Internet of Things: Securing the Identity by Analyzing Ecosystem Models of Devices and Organizations

Kai Chih Chang
Razieh Nokhbeh Zaeem
K. Suzanne Barber

April 2019
UTCID Report #19-02
Sponsored By

HID

gemalto
security to be free

GENERALE
GLOBAL ASSISTANCE

IDENTITY PROTECTION SERVICES

IDEMIA
augmented identity

LifeLock
Guarantee Your Good Name

TransUnion
Internet of Things: Securing the Identity by Analyzing Ecosystem Models of Devices and Organizations

Kai Chih Chang and Razieh Nokhbeh Zaeem and K. Suzanne Barber
Center for Identity
The University of Texas at Austin
{kaichih, razieh, sbarber}@identity.utexas.edu

Abstract

The Internet of Things has become an integral part of our daily life. Its combination of network and emerging technology interlaced with each other results in a complicated environment that is left to us to understand and interact with. Information travels in the cyber world, not only bringing us convenience and prosperity but also jeopardy. Protecting this information has been an issue and commonly discussed in recent years. One type of this information is Personally Identifiable Information (PII), often used to perform personal authentication. With total cost of more than $40 billion since 2006, several reports of theft and fraudulent use of PII have been released. An all-embracing technique and system is needed in order to protect users from identity theft. In this paper, we present the Identity Ecosystem, a comprehensive identity framework that contains a mathematical representation of a model of Personally Identifiable Information attributes for people, and two novel models, devices and organizations, that have strong connections with the PII model of people. This research aims to combine the above three models and leads to better prevention against identity theft and fraudsters.

Data protection has been a problem since the network began to evolve. With the commercialization of the Internet, security issues have been extended to cover personal privacy, financial transactions and cyber-theft threats. In the paradigm of the Internet of Things, security and safety are inseparable. Whether it is accidental or malicious, interfering with personal mobile phones, hacking into the computers of an organization, and other similar acts pose a threat to human privacy, property, and even life. Even arbitrary data, like a temperature, might be related to a user when it is combined with other data like location or is profiled over a period of time. Privacy becomes crucial in the Internet of Things. How to protect the privacy of individuals, that is, to safeguard this identity information to prevent identity theft, has become one of the mainstream topics discussed today. Federal Trade Commission (FTC) has estimated an annual loss of over 15 billion dollars from identity theft in 2006 (Synovate 2007). In 2010 this figure had more than doubled, as 8.1 million U.S. adults were the victims of identity theft or fraud, with total costs of $37 billion (Miceli and Vamosi 2011). Identity theft, according to the National Institute of Justice, has become a prime crime in the information age, with an estimated 9 million or more incidents each year (Newman and McNally 2005). Identity theft threatens our safety and property, unless we can truly prevent fraudsters from identity breaching.

In this paper, we seek to discuss the identity theft issues most relevant to people, businesses, and devices. The first aspect of identity is the one that identifies people, or Personally Identifiable Information (PII). The IoT world would benefit from this kind of identifiers. However, there is no special identifier in the IoT world and there will never be one in the near future. For example, public classic IP-addresses (IPv4 addresses) are a rare resource. Access providers use IP-address pools and “re-use” IP-addresses by dynamic assignment, which means that with every mobile phone login, the mobile client might be assigned to an IP-address different from the one that was assigned from last login (Friese, Heuer, and Kong 2014). Our understanding of this personally identifiable information is not enough.

In the pursuit of security, this information needs to be understood and valued. Being merged with online attributes and offline attributes, the cyber world has been assimilated into people’s everyday world. Online attributes are composed of one’s social media accounts, online shopping patterns, passwords, email accounts and so on. Offline attributes are those related to the physical world such as bank accounts, credit and debit cards, social security number, fingerprint, blood type, etc. A more comprehensive online identity framework is needed based on a sound understanding of PII (Liang 2014).

The current Identity Ecosystem is limited to a single general model that hypothesizes only individuals have PII. But in fact, a mobile phone tracks its owner’s current location. A laptop stores plenty of one’s private information. Even one’s sports watch or e-health equipment are transmitting his/her body status such as body temperature and heart rate. This information travels in the cyber world through the Internet. Eventually, it flows into the server of a company or an organization. A security incident at that organization may expose personal information that belongs to a large number of people and result in monetary loss. Taking the above scenario into consideration, in this paper, we introduce two extra models: devices and organizations. Only by combining the graphic model of people, devices, and organizations will we obtain comprehensive knowledge of the operation of PII in the cyber world.

