SMART GRID PROTECTION AGAINST CYBER ATTACKS

Contract No 608224

Deliverable D4.3

High-level design documentation and deployment architecture for security information analytics

AIT Austrian Institute of Technology • Fraunhofer AISEC • The Queen’s University Belfast
Energieinstitut an der Johannes Kepler Universität Linz • EMC Information Systems International Ltd
Kungliga Tekniska högskolan (KTH) • Landis + Gyr
United Technologies Research Centre • SWW Wunsiedel GmBH

Document control information	
Title	High-level design documentation and deployment architecture for security information analytics
Editor	Silvio La Porta (EMC)
Contributors	Robert Griffin (EMC), Rohan Chabukswar (UTRC)
Description	This deliverable consists of both design documentation and the deployment architecture for the Security Information Analytics mini-project. We also introduce the Security Analytics canonical architectures and approaches.
Requested deadline	30/06/2015

SPARKS Security Scrutiny Committee Assessment

This deliverable has been examined by the SPARKS Security Sensitivity Committee (SSC), in accordance with the process outlined in Deliverable D2.1 on SPARKS Security Management. According to the SPARKS Description of Work, this deliverable has a dissemination level of PU (Public).

The SPARKS SSC understands there are a number of potentially sensitive pieces of information in the discussion of security analytics in this deliverable. It is possible that cyber attackers who gain access to the deliverable could take advantage of this information to define and execute attacks against Smart Grid infrastructure or other environments. However, all of the information presented herein is already available in various publications, so its publication here does not significantly change the level of risk relative to cyber-attacks.

Executive Summary

Even with the most resilient security system in place to protect Smart Grids, some threats will elude those defensive measures with potential catastrophic results. It is therefore paramount to detect and understand those kinds of threats to shorten response time, minimize damage, and highlight potential latent threats and deficiencies in the security system.

A Smart Grid system is the result of the convergence of power and communications infrastructure, reaping the benefits of comprehensive and intelligent management of an increasing variety of electric devices and sources of power. However, this increase in connectivity also leads to an increase in the potential for security risks and failure.

In this report several incidents are highlighted along with numerous known methods for intrusion, and potential deficiencies in both the communications infrastructure, (software security hole) and the power infrastructure (insecure monitoring and control systems). All of these factors, along with the increase in attack sophistication and attacker knowledge, underline the need for a holistic approach to data analytics, in order to reduce response time, improve recovery or remediation efforts, expose potential security deficiencies and provide accurate and relevant insight to inform governance and policy making.

A Smart Grid data analytics platform must be able to examine a large number of disparate data sources both in real-time, allowing for early warning systems, as well as support for forensic investigation, through access to historical data. This is a marked shift from previous approaches which were based on a limited number of different data sources, examining relatively small amounts of data and trying to match them to known attack signatures. This new approach is necessary to detect and analyse more subtle and sophisticated forms of stealth attack or creeping failures that affect several aspects of the system but otherwise remain undetected.

The ever changing threat landscape requires the security analytics system to differentiate normal behaviour from anomalous, by leveraging multiple and more complex algorithms. The desired result is to reduce the number of false positives while improving detection rate, making the role of the operator much easier while increasing the potential for corrective action. In addition, the analytics platform should also simplify coordination across multiple disciplines in the context of threat analysis and detection.

The general design approach presented here includes several components:

- Visibility and access to relevant information to guide the analytics work and the data to be analysed resulting in greater insight
- Establishing normal patterns and detecting anomalous behaviour, prioritizing the most pertinent for investigation
- Carrying out mitigation, remediation and recovery actions on priority tasks. Including the engagement with people in the relevant organization to create an improved culture of security

The implementation of the system was carried out leveraging several Smart Grid systems, primarily the NIMBUS microgrid. This site is made up of a variety of electrical sources and information sources. Having a wide ecosystem of devices available enables the development of the system in a device agnostic way. The SPARKS Security Information Analytics (SIA) prototype was developed with the potential for customization in mind, enabling domain-experts to adapt its operation and decide what is considered anomalous behaviour. Currently the design of the platform includes two components, a static rule analyser that examines the data against an expected set of behaviours. The second, currently in development, is a smart auto analyser that will leverage machine-learning algorithms to identify patterns and clusters in the data, establishing normal behaviour and thus facilitating the detection of any divergences from that pattern.

Table of Contents

Table of Figures

1 Introduction to Security Information Analytics for Smart Grid

Assuming we have a robust architecture for Smart Grid security and have put in place defensive mechanisms designed to thwart intentional attacks and inadvertent exposure or loss of data, how do we discover and recover from attacks, losses and security failures that elude those defensive mechanisms? This document presents an architecture and deployment model for the effective use of capabilities and processes that enable the discovery and analysis of security issues to determine an appropriate response, and remediation or recovery strategies for those issues on which action must be taken.

It is essential to put in place a security information analytics system for Smart Grids that can respond effectively to attacks, intrusions and data exfiltration. This is important even with the most resilient and attack-resistant architecture possible in place. The operational infrastructure includes a number of systems:

- Enterprise Geographic Information System (GIS) is the platform that creates and collects information about utility assets (cables, transformers, customers, etc…) and makes that information available to the enterprise for monitoring and analysis. (ESRI, 2011)
- SCADA systems that control and monitor devices used in power generation, management and distribution.
- Customer information systems that monitor usage, perform billing, handle customer relationships and so on.
- Interfaces with external systems such as weather information, traffic information, satellite imagery, threat intelligence and so on.

Utilities use this combined information for a broad range of applications, including managing a comprehensive picture of the operating environment, detecting and analysing faults, capacity planning, predicting load, analysing the network, and managing security operations. For all these purposes, the utility must understand the relationship of its assets to each other. Since the smart grid is composed of two networks – the electrical distribution network and the communications network – utilities must understand the physical, spatial and electronic relationships both within each of these networks and between them. For example, the communications network not only enables the collection and consolidation of information from the electricity distribution network, but also provides the means of distributing control information to substations, smart meters and other components in the electricity distribution network. Understanding these interconnections is essential not only for effective operational management but in security management as well.

The GIS is particularly important in enabling the utility to understand the electrical and communication networks and the relationships between them. It provides a means to monitor the operational and security health of the system, answering such questions as which sensors have reported anomalous values for the past hour, where the anomaly may be in terms of the historical record for a particular sensor for a particular time period (day, week, month) or in terms of an abstracted pattern of values for sensors providing that particular function. This security information analytics for the Smart Grid takes advantage of device information provided by the SCADA system. But it needs to go beyond that device-specific information to present both a comprehensive perspective on the grid, and insight into specific operational and security issues that could affect the availability and safety of the grid.

This perspective into the health of both the electrical and the communication networks, as well as components within those networks, enables the Smart Grid to adapt quickly to prevent outages, whether those are the result of equipment failure, weather conditions, accidents, physical attacks or cyber security incidents. For example, there have been a number of instances in the United States of substation failures, including the explosion of Ives Dairy substation in Miami, Florida in 1993. (Ragan, 2010) That explosion was caused by the coincidental failures of several control and monitoring systems, including the emergency response system that would have notified the grid dispatcher of a serious problem. The more comprehensive monitoring and control systems enabled by Smart Grid, particularly through the instrumentation of a larger number of more diverse sensors within the substation, can reduce the risk of such events. Physical attacks on the grid, such as the April 2013 attack on the PG&E (Pacific Gas and Electric) transmission station in Metcalf, CA, demonstrate the importance of more comprehensive operational systems that can detect and respond to damage more quickly in order to limit the impact on the grid as a whole. (Tweed, 2014)

Malware like Havex with the capability to target control systems have recently emerged. Additionally, some crime-ware trojans such as BlackEnergy—used to automate cybercriminal activity—have been adapted and extended to operate in industrial control systems (ICS). The Havex malware was used between 2011 and early 2013 during the “DragonFly” campaign to target energy, gas and oil companies. (Symantec, 2014a) In this campaign one of the infection vectors used to spread the malware was the water hole technique, which consisted of compromising SCADA software company’s websites by repacking the malware with their legitimate software. BlackEnergy has a module that scans and targets machines in an IP block searching for well-known open ports used by SCADA control systems. Additionally, the Sandworm crew used a zero-day exploit (CVE-2014-0751) to spread their malware and target HMI servers. (Symantec, 2014b) The aforementioned malwares are able to capture screenshots and record operators’ activities in the compromised machines.

Another challenge is the state of the ICS machine involved in these architectures. These legacy control systems were typically designed with no security features since they were typically isolated from the Internet. These control systems often run old operating systems with archaic software and are not updated to support new operating systems or new libraries as these would potentially expose them to old unpatched vulnerabilities.

The aforementioned incidents show that cyber criminals are increasingly targeting critical IT infrastructures while gaining knowledge on how to use and interact with these systems. These are extreme examples of the kinds of risks that the security information analytics for Smart Grid must take into account. They point to an important principle that should be generally instrumented in the smart grid operational model: the smart grid must have the information, the analytics and the control capabilities to prevent outages as much as possible, detect outages quickly when they do occur, and to respond to those outages quickly to minimize their impact on the greater grid. To accomplish this, the security information analytics for

Smart Grid must be able to detect and respond to failures in equipment, sensors and communications, such as those that resulted in the Miami substation explosion. For example, the operational systems (and the GIS in particular) should detect and monitor power surges that might reduce the life expectancy of a particular transformer. The systems can then institute appropriate remedial actions, and also perform a spatial analysis to enact an alternative power distribution path, load allocation and usage scenarios that would mitigate the impact of the failure in that component. Furthermore, security information analytics for Smart Grid must be flexible and able to quickly adapt as new threats emerge.

2 Requirements for Security Information Analytics for Smart Grids

A major challenge in achieving security information analytics for Smart Grids is ensuring the quality of the data within the various systems. The grid is transitioning from a relatively static topology—in which changes to the physical infrastructure were relatively infrequent—into a more dynamic model where power and information sources and consumers change much more frequently. Both the sources and consumers will also vary more significantly in terms of reliability and complexity than in the past. For example, power sources will range from very large commercial power stations to small household renewable energy capabilities. Information sources will range from comprehensive sensors installed by utility companies within the distribution network, to highly variable information sources in connected cars and home area networks. The utility must be able to maintain up-to-date security information analytics for Smart Grids that reflects this dynamism in the communication and the electricity networks, so that accurate analysis can be performed and appropriate decisions taken when anomalies that could indicate equipment failure, cyber-attacks or other issues are detected.

As discussed in the PSERC research paper by Govindasaru et al. (2012), the security information analytics for Smart Grid must go beyond the conventional focus on distribution and generation infrastructure for fault isolation, remediation and recovery to a focus on information and control across three levels:

- Internal data sources, including event logs, configuration databases, vulnerability scans, etc…
- Applications, including generation control, transmission control, distribution control, customer relationship, etc…
- Systems data sources, including SCADA, distribution management, customer information management, etc…
- Network data sources, including communications network (routers, firewalls, packet analysis, etc…) and distribution network (generators, transformers, cables, etc…)
- External data sources, such as threat intelligence, vulnerability reports and other resources that complement the organization’s own data

This more comprehensive approach to information and control both enables and requires a new understanding of data analysis. As discussed by Cardenas (2013), it requires the ability to handle huge amounts of data, to process that data using new analytics and visualization techniques, and to integrate the results of that analysis with governance processes that make those results readily actionable.

