Modeling Privacy Preservation in Smart Connected Toys by Petri-Nets

Benjamin Yankson[1,2], Farkhund Iqbal[3]
Zhihui Lu[4], Xiaoling Wang[5], Patrick C. K. Hung[1]
[1] Faculty of Business and IT, University Of Ontario Institute of Technology, Canada
[2] Sheridan College, Canada
[3] College of Technological Innovation, Zayed University, UAE
[4]School of Computer Science, Fudan University, China
[5]School of Computer Science and Software Engineering, East China Normal University, China
{benjamin.yankson, patrick.hung}@uoit.ca; farkhund.iqbal@zu.ac.ae
lzh@fudan.edu.cn; xlwang@sei.ecnu.edu.cn

Abstract

Children data privacy must be considered as integral and factored into the system design of Smart Connected Toy (SCT). The challenge is that SCTs are capable to gather significant amount volunteered and non-volunteered data, which lacks privacy considerations. It is imperative to adopt a modeling technique that autonomously preserves privacy and secure children's data in SCT transactions. This paper surveys the current data flow modeling techniques, which most of them do not have elements to address the privacy of Personal Identifiable Information (PII). This paper shows a Petri-Net simulation which provides privacy assurance in order to minimize the risk of privacy violation of a child's PII and related data.

Keywords: Smart Connected Toys (SCT), Petri-Nets, Privacy, Data Flow Modeling, Simulation

1. Introduction

With the advent of Smart Connected Toys (SCTs) such as Hello Barbie and Cognitoy Dino, the privacy of children's information has now become a growing concern. The child's user-generated sensitive data, their context data and the Personally Identifiable Information (PII) provided by the parent can be directly linked to the child and ultimately their safety. Referring to the National Institute of Science and Technology (NIST) Special Publication 800-128, PII is defined as "any information about an individual maintained by an agency/organization, including: (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, and biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information" [1]. Many SCTs are integrated with sensory and networking capabilities, allowing for new opportunities, and extending capabilities outside the confinement of the toy itself, as well as providing an opportunity for a child's information to be transmitted to an external source, such as Cloud services. SCTs allow the integration of a personalized application, which in turn provides additional opportunities and added value to each individual child's gameplay or learning experience. SCTs are built as part of the Internet of Things (IoT) with the capability of providing Location-Based Services (LBS), Mobile Advertisement (MA), Geo-Social Network applications (GeoSNs), and contextual data collection [2].

Children provide a unique user-base, for toy manufactures, which requires special attention in several key areas related to privacy. Firstly, it is widely accepted in most jurisdictions that a child's data is considered particularly sensitive and should be treated with extreme care. Online privacy for children has been a great concern. This concern is inherited into the SCT computing environment; particularly when the child's location can potentially be shared with other parties resulting in harassment, stalking, grooming, sexual abuse, exploitation, or personal data misuse [4]. Sexual solicitation and Internet-initiated offline encounters, which SCTs can provide an avenue for, are a chief concern for the online safety of children. The U.S. Department of Justice indicates that "1 in 25 youth received an online sexual solicitation in which the solicitor tried to make offline contact" [5]. Other concerns include market researchers attempting to collect a child's personal data and usage patterns for targeted advertising and third-party advertisers inferring

URI: https://hdl.handle.net/10125/59610
ISBN: 978-0-9981331-2-6

information about a child based on their location and detailed behavioral patterns that may be used for the undesirable purposes [6]. Referring to de Carvalho and Bandiera-Paiva [17], the privacy and security requirements to protect any data, and restrict principals' interaction, must consider all users [17].

In the SCT design and system development, a key component to consider, and one which must be of focus, is the privacy of the user-generated data created by the child during play or involuntarily collected from SCT's context. For example, the user-generated data in the SCT can range from basic audio recordings to a video of an activity which was unintentionally recorded during play. Contrary to the user-generated data, context data mainly focuses on a composition of various data points generated by the SCT from its operating environment. For example, the location data based on Global Positioning System (GPS) coordinates, and other historical use of the SCT automatically generated and stored or shared by the SCTs. Taking into consideration the importance of the privacy of children's PII, through either context or user-generated data, a degree of emphasis must be placed on the overall system design of the SCT, and its ability to factor in privacy preservation elements right from the initial stages of the SCT system design. Systems Development Life Cycle (SDLC) is a process consisting of four phases and adopted for the development of a software system including analysis phase, design phase, implementation phase, and testing phase [3]. With the current SDLC approach, the concept of the end user's information privacy, or information security, is often an afterthought brought on by an information breach or when a vulnerability has been identified that has potential to result in undesirable consequences.

Considering the significance of the data created and collected through a child's interaction with SCT, PII collected by toy makers, and the contextual data generated by the toy; any improper data flow models employed in the system development without a proper privacy element mechanism within the toy's security infrastructure can lead to improper disclosure of a child's information and potentially put the child's safety at risk. Data-flow modeling and verification is an important challenge for traditional system workflow management[21]. The criticality of data-flow verification was first mentioned in [22]. Currently, Data Flow Diagram (DFD), Privacy-Aware Data Flow Diagram (PA-DFD), and Privacy by Design (PbD) are widely adopted preferred models used within the industry to model data flow[3]. These existing approaches although are widely accepted have common drawbacks including but not limited to: lacking specific privacy element necessary to support sensitive data as part of the design element; lacking basic characteristics that support the concept of privacy of PII or context data that flows through an information system like a SCT; and finally lacking concrete semantics, and verification, which makes it challenging to truly identify privacy violation. Our proposed approach integrates privacy elements in the modeling of the system flow, offer a well-defined mathematical semantics for verification enforcing a privacy to provide a sound level assurance to sensitive user information, and ability to simulate the data flow to project the behavior of the SCT data flow and possible privacy violation. This paper makes the following contributions: (1) Conducting a survey of the current data flow modeling techniques; and (2) Simulate a data flow model with privacy elements for privacy preservation by Petri-Nets. This paper is divided into the following sections: (1) Introduction; (2) Data Flow Model and Petri-Nets; (3) SCT Data Flow Simulation; and (4) Conclusions and Future Works.

