THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE

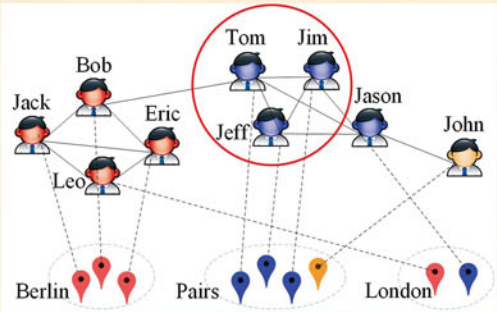
Continuous Spatial-Aware Community Search

Wang Zheng **Supervisor: Prof. Reynold C.K. Cheng**
Department of Computer Science, The University of Hong Kong

INTRODUCTION

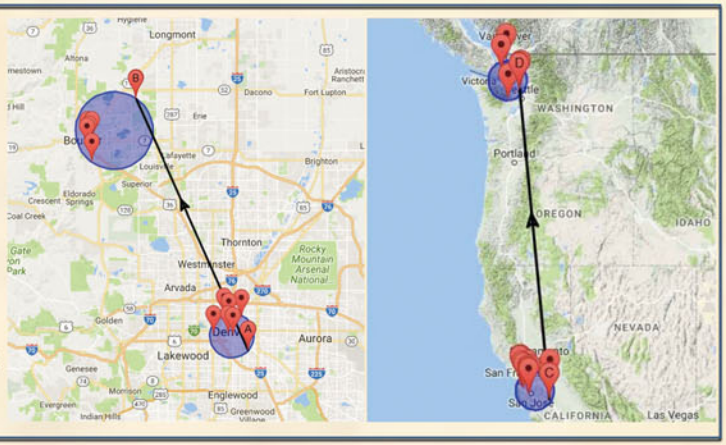
- **Can we find a community that is close socially and geographically?**
- ***User's communities could evolve over time? Continuous SAC Search***
- ***Motivation of CSAC Search?***

1. Event recommendation & High-performance
2. Social marketing
3. Geo-social data analysis



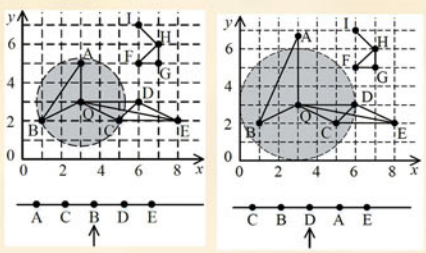
PROBLEM

- **[SAC search] Given a spatial graph G , an integer k, and a query vertex q, return a subgraph G_q such that**
  - ❑ It is connected and contains q;
  - ❑ Each vertex has at least k neighbors;
  - ❑ The minimum covering circle (MCC) of vertices in G_q satisfying Properties 1 and 2 has the minimum radius.
- **[CSAC search] find a list of SACs in a time period from Ts to Te**
- **Real examples in Brightkite**

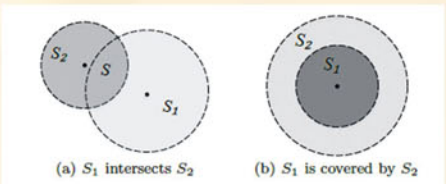


ALGORITHMS

CAppFast



CAppAcc



Cexact+

- ❑ Prune fixed vertices
- ❑ Narrow the range of optimal radius
- ❑ maintain variables
- ❑ improve the efficiency of query

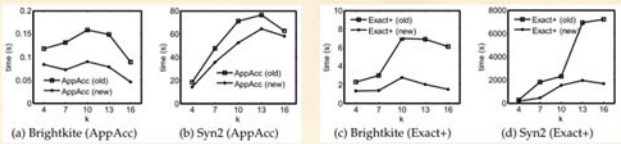
Improvement of AppAcc

- ❑ A new lemma
- ❑ A new pruning

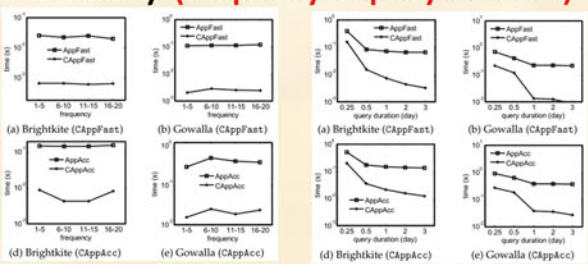
EXPERIMENTS

- ❑ Four real datasets (Brightkite, Gowalla, Flickr, Foursqure)
- ❑ Two synthetic datasets

Improvement SAC algorithms



Efficiency (frequency & query duration)



Wearable Computing System

Sheshan Aaron

Supervisor: Dr. Dirk Schnieders

THE UNIVERSITY OF HONG KONG

DEPARTMENT OF COMPUTER SCIENCE

Introduction

Over the last few years, smartglasses have grown in popularity, however, there are still many interaction challenges that need to be overcome before they become widely adopted. Furthermore, social media has enabled billions of people to report and share life and news events immediately. One potential application of smartglass technology for mainstream consumer usage is lifelogging and reporting via social media.

Objectives

In this project, we seek to develop a complete smartglass system, together with user interaction methods and software for a mobile twitter reporter and evaluate the system's usability.

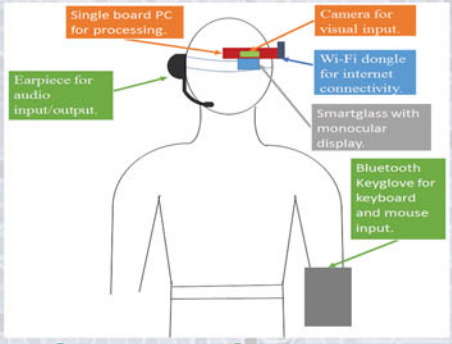
Aims

1. Development of a smartglass with audio and visual output.
2. Development of textual and command input methods.
3. Development of an application that will interact with Twitter.

Methodology

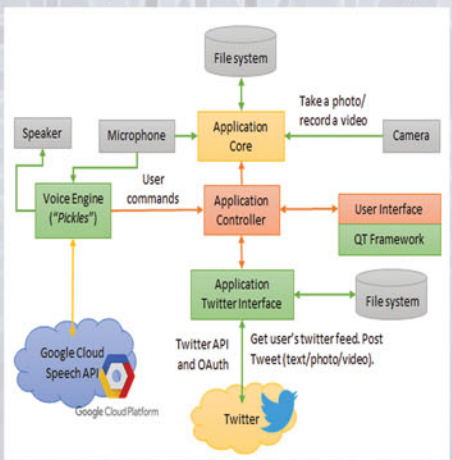
Hardware Architecture

We developed a keyglove and a smartglass, and incorporated other hardware elements.