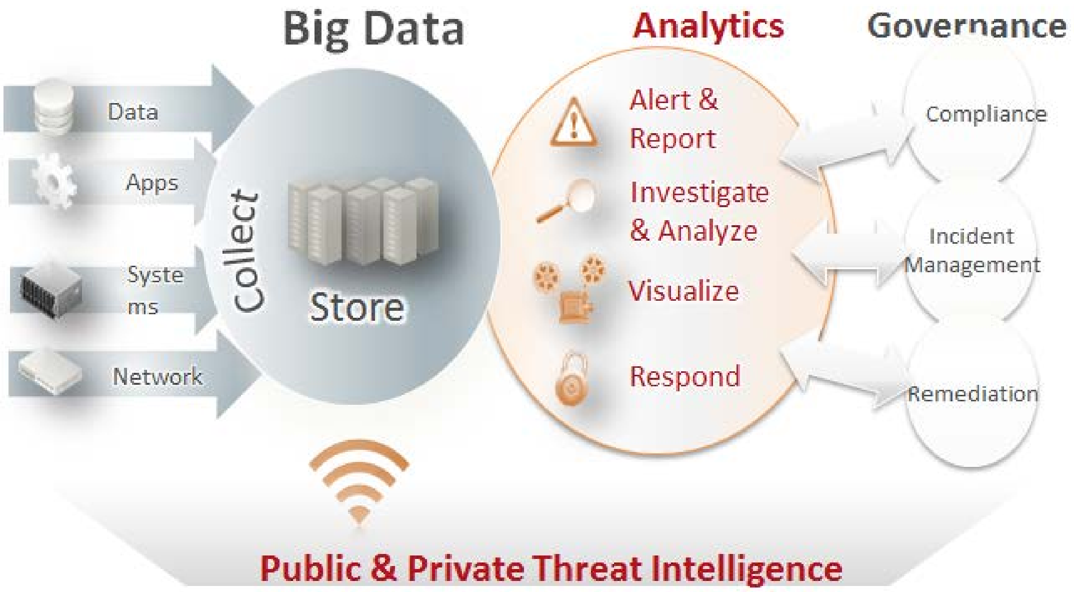


Figure 1: Comprehensive approach to Data Analytics

These three essential areas of information analysis have been discussed extensively in a recent study by ESG analyst Jon Oltsik (2014). Although not focused on Smart Grid, the study is very relevant to the Smart Grid operational model.

The data sources that utilities must use in order to gain visibility include four major areas. The first area is data sources providing information about risk, including sources within the organization and external sources. Second is visibility into the physical and cyber infrastructure i.e. status monitoring of the electrical and communications networks. Third is identity information i.e. a full inventory of devices and components present in both the physical and cyber environments. Fourth is the extraction of information that provides insight into applications i.e. their performance and activity levels. These sources represent a much greater volume of data than utilities have dealt with in the past, requiring new models for the effective storage and use of this information. Equally importantly, this set of sources represents a much broader diversity of inputs than utilities have dealt with in the past, particularly with external sources of risk information such as threat intelligence. Utilities will have used comparable operational information, such as MTBF (Mean Time Between Failures) information provided by equipment manufacturers. But they rarely, if ever, will have integrated that operational information with information regarding cyber threats that might target that same equipment. But this correlation of operational and security information is essential in order to understand what the anomalies in equipment behaviour might mean.

The **data analytics capabilities** must be able to work with this larger and more disparate set of information, requiring not only new approaches such as massively parallel processing but also new algorithms. Current security analytics, for example, can spot obvious policy violations through applying a priori rules that look for correlations across events. However some attacks remain invisible when examining a single perspective or action in isolation. Most current tools are unable to analyse the required number of disparate datasets in combination to detect such intrusion. The data analytics systems in the Smart Grid security information analytics for Smart Grids must be able to model “normal” behaviour and detect anomalies against those derived patterns, regardless of whether those anomalies reflect equipment failure or cyber-attack. The systems will need to draw on a broader range of analytics algorithms than rules based on correlation, including “nested-algorithms” that analyse anomalies across multiple algorithms in order to eliminate false positives, pinpoint problems, and provide security analysts with the right details to prioritize and expedite remediation processes. The analytics must enable the utility to understand normal state behaviour and look for anomalies that result from malicious activity, performing more detailed contextual analysis to determine the appropriate response.

Moreover, the analytics must support a continuum in terms of the depth of analysis and speed of response. The analytics capabilities must include real-time analytics used for fast incident detection/response and based upon events, logs, NetFlow, network packets, and activity on endpoints. But they must also support historical data analysis of security anomalies and the associated investigations that may span several months or years. This type of asymmetric investigation may require analysts to sift through massive quantities of historical data to piece together patterns, detect malicious behaviour, and trace the root of these activities. With asymmetric big data security analytics, security analysts will use an assortment of methods to discover and investigate “low-and-slow” attack patterns.

Improved visualization and query technologies must provide easy-to-use GUIs, comprehensive reporting, and intuitive navigation. They must provide visual analytics with 3D images and simple pivoting across dimensions rather than just spread sheets and charts. The visualization capabilities, like the analysis capabilities, must support both real-time response and longer-term investigation, including through integration with workflow and data sharing to align incident detection/response with operations. Analysts must be able to work together with IT and security operations teams once incident detection processes transition to incident response, leveraging effective communication capabilities, investigation tools and end-to-end oversight. This will require workflow tools to manage the incident detection/response process independently but also in interoperation with IT operations and security tools.

Integrating analytics with effective governance processes, including policy management, compliance, risk management, vulnerability management and incident management, is also essential to an effective Smart Grid operational model. For example, the best strategy for dealing with both operational and security vulnerabilities is to manage them within the context of the overall risk management process for the organization. Managing security vulnerabilities has to start with understanding the potential for exploiting the vulnerability, the motivation of attackers in taking advantage of that vulnerability, and the impact that such an attack could have. In the context of that more robust understanding and management of risk often termed “risk intelligence”, the organization can make the necessary business decisions regarding how to prioritize remediation of the vulnerability, in each of the places where it occurs, against the numerous other competing and complementary security actions that the organization is considering.

Similarly, security analytics must integrate with IT and security incident detection processes. This requires effective workflow tools to manage the incident detection process and to achieve smooth hand-off to incident response processes. The integration requires interoperation with existing security and IT operations tools that facilitate incident analysis and response planning by providing context and insight regarding the source, scope and impact of the incident in order to understand both its immediate level of risk and the potential for risk related to similar incidents that may have occurred but were not detected or remediated.

As noted in a recent IDC report (Feblowitz, 2013), the smart grid has also put focus on the privacy of information. Many consumers are concerned that their consumption data collected through the utilities' meters will be shared with third parties or will be accessible to hackers. Privacy of customer data extends to competition as well. Vertically organized utilities are required to keep customer consumption data collected by their distribution subsidiaries separate from their competitive retail subsidiaries. The visibility, analytics and action required for an effective security information analytics system for Smart Grids must respect and support these requirements for the privacy of information not only in order to conform to regulatory requirements, but also to engage the human community in effective security.

3 Architecture and Design Approaches to Security Information Analytics for Smart Grids

3.1 Introduction to Security Information Analytics Architecture and Design

As discussed in deliverable D3.1, there is a significant gap in existing Smart Grid reference architectures with regard to security analytics, a gap that is a symptom of a more fundamental issue. For example, NISTIR 7628, while providing a valuable overview of the range of security capabilities that should be considered by any organization designing, deploying, evaluating or auditing smart grids, NISTIR 7628 is based on assumptions about security strategy that, while best practice at the time NISTIR 7628 was written, are no longer in step with the transformations that have occurred in the threat landscape, business models, and in infrastructure technology in the past several years. In particular, security strategies have to include responding to targeted attacks by aggressive and well-funded adversaries who will succeed in bypassing defensive mechanisms. Present day security strategies must be augmented with detection, analysis and response capabilities that go well beyond those described in NISTIR 7628.

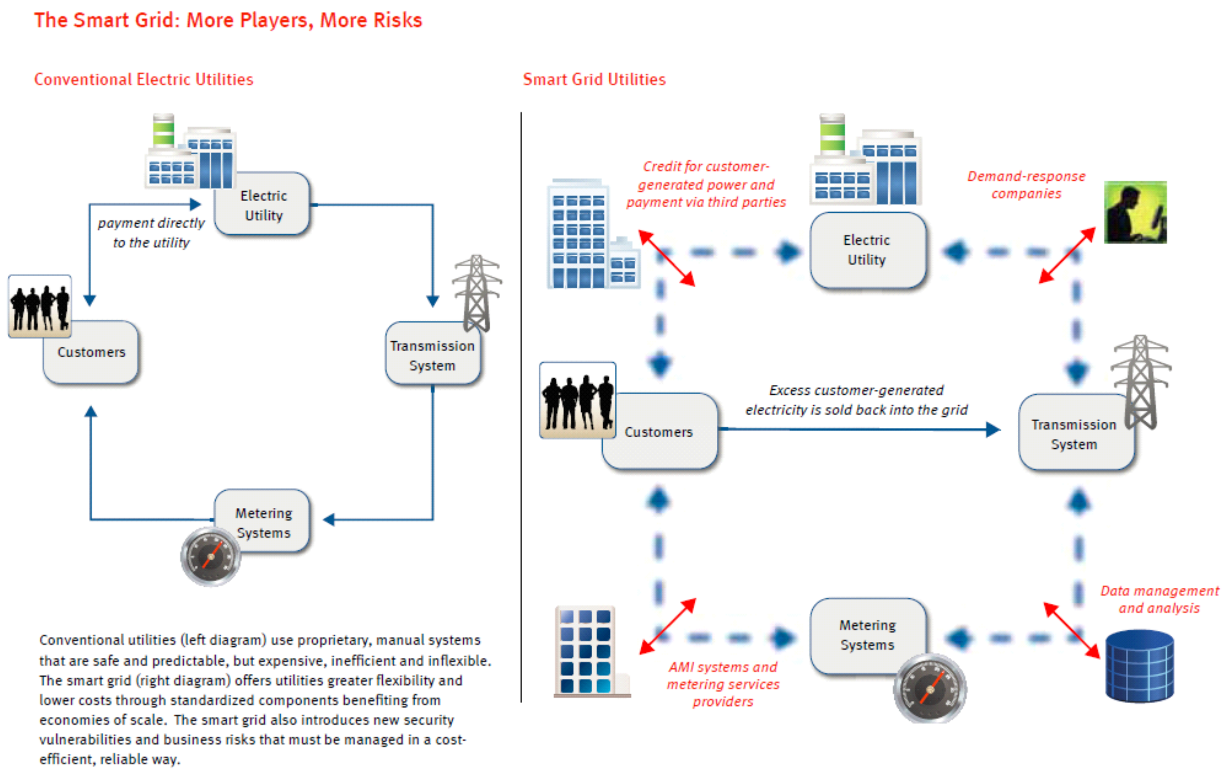
This section describes the requirements for visibility, analytics and response in a Security Information Analytics capability for Smart Grids.

3.2 Visibility: Data Identification, Collection and Preparation

3.2.1 Visibility into Risk

Power and utility companies are facing a wide range of new and emerging risks. For example, they are seeing a rise in operational compliance requirements amid a climate of transformations in national and international energy policy. They must address issues regarding public acceptance of initiatives such as AMI (Automated Metering Infrastructure) and the development of new renewable energy capabilities such as wind farms.

Operational risks continue to change as new technologies are employed and as existing relationships with suppliers and consumers change. Smart Grid enhances the traditional electrical grid with communications capabilities that converge utility operations with IT. There are many benefits to this convergence, including increased data for real-time situational awareness, historical analysis and improved customer service. But as noted by Vermesan and Friess (2013), in gaining the benefits that come with automated meter reading, demand response management and interactive home networks, utilities also open themselves up to new security risks:



The Smart Grid: More Players, More Risks

Conventional Electric Utilities
Smart Grid Utilities
Electric Utility
payment directly to the utility
Customers
Transmission System
Metering Systems
Credit for customer-generated power and payment via third parties
Electric Utility
Demand-response companies
Customers
Excess customer-generated electricity is sold back into the grid
Transmission System
Data management and analysis
AMI systems and metering services providers
Metering Systems

Conventional utilities (left diagram) use proprietary, manual systems that are safe and predictable, but expensive, inefficient and inflexible. The smart grid (right diagram) offers utilities greater flexibility and lower costs through standardized components benefiting from economies of scale. The smart grid also introduces new security vulnerabilities and business risks that must be managed in a cost-efficient, reliable way.

