

# Transformation through Information – Secure Big Data for Men and Machines

---

**T. Venkata SubbaReddy, B. Madhava Reddy and B. Sai Venkatesh**  
**SASTRA University, TamilNadu, India**

## **Project Description:**

The current era of data explosion entails the necessity of high efficiency in terms of data capacity [1] and data security [2]. This scenario of Big Data [3] inevitably leads to the technology of Internet of Things (IoT) [4] in the future.

The present project purports to the effective harnessing of nonlinear signal processing [5] principles leading to enhanced security of data without compromising on capacity. The advantage of using nonlinear signal processing lies in the fact that the nonlinearity of a single NMOS transistor [6] is able to provide robust security by generation of chaotic signals [7]. This results in low power dissipation and simplicity of circuitry.

The enhanced secure communication techniques are then studied giving importance to the phase variations in the signal and are then applied to real world information systems [8]. Also, the possibility of introducing such techniques in conventional big data systems such as RDBMS [9] and Hadoop [10] are considered.

After significantly demonstrating the capabilities of the nonlinear signal processing approach in terms of fidelity, capacity and robustness [5], the techniques are extended even further to include an Internet of Things (IoT) based environment [4]. The implementation of nonlinear signal processing techniques to IoT based systems such as RFID are explored [11].

At the final stage, the change in the managerial perspective required to handle the IoT dominated environment is discussed [12]. The business level implications of such a technology shift are studied. This study of IoT is termed as “Management of Things” (MoT).

The principal aim of this project is to provide a feasible, efficient, innovative yet cost-effective solution to the biggest problems of the telecommunication world today – data capacity and data security [1, 2].

This project thus follows from the motto “Transformation through Information” and leads us gently to become effective citizens of a smarter planet [13].

## **The Grand Challenge:**

The current era is the era of information technology and the one phrase that keeps echoing in every nook and corner of the world is ‘Big Data’ [3]. As the data collection and processing capacity of the world as a whole is increasing at an exponential pace, there is a huge demand for attaining very high standards in terms of data throughput, processing capabilities and security, all at the same time [1, 2]. There is always a trade-off between

security and capacity. However, if data increases, high security is required to protect the data which imposes a Herculean demand on the signal processing, storage and communication equipment [14]. The most prominent beneficiaries in this era of Big Data are social networking sites and internet enabled services such as banking where privacy and security in data transmission is most crucial [14]. Even though there are good protection measures currently such as steganography [15], the data storage and transmission capacity offered by such systems is dismal in Big Data terms. Typical examples are recent events such as the Gmail account hacking on September 11, 2014. Google, a very big multinational company with high security protection measures for data privacy, has succumbed to the hacking menace. This shows that there is a great need for security of big data.

Internet of things (IoT), which is also called as “Machine to Machine” or “Machine to Man” or “Man to machine” is a major contributor of big data [4].



Figure 1 The infamous Hacker Logo



Figure 2 Internet of Things

The machines communicate among themselves without intervention of humans. In such situations there is a great possibility of hacking occurrence. So IoT needs much more security and data handling capabilities than any other contemporary technology.

### **Project Proposal:**

The present project purports to providing a solution which offers both high security and high capacity with low power dissipation by using basic principles of signal processing. The bottom line is that a small change in the way signals and data is processed will go a long way in enhancing both the security and data capacity, without causing significant changes to the existing infrastructure.

### **The Project Milestones:**

1. Achieve high security in a basic one to one communication system using principles of signal processing.
2. Elevate the secure information storage/transmission/retrieval system to the level of big data, focusing on contemporary infrastructures such as RDBMS and Hadoop.
3. Extend the twin competencies of high security and high capacity to IoT.
4. Introduce the managerial perspective required to handle a high capacity, high security IoT dominated environment.

## The Project Details – Actionables

The project milestones provided above provide a convenient way to structure the project into four stages, each developing from the previous stage and adding its own contribution to the solving of the grand challenge.

### Stage 1: Enhancing Security in a One-to-One Communication System

A nonlinear signal processing approach will greatly benefit handling and manipulating large amounts of data by effectively harnessing their phase relationships collectively which is the hallmark of Big Data. The present project effectively harnesses the nonlinearity present in a semiconductor N-MOSFET to increase the signal complexity so as to give a pseudo random appearance [6]. This forms a security key in the presenting work. By giving two sinusoidal signals as inputs to the gate and the drain of the MOSFET and suitably selecting the frequencies of the gate and drain signals, proper disharmony between the signals is achieved leading to chaos [7].

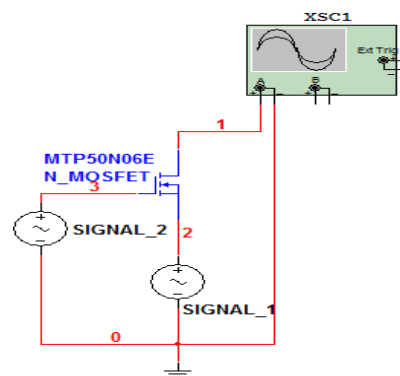


Figure 3 Schematic for Generation of Chaos

The ratio of frequencies at drain and source play a main role in generation of the chaos pattern and it is so sensitive to the initial conditions such that even a 1% difference in one of the frequencies renders it impossible to generate the same chaos pattern. The main advantage using this method is that the power dissipation is very low and the hardware implementation is very simple. The chaos generated is used as a carrier to modulate the incoming data. At the retrieval stage, the user should have the same frequencies in order to generate the same chaos pattern to demodulate the received signal.

