Protect your privacy in Big Electrical Data

Amandeep Singh Virdi

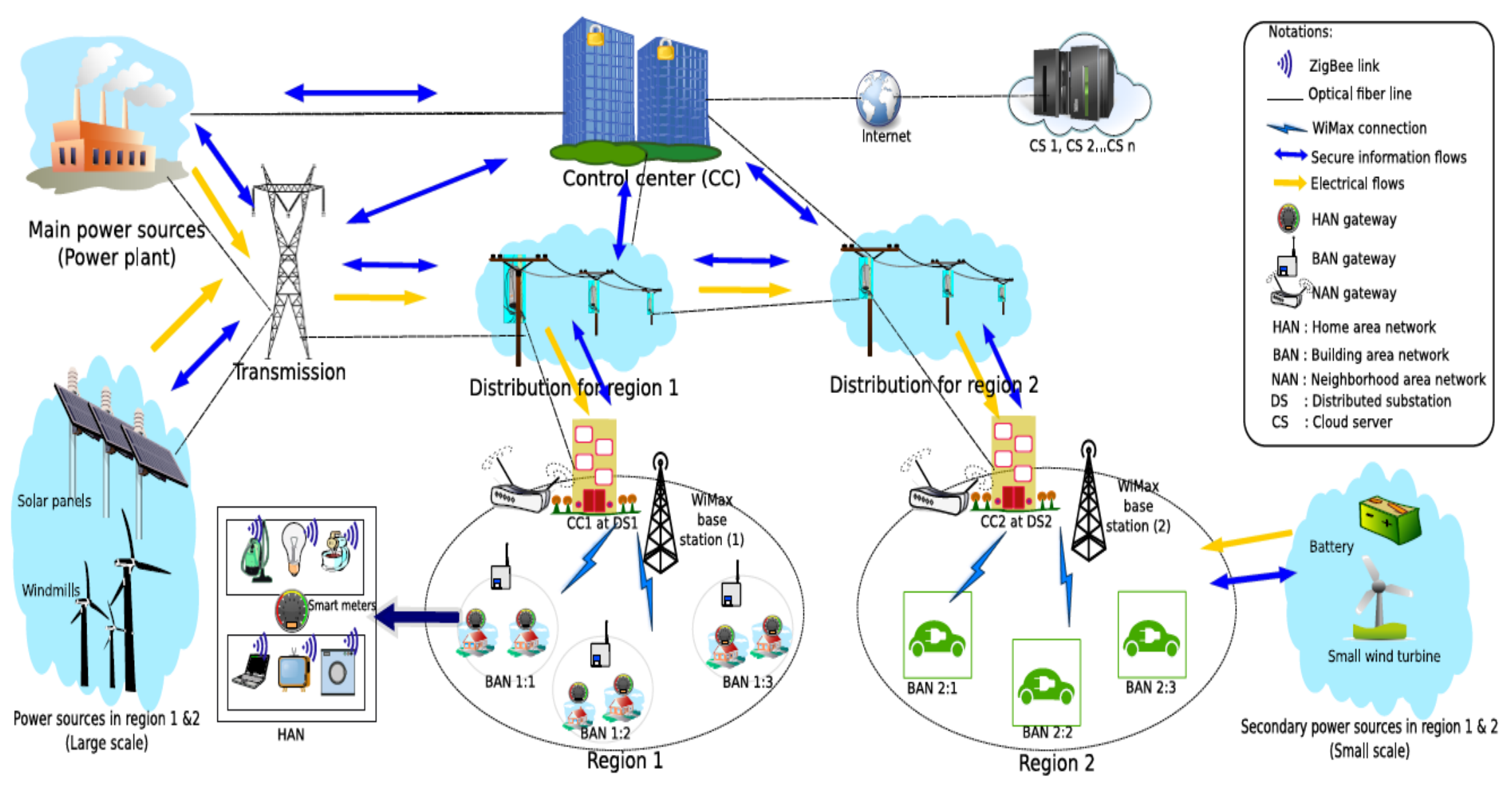
INTRODUCTION

For the last few years, the adoption and deployment of smart meters that measure power consumption and various other parameters at an almost real-time rate has increased at unprecedented rate. This, added to the fact that these smart grid architecture would be using cloud computing technology based on a shared infrastructure platform for the storage and retrieval of the data collected from these smart meters, has lead to serious privacy concern. Privacy Preserving Data Mining schemes in context of smart grids communication is the need of the hour. With this proposed research, we argue that data repositories hosted on a public cloud infrastructure that follow the *StrongBox* model of key management provide a convenient framework to deploy Searchable Encryption schemes in order to mitigate concerns regarding user privacy.