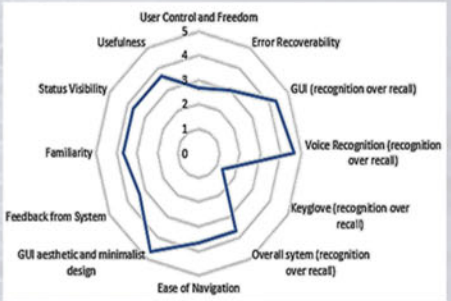
Software Architecture

The developed software system consists of a voice recognition engine ("Pickles"), twitter application software ("Twiptor"), a user interface, and the application core functionality.



Results

User tests were performed on different subjects, with users using the full system to perform a range of tasks. The users were then asked to rate the application on different metrics while comments were also taken on different usability aspects of the system. The following diagram illustrates the overall feedback



Discussion

Analyzing the results, we can see that users particularly liked the voice recognition feature and mostly disliked the key glove. We think this is due to the learning curve involved for voice control compared to the key glove. Therefore, we need to rethink our strategy for key input.

Further Work

The following is a list of areas we would like to work on:

1. Smaller, lighter, more ergonomic design for smartglass and key glove.
2. Have a better layout for keys and have less keys for key glove.
3. Reduce system latency.

Reconstruction of Display and Eyes from a Single Image

Dirk Schnieders, Xingdou Fu, and Kwan-Yee K. Wong
Department of Computer Science, The University of Hong Kong, China

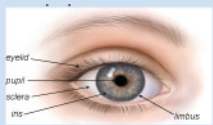
Introduction

- A novel method for reconstructing human eyes and visual display from reflections on the cornea.
- It is useful for point-of-gaze estimation, which can be approximated from the 3D positions of the iris and display.
- Measuring our eye gaze has a large number of applications in several fields including psychology, industrial engineering and advertising.
- It is shown that iris boundaries and display reflections in a single intrinsically calibrated image provide enough information for this.

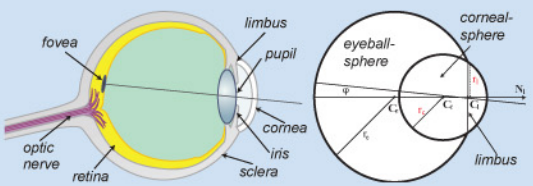
Contributions

- Estimate the positions of eyes and display from a single image.
- No subject-specific parameters for determining where a user is gazing relative to a computer display.
- Active illumination is not employed in this work, and as a result, off-the-shelf equipment can be used.

Human Eye Model

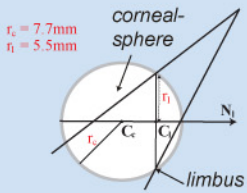


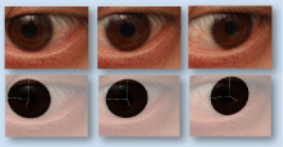
- Model the eye by segments of two quasi spheres.
- The optical axis is assumed to be parallel to the supporting plane normal of the limbus.
- It has been found that the difference in eye parameter values among adults is small and it is reasonable to assume that some parameters of an adult human eye are close to certain anatomical constants.



Reconstructing the Limbus

- The circular limbus can be reconstructed *(up to a sign ambiguity)* from its perspective projection.
- The corneal sphere center can than be determined from the anatomical constants.

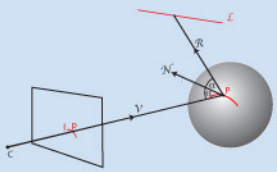


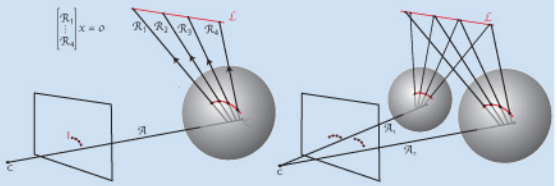


$$|\mathbf{C}_l\mathbf{C}_c| = \sqrt{r_c^2 - r_l^2}$$
$$\mathbf{C}_c = \mathbf{C}_l - |\mathbf{C}_l\mathbf{C}_c|\mathbf{N}_l$$

Display Edge Reconstruction

- A single display edge is reconstructed from its reflection on the cornea.
- The reflection enables single view line estimation by solving for the nullspace of a Plücker line matrix.

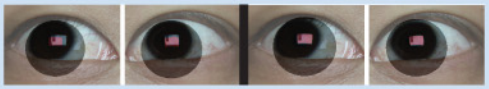




- A single corneal sphere results in two possible candidates for the display edge.
- Two corneal spheres determine the display edge without ambiguity.

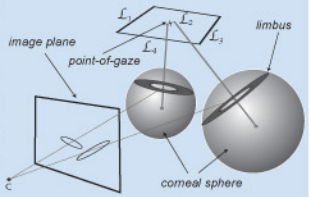
Optimization

- Display edges are independently estimated as four lines with the closed form solution.
- In general the four lines will not intersect, nor form a perfect rectangle.
- Limbus boundaries and anatomical parameters are optimized such that the four lines form a rectangle.

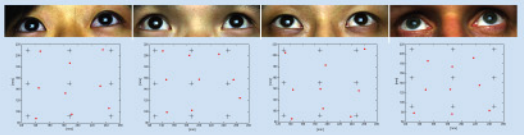


Gaze Estimation

- A single point-of-gaze is approximated as the point on the display that minimizes the sum of the distances to the two optical axes.



Approximate Point-Of-Gaze



Conclusions

- Eye gaze estimation has a history of more than 100 years, yet it is still not widely used, because the technology lacks usability requirements that hinder its applicability.
- One of the main challenges is how to make the calibration of a commercial eye gaze tracker simple and easy.
- Minimal setup and calibration requirements of our method enable this technology for everyone.

ONLINE MULTI-USER TRAVEL PLANNER APP

Introduction

The dissertation aims at building an iOS app that helps users collaboratively plan their trips.

