Redbooks
ibm.com/redbooks

Smart Grid Cyber Health Assessment in a Big Bad Data World

Dr. Lisa Sokol

Dr. Steve Chan

IBM

Redguide

Introduction

We expect our electricity (for heat, air conditioning, and lights), water, and other utilities to be available whenever we want them. But our everyday essentials have become the target of our adversaries. Exacerbating the situation, when a part of the grid fails, we don't know whether it is from natural causes or the actions of bad actors. Regardless, the failure needs to be fixed.

Despite our diligence in the creation of new electric grid standards, these standards are not sufficient to address the urgent cyber threats and challenges that critical infrastructures now face. The lack of electric grid standard granularity can result in a failure. There are also several other factors at play, including these:

- Complexity and sophistication of a smart grid
- Large number of electric grid components
- Wide variety of involved actors
- Lack of time stamp standardization among grid components

An assortment of smart grid standards, such as the International Electrotechnical Commission (IEC) and the Institute of Electrical and Electronics Engineers (IEEE), allow a viable approach vector to insert disinformation into the grid via a myriad of threat vectors.

Innovative analytic approaches are required for the detection of one type of threat, known as misinformation or disinformation or *astroturfing*. This paper proposes a strategy that combines contextual analytics for version verification (current component state, component history, graphical knowledge of grid connectedness, a decay function for impact of other components), predictive modeling, and a computing model assessment using edge computing.

This IBM® Redguide™ publication describes the various issues that can impact the energy grid and provides examples of grid failures. It discusses the value and possibilities of a smart grid and how analytics can play a key role in the overall solution. It also introduces the combination of Irwin technology from Mehta Tech, Inc. and IBM Watson™ cognitive system, which form a technology stack to monitor the electric grid.

Components and attributes of the electric grid

A physical electric power grid has three basic sets of components:

- Supply
- Transmission and distribution (T&D)
- Demand

Within each of these sets, there are many more subentities. For example, a power and distribution infrastructure could include generators, substations, distribution substations, transformers, digital fault recorders (DFRs), Phasor Measurement Units (PMUs), flywheels, and Battery Energy Storage Systems (BESSs). In general, electric grid networks support the transfer of power between grid components, and grid networks can vary in size (from small to large), density (from sparse to a huge number of nodes), and topology (from those with highly modular structures to those with highly overlapping structures).

A generic physical electric power grid system (a set of entities) is joined together by *nodes* and *edges* (also known as *links* or *relationships*). An *entity* refers to something that is real and can be a thing, such as a DFR. Entities have *attributes*, which are values that are specific and closely related to that entity. Examples of DFR attributes include voltage, status of key devices, and breaker position. The attributes for an entity, such as a DFR type, can be of various kinds:

- Permanent (for example, make and model)
- Temporary (for example, location or configuration)
- Exclusive (for example, asset numbers)
- Non-exclusive (for example, swing recorder functionality, PMU capability)

What makes an electric grid smart

Smart grids are networks that have computerized power transmission and distribution infrastructures. Smart sensors and meters along energy production and transmission and distribution pathways generate large amounts of granular, real-time, streaming data. The grids support the generation and two-way transfer of petabytes of heterogeneous data that pertains to the production and consumption of electricity. Grid management systems, including computer-based remote control and automation, use the data to enable smarter operational decisions. The successful use of these systems results in gains for grid reliability, efficiency, flexibility, and resiliency.

Different smart grid networks can be banded together to create what is commonly called a *system of systems*. Each network group has a set of resources that it applies to electric generation, transmission, and distribution tasks. Each of these systems can pool their resources and capabilities to create more complex systems. This approach has the effect of creating a high-level entity that has more flexibility, functionality, performance capability, and resiliency.

The communication patterns between nodes and networks can be described as sets of binary relationships between the nodes. The level of interaction between the different nodes or networks varies with smart grid function and location within the network. The amount and frequency of the interaction between entities are dictated by an entity's role within the smart grid network.

