User privacy risk analysis for the Internet of Things

User Privacy Risk Analysis For The Internet of Things

Akash Aggarwal[1], Waqar Asif[2], Habibul Azam[2], Milan Markovic[3], Muttukrishnan Rajarajan[2], and Peter Edwards[3]

[1]Department of Mathematical Sciences, Indian Institute of Technology (BHU), Varanasi, India 221005
[1]Email: akash.aggarwal.mat15@itbhu.ac.in
[2]School of Engineering and Mathematical Sciences, City, University of London, UK
[2]Email: waqar.asif@city.ac.uk, habibul.azam@city.ac.uk, r.muttukrishnan@city.ac.uk
[3]School of Natural and Computing Sciences, University of Aberdeen, UK
[3]Email: milan.markovic@abdn.ac.uk, p.edwards@abdn.ac.uk

***Abstract*—The Internet of Things (IoT) refers to a large network of devices such as sensors and actuators in which diverse types of data is generated and shared. Data can be shared in its raw form or as a result of data processing activities performed by an IoT device (e.g. anonymization, aggregation, etc.). However, sharing such data introduces a multitude of risks which are influenced by data type, data harvesting granularity, user demographics and the device under consideration. In this work, we propose a novel extension to our attack tree risk model [1] to consider user preferences for sharing personal data. We enrich our earlier work by exploring more attacks and complimenting them with a user privacy-risk model. We evaluate this proposed model and identify a range of scenarios which can result in personal information privacy violation and thus provide a model for estimating the potential risk of an IoT ecosystem.**

***Index Terms*—Privacy, Attack Cost, Attack Impact, Attack Attributes, Attack Scenarios.**

I. Introduction

A constantly growing number of physical devices are being connected to the internet at an unprecedented rate thus realizing the idea of the Internet of Things. According to an estimate, a mere 24 billion devices will be connected with each other by the end of year 2020 [8]. IoTs owe this huge growth to the large variety of devices that can communicate over the internet. This involves simple devices such as a Radio Frequency Identification (RFID) tags to more complex devices such as smart meters, smart TVs and smart Heating, Ventilation & Air-conditioning (HVAC) systems. This variety of devices brings with itself a vast range of security and privacy concerns. One of the most significant concerns associated with such an interconnected heterogeneous network is the loss of personal information.

IoT devices may gather highly granular data about an individual which once combined and analyzed, can reveal vital information. An RFID card reader, such as the one used by the Transport for London (TFL) authorities [10] for paperless travel ticketing, is deemed as a simple IoT device. The information generated by such a reader can be used to reveal travel habits of an individual thus becoming a huge privacy concern. Similarly, a smart meter which is advertised as a tool for sharing monthly energy consumption of a household for accurate energy billing can be used to identify living patterns [2]. Existing approaches that are used to mitigate privacy violations include the use of efficient data anonymization techniques. These include but are not limited to *k-anonymity*, *l-diversity*, *t-closeness* and ϵ-*differential privacy* [12]. Data anonymization approaches are known to work well in their considered scenarios but most of these assume having a tabulated form of incoming data. This is viable in the case of an RFID tag reader but is highly unsuitable for a smart meter, where data anonymization is done in real time. In an IoT context, an adversary can tweak the hardware of an IoT device forcing it to relay all information to a third party [4], thus rendering these data anonymization approaches ineffective.

The increasing number of incidences such as that of a smart TV reported in 2017 [13], or of a smart meter delineated in 2012 [9] has highlighted the need for a device user to analyze and access all possible risks associated to an IoT device before using it. This risk analysis has also become a necessity after the introduction of the General Data Protection Regulation (GDPR) [6] in the European Union, which focuses on providing a user with the right authority/control over his/her information by ensuring transparency in the system that is gathering their information. In this paper, we propose a new way of evaluating and identifying potential risks of an IoT system based on different user preferences, gathered data and probability of occurrence of an event. We extend our prior work on attack tree risk model [1], which identified a number of sources of risk and enhance it with the addition of a few new risks. We also couple it with a novel user privacy-risk model that encapsulates the influence of a user on the perceived risk of a device.

The remainder of the paper is organized as follows: Section II introduces the user privacy-risk model, Section III explains, expansion of the attack tree model and Section IV presents the results. Conclusions are presented in Section V.