Figure 2: More Players, More Risks

These risks can be understood in terms of changes in attackers, attack methods, and exploitable vulnerabilities.

- **New kinds of attacks:** Electricity utilities become susceptible to sophisticated electronic attacks as power grids integrate communications and distribution systems, employing implementations built on widely-accepted standards. The integration of systems may attract criminal, competitive and nation-state attacks related to stealing customers’ payment account information, disrupting electricity supply and damaging electricity infrastructure.
- **New attack points:** The integration of new technologies into delivering smart grid services – from electronic metering systems to demand response services – means there are far more potential points of entry for attackers into Smart Grid than into the closed-grid systems of the past. In addition, the integration of new partners and the establishment of new relationships with existing partners in the Smart Grid supply chain, including cloud services, demand response, equipment suppliers and power resources, increase the number and kinds of attack points. As has been demonstrated by attacks in other industries, electric utilities have become only as secure as the most vulnerable partner in their supply chain.
- **New vulnerabilities:** For example, meter readings in the Smart Grid that are transmitted electronically could be tampered with to report lower-than-actual energy use. (Searle, 2012) Fraud has always been a possibility with mechanical meters. But the use of smart meters increases the likelihood of vulnerabilities that can be exploited by attacks, such as vulnerabilities in the transmission protocol standards. The disclosure of long-standing vulnerabilities in fundamental security standards such as Secure Sockets Layer (SSL) shows that even well-established technologies can be

sources of vulnerabilities that attackers can exploit. The rapid development of new technologies for Smart Grid increases the risk that the technical capabilities used by utilities and their supply chain will have exploitable vulnerabilities.

The Smart Grid industry will need to have much greater understanding of and visibility into these risks. A utility will need to participate in security communities that share information about attackers and attacker strategies in order to benefit from the experiences and research of industry, government and academia. The benefit of this insight into attackers has been demonstrated by information sharing models for the financial services industry, for example, in which each institution can share experiences with new phishing attacks with other financial institutions to share insight into new attackers and new attack methods. The VERIS Community Database Project (VCDB), a community data initiative to catalogue security incidents in the public domain using the VERIS framework, provides another such information-sharing community. (http:// www.vcdb.org/) Threat reports from government, industry and academic sources also provide important insights into changes in the attacker community, as well as into existing and emerging attack methods.

Reports such as the annual Verizon Breach Reports (Verizon, 2014), the Common Vulnerabilities and Exposures (CVE) list and the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) continue to demonstrate the importance of visibility into vulnerabilities and processes for addressing those vulnerabilities. Visibility into vulnerabilities continues to be critical in order to understand, prioritize and mitigate the risks related to vulnerabilities. But as a report by Enterprise Management Associates has shown (Oltsik, 2013), enterprises are already overwhelmed by the volume of information they are collecting and analysing. Visibility into vulnerabilities, therefore, while being as comprehensive as possible, must also leverage analytics capabilities and integration into the larger risk governance processes to enable insight into what vulnerabilities should be addressed, within what time frame. Visibility—not only into risk, but into all the areas explored in this section—means not just the collection of data, but the transformation of that data into information that can be acted on effectively.

3.2.2 Visibility into Infrastructure

As discussed in detail by Musser (2009), Smart Grid enables the collection and integration of both operational and non-operational data on an unprecedented scale across the grid infrastructure.

Transmission systems for Smart Grid provide much more accurate, timely and extensive information than earlier systems that relied on analog devices for monitoring and control. This includes the ability to detect and record transient disturbances, to isolate the location of disturbances and to provide more precise measurement of disturbances. Phase measurement units (PMUs) and dynamic line rating units are particularly important in terms of providing this monitoring not only for operational purposes, but also as part of the security monitoring and response. For example, physical attacks on critical power lines could result in a cascade of impacts that disrupt service across a broad geographic area. Research on fault isolation, such as the paper by Zhang et al (2009) has shown the effective use of multivariate analysis of PMU information in order to identify faulty components and faulty sections, then analysing that information to accomplish fault isolation. Such faults can be the result not only of mechanical failure or weather conditions, but also of both physical and cyber-attacks. For example, cyber manipulation of the distribution system could result in surges that damage transformers as effectively as natural events such as lightning strikes. Visibility into these faults and into conditions that might trigger such faults is as vital in the identification and remediation of cyber-attacks on the transmission system as in the identification and remediation of such faults from other causes.

As noted above, visibility into substations is also critical to an effective Smart Grid operational model. Substation sensors enable the monitoring of power flow conditions and equipment performance. They enable monitoring across the substation component, such as transformers, transformer cooling, circuit breakers, switches, relays, batteries, surveillance equipment and meters. They also enable the monitoring of environmental conditions, including temperature in the substation and meteorological conditions that might impact the substation. These monitoring capabilities can identify individual situations that could indicate a cyber-attack, such as fluctuations in transformer cooling, unusual behaviour in circuit breakers and disabled sensor or control mechanisms. In combinations, the capabilities provide a rich set of data to which analytics can be applied both to detect potential cyber physical attacks and to investigate those potential attacks to disregard false indicators of compromise and to understand actual attacks more fully.

Distribution monitoring has already been impacted significantly by the deployment of smart meters, providing a much more extensive and varied set of information than was available through manual reading of mechanical meters. For example, distribution applications can use field data to establish baseline patterns of real-time power flow conditions. Variations in actual power flow can then be identified quickly and analysed to determine whether manual or automated changes should be made to line regulators, capacitor banks and other distribution components to address operational requirements. These same variations, however, can also be indications of cyber or physical attacks. Information regarding the variations can be combined other sources to identify potential attacks and appropriate response.

Advanced Metering Infrastructure (AMI) information, particularly in combination with other sets of information such as home area network configuration and household information, is already being used to identify potential fraudulent use of electric power. (Tweed, 2012) In this case, normal patterns of usage abstracted for the aggregate information from the AMI can be compared against the usage of a particular household, industrial user or other consumer. Anomalies between the abstracted pattern and power consumption by a particular consumer can indicate not only faults in the meter but also meter tampering for purposes of power theft.

Home Area Network (HAN) and Energy Management Control (EMC) Systems have developed in response to consumer interest in optimal control of discretionary power usage and utility interest in optimal control of load conditions. Monitoring information from the HAN can be valuable to the consumer in quickly identifying ways to optimize their power consumption. But here too, patterns of usage can be derived that provide opportunities to detect such issues as attacker compromise of an EMC that could result in significant damage within the consumer environment, such as in the case of an industrial consumer.

Work management systems are increasingly mobile, subject to attacks both in the reception of service instructions and in the transmission of service and operational data. Widespread disruption of a work management system could be part of an attack strategy focused on disruption of service by delaying response to emergency conditions. Monitoring to ensure the health of the work management system is valuable in rapidly identifying and responding to such disruptions.

With Smart Grid, customer information systems contain substantially more data than previously. This can contain customer financial information that, if not well-protected, could be of interest to cyber attackers looking for financial gain or doing reconnaissance in preparation for fraudulent activity. The recent cyber-attacks against retail enterprises for the purposes of stealing credit card information and user financial account information indicate that such attacks could increasingly be targeted at Smart Grid customer information systems as well as the data in those environments become more extensive.

Generation resources typically already have extensive monitoring capabilities for operational purposes such as automated fault detection and capacity management. Resilience against cyber-attacks require enhancements to these capabilities, such as protocol inspection to detect spurious or malformed control commands that could result in disruption of service or damage to the generation resources. The cascading effect of unexpected control outages in the 2006 power disruption in Europe, as analysed in the UCTE 2006 report (p. 20) resulted in "a significant amount of generation units tripped due to the frequency drop in the Western area of the UCTE system." This situation can be caused by a cyber-physical attack, rather than by a change such as the switching-off that initiated the cascade of failures. The report also noted that "most of the TSOs do not have access to the real-time data of the power units connected to the distribution grids. This did not allow them to perform a better evaluation of the system conditions." (UCTE, 2006 p. 20) Visibility into the relationship of power generation to generator capacity will be critical in identification of and automated response to prevent cascading failure triggered by a cyber-event.

Enabling distributed energy resources, a major goal of Smart Grid, creates significant challenges with regard to managing grid capacity and resilience due to the complexity of bi-directional energy flows and communications. Disruption of or error in the communication flow as the result of a cyber-attack, could result in the propagation of inappropriate load commands or measurements that could result in disruption of service or damage to components. Fraudulent claims regarding energy production by distributed energy resources could also be introduced into the system if there is inadequate visibility of generation, distribution and consumption data.

Enterprise Information Management (EIM) systems, including the Geographical Information Systems (GIS) discussed earlier in this chapter, are an important source of visibility within Smart Grid. Financial management of the utility may be disrupted by attacks on the EIM. Emergency response may be disrupted by attacks on the GIS.

Finally, Data management and analytics systems provide the opportunity for integration of data across all of the sources of visibility discussed in this section. This integration has great benefits for operational control, customer satisfaction and realization of business opportunity.

But there also needs to be visibility into these analytics systems themselves in order to detect and respond to cyber-attacks against these systems. The rise of targeted attacks, particularly those that have been characterized as Advanced Persistent Threats employing extensive reconnaissance, multiple attack vectors and detection-evading tactics (SBIC, 2011), have made security analytics increasingly important in cyber defence, as much for data management and analytics systems as for the other systems discussed above.

As discussed by Popovik et al. (2013), visibility across all these systems has created the opportunity for better observability of power systems, cross-correlation between related measurements and improved decision-making capability when operating the system. Equally importantly, this visibility into infrastructure creates the opportunity for the application of security analytics to detect and investigate cyber physical attacks, particular in combination with the visibility into risk already discussed and the visibility into identity, information and applications that we discuss in the remaining sections of this chapter.

3.2.3 Visibility into Identities

Effective cyber-physical security for Smart Grid, as for any other environment, requires clear and comprehensive visibility into who has access to which resources, including infrastructure, applications and information, who approved that access and whether that access is appropriate based on the individual's business relationship with the organization.

Visibility into identities starts with collecting identity information and entitlements both from structured and unstructured data resources, including cloud and mobile apps. This information should relate to the full spectrum of both internal and external users (employees, partners, customers) across the full range of devices (laptops, tablets, mobile devices) and applications that they use. It should include a comprehensive understanding of the attributes of each user; who they are, what they do and the policies that apply to them.

This visibility must be balanced against requirements for privacy that come both from the external regulatory environment and internal policies. Both consumers and government agencies are concerned that personal information may be visible within the Smart Grid enterprise in situations where the information is not needed and to individuals who should not have access to it. For example, supply chain relationships may encourage the utility to share customer information with third parties and across national boundaries. It is very important for Smart Grid enterprises, therefore, to be circumspect in what identity information they collect, where and how they store it, and when and with whom they share it.

Identity intelligence provides essential context that can dramatically improve the effectiveness of security investigations. A security analyst can use identity context to see if the user's access is appropriate, and how the user relates to the application in question. This intelligence is especially important for such issues as detecting Segregation of Duties (SoD) violations, finding orphaned accounts and identifying inappropriate assignment of access privileges for particular users.