The above mentioned process has been implemented in order to ascertain the validity of the hypothesis. The details are given in the Implementation section of this report.

**Benefit to the Society:** The tools and techniques developed at this stage enable to construct effective one to one communication systems using the basic principles of signal processing and chaos theory with the added advantages of low cost and low power dissipation, owing to the simplistic circuitry involved. By combining the basic chaos enabled modulation techniques outlined here with the conventional multiplexing and multiple access technologies such as OFDM and CDMA [16], power efficient and secure communication systems can be developed.

**Skills and Learning Intent:** The generation of chaos using the above mentioned schematic provides ample opportunities for exploring how the interaction of signals in a nonlinear system such as a transistor results in a change in the complexity and stability of the system. A good understanding of the transistor characteristics such as the operating point, and the harmonics generated due to the nonlinearity is also obtained. Finally, the application of

purely theoretical concepts such as Chaos Theory to the real world will be learnt [7]. The skills required for this stage would include the basic knowledge of transistor operation with good clarity on the various linear and nonlinear regions of the output and transfer characteristics, coupled with a basic understanding of signals and systems in general, and about the frequency concept and Fourier Spectrum in particular [17].

### **Stage 2: Incorporating Chaos enabled Security in Big Data Systems**

The high level of security achieved in Stage 1 is used as the basis for this stage. Here, the focus is on increasing the data storage capacity without compromising on the security.

Most of the contemporary data management in large corporations such as Facebook, Banking sector etc. is carried out using Relational Database Management Systems (RDBMS) [9] or using contemporary “cloud specialized” [18] Distributed File Systems such as Hadoop (HDFS) [10]. The key factors regarding these technologies are the relationships between entities which enable efficient storage of data. While in RDBMS it is the relationships between various fields and entities of the stored data, in Hadoop it is the relationship in the indexing and the map-reduce procedures.

The output data modulated with chaos of Stage 1 is thus a perfect candidate for optimizing data capacity in RDBMS and HDFS. The perspective is that the phase relationships obtained in Stage 1 form the basis for the entity relations for this stage.

As a local prototype, small scale example of this, the Stage 1 output can be embedded in an image and the various security and capacity based capabilities can be analysed in order to validate the robustness of the modulated data. This prototype thus pertains to a Steganographic system where a phase distribution is mapped onto a pixel distribution [8].

**Benefit to the Society:** The implementation of the techniques outlined in this section can ensure the security of user data in critical systems such as banking and transaction based sites and also in the social networking platforms. By expanding the scope of “Secure Big Data” to media such as images and video, the concept of cloud computing can be revolutionized and major disasters such as the Gmail hack of 11/9/2014 can be prevented.

**Skills and Learning Intent:** The learning intent of this stage mainly purports to extending the results of the previous stage to involve the data in concern. The relationship concept, which is the main highlight of RDBMS and HDFS can be understood clearly, and an effective mapping can be made from the phase relationships of the previous stage to the data relationships of the present stage. The skills required would then be a basic understanding of the concept of data – bits and bytes, and the storage systems (such as pixels in images), coupled with an elementary understanding of the concept of database management, fields and entities.

### **Stage 3: Extending the Secure High Capacity Techniques to Internet of Things**

In the near future the Internet and wireless technologies are predicted to connect different sources of information such as sensors, mobile phones and cars in an ever tighter manner. The number of devices which connect to the Internet is exponentially increasing. This can be clearly witnessed by the fact that by 2013, 1 Trillion devices have been connected to the internet. A wide adoption of the Internet of Things model will result in the generation of a large amount of data requiring storage, processing and retrieval. This challenge requires the ubiquitous presence of cloud computing since a tremendous amount of data will be generated by IoT. The million dollar question then is: “Who owns the data?”.

Will the data from the car go to the car manufacturer, local law enforcement, a person’s insurance company, and eventually into court if he/she meets with an accident?

Could the service providers collect information and sell it? A second consideration is the ability to hack the IoT objects. The hacker could read the information, manipulate it to the victim's disadvantage, steal the device identity, or change GPS coordinate information.