Users can browse through tourist spots with advanced operations such as searching, sorting and filtering.

The app collects users' inputs on their travel preferences, and automatically constructs a travel timetable which is freely customizable by multiple users in real-time.

User

- Log in and save trips on cloud •
- Share travel plans to other users •

Travel Plan

- A travel plan represents a multi-day trip •
- Choose one from a list of plan to work on •
- Control your overall plan •

Spot

- A spot represents a tourist attraction •
- Scroll through a list of spots •
- Find spots by searching, sorting and filtering •
- View spot details •
- Indicate preferences for auto-scheduling •

Timetable Planner

- A timetable represents a detailed itinerary •
- Spots are placed on the timetable as blocks •
- Drag and drop from a list of favourites •
- Modify durations by adjusting the circular handles •
- Move blocks around with long-press gestures •
- Landscape view to enable multi-day editing •

Collaboration

- Save preferences and itineraries asynchronously to cloud •
- Real-time synchronization across shared users •

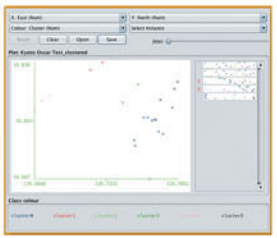
Swift

amazon web services

Java EE

Auto-scheduling

- Users provide priorities and time preferences •
- Server algorithmically generates timetables •
- Balance between computational cost and optimality •
- Use GPS as distance measure to find shortest paths •



Fong Oscar Hok Kar
Supervised by Dr. Ronald H.Y. Chung

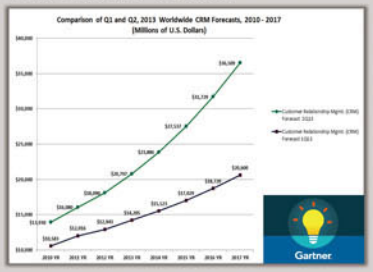
Real-Time Customer Relationship Management Using Elasticsearch

Xiyao Zhang

Supervisor: Prof. C.L. Wang

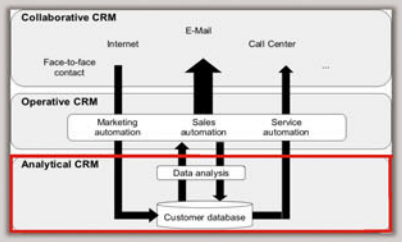
Motivation

In a Market's View:



- Expecting a $36.5B worldwide market by 2017
- Huge Market and Promising Prospects !

In a Product's View:



- Analytical CRM is the engine to high-level operational and collaborative CRM

Problems Exist:

- Offline batched model learning and result processing
- Only single party data analytics supported
- Fixed model learning and processing

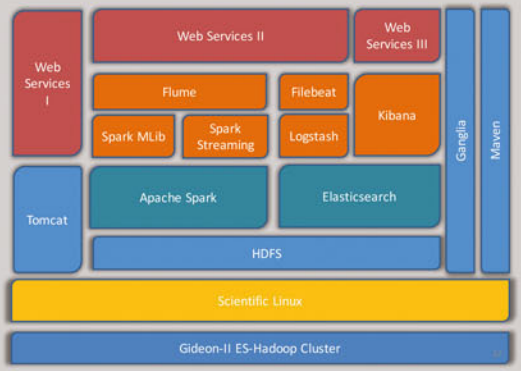
Objective

- To design and implement a SaaS analytical CRM platform with real-time learning and processing capability from bottom to top
- An Enterprises-oriented big data analysis application

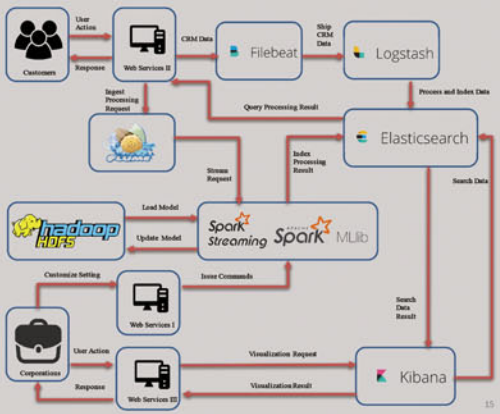
Key Features

- Real-time incremental learning and response with data and visualization
- Multi-parties data networking coordination
- Customizable knowledge model learning and processing

System Architecture

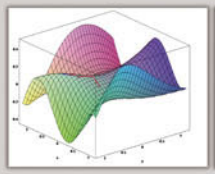


System Workflow



Real-time Enablement

- Real-time Incremental Learning using SGD



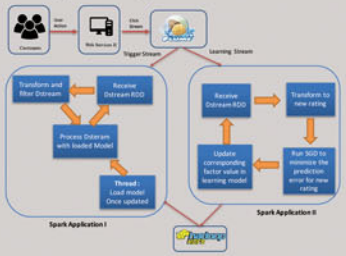
- Minimization(prediction error)

$$w := w - \eta \nabla Q(w) = w - \eta \sum_{i=1}^{n} \nabla Q_i(w)$$

- Learning rules

$$p'_{ik} = p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e^2_{ij} = p_{ik} + \alpha(2e_{ij}q_{kj} - \beta p_{ik})$$
$$q'_{kj} = q_{kj} + \alpha \frac{\partial}{\partial p_{kj}} e^2_{ij} = q_{kj} + \alpha(2e_{ij}p_{ik} - \beta q_{kj})$$

Updating learning models in real-time



Usage Scenarios

E-commerce

Social Network

Visual Monitor

Environment

Cluster Name	Elascticsearch-Hku
Nodes	Se003-se008, se041-se050 (total 16 machinces)
CPU	Intel(R) Xeon(R) CPU E5540 @ 2.53GHz
Memory	16G
Operating System	Scientific Linux release 6.6 (Carbon)

Main Results

- **Speed Improvement** : Achieve 100X faster processing speed than batched version
- **Accuracy Improvement** : 20% with incremental learning

Low-light Obstacle Avoidance & Navigation for an Autonomous Quadrotor Robot

Supervisor: Dr. Robert C. Roberts
Author: Ngan Wai Kong, Tony

Mar 2016 ~ Jul 2016

DJI Developer Challenge - TeamHKU

A team with 4 MSc and 2 UG CS students participated in DJI Developer Challenge.
It's about autonomous rescue operation by locating tags and target landing in a diasaster site.
More than 30 testings were done, robust tag detection, landing and simple avoidance were made.
TeamHKU was ranked top 15 over 130 competitive teams from startups and universities.