In the following section, we briefly introduce how Ecosystem works, and then introduce our two models. Then we discuss our data resources. Finally, we present the conclusion and proposed future work.

Ecosystem Models

As mentioned in section 1, the Identity Ecosystem developed at the Center for Identity at the University of Texas at Austin has constructed a graph-based model of people. It provides a statistical framework for understanding the value, risk and mutual relationships of personally identifiable information attributes. It uses a Network Model to simulate the relationships among PIIs for individuals. It allows predictions in the presence of interventions and it is able to handle incomplete data sets. It is visualized in a 3D graphic model and can be moved and rotated. The Ecosystem allows the users to choose a node property, such as value or risk, to determine node sizes and colors in the 3D graphic model (Nokhbeh Zaeem et al. 2016). Figure 1 shows the graph visualized in Ecosystem. Three interesting questions that Ecosystem can answer are inferring probability of breach based on evidence, detecting most probable origin of the breach and finding breach hot-spots.

- Effect of exposure: Assuming a set of attributes is exposed, the Bayesian inference model of Ecosystem calculates the change in the probability of exposure of other attributes. The Ecosystem can also show the predicted expected loss of the set of attributes compromised. Figure 2 shows how the probability of breach for other attributes changes, once the Social Security Number and Social Security Card attributes have been breached. Multiple attributes can be selected as evidence at the same time. It also shows potential loss after such a breach scenario.
- Cause: If an individual finds out that his/her PII is compromised, the Ecosystem can help to detect the most probable origin of the breach through selecting identity information as the evidence and running the query.
- Cost/Liability: The Ecosystem can calculate attributes which have the highest cost (breach hot-spots) and should be best protected.

So far, Ecosystem can answer these questions for the model of people’s PII. In this section, we introduce two novel models that have a strong connection with the PII graph of people: devices and organizations. In an IoT world there will exist a vast amount of raw data being continuously collected. It will be necessary to develop techniques that convert this raw data into usable knowledge (Stankovic 2014). The identity data would be one of these types of data. We define a person’s identity as a set of information that are linked to the person. The identity data not only exist for people, but are also extended into our mobile phones, vehicles, online applications, and so on. Hence, it is important to build the concept of identity for our devices.
Recently the concept of “Smart City” has rapidly risen (Dohler et al. 2011). Smart Cities consists of smart phones, mobile devices, sensors, embedded systems, smart environments, smart meters, and instrumentation sustaining the intelligence of cities (Schaffers et al. 2011). As a result, the relationship between people and devices has become blissfully tight. From mobile phones and laptops to GPS, sports watches and even to baby monitors, technical devices are collecting our PII anytime and anywhere.

We constructed a list of PII of items according to devices’ characteristic, function, affordances and other documents (Gubbi et al. 2013) (see Table 1). Then we endeavored to manually find the links between these nodes. As a result, we generated a model graph of devices’ PII. Figure 3 is a snapshot of the device graph presented by Ecosystem. Its main point lies in the links to the person’s PII graph.

In fact, it is not uncommon to see the relationships between devices and people in our daily lives. The IP and MAC address and the vehicle’s GPS imply one’s location. Plenty of personal information have been stored in applications in one’s mobile phone and computers. Moreover, sports and health devices are collecting one’s body temperature and heart rates. Recent advances in mobile technology and cloud computing have inspired numerous designs of cloud-based health care services and devices. Within the cloud system, medical data can be collected and transmitted automatically to medical professionals from anywhere and feedback can be returned to patients through the network (Deshpande and Kulkarni 2017). This progress presages the growing convenience of collecting PII through devices, while it concerns many with respect to privacy of personal information.