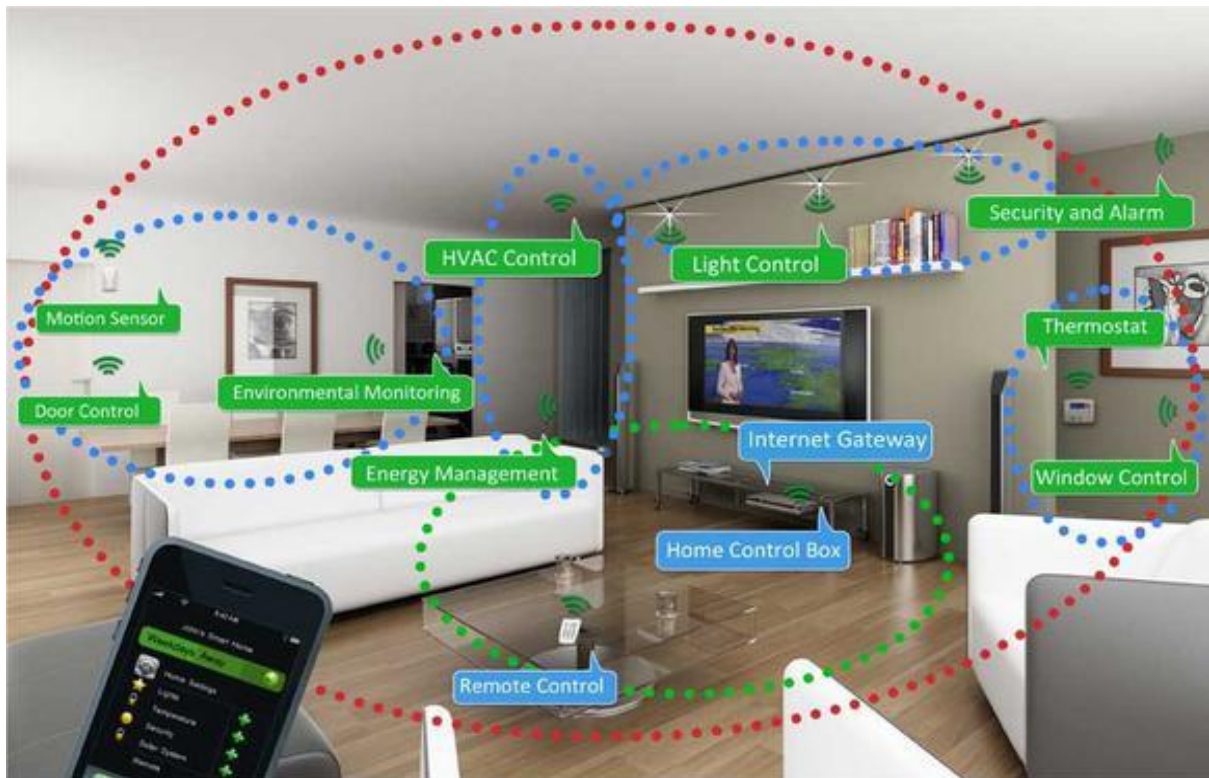


Figure 4 An IoT enabled Smart House using Chaos based Secure Big Data

As an example, consider the Google self-driving car, which uses the internet while in motion, for both recording of vital statistics as well as for navigation purposes. If someone hacks the signal and controls it by themselves, then what would be the resulting impact on the possibility of accidents? This presents new security challenges to both the owner of the object as well as the collector of the data. To overcome these, a nonlinear signal approach can be used. Consider that the car is communicating using chaos signal as carrier whose frequency ratios remain constant only for a particular time period (similar to frequency hopping principle in CDMA) [16]. A hacker cannot hope to predict which frequencies are used by the devices at a particular instance, which means that this approach provides high security.

Thus the methods outlined in Stages 1 and 2 can be put to effective use in this stage to enhance both the capacity and security in the IoT dominated world.

At the local level, a prototype IoT setup such as a Radio Frequency Identification System [11] can be constructed incorporating the security and capacity enhancement techniques proposed in stage 1 and 2. This can then be used as a lab or classroom based demonstration session, where the various frequency ratios can be explored and the ensuing security improvement can be characterized.

**Benefit to the Society:** The implementation of the techniques outlined in this section will go a long way in enabling a highly reliable, high capacity, highly secure Internet of Things based environment. Of particular mention here is the drastic change in the trend of automation IoT is bound to make. The world will proceed from Smart Technology to Smart Devices to Smart Homes and eventually to a Smarter Planet.

**Skills and Learning Intent:** The present stage of the project will provide ample learning insights on the working and control of present day automation “things”. As a consequence of this understanding, the vulnerabilities and security based necessities of IoT will be understood clearly. The skills required here are basic understanding of automation technologies, “smart sensors”, coupled with a basic understanding of any one real time IoT system (RFID would be a good starting point).

**Stage 4: A Shift of Paradigm – From ‘Internet of Things’ to ‘Management of Things’**

The motivation for this stage stems from the fact that the successful implementation of the techniques outlined in the previous stages is bound to enhance the security of big data systems. This progress will inevitably result in a large number of corporate organizations hopping on the IoT bandwagon. This will eventually call for a shift of paradigm in the managerial and corporate levels. The Business Strategy [19] will have to be tuned in accordance with the developing trends to ensure sustainable competitive advantage.

As a starting step of this stage, the “6 M’s” of Management [12] are enhanced and extended to a more comprehensive model including 9 Stages, as shown below:

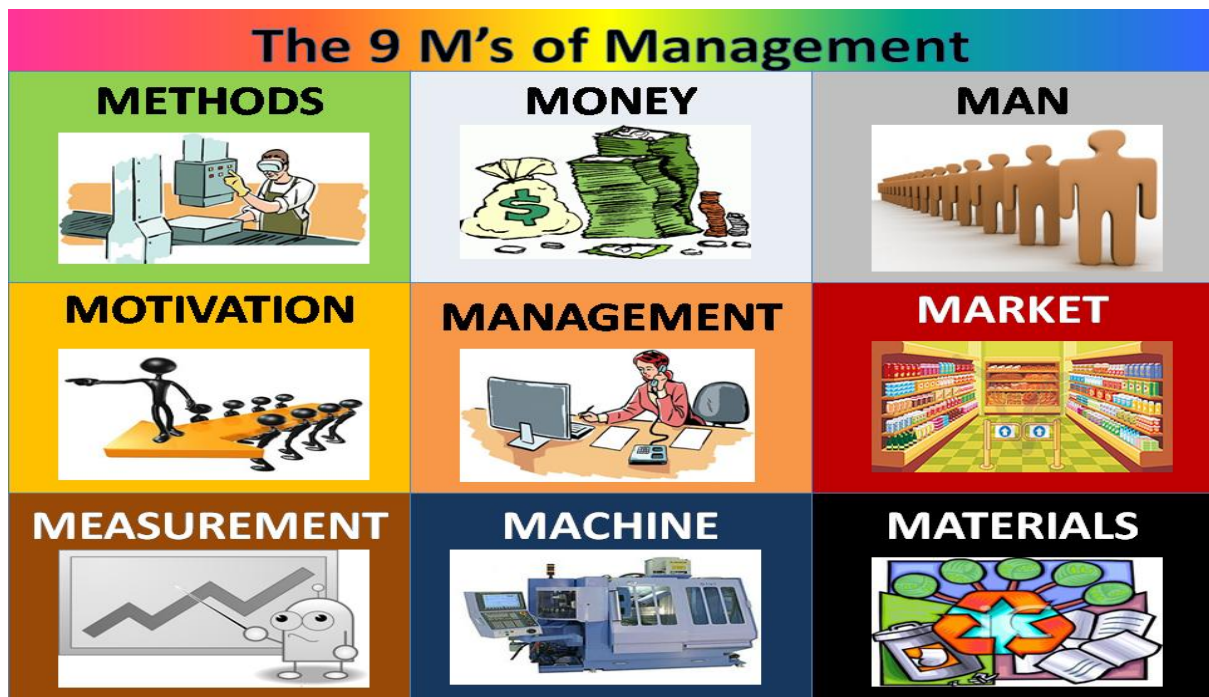


Figure 5 The 9 M's of Management

The understanding of how the IoT dominated environment brings about a change in the strategic perspective of each of the 9 M's [20, 21], leading to a different approach to the framing of policies and decisions [22], will form the major highlight and learning experience of this stage.

The changes brought about in the 9M's due to dominance of IoT with enhanced security and capacity can be tabulated as follows:

<b>Dimension</b>	<b>Effect of Enhanced IoT Technologies</b>
Management	Tuning of strategies to prevent ethical issues (privacy, safety etc).
Man	Decrease of Labour and increase of machine-maintenance tasks.
Motivation	Challenging tasks and R&D opportunities in development of IoT.
Materials	Reduced reliability on materials due to simpler circuits.
Methods	Enhancing Efficiency using Signal and Information Processing.
Money	Lower cost due to circuit simplicity and low power dissipation.
Measurement	Increased speed and precision albeit with more maintenance.
Machines	Increased usage of machines mitigates certain operational risk factors.
Market	Heavily dependent on right positioning and market segmentation.

**Benefit to the Society:** As the dominance of Things becomes more pronounced with the advent of IoT, ethical issues related to privacy and safety become inevitable concerns. Also, the approach to the managing of such IoT equipments and optimizing their use in order to maximize shareholder's wealth becomes crucial [20]. The MoT perspective helps address such issues.

**Skills and Learning Intent:** The MoT developed in this stage will provide useful insights to the students/staff on applying core management concepts and policies to effective use in a near-future situation. The "industry awareness" on such ethical issues can be obtained. The skills required for this stage include basic understanding of the ethical issues that may arise out of an IoT dominated world.

### **Materials and Resources:**

Hardware:

- N-MOSFET - MTP50N06E
- AFO
- Cathode Ray Oscilloscope (CRO)
- Connecting wires
- Bread Board/ Printed Circuit Board

Software:

- MATLAB
- Nonlinear Analysis Toolkits

Others:

- Internet Facility
- Memory Storage Facilities

## Project Plan and Budgeting:

The major stages in the proposed project and the steps and corresponding time budgeting for each step are tabulated below.

STAGE	STEPS	TIME BUDGET
Enhancing Security in a One-to-One Communication System	Characterize the transistor (DC, AC)	15 minutes
	Determine the knee point	15 minutes
	Setup the Circuit to generate chaos	15 minutes
	Explore different frequency ratios	30 minutes
	Characterize the generated chaos	30 minutes
	Transmit it using proper modulation	30 minutes
	Set up the receiver with exact ratio	30 minutes
	Demodulate the signal	15 minutes
	Characterize the fidelity	15 minutes
	Characterize the error due to wrong ratio	15 minutes
	<b>TOTAL BUDGET</b>	<b>3.5 hours</b>
	Incorporating Chaos enabled security in Big Data Systems	Select the medium (ex. Image)
Choose the embedding algorithm (LSB etc)		15 minutes
Embed the modulated signal in the image		15 minutes
Characterize the histograms		15 minutes
Nonlinear analysis of the embedded image		30 minutes
Implement receiving and de embedding		15 minutes
Characterize the sensitivity of embedding process		15 minutes
Explore the architectures of RDBMS and HDFS		30 minutes
Frame algorithms for embedding modulated signal		60 minutes
Implement small scale prototypes		60 minutes
<b>TOTAL BUDGET</b>		<b>4.5 hours</b>
Extending the Secure High Capacity Techniques to Internet of Things	Explore the concept of IoT	60 minutes
	Explore potential avenues for chaos induced security	60 minutes
	Select a small scale implementable model	60 minutes
	Design the circuit/system	1 Day
	Implement the proposed technology	7 Days
	Characterize	60 minutes
	Summarise the results	60 minutes
	<b>TOTAL BUDGET</b>	<b>8 Days, 5 Hours</b>
From Internet of Things to Management of Things	Explore the 9M concept	60 minutes
	Understand the Managerial Stance	60 minutes
	Formulate a futuristic IoT corporate environment	60 minutes
	Frame a Case Study for Analysis	60 minutes
	Analyze perspectives on each of the 9 M's	60 minutes * 9
	Tabulate the results	60 minutes
	<b>TOTAL BUDGET</b>	<b>14 Hours</b>
<b>GRAND TOTAL BUDGET</b>		<b>9 Days, 3 Hours</b>



### **Implementation Results:**

As a proof of concept for stages 1 and 2 of the proposal, we have implemented a prototype based on the nonlinear signal processing concept.