II. User Privacy-Risk Model

An attack tree model is a hierarchical representation of a set of incidences that can be combined to launch an attack on an IoT infrastructure. The key goal of such a tree is the privacy violation of a user with leaf nodes represent the possible attacks. The path generating from the leaf nodes up to the goal of the tree is referred to as scenarios. A single tree can have a large number of scenarios based on the number of different possible paths. The risk then is a mere approximation of the relative harm a particular attack can cause in relation to all other possible scenarios [1]. The impact of an attack highly depends upon two metrics, the granularity of data being gathered and the type of a user profile.

Data granularity defines the frequency of the data being collected and the processing that is performed. This can thus lead to three broadly defined data types:

- **Raw data:** Data originating from a single IoT device that has not been altered by any additional post-processing.
- **Aggregated data:** Data representing some abstracted values calculated using raw data produced by a single IoT device, over some time period.
- **Pooled data:** An abstraction of aggregated data obtained from more than one IoT device.

On the other hand, the user profile is a bit treacherous. These profiles depend upon the demographics, general privacy beliefs and attitude towards an IoT ecosystem. The demographic characteristics include gender, education and age of an individual. The belief characteristic encompasses the privacy perception of an individual such as: privacy as a right guaranteed by law, privacy as an individual's responsibility and privacy as a need associated to people who are involved in wrong doing. Attitudes include how they react when they are supposed to provide their personal data to the system. A combination of these characteristics can be classified into three user profiles, namely *fundamentalists*, *pragmatists* and *unconcerned* [11].

The core user characteristics associated with such profiles can be summarized as follows:

- **Fundamentalists:** Users who are unwilling to provide their personal information even in return for the service enhancement. They perceive anything as a threat to their privacy.
- **Pragmatists:** Users who willingly take part in sharing some amount of personal information in exchange for better services. They weigh the risk between the information being offered to the services being rewarded.
- **Unconcerned:** Users who are not concerned with their data privacy and would be willing to share their information with anyone. They don't perceive any risk factor.

Users react in different way to ensure data privacy compliance. Where some mitigate privacy violation risk by adding random noise to their data set [7], other rely on slight manipulations to ensure utility in the remaining data [2]. The two widely used user-data privacy models are referred to as *Data Eradicator* and *Data Manipulator*. These models are defined as:

- **Data Eradicator:** A data eradicator module ensures that data generated by an IoT device is obfuscated with the addition of noise before it is published. This can be achieved with the aid of external hardware, for instance, smart meter data can be altered by randomly turning an HVAC system ON and OFF.
- **Data Manipulator:** The data manipulator module enables sharing of selected parts of the data and obfuscating the rest. This provides flexibility in selecting the amount of data that is shared in exchange for some benefits (e.g. discounts on utility bills when sharing smart meter data with the utility provider).

The user can control the level of data privacy with the help of a tuneable ϵ parameter. For example, considering the aforementioned smart meter scenario, the ϵ parameter may represent an Energy Management Unit (EMU) used to obfuscate raw energy readings shared by the smart meter [2]. Users can control the level of obfuscation by altering EMU settings based on their data sharing preferences.

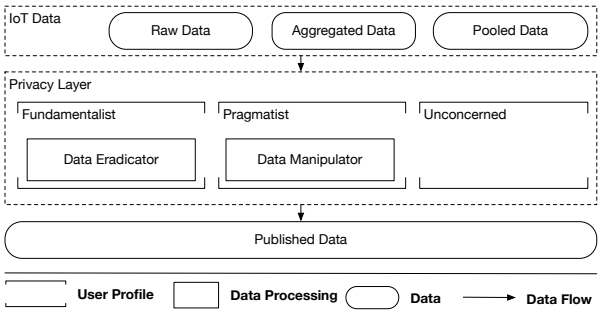


Fig. 1: A high level overview of the user privacy-risk model.

In our model represented in Fig 1, we consider a wide range of devices that can either produce, raw data, aggregated data or polled data. This data is then relayed to the privacy layer, where based on different user profiles, data is altered. For instance, a fundamentalist user would use the data eradicator to ensure user data privacy. On the contrary a user who prefers a data eradicator would be referred to as a fundamentalist user. Similarly, a pragmatist user would prefer a data manipulator to ensure a balance between utility and privacy of a data set. An unconcerned user would not alter any incoming data. The output of the privacy layer becomes the published data where the granularity of this data set is dependant upon the incoming data from the IoT device layer. The output of this user privacy-risk model is then fed into the attack tree to highlight the influence of an attack.

