Second order sliding mode observer for anomaly detection in TCP networks: from theory to practice.

Sandy Rahmé, Yann Labit, Frédéric Gouaisbaut and Thierry Floquet

Abstract—We investigate the design of a second order sliding mode observer for detecting anomalies in TCP networks. More precisely, based on a fluid model of TCP network, we aim at designing an observer, whose state variables reconstruct some external signals which perturb the nominal behavior of TCP. The proposed methodology is validated via NS-2 simulations. Finally, an experimental setup is proposed to show the effectiveness of our approach.

I. INTRODUCTION

Transmission Control Protocol (TCP) is a protocol that handles reliable data transfer through networks. Many events in TCP network deviate from the expected behavior such as network overload, flash crowds, worms, port scans, risky internal user behavior, denial of service attacks (DoS), network intrusions, etc. These abnormal patterns are in general called network anomalies (see for example [1], [2], [3] and references therein). Whether malicious or legitimate, anomalies can have an impact on a router by creating congestion and can reduce significantly the Quality of Service (QoS) of a whole network. That is the reason why, nowadays, anomaly detection is an important issue. In [2], a broad overview is provided spanning multiple research areas and application domains.

Anomaly detection techniques can be categorized into Intrusion Detection Systems (IDS) and Anomaly Detection Systems (ADS) [4], [5]. Intrusion Detection Systems use the misuse signature to identify well-known intrusions, but its main disadvantage is the disability of detecting "zero-day" attacks or unknown attacks. On the other hand, Anomaly Detection Systems are only concerned in activities that lead to unprovable network changes. The main advantage of anomaly detection is that it does not require a prior knowledge of intrusion and can thus detect unknown intrusions.

Following the seminal work of [6], which had introduced feedback theory in network control, we propose to use the control theory and especially the design of observers to cope with anomaly detection. To the best of our knowledge, for observation/detection purpose, this framework has been barely adopted. In [7], a flatness based detection is developed for the non linear fluid model of TCP. In [7], a classical Luenberger observer coupled with an extended model is designed for detecting CBR anomalies.

The present paper aims at developing a novel sliding mode observer. Sliding modes [8] are often used to design robust nonlinear observers or control laws because of their robustness against various kinds of uncertainties such as parameter perturbations, external disturbances and measurement errors. The observer trajectories are constrained to evolve after a finite time on a suitable sliding manifold by the use of a discontinuous output injection signal (the sliding manifold is usually given by the difference between the observer and the system output). Subsequently the sliding motion provides an estimate in finite time of the system states. The unknown inputs can be explicitly reconstructed by analyzing the so-called equivalent information (or output) injection [7], [9].

However, first order sliding modes exhibit undesirable high-frequency oscillations, typically referred to as the chattering phenomenon, caused by fast switching in the discontinuous control signal. One common solution to bypass the chattering problem is to use smoothing techniques based on the principle of equivalent control [10]. In our previous work on anomaly detection in TCP networks ([10] and [11]), a first order sliding mode observer was designed and, because of high frequency oscillations, first and third order low pass filtering of the equivalent output injection were used to reconstruct anomalies in the network. The drawback lies in the intuitive regulation of the filter parameters.

Another relevant method proposed in the literature is based on higher order sliding modes [12]. The principle is to act on the higher order time derivatives of the system deviation from the constraint, instead of influencing the first deviation derivative. In this paper, a higher order sliding mode based scheme is adopted for chattering avoidance while detecting anomalies. Mathematical formal model of the TCP network is used for testing the observer’s performances via the network simulator NS-2 [13], the simulator targeted at networking research. However these presentations of the TCP dynamics are, at present, far from representing all the real traffic characteristics. Therefore, one better means consists of replaying a real TCP traffic in a simulator (NS-2 for this work) so that all traffic characteristics might be taken into account. The TCP replay technique represents a first step towards detecting anomalies in a real traffic based on control theory techniques.

