily by interchanging the Sent and Received actions, as above.

**Property 13.** A TAMARIN protocol model ensures Injective Message Delivery (IMD) between sender \( A \) and receiver \( B \) for data \( D \) if the following formula is satisfied on all traces:

\[
\forall i : \text{time}, A, B, m : \text{msg.Sent}(A, B, m)@i \\
\Rightarrow (\exists j.\text{Received}(A, B, m)@j \land i < j \land (\exists i_2 : \text{time}, A_2, B_2 : \text{msg.Received}(A_2, B_2, m)@i_2 \land i_2 = i))
\]

To verify Injective Message Integrity, we simply check both properties at the same time.

Flow Authenticity and Flow Delivery are expressed in TAMARIN as follows: we first verify that Injective Message Authenticity or Injective Message Delivery hold, respectively, and then check whether the order of messages is preserved.

\^4 We use unique fresh messages on the sender side to prevent false attacks that would result from the same message being sent twice and received twice.

---

**Property 14.** A TAMARIN protocol model ensures Flow Authenticity (FA) between sender \( A \) and receiver \( B \) for data \( D \) if it ensures Injective Message Authenticity and if the following formula is satisfied on all traces:

\[
\forall i, j : \text{time}, A, B, m, m_2 : \text{msg.} \\
(\text{Received}(A, B, m)@i \land \text{Received}(A, B, m_2)@j \land i < j) \\
\Rightarrow (\exists k, l.\text{Sent}(A, B, m)@k \land \text{Sent}(A, B, m_2)@l \land k < l)
\]

**Property 15.** A TAMARIN protocol model ensures Flow Delivery (FD) between sender \( A \) and receiver \( B \) for data \( D \) if it ensures Injective Message Delivery and if the following formula is satisfied on all traces:

\[
\forall i, j : \text{time}, A, B, m, m_2 : \text{msg.} \\
(\text{Sent}(A, B, m)@i \land \text{Sent}(A, B, m_2)@j \land i < j) \\
\Rightarrow (\exists k, l.\text{Received}(A, B, m)@k \land \text{Received}(A, B, m_2)@l \land k < l)
\]

Again, to verify Flow Integrity, we simply check both properties at the same time.

### 4.3 Resilient Channels, Counters and Timestamps

As noted above, delivery properties typically require a resilient channel as an unrestricted Dolev-Yao intruder can simply delete all messages and thus prevent any message from arriving. We can model a resilient channel in TAMARIN by adding a restriction that enforces that all messages are eventually delivered. A restriction is a trace formula that TAMARIN will assume true, i.e., it will discard all traces violating the restriction when trying to prove a property.

To model a resilient channel, we add a new action \( \text{Ch\_Sent}(m) \) to all rules that send out messages on the resilient channel, and an action \( \text{Ch\_Received}(m) \) to all rules that receive messages from the resilient channel. Note that here \( m \) is not only the payload of type \( D \), but the entire protocol message, and that we do not include senders or recipients. Using these actions, we can express the fact that the channel is resilient using the following formula:

\[
\forall i : \text{time}, m : \text{msg.}\text{Ch\_Sent}(m)@i \\
\Rightarrow (\exists j.\text{Ch\_Received}(m)@j \land i < j)
\]

Note that this restriction on the intruder’s capabilities does not prevent him from delaying messages for a certain time, reordering or duplicating them, or injecting new messages. This means that even when assuming a resilient channel our security properties do not hold vacuously.

We also use restrictions to model sequence numbers and timestamps. An intuitive way of modeling
sequence number in TAMARIN would be to use state facts to implement a counter using a constant (e.g. zero) and a function (e.g. inc(·)). Consider a protocol that simply sends out a message together with its counter:

\[ \text{Counter}_\text{Init} : [] \rightarrow \text{Counter}(\text{zero}) \],
\[ \text{Send}\_\text{Message} : [\text{Counter}(n), \text{Fr}(m)] \rightarrow \rightarrow \text{Counter}(\text{inc}(n)), \text{Out}((m, n)) \]

Such a model usually results in non-termination. When TAMARIN tries to prove a property, it tries to find a counterexample using a backwards-search approach. More precisely, it starts from the negation of the formula, and tries to construct a valid execution by resolving the premises of all rule instances mentioned in the formula until it either has found a counterexample or a contradiction.

When analyzing the above counter model, resolving the first premise of the Send\_\text{Message} rule results in two cases: either the premise is the conclusion of a Counter\_\text{Init} rule, or it results from a Send\_\text{Message} rule itself. In that case we need to resolve the same premise again, and enter a loop.