The nodes that are adjacent to a specific entity tend to have a high level of interaction. Typically, the strength of interaction between a node of interest and other nodes (which could also represent an entire network as part of the *system of systems* paradigm) is a function of node functionality and the distance between nodes. To some degree, the interaction strength decays as the distance increases. However, if a smart grid node suffers a catastrophic event (for example, Category 3+ storm event), it has the potential to have a negative impact on all of the other grid nodes, regardless of distance. The exact strength relationship can be assessed by using a dynamic form of network science and graph analytics.

Of particular note, system health is akin to an epidemiology model, wherein an individual's health is a function of the health of the individuals with whom that person interacts. The closer the relationship and the more serious the contagious disease, the more likely that an individual's health might be compromised. The referenced closeness can be assessed via an algorithmic and heuristic determination, such as by using Irwin technology and IBM Watson solution.

In fact, the epidemiological concept can be applied to an electric grid in the form of a Bak-Tang-Weisenfeld Model[1] of *positive influence dominating sets* (PIDS) and *negative influence dominating sets* (NIDS). The concept of PIDS is taken from graph theory.[2] A PIDS is a network composed of entities that might share the same purpose to provide a powerful medium for disseminating data and influence.

An influence diagram[3] consists of a network of nodes that represent events that are joined by directed links, which, in turn, represent direct influence between events. Events have both *state* and *strength*. Each event can be modified by the influence of other events. Influence can be passed along links between events over time with differing degrees of certainty. Influence becomes further complicated by cycles and feedback loops within a network graph.

NIDS is a formalism that can be used to support the reasoning about certain actions and the decision-making processes for smart grid agents. The Bak-Tang-Wiesenfeld model can use the dominating set and influence dominating sets to create a model that describes real phenomena, focusing on transitions near critical points. This type of model can be invaluable from an emergency preparedness planning standpoint.

The size of these smart grids (even microgrids), their pervasiveness, and their critical roles within the critical infrastructure paradigm make it of critical importance that the grids maintain optimal levels of health. Smart grid health is a function of power throughput, grid resiliency, reliability, and flexibility.

Pattern and model creation by using historical big data

A key component to smart grid health is learning (for example, machine learning) what patterns and models are associated with both smart grid health and smart grid failures or stress. Examples of interesting outcomes are potential failures indicated by the phase angle differences presented by Phasor Measurement Units (PMUs) due to events, such as demand spikes. Interesting behaviors and actions include entity failure order, failure relationships between entities, and predictors (entity values) associated with failures.

An exploration of large amounts of historical data can discern the valuable relationships between outcomes of interest and data variables and the values of these variables. The actual analytics require a rich set of data that consists of both fine-grained data and data aggregates. Predictive analytics software, such as the IBM SPSS® Modeler, can use historical big data to create of a set of models or rules. These models or rules are indicative of

[1] Christian Zehetner, *Gradient Based Bak-Tang-Wiesenfeld Sandpile Model, 2010*

[2] Barry Markovsky, David Willer, and Travis Patton. "Power relations in exchange networks." American Sociological Review (1988): 220-236.

[3] Ronald A. Howard and James E. Matheson. "Influence diagrams", Decision Analysis 2.3 (2005): 127-143

two types of behaviors or actions: Good actor behavior or normal actions and bad actor behavior or threatening actions. Predictive analytics can use historical data about each entity, each network, and each subnetwork to discover relationships, trends, patterns, models, and predictors that are associated with both the normal activities and the known threat activities. The wide repertoire of algorithms might address grid-specific aspects, such as fault location algorithms, using both geography (for example, spatial analysis) and time (for example, temporal analysis). These historical predictive models could be deployed by the analytics assessment portion of the grid.