The paper is organized as follows. Section II introduces the TCP network topology which represents the problem statement. In section III, the high order sliding mode observer...
is conceived for anomaly detection and reconstruction in TCP networks. Then the efficiency of the proposed observer is evaluated via NS-2 in section IV. Section V presents the methodology adopted for replaying TCP traces in NS-2 while preserving the real characteristics of the real traffic. Finally, remarks concerning our techniques and some research fields conclude the paper.

II. TCP NETWORK MODEL

A network topology consisting of \( N \) homogeneous sources connected to a destination through a router is considered. Many researches have been concerned with mathematically modeling TCP behavior \([7]\), \([7]\). This paper focuses on a simplified fluid flow model developed in \([7]\) where the average values of the key network variables are described by the following coupled nonlinear differential equations:

\[
\begin{align*}
\dot{W}(t) &= \frac{1}{R(t)} - \frac{W(t)W(t - R(t))}{2R(t - R(t))} p(t - R(t)), \\
\dot{q}(t) &= \frac{1}{R(t)} N - C + d(t), \\
R(t) &= \frac{q(t)}{C} + \tau,
\end{align*}
\]

where \( W \) is the average TCP congestion window size [packets] of \( N \) TCP connections, \( q \) is the queue length of the router buffer [packets] and \( R(t) \), the round trip time (RTT) \([s]\), is a function of the link capacity \( C \) [packets/s] and the propagation delay \( \tau \) \([s]\). The variable \( p \) is the dropping probability of a packet entering the buffer queue. The term \( d(t) \) is introduced to the queue length dynamics to represent additional traffic perturbing the normal TCP network behavior. This latter TCP model \([7]\) is a nonlinear system with a variable time delay. Therefore, conceiving an observer for traffic monitoring and anomaly estimation using the sliding mode approach remains a difficult task. We propose to study the linearized fluid-flow system, represented by \([7]\):

\[
\begin{align*}
\delta \dot{W}(t) &= -\frac{N}{R(t)} (\delta W(t) + \delta W(t - h(t))) \\
&\quad - \frac{1}{R(t)} (\delta q(t) - \delta q(t - h(t))) \\
&\quad - \frac{W(t)}{2R(t)^2} \delta p(t - h(t)), \\
\delta \dot{q}(t) &= \frac{N}{R(t)} \delta W(t) - \frac{1}{R(t)} \delta q(t) + d(t),
\end{align*}
\]

where \( \delta W = W - W_0 \), \( \delta q = q - q_0 \) and \( \delta p = p - p_0 \) are the state variables varying around the equilibrium point obtained from:

\[
\begin{align*}
W_0^2 p_0 &= 2, \\
W_0 &= \frac{R_0 C}{2}, \\
R_0 &= \frac{2}{N} + \tau_p, \\
d_0 &= 0.
\end{align*}
\]

Supposing that \( q(t) \) is continuously measured, thus defined as the system output, the linearized TCP model \([7]\) is given by a time-delay system of the form:

\[
\begin{align*}
\dot{x}(t) &= M x(t) + M_d x(t - h) + D y(t) + D_d y(t - h) + E_d u(t - h), \\
\dot{y}(t) &= G x(t) + H y(t) + d(t),
\end{align*}
\]

where \( x(t) = \delta W(t) \) is the state, \( y(t) = \delta q(t) \) the output, \( u(t) = \delta p(t) \) the input, and

\[
\begin{align*}
M &= M_d = \frac{N}{R_0 C}, \\
D &= \frac{1}{R_0 C^2}, \\
D_d &= \frac{1}{R_0 C^2}, \\
E_d &= \frac{R_0 C^2}{2N^2}, \\
G &= \frac{N}{R_0}, \\
H &= -\frac{1}{R_0}.
\end{align*}
\]

For TCP congestion control, the router queue length is regulated by an Active Queue Management (AQM). By controlling the dropping probability \( p(t) \), AQM informs traffic sources (either implicitly or explicitly) in order to adjust their data sending rate into the network. Many mechanisms such as RED, Stabilized-RED (SRED), BLUE, REM, Adaptive Virtual Queue (AVQ) can be used (see \([7]\) for a complete evaluation). Feedback control methods can also be an alternative to achieve satisfactory control performance such as classical PI, PI, PID \([7]\), static state feedback \([7]\), and references therein.