CONTACT

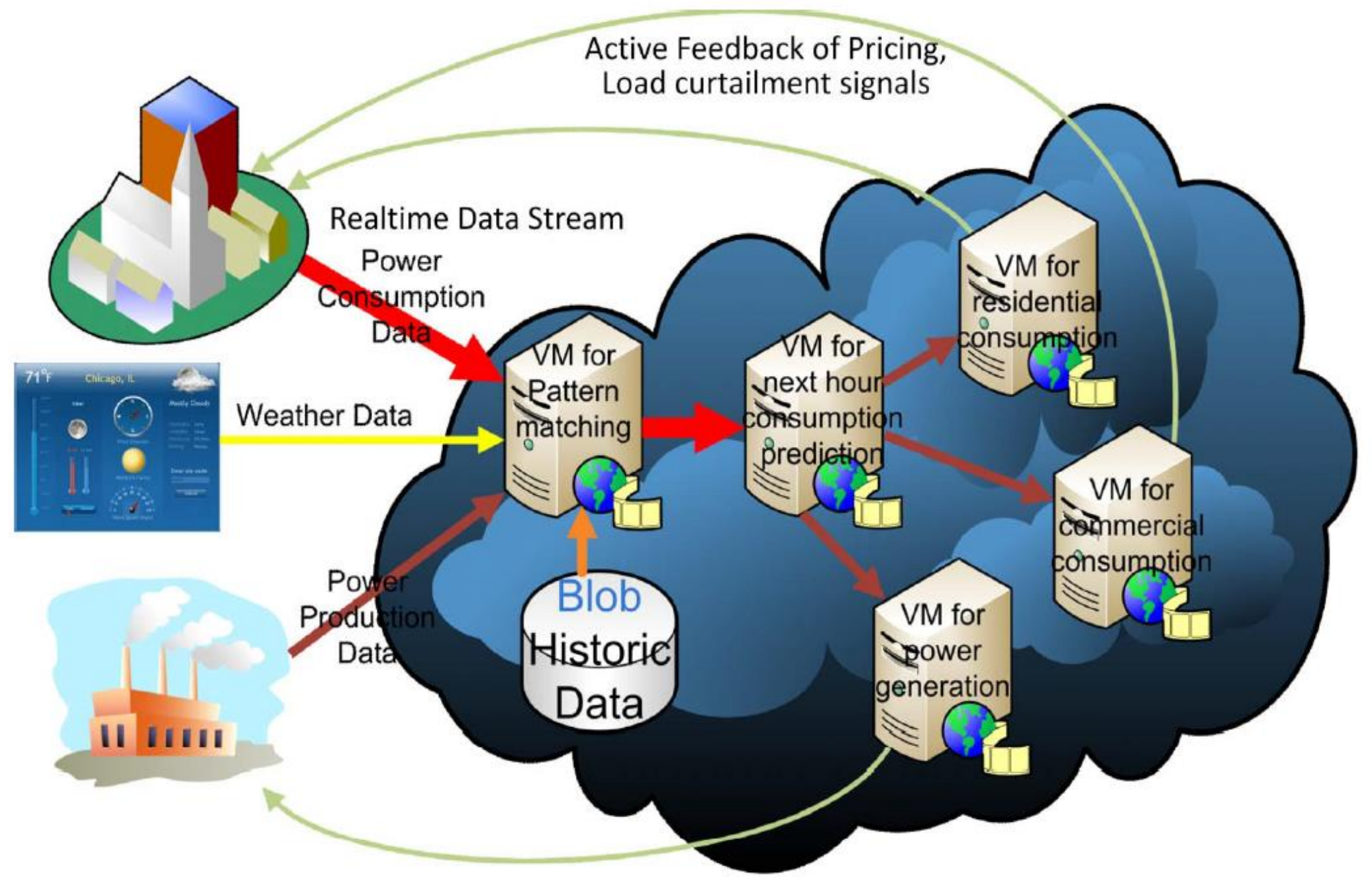
Amandeep Singh Virdi
NTNU Trondheim
amandeev@stud.ntnu.no