There is a third set of behaviors of interest: Those behaviors that are different (that is, anomalous) from the defined good and bad behaviors. Anomalous actions are not good or bad. Restated, instances of anomalous behavior must be illuminated and evaluated. Over time, vetted anomalous behavior assessments can be added to either the set of good actor behavior or normal actions and bad actor behavior or threatening actions (a *threat vector*).

Continuous smart grid health assessment

Smart grid health status cannot be discerned from a single data element value. As part of the system of systems concept, entity health is a function of its own health and the health of other entities that it is connected to. Graph distance and relative health status of other entities must be included in the assessment. If a particular entity is not healthy, it might contribute to a health status downgrade for other connected entities. A decay function analytic must include both how far away another entity is within the graph and the health status of that entity. The decay function is conceptually equivalent to the epidemiology metaphor, which is also a complex network problem.

An accurate picture of a smart grid's health requires a contextually correct picture of each grid entity (singular components, hierarchical combinations of entities, and each *network of networks*). A contextual picture includes information, such as current component state, component state history, edge distance, graphical knowledge of grid connectedness, and a decay function for impact of other components.

Health assessments for smart grid entities must take place in real time or quasi-real time. The assessment must make sense of new data observations as they are generated. The grid analytics environment must be edge-based so that it can both assess and respond in real time or quasi-real time. The edge-based analytics can determine whether the data observation for that grid entity or system or system of systems now matches the previously defined good actor behavior/normal actions, bad actor behavior/threatening actions, and the associated patterns. The analytics environment can determine whether the addition of this new data point changes the existing scores or the likelihood for models, trends, behaviors, scenarios, and situations. The assessment also determines whether the entity no longer matches known actions and behaviors, which results in identifying that entity as *anomalous*.

A cumulative, cohesive picture of the nodes and the networks allow the analytics to use a combination of internal relevance detection models, rules, and situational assessment algorithms to make sense of and to evaluate various aspects of the grid. The substantive portions of current smart grid assessment analytics are either pattern match assessors or statistical assessments. Neither of these approaches use context-based analytics.

As the real-time analytics reveal discoveries, including anomalies that matter, automatic decisions can be executed or alerts can be sent to users. Alerts can trigger real-time responses or a lengthier replanning event. Given real-time assessment, the analytics can determine whether an interesting event has occurred, such as drift in the center of inertia (COI). A matching of a pattern assessment rule or a sufficient change in a variable's value triggers either a functional change or an alert to an operator.

When a surprising event occurs, for example, a drop in demand on a sunny hot summer day, the analytic environment could create an indicator of the presence of possible false information. Those changes or discoveries that are deemed relevant and interesting can be pushed to appropriate users. Some triggers may initiate the gathering of additional data to generate new evidence for a hypothesis or to confirm or deny an hypothesis. Another reaction might be that of automatic system adaptation or reconfiguration or changes in the network to counteract the threat of system instability. Other grid parameters can be modified, such as security settings, bandwidth allocation, and pathway selection. The dynamic modification of these parameters can enhance system reliability, stability, and resiliency.

The continuous smart grid health assessment enables the optimization of the grid and the avoidance of failures (through early detection) by sending actionable insights to substation controllers for local action, to automated systems for global optimizations, or to human operators for situational awareness and further action. This sense-and-respond system implements a new paradigm, thereby transforming what would traditionally be simply a postmortem analysis of system faults to a paradigm of dynamic control, which enhances decision support for more robust decision engineering pathways to prevent adverse situations and optimize operating conditions.