III. HIGH ORDER SLIDING OBSERVER: TCP TRAFFIC MONITORING

A. Concept of high order sliding modes

Higher order sliding modes generalize the basic sliding mode idea while keeping its main advantages. The sliding surface is defined by the vanishing of a corresponding sliding variable \( s \) and its successive time derivatives up to a certain order, defining the \( r^{th} \) order sliding surface:

\[
S_r = \left\{ x \in \mathbb{R}^n : s = s = \ldots = s^{(r-1)} = 0 \right\}.
\]

A control law leading to such a behavior is called a \( r^{th} \) order ideal sliding mode algorithm with respect to \( s \). Higher order sliding modes, that are characterized by a discontinuous control acting on the \( r^{th} \) \( (r > 1) \) time derivatives of the sliding vector (instead of the first time derivative in classical sliding mode where \( r = 1 \)), can reduce the chattering phenomenon while preserving the robustness properties.

Higher order sliding mode designs for either observation or control, with applications in mechanics, robotics or electric machines, can be found in the literature (see for example \([7]\), \([7]\)). In what follows, higher order sliding mode observers will be designed because of the possibility to obtain finite time estimates of anomalies in TCP networks without introducing a low-pass filter (which is not the case with a first order sliding mode, for which it is necessary to
We focus in this paper on a specific second order sliding mode algorithm called super-twisting algorithm. This algorithm is developed to control systems with relative degree one in order to avoid the chattering phenomena [7]. The continuous control law \( \nu \) consists of two terms:
\[
\begin{align*}
\nu(s) &= \nu_1 - \lambda |s|^{\frac{3}{2}} \text{sign}(s), \\
\dot{\nu}_1 &= -\text{sign}(s),
\end{align*}
\]

where \( \alpha > 0, \lambda > 0 \). The main advantage of the super-twisting algorithm is that it does not need the knowledge of the time derivative of the sliding variable.

**B. Observation scheme for TCP network**

In this section, a second order sliding mode observer is designed, first to estimate the average TCP congestion window size and then, to reconstruct any anomaly arising into the router buffer. For the design, the anomaly \( d(t) \) and its first time derivative are supposed to be bounded by upper bound \( d_{max} \) and \( \dot{d}_{max} \) respectively. Let us consider the following observer structure based on the super twisting algorithm described in the previous section:

\[
\begin{align*}
\dot{x}(t) &= \frac{1}{M} [M \dot{x}(t) + M_a \dot{x}(t - h) + D_y(t) + D_d(t - h) + E_d \nu(t - h)], \\
\dot{y}(t) &= \frac{1}{M} [G \dot{x}(t) + H \nu(t - z) - \lambda |y - y|^{\frac{3}{2}} \text{sign}(y - y), \\
\dot{z} &= -\text{sign}(y - y).
\end{align*}
\]

The observation error is defined by the vector \( E = [e_x, e_y, e_z] \) with

\[
\begin{align*}
e_x(t) &= \dot{x} - x(t), \\
e_y(t) &= \dot{y} - y(t), \\
e_z(t) &= z + G e_x - d(t).
\end{align*}
\]

Then, using (12) and (13), the observation errors dynamics can be written as:

\[
\begin{align*}
\dot{e}_x(t) &= M e_x(t) + M_d e_x(t - h), \\
\dot{e}_y(t) &= G e_x(t) + z - \lambda |e_y|^{\frac{3}{2}} \text{sign}(e_y) - d(t) \\
\dot{e}_z(t) &= G(M e_x(t) + M_d e_x(t - h)) - \dot{d}(t) -\text{sign}(e_y).
\end{align*}
\]

The following theorem proves the asymptotic stability of the sliding mode observer (14).