Can the drone achieve collision-free navigation under insufficient light condition ?
Modeling a low-cost quadrotor robot with one monocular camera and real-time sensor
Feasibility analysis of classic optic flow algorithms with different brightness levels.
Testing-driven development on top of ROS and Gazebo simulator.

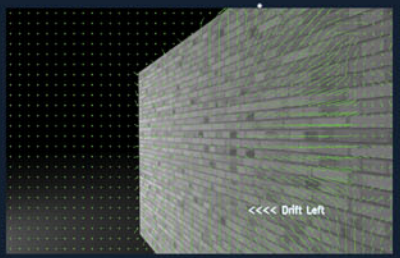
Sep 2016 ~ Mar 2017

Beyond the challenge - Thesis

TeamHKU


Avoidance Test in Challenge


Simulation World



Reactive Controller (Optic flow)

Retrieve real time sensor data from ROS topics through ARDrone driver

Generate the optic flow field in each consecutive image frame (OpenCV)

Divide image into 3 smaller windows and compare the distribution of flow vector

Achieve collision-free navigation using statistical based threshold control

Simulate the operation using Gazebo and schedule field testings

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE

Student: Jasper Barzilaij
Supervisor: Dr. S.M. Yiu

Secure Testing for Genetic Susceptibility on the Cloud

Background

The rapid increase in the availability of genetic data allows researchers and other parties to access a vast amount of genetic data. Access to such data supports the development and innovation of various applications. However, genetic data is highly sensitive and has major implications for one's **privacy**. To give an example of the sensitivity of the data, **genetic susceptibility** can be determined by one's genome. We can eliminate the risks and preserve the privacy of geneticdata by using **homomorphic encryption**.

Goal

To develop a secure system with which a user can determine if they are **genetic susceptible** for one or more diseaes, while preserving one's **privacy**, in a practical manner.

The genetic data (collection of encrypted **SNPs**) of the user resides with an untrusted hosting service, which is encrypted with a homomorphic encryption scheme and send to the host during installation/registration. With the system all set up, the user can query the host to determine his genetic susceptibilty based on the submitted genetic data.

Genetic Susceptibility

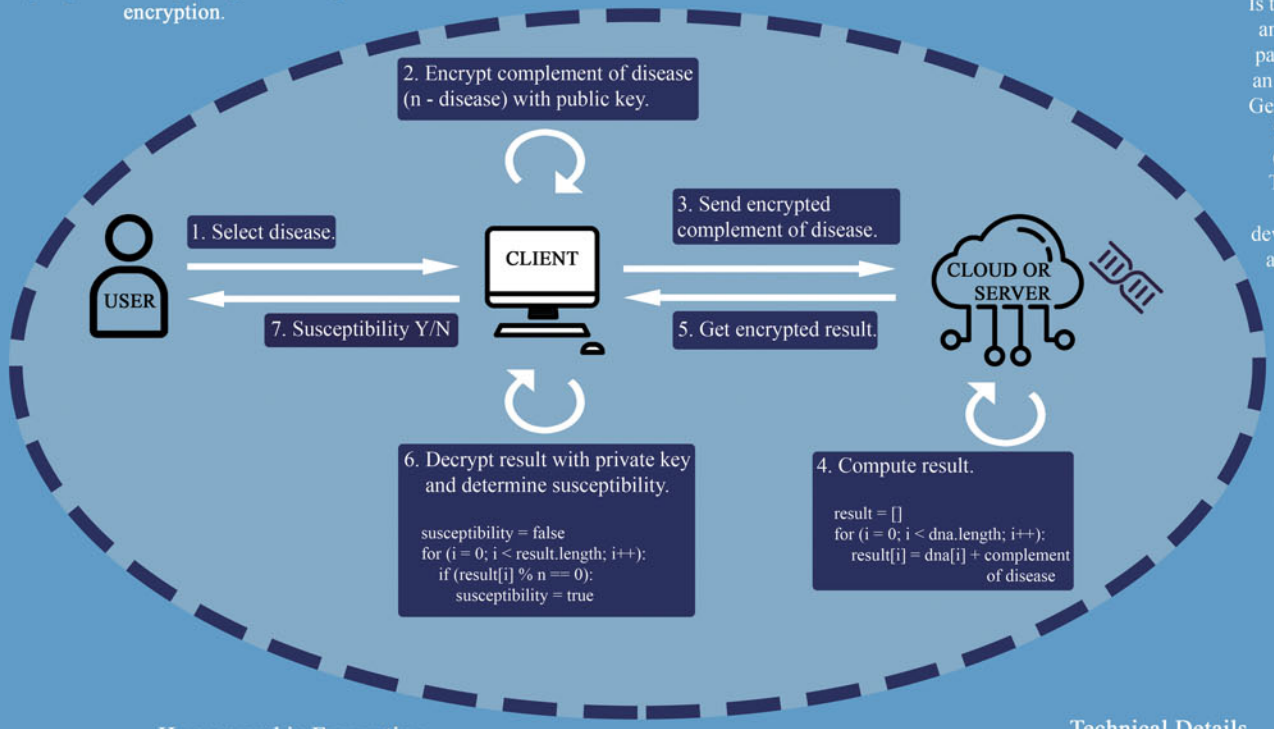
Is the increased likelihood of an individual developing a particular **disease** based on an individual's genetic data. Genetic susceptibility results from genetic variations (represented by **SNPs**). These genetic variations contribute to the development and severity of a disease but not directly cause it.

Genome

A genome is an organism's complete set of DNA, including all its genes

Gene

"a part of the DNA in a cell that controls the physical development, behaviour, etc. of an individual plant or animal and is passed on from its parents"
[Cambridge dictionary]



SNP

A single-nucleotide polymorphism (SNP) is a genetic variation that occurs at a specific position in the **genome**. And underlies the **genetic susceptibility** to **diseases**.

Homomorphic Encryption

Homomorphic encryption is a type of encryption that allows you to perform computations on encrypted data while preserving its **privacy**.

operation(data) = decrypt(operation(encrypt(date))

With fully homomorphic encryption, any operation on the ciphertext is possible. However, due to the processing time, it is impractical. Partially homomorphic encryption supports either addition or multiplication, Although it supports a limited set of operations, it is less complex and far more practical.