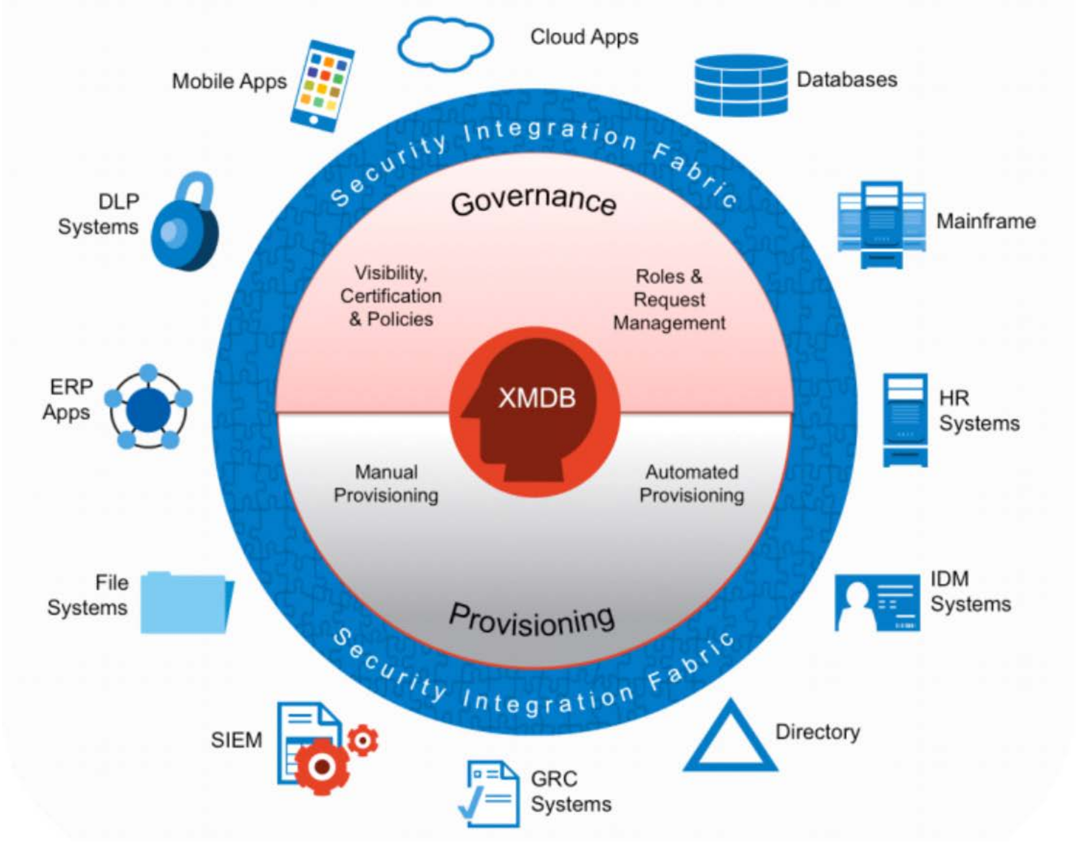


Figure 3: Visibility into Identities

3.2.4 Visibility into Information

The concern about ensuring the protection of identity information raised in the previous section applies in more general terms to all information that is collected, stored and used by a Smart Grid enterprise. As is already apparent in the discussions of visibility into risk, infrastructure and identity, the Smart Grid depends on the collection and use of a broad range of information.

As noted by IDC, the enterprise needs to develop a data and information strategy that ensures that information is available as needed for both current and future requirements. (Feblowitz, 2013) At the same time, the strategy must also ensure the confidentiality, integrity and availability of that information in conformance with both external regulation and internal policy and strategy. This requires identifying sensitive information across the enterprise: where it is stored, how is it used and who has access to it. This needs to be a dynamic process that reflects the constantly evolving nature of both the information in the Smart Grid enterprise and in the attackers who may want to gain access to that information to steal, corrupt or destroy it.

Also as discussed in the IDC report, this visibility into information is essential not only in order to understand its sensitivity, use, and lifecycle, but also to leverage it for collaborative aspects of the security information analytics for Smart Grid as a whole and security operations in particular. Information plays a critical role in the driving decisions both in on-going operations and in longer-term projects, in meeting compliance requirements, in evaluating risk and many other aspects of the business. Visibility into the availability and quality of this information empowers the organization to address these requirements. In order to achieve this visibility, the Smart Grid enterprise should have capabilities such as the following:

- **Knowledge management**: This capability ensures that information is captured across all appropriate sources, managed effectively across its lifecycle and disseminated as appropriate to all participants who should have access to that information.
- **Collaboration support**: This capability enables the secure exchange of information across all appropriate participants, including external partners.
- **Information integration**: This capability employs analysis, taxonomies, ontological frameworks and other processes and technologies to enhance the value of managed information and to incorporate it into the business processes.

Many Smart Grid enterprises have seen an increase in collaboration between operations and information technology, particularly for sets of data such as automated meter data that is valuable for outage management and for security analytics. There are still major opportunities in this area and in integration, particularly in areas such as transmission data such as synchro-phasor measurements that are already employed in operations but only rarely for purposes of security analytics.

The IDC paper “Making Good on the Promise of Smart Grid” discusses the importance of data archiving strategy as part of this opportunity for convergence of operational technology (OT) and information technology (IT). (Feblowitz, 2010) Utilities typically archive meter data for a minimum of three years and often for seven years. This reflects regulatory requirements for archived data retention. But there is significant value in archived data that is not realized unless it is readily visible for longitudinal analytics, including for security purposes. Storage tiering represents an important strategy for addressing visibility requirements while minimizing data storage costs. While flash storage may be used for near-real-time analysis of operational safety, reliability and performance, or real-time-analysis related to security, less expensive disk-based technologies can be used for a mid- and long-term data storage, making the information visible even though at a slower access rate.

3.2.5 Visibility into Applications

The Electric Power Research Institute (EPRI) has suggested that that the goal of the smart grid is to enable a broad range of advanced grid applications, including outage management, real-time contingency management, AMI management, SCADA management, and so on. (EPRI, 2010) Visibility into applications is essential to assessing risks comprehensively and to making informed enterprise-level decisions.

The transformation taking place with advanced metering is a good example. In conventional metering systems, meters were read once a month, totalling 12 reads a year per customer. With advanced metering infrastructure (AMI) meters will be read every 15 minutes, increasing the number of reads per customer by 3,000 times per year. We have touched on the importance of visibility into AMI as an infrastructure and on the importance of visibility into this information from AMI. But equally important is visibility into the AMI applications from both operational and security perspectives. Is the application for receiving and processing information from smart meters encountering connectivity issues that result in lost information? Is it participating in a patch management strategy to ensure it is fully operational and secure? This visibility into the application is critical in ensuring their effective operation and security.

3.3 Analytics: Deriving Patterns and Understanding Anomalies

3.3.1 Introduction to Security Information Analytics for Smart Grid

Most Smart Grid enterprises already monitor and analyse at least some of the sources of information discussed above for signs of unusual behaviour of people, applications, infrastructure, and communication. But often this analysis is focused on explicit indicators such as previously identified malware signatures or blacklisted IP addresses or domains. Sophisticated attackers can circumvent such tell-tale, static monitoring approaches by modifying lines of code, by provisioning a new virtual machine in a public cloud, or by registering a new Internet domain as a command and control or drop site.

Attackers typically operate by collecting information on the security systems and software installed on the target network. This allows them to test their malicious code and also to verify they will evade detection by the target network systems before launching an attack. It's much harder though for attackers to circumvent monitoring and analysis systems that are watching for unusual patterns and behaviours. Sooner or later, hostile malware or users must do something unusual that breaks with system norms, and that is when Intelligence Driven analytic systems, often called "Intelligence-Driven Security", will find them. (RSA, 2013)

For example, when it comes to detecting malware, endpoint threat detection solutions do not look for "known bad" files; they look for suspicious behaviours. By comparing what is actually running in memory with what should be running based on the files residing on the local disk, malware detection tools are better able to identify discrepancies and get a direct and more reliable view of whether illicit code is present.

3.3.2 Analytics: Establishing Patterns of What is Normal

Intelligence Driven Security systems establish what "good" behaviour looks like within an IT environment by monitoring and learning a variety of machine and human activities, from what ports on servers are typically used for outside communications to employees' individual log-in locations and habits. Analytics solutions often rely on logs and configuration information as data sources. But they achieve far greater visibility by also incorporating other sources. Figure 4 shows an example of data integration and information exchange for operational and security analytics for a substation, adapted from Figure 1 to include the broad range of input sources described in Popovic's 2013 discussion of Smart Grid data analytics. These input sources include digital protective relays (DPR), digital fault recorders (DFR), digital disturbance recorders (DDR), sequence event recorders (SER), remote terminal unites (RTU), phase measurement units (PMU), and a number of other sources.

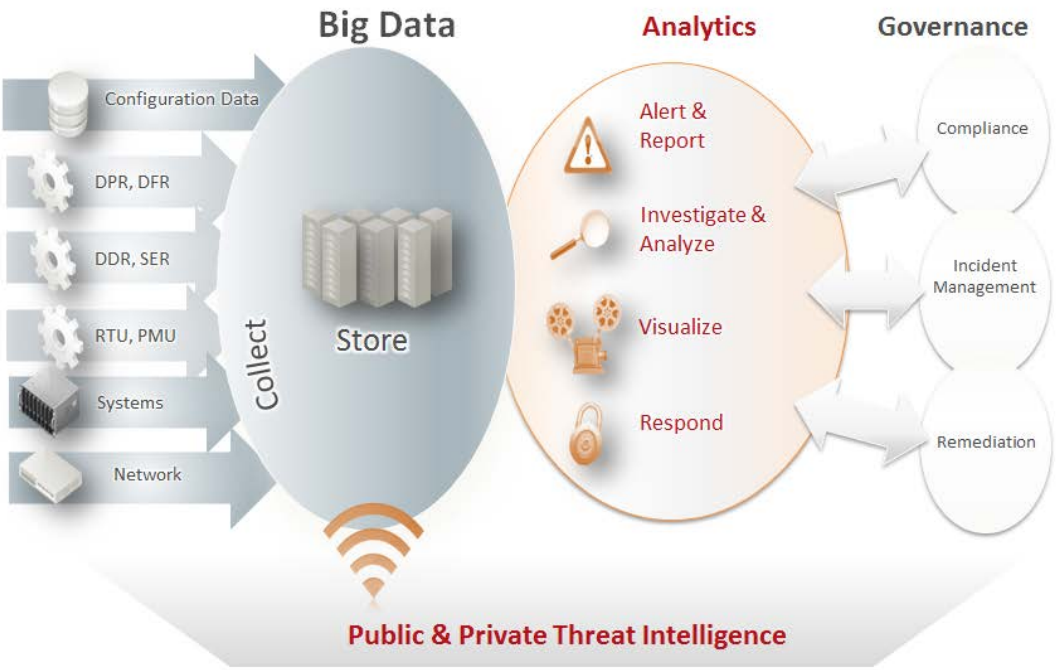


Figure 4: Deriving Patterns of Normal Behaviour for Substations

Similarly, capabilities such as network packet-capture are important in establishing normal behaviour in the IT infrastructure. Full network packet-capture means recording, parsing, normalizing, analysing, and reassembling all data traffic at every layer of the network stack. As network traffic is captured, it's analysed and tagged to facilitate subsequent threat analysis and investigation. Capturing and tagging network data enables security analysts to reconstruct users' sessions and activities to understand not just basic details such as what time or to which IP address specific data packets were transmitted, but exactly what information was sent in and out and the resulting damage. These techniques help organizations learn what is "typical" within an IT environment so that future deviations from normal—which often indicate problems—can be identified and investigated as they arise.

3.3.3 Analytics: Detecting Anomalies within the Operational Environment

With patterns of normal behaviour in hand, activities outside the norm can be detected, analysed and appropriately acted upon. For example, if an anomaly is flagged as a potential security issue, it can be passed to an analyst for further investigation. If the analyst determines that the event is a false positive, the analytics tools can "learn" from that experience so that it is less likely to flag future recurrences of that event as a potential security violation.