**Theorem 1** The origin of the system (14), (15) is asymptotically stable if there exist a positive definite matrix \( P = \begin{pmatrix} p_1 & p_3 \\ p_3 & p_2 \end{pmatrix} \) and \( W \in \mathbb{R}^{2 \times 1} \) such that the following Linear Matrix Inequalities (LMIs) are verified:

\[
\frac{1}{2} E_{12}^T P + \frac{1}{2} P E_{12} - C^T W^T - WC \geq 2 \Pi \begin{pmatrix} p_3 & \frac{1}{2} p_2 \\ \frac{1}{2} p_2 & 0 \end{pmatrix} < 0
\]

where \( \Pi = \text{sup}_{t \geq 0} |M e_x(t) + M_d e_x(t - h)| - \dot{d}_{max} \), \( C = [1 \ 0] \), and \( E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \).

The observer gain is obtained from \( L = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{\alpha} & 0 \end{pmatrix} = P^{-1} W \).

**Proof:** The convergence of the observer is proved by studying (14) and (15) separately.

First, it is shown in [7] that since \( M = M_d < 0 \), the stability of the origin of (14) is guaranteed independently of the delay \( h \). This implies that the quantity \( G(M e_x(t) + M_d e_x(t - h)) \) is ultimately bounded. Then, let us study the stability property of (15) using Lyapunov arguments. To this end, let

\[
\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} |e_y|^{\frac{3}{2}} \text{sign}(e_y) \\ e_z \end{pmatrix}
\]

be the new state vector.

Assume that \( \pi(t) = G(M e_x(t) + M_d e_x(t - h)) - \dot{d}(t) \) is a bounded function, i.e. for all \( t \), \( |\pi(t)| < \Pi \). The dynamics of the system in the \( \phi \) coordinates is given by:

\[
\dot{\phi} = |\phi_1|^{-1} \left( K \phi + \begin{pmatrix} 0 \\ |\phi_1| \pi(t) \end{pmatrix} \right),
\]

where \( K = \begin{pmatrix} -\frac{1}{2} & 0 \\ 0 & -\frac{1}{\alpha} \end{pmatrix} \) is Hurwitz.

In order to prove the finite time convergence of \( \phi \) towards zero, introduce the following candidate Lyapunov function:

\[
V = \phi^T P \phi,
\]

for a positive definite matrix \( P = \begin{pmatrix} p_1 & p_3 \\ p_3 & p_2 \end{pmatrix} \). Because of the stability condition (14), there exists a positive definite matrix \( Q \) such that:

\[
\dot{V} = -|\phi_1|^{-1} \phi^T Q \phi.
\]

This implies that \( \phi \) tends towards zero in finite time, leading to \( e_y = 0 \) and \( e_z = 0 \). Since \( \lim_{t \to \infty} e_x = 0 \), and \( e_z = z + G e_x - d(t) = 0 \), an estimation of the anomaly \( d(t) \) is obtained by integrating \( \dot{z} = -\text{sign}(e_y) \).

**IV. SIMULATIONS VIA NETWORK SIMULATOR NS-2**

For the validation of the proposed observer (14) in NS-2, the topology shown in Figure ?? in considered. The normal TCP traffic is generated by 60 TCP sources generating long lived TCP flows (FTP connections) to a receiver through a router with a link capacity \( C = 3750 \) packets/s (which is equivalent to 15 Mbits/s with a mean packet size of 500
bytes), and $T_p = 0.2s$ the propagation delay. Anomalous traffic is generated by 3 sources attacking the router. The maximal buffer size is set to 800 packets. To regulate the queue length of the router to the desired level $q_{ref} = 175$ packets, different AQMs are tested: RED [?] and Gain-K [?]. RED is an AQM designed to randomize packet drop by setting a marking probability according to the average queue length of the router. While the static state feedback Gain-K developed to stabilize the router queue length around an equilibrium point.