2. Data Flow Model and Petri-Nets

This section surveys and analyzes Data Flow Diagram (DFD), Privacy-Aware Data Flow Diagram (PA-DFD), and Privacy by Design (PbD) widely adopted preferred models use in the SCT design. For example, de Cavalho, and Eler [18] used high-level DFD to model data flow in order to identify privacy and security threats within the smart toy environment. To ascertain detail requirements of system development, DFDs are used to produce the process model [8]. DFDs have been the industrial most widely used approach to information systems design. A "Complex process" also exists which consists of multifaceted functionality, or computation, that is detailed in an additional DFD [3], and "Data deletion" which is an extension of another type of flow, which acts as a *data store for the incoming* flow of information [7]-[23]. DFDs do not have the basic characteristics that support the concept of privacy of PII or context data that flows through the SCTs. Although DFDs are the most popular modeling technique, they fail to be adequate in designing a system which maintains the required privacy protection and handles sensitive user data like that of children; due to lack of specific elements to address privacy [7]. On the other side, PA-DFDs require the system analyst, or architect, to identify a classification of the data flows as personal data or non-personal data [3]. Additionally, information for personal data flow must include the name of the external entity as well as which personal data will flow. A typical PA-DFD identifies (i) the purpose for the data to flow, and (ii) the retention time for the use of the personal data [7]. This new information plus the existing DFD annotation is needed to detect the part in the model and transformed to privacy aware notation for PA-DFDs

[3]. Although PA-DFDs seems to be a good solution for modeling SCT data, it is found that this modeling solution lacks concrete semantics, and verification, which makes it challenging for its intended usefulness or the ability guarantee that no privacy violations will occur within the SCTs [7].

In evaluating DFD, PA-DFD, and PbD, these methodologies lack certain requirements to be used to successfully model the SCT privacy framework. The required stature of DFDs elements do not have a symbol or element which to represent privacy; and lacks, classification of the data flows as personal data or non-personal data [3]. Alternately, PA-DFDs which seem to be a good solution for modeling SCT data, lack concrete semantics and verification required to test [7]. Lastly, although PbD has the cited principles above, and it is expected to be integrated prior or as part of to design, development, and implementation of any information system serving as insurance for privacy assurance, there is no enforcement mechanism to guarantee its implementation. The alternative option to DFDs, PA-DFDs, and PbD is to explore other modeling techniques or approaches, such as Petri-Nets. Petri-Nets have the capacity to integrate privacy elements in the modeling of the system flow and offer a well-defined mathematical semantics for verification enforcing a privacy policy to provide a sound level assurance to sensitive user information.

Petri-Net is a well-known mathematical modeling language that can be formally tested and verified. They are powerful modeling formalisms in computer science, and many other disciplines, which can address all the shortcomings of DFDs, PA-DFDs, and PbD. Petri-Nets utilize a “token”, as a primitive concept, and it is depicted with black dot residing inside a “place” of a Petri-Net graph [7]. Tokens can be at or can be absent in certain places, stipulating whether conditions associated with those places are true or false [11]. In modeling Petri-Nets, a change of state is denoted by a movement of the token from place to place; which is triggered by the firing of a transition; representing an occurrence of an action or an event. Generally, a transition is *enabled* when there are sufficient tokens in its input place, but the firing of any transition is subject to token availability and input condition. After firing, tokens will be transferred from the input places (old state) to the output places, denoting the new state [13].

Colored Petri-Net (CPN) was introduced by Kurt Jensen [19] to address the issue with unstructured Petri-Nets due to inability to distinguish between tokens in basic Petri-Nets. CPN is considered as a discrete-event modeling language; which has been under development since 1979 by the CPN group at Aarhus University, Denmark [7]. In a Colored Petri-Net diagram, a token is distinguishable from other tokens by using a unique color for each token. In addition, CPN addresses the issue undistinguishable tokens in basic Petri-Nets, by attaching a place of a CPN with a color set and allowing multi-color tokens. CPN has the same elements as basic Petri-Nets, and the transition functions operate in the same manner. Along with the characteristics of basic Petri-Nets, CPN uses high-level programming language based on the functional programming language Standard Markup Language (SML) [14]. CPN SML provides primitives for defining data types and various data manipulation, which makes models be compact [19].

3. SCT Data Flow Simulation

We performed simulations of the model of the SCT system in learning about different states and behaviors regarding privacy of data (tokens) [20]. The SCT privacy is expected to have multiple states to form a global state of privacy for the system, so the ability to model concurrency and synchronization is needed. A system may have many local states to form a global state. This section has been broken into three: Model Configuration, Data Flow Simulation, and Discussion.

3.1. Model Background & Configuration

In our experiment, the model is created using a Petri-Net diagram. In our model, a change of state is denoted by a movement of the token from place to place; which is triggered by the firing of a transition; representing an occurrence of an action, such a turning the SCT on or recording audio.

Definition 1 - A Petri-Net is formally defined as a five-tuple N= (P, T I, O M_0), where [11]:

I. $P = \{p1, p2 ,\ldots, pm\}$ is a finite set of places;

II. $T = \{t1, t2, \ldots, tn\}$ is a finite set of transitions, $P \cup T \neq \emptyset$, and $P \cap T = \emptyset$;

III. $I{:}P \times T \rightarrow N$ is an *input function* which defines directed arcs connecting places to transitions. Here, N is a set of nonnegative integers;

IV. $O{:}T \times P \rightarrow N$ is an *output function* defining directed arcs from transitions to places; and

V. $M_0{:}P \rightarrow N$ is the *initial marking*.

Definition 2. Assume there is a non-empty set S = {s1, s2, s3… sN}. A multiset over S is a function $m\colon S \rightarrow N$ that maps each element $s \epsilon S$ into a non-negative integer (s) ϵ N called the number of appearances (coefficient) of s in m [6]. The *net structure* consists of a finite set of places, P, a finite set of transitions, T, and

Definition 3. A Colored Petri Net (non-hierarchical) can be represented as a nine-tuple CPN= (P, Σ, V, C, G, E, I) [6], where (Adopted from [11]):

i. The finite set of *places* is denoted by P;
ii. The finite set of *transitions* is denoted by T;
iii. The finite set of directed *arcs* is denoted by $A \subseteq (T \times P) \cup (P \times T)$;
iv. The finite set of *color sets* is denoted by Σ;
v. The finite set of *typed variables* is denoted by V, where $\forall v \epsilon V$. Type$[v]$ ϵ Σ;
vi. A *color set function, which assigns a color set to each place,* is denoted by C:$P \rightarrow \Sigma$;
vii. A *guard function* which assigns a guard to each transition t is denoted by G:$T \rightarrow EXPRv$ such that $[G(t)]$=$Bool$; and
viii. An *arc expression function* is denoted by E:$A \rightarrow EXPRv$. For each arc $a \ \epsilon \ A$, this function assigns an expression such that $[E(a)]$=$C(p)MS$. Here, $p \ \epsilon \ P$ is connected to the arc a.