Smart grid under cyber attack

Smart grid vulnerabilities allow attackers to penetrate a network and gain access to control software. *Cyber threat*, for the purposes of this guide, is defined as the modification or disruption of the information that deals with the power generation, power supply, or power demand. The follow situations are examples of information manipulation:

- Creating disinformation about power generation and causing automatic measures to activate to deal with the perceived power generation problems
- False tweets related to fabricated threats, such as that of a power plant leaking radiation
- Manipulation of energy market data to create profit
- Creating grid instability by creating false demand and incorrectly modifying load strategies

There are a large number of formal players within the electric grid ecosystem, not all of whom have the same goals. These players include consumers, producers, distribution operators, equipment manufacturers, transmission operators, electric power operators, and energy market exchanges. Each of these players typically has access points to the smart grid communication network and can be a source of cyber threat. Of course, bad actor types can also create access to present threats. These components are among those that are targeted:

- A single piece of hardware, such as a single PMU or a Phasor Data Concentrator (PDC)
- Sets of hardware (for example, a substation and its generators)
- Information or processes associated with the supply and demand of power

The Industrial Control Systems Cyber Emergency Response Team (ICS-CERT)[4] responded to 256 incidents that targeted critical infrastructure sectors in fiscal year 2013, and 59% of those incidents involved the energy sector.[5]

Information-based attacks on smart grids

Data has become important in almost all aspects of life; however, data is not always what it seems to be. The reach of social networking (for example, Facebook) and microblogging services (for example, Twitter) has extended to hundreds of millions of users. The popular US-based Twitter service generates more than 500 million tweets a day, which translates to about 5,700 tweets a second, on average. These valuable new venues are frequently mined in real time to detect new information or changes in patterns or values of variables. Social networking venues are becoming increasingly subject to illegitimate use, such as spamming and *astroturfing*, the promulgation of disinformation under the cover of legitimate grassroots behavior. As an example, a known channel was compromised by a cyber attack and generated a false tweet. This case of astroturfing occurred on April 23, 2013 when a false tweet was generated by @AP, the official twitter handle of the world's oldest and largest news-gathering organization, Associated Press. The tweet asserted that there were "Two explosions in the White House and President Barack Obama had been injured." This disinformation caused the market to drop by more than 100 points. Reuters estimated that the temporary market loss in the S&P 500 was approximately $140 billion.[6]

Similarly, information-based attacks are occurring on smart grids. Smart grid malfunctions have enormous costs, potentially triggering power outages, communications disruptions, scarcity of food and water, and raw sewage contamination of water supplies. In March of 2014, *The Wall Street Journal* reported that a Federal Energy Regulatory Commission (FERC)[7] study indicated that if a certain combination of 9 out of the 30 major electric substations were knocked off the grid, a nationwide blackout would likely occur. In 2009, *The Wall Street Journal* reported[8] that cyber spies from China and Russia had hacked into the US electric grid and inserted software that was capable of disrupting the power supply. In June 2010, malicious Stuxnet[9] software aimed at disrupting Siemens SCADA (Supervisory Control and Data Acquisition systems) installations at nuclear power plants, was introduced.

Detecting questionable information within a smart grid

The real-time assessment must determine whether an interesting event has occurred and whether that event is naturally occurring or, instead, had cyber disinformation underpinnings. For the astroturfing example, given the trust level that most individuals give to Associated Press[10] data, the use of traditional veracity assessment techniques, such as rumor centrality measures (detecting the source of the information diffusion or the origin of the rumor for this case) would not have been particularly revealing.

[4] ICS-CERT, https://ics-cert.us-cert.gov/

[5] Meghan McGuinness, "A new organization for cybersecurity across the electric grid," *Bulletin of the Atomic Scientists*, http://thebulletin.org/new-organization-cybersecurity-across-electric-grid7046

[6] Steven C. Johnson, "Analysis: False White House tweet exposes instant trading dangers," http://www.reuters.com/article/2013/04/23/us-usa-markets-tweet-idUSBRE93M1FD20130423

[7] Rebecca Smith, "U.S. Risks National Blackout From Small-Scale Attack,"*The Wall Street Journal,* http://www.wsj.com/articles/SB10001424052702304020104579433670284061220

[8] Siobhan Gorman, "Grid Is Vulnerable to Cyber-Attacks," *The Wall Street Journal*, http://www.wsj.com/articles/SB123914805204099085