#### **Implementation of stage 1:**

According to the circuit schematic shown in fig. 6, the experimental setup is as shown in the fig.5. Two sinusoidal signals are given as inputs to the gate and the drain of the power N-MOSFET MTP50N06E [6]. The amplitude of the drain and the gate signals are chosen as 5V<sub>rms</sub> and 2V<sub>rms</sub> respectively. These values are chosen so as to bias the transistor in the knee region. This region corresponds to a lower power dissipation of around 90mW as opposed to 240mW outside the knee region.

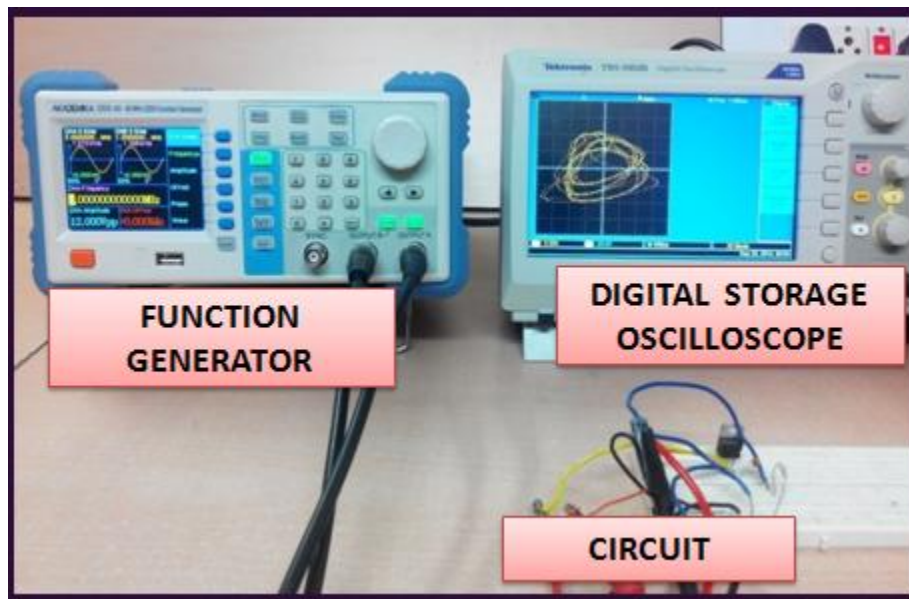


Figure 6 The Experimental Setup

By suitably selecting the frequencies of the gate and drain inputs, proper disharmony between the signals is achieved leading to chaos. In order to demonstrate this keep both signal frequencies as 1MHz, then as shown in fig.7 for the 1:1 case, the sine signal is deformed into pulse (wave shaping), because the transistor is operating in nonlinear region. Now the gate signal frequency is set to 1MHz and the drain signal frequency is set to 2MHz so that it is twice that of the gate signal frequency. In this case, the nonlinearity of the MOSFET gives rise to periodic doubling which is the starting point of instability. Similarly three loops are formed for frequency ratio 1:3, four loops are formed for 1:4 and six loops for 1:6 respectively. When the drain frequency is set to a non-integral multiple of the gate frequency such as 2.9MHz, chaos is observed. This “Road to Chaos” [7] is shown as follows:

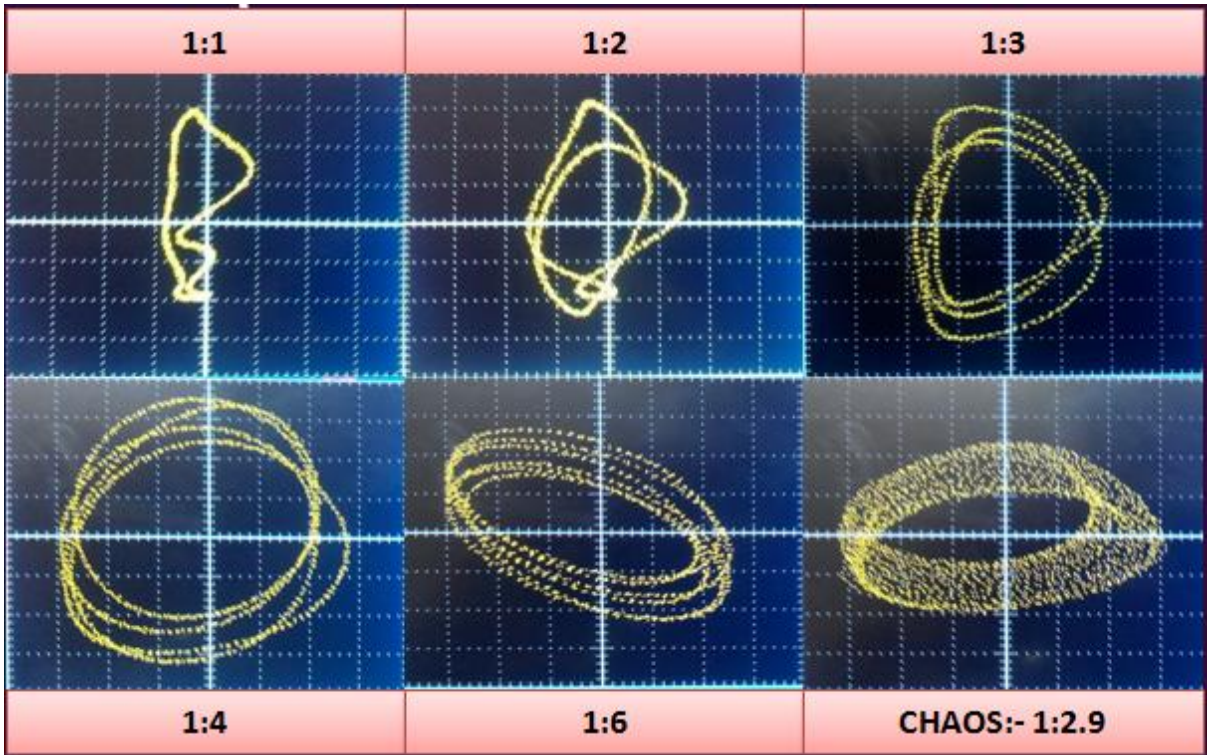


Figure 7 The "Road" to Chaos

In order to better understand the properties of the generated chaotic signal, the phase portrait, waveform and spectrum are plotted as follows:

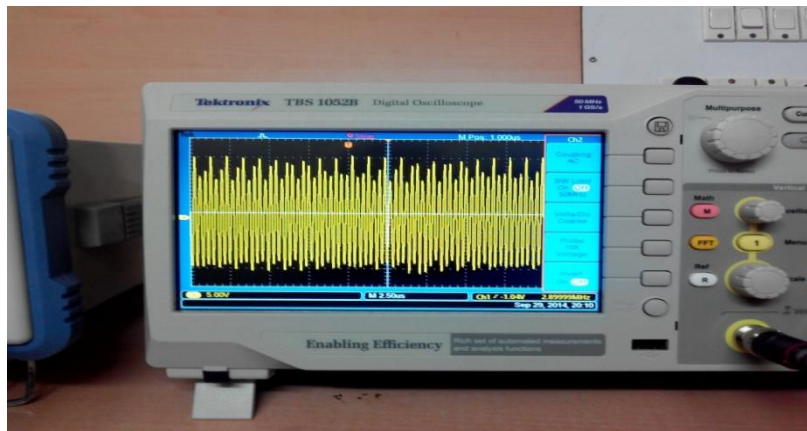


Figure 8 Voltage Waveform



Figure 9 Fourier Spectrum

**Characterization:** Two parameters are used to characterize the chaos obtained. The first is the Lyapunov exponent [23] which is a measure of sensitivity or complexity. In our case we achieve a reasonably high value of 9.46. The other parameter is Kolmogorov Entropy [24] which is a measure of difficulty in reproducing and information capacity. A significantly high value of 8 bits/symbol is obtained indicating that the chaotic carrier signal serves to carry a reasonably large amount of information, thus testifying to the concept of ‘Big Data’. The other examples of chaos generated using various frequency ratios are as follows:

To characterize the sensitive dependence on frequencies as mentioned in stage 1 of proposal, chaos generated with 1:2.8 and 1:2.9 are overlaid and compared as can be seen in fig, after first few cycles, drastic change in the signals can be observed, this ensures high security.

