2nd Floor

P1

PS

C3

C3

W

L4

Testbed 3

B

C1

Testbed 1

Testbed 2

Testbed 4

Figure 7: Floor plans of four testbeds and device deployment layouts (the device abbreviation labels are illustrated in Table 3).

approval for the study. All participants are fully aware of

the bedroom door (with C3) does not have this pattern.

all the installed devices and apps. We do not use any sensi-

Observation 2: The e2s correlation C23 means that MS3's

tive devices such as cameras and microphones. The sound

illuminance goes high only when L4 is on. This is because

sensor of Device A in Table 3 only reports the sound level

there are no other light sources near MS3. Other illuminance

rather than the raw audio. All data is considered sensitive

sensors do not have such a correlation as the high illumiand personal identifiable information (PII) is removed right

nance value can be caused by multiple lights or natural lights.

after collection for long-term storage. We store all the data in

an encrypted hard drive mounted to our lab's server, which

Observation 3: Smart plugs P2 and P4 are to turn on/off a

is only accessible to accounts of the paper authors.

fan and a lamp, respectively. Whenever P2 and P4 are turned

For the purpose of testing, we need to inject anomalies (see

on, higher power use is observed (see e2e correlations C16

and C10 in Table 5). However, for P1 that is connected to a

Section 6.3). To avoid safety issues, the injected anomalies

switch(P1)

do not target any safety-sensitive devices, such as heaters.

TV,

Eon

is not followed by a power-high event, as the

We notify participants of incoming testing one day ahead

TV needs to be further turned on manually by the residents.

but do not disclose the details (e.g., device and time) of the

Observation 4: Physical- and user activity-channel correla-

anomaly cases. We also ask participants to keep their normal

tions cannot be obtained without mining, since they are not

living habits and do not panic if they notice any anomalies.

included in any smart apps. On the other hand, some corre-

The purpose is to avoid their behavioral bias during testing.

lations can be easily extracted from smart apps but difficult

Details of the injected anomalies are presented to participants
to mine. For example, correlations that involve delays are
after the testing.
difficult to be mined accurately, but can be precisely derived
from rules, such as R4, R6, R8, and R10.

6.2 Training

Training Baseline Approaches. We select the Association
Training HAWatcher. From Testbed 1, we generate 46 e2e
Rule Mining (ARM) [24] and the One-class Support Vec-
correlations from the automation rules. In addition, we gen-
tor Machine (OCSVM) [67] based detectors as two base-
erate totally 2,398 hypothetical correlations, including 46
line approaches. We choose OCSVM because it is wiedly
e2s correlations from the smart app channel, 544 from the
used for anomaly detection and trained with one class of
physical channel, and 1,808 from the user activity channel.
input data, which is suitable for our training data containing
Then, the hypothetical correlations are checked using 22,655
no or few anomalies [53]. ARM is selected because it is a
events collected from the three weeks' training phase. In
well-established method for mining correlations/rules, and
total, 146 correlations are accepted by hypothesis testing,
HAWatcher is also based on correlation mining.
and 130 remain after refining. On other three testbeds, the
We perform ARM [24] on the same training dataset for
portion of smart app channel correlations are 32/109, 15/55,

comparison. Since ARM algorithms require transaction-form

and 8/26, respectively. Table 5 lists a portion of the corre-

inputs, we segment the training dataset at places where the

lations after refining. Some correlations reveal interesting

time interval between two consecutive events is longer than

facts that are confirmed by the residents.

60s (the same as the threshold d used for hypothesis testing).

Observation 1: While C1 and C3 are both contact sensors,

By using the library pymining [22], we mine 221 association

C1 has one additional correlation C11 =

acceleration(C1)

rules with the confidence threshold of 0.95. Unlike our cor-

,contact(C1)

which

means

the

,acceleration(C1)

relation mining method that covers various attributes and

event

active

should

closed

devices, rules produced by the association rule mining are

be

followed

by

contact(C1)

This is because the front door

'closed

dominated by motion sensors MS3 and MS4. All the 221 rules

(with C1) is typically closed right after being opened, while

have either MS3 or MS4's motion attributes in their conse-