In this paper, the SCT privacy framework depicting various component of SCT environment is proposed to model in CPN. Referring to Figure 1, Conceptual Model of Toy Computing Environment includes the standard real-life situational environment of an SCT including *Physical and Social Environment*, *Cloud Service Environment*, and *Monitoring Environment*. The *Physical and Social Environment* of the SCT and includes similar SCTs and an online connectable device such as WiFi. Within this environment, context data such as geo-location, original demographic registration information (name, age, gender, and address), directly created interaction data, and activity data is available. The expected types of data including but not limited to interaction captured data by the SCTs through a microphone, camera, etc. The SCTs will be generally equipped with a camera, microphone, GPS, and sensors for face and sound detection which allows the device to create and collect such data. Within the *Cloud Service* environment, the SCT manufacturers provide external services through Cloud services outside the immediate environment of the SCTs. This allows data to be exchanged or sent across from the SCTs to the cloud. For example, text, picture, video, sound (voice), and location and sensing data to the SCT manufactured services provider. Generally, they may be other information which can gather and infer from SCTs involved prior activity including historical data on the child such as the SCTs move around. Within the SCT, a tremendous amount of information is gathered, exchanged and transmitted to a connected to Cloud services. Within the *Monitoring Environment*, the idea is prior to any communication to the Cloud service provider, a parent/guardian will have configured a privacy preference file which then is incorporated into the privacy policy for notification of any noncompliance. This attests that the guardian will be in charge to monitor child activities and be alerted in case any of the rules in the privacy policy is breached. Generally, a child (data subject) is associated with an identity, but the parent is the data owner and control access (read, write, modify) and use of the data other than a privilege granted to the child. It is because context data including location data, and it can lead to the identification of the child and his or her location. It is incumbent on the system to provide a level of initial privacy preference through the data flow model. This means that the SCTs cannot be used until the preference file is configured for access by the parents.

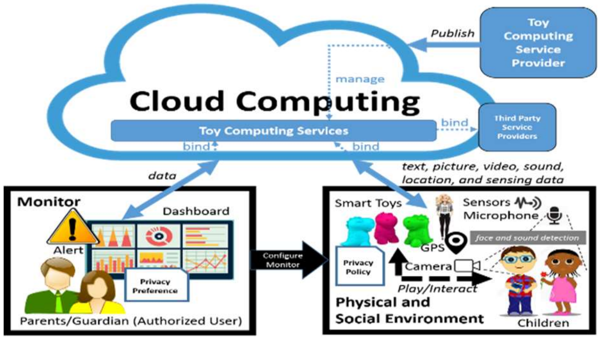

Figure 1: Conceptual Model of Toy Computing Environment

3.2. Data Flow Simulation

Figure 2 and 3 demonstrate the privacy of SCT physical and social environment. In order to simulate the SCT physical and social environment, we defined unique tokens represented by a color set. This color set represents the various tokens illustrating tokens within places. The color set as defined for SCT includes:

- "A" (Activate SCT): This is the token which can be a trigger to turn on the SCT.
- "MED" (Media (Audio, Video)): This is the type of data the SCT can transmit.
- "CON" (Context Data): This is context data such as GPS data gathered by the SCT.
- "TXT" (Text Data): This is generally text data in a situation where the SCT can.
- "MEDPP" (Media with Privacy preference): Audio, video, an image with privacy preferences.
- "CONPP" (Context Data with Privacy Preference): Geolocation data with privacy preference.

- "TXTPP" (Text with Privacy Preference): Text data with privacy preference.

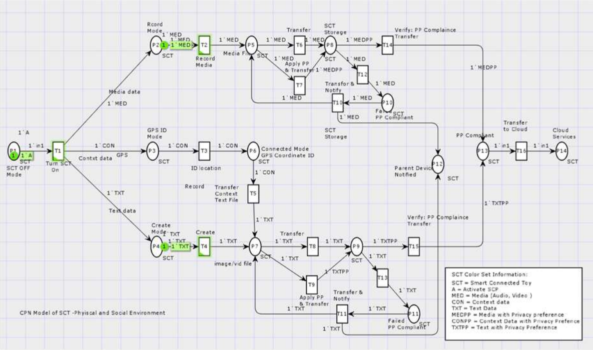
CPN Model of SCT -Phyiscal and Social Environment
SCT Color Set Information:
SCT = Smart Connected Toy
A = Activate SCP
MED = Media (Audio, Video)
CON = Context data
TXT = Text Data
MEDPP = Media with Privacy preference
CONPP = Context Data with Privacy Prefence
TXTPP = Text with Privacy Preference

Figure 2: CPN Model of SCT – Physical and Social Environment with Token at P1, P2, and P4

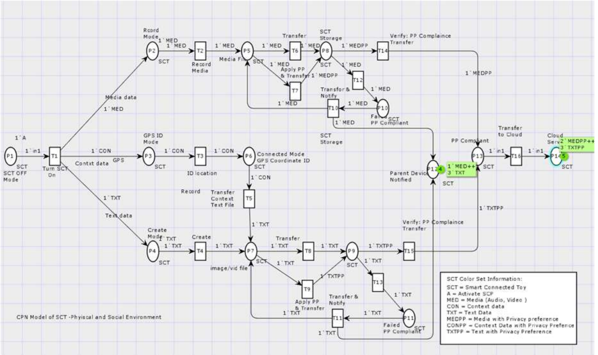
CPN Model of SCT -Phyiscal and Social Environment
SCT Color Set Information:
SCT = Smart Connected Toy
A = Activate SCP
MED = Media (Audio, Video)
CON = Context data
TXT = Text Data
MEDPP = Media with Privacy preference
CONPP = Context Data with Privacy Prefence
TXTPP = Text with Privacy Preference

Figure 3: CPN Model of SCT – Physical and Social Environment after the first full transition