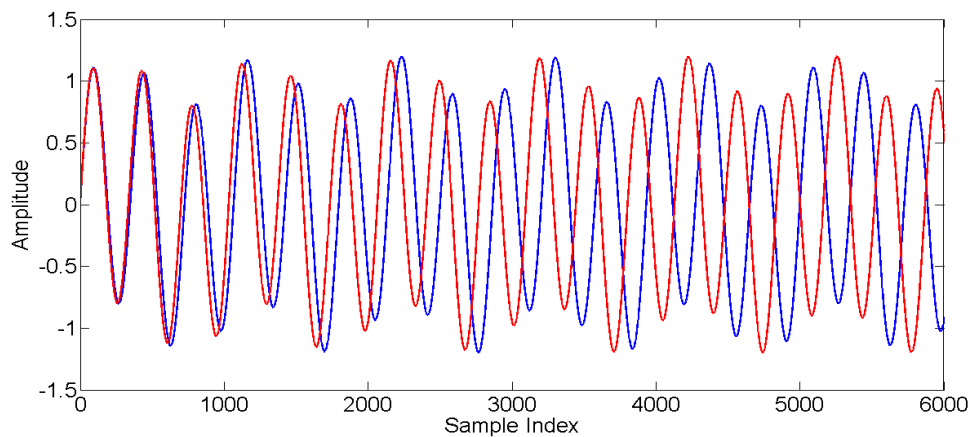


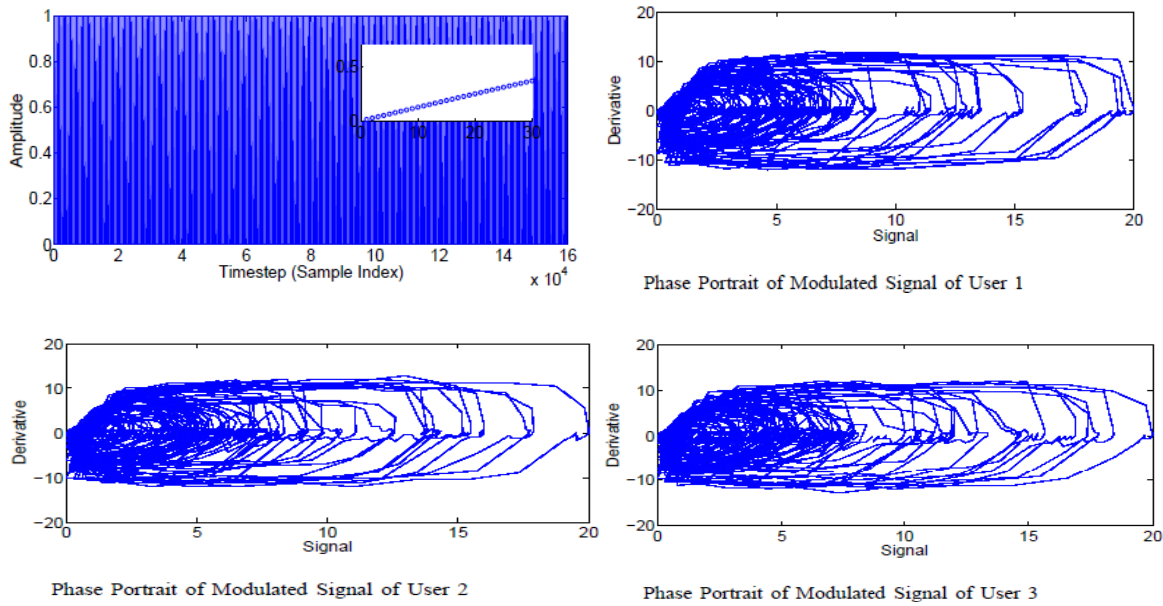
Figure 10 Comparison of Chaotic waveforms of ratios 1:2.8 (blue) and 1:2.9 (red)

### Implementation of stage 2:

This Phase is a proof of concept implementation of Stage2 in proposal. Specifically, one to one communication is dealt with here. Core essence of this stage is that we use the relationships between data to achieve high capacity in storage and transmission. The medium of storage and transmission is image, where pixels and pixel relationships are used to increase capacity. Extreme sensitivity to initial conditions renders chaotic signals perfect choices as the carriers for such data transmission. While on one hand, conventional steganographic techniques rely on non-raster customized embedding patterns for achieving robustness, the steganographic technique proposed in this work, on the other hand, relies on the nonlinear phase relationships between various pixels to securely embed information, thereby enhancing the robustness in security. This spatial phase distribution forms an ornamental pattern by itself, as seen in the phase plane portrait of the chaotic carrier signal. Embedding this onto the image is akin to embedding an innocent image (phase pattern) onto another innocent image (pixel pattern). The chaotic carrier has a pseudo random appearance and hence forms the ‘key’ in the present work.

The next step is to modulate the generated chaotic carrier using the message signal. The modulation used here is simple Amplitude modulation [25]. In the case of analog message signals it can be performed using a transistor or diode based mixer circuit. For digital messages, a suitable software code can perform the modulation. In the present work, the modulation operation is mathematically represented by multiplication. In order to completely utilize the data embedding capabilities, 3 different messages are generated. The three messages come from different users and are in different frequencies. These messages will be embedded into the red, green and blue data of the image respectively. In the present work, the

three message signals are modelled by three sinusoids of the frequency ratio 1:2:3. As an illustration, the message signal of User 1 is shown as follows:



In the present work the carrier image is chosen as Baboon of 400x400 resolution(.png), which can thus hold 160000 samples. The scaling factor is selected as 2. To embed the modulated signal, the image is first split into three arrays of red, green and blue. For each array, each sample of the corresponding modulated signal is scaled down by a factor of around 1 percent of the maximum pixel value intensity 255, thus resulting in a scaling factor of 2. These scaled down samples are added to the image pixel values. After repeating this process for the other two colours, the components of the image are re-joined to give the embedded image. The original carrier and the embedded steg-image is shown. The chaotic properties of the embedded image are studied.

In order to better understand the changes in the carrier image induced due to the embedding process, the red, green and blue histogram of the carrier and embedded images are plotted. The embedded image is now transmitted through the internet or equivalent transmission channel.