A. User Scenario

Similar to the attack tree scenario, a user scenario represents all possible combination of the IoT data layer, the privacy layer and the selection of ϵ for the data manipulator and/or data eradicator module. Each user scenario is evaluated for an entropy (α) measure where, α is a measure of the personal information that is being shared in an IoT ecosystem. This entropy lies between 0 and 1 ($0 \leq \alpha \leq 1$), where $\alpha = 0$ implies

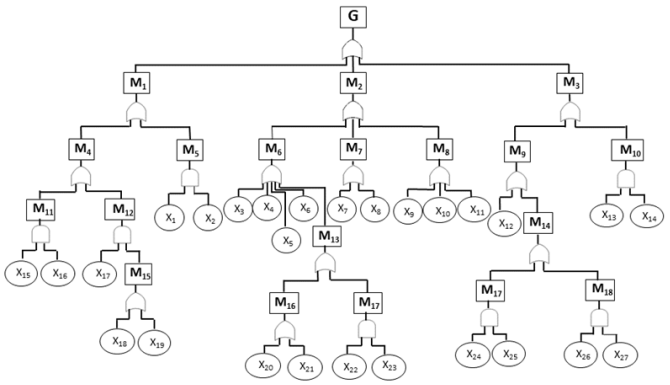


Fig. 2: Attack tree risk model.

minimal personal information is being shared and $\alpha = 1$ implies all personal information is being shared. In order to realize this and factor the aforementioned definition of raw data, aggregated data and pooled data, a set of utility functions are defined here. These functions highlight the associated risk that follows the pattern, *Raw data* $>$ *Aggregated data* $>$ *Pooled data*:

$$u_R(x, \epsilon) = 1 - 10^{-x\epsilon} \tag{1}$$

$$u_A(x, \epsilon) = 1 - exp^{-x\epsilon} \tag{2}$$

$$u_P(x, \epsilon) = 1 - (x + 1)^{-\epsilon} \tag{3}$$

Here $u_R(x, \epsilon)$, $u_A(x, \epsilon)$, $u_P(x, \epsilon)$ represents the utility function of raw data, average data and pooled data and ϵ represents the tuneable privacy metric. On the other hand x is a non-tuneable parameter that defines the percentage of raw data, aggregated data and pooled data that the sensor is harvesting.

In case of *fundamentalist users*, who are highly concerned about their privacy, the probability of selecting a data eradicator would be higher than selecting a user controllable privacy module. This means that restricted personal information would be shared by simply tuning ($\epsilon_F \longrightarrow 0$). On the other hand, a pragmatist user would use the user controllable privacy module and tune ($0 < \epsilon_P < 1$) with a lower value representing minimal information sharing and value of 1 representing maximal information sharing. The entropy factor associated with the fundamentalist users (α_F), pragmatist users (α_P) and unconcerned users (α_U) can be defined as:

$$\alpha_F = u_R(x, \epsilon_F) + u_A(x, \epsilon_F) + u_P(x, \epsilon_F) \tag{4}$$

$$\alpha_P = u_R(x, \epsilon_P) + u_A(x, \epsilon_p) + u_P(x, \epsilon_P) \tag{5}$$

$$\alpha_U = u_R(x, \epsilon_U) + u_A(x, \epsilon_U) + u_P(x, \epsilon_U) \tag{6}$$

The overall entropy (α) of the user privacy-risk model can then be defined as:.

$$\alpha = P(F) * \alpha_F + P(P) * \alpha_P + P(U) * \alpha_U \tag{7}$$

Here, $P(F)$, $P(P)$ and $P(U)$ represent the probability weights of respective user profiles in an IoT ecosystem.

III. Attack Tree Model

An attack tree model (Fig 2) is an analytical way of analyzing the potential risk of an IoT system. The model is represented as a decision tree with root node representing the attacker goal (i.e. to violate user privacy). The leaf nodes representing the possible attacks and the path from the leaf nodes to the attack goal defining the attack scenarios. The intermediary nodes are referred to as sub-goals and they are either a resultant of other sub-goals or a combination of leaf nodes. A detailed explanation of the attack tree model is defined in [1]. This section, makes use of our earlier work and compliments it with a two more attacks for the non-physical device (M_2) category, which are explained in detail in the next section.