Our solution to avoid this problem is to not model the counter explicitly, but to let the intruder choose the sequence number, while limiting his choice using a restriction. Consider the rule

\[ \text{Send}\_\text{Message} :
\[ [\text{Ln}(n), \text{Fr}(m)] \rightarrow \rightarrow \text{Seq}\_\text{Sent}(A, B, n) \rightarrow \rightarrow \text{Out}((m, n)) \]

and the following restriction

\[ \forall i, j : \text{time}, A, B, \text{seq}_1, \text{seq}_2 : \text{msg}.
\[ (\text{Seq}\_\text{Sent}(A, B, \text{seq}_1)) @ i \land (\text{Seq}\_\text{Sent}(A, B, \text{seq}_2)) @ j
\[ \land i < j \rightarrow (\exists \text{dif}, \text{seq}_2 \approx \text{seq}_1 + \text{dif}) \]

where “+” is an associative and commutative infix operator provided by TAMARIN. Note that “+” does not have any other associated equations and thus does not exactly correspond to an addition of numbers. In particular we do not have a neutral element 0, so that seq\(_1 + 0 \neq \text{seq}_1\). The restriction ensures that the term representing the sequence number in any two subsequent messages increases by including a new term dif, but without fixing dif precisely. Although this abstraction allows jumps (for example increments by 2 or more) in the sequence number which would not occur in reality, it fixes an order on the sequence numbers which is sufficient to prove the properties we are interested in, as we will see in the case studies. Finally timestamps can be modeled in the same way, which means that the intruder controls the timing, but cannot go back in time.

5 Applications to SCADA Protocols

We verify the security of multiple variants of two industrial communication protocols (namely MODBUS and OPC-UA) to check if they guarantee the properties we defined in Section 3. The TAMARIN code is available online\(^5\), all verifications where completed on a standard laptop within a few minutes. As mentioned earlier, all protocols presented in this Section are transport protocols which carry a request from a client to a server (the responses from the server to the client can be considered as another instance of the same protocol) over a potentially asynchronous and insecure network. We consider an unbounded number of sessions of the protocol, where each session is an arbitrary long sequence of messages.

5.1 MODBUS

Description: MODBUS (MODBUS, 2004) is an industrial communication protocol designed by Modicon (now Schneider Electric) in 1979. It has become one of the most popular protocols in the domain and can be used either on serial bus or on TCP communication. We focus on the TCP version of the protocol, which is nowadays more popular than the serial version. In all MODBUS protocols, only the client is able to send requests to which the server answers (meaning that the server does never send a message on its own). In the TCP version, the message includes a sequence number in addition to the TCP sequence number. This number is called a transaction identifier and only increased by one at each client request. Some other terms are also part of the message, i.e. (i) a protocol identifier only used for compatibility with non-TCP versions, (ii) the length of the message and (iii) the unit identifier which is used to dispatch the command to actuators and sensors. Those three terms are public values that do not impact the security of the protocol. We choose to model them as single public header \( ph \). A generic session of the protocol is displayed in Figure 2, where \( n \) is the transaction identifier, \( req \) is a request from the client and \( resp \) is the corresponding response from the server.

The protocol relies on TCP to provide countermeasures against network errors (e.g. checksums such as CRC or LRC), and does not implement any protection against malicious adversaries. Thus anyone is able to forge a fake message or modify an existing one, allowing an adversary to run arbitrary commands on servers. To avoid such attacks, two secure versions were proposed. In (Fovino et al., 2009), the authors

\(^5\)http://indusprotoverif.forge.imag.fr/DPFLR17.tar.gz
proposed a version of MODBUS based on well-known cryptographic primitives such as RSA and SHA2. Figure 3 presents the same session than in Figure 2 plus the counter-measures proposed in (Fovino et al., 2009) where ts_i is the timestamp of the i-th message.

In (Hayes and El-Khatib, 2013), they designed another secure MODBUS protocol based on SCTP (Stream Control Transmission Protocol). SCTP is a transmission layer protocol as TCP and UDP which provides protection against Denial-of-Service attacks. However, like TCP it provides counter-measures against network errors but none against malicious intruders. To avoid an adversary forging fake messages or modifying existing ones, Hayes et al. added message authentication codes (MACs). Moreover, to avoid replay attacks a nonce (called verification tag) provided by SCTP is included in the MACs of the messages. Figure 4 details the session in Figure 2 plus the counter-measures proposed in (Hayes and El-Khatib, 2013) with vt the verification tag of the SCTP session.