SMART GRID COMMUNICATIONS



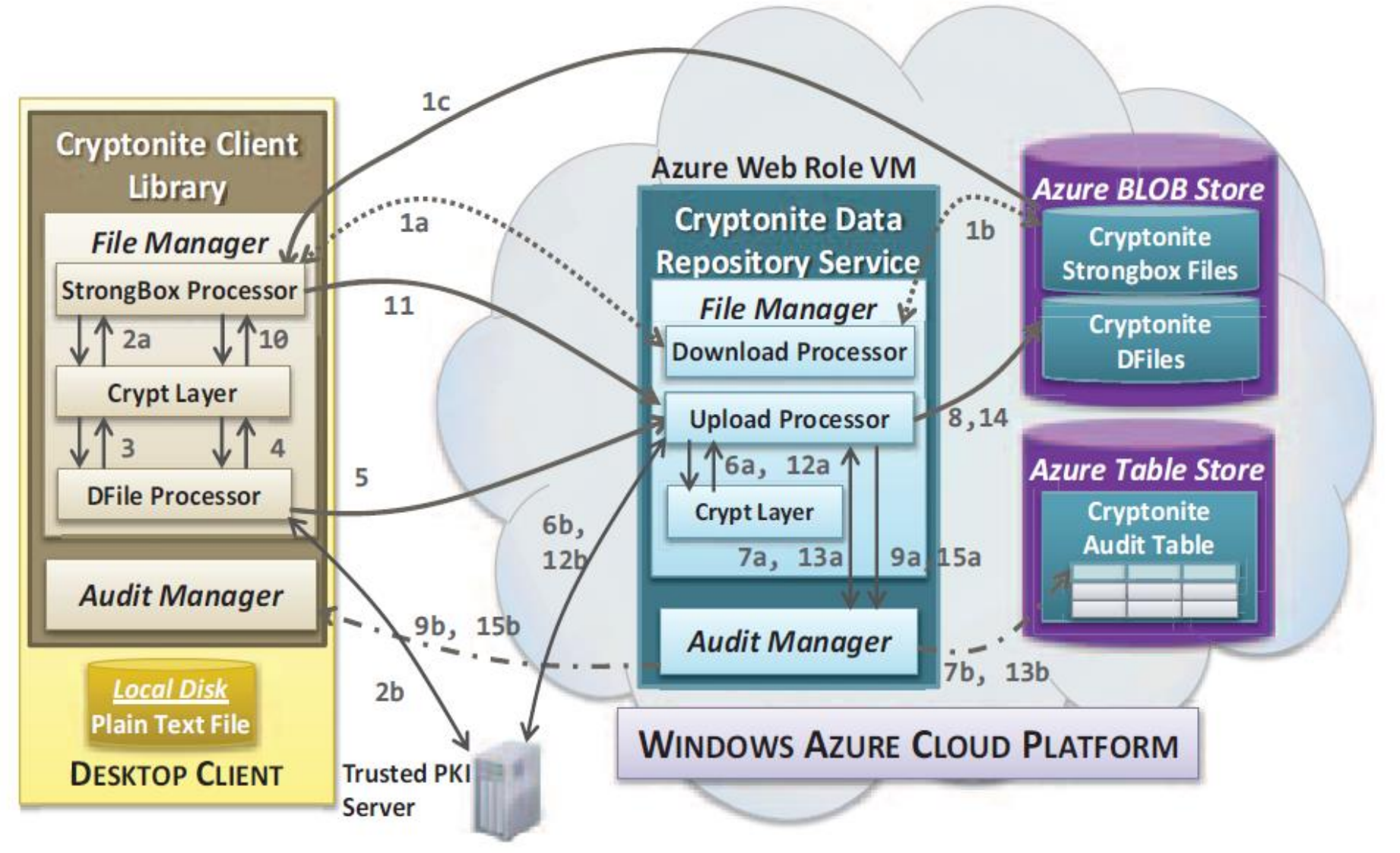
Not only do smart meters record electricity consumption, they can also be configured to measure events such as power quality and meter status. Therefore, smart meters capture two key types of data: Measurement data, which is the regular consumption data; and Events data, which refers to unscheduled alerts and logs corresponding to unusual situations. This, however, generates an unprecedented amount of data arriving in rapidly changing streams and of complex structure and variance, leading to what can be described as a big data challenge.