A. Non-physical Device Attack

These attacks are centred on the IoT system where the attacker does not necessarily have physical access to the IoT device. These attacks can be divided into three categories: *Active Attack* (M_6), *Malware Attack* (M_7) and *Passive Attack* (M_8). In this section we only highlight the newly added sub-goal namely, Malware Attack (M_7) and the leaf node namely, Distributed Denial of Service attack (DDoS) (X_5).

An active attack is the one in which an adversary tries to penetrate the network by performing a set of operations either on a device itself or the communication channel. One such penetrating attack is referred to as the Distributed Denial of Service (DDoS) attack. In such an attack, an adversary can either flood the gateway node of the IoT network or the web-host to which it is connected to ensure that no or limited information is sent by the IoT device to the receiving end. This as a result would overload the IoT device with re-transmissions and thus cause loss of vital personal information. On the other hand, an adversary can also exploit the outdated security definition of an IoT device (X_7) or gain unauthorized access (X_8) to a device to launch a Malware Attack. Such an attack can cause multiple issues such as relaying personal information to a third party [3], forcing the IoT to produce faulty data or using the IoT device to launch a botnet attack on the remaining connected IoT devices [5].

B. Attack Scenario

An attack scenario defines a set of possible attacks that can be used in conjunction or in isolation to obtain personal information from an IoT system. These attack scenarios aid in identifying the potential attack path an adversary can choose to compromise a user's data privacy. In order to evaluate such scenarios each sub-goal is assigned a separate weight w_i and a probability m_i where i represents the number of sub-goals. Each leaf node is assigned a probability x_j where j represent the number of leaf nodes. An AND/OR gates structure is used in Fig 2, where OR-gate is the maximum of inputs and AND-gate is the multiplication of input. According to these gates functionality the probability g for the overall risk of an IoT ecosystem can be calculated using the following equations:

$$g = max(w_1 m_1, w_2 m_2, w_3 m_3) \tag{8}$$

On solving Eq 8, we get the overall risk probability g as:

$$\begin{aligned} g = max(&w_1w_4w_{11}x_{15}x_{16}, w_1w_4w_{12}x_{17}w_{15}x_{18}, \\ &w_1w_4w_{12}x_{17}w_{15}x_{19}, w_1w_5x_1x_2, w_2w_6x_3, w_2w_6x_4, \\ &w_2w_6x_5, w_2w_6x_6, w_2w_6w_{13}w_{16}x_{20}, \\ &w_2w_6w_{13}w_{16}x_{21}, w_2w_6w_{13}w_{17}x_{22}x_{23}, w_2w_7x_7, \\ &w_2w_7x_8, w_2w_8x_9, w_2w_8x_{10}, w_2w_8x_{11}, w_3w_9x_{12} \\ &w_3w_9w_{14}w_{18}x_{24}x_{25}, w_3w_9w_{14}w_{19}x_{26}x_{27}, \\ &w_3w_{10}x_{13}x_{14}) \end{aligned} \tag{9}$$

Eq. 9 identifies 20 attack scenarios that can result in a potential user data/identity privacy risk. These attacks are enumerated in Table I.

TABLE I: Probabilities for possible data breach scenarios

Scenarios	Leaf Nodes	Probability
S_1	x_{15}, x_{16}	$w_1w_4w_{11}x_{15}x_{16}$
S_2	x_{17}, x_{18}	$w_1w_4w_{12}x_{17}w_{15}x_{18}$
S_3	x_{17}, x_{19}	$w_1w_4w_{12}x_{17}w_{15}x_{19}$
S_4	x_1, x_2	$w_1w_5x_1x_2$
S_5	x_3	$w_2w_6x_3$
S_6	x_4	$w_2w_6x_4$
S_7	x_5	$w_2w_6x_5$
S_8	x_6	$w_2w_6x_6$
S_9	x_{20}	$w_2w_6w_{13}w_{16}x_{20}$
S_{10}	x_{21}	$w_2w_6w_{13}w_{16}x_{21}$
S_{11}	x_{22}, x_{23}	$w_2w_6w_{13}w_{17}x_{22}x_{23}$
S_{12}	x_7	$w_2w_7x_7$
S_{13}	x_8	$w_2w_7x_8$
S_{14}	x_9	$w_2w_8x_9$
S_{15}	x_{10}	$w_2w_8x_{10}$
S_{16}	x_{11}	$w_2w_8x_{11}$
S_{17}	x_{12}	$w_3w_9x_{12}$
S_{18}	x_{24}, x_{25}	$w_3w_9w_{14}w_{18}x_{24}x_{25}$
S_{19}	x_{26}, x_{27}	$w_3w_9w_{14}w_{19}x_{26}x_{27}$
S_{20}	x_{13}, x_{14}	$w_3w_{10}x_{13}x_{14}$

All these attack scenarios highlighted in Table I will have different attack impact on the system. There is no single method to determine the weight value for each node in the attack tree. So in this work we use a quadruple attack grade standard to find risk associated with attacks. These include: Attack Impact I_L (estimates the number of affected nodes due to an attack), Attack Cost c_L (reflects the costs associated with an attack), Technical Difficulty d_L (estimates the difficulty associated with an attack) and the Probability to be Discovered s_L.