Referring to Figure 2, prior to the SCT being turned on, at P1, the SCT is in "off" mode and as such no concern about the privacy of any data set can be garnered from the SCT. Upon the T1, the first transition, the SCT is put in an "on" mode, with possibly three places activated simultaneously. At this stage, as depicted in Figure 2, the SCT has the capacity to be in the record, context data gathering, or text input mode. The SCT in its current place (P2, P3, and P4) and the next transition (T2, T3, T4) can occur simultaneously. Now if we consider T2, and that the SCT is recording audio/video information, the next place P5 now contains audio or video file. At this juncture, the SCT can fire transition through T6, by sending the video or audio file directly to SCT internal storage P9, or by T7, applying the privacy preference rules and transition to P9 (SCT internal storage). Another option of transition from P5 is the audio/video created content through T5, which delete results in a sink transition. In this case, if the SCT did not have any privacy policy attached, and transition from P5 (audio/video file) to P9 goes through T6, the next transition at P9 is to through T14 (to delete) and back to P2 (record mode). The other option from P9 would have to proceed to T15 (verify PP compliance), which will end up in P12 (failed PP complaint store) considering that the SCT transition through T6 without applying the privacy preference. At this point in P12, the SCT will transfer, T12, audio file back to P5. Assuming that after failed application from the initial place, transition run, P5 containing the audio file transition to P9 through T7, the SCT will still have equal option to transition to T14 (delete) back to P2 or transition through T15 to P14 (PP complain store). In this state, the SCT will transition T18 (verify Wi-Fi connection & transfer) to P15 (Cloud services or store) if the condition of Wi-Fi connectivity is met. If the condition fails, T18 returns to P14 the repeat again until it is successful or if to delete the file through T14 based on the privacy preference condition with data retention requirement, which is a sink transition. Text data goes through the similar place and transition stages as illustrated in Figure 2 except for text data the next Place after P1 is P4. Similarly, context data gathered in the SCT will have to go through a comparable transition as described during the recording of information. Based on the initial policy, the SCT transitions from P1 through T1 to P3 (GPS ID). P3 is a place for context data gathering, and results in transition through T3 where the system tries to identify context information and based on the result will end up in P6 (not connected mode; no context information) or P7 store of context information. If the system is at P7, the SCT system is forced to apply privacy preference and transfer content to storage P10. The next transition T13 verifies PP compliance and transfer to P14 (PP compliant store) before it transitions to T18 onto P15 (Cloud service). With the scenario above, there was only 1 initial token at P1; we assume that this is the first time SCT is been used or there is no existing data anywhere within P2, Pi (where i represents highest number place within the system). Considering how SCT is been used, there can be multiple tokens at any given time within P1 through P15. The results are shown in Figure 3.

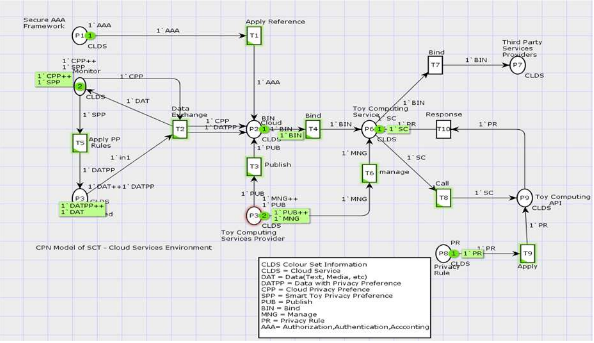
CPN Model of SCT - Cloud Services Environment
Secure AAA Framework
Apply Reference
Third Party Services Providers
Toy Computing Service
Toy Computing Services Provider
Toy Computing API
Privacy Rule
CLDS Colour Set Information
CLDS = Cloud Service
DAT = Data(Text, Media, etc)
DATPP = Data with Privacy Preference
CPP = Cloud Privacy Prefence
SPP = Smart Toy Privacy Preference
PUB = Publish
BIN = Bind
MNG = Manage
PR = Privacy Rule
AAA= Authorization,Authentication,Accconting

Figure 4: CPN Model of SCT - Cloud Services Environment Initial State

The Cloud service of SCT can be modeled with Petri-Net to provide privacy assurance of data as illustrated in Figure 4 and 5. The CLDS (Cloud Service) color set includes: "DAT" (Data(Text, Media)), "DATPP" (Data with Privacy Preference), "CPP" (Cloud Privacy Preference), "PP" (Smart Toy Privacy Preference), "PUB" (Publish), "BIN" (Bind), "MNG" (Manage), "PR" (Privacy Rule), and "AAA" (Authorization, Authentication, Accounting).

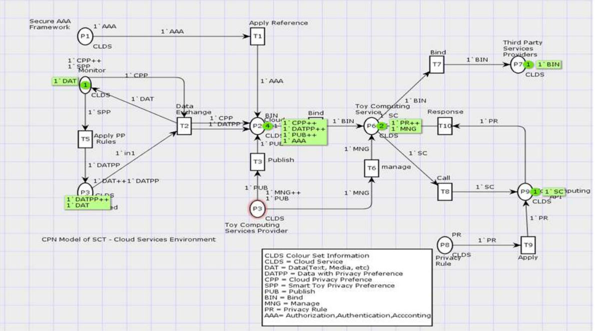

Figure 5: CPN Model of SCT - Cloud Services Environment after First Full Transition

While the data exchange occurs between P4 and P5 as well as P1 has referenced on P2, P2 will transit through T4 (bind) to P6 (Toy Computing Service, such as a Cloud storage). At this point, P6 can do the transition from T7 to P7 (Third Party Service Providers), or from P6 through T8 (call) to P9 (Toy Computing Application Programming Interface (API), which the services can use to access the privacy preference. Prior to the transition of P6 to P9, it is expected that P8 (Privacy Rule) would have transitioned through T9 (Apply) onto P9 before P9 would have transition T10 (Response) back to P6. At an initial offset, there is a possibility of multiple tokens and the current token which can be initialized at the same time. For example, the initial start can contain tokens at P1, P4, P5, and P8.