Paillier Cryptosystem

Paillier's cryptosystem is a partially homomorphic encryption scheme and is the best candidate for our system. It exhibits **additive homomorphism**, multiplying two or more ciphertexts is equivalent to the addition of the plaintext values.

$$c1 = E(k, p1)$$
$$c2 = E(k, p2)$$
$$c1 * c2 = m1 + m2 \bmod n$$

Technical Details

In our system, the variable 'disease' is a **SNP** that is associated with a genetic disease. And the genetic data of an individual, encrypted and uploaded to a untrusted hosting service is a **collection** of the individual's **SNPs**.

In order to determine, with **Paillier's cryptosystem**, if a disease is present in one's genetic data, we need to be able to check for equality between two **SNPs.** This can be achieved by using complements.

If snp-disease equals snp-user:
$E(n - snp\text{-}disease) + E(snp\text{-}user) \bmod n = 0$

THE UNIVERSITY OF HONG KONG

DEPARTMENT OF COMPUTER SCIENCE

COMP7704 MSc(CompSc) Dissertation

IP Camera Security Study

Supervisor: Dr. Lucas C.K. Hui

Student: Wong Po Shing

Objective

The objective of this dissertation is to develop a set of systematic security evaluation and configuration guidelines for the IP cameras. These guidelines can assist the end users to select and protect their IP cameras. Furthermore, I hope the information of this dissertation will help the IP camera manufacturers to improve their products.

Methodology

Knowledge

1. IP cameras' vulnerabilities study

- XSS
- DDNS poisoning
- Arbitrary shell commands execution
- Hard-coded credentials
- Backdoor accounts

2. Attack and defend study

- Cyber Kill Chain [1] study

- Experiment of repeating attacks following Cyber Kill Chain stages

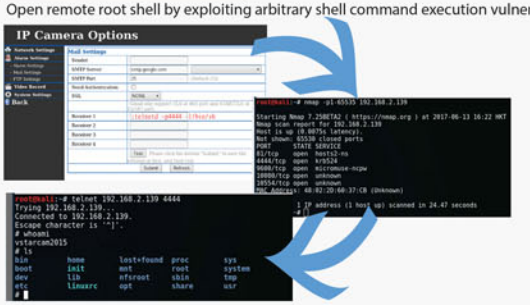
Get admin password by sending malformed HTTP request

```
root@kali:~# python getPassword.py 192.168.2.139 81
** Generic ipcam exploit - bind shell **
connecting...
[*] Sending malformed HTTP request

[*] Packet sent!
[*] Got info leak.. Parsing...
[*] Success! Admin password is: 'password'
root@kali:~#
```

Malformed HTTP request	Legitimate HTTP request
GET index.asp HTTP/1.1	GET /index.asp HTTP/1.1

Open remote root shell by exploiting arbitrary shell command execution vulnerability



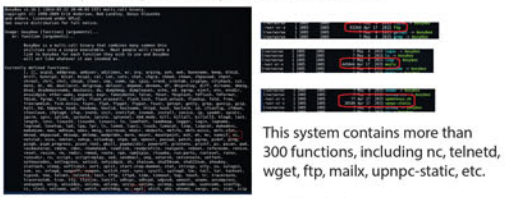
Framework

OWASP IoT security assessment framework [2] study

I1: Insecure Web Interface	I2: Insufficient Authentication/Authorization
I3: Insecure Network Services	I4: Lack of Transport Encryption
I5: Privacy Concerns	I6: Insecure Cloud Interface
I7: Insecure Mobile Interface	I8: Insufficient Security Configurability
I9: Insecure Software/Firmware	I10: Poor Physical Security

3. Operating system study

- Identify dangerous system functions



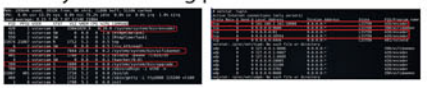
This system contains more than 300 functions, including nc, telnetd, wget, ftp, mailx, upnpc-static, etc.

- Identify design & configuration flaws

Wrong place for ini files

User credentials are in plain text

- Study the running processes

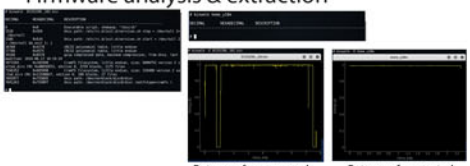


4. Application & firmware study

- Disassemble and analyze applications

No sanitation for parameters makes arbitrary shell command execution possible.

- Firmware analysis & extraction



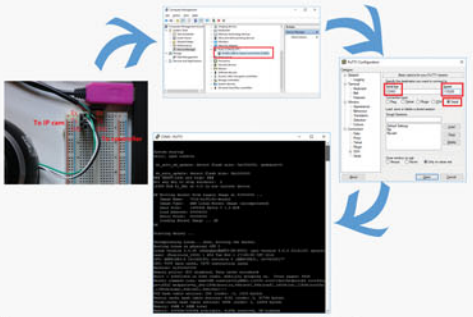
Entropy of compressed firmware

Entropy of encrypted firmware

5. Physical security study

- Discover serial port on the circuit board (PCB)

- Connecting to the IP cameras' serial ports



Assessment checklist

7 categories are selected from OWASP IoT security framework to create checklist template; including 1)web interface, 2)authentication/authorization, 3)network services, 4)transport encryption, 5)security configura--bility, 6)software/firmware and 7)physical security. Selection of the categories is based on the areas that I had studied in this project.

Part of the assessment results of P2P Camera S6203Y, Yi Home camera and Yi Dome 1080P camera.

Conclusion

About the checklist

- The checklist provides the baseline for IP camera security assessment in a systematic way
- The items in the checklist should vary with the features of the IP camera.

Future work

- Study the categories of Mobile and Cloud interface and add them to the checklist.

1. Lockheed-Martin. Cyber Kill Chain. Available from: http://www.lockheedmartin.com/us/what-we-do/aerospace-defense/cyber/cyber-kill-chain.html.
2. OWASP. IoT Testing Guides. Available from: https://www.owasp.org/index.php/IoT_Testing_Guides.

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE

Efficient Classification Algorithm K-NN for Encrypted Data

OUR CONTRIBUTION

1. Data privacy and confidentiality: Contents of database A and database B or any other intermediate results would not be revealed to both the SHTP1 and SHTP2.