TABLE II: Grade standard

Attack Impact/device		Attack Cost/ten thousands		Technical Difficulty		Probability to be Discovered	
I_L	grade	c_L	grade	d_L	grade	s_L	grade
Individual	2	>10	5	quite difficult	5	quite difficult	1
Multiple	1	6-10	4	difficult	4	difficult	2
		3-6	3	intermediate	3	intermediate	3
		0.5-3	2	simple	2	simple	4
		<0.5	1	quite simple	1	quite simple	5

Table II highlights the grade level standards of the attributes. After value assignment for leaf nodes, the combined influence of the user and these four attributes associated to a single leaf is formally defined as:

$$r_L = a_1 * u_1(i_L) + a_2 * u_2(c_L) + a_3 * u_3(d_L) + a_4 * u_4(s_L) \tag{10}$$

Where, a_1, a_2, a_3, a_4 are the corresponding attribute weights of i_L, c_L, d_L and s_L respectively. Here $a_1 + a_2 + a_3 + a_4 = 1$ and $u_i(x)$ represent the utility function corresponding to each attribute $i = 1, 2, 3, 4$. For the sake of simplicity, in this work, all four attributes use the same utility function:

$$u_i(x) = \frac{c}{(x+1)^{1/\alpha}} \tag{11}$$

This utility function is based on the observation that all four attributes are inversely proportional to their grade values and the entropy of the user privacy-risk model would directly influence the attack attributes of a network. For instance, a sensor gathering average values, owned by a *fundamentalist user* would pass through a *data eradicator* and would generate *aggregated data* which would have lower entropy factor α_F. This as a result would directly influence the impact of an attack as an adversary would be able to gain limited insight into personal information of the user thus reducing user privacy risk r_L.

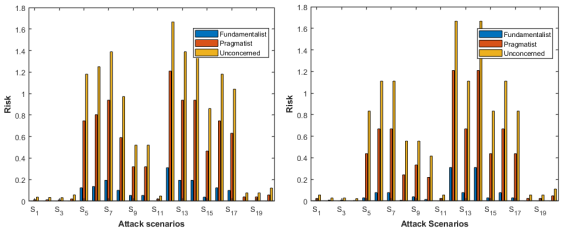
TABLE III: Attack attributes and grades

Leaf node	Attack impact	Attack cost	Technical difficulty	Probability to be discovered
X_1	2	5	5	1
X_2	2	4	2	2
X_3	1	3	2	2
X_4	2	2	1	2
X_5	1	2	2	1
X_6	1	5	5	2
X_7	1	1	1	1
X_8	1	2	2	1
X_9	2	1	2	1
X_{10}	2	2	2	1
X_{11}	1	2	3	2
X_{12}	1	3	3	3
X_{13}	1	1	1	3
X_{14}	1	2	2	1
X_{15}	2	1	1	4
X_{16}	2	2	3	3
X_{17}	2	3	3	1
X_{18}	2	2	3	2
X_{19}	2	2	3	3
X_{20}	1	2	3	5
X_{21}	1	3	2	5
X_{22}	2	1	2	1
X_{23}	1	2	4	2
X_{24}	1	1	1	1
X_{25}	1	2	1	1
X_{26}	1	1	1	1
X_{27}	1	2	1	1

IV. RESULTS

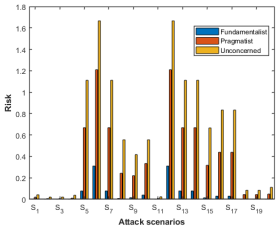
In this section, we report results of simulation experiments testing the ability of our user privacy model to estimate privacy risks associated with an IoT system. Privacy risks depend upon the grades associated with each leaf node of Fig 2. These grade values depend upon the system dynamics and changes based on the selection of an IoT device. In this work, the belief and plausibility metrics proposed by Dampster and Shafer [14] are exploited to identify the risks associated with each leaf node. These grades are presented in Table III.