<table>
<thead>
<tr>
<th>Administrator</th>
<th>AdministratorPassword</th>
<th>AdministratorUserID</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApplicationType</td>
<td>ApplicationVendor</td>
<td>BusType</td>
<td>Cache</td>
</tr>
<tr>
<td>CircuitDesign</td>
<td>Color</td>
<td>CookieWipe</td>
<td>GeolocationStreetAddress</td>
</tr>
<tr>
<td>GeolocationZipCode</td>
<td>InventoryTag</td>
<td>IPAddresses</td>
<td>MACAddress</td>
</tr>
<tr>
<td>ManufactureLocation</td>
<td>Manufacturer</td>
<td>MemorySize</td>
<td>MemoryType</td>
</tr>
<tr>
<td>ModelNo</td>
<td>NetworkCards</td>
<td>NetworkConnectionSpeed</td>
<td>NetworkConnectionType</td>
</tr>
<tr>
<td>NetworkProxySettings</td>
<td>NumberOfAssociateEmails</td>
<td>NumberOfPorts</td>
<td>NumberOfTransactions</td>
</tr>
<tr>
<td>OpenDeviceIdNumber</td>
<td>OperatingSystemType</td>
<td>OperatingSystemVendor</td>
<td>OrganizationalLocation</td>
</tr>
<tr>
<td>Owner</td>
<td>PortNumbers</td>
<td>PowerFrequency</td>
<td>PowerUsage</td>
</tr>
<tr>
<td>ProcessorType</td>
<td>RegistryProperties</td>
<td>Reputation</td>
<td>SerialNo</td>
</tr>
<tr>
<td>ShipDate</td>
<td>SwitchingRate</td>
<td>TimeZone</td>
<td>TransactionFlags</td>
</tr>
<tr>
<td>TransactionsByCountry</td>
<td>TransactionsProfile</td>
<td>TransactionVolCountry</td>
<td>TransactionVolumeTotal</td>
</tr>
<tr>
<td>UniqueDeviceIdentifier</td>
<td>Users</td>
<td>Watermark</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: List of all nodes of devices

**Devices**

Recently the concept of “Smart City” has rapidly risen (Dohler et al. 2011). Smart Cities consists of smart phones, mobile devices, sensors, embedded systems, smart environments, smart meters, and instrumentation sustaining the intelligence of cities (Schaffers et al. 2011). As a result, the relationship between people and devices has become blissfully tight. From mobile phones and laptops to GPS, sports watches and even to baby monitors, technical devices are collecting our PII anytime and anywhere.

We constructed a list of PII of items according to devices’ characteristic, function, affordances and other documents (Gubbi et al. 2013) (see Table 1). Then we endeavored to manually find the links between these nodes. As a result, we generated a model graph of devices’ PII. Figure 3 is a snapshot of the device graph presented by Ecosystem. Its main point lies in the links to the person’s PII graph.

In fact, it is not uncommon to see the relationships between devices and people in our daily lives. The IP and MAC address and the vehicle’s GPS imply one’s location. Plenty of personal information have been stored in applications in one’s mobile phone and computers. Moreover, sports and health devices are collecting one’s body temperature and heart rates. Recent advances in mobile technology and cloud computing have inspired numerous designs of cloud-based health care services and devices. Within the cloud system, medical data can be collected and transmitted automatically to medical professionals from anywhere and feedback can be returned to patients through the network (Deshpande and Kulkarni 2017). This progress presages the growing convenience of collecting PII through devices, while it concerns many with respect to privacy of personal information.

**Organizations**

We are also interested in the relationship between people and organizations since activities that people trigger or
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Table 2: List of all nodes of organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>401KAdministrator</td>
<td>AccessCards</td>
</tr>
<tr>
<td>Attorney</td>
<td>BalanceSheet</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>BetterBusinessBureauRec</td>
</tr>
<tr>
<td>BusinessType</td>
<td>Buyer</td>
</tr>
<tr>
<td>CreditCardNumber</td>
<td>CreditCards</td>
</tr>
<tr>
<td>Customers</td>
<td>DateEstablished</td>
</tr>
<tr>
<td>Employees</td>
<td>FacebookAccount</td>
</tr>
<tr>
<td>InStorePurchasingPatterns</td>
<td>Investors</td>
</tr>
<tr>
<td>License</td>
<td>LoanNumber</td>
</tr>
<tr>
<td>LoyaltyCards</td>
<td>Name</td>
</tr>
<tr>
<td>OperatingSystem</td>
<td>Owner</td>
</tr>
<tr>
<td>PLStatement</td>
<td>PurchasingPatterns</td>
</tr>
<tr>
<td>StockExchTickerSymbol</td>
<td>Stockholders</td>
</tr>
<tr>
<td>VendorAddress</td>
<td>VendorName</td>
</tr>
<tr>
<td>WorkforceCommissionID</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: List of all nodes of organizations

Figure 4: The model of organizations shown in Ecosystem.

be part of everyday are related to companies and organizations. Identity data breach through organization is now a widespread problem around the globe. A security incident at an organization may expose personal information that belongs to a large number of people. The goal here is to construct a graph-based model of organization PII attributes and analyze its linkages to the people PII graph in order to help the Identity Ecosystem's investigation.