Security Analysis: We modeled the three versions of MODBUS described above and analyzed them with TAMARIN to check if they satisfy the properties we defined in Section 3. We performed a first analysis assuming an insecure network, and the results are presented in Table 1.

TAMARIN finds attacks for all properties against the standard version. This is not surprising since this version of the protocol was not intended to provide any security. However, the version with digital public key signatures from (Fovino et al., 2009) is subject to attacks since the identity of the receiver of the message is not specified in the signature. Thus an intruder is able to reroute a message to different recipient which accepts the message, violating all of our properties. This attack could be prevented by adding the receiver inside the signature, or using a different public and private key pair for each connection, which however would be equivalent to using a symmetric authentication technique such as MACs, which is done in version with MACs from (Hayes and El-Khatib, 2013). In this version the attack is prevented since the symmetric authentication keys are restricted to a specific session between a specific client and a specific server. Thus if an intruder changed the destination of a message, the new recipient would not be able to verify the MAC. This version of the protocol ensures all authenticity properties, however it still fails on all delivery properties as the intruder can simply delete all message. When assuming a resilient channel, it also ensures all delivery properties (see Table 2). Note that even when assuming a resilient channel the first two variants do still not ensure any property as messages are still not guaranteed to be delivered at the right recipient, as in the above attack.

5.2 OPC-UA

Description: OPC-UA is one of the most recent industrial communication protocols, being released in 2006 (OPC Unified Architecture, 2012). It is developed by the OPC Foundation\(^6\), and is often referred to as the next industrial communication standard. It is a multi-level protocol, including transport and session layers. The security layer implements key agreement through a handshake. Then the client is invited to provide an authentication method such as a password or a certificate using the generated key. The transport layer consists in sending messages from the client to the server using the security keys negotiated. A generic session of the protocol is displayed in Figure 5 where:

- mh is a message header containing public values.
- sh is a security header consisting of a fresh nonce called security token.

\(^6\)A consortium of the main stakeholders of the domain.
was not intended to not provide any security. However

Table 3.

responding delivery properties, we again need to assume
guarantee Flow Authenticity. To also have the corre-
for all authenticity properties. This means that having
the case of an insecure network are presented in
for the three security
MODES (None, Sign and SignAndEncrypt). Results
of OPC-UA presented in Figure 5 for the three security

Finally, three security modes exist in OPC-UA:

• SignAndEncrypt (Figure 5): messages are signed
\( \text{mac} (m, K_{XY}) \) and encrypted \( \{ m \} _{K_Y} \), where
\( \text{mac}(\cdot, \cdot) \) is a message authentication code function,
\( K_Y \) the symmetric encryption key shared by \( X \) and \( Y \),
\( K_{XY} \) the symmetric signature key shared by \( X \) and \( Y \).
• Sign: it is the same as SignAndEncrypt but
messages are only signed using \( \text{mac}(m, K_{SigCS}) \),
and not encrypted. Thus message 1 (respectively 2, 3, and 4) of Figure 5 becomes:
\( mh, sh, n, rID_1, \text{req}_1, \text{pad}, \text{mac}((mh, sh, n, rID_1, \text{req}_1, \text{pad}), K_{SigCS}) \)
• None: messages are neither signed nor encrypted
(mainly used for compatibility). Thus message 1 (respectively 2, 3, and 4) of Figure 5 becomes:
\( mh, sh, n, rID_1, \text{req}_1, \text{pad} \)

Security Analysis: We model the transport layer of OPC-UA presented in Figure 5 for the three security modes (None, Sign and SignAndEncrypt). Results for the case of an insecure network are presented in Table 3.

TAMARIN finds attacks on the version with security mode None. This is not surprising since this version was not intended to not provide any security. However both the Sign and SignAndEncrypt versions are safe for all authenticity properties. This means that having only the MACs added in the Sign version is enough to guarantee Flow Authenticity. To also have the corresponding delivery properties, we again need to assume a resilient channel (see Table 4).

Out of curiosity, we also checked a variant of the protocol with only symmetric encryption and no MAC (thus not an official version). It appears that we obtain the same results as for signatures. This is due to the fact that the symmetric keys are only shared by two participants and any message with its destination changed would not be readable by its new recipient.

OPC-UA in case of bounded sequence numbers.