Analysis systems capture and analyse massive amounts of rapidly changing data from multiple sources, pivoting on terabytes of data in real time, organized into various levels to enable different types of detection. For example, data can be captured and analysed for potential security issues as it traverses the network. This **capture time analysis** identifies suspicious activities by looking for the tools, services, communications and techniques often used by attackers without depending on logs, events, or signatures from other security systems. Examples of this capture time analysis includes the detection of non-browser software programs running HTTP, protocols over non-traditional ports, and executables embedded in PDF files. Additionally, these sophisticated tools can detect subtle signs of attack by correlating events that seem innocuous in isolation but that are problematic when strung together. Analytical techniques fuse internal inputs from various sources using metadata. These advanced detection mechanisms also act as trip-wires that can provide early warning of potential infiltration. Processing of these information flows happens as they occur, meaning suspicious activities are spotted while there's still time for security teams to stop attacks in progress.

This is comparable to the real-time operational response that is a fundamental capability in Smart Grid systems. Papers such as the California ISO Smart Grid Roadmap and Architecture (30) describe the application of synchro-phasor technology in real-time fault isolation and remediation: "Phasor units measure voltage and electric current physical characteristics. This data can be used to assess and maintain system stability following a destabilizing event within and outside the ISO footprint, which includes alerting system operators to take action within seconds of a system event. This capability reduces the likelihood of an event causing widespread grid instability." Such "capture time analysis" is shown in Figure 4 of the Popovic paper (2013), illustrating the extraction of phase current features to determine if a fault has occurred.

This kind of capture analysis and response is important in terms of real-time faults caused by cyber-attacks rather than natural disasters or equipment failure. The Aurora attack, in this case referring to the demonstration by Department of Homeland Security conducted at the Idaho National Laboratory (INL) in 2007, showed the creation of an out-of-phase condition that could damage alternating current (AC) equipment. (Swearingen, 2013) The attack forced the repeated opening and closing a circuit breaker or breakers to rapidly disconnect and reconnect a generator to the grid, but out of phase. Many circuits of utilities carry varying load profiles, from resistive to inductive loads. These circuits may include rotating equipment. This load profile allows for the real-time failure demonstrated in AURORA to occur. Analytics that detect the anomalous behaviour in circuit breakers can enable automated responses that prevent equipment damage or worse.

Analysis systems can also perform **batch analysis** on large volumes of historical security data. In the case of security analytics, such data are needed not only to fulfil most companies' data retention and audit requirements but they are also invaluable in uncovering adversarial tactics that may have taken many months to execute and may even be ongoing. For instance, batch analysis of security data archives can help uncover previously overlooked cyber-attacks in which illicit data was transmitted only sporadically in small, stealthy streams over weeks or months. These types of "low and slow" attack techniques are hard to spot when they are occurring, because they are designed to seem innocuous by taking cover under existing processes and communication streams. These techniques usually become suspicious only when executed in a particular pattern over a specific window of time. Detailed, automated analyses of security data archives can discover attackers in the midst of establishing a foothold, as well as reveal information losses those organizations may not even realize they sustained.

An example of batch analysis is the identification of compromised hosts through the use of large volumes of historical information to establish a pattern of normal behaviour for hosts in an enterprise, then the review of that information to identify hosts that diverge from that pattern. In this first example, HTTPS packet data is used as input. The security analytics tool looks for anomalous HTTP access, DNS lookups, accessed domains, traffic, users, IP addresses, beaconing activities, event timestamps and other network information. Using this information, the tool can create a ranked list of likely malicious IP addresses that require further investigation. Further, the tool can create reports containing additional information regarding the IP addresses that can help in forensic investigations and threat detection. The tool can then also search connections to IP addresses belonging to the same malicious IP subnet to identify other machines that require further investigation.

Batch analysis can also use rapid transitions in DNS addresses to identify potentially compromised hosts, In this case, DNS packet data is used as the input. The security analytics tool looks for anomalous subdomains, users, IP addresses, ISP domains and other network information. Using this information, the tool can create a ranked list of likely fast-fluxing domains that are strong candidates for further investigation. Further, the tool can create reports containing historical visibility for about each domain (for example, 30-day history) and additional information regarding the domain that can help in forensics and threat detection.

In summary, batch analyses can uncover attacker techniques and indicators of compromise that security teams can use in the future to detect similar attacks. More generally, batch analysis enables organizations to detect operational and security anomalies, and reconstruct incidents with certainty and detail so they can investigate their losses and remediate problems faster and more effectively.

3.3.4 Analytics: Detecting Anomalies within the Administrator Environment

Among the most difficult attacks to detect are those in which the attacker exercises legitimate use cases and capabilities in order to accomplish malicious goals. The attacker hides in plain sight, making it difficult for security personnel to distinguish between legitimate activity and attacker reconnaissance, infection of target systems and other attacks.

This technique is especially dangerous in attacks against administrative environments that may have been compromised as a first step in subversion or destruction of operational systems. Advanced attacks often use social engineering approaches to steal user credentials, trick the user into opening an attachment containing malware or to open URLs that would exploit the user's web browser, leading to potential malware infection. Having established a foothold in the administrator environment through such a technique, the attacker can then leverage any of a number of legitimate system capabilities in order to move laterally across the organization, increasing both the scope and kinds of infection of the Smart Grid environment. This lateral movement can be accomplished by manipulating task scheduling, by creating or modifying services, and by remote reconnaissance or image execution. After the system infection phase, the next steps are 1) to make the attacker presence persistent on the user machine and 2) to start privilege escalation, using a range of techniques that are well-documented in the literature.

In the Windows environment, for example, one of the primary methods by which APT actors utilize existing architecture is by using Windows interconnectivity to increase their foothold in the network. Most commonly, allowed interaction between systems allows attackers to conduct lateral movement. This allows attackers to navigate from the initial compromised hosts to the more high-value targets in the environment.

Before starting with lateral movement, the attacker needs to identify the Windows environment, for example, if it is running in an Active Directory domain or not. In the Active Directory case, the attacker can use different techniques to get an administrator account credential, for example, getting a credential from the NTLM (Windows Challenge/Response authentication protocol) that is stored in the machine memory inside MSV1_0.dll or if the administrator login to the machine was remote, the attacker could find the credential in Kerberos.dll, wdigest.dll or tspkg.dll.

It is important to highlight that the attacker does not need to brute force the credential because they can use the “Pass the hash” technique. This enables the attacker to authenticate at a target server/service using directly the dumped password hash string. This is possible due to a Windows vulnerability which incorrectly uses a *salt* value in the authentication implementation. *Mimikatz.exe*, an open source tool, is often used to accomplish this type of attack.

There are a number of Windows administration tools that are frequently used by attackers in order to move laterally within an organization, infecting more and different kinds of systems. We’ll look at four such tools. The first of these is the “at” command (at.exe).

The *at* command was implemented in Windows 2000 as a method of manually scheduling tasks from the command. It was designed to allow administrators to schedule tasks on hosts or remote systems. However, the command still exists, and is operable, up to Windows 7 and Server 2008 systems. Attackers can use this command to create immediate or longer-term tasks that allow them to execute commands on remote systems. The *at* command works by:

1. Connecting to the IPC$ share on the remote system over TCP/445
2. Creating a named pipe named “atsvc”
3. Sending a JobAdd request with the command to be run.

Once the remote system receives the request, it creates a .job file in the Tasks folder under C:\WINDOWS\System32\ with a file name of At<*job number*>.job.

Analysing the JobAdd request enables the security environment to detect malicious use of the *at* command. Since the *at* command will submit a known UUID in the bind request to port TCP/445, this value can be detected, and parsed through the stream to the JobAdd. In order to detect malicious use of the *at* command, a security analytics environment that has captured the command can parse its location and length fields and job file name to detect discrepancies from normal use of this command, notifying security personnel of the anomaly.

The second tool frequently used by attackers in the administrative environment is the *schtasks* command. The *schtasks* command was introduced with Windows XP/Server 2003 as a replacement for the *at* command to allow administrators to schedule jobs from the command-line. Adversaries can utilize this functionality to exploit the same attack vector as with the *at* command. While the *at* command uses a well-known endpoint in the form of the named pipe "atsvc", the *schtasks* command uses dynamic RPC endpoint mapping to determine its communication port. As such, it submits the *ITaskSchedulerService* UUID to the RPC endpoint mapping port (TCP/135) and receives a port assignment with which to begin communication. As the UUID for this service is known, a security analytics tool can detect when this UUID is seen passed with the appropriate protocol payload. Once the job is scheduled with the Task Scheduler, security analytics can recognize it as the command being executed by the job, identify it as an anomalous use of *schtasks* and notify security personnel of the potential attack, including displaying the binary that is scheduled to be run.

A third tool frequently used by attackers in the administrative environment is the *sc* command, or *service control*. The *sc* command was introduced with Windows NT 3.51 and included on endpoints as of Windows XP. The *sc* command uses the Service Control Manager Remote Protocol to interact with the Service Control Manger on the local, or remote, system. This command was designed to allow administrator to create, control, or remove services from the command-line. It is commonly used by attackers to create new malicious services, disable services that may prove problematic, or remove services as necessary. While there are other command-line binaries that allow for control of services remotely (namely *netsvc* and *instsvc*), these do not allow for the creation of new services remotely and as such, the *sc* command is preferred by attackers. Microsoft has allocated both an RPC interface UUID to present to the RPC endpoint mapping port, as well as a pipe named "svcctl". As the UUID for this service is known, a security analytics tool can identify when this UUID is seen passed with the appropriate protocol payload. As the purpose of this command is to interact with the database of installed services, the security analytics tool can then scan this database to allow the review of host artefacts resulting from the command's execution. The security analytics tool can then notify security personnel of the potential attack, including displaying the artefacts related to the *sc* command.

A fourth tool frequently used by attackers in the administrative environment is the windows management interface. Windows Management Instrumentation (WMI) is described by Microsoft as "the infrastructure for management data and operations on Windows-based operating systems." (Micorsoft, n.d.) It is the latest method provided by Microsoft to conduct administrative tasks on host and remote systems via C++, Visual Basic, or the *wmic* command-line utility. In order to connect to remote systems, one of two primary methods is used. The first is to specify the connection information in a SWbemServices object using a moniker string. This allows the script to communicate with a remote system using the user's current credentials. However, this approach is limited as alternate credentials cannot be supplied and this method is unable to connect to systems across domains. The second method is more flexible, as it allows for specification of target computer, domain, username and password. As the first method doesn't allow for on-the-fly use of stolen credentials, this second method is more attractive to adversaries. This method involves using the

ConnectServer method from SWbemLocator (IWbemLocator for C++ applications) to specify and conduct connection operations.

Both of these methods use the IWbemLevel1Login interface to connect to the management services interface within the requested namespace. This is useful in detecting this traffic as the interface must use the pre-designated UUID from Microsoft. As the UUID for this service is known, a security analytics tool can identify the UUID being passed with the appropriate protocol payload. In addition to passing the UUID to the RPC endpoint mapper, the UUID is also sent in remote requests along with parameters containing queries and commands to be executed on the remote system. In the case that *pktPrivacy* is not set, these parameters are sent to the remote system in the clear with a leading value of “__PARAMETERS”. This value is a result of the use of the *_PARAMETERS* class to set input parameters for the aforementioned WMI methods. By detecting the “__PARAMETERS” values and removing the “abstract” values (*__PARAMETERS* is an abstract class), the security analytics tool is able to retrieve and register the parameters given to the WMI remote command, including queries and commands to be executed on the remote system. The security analytics tool can once again notify security personnel of the potential attack, including displaying the detected artefacts.

This is just a brief overview of the most popular tools used to accomplish lateral movement. Other tools like *psexec.exe*, *wevtutil.exe*, *winrs.exe* and *net.exe* are also used by attackers to expand their network foothold.