[9] David Kushner, "The Real Story of Stuxnet," *IEEE Spectrum*, http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet

[10] Steven C. Johnson, "Analysis: False White House tweet exposes instant trading dangers," http://www.reuters.com/article/2013/04/23/us-usa-markets-tweet-idUSBRE93M1FD20130423

However, if we examine complex models based upon the intersection of big data indicators, there is a higher likelihood that questionable information would have been discovered through the lack of high volume or variety of other micro-narratives about the alleged White House explosions. The absence of supporting narratives indicates questionable tweet veracity. Use of analytics within unstructured data, for example text, requires sophisticated natural language processing (NLP) software that can support the discovery (or lack of) of pertinent micro-narratives or “needles within the haystack” for validating context. Because much content is also digital media-based, analytics software for this ecosystem must be applied.

In 2012, the communications system of a German power utility was hit by a denial-of-service (DoS) attack,[11] which had a botnet as the principal underlying attack vector (a DoS attack involves thousands of false requests sent to a server with the intention of overwhelming a system). DoS attacks on an electric grid can result in severe instability of power systems. The jamming can result in incorrect system state information being presented to the controller of the involved wide area system monitoring and control. If a controller is receiving real-time information from several different overlapping channels, it can potentially detect the malicious intent of the jammer. A DoS attack might insert corrupted data, such as incorrect format and content or correct format and incorrect time. In this case, the customers’ electricity supply was not affected, but it was reported that it took several days to repair the affected systems.

Ideally, the decision to modify a power system must be made on the basis of an assessment of current node (generators, systems, PMUs, or the related fabric of an *Internet of Things* measurements and the time-stamped history of each of these grid measurements. However, given the knowledge that there is the potential for some of the system data to be intentionally false, for example the insertion of false power control commands or the modification of grid status data, additional types of assessment strategies must be included in the smart grid health analytics portfolio.

Model assessment under questionable information veracity

The real-time analytics environment can discover whether the cumulative data (new streaming data conjoined with historical) on that grid entity matches the complex models and patterns that have been developed in the deep analytics portion of the process. Multiple, complex models are used, because adversaries endeavor to obfuscate their efforts and, unfortunately, simple models are relatively easy to fool. This is an instance where the “bigness” of the data or *bigger data* that is associated with a smart grid brings an advantage. The more parameters that are tracked, the more likely it is that anomalies and “strangeness” can be detected. Experience has shown that the networks that link more frequently to other networks introduce common vulnerabilities. Analytic models fall into three categories:

- Normal behavior
- Known-threat behaviors
- Anomalous behavior

Two entity attributes that are especially useful for the detection of questionable information are time stamps and geospatial data. The Northeast Blackout of 2003[12] affected regions of the Northeastern and Midwestern United States and the Canadian province of Ontario slightly before 5:10 p.m. EST on Thursday, August 14, 2003. Although some power was restored by 11 p.m., many did not get power back until two days later. The blackout affected approximately 10 million people in Ontario and 45 million people in eight US states. The

[11] “Europe's power grid hit with denial-of-service cyber attack,” *Electric Light&Power*, http://www.elp.com/articles/2012/12/europes-power-grid-hit-with-denial-of-service-cyber-attack.html

[12] JR Minkel, “The 2003 Northeast Blackout--Five Years Later,” *Scientific American*, http://www.scientificamerican.com/article/2003-blackout-five-years-later/

blackout's primary cause was a software bug in the alarm system at a control room of FirstEnergy Corporation in Ohio (whose 10 electric utility operating companies comprise the largest investor-owned utility system in the US). Operators were unaware of the need to redistribute power after overloaded transmission lines made contact with unpruned foliage, and what could have been a manageable local blackout quickly cascaded into increasing distress on the electric grid and ensuing massive blackout.