2. Query privacy and confidentiality: Contents of query P or any other intermediate results would not be revealed to both the SHTP1 and SHTP2.

3. Accuracy and correctness: The final result {d1, d2, d3…dk} would be only revealed to Peter from SHTP1 and SHTP2.

4. No computations needed on the query provider: After sending his encrypted query P to SHTP1, Peter does not need to participate any computations in the whole progress. That is a real improvement compared with previous secure models.

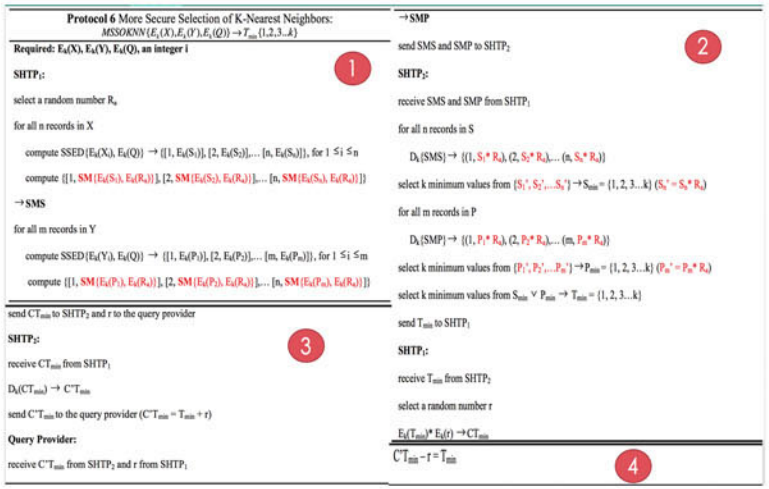
Paillier Cryptosystem Algorithm: *compared to RSA Cryptosystem (not additive homomorphic), and Distance-preserving transformation (DPT) is only a data transforamtion Ek (p) = Np + t (N is a dxd orthogonal matrix, t is a d-dimensional column vector)*

More Secure Selection of K-Nearest Neighbors Protocol:

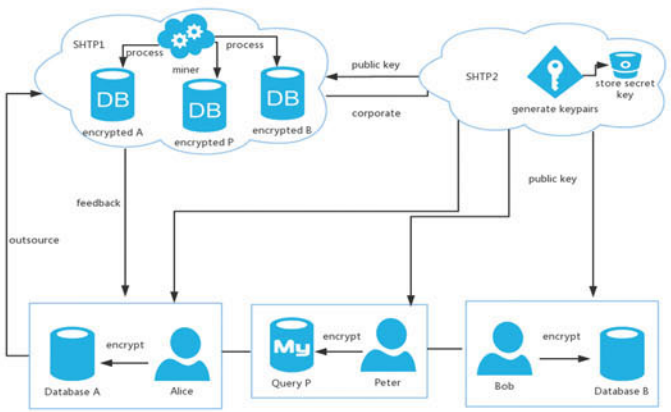
1. Improve the security level of SKNNS Model

2.The vulnerability of SKKNS Model is to reveal the real Euclidean distances to SHTP2

More Secure Selection of K-Nearest Neighbors Protocol



Secure K-Nearest Neighbor Scheme Model (SKNNS)



BACKGROUND

Due to the increased tendency of cloud computing, there are some service providers like Google and Amazon moving their institutions to the cloud. That means to share knowledge or work collaboratively is an emerging and key technique to rescue some traditional companies and organizations. Thus, many companies and organizations are willing to outsource their own private data to save the storage and data management cost. In order to enforce the privacy and data confidentiality, the direct approach is to hide the original data on client provider before outsourcing. However, to hide the original data and mine them straightforwardly is a big challenge in Information Security research area. We need to construct a new secure data mining algorithm to protect the privacy and data confidentiality of customers' own data.

PAILLIER CRYPTOSYSTEM

The Paillier cryptosystem is an additive homomorphic and probabilistic asymmetric encryption scheme for public key cryptography. The scheme is an additive homomorphic cryptosystem, that means given only the public key and the encryption of m1 and m2, one can compute the encryption of m1 + m2. The properties of the Paillier cryptosystem could be included as three formula.

1. Homomorphic Addition $E_k(a+b) = E_k(a) * E_k(b)$

2. Homomorphic Multiplication $E_k(a*b) = E_k(a)^b$

3. Semantic Security to against chosen-plaintext attacks (If given a set of cipertexts, an adversory could not deduce any additional information about the plaintext.)

MSc Dissertation 2016-2017

Supervisor: Dr. S.M.Yiu

Author: YAN Jiazheng (Clark)

Designing a New Security Protocol for V-NDN Communications

Student: Wong Lik Yan Supervisor: Dr. T.W. Chim

Introduction

V-NDN is a vehicular network architecture that naming the data packet with the data name rather than the end point. The self-contained data is believed to be well-suited to the vehicular network.

In the project, we propose to use HIBS in V-NDN to authenticate the data packet. We suggest using bloom filter to notify the user about the list of the malicious user.

Hierarchical Identity-Based Signature (HIBS)

The security of this signature scheme is based on CDHI assumption and selective-ID security model. The scheme comprises 6 functions as follows:

Setup:

Pick random generator g, $h_i \in G$ where G is cyclic group of prime order p

Pick random $\alpha, x_j, z \in Z_p^*$

Set $g_1 = g^{\alpha}$, $u_j = g^{x_j}$, $Z = g^z$

Public Key = (g, g_1, h_i, u_j, Z)

Master Key = (α, x_j , z)

where i = 0, …, number of message (n_m) and j = 1, … , maximum hierarchy level (l)

Key extract:

$$d_{id_k} = \left(g^{\frac{1}{\alpha + r_k \left(\left(\sum_{i=1}^{k} x_i v_i \right) + kz \right)}} \left(\prod_{i=1}^{k} g^{v_i} \right)^{r}, r_k, g^r \right)$$

$$d_{id_k} = (d_0, d_1, d_2)$$

where r and r_k are random Z_p^* , ID = $(v_1, \ldots, v_k) \in Z_p^*$ with $k \leq l$

Key delegate:

$$d_{id_{k+1}} = \left(d_0 \, d_2^{v_{k+1}} \left(\prod_{i=1}^{k+1} g^{v_i} \right)^{r'}, d_1, d_2 g^{r'} \right)$$

where r' is a random Z_p^* , ID = $(v_1, \ldots, v_{k+1}) \in Z_p^*$ with $k+1 \leq l$