Until now we assumed sequence numbers to be unbounded integers from \( N \). However, in reality machine integers are obviously bounded and this can have an impact on properties such as Flow Integrity. To evaluate this impact, we tested a modeling of OPC-UA SignAndEncrypt (described in Figure 5) with explicitly bounded sequence numbers (in our example we bound it to four). This means that if a client sends four messages, then the fourth message has the same sequence number as the first one.

We checked the properties described in Section 3 on this version with TAMARIN and obtained the results presented in Table 5: it turns out that Flow Integrity is no longer verified. The attack works as follows: the client sends out four messages, thus the fourth message has the same sequence number as the first one. The intruder delays the first three messages so that the first message received by the server is the forth with sequence number zero. He then transmits the second message which has a sequence number of one, and is thus accepted by the server although it was actually sent earlier than the message he accepted previously.

Interestingly the described attack disappears if we assume a resilient channel. The server will accept each sequence number only once\(^7\), and if we have two messages with the same sequence number this leads to a contradiction since both of them have to be received. This however implies that each sequence number can be used only once also on the client side, thus there can be only a finite number of messages, bounded by the range of the sequence numbers.

\(^7\)Note that allowing a sequence number to be reused leads to attacks on Injective Message Authenticity as the same message can be accepted multiple times.
This analysis illustrates the need for a big range of sequence numbers. If more messages than the range of sequence numbers allows need to be exchanged, one has to reinitialize the session (i.e., exchange new keys) before running out of sequence numbers. This is the solution adopted by OPC-UA: in (OPC Unified Architecture, 2012, p. 36) it is stated that “A SequenceNumber may not be reused for any TokenId. The SecurityToken lifetime should be short enough to ensure that this never happens [...]”. Our analysis underlines the importance of this requirement.

6 Conclusion

We provided a formal definition of Flow Integrity and other related properties in industrial systems. Flow Integrity ensures that all messages are received without alteration, and in the same order as they were sent. We checked Flow Integrity on multiple variants of two real industrial protocols: MODBUS and OPC-UA. Our analysis confirms that most of the secure modes of these protocol ensure Flow Integrity given a resilient network. However, we also identified a weakness in a supposedly secure version of MODBUS, due to an insufficient use of cryptography. Moreover, our analysis of bounded sequence numbers highlighted the importance of the renewal of session keys to avoid the reuse of sequence numbers. Unsurprisingly, the insecure modes of these protocols did not ensure any of our security properties. Moreover it turns out that to ensure delivery one has to assume a resilient channel, as the intruder can otherwise always block messages. At the same time, our results show that a resilient channel alone is not sufficient to ensure Flow Integrity: one still needs to use cryptography to prevent the intruder from rerouting or injecting messages.

In the future, we aim at testing this property on real implementations to see if we can reproduce the theoretical results. We would also like to study other industrial protocols such as DNP3 or IEC 61850. Finally we are interested in formalizing properties similar to Flow Integrity for protocols with encapsulation. Such protocols permit for example to transfer MODBUS packets through an OPC-UA channel. They are a real challenge for formal verification as there is few work on protocol composition, and it has turned out that verifying the composition is more complicated than verify the protocols independently.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>NIMA</th>
<th>IMA</th>
<th>FA</th>
<th>NIMD</th>
<th>IMD</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPC-UA None</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
</tr>
<tr>
<td>OPC-UA Sign</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
</tr>
<tr>
<td>OPC-UA SignAndEncrypt</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
</tr>
</tbody>
</table>

Table 3: Results for OPC-UA (OPC Unified Architecture, 2012), assuming an insecure network.

REFERENCES

ANSSI (2012). Managing cybersecurity for ICS.
<table>
<thead>
<tr>
<th>Protocol</th>
<th>NIMA</th>
<th>IMA</th>
<th>FA</th>
<th>NIMD</th>
<th>IMD</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPC-UA None</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
</tr>
<tr>
<td>OPC-UA Sign</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
</tr>
<tr>
<td>OPC-UA SignAndEncrypt</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
</tr>
</tbody>
</table>

Table 4: Results for OPC-UA (OPC Unified Architecture, 2012), assuming a resilient channel.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>NIMA</th>
<th>IMA</th>
<th>FA</th>
<th>NIMD</th>
<th>IMD</th>
<th>FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPC-UA SignAndEncrypt with bounded numbers Insecure Channel</td>
<td>SAFE</td>
<td>SAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
<td>UNSAFE</td>
</tr>
<tr>
<td>OPC-UA SignAndEncrypt with bounded numbers Secure Channel</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
<td>SAFE</td>
</tr>
</tbody>
</table>

Table 5: Results for OPC-UA with bounded counters.