One of the more useful cyber disinformation detection mechanisms is to look for new entries or entries that normally are not part of a sequence. Ideally, sequences are accurately time stamped, but many of the smart grid components lack the ability to generate a consistent, synchronized time stamp.

Another strategy is the insertion of PMUs within the grid that have overlapping collection areas. Centralized computing nodes are used to compare assessments of data for each observed entity. When there are two different measurements for the same entity at the same time, an alert needs to trigger a deeper assessment.

The Irwin technology and IBM Watson solution applied to bigger data for a deeper assessment

PMUs are going to play a significant role in improving situation awareness of the utility grid, making it more reliable and resilient. The large amount of data produced by PMUs is difficult to capture and difficult to analyze and use for decision making. With this solution, the data can be analyed at an estimated 60 samples per second, which makes decision making closer to real time. The volume, velocity, variety, and the complications of value and veracity of this data not only qualify it as big data, but as *bigger data*.

After all, these PMUs, which are also known as *synchrophasors*, are, in essence, devices that sit on transmission lines, at substations, and on distribution networks. The PMUs are tasked with checking the magnitude and angle of electrical sine waves at the speed of the grid (60 cycles per second in North America). The PMUs also synchronize data across various entire grid networks by using global positioning system (GPS) radio clocks.

The amount of data generated by smart grid entities results in a massively big data paradigm, with everything moving very fast and with accuracy demands that require a high-speed communications system to capture what the PMUs are sensing. Determining what the data means and providing the requisite decision support to grid operators definitely puts it into the bigger data category. The pertinent actors for realizing a smarter grid have been diligently endeavoring to translate this mass of bigger data into enhanced context-awareness.

Wide-area electric power system outages are often caused by lack of context awareness regarding the systems, which can destabilize the grid. Yet certain bigger data solutions, in the form of the Irwin technology and IBM Watson solution, can be used to enable wide-area situational awareness and robust decision engineering[13] and to power system operations pathways.

[13] S. Chan and S. Sala, "Sensemaking and robust decision engineering: Synchrophasors and their application for a secure smart grid," DOI:10.1109/DEST.2013.6611337 Conference: Digital Ecosystems and Technologies (DEST), 2013 7th IEEE International Conference, http://bit.ly/1zasM3F

The following scenarios are among those to consider:

- Phase angle problem

  When certain transmission lines trip, the voltage phase angle difference becomes so great that the protective relaying does not allow the transmission line to be reconnected and thus re-energized.

- On-grid photovoltaic (PV) systems dilemma

  Small PV systems have settings that trip them offline during a disturbance, leaving the utility to deal with the loss of the PV generation while also servicing the load for the customer (who had been served from the PV), perhaps all the while dealing with the loss of a large generating unit.

- Under-frequency load shedding problem

  For a significant loss of generation, under-frequency load shedding attempts to match the load to available generation to prevent a collapse of the power system. Typically, customers are grouped into different “tiers” (forming a specific-sized block of load). Each tier has a specific frequency setting for which the block of load is shed when the frequency drops to that level:

  – Tier 1 is the first frequency level at which a block of load is shed.
  – If the frequency decays to the Tier 2 level, another block of load is shed.
  – If the frequency further decays to the Tier 3 level, the final block of load is shed.

Normally, the frequency recovers before the Tier 3 level is shed. However, utilities are increasingly experiencing a phenomenon where the frequency of the system has decayed low enough that Tier 1, 2, and 3 customers are all shed.

For each of the listed cases, an astroturfing campaign that promulgated disinformation would have direct impact on decisions made regarding the electric grid. For the first case, if an erroneous greater-than-normal phase angle were presented to the control room operator, decisions might be made to disconnect a transmission line or not to reconnect one. On an even more granular level, each PMU is carefully instrumented and calibrated to the particular geographic location, so local conditions are considered, for example. Large-scale disinformation might have a greater impact on PMUs that have lower phase angle thresholds~~,~~ than others. In another case, if the small PVs were forced to trip offline due to disinformation, the electric utility would experience a double whammy: It is no longer receiving power from the small PV systems, and it also has to provide load to compensate for the small PV systems being offline. If today’s trends continue, where an increasing amount of renewables are used and PV penetration is increasing, this exposure will increase.