ROLE OF CLOUD COMPUTING


Active Feedback of Pricing, Load curtailment signals
Realtime Data Stream
Power Consumption Data
Weather Data
Power Production Data
VM for Pattern matching
Blob Historic Data
VM for next hour consumption prediction
VM for residential consumption
VM for commercial consumption
VM for power generation

Smart meters are being deployed across the world at an increasingly high rate. These always-on devices use two-way communication to collect data at an unprecedented rate. Due to their dynamic and always-on nature, it is necessary that they are paired with a similar feature at the other end of the spectrum.
Clouds offer the advantages of scalable and elastic resources paired with an always-on nature as a solution to our problem

STRONGBOX MODEL OF KEY MANAGEMENT



Cryptonite uses several well known and emerging cryptographic and security techniques in its architecture:

- Public Key Encryption
- Digital Signatures
- Broadcast Encryption
- Lazy Revocation
- Key Rotation

It provides a perfect framework to develop privacy preserving schemes especially in context to data mining smart meters but it has some gaps.

SEARCHABLE ENCRYPTION

Since the cloud computing model we are interested in would have a single owner/multiple readers/multiple writers scheme, Symmetric Searchable Encryption is out of question. This is because Symmetric Searchable Encryption only allows that user to search through the encrypted file contents that encrypted it in the first place. Therefore, for efficiently searching through the cloud contents, we need to look at the multiple-user access offered by Asymmetric Searchable Encryption.

PRIVACY PRESERVING DATA MINING

With the advent of big data and various analytic techniques to extract meaningful information from such data, there is a growing concern regarding user privacy. The basic idea for privacy preserving data mining is "to modify the data in such a way so as to perform data mining algorithms effectively without compromising the security of sensitive information contained in the data". Considering how expensive the data distortion dependent scheme of privacy preserving data mining is, we favour the association rule based PPDM.

THE PROPOSED RESEARCH

The goal of our research is to allow for user privacy in the smart grid architecture through access control to the cloud contents. Regardless of the fact that the contents on the cloud may be encrypted and can therefore be considered secure to a reasonable degree, we do require adequate methods for enabling restricted access to it
Since privacy preserving schemes should follow certain realization process, we define two goals: 1) provide user privacy, which is done through a combination of *StrongBox*, searchable encryption and association-rule based privacy preserving data mining; and 2) maintain availability, which is provided by the always-on nature of the cloud computing environment.
There are various challenges we need to consider too. For example, *Cryptonite* has considerable overhead in the sense that it doesn't scale as well as expected. Making sure that our addition to the architecture does not result addition to this overhead would quite the task. Apart from this, we have a two-fold challenge: 1) define "trap door": for example, if we another encryption scheme to secure this mapping of UUID and trap door, we add another layer of encryption/decryption to the architecture; 2) build association-based rules: for example, we need to explicitly specify 'support' and 'confidence' among X and Y and this can be difficult

REFERENCES

[1] Simmhan, Y., Aman, S., Kumbhare, A., Liu, R., Stevens, S., Zhou, Q., & Prasanna, V. (2013). *Cloud-Based Software Platform for Big Data Analytics in Smart Grids*. Computing in Science & Engineering, 15(4), 38-47.
[2] Kumbhare, A., Simmhan, Y., & Prasanna, V. (2012). *Cryptonite: A Secure and Performant Data Repository on Public Clouds*. 2012 IEEE Fifth International Conference on Cloud Computing.
[3] Song, D. X., Wagner, D., & Perrig, A. (2000). *Practical techniques for searches on encrypted data*. Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000.
[4] Xu L., Jiang C., Wang J., Yuan J., Ren Y. (2014). *Information Security in Big Data: Privacy and Data Mining*. IEEE Access, 2, 1149 – 1176.

…and more.