Within Windows management, the use of the *ExecQuery* method is also interesting, as it allows adversaries to conduct actions and reconnaissance on remote systems by conducting WQL (WMI Query Language) queries against the remote system. The *strQueryLanguage* parameter defines the query language to be used. This parameter must be defined with the value “WQL”. The *strQuery* parameter is required, as it contains the value of the query to be executed. Without *pktPrivacy*, this value is sent to the remote system in the clear. Knowing that the value of *strQueryLanguage* must be “WQL”, and that the following parameter is the query to be executed, the query itself can be parsed and analysed for any session that contains the WbemLevel1Login UUID.

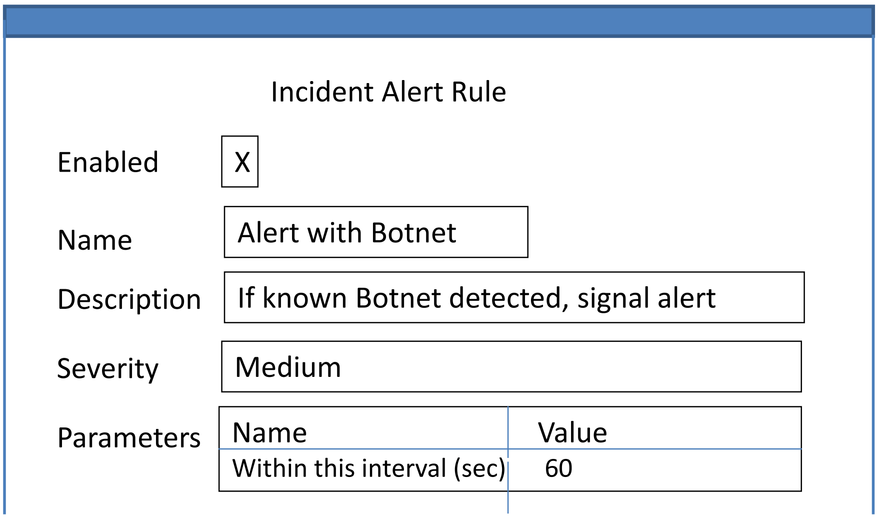
In all four of these examples of attacks within a Windows environment, the difficulty with combatting attackers using these methods is that all four of the capabilities have legitimate use cases within the administrator environment. This serves to increase the utility of these methods to attackers, since the valid use of these capabilities is not detected by most security tools that look for static indicators of compromise, rather than doing dynamic analysis of the use of these kinds of capabilities. Given this exploitation of legitimate communication between systems, it is extremely important for the success of analytical approaches to security that the normal use of these capabilities is well understood and incorporated into the analysis being done, in order to minimize false positives in the detection process. It is even more important to understand the environment in which this activity is seen if response personnel are to combat these threats. Understanding the known methods within the protocols used by these capabilities and determining their typical use and legitimate use cases within the organization’s actual environment is essential to applying security analytics in the detection of possible malicious use of these capabilities.

3.3.5 Analytics: Investigation and Prioritization

Advanced analytics tools examine the behaviour of machines, networks, and processes to determine whether they are evidencing operational problems, have been compromised by malware and so on. But such tools do more than detect incidents; they assess risk and prioritize alerts for remediation.

For example, files determined to be malicious may warrant a lower prioritization score if they're determined to be malware that causes more of a nuisance than a true threat. Conversely, files that bear no outward signs of tampering may contain a custom-compiled executable designed to run only when it reaches certain systems or when a covert command is given. To uncover this type of dangerous, customized malware, advanced threat detection systems use a series of analytical techniques to rate the risk levels of suspicious files.

For example, an organization may set a rule requiring the security system to analyse every new executable coming into its networks. The malware detection system would then "sandbox" new executables, running them in a quarantined environment, recording everything they do, and elevating their risk score if suspicious behaviours are observed, such as changing registry settings or replacing operating system DLLs. Of course, legitimate software could also perform these actions: to integrate functions with existing software or to install a patch, for instance. But if the new executable demonstrates one of these behaviours along with initiating unusual network connections, then its overall risk score skyrockets. In that case, the analytics system should send an alert to an incident management system that could notify the security team of the issue automatically to prevent further introduction of the suspicious file and quarantine it in any systems where it's found.



Incident Alert Rule

Enabled ☒

Name: Alert with Botnet

Description: If known Botnet detected, signal alert

Severity: Medium

Parameters

Name	Value
Within this interval (sec)	60

Figure 5: Responding to Detected Security Anomalies

Similarly, anomalies that indicate operational issues (not necessarily security issues) can be assessed and prioritized in terms of risk. As discussed in the California ISO paper (Cal ISO,

2010) such analysis enables responses that can have a substantial effect on stability of the Smart Grid:" Phasor data is also useful in calibrating the models of generation resources, energy storage resources and system loads for use in transmission planning programs and operations analysis, such as dynamic stability and voltage stability assessment. The technology may have a role in determining dynamic system ratings and allow for more reliable deliveries of energy, especially from remote renewable generation locations to load centres." (p. 8)

3.4 Action: Mitigation, Remediation and Recovery

3.4.1 Introduction to Mitigation, Remediation and Recovery for Smart Grid Security Analytics

Visibility and analytics enable effective action for recovery from incidents, remediation of vulnerabilities and mitigation of risk. For example, a cyber-attack launched via a compromised communication network against the circuit breakers in a substation results in damage to transformers in the station. Because the disabled substation could result in loss of power for millions of customers, this is assessed as a high priority issue. The response team confirms that the right people within the company are handling the incident properly: managers are conducting forensic analyses to figure out what happened, crews have deployed to fix the problem and the response teams are coordinating with each other to restart systems and restore power.

Once the immediate crisis is over, remediation activity determines how the attack occurred and whether there is a vulnerability that can be addressed to reduce the risk of similar attacks in the future being able to succeed. At the same time, the risk management team investigates the failure scenario to determine whether there are mitigation strategies that would reduce the likelihood and impact of transformer failure, for whatever reason.

3.4.2 Action: Recovering from and Managing Incidents

Even in the best instrumented and most secure operational model, incidents have to be expected. In fact, the number of incidents queued up for handling may be extremely large. It is essential for the incident response system not only to prioritize the incidents in the queue, but also to eliminate false positives and other clutter from the queue. To do this, incidents should be automatically checked against a global repository of items that analysts have previously investigated and resolved, before the new incidents are put on the queue. The incident response system should continually learn from these previous incidents, updated threat information, changes to operational configurations and other data sources in order to simplify the analyst's job as much as possible.

The incident response system should, as much as possible, prioritize the incidents in terms of known or potential impact on the business. The "Security Engineering Report on Smart Grids" by Hwang et al (2012) explores this impact from two perspectives. The first is technical impact factors:

- Loss of confidentiality. How much data may have been disclosed and how sensitive is it?

- Loss of integrity: How much data could have been corrupted and how damaged is it?
- Loss of availability: How much service may have been lost and how vital is it?
- Loss of accountability: Is the incident traceable to one or more individuals?

The second is business impact factors:

- Financial damage: how much financial damage may result from the incident?
- Reputation damage: How much reputation damage that would harm the business may result from the incident?
- Non-compliance: How much exposure to and risk of non-compliance does the incident introduce?
- Privacy violation: How much personally identifiable information may have been disclosed?

A full understanding of the impact may not be reachable until after the investigation is complete, or even for some length of time after that. But the incident response system should provide what prioritization it can, while ensuring that information related to the periodization decision is available to the response team so that they can ensure that the most important incidents are addressed as quickly as possible.

Incident response system should provide a rich set of context about prospective problems. For instance, for an incident related to a suspicious file that may represent malware, the system should correlate suspicious behaviours about the file (e.g., a driver, a process, a DLL), capture what's known about the file (e.g., file size, file attributes, MD5 file hash) through static and heuristic analysis, provide context on the file owner or user, and so on. Security analysts can then use this information to investigate if the file is malicious, and should be blacklisted, or non-malicious, and should be whitelisted. If an item is deemed malicious, all occurrences of the problem across the entire IT environment can be instantly identified. Then, once a remedy is determined, the security operations team can perform any necessary forensics investigations and/or clean all the affected endpoints.

Incident response systems should also ingest information from external sources to enrich the organization's internal data sources for purposes of incident investigation and response. For example, the security analytics platform and management dashboard should aggregate and operationalize the best and most relevant intelligence and context from inside and outside the organization to accelerate the analysts' decision making and workflows.

Remediation after malware infection is a complicated task. If the compromised machine is vital for the system availability, it is usually not always possible to use previously saved safe machine images. The incident response team should check all the machine environments to remove all potential access points that attackers could utilize. Cyber criminals tend to use different entrench techniques in a victims' network. The term entrenchment is used to describe a technique used by the attackers that allows them to maintain unauthorized access into an enterprise network despite attempted remediation efforts by the victim. The victims' machine can be compromised in a variety of ways; for example the attackers could install web shells, they can add malicious or modified DLL to running web servers, they can utilize RDP backdoors, hide malware that will commence malicious activity after a fixed period of time and so on.

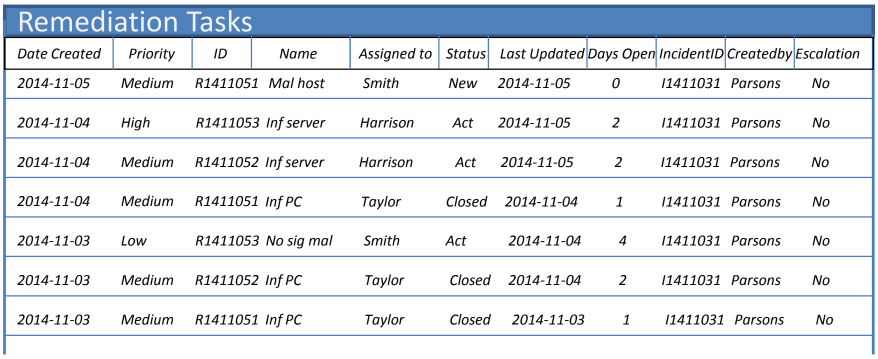
It is always good practice before starting to clean infected machines, to monitor network traffic and search for similar traffic patterns or similar IP connections as well as checking for all the possible lateral movements analysing forensically the victim's machine. The aim is to perform this in a stealth way so the attacker is unaware they are under surveillance thereby avoiding the possibility that the attacker will lunch countermeasures to cover their evidence or start to deploy other entrenchment techniques to remain in the network. Forensics analysts should search not only for malicious software but also for legitimate software that could be installed on the machine for malicious purposes as well for misconfigurations created by the attackers.

An effective incident response team should be composed of malware experts, IT forensics analysts and network experts thereby giving the organization a wide competency and skills blend to enable successful detection, protection and investigation. Operations teams may be confronted with potential evidence of an infiltration or breach, but find themselves exploring potential causes without success for weeks or even months. When that happens, it helps to bring in people with specialized expertise and tools in incident response (IR). IR specialists can deploy technologies that capture activity on networks and endpoints in key segments of the IT environment. Based on the scans, analyses and supplemental information these technologies generate, experienced IR professionals can usually pinpoint where and how security breaches are occurring and shut down ongoing cyber-attacks much faster than organizations can do on their own.

3.4.3 Action: Remediating Vulnerabilities and Anomalies

Responding to the incident itself is essential. But equally important is determining whether there were vulnerabilities that contributed to the incident's occurrence or impact. These vulnerabilities may have been technological, such as software vulnerabilities that provided access for an attacker or that caused unexpected behaviour in operation of a component. They may have been process issues that prevented an issue from being recognized until it had reached a critical level or that resulted in the initiation of a failure condition. Or they may be organizational, educational or other issues related to the structure and people of the organization, such as in individual vulnerability to social engineering attacks that resulted in malware infections.