Sign:

$$\sigma = \left(\frac{1}{d_2^{\sum_{i=1}^{k} v_i}} d_0 \left(h_0 \prod_{i=1}^{n_m} h_i^{m_i} \right)^{s}, d_1, g_1^{s}, \left(Z^k \prod_{i=1}^{k} u_i^{v_i} \right)^{s} \right)$$

$$\sigma = (\sigma_0, \sigma_1, \sigma_2, \sigma_3)$$

where s is a random Z_p^* , message = $(m_1, \ldots, m_{n_m}) \in Z_p^*$

Verify:

$$e\left(\sigma_0, g_1 \left(Z^k \prod_{i=1}^{k} u_i^{v_i}\right)^{\sigma_1}\right) == e(g,g)\, e\left(h_0 \prod_{i=1}^{n_m} h_i^{m_i}, \sigma_2\, \sigma_3^{\sigma_1}\right)$$

where e is a bilinear map

Batch verify:

$$e\left(\prod_{j=1}^{N} \sigma_{0,j}^{\delta_j}, g_1\right) e\left(\prod_{j=1}^{N} \sigma_{0,j}^{\delta_j \sigma_{1,j}}, Z^k\right) \left(\prod_{i=1}^{k} e\left(\prod_{j=1}^{N} \sigma_{0,j}^{\delta_j \sigma_{1,j} v_{i,j}}, u_i\right)\right) =$$

$$= e(g,g)^{\sum_{j=1}^{N} \delta_j}\, e\left(h_0, \prod_{j=1}^{N} \left(\sigma_{2,j} \sigma_{3,j}^{\sigma_{1,j}}\right)^{\delta_j}\right) \left(\prod_{i=1}^{n_m} e\left(h_i, \prod_{j=1}^{N} \left(\sigma_{2,j} \sigma_{3,j}^{\sigma_{1,j}}\right)^{m_{i,j}\, \delta_j}\right)\right)$$

Performance:

sign	verification	batch verification
7.68ms	10.94ms	*9.30ms

Table 1: The average computation speed per signature

* Remarks: The batch verification is tested against 100 nos. message-signature pairs.

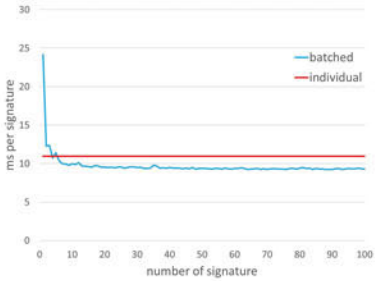
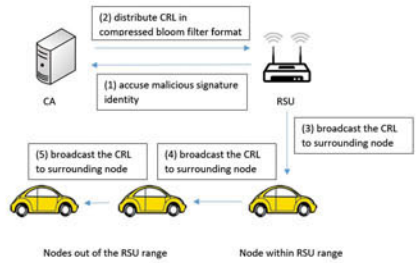


Figure 1: The comparison of individual and batched scheme

Certificate Revocation List (CRL) with bloom filter



Handling false positive

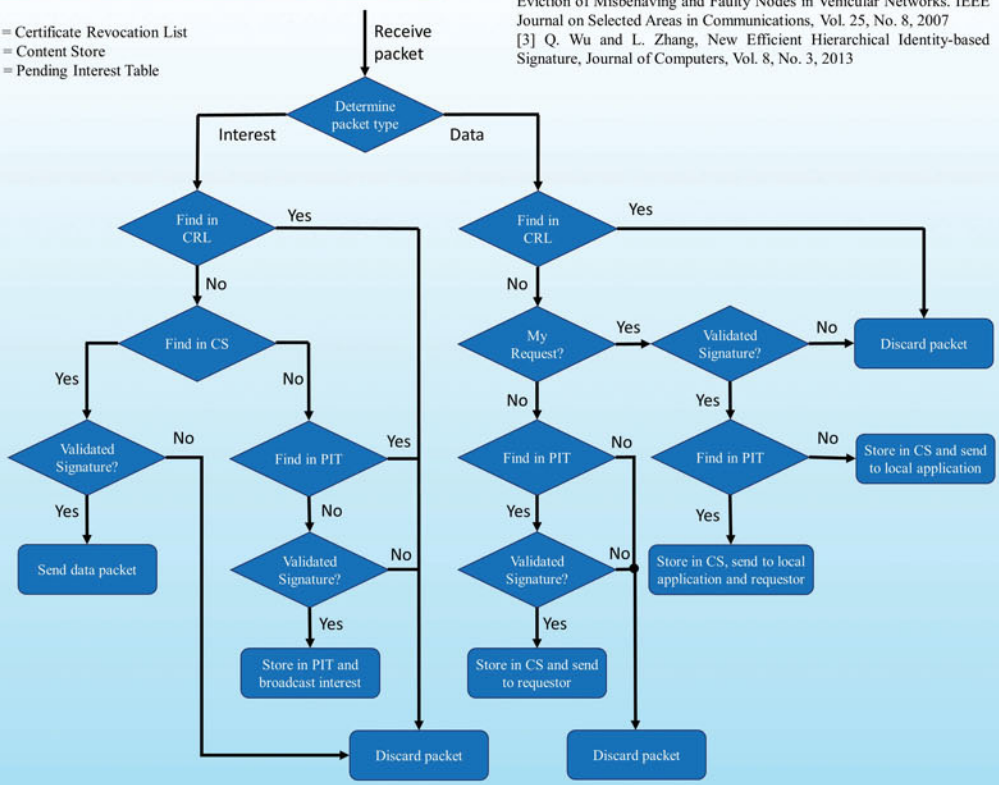
Bloom filter is a probabilistic data structure, it is possible to find false positive.
To mitigate the impact, we suggest:
1. Concatenate a random number at the end of the identity.
2. Every OBU possesses more than one key pairs
3. Request a new key if false positive found

Reference

[1] G. Grassi, D. Pesavento, L. Wang, G. Pau, R. Vuyyuru, R. Wakikawa and L. Zhang, Vehicular Inter-Networking via Named Data, arXiv:1310.5980.
[2] M. Raya, P. Papadimitratos, I. Aad, D. Jungels, and J.-P. Hubaux. Eviction of Misbehaving and Faulty Nodes in Vehicular Networks. IEEE Journal on Selected Areas in Communications, Vol. 25, No. 8, 2007
[3] Q. Wu and L. Zhang, New Efficient Hierarchical Identity-based Signature, Journal of Computers, Vol. 8, No. 3, 2013