The parental model interface of the privacy preservation framework component is modeled by Petri-Nets in Figure 6 and Figure 7. The Monitoring Interface begins at P1 (Blank Template) within the initial token, the next transition T1 (create) allows a transition from P1 to P2 (Privacy Preference template). At this point, the system will do a transition through T2 (Apply P. Preference) to P3 (Monitor). At P3, the system will transition through T3 (Configure Rules) to P4 (Dashboard), and transition T4 (Apply Rules) to P5 (SCT). On P5 at any time the apply rules for privacy preference are a breach; the system will trigger an alarm through T5 and alert back on P4, at any time P6 transition through T5 and the collected information on P5 breaches Privacy Preference rules applied. At the initial stage (initial marking) of this Petri-Net diagram, P1 to P5 can contain tokens (e.g., M (1, 0, 0, 1, 0)). The privacy state of data is defined as a possible state where a privacy preference has been applied, and the data itself can be identified as private data, public data, anonymous data, encrypted data, confidential data, and de-identified data. Based on the modeling activities such set of data can transition from one state to the other.

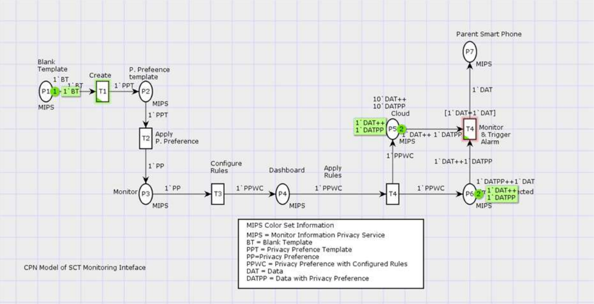

Figure 6: CPN Model of SCT - Monitoring Interface Initial State

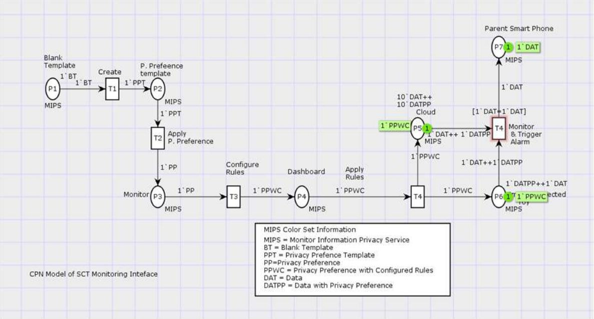

Figure 7: CPN Model of SCT - Monitoring Interface First Full Transition

The color set definition can be found on the color set information on the CPN model diagram in Figure 8 and 9. At the initial state P1, all the data is private and created either by the SCT through the children. If the SCT fires transition T1 by inferring information from the SCT, the data is still private and returns to P1. If P1 decides to process data and do transition through T2 to P2 (Data Storage), the data can be subject to various activities or possible transition. If P2 does transition to T3 (Release of Infor) and releases Information in this move, the data is now in P4 (Public) which has become public record. At this point, any transition which occurs T6 returns to P4 (Public). That record will become public record. Now considering that at P2, the System does transition through T4 (De-identified Data) to P5 (Anonymous Data), the data has become anonymous. At P5 there is a possibility to do transition through T8 to P6 (Unidentified information). With P6 any transaction which can result in a transition such as T9 will return to

P6. Considering at P5, the system transition through T7 (Aggregate data) to P3 (Aggregated Data), the information is reassembled, and they system can Release information by transition T3 to P4 which is also public record. As per stated earlier, at P4 all transaction will result in a back as a public record.

The other option, other than T4 and T3, is for the system to do transition through T5 (Encrypt Data) to position P7 (Encrypted Data). Once at P7, the system can do transition through T10 (Release of Info) to P8 (confidential Data storage). The next Transition will be T11 (Decrypt) which allow information to be decrypted back to P1. At any time within the lifecycle of information within SCT, the initial Petri-Net can have a token at P1. If this is not the first time of use, we can have concurrent and multiple tokens within various places of the system. Figure 8 provides details information after a full transition has occurred. Table 1 depicts sequences of increasing tokens to demonstrate, the initial state of "CPN Model of SCT – Physical and Social Environment" and Table 2 shows the result of the first transition. Subsequently, Figure 5 above shows CPN Model of SCT – Physical and Social Environment with initial M (1,1,0,1,0,0,0,0,0,0,0,0,0,0,0). Table 2 and Figure 8 provide information after the first transition has occurred. All the "places" with zero tokens, in Figure 1, indicates that tokens have not reached a "place" as no transition has occurred.

Total Number of Tokens	Number of SCT Activate Tokens	Number Of Media Data Tokens	Number of Context Data Token	Number of Text Data Tokens	Number of Transition	Number of Parental Notification Sent (TEXT)	Number of Parental Notification Sent (MEDIA)	Total Number Parental Notificatio n Sent	Number of PP Compliant (TEXT)	Number of PP Compliant (MEDIA)	Total Number of PP Compliant
3	1	1	0	1	0	0	0	0	0	0	0
5	1	2	0	2	0	0	0	0	0	0	0
7	1	3	0	3	0	0	0	0	0	0	0
9	1	4	0	4	0	0	0	0	0	0	0
11	1	5	0	5	0	0	0	0	0	0	0
13	1	6	0	6	0	0	0	0	0	0	0
15	1	7	0	7	0	0	0	0	0	0	0
17	1	8	0	8	0	0	0	0	0	0	0
19	1	9	0	9	0	0	0	0	0	0	0
21	1	10	0	10	0	0	0	0	0	0	0
201	1	100	0	100	0	0	0	0	0	0	0
1001	1	500	0	500	0	0	0	0	0	0	0
2001	1	1000	0	1000	0	0	0	0	0	0	0
3001	1	1500	0	1500	0	0	0	0	0	0	0
4001	1	2000	0	2000	0	0	0	0	0	0	0
5001	1	2500	0	2500	0	0	0	0	0	0	0
6001	1	3000	0	3000	0	0	0	0	0	0	0

Table 1: CPN Model of SCT – Physical and Social Environment Initial State Prior Transition

Total Number of Tokens	Number of SCT Activate Tokens	Number Of Media Data Tokens	Number of Context Data Token	Number of Text Data Tokens	Number of Transition	Number of Parental Notification Sent (TEXT)	Number of Parental Notification Sent (MEDIA)	Total Number Parental Notificatio n Sent	Number of PP Compliant (TEXT)	Number of PP Compliant (MEDIA)	Total Number of PP Compliant
5	0	2	1	2	1	0	0	0	0	0	0
7	0	3	1	3	1	0	0	0	0	0	0
9	0	4	1	4	1	0	0	0	0	0	0
11	0	5	1	5	1	0	0	0	0	0	0
13	0	6	1	6	1	0	0	0	0	0	0
15	0	7	1	7	1	0	0	0	0	0	0
17	0	8	1	8	1	0	0	0	0	0	0
19	0	9	1	9	1	0	0	0	0	0	0
21	0	10	1	10	1	0	0	0	0	0	0
23	0	11	1	11	1	0	0	0	0	0	0
201	0	101	1	101	1	0	0	0	0	0	0
1001	0	501	1	501	1	0	0	0	0	0	0
2001	1	1000	0	1000	1	0	0	0	0	0	0
3001	1	1500	0	1500	1	0	0	0	0	0	0
4001	1	2000	0	2000	1	0	0	0	0	0	0
5001	1	2500	0	2500	1	0	0	0	0	0	0
6001	1	3000	0	3000	1	0	0	0	0	0	0