The observer scheme is implemented in the TCP network on the router level as shown in Figure 3. It is designed to detect the congestion window of the sources using, as inputs, the actual and retarded measured values of the queue length on the router and the dropping probability of a packet. This latter quantity representing the unknown input of the observer can be obtained by the adopted AQM.

In the simulations, two types of anomaly shapes are introduced using User Datagram Protocol (UDP) within the interval $50 - 100s$: a Constant Bit Rate traffic (CBR) and a Triangular Bit Rate (TBR), a traffic generator that we have implemented in NS-2 to simulate the triangular shape. This profile is justified in [?] where experimentations were held on softwares generating anomalous traffic like TFN2K [?]. The performance of the second order sliding mode observer is shown in Figure 3 comparatively to the first order observer developed in [?]. The improvement brought by the second order is proved by the fast detection of the presence of an anomaly in the network especially for the CBR detection in association with the Gain-K queue regulation. Furthermore, the instantaneous response for the disappearance of the anomaly is noticed. Besides of the convergence criteria, the chattering phenomenon is well reduced by the second order without introducing low pass filters. The high order sliding mode observer allows a better tracking of the real anomaly reaching the router buffer. Note that to reduce more the oscillations caused by the switching in the sign function (??), we replaced this latter by a continuous approximation with a high gain in the boundary layer like a saturation function:

$$sat(e_y) = \begin{cases} \frac{e_y}{\epsilon} & \text{if} |s| \leq \epsilon \\ \text{sign}(e_y) & \text{if} \ |s| > \epsilon \end{cases} \quad (12)$$

V. REAL TRAFFIC REPLAY

In this section we present experimental results based on real traffic trace collected on the RENATER1 network. To obtain our traffic trace, the capture tools installed in the experimental platform “LaasNetExp” at LAAS-CNRS2 are used. LaasNetExp [?] is a generic polymorphic platform for network emulation and experiments. From a specific machine, chosen to be the router for our network topology as in figure ??, incoming data and outgoing data are captured via the network analyser ”Wireshark” [?]. While capturing, anomalous packets have been sent from a machine in Mont-de-Marsan Institute of technology using TFN2K [?], a software for real attacks generation. The properties of the data transmitted are detailed later after recalling the procedure required for flow replaying tools in NS-2.

NS-2 contains the necessary features to replay traces. The simulator supplies only a single class to replay a real trace: “trafficTrace” which allows to generate flows from previously created trace files representing the list of the packets

1RENATER is the French National Network for Education and Research.
2The Laboratory of Analysis and Architecture of Systems (LAAS) is a CNRS research unit associated with the University of Toulouse, France.
sizes in each corresponding flow. Therefore in [?] a certain number of treatments is defined to be applied on metrological traces before the replay. Once the trace is obtained, an offline analysis allows extracting flows (TCP, UDP and others) from the trace as well as determining non trivial properties of every flow transmitted like the packets sizes, average round trip time, packets loss rate and links capacities. Certain modifications of the NS-2 code were also necessary to let the sources generate packets of predefined length.

Returning back to our experiments, the trace analysis showed 6 TCP flows including packets of different sizes. It is noticeable that the proposed observer (??) is conceived for a system based on average TCP dynamics (??). Therefore the methodology proposed in [?] will be taken into consideration for extracting the average characteristics of the real traffic. The average TCP packet size is found to be equal to 1 KBytes, the average values for the links capacities and the RTT are determined for each flow. The anomalous traffic sent from Mont-de-Marsan with a constant bit rate of 100 Kbits/s consists of several short then long sequences. The short attacks last 7 seconds each with 5 seconds between two consecutive attacks, the long attacks are 4 minutes each with a variable duration from 1 to 3 minutes between them. For now the characteristics of the sources and incoming links to the router are defined for the simulation topology as presented in figure ??.