The first set of simulations highlights the change in probability of risk for each attack scenario for a particular user profile. The results highlight the difference in risk when different attack attributes are selected. We assign weights to each sub goal of the attack tree as $w_1 = w_2 = w_3 = 1/3$, $w_4 = w_5 = w_6 = w_7 = w8 = w_9 = w_{10} = w_{11} = w_{15} = w_{16} = w_{17} = w_{18} = w_{19} = 1/2$ and give the rest a weight of 1 for reducing the influence of unknown sub-goal probabilities. We assign weights for the user tree model as $\epsilon_F = 0.01$,

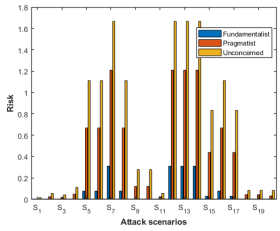


(a) $a_1 = 1, a_2 = 0, a_3 = 0, a_4 = 0$

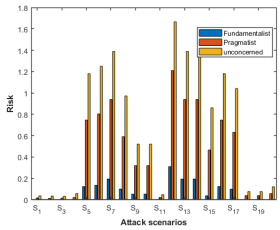
(b) $a_1 = 0, a_2 = 1, a_3 = 0, a_4 = 0$



(c) $a_1 = 0, a_2 = 0, a_3 = 1, a_4 = 0$

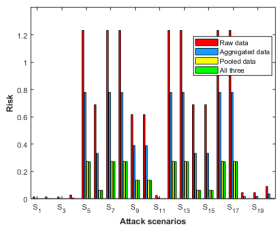


(d) $a_1 = 0, a_2 = 0, a_3 = 0, a_4 = 1$

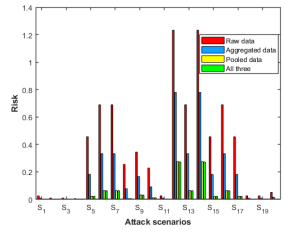


(e) $a_1 = a_2 = a_3 = a_4 = 0.25$

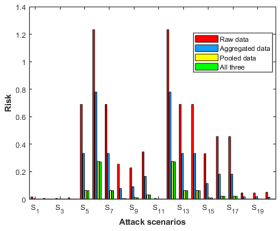
Fig. 3: Attack risk of a particular user profile for varying attack attribute selections.



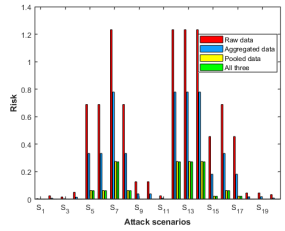
(a) $a_1 = 1, a_2 = 0, a_3 = 0, a_4 = 0$



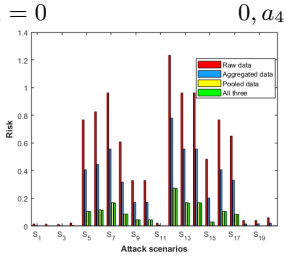
(b) $a_1 = 0, a_2 = 1, a_3 = 0, a_4 = 0$



(c) $a_1 = 0, a_2 = 0, a_3 = 1, a_4 = 0$



(d) $a_1 = 0, a_2 = 0, a_3 = 0, a_4 = 1$



(e) $a_1 = a_2 = a_3 = a_4 = 0.25$

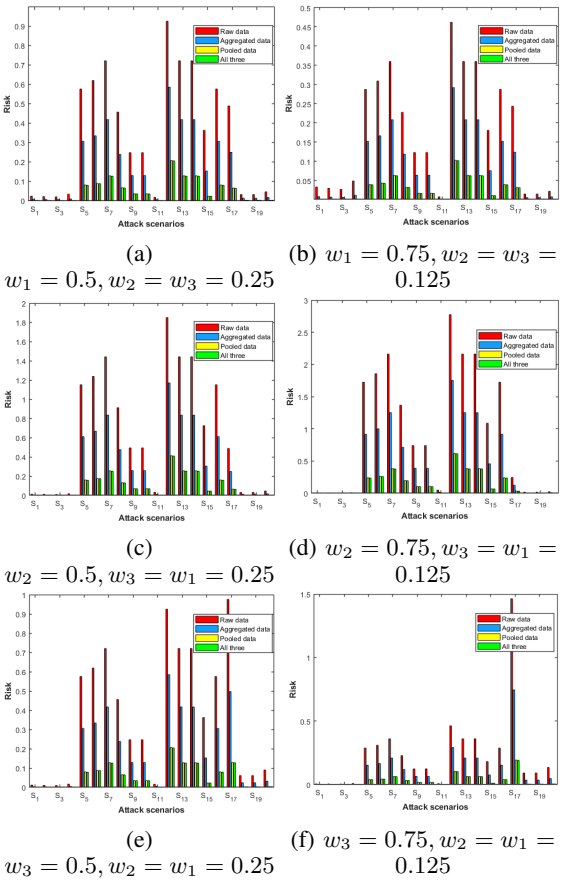
Fig. 4: Attack risk of a user for different data type by varying attack attribute selections.