Table 2: CPN Model of SCT – Physical and Social Environment Result after 1st Transition

The other test begins with sequencing different number of MED, TXT data tokens which are added to the first transition place (P2, P4). Once the plain data is added, the "CPN Model of SCT – Physical and Social Environment" model will process the tokens applying well-defined privacy preference, assuring that all data is protected before moving to the cloud. In addition, in cases where the model fails to apply privacy preferences, the parent or guardian gets notified. Table 3 demonstrates results of the final nth transition of a defined set of tokens, ranging from 1 through 6000, and results of "Number of PP Compliant (TEXT)", "Number of PP Compliant (MEDIA)" and "Total Number Parental Notification Sent" after compliance fails. Figure 9-11 show a demonstration of model processing a 1001 token ("A", "MED", "TXT"), at the initial state, at the middle of the process and at the final transition. Figure 12 provides a comparative analysis of Total Number of Parental Notification Sent vs. Total Number of PP Compliant.

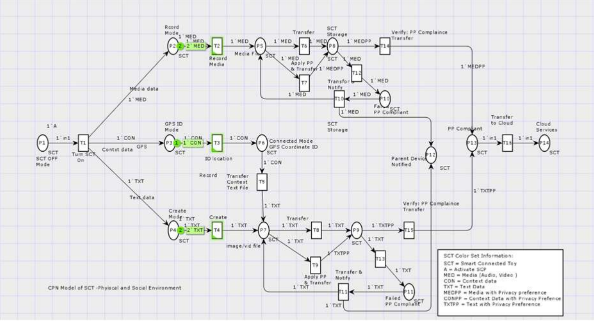

Figure 8: CPN Model of SCT – Physical and Social Environment Result after First Transition

Total Number of Tokens	Number of SCT Activate Tokens	Number Of Media Data Tokens	Number of Context Data Token	Number of Text Data Tokens	Number of Transition	Number of Parental Notification Sent (TEXT)	Number of Parental Notification Sent (MEDIA)	Total Number Parental Notificatio n Sent	Number of PP Compliant (TEXT)	Number of PP Compliant (MEDIA)	Total Number of PP Compliant
5	1	1	0	1	100	6	0	6	3	2	5
7	1	2	0	2	100	4	6	10	4	3	7
9	1	3	0	3	100	6	1	7	5	4	9
11	1	4	0	4	100	6	8	14	6	5	11
13	1	5	0	5	100	7	7	14	7	6	13
15	1	6	0	6	100	9	3	12	8	7	15
17	1	7	0	7	1000	16	10	26	9	8	17
19	1	8	0	8	1000	18	15	33	10	9	19
21	1	9	0	9	1000	14	12	26	11	10	21
23	1	10	0	10	1000	13	14	27	12	11	23
201	1	100	0	100	10000	112	103	215	102	101	203
1001	1	500	0	500	10000	531	504	1035	502	501	1003
2001	1	1000	0	1000	100000	981	1025	2006	1002	1001	2003
3001	1	1500	0	1500	100000	1500	1588	3088	1502	1501	3003
4001	1	2000	0	2000	100000	2048	1973	4021	2002	2001	4003
5001	1	2500	0	2500	100000	2569	2537	5106	2502	2501	5003
6001	1	3000	0	3000	100000	2954	3068	6022	3002	3001	6003

Table 3: CPN Model of SCT – Result After nth Transition for various set of Tokens

3.3. Discussion

Based on the experimental results and discussion provided the following findings: (1) Excessive number of notifications more than expected; (2) Any type of notification system implemented for non-compliant within the Privacy Preservation engine needs to be tailored to exactly what the users is concern about. Example if users it only concerns about GPS information be shared by the SCT, then the notification should be tailored only to be alarming Guardian of non-compliance Notifications need; and (3) Considering there is no constraint on the number of tokens that can be processed by the SCT, the limited resource can fully exhaust causing the system to crash or fail. This result will help us in designing a theoretical model of privacy preservation engine as the result of the experiment shows that: (1) Petri-Net allows model my privacy preservation engine system design, simulate how the system will process information by the SCT, and provide us an opportunity to identify unsuspected flaws within the operations of the theoretical model of privacy engine; (2) It will allow us to identify areas within the engine, where unique information security technical, administrative or operational controls needs to be implemented to provide an extra mechanism to address threat and vulnerability within the theoretical privacy engine based on any information flow within the SCT; and (3) This provides us an opportunity to on an ongoing basis further simulate the behavior of the SCT privacy engine anytime a change is made to the SCT information flow prior to deploying it into an actual production environment.

The advantages of our proposed models, as presented in our simulated experiment, over existing models such as DFD, PA-DFD and PbD include: First, our proposed Petri-Nets models' ability in dealing with concurrences and conflicts during data flow. As demonstrated in our simulated experiment, the models are able to handle concurrent tokens going through the system at the same time. Comparatively, in modeling DFD, or PA-DFD there is no simple way to model, simulate and verify the behavior concurrency or conflict. Second, compared DFD, PA-DFD, and PbD, a formal semantics have been defined for Petri-Nets model in our experiments making it possible to verify and test to tokens for privacy violation. Our Petri-Nets models have the capacity to integrate privacy elements in the modeling of the system flow and offer a well-defined mathematical semantics for verification enforcing a privacy policy to provide a sound level assurance to sensitive user information. The other discuss modelling approach, such as DFD, lacks concrete semantics and verification required to test privacy [7]-[24]. Without the ability to do this, it makes it challenging for its intended usefulness or the ability guarantee that no privacy violations will occur within SCT data flow. Basically, the required stature of DFDs elements do not have a symbol or element which to represent privacy; and lacks, classification of the data flows as personal data or non-personal data [3]. Third, our Petri-Nets models are state-based instead of event-based, so each state of an instant case can be modeled explicitly and simulated to determine the behavior and final result of each token. In our experiment, we are able to determine what will happen if a defined token fails privacy verification test within the system. Such a result cannot be achieved with DFD, PA-DFD, or PbD model. Lastly, another advantage worth noting here is that although modeling techniques such as DFD, PA-DFD can describe the boundaries of the system data flow[21]-[25] it fails to be able to test the boundaries to determine its behavior as Petri-Net models can be simulated to present the results.