The incident management system should support the determination of such vulnerabilities and assist in the remediation of those vulnerabilities, such as through the remediation planning shown in the figure below.



Remediation Tasks

Date Created	Priority	ID	Name	Assigned to	Status	Last Updated	Days Open	IncidentID	Createdby	Escalation
2014-11-05	Medium	R1411051	Mal host	Smith	New	2014-11-05	0	I1411031	Parsons	No
2014-11-04	High	R1411053	Inf server	Harrison	Act	2014-11-05	2	I1411031	Parsons	No
2014-11-04	Medium	R1411052	Inf server	Harrison	Act	2014-11-05	2	I1411031	Parsons	No
2014-11-04	Medium	R1411051	Inf PC	Taylor	Closed	2014-11-04	1	I1411031	Parsons	No
2014-11-03	Low	R1411053	No sig mal	Smith	Act	2014-11-04	4	I1411031	Parsons	No
2014-11-03	Medium	R1411052	Inf PC	Taylor	Closed	2014-11-04	2	I1411031	Parsons	No
2014-11-03	Medium	R1411051	Inf PC	Taylor	Closed	2014-11-03	1	I1411031	Parsons	No

Figure 6: Remediation Planning

On a more comprehensive level, an incident may indicate a more fundamental issue in the operational model, such as in terms of missing or improperly instrumented controls. For example, the 2009 paper by the US Department of Energy calls out the vulnerabilities inherent in the older control systems that are currently deployed throughout the United States and that may have to be replaced: “The electric power industry relies heavily on control systems to manage and control the generation, transmission, and distribution of electric power. Many of the control systems in service today were designed for operability and reliability during a time when security was a low priority. Smart Grid implementation is going to require the installation of numerous advanced control system technologies along with greatly enhanced communication networks.” (p. 4) This book has disused many of these advanced technologies that may need to be considered as part not only in the initial development of the security information analytics for Smart Grid but also in addressing incidents that occur.

3.4.4 Action: Mitigating Risk

The Smart Grid security information analytics for Smart Grid includes the effective risk management discipline discussed earlier, employing a broad range of factors to make probabilistic decisions about risk and take prioritized actions, including alerts to response teams, to recover from incidents and remediate vulnerabilities. But an incident may also indicate the opportunity to take actions to mitigate the risk associated with that incident.

For example, a well-prepared security teams will know what the organization’s valuable information assets are and which systems, applications and users have access to them. Awareness of these parameters help security analysts, to narrow their field of investigation during a breach so they can address problems faster and with greater confidence. But a given incident may indicate that the security operations teams should conduct a breach readiness assessment or institute

Practice drills to improve the speed and efficacy of their reactions to cyber-attacks. They may need to revise their inventory of high-value assets that must be protected based on their new knowledge of what is attractive to an attacker. They may need to review their security policies again business priorities and regulatory requirements.

The diagram below, expanding on a similar diagram in the Popovik paper (2013), provides an example of a process for mitigating risk in response to incidents such as detected intrusions.

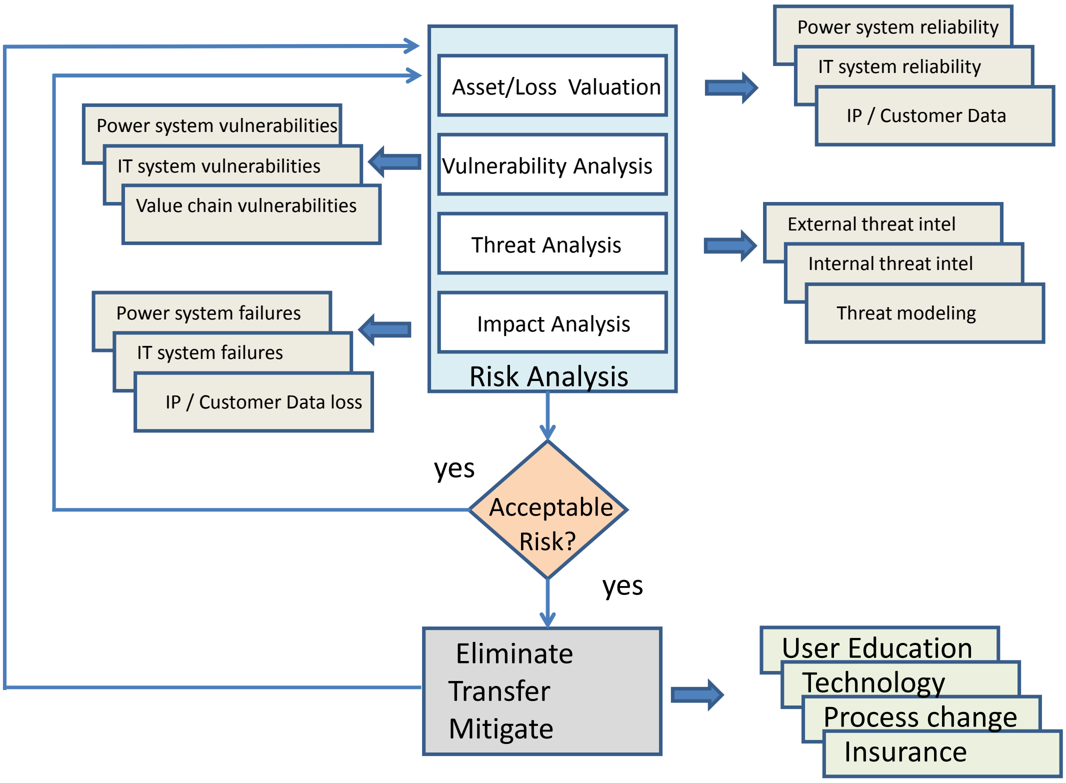


Figure 7: Risk Mitigation

Such a process can take advantage of operational incidents, regardless of whether they result from a security incident, equipment failure, natural disaster or any other cause, to enable organizations to take new actions to progressively improve their processes, optimize staffing and skills, modify their technology platforms, change their supplier relationships or take any other of the multiple of actions that could help them better address the risk of such an incident.

These improvements can be assisted by the technology advancements in big data and security analytics systems that deliver “imagine if” capabilities. The bounds of what’s imaginable are now being explored by operations professionals and business leaders together. For organizations concerned about an effective operational model, these “imagine if” scenarios often focus on injecting better intelligence and context into both operational and security practices. For example, if we apply new analytic approaches to historical data, what could we learn? What do the cyber-attacks we’ve encountered tell us about our business and operational risks? If we add new log sources or external intelligence feeds to our data warehouse, what patterns could we look for that we couldn’t even imagine seeing before? What types of intelligence might help us hunt down threats or respond to operational incidents more quickly, including through automated capabilities that do not require human intervention?

An effective security information analytics for Smart Grid for Smart Grid should ensure that this connection between incident response and the risk management process is established and effective.

3.5 The Human Factor as an Asset

3.5.1 Engaging the Employee Community

The complex interplay among people, process and technology in security operations makes it challenging to adjust any one element without also adjusting the others. Harmonizing tools, skills and methodology in operations is essential to providing defence-in-depth and to protecting the organization's critical information assets by enabling the employee community to be more effective. Additionally, perfecting the people-process-technology triad can unlock operational efficiencies by automating routine tasks and streamlining workflows.

If this harmonization is achieved, the result is that operations administrators, security analysts and other staff will spend far less time tracking down information for an investigation or researching the status of an incident. Instead, they can focus their time on enriching intelligence sources, uncovering subtle irregularities in their IT environments that point to serious problems, or hunting down covert threats faster.

Process integration is an important aspect of improving the impact and contributions of the employee community. It eliminates many routine steps, such as copying-and-pasting incident information, that go along with manually joining disparate security operations workflows. Integration also reduces opportunities for error, because activities for complex processes such as incident response can be programmed to follow a deterministic sequence of actions based on best practices. Finally, process integration can facilitate cooperation among different parts of the business—among audit, information security and compliance, for example—and help organizations create a unified view of conditions and risks throughout the organization.

This collaboration across the enterprise is essential in creating a culture of engagement that enables effective operations. Such a culture can have a dramatic impact on security as well, by making encouraging a personal commitment to security by every employee. While social engineering attacks continue to remain a significant threat, the awareness of individual responsibility for security can be a powerful force in helping each employee recognize and avoid responding to such phishing, pharming and vishing attacks.

3.5.2 Engaging the User Community

The importance of engaging the user community is a significant focus in the 2010 "Smarter Energy and Utilities" paper by IBM. It calls out the importance of educating the consumer, for example, pointing to the Mobile Experience Center created by Oncor: "The Center is essentially a "smart home" on wheels that travels throughout the state, allowing customers to experience first-hand how they are now able to make more informed choices about their energy consumption and expenditures than ever before." (p. 4)

Approaches to engaging the user community have been explored in detail in "A 15-Minute Guide to Smart Grid Correspondence with Utility Customers". (EMC, 2010) The paper recommends an integrated communications strategy that includes the following five components in order to engage the customer in a positive attitude toward Smart Grid:

- **Integrated content**. Regardless of the content, information shared with the customer must accurately and consistently represent product and service capabilities. Managing the content centrally helps to ensure that this is true of all contact with the customer.
- **Effective design tools.** Communications should be personalized with specific customer information and content, but should also follow business rules governing design variables, customization parameters and the use of customer data.
- **Multichannel content generation**. Composition and formatting capabilities must support a broad set of electronic and print formats, while meeting scalable production demands.
- **Content archiving**. Legal and regulatory constraints demand the archiving of every piece of utility correspondence. A comprehensive achieve also supports an effective customer relationship.
- **Enterprise integration**. Utilities must be able to integrate document personalization and generation services into existing enterprise applications.

Engaging the user community has significant benefits in terms of their confidence in the value and effectiveness of Smart Grid. As with the employee community, it also has significant benefits in terms of their confidence and participation in Smart Grid security, as essential for the user as it is for the utility.

3.5.3 Security Shared is Security Strengthened

This chapter has touched a number of times on the importance of collaboration as part of the Smart Grid operational model. For example, if security analytics platforms integrate threat intelligence from outside sources, organizations can see the threat landscape as a panorama, not just from the narrow aperture of their own internal IT environments. Enhanced visibility will lead to enhanced security capabilities, vastly expanding options for how security operations centres (SOCs) act and respond to prospective threats.

Most utilities have seen an increase in cooperative efforts between engineering, owners of the operational technology; and IT, owners of information and communications technology. Utilities have not yet crafted a formal organizational structure to handle ownership of OT and IT. At this point, "convergence" is taking place for select sets of data (e.g., meter signals fed to outage management). Convergence is also occurring from a data security perspective to ensure CIP compliance and adherence to applicable privacy laws. To a lesser extent, utilities are also seeing convergence when it comes to grid asset management. In other cases, data convergence between OT and IT is still future oriented, unfolding, and less transparent, especially for transmission (e.g., synchro-phasor sensor/apps, renewables), engineering/system performance, and marketing.

The “Smarter Energy and Utilities” paper calls out the importance of extending this “collaboration mindset” (IBM, 2010, p. 3) to collaboration across the industry. A number of collaborative forums already exist that are facilitating this collaboration, such as the European Atomic Forum (FORATOM), the Electric Power Research institute (EPRI), the Global Intelligent Utility Network Coalition, and the Global Smart Grid Federation. Participation in forums such as these is part of an effective Smart Grid operational model, as should participation in academic and industry research and collaborative engagement with both existing and new partners in risk management, operational effectiveness and cyber security.

4 The SPARKS Security Information Analytics Architecture

The SPARKS's Security Information Analytics mini-project is designed as an external component, which will be connected to the SCADA Open Platform Communication (OPC) [1] server and the other sources of information available in the system (for example network logs, system access log. or the others Sparks' mini-projects,..).