For the capture and during the replay in NS-2, the router is configured to regulate the queue length using the mechanism of Token Bucket Filter (TBF). TBF is a common algorithm used to regulate the queue length of the router [?]. Moreover, to construct the proposed observer in (??), the equilibrium point is required. It is defined during the normal behavior of the TCP network, i.e. before the beginning of the attacks. The values of the congestion window size and the queue length at the equilibrium, i.e. \( W_0 \) and \( q_0 \), are determined from the mean value around which \( W(t) \) and \( q(t) \) oscillate respectively during intervals of 5 seconds. To summarize, the router characteristics as well as the equilibrium point (??) are presented in the following table: The results of the real traffic trace in figure ?? show

<table>
<thead>
<tr>
<th>Equilibrium point</th>
<th>Router configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W_0 )</td>
<td>5.8 packets</td>
</tr>
<tr>
<td>( q_0 )</td>
<td>17.08 packets</td>
</tr>
<tr>
<td>( p_0 )</td>
<td>0.05945</td>
</tr>
<tr>
<td>( R_0 )</td>
<td>4.057 s</td>
</tr>
<tr>
<td>( C )</td>
<td>0.1 Mbits/s</td>
</tr>
<tr>
<td>( q_{max} )</td>
<td>20 packets</td>
</tr>
</tbody>
</table>

**TABLE I** CONFIGURATIONS FOR TRACE REPLAY IN NS-2.

the efficiency of the second order sliding mode observer tested for a real TCP traffic replay. On the other hand, the graphs show the limitations of the theoretical TCP fluid flow model (??) in representing real TCP flows. In figure ??, before the attacks and during the short ones, the average congestion window \( W(t) \) is correctly estimated, but during the long attacks, the estimation does not follow the real mean evolution. This phenomenon can be explained by two real aspects:

- the sources are frequently forced to reduce their windows sizes towards zero, whereas one fundamental hypothesis in the model (??) is the persistence of emitting packets.
- the average size of the TCP packets during the attack is different than the average size taken for the observer design, thus varying the value of the congestion window at the equilibrium.

Consequently, these two practical facts perturb the estimation of \( W(t) \) thus the estimation of \( d(t) \). However the anomaly reconstruction reaches 12 packets/s instead of 15 packets/s as illustrated in the figure ??.

![TCP congestion window W(t)](image)

![queue length q(t)](image)

![anomaly d(t)](image)

Fig. 4. Estimations with real traffic replay.
mode observer [2] on real traffic conditions with respect to the hypothesis considered in the theoretical model.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, a second order sliding mode observer has been designed to detect anomalies in TCP/IP networks. Simulations via Simulink and NS-2 exhibit a guaranteed Quality of Service (QoS) with different Active Queue Management (AQM) mechanisms like RED and Gain-K. Our methodology on real traffic properties shows a successful anomaly detection, leading thus to an open way towards applying the control theory techniques and especially sliding modes to the anomaly detection problematic.

This research work applied on TCP network can be improved in both network and control domains. The fluid flow model adopted for this study showed some practical limitations. Future work will be focused on an improved model representing more realistic TCP behavior. Besides, different AQMs in addition to TBF could be considered for other trace replay tests. Another proposal is to address the problem of anomalies in a large-scale distributed manner.

ACKNOWLEDGEMENTS

The authors would like to thank David Gauchard, an Engineer in Computer Sciences in LAAS, for the essential technical help in configuring the LaasNetExp platform. They would also like to acknowledge the work of Laurent Gallon, an Associate Professor in Mont-de-Marsan Institute of Technology for his help in generating network anomalies. Without their support, it would have been very difficult to realize the open experiments.

REFERENCES