$\epsilon_P = 0.5$, $\epsilon_U = 1$ where F, P and U are a subscript for a particular user profile. Fig 3 represents the attack risk of a particular user profile for varying attack attribute selections.

Fig 3 highlights the risk pattern among user profiles as: unconcerned $>$ pragmatist $>$ fundamentalist. This implies unconcerned users would have the highest attack risks as compared to other user profiles. The overall risk factor is dependent upon the utility function of the attack attributes and their corresponding utility weights. The utility function (Eq 11) is dependent upon the $\alpha - factor$ and the grade value of a particular attack attribute. Results demonstrate that the α is dependent upon the user profile and it is highest for an unconcerned user. Moreover, Table II and III justify these results by highlighting that for any attack if the attack attributes have low grade standards then it would have higher risk compared to others. So for a higher α value the overall risk will be relatively high. This leads to the hypothesis that unconcerned users would have the highest risk compared to others thus proving the correctness of the proposed model. For instance: a fundamentalist user would have low risk due to eradicated data and a pragmatist user would have control over the data shared within the IoT ecosystem. However, an unconcerned user would share the data without any level of data distortion. As a result, the risk value for an attack increases.

Fig 3a shows that when attack impact is of more significance a user would have higher risk of losing personal data through S_5, S_7, S_8, S_{12}, S_{13}, S_{16} and S_{17}. On the other hand if attack cost is a concern then S_{12} and S_{14} take precedence over other possible risks as shown in Fig 3b. S_6 and S_{12} are of higher importance when technical difficulty is of concern and when probability to be discovered is of high importance as shown in Fig 3c and Fig 3d respectively. Furthermore, when all attack attributes are given equal weight, S_{12} takes precedence over all other attacks and is followed by S_5, S_6, S_7, S_{13} and S_{14} as shown in Fig 3e. This means in order to ensure data privacy a user needs to take preemptive measures to avoid side channel attack (S_5), inquiry attack (S_6), DDoS attack (S_7) and needs to ensure that security definitions are up to date so that unauthorized access (S_{13}) and physical layer eaves dropping (S_{14}) can be avoided.

In the second set of simulations we highlight the overall risk to a user when raw data, aggregated data and pooled data are shared. The results highlight the change in risk after different attack attributes were selected. We assigned the similar weights to each sub-goal of the attack tree and ϵ value for each particular user as in previous simulation. Here we incorporate outcomes from the survey carried out in [11] for identifying



(a) $w_1 = 0.5, w_2 = w_3 = 0.25$

(b) $w_1 = 0.75, w_2 = w_3 = 0.125$

(c) $w_2 = 0.5, w_3 = w_1 = 0.25$

(d) $w_2 = 0.75, w_3 = w_1 = 0.125$

(e) $w_3 = 0.5, w_2 = w_1 = 0.25$

(f) $w_3 = 0.75, w_2 = w_1 = 0.125$

Fig. 5: Attack risk of a user for different data type by varying branch weights.

an approximate percentage of fundamentalist, pragmatist and unconcerned users in a group of individuals. The outcome of the survey define probability values as: $P(F) = 0.12$, $P(P) = 0.69$ and $P(U) = 0.19$. These are then used to calculate the overall entropy factor (α) as discussed earlier in Eq 7.