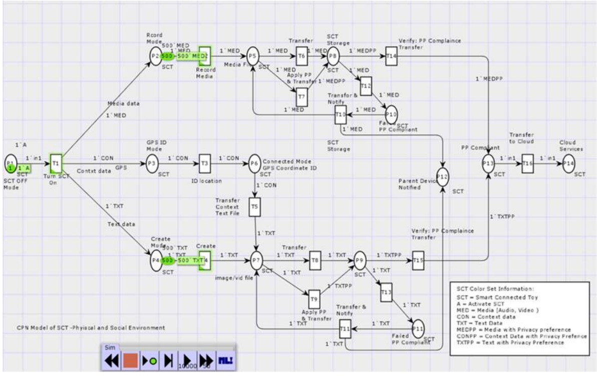

Figure 9: CPN Model of SCT – Physical and Social Environment with 1001 Tokens at Initial State

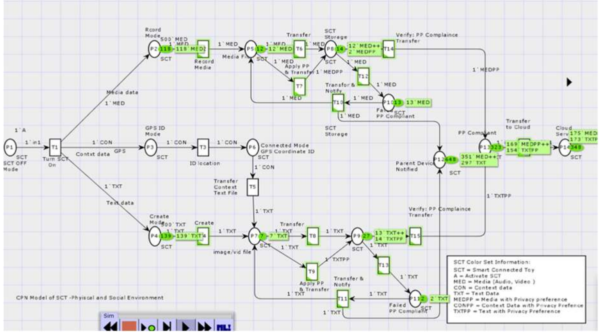
Rcord Mode
Record Media
Media File
Transfer
Apply PP & Transfer
SCT Storage
Verify: PP Complaince Transfer
Transfer & Notify
SCT OFF Mode
Turn SCT On
Contxt data
GPS ID Mode
ID location
Connected Mode GPS Coordinate ID
Transfer Context Text File
Parent Device Notified
PP Compliant
Transfer to Cloud
Media data
Text data
Create Mode
Create
image/vid file
Failed PP Compliant
CPN Model of SCT -Physical and Social Environment
Sim
SCT Color Set Information:
SCT = Smart Connected Toy
A = Activate SCT
MED = Media (Audio, Video)
CON = Context data
TXT = Text Data
MEDPP = Media with Privacy preference
CONPP = Context Data with Privacy Prefence
TXTPP = Text with Privacy Preference

Figure 10: CPN Model of SCT – Physical and Social Environment with 1001 Token at Mid Processing

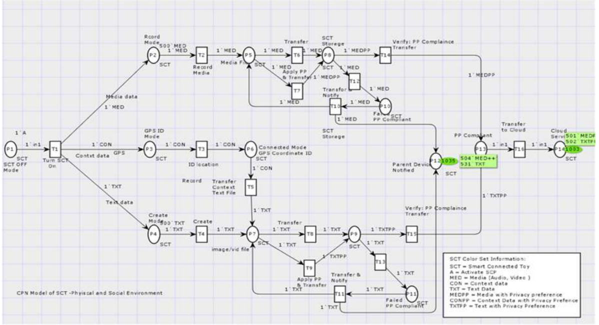

Figure 11: CPN Model of SCT – Physical and Social Environment with 1001 at Final Transitions

4. Conclusions and Future Works

Moving away from traditional system development modeling techniques like DFD, PA-DFD, PbD and adopting Petri-Nets as a modeling technique for privacy preservation in SCT system development is necessary in order to address privacy concerns. These privacy concerns can result in a breach of data, which can bring about catastrophic consequences- such as child abduction or death. In this paper, we discussed the SCT capacity to collect, process, and store PII, context data, or user-generated data. User-generated and context data are an increasing privacy landmine, which current SDLC popular modeling techniques such as DFD, PA-DFDs, or other concepts (such as PbD), do not provide adequate privacy elements for. Privacy elements are necessary to establish a degree of assurance, confidentiality, and Integrity. In evaluating traditional modeling, it is clear that DFD, PA-DFD, and PhD methodologies lack essential privacy requirements to successfully model the SCT privacy framework. Although the PbD concept presents as an effective way of addressing privacy, in principle, its implementation is subjective considering that there is no mechanism for enforcing it integration into system development. Similarly, existing literature research [7] confirms that DFD lacks specific elements to address privacy that can introduce core vulnerabilities for modeling the SCT system or data flow. Although PA-DFDs can tackle privacy of personal data from the earliest stages of information system design, it fails to perform formal verification due to lack of concrete semantics, and it may not be appropriate to guarantee privacy assurance in the SCTs. On the other hand, Petri-Nets, have the capacity to integrate privacy elements in the modeling of the system flow including a well-defined mathematical semantics for verification and an enforced privacy policy to provide a sound level assurance to sensitive user information. As a result, it is a far superior modeling technique to ensure autonomous privacy preservation for SCT.

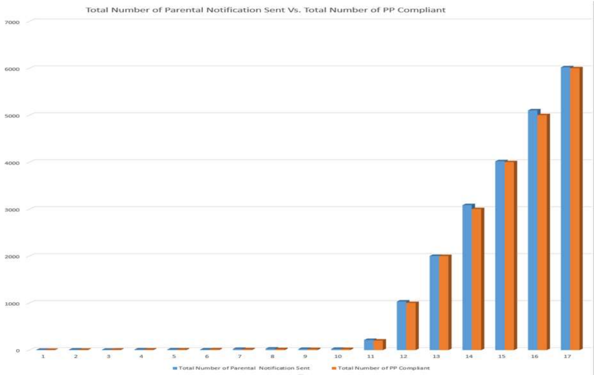

Figure 12: Total Number of Parental Notification Sent Vs. Total Number of PP Compliant