Data transmission flow of the vehicle

CRL = Certificate Revocation List
CS = Content Store
PIT = Pending Interest Table



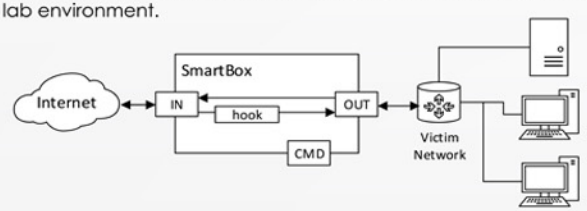
香港大學計算機科學系
THE UNIVERSITY OF HONG KONG
DEPARTMENT OF COMPUTER SCIENCE

Anti-DDoS Technique Using Self-learning Bloom Filter

C.Y. Tseung, K.P. Chow, X. Zhang

Introduction

DDoS attack is still one of the major threats from Internet. We propose a new technique to mitigate different types of DDoS, combining and taking advantages of both machine learning algorithms and Bloom filter. We use machine learning to extract features of attacks, then use a customized Bloom filter to defend attacks based on selected features. We implemented and tested the performance of the proposed technique in a lab environment.



Machine learning and Bloom filter

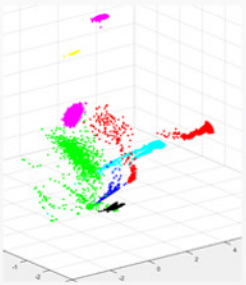
Machine learning is used in detecting whether an attack is happening. Once an attack is detected, the packets will be further analyzed and best determining features will be selected by the machine learning algorithms. The resulting feature list essentially means that by looking at these features, we can determine if a packet is good of bad.

Bloom filter is a space-efficient algorithm for identifying consecutive similar objects. Making use of this feature we can classify packets into "normal packet" or "attack packets". For attack packets, we will simply drop them inside the SmartBox.

Leaky bucket algorithm is also used with the Bloom filter to decrease all bin counters as a "cool-down" method, allowing legitimate packets to pass through continuously.

Characteristic of attack packets

A simple "principal component analysis" done on attack packet samples shows that these packets have some feature in common. Our idea is to automatically separate bad traffics and block them.

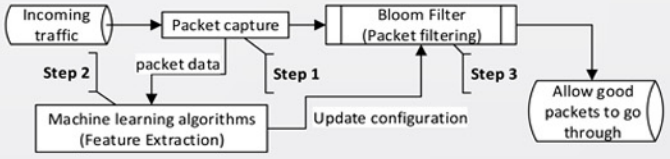


Defense flow

Step 1 – Capture and manipulation of incoming packets
To avoid changing the topology of victim's network, we make our interception transparent to the network. We utilize the built-in network bridge and hook of Linux system to inspect and modify network flow transparently.

Step 2 – Self-learning Feature Extraction
Our machine learning algorithm runs in background to identify whether the victim is being attack. A list of features of packets will be extracted and automatically applied to the Bloom filter for attack packet blocking.

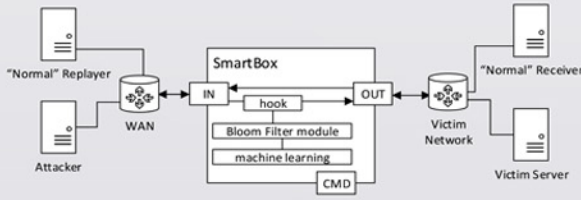
Step 3 – Defend with Bloom filter
Based on features given, such as network protocol, source address, TTL, our tailor-made Bloom filter identifies attack traffics and block them.



Experiment result

A testbed was built to test our proposed technique. PCAP files containing normal traffics and attack traffics were replayed to simulate the attacking scenario. We use 2 servers for receiving traffic of normal and traffic of attack to calculate the blocking rate of attack traffic and passing rate of normal traffic. As all traffics will go through the SmartBox, it simulates the situation that the victim is being attack during normal business operations.

By calculating the traffic received in "receivers" and compare to the original traffic "replayed", we can calculate the blocking rate for attack and normal traffic.



Input Attack Type	Learning Algorithm	Feature Automatically Selected	Attack Block Rate	Normal Pass Rate
ICMP	Baseline	ip.src, ip.proto, ip.srcport	0.00%	98.39%
	ANOVA	frame.len, ip.srcport, tcp.flags	94.09%	100.00%
	Median	frame.len, ip.srcport	94.07%	100.00%
NTP	Baseline	ip.src, ip.proto, ip.srcport	95.70%	98.39%
	ANOVA	frame.len, ip.srcport,ip.dstport, tcpflags	95.70%	98.39%
	Median	frame.len, ip.srcport	95.70%	100.00%
UDP	Baseline	ip.src, ip.proto, ip.srcport	0.55%	100.00%
	ANOVA	tcp.flags	100.00%	91.94%
	Median	ip.srcport	41.16%	100.00%
DNS	Baseline	ip.src, ip.proto, ip.srcport	1.34%	100.00%
	ANOVA	ip.srcport, tcp.flags	91.59%	91.94%
	Median	ip.srcport	91.20%	93.55%
TCP	Baseline	ip.src, ip.proto, ip.srcport	0.00%	100.00%
	ANOVA	frame.len, ip.ttl, ip.proto, ip.dstport	71.48%	100.00%
	Median	frame.len, ip.dstport	72.13%	100.00%
DARPA	Baseline	ip.src, ip.proto, ip.srcport	0.00%	98.39%
	ANOVA	frame.len, tcp.flags	100.00%	96.77%
	Median	frame.len	100.00%	96.77%

Summary

We proposed a low-cost, plug-and-play and self-adapting anti-DDoS technique. We applied the Machine learning algorithms, which are relatively computationally intensive, in the background, to identify features of attack packets. We implemented a customized Bloom filter, which is fast and space-efficient, to block spam traffic based on selected features. We showed a sample implementation of our proposed technique and tested the performance and effectiveness of our technique by experiments.

It is difficult to launch a real attack on the Internet to test our proposed mechanism, but we believed by our experiments result, a better and real industrial SmartBox using our technique can be made to protect our Internet society.