Fortunately, with this solution, it is possible to leverage higher-order derivatives at the edge (that is, edge analytics) to discern the various slopes and shapes over time. In this way, it is possible to discern synthetic from natural conditions. With this capability, it is now possible to better determine normal versus anomalous activity and to more granularly distinguish between good actor behavior/normal actions and bad actor behavior/threatening actions.

Summary

Smart electric grid health assessments are made challenging by the complexity and sophistication of a smart grid. As reported in the news, critical infrastructures are being targeted by adversaries. This guide presented the Irwin technology and IBM Watson solution, a bigger data solution that can support wide-area situational awareness and facilitate robust decision engineering and power system operations pathways. This approach relies on contextual analytics for version verification, predictive modeling, and a computing model assessment that uses edge computing. This solution can detect a cyber threat, such as misinformation or disinformation, by providing a Smart Grid Cyber Health Assessment surrounded by the realities of a big bad data world.

Other resources for more information

For more information about the concepts that are highlighted in this guide, see the following resources:

- IBM InfoSphere® Sensemaking

  https://www.ibm.com/industries/publicsector/fileserve?contentid=235174
- Business Analytics for Big Data

  http://www.ibm.com/software/analytics/solutions/big-data/
- IBM SPSS Modeler

  http://www.ibm.com/software/analytics/spss/products/modeler/
- Jeff Jonas, IBM Fellow and Chief Scientist of the IBM Entity Analytics Group, blogs on sensemaking and context analytics

  http://jeffjonas.typepad.com/jeff_jonas/
- *Context-Based Analytics in a Big Data World: Better Decisions*, REDP-4962
- *A Framework for Smart Grid Analytics and Sensemaking: The Mehta Value*, REDP-5082
- *Analytics in a Big Data Environment*, REDP-4877
- *Turning Big Data into Actionable Information with IBM InfoSphere Streams*, TIPS0948

Authors

This guidepaper was produced by a team of specialists working with the International Technical Support Organization (ITSO).

Dr. Lisa Sokol is a Data Scientist within the Office of the Chief Technology Officer, IBM Software Group, and US Federal Government Services. Her primary areas of interest are assisting government communities in solving hard problems and using context analytics to discover actionable information buried within large amounts of data. She has designed several systems that detect and assess threat risk relative to fraud, terrorism, counter intelligence, and criminal activity. She has a doctorate in Operations Research from the University of Massachusetts.

Dr. Steve Chan is a Data Scientist who serves as the Director of the IBM Center for Resiliency and Sustainability, which includes the IBM Network Science Research Center, under IBM i2®. He is also Chairman of the Board for Mehta Tech. He is the Professorial Chair of the Network/Relationship Science Analytics PhD Program and Director of the Network Science Research Center at Swansea University, Research Professor of Sensemaking and Visualization Analytics, and Director of Asia-Pacific Institute for Resiliency and Sustainability at Hawaii Pacific University/Swansea University, and Professor of Network-Relationship Science for the CyberPsychology Research Center at the Royal College of Surgeons in Ireland. He is a Chief Technology Officer at MIT and a Senior Fellow at Harvard. He is also an alumnus of MIT and Harvard University.

Thanks to LindaMay Patterson and Phil Monson for their contributions to this project.

Now you can become a published author, too

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time. Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online:

ibm.com/redbooks/residencies.html

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new IBM Redbooks® publications, residencies, and workshops with the Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

i2®
IBM®
IBM Watson™
InfoSphere®
Redbooks®
Redguide™
Redbooks (logo) ®
SPSS®

The following terms are trademarks of other companies:

Other company, product, or service names may be trademarks or service marks of others.