Future research will focus on Petri-Nets semantics by defining an appropriate transformation algorithm for DFD to PA-DFD, and PA-DFDs to CPN, which demonstrates the effectiveness of transformations such as an avenue for current SCT systems with full privacy assurance. Areas of interest include: (1) Construct SCT data transformation algorithm for conversion of existing SCTs PA-DFD models to CPN models. Emphasis will be on defining an algorithm by using pseudo code to transform a PA-DFD into a CPN model based on Definition 3 represented by the nine-purple (P,T,A,Σ,V,C,G,E,I). In conducting this transformation, part of what will be addressed is to parse the PA-DFD model, store information, define the color set for SCT data, and transformation of data flow. The new define color set of SCT data includes data privacy states where colset SCT = with Public | Anonymous | Private | Confidential | Unidentified | Aggregate|; and (2) Study SCT presented in DFD, PA-DFD, and CPN to demonstrate the core difference in handling privacy element essential. The sequence of the modeling will begin with modeling the DFD for the SCT with all privacy hotspot demonstrated. The next step will apply transformations on the identified hotspots in the DFD to obtain the PA-DFD. Finally, we will apply a transformation to the generated PA-DFD model to a CPN model.

5. References

[1] NIST. "Guide for Security-Focused Configuration Management of Information Systems". [Online]. Available: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-128.pdf

[2] L. B. Rafferty, "Toy Computing Background: Mobile Services for Toy Computing". Springer International Publishing, 2015. Page(s):10-37

[3] R. Ibrahim and S. Yen, "Formalization of the Data Flow Diagram Rules for Consistency Check", Int. J. of Software Eng. & Applicat, vol. 1, no. 4, pp. 95-111, 2010

[4] D. J. Weitzner, "Free Speech and Child Protection on the Web," IEEE Internet Computing, Volume: 11, Issue: 3, 2007, Page(s): 86 – 89.

[5] Dept. Of Justice . "Sex Offender Registry". [Online]. https://www.nsopw.gov/en/Registry. [Accessed: 12 July 2017].

[6] D. Salomon, "Privacy and Trust," Elements of Computer Security, Undergraduate Topics in Computer Science, Springer, 2010, Page(s): 273 – 290.

[7] M. Rahmana, M. "Petri Nets Semantics for Privacy-Aware Data Flow Diagram". CHALMERS UNIVERSITY OF TECHNOLOGY UNIVERSITY OF GOTHENBURG Gothenburg, Graduate Thesis, 2017. [Online]. Available: https://gupea.ub.gu.se/handle/2077/53077. [Accessed March. 2018]

[8] A. Dennis et al., System Anal. And Design, 5th ed. John Wiley & Sons, 2012

[9] A. Cavoukian, “Privacy by Design: The 7 Foundational Principles”. IAB. 2016.

[10] A. Cavoukian, “Strong Privacy Protection – Now, and Well into the Future”. 2017 [Online]. https://www.ipc.on.ca/wp-content/uploads/Resources/PbDReport.pdf. [Accessed August 5, 2017]

[11] Handbook of Dynamic System Modeling, Taylor & Francis Group, LLC, 2007.

[12] C. A. Petri, “Kommunikation mit Automaten.” Bonn: Institut fur lnstrumentelle Mathematik,

[13] Ling C. “The Petri Net Methods”. School of Computer Science & Software Engineering, Monash University. 2015

[14] R. Milner, M. Tofte, R. Harper and D. MacQueen, The definition of standard ML (revised). London: MIT Press, 1997

[15] T. Kosa, “Towards Measuring Privacy”. University of Ontario Institute of Technology. Thesis- Doctor of Philosophy. The Faculty of Science. 2015. [Online]. Available: https://ir.library.dc-uoit.ca/bitstream/10155/609/1/Kosa_Tracy%20Ann.pdf. [Accessed January 5, 2017]

[16] Q. Su, F. He, N. Wu and Z. Lin, "A Method for Construction of Software Protection Technology Application Sequence Based on Petri Net With Inhibitor Arcs," in IEEE Access, vol. 6, pp. 11988-12000, 2018.doi: 10.1109/ACCESS.2018.2812764.

[17] M. A. de Carvalho and P. Bandiera-Paiva, "Evaluating ISO 14441 privacy requirements on role-based access control (RBAC) restrict mode via Colored Petri Nets (CPN) modeling," 2017 International Carnahan Conference on Security Technology (ICCST), Madrid, 2017, pp. 1-8. doi:10.1109/CCST.2017.8167833.

[18] L. de Carvalho, and M. Eler, “Security Requirements for Smart Toy”. [Online] Scitepress.org. Available at: http://www.scitepress.org/Papers/2017/63370/63370.pdf [Accessed 1 May 2018].

[19]K. Jensen and L. Kristensen, Coloured Petri Nets. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009

[20] “CPN Tools homepage,” *CPN Tools*. [Online] Available: http://cpntools.org/2018/01/16/download/. [Accessed August 5, 2017]

[21] C. Liu, Q. Zeng, and H. Duan, “Formulating the Data-Flow Modeling and Verification for Workflow: A Petri Net based Approach”. International Journal of Science and Engineering Applications Volume 3 Issue 4, 2014, ISSN-2319-7560

[22] S. Sadiq, M. Orlowska, W. Sadiq, and C. Foulger. “Data flow and validation in workflow modelling”. In *Proceedings of the 15th Australasian database conference - Volume 27* (ADC '04), Klaus-Dieter Schewe and Hugh Williams (Eds.), Vol. 27. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 207-214.

[23] T. Murata, “Petri Nets: Properties, Analysis and Applications,” *Proceedings of the IEEE*, vol. 77, no. 4, pp. 541–580, April, 1989.

[24] W. M. P. van der Aalst, “The application of Petri nets to workflow management,” *Journal of Circuits, Systems and Computers*, vol. 8, no. 1, pp. 21–66, 1998.

[25] C. C. Dolean, and R. Petrusel, “Data- Flow Modeling: A Survey of Issues and Approaches”. December 2012, Informatica Economica; 2012, Vol. 16 Issue 4, p117.

5. Acknowledgment

The work of this paper is supported by National Natural Science Foundation of China under Grant No. 61728202- Research on Internet of Things Big Data Transmission and Processing Architecture based on Cloud-Fog Hybrid Computing Model, and Shanghai 2018 Innovation Action Plan- Hong Kong, Macao and Taiwan Science and technology cooperation project under Grant No. 18510760200- Research on smart city big data processing technology based on cloud-fog mixed mode.