Figure 11 Carrier and Embedded Images

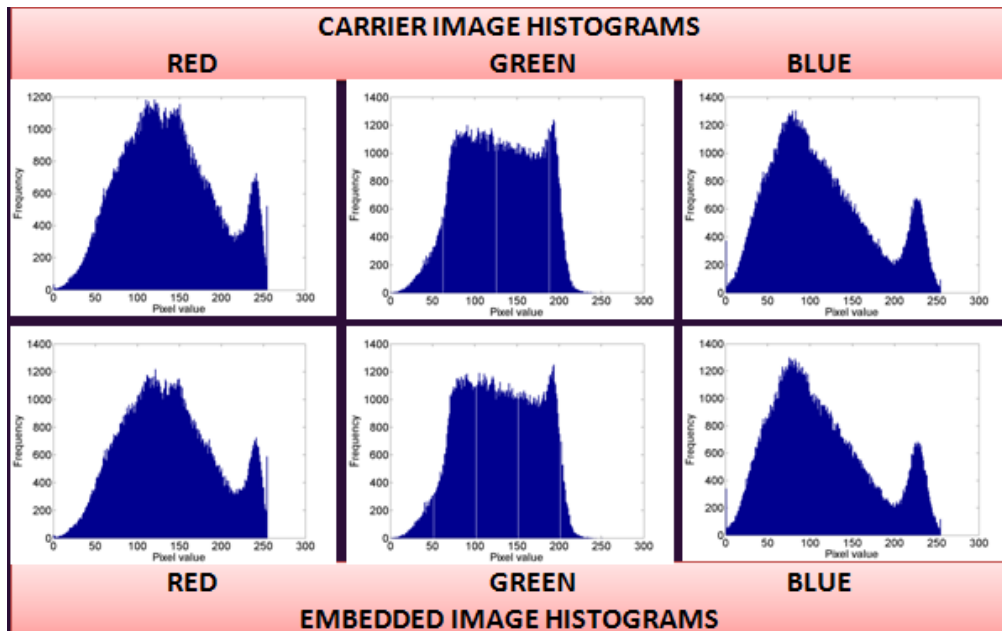


Figure 12 Comparison of Histograms

A similar concept can be implemented by considering the phase relationships in the Hadoop platform. Thus, the concept of RDBMS can be extended to a more chaotic flavour, utilizing the phase relationships to enhance capacity.

The implementation for the stages 3 and 4 can be carried out as per the project plan given earlier.

**Acknowledgement:**

The bottom line of the project is that a small technological change from a linear signal processing approach to the nonlinear signal processing approach yields a number of significant advantages including,

1. Simple Circuitry
2. Low Power Dissipation
3. High Fidelity and Low Distortion
4. Enhanced and Robust Security
5. Enriched Data Capacity

The present project proposal thus highlights the various steps in achieving these golden advantages pertaining to the world of Big Data and Internet of Things.

This project is thus in accordance with the statement made by the Honourable Prime Minister of India, Shri Narendra Modi: “...since I have begun life as a small man, I like to do small things... and big things for small people”.

**Acknowledgement:**

The team wishes to acknowledge SASTRA University for providing research facilities. The implementation was supported by the Research and Modernization fund of SASTRA University, India.

## **References:**

- [1] D. Salomon, D. Bryant and Giovanni Motta, *Handbook of Data Compression*, (Springer, California, 2010).
- [2] J. Jacobs and B. Rudis, *Data-Driven Security: Analysis, Visualization and Dashboards*, (Wiley, Indiana, 2014).
- [3] X.Wu, X.Zhu, G.Q.Wu and W.Ding, *Data mining with big data*, IEEE Trans. on Knowledge and Data Engineering **26**,97-107(2014).
- [4] A.McEwan and H.Cassimally, *Designing the Internet of Things*, (Wiley, 2013).
- [5] K. E. Barner and G. R. Arce, *Nonlinear Signal and Image Processing: Theory, Methods, and Applications*, (CRC Press, U.S, 2003).
- [6] M. E. Inchiosa, A. R. Bulsara, A. D. Hibbs, and B. R. Whitecotton, *Signal Enhancement in a Nonlinear Transfer Characteristic*, Phys. Rev. Lett.**80**,1381(1998).
- [7] E.Bilotta and P.Pantano, *A gallery of Chua attractors*,(World Scientific, Singapore, 2008).
- [8] M. Cross and H. Greenside, *Pattern Formation and Dynamics in Nonequilibrium Systems*, (Cambridge University Press, Cambridge, 2009).
- [9] Cognos Inc. Information: <http://www.answers.com/topic/cognos-incorporated-usa>.
- [10] T. White, *Hadoop: The Definitive Guide*, (O'Reilly, 2010).
- [11] V. Hunt, A.Puglia and M.Puglia, *RFID: A Guide to Radio Frequency Identification*, (Wiley, 2007).
- [12] A.Adekola and B.S.Sergi, *Global Business Management*, (Ashgate, 2012).
- [13] E.Schouten, *IBM SmartCloud Essentials*, (Packt, 2013).
- [14] J.M.Guo and T. N. Le, *Secret Communication Using JPEG Double Compression*, IEEE Signal Processing Letters **17**,879-882(2010).
- [15] N.Hopper, L.VonAhn and J.Langford, *Provably secure steganography*, IEEE Trans. on Computers **58**,662-676(2009).
- [16] T.Rappaport, *Wireless Communications: Principles and Practice*, (Prentice Hall, 2002).
- [17] S. Marshall and G. L. Sicuranza, *Advances in Nonlinear Signal and Image Processing*,(Hindawi, New York, 2006).
- [18] K.Jamsa, *Cloud Computing*, (Jones and Bartlett, 2013).
- [19] G.V.S.Sekhar, *Business Policy and Strategic Management*, (I.K.International, 2009).
- [20] R.J.Elliot, W.C.Hunter and B.M.Jamieson, *Financial Signal Processing*, (Federal Reserve Bank, 2000).
- [21] K.Hegar, *Modern human relations at work*, (Cengage, 2011).
- [22] A.L.Rice, *The Enterprise and its Environment: A system theory of management organization*, (Psychology Press, 2001).
- [23] M.T.Rosenstein, J.J.Collins and C.J.DeLuca, *A practical method for calculating largest Lyapunov exponents from small data sets*, Physica D, **65**, 1993.
- [24] R.M.Gray, *Entropy and Information Theory*, (Springer, 2011).
- [25] L.S.Packer, *Graphical Phase plane methods in dynamics and vibrations*, Cornell, 1956.