Fig 4a highlights that when attack impact is of high importance S_5, S_7, S_8, S_{12}, S_{13}, S_{16} and S_{17} take precedence over all other risks and among these raw data is more prone to risk compared to aggregated and pooled data. A key highlight here is that when all three components of IoT data layer are combined together the overall risk is similar to that of pooled data. In case of attack cost S_{12} and S_{14} causes of higher risk whereas, in case of technical difficulty S_6 and S_{12} are of a concern as shown in Fig. 4b and Fig. 4c respectively. Fig. 4d shows that S_7, S_{12}, S_{13} and S_{14} pose a high risk whereas Fig. 4e shows that S_5, S_6, S_7, S_{12}, S_{13}, S_{14}, S_{16} and S_{17} are a big concern. This means that alongside side channel attack (S_5), DDoS attack (S_7) and inquiry attack (S_6) the user also needs to take care of security definition, unauthorized access, eaves dropping and privacy policy.

In the final set of analysis, the evaluation concentrates on the overall risk influenced by a user sharing only raw data, aggregated data, pooled data and all three in equal ratio while the weights of sub-goals (M_1, M_2, and M_3) are varied. Each attack attribute is assigned with the same weight as follows: $a_1 = a_2 = a_3 = a_4 = 0.25$. In each chart, a specific sub-goal is out weighed and the remaining two sub-goals are assigned with a same weight such that $w_1 + w_2 + w_3 = 1$ and the similar probability weights of user profile and other sub-goals weights are used as those in the previous simulation.

Fig 5a, Fig 5b, Fig 5c and Fig 5d shows that changing the weight of physical device attack (M_1) and non-physical attack (M_2) has little to no influence on the risk attack scenarios. S_{12} has precedence over all attack risk with S_5, S_6, S_7, S_{13}, S_{14}, S_{15} and S_{16} having slightly lower risk. A major change is observed when data storage attack (M_3) is given importance where a slightly higher value of $w_3 = 0.5$ and $w_1 = w_2 = 0.25$ reports a higher risk of S_{12} and S_{17} as shown in Fig 5e. On the other hand S_{17} becomes the biggest concern when $w_3 = 0.75$ and $w_1 = w_2 = 0.125$. This means that purchasing privacy from a third party organization (S_{17}) can pose a high privacy risk.

V. Conclusion

In this work we present a novel way of estimating attack risks associated with IoT deployments by exploring user preferences, data harvesting granularity of a device, attackers preferences and the plausibility of an attack. We highlight what attack could result in a higher probability of personal data loss and thus enable a user to take preemptive measures based on his/her scenario. We run simulation experiments to justify our scenarios and provide a plausible argument for each outcome.

References

[1] W. Asif, I. Ghosh Ray, and M. Rajarajan. An attack tree based risk evaluation approach for the internet of things. pages 1–8, 10 2018.

[2] W. Asif, M. Rajarajan, and M. Lestas. Increasing user controllability on device specific privacy in the internet of things. *Computer Communications*, 116:200 – 211, 2018.

[3] A. Aziz. Prospective client identification using malware attack detection, May 5 2015. US Patent 9,027,135.

[4] C. Bekara. Security issues and challenges for the iot-based smart grid. *Procedia Computer Science*, 34:532–537, 2014.

[5] E. Bertino and N. Islam. Botnets and internet of things security. *Computer*, (2):76–79, 2017.

[6] E. GDPR. General data protection regulation. https://eugdpr.org/.

[7] G. Kalogridis, R. Cepeda, S. Z. Denic, T. Lewis, and C. Efthymiou. Elecprivacy: Evaluating the privacy protection of electricity management algorithms. *IEEE Transactions on Smart Grid*, 2(4):750–758, 2011.

[8] A. Majeed. Internet of things (iot): A verification framework. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 1–3. IEEE, 2017.

[9] B. News. Warning over smart meters privacy risk. https://www.bbc.co.uk/news/technology-18407340, June 2012.

[10] Oyster. Transport for london - oyster cards. https://oyster.tfl.gov.uk/oyster/entry.do.

[11] L. Ponciano, P. Barbosa, F. Brasileiro, A. Brito, and N. Andrade. Designing for pragmatists and fundamentalists: Privacy concerns and attitudes on the internet of things. 08 2017.

[12] D. Rebollo-Monedero, J. Parra-Arnau, C. Diaz, and J. Forné. On the measurement of privacy as an attackers estimation error. *International journal of information security*, 12(2):129–149, 2013.

[13] J. Saarinen. Smart tv vendor penalised for massive privacy violation. https://www.itnews.com.au/news/smart-tv-vendor-penalised-for-massive-privacy-violation-450210, February 2017.

[14] R. R. Yager. On the dempster-shafer framework and new combination rules. *Information sciences*, 41(2):93–137, 1987.