The most fundamental PII of organizations is people (see Table 2). Employees, officers, supervisors, board of directors, and even CEOs are integral parts of the model. They have the ability to access most machines in the company or factory which store most customers’ information. Hence, any information that is related to the machines would be treated as an organization attribute. We have also focused on documents that organizations would use in various situations by investigating the Certification of Formation from Texas Secretary of State\(^2\). Figure 4 is a snapshot of the organization graph presented by Ecosystem.

Community websites spread rapidly, not to mention the shopping websites. Every time one applies for a membership, he/she gives personal information to the organization that owns the website. Once the data has been received, the organization has the duty to keep these PIIs safe. However, breaches happen everywhere. Through the servers of an organization, customers’ banking accounts could be exposed and misused by others. It is through these means

\(^2\)“Texas Secretary of State”, Rolando B. Pablos. [https://www.sos.state.tx.us](https://www.sos.state.tx.us) (accessed August, 2017).
Data Sources

Modeling Identity Attributes (Nodes)
The Ecosystem distinguishes various properties of identity attributes. Take attribute’s type for instance; we divided the attribute’s type for a person into four categories: What You Are, What You Have, What You Know, and What You Do. In our previous work (Nokhbeh Zaeem et al. 2016), we briefly introduced every property of identity attributes in detail. Here we only introduce the way we came up with nodes for devices and organization using this classification. We also refer to a list of documents from Texas Secretary of State\(^2\) and (Gubbi et al. 2013) in our methodology.

What You Are For a person, it means a person’s physical characteristics, such as fingerprints and retinas. For a device, it means the type of a device. It can be a laptop, a smart watch, a sensor, and so on. It is also related to a device’s hardware configuration, such as circuit design and power usage. For an organization, it can also be its type. Also, it can be an organization’s icon, such as stock market icon.

What You Have For a person, it means credentials and numbers assigned to the person by other entities. For a device, it can be its model numbers, serial numbers, and inventory tags. For an organization, it can be its sales tax number and DUNS number.

What You Know For a person, it means information known privately to the person, such as passwords. For a device and an organization, it means any information that is stored in them. So all information stored in an app or customer information stored in a server of an organization are all related to this type.

What You Do For a person, it means a person’s behavior and action patterns, such as GPS location. What a device can do is often related to its application type, but for an organization of an online shopping website, it can be its online shopping pattern.

Modeling Identity Relationships (Edges)
The Ecosystem displays each attribute as a node. These nodes are related to each other in many different ways. The Ecosystem displays each relationship as an edge. We divided the type of relationships between a person’s PII into 7 categories (Nokhbeh Zaeem et al. 2016). According to this classification, we are able to assign edges between nodes for devices and organizations.

Breeds α Breeds β means that an instance/value of α may be used in order to create an instance/value of β. For example, a driver’s license breeds a boarding pass. A publication in an organization breeds its patent.

Composed Of α Composed Of β means that for any value \(\alpha_i\) of the attribute α there is a value \(\beta_j\) of the attribute β such that \(\beta_j\) is a proper part of \(\alpha_i\). For example, a device’s circuit design is composed of bus type and memory type.
Conclusion and Future work

In this paper two novel graphic-based models were introduced which offer an insight into how personally identifiable information is utilized within the cyber world. The model of devices and organizations imply the proliferation of technology as the Internet brings closer the vision of the Internet of Things. We are interested in the connections of these models to the PII model of people. Previously the Identity Ecosystem could answer three interesting questions, which were based only on the PII attributes of people. By combining and analyzing the people, device, and organization models together, we expect to derive more accurate and comprehensive results from the Identity Ecosystem. The potential structures and types of this cooperation framework and innovation resources from ITAP need further examination and interpretation since the Center for Identity envisions using low risk, low value, and high uniqueness PII for identifying and authenticating people in the future IoT-based society.

References