SCADA networks have a star topology with an OPC Server as the central node. The OPC Server for SCADA provides connectivity to compliant devices such as any PLCs, RTUs, DCSs and other devices. Commonly, the OPC Server has features such as Advanced Polling Engine that polls all devices and excludes the disconnected devices from the polling cycle. These kinds of features are designed to increase the resilience of the system. For example, once the disconnected device comes back online it gets re-added to the polling cycle in order to prevent any delays when multiple devices are connected via a common communications channel.

The purpose of OPC is to define a common interface that is written once and then reused by multiple components such as business applications, SCADA severs, Human Machine Interface (HMI), or custom software packages.

As shown in Figure 8 the mini-project receives data from the OPC server and uses a web-based graphical user interface to present alerts and anomalies to operators.

An additional set of APIs will be provided so that the Security Information Analytics (SIA) server can be easily interrogated from external security software or other enterprise application software.

The SIA system is passive, in that it only detects errant behaviour in the broader system, but has no reactive capability. For this reason integrating SIA systems with reactive systems such as the SPARKS' Resilient Control System is a valuable enhancement opportunity.

The mini-project uses the data provided from the Nimbus microgrid test-bed, which contains and monitors many different devices. The Security Data Analytics development is carried out with that device diversity in mind, making the system more generic and enabling the testing of different algorithms for each subsystem connected to the microgrid.

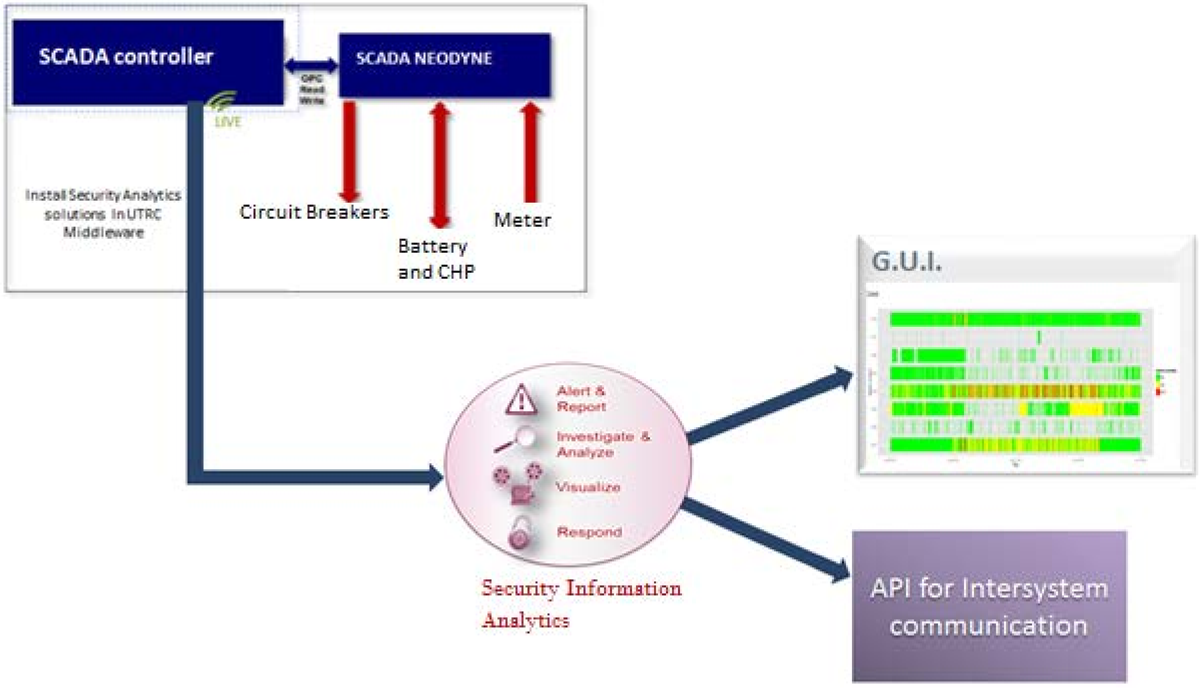


Figure 8: SPARKS' Security Information Analytics high-level architecture

4.1 High level approach to Nimbus

The following section describes the NIMBUS micro grid dataflow, and introduces the Security Data Analytics architecture, internal structure and finally the GUI.

4.1.1 Nimbus Microgrid Dataflow for Security Analytics

The NIMBUS microgrid integrates an electrical microgrid and a thermal system. The test-bed is managed using a hierarchical control strategy, with local controllers of subsystems feeding up to power system coordination and energy optimization systems. This structure incorporates all of the major components in microgrids, such as renewable generation, storage and different kind of interfaces. Such a small-scale power plant requires a hierarchical control strategy to assure the optimal operation of all of the sub-systems. It involves a multidisciplinary approach which includes information and communication technologies (ICT), power systems, power electronics, controls and optimization, and diagnostics. Figure 9 shows an overview of the energy test-bed.

The electrical microgrid incorporates:

- A 10kW wind turbine,
- A 35kWh (85kW peak) Li-Ion battery,
- A 50kW electrical/82kW thermal combined heat and power unit (CHP),
- A feeder management relay to manage the point of coupling between the microgrid and the national grid, and
- A set of local loads.

An extensive network of meter and sensors has been deployed to facilitate the measurement and data collection necessary for control of the electrical and thermal systems. These measurements together with relevant information about gas and electricity power consumption measurements and prices, as well as thermal/electrical loads and weather and wind forecasts, are continuously available from the system. The dataflow cascades in a hierarchical manner as described below:

Data Measurement:

The system data is collected by a wide range of meters and sensors. The meter measurements make up the bulk of the collected data, accounting for 232 variables.

1) The primary points of data collection are the eight 3-phase electrical meters (labelled EM-01 through EM-07 and EM-10 in Figure 9). Each of these meters continuously measure 29 variables:
   a) Three phase-neutral voltages,
   b) Three phase-phase voltages,
   c) Four line currents (three phases and neutral),
   d) Active power, reactive power and apparent power per phase,
   e) Total active power, reactive power and apparent power,
   f) Power factor,
   g) Frequency,
   h) Active energy (positive and negative), reactive energy (positive and negative) and apparent energy

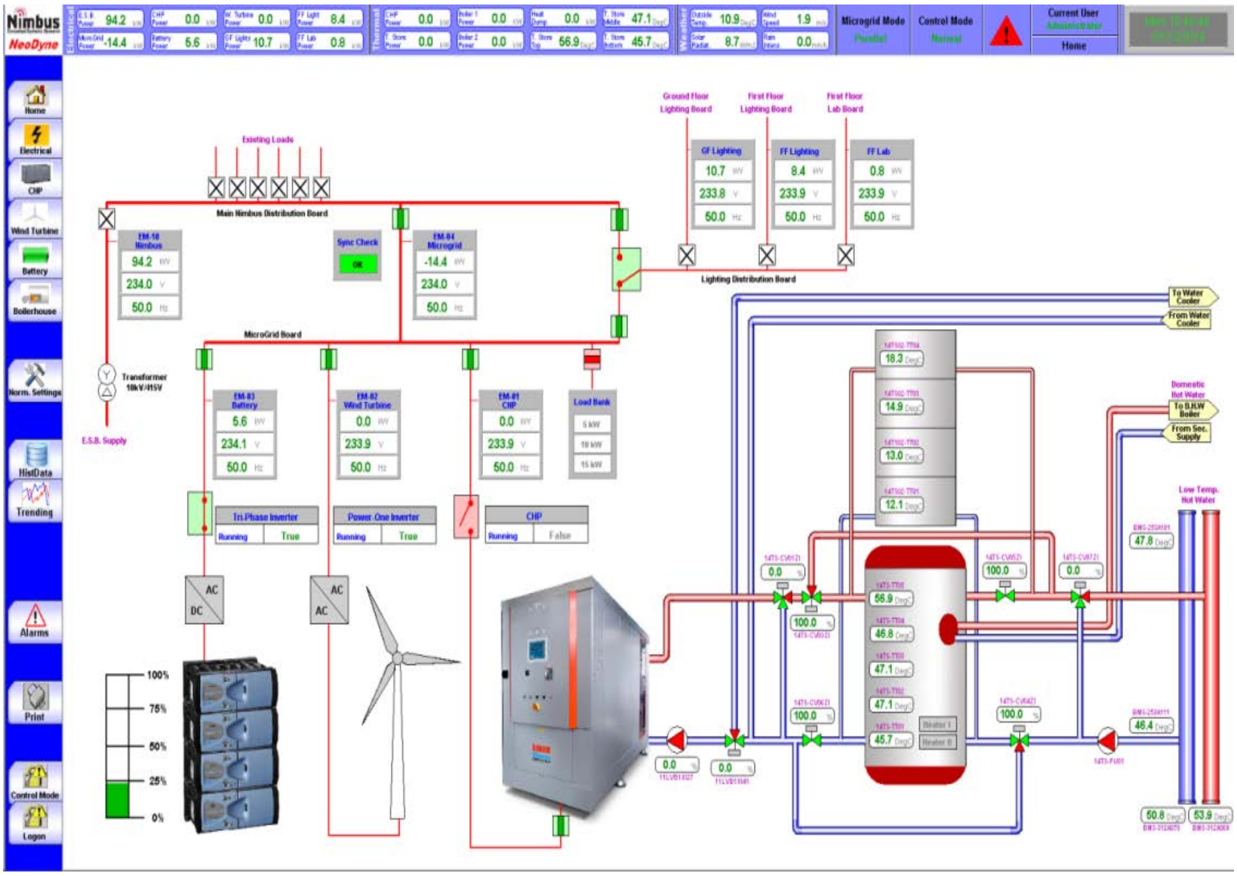


Figure 9: Nimbus SCADA Overview

Apart from the meters which measure the general health of the system, each component records the values pertinent to its operation:

2) The battery and its AC-DC inverter record the system AC and DC voltages, currents, AC active, reactive and apparent powers, AC power factor, AC frequency, DC power, internal battery voltage and current, the battery state-of-charge, state-of-health, the inverter efficiency, and the current active and reactive power set points. They also communicate the recommended and maximum charge and discharge voltages, currents and powers for on-the-fly comparisons.
3) The CHP and the synchronous generator that connects it to the main grid together record active and reactive powers and the power factor of the generator, the phase angles, voltages, frequencies of the mains and generator, the phase-phase and phase-neutral voltages of the generator, along with the electrical and thermal power, and the volume of gas consumed and the power, thermal and total efficiency, power set point (active and reactive) and planned power. Apart from these electrical measurements, it also records a host of thermal and mechanical measurements, including water temperatures at flow and return, temperatures before and after mixture cooling, mixture temperature, cooling water temperature (before and after mixture cooling), exhaust temperature, engine cool water temperature and pressure, air intake temperature, temperatures before and after catalyst, oil and charge pressure, flow regulation, circulating water temperature change, exhaust counter pressures, emergency and mixture cooler regulation, the external and casing temperature, and the engine RPM.
4) The wind turbine is connected to the grid through three synchronised single-phase inverters, each of which records the grid and input voltage, current, power, the grid frequency, and energy for the day as well as total, and the inverter and booster temperatures.
5) The two thermal tanks (water and phase-change material) record the material temperatures at different levels in the tanks.
6) The battery, CHP, wind turbine inverters, and the thermal storage also contain internal checks, which generate alarms and warnings. These alarms and warnings are also communicated to the system.

Data Collection:

The data variables collected above are communicated by the meters and system components to the Siemens S3700 PLC. The PLC also logs other data from the system, such as the position indicators of all the control valves, and the breaker statuses. Since the PLC also acts as the conduit for the commands sent to the system, such as changes to the system mode (normal or experimental), and the power, valve and pump set points, close/open commands to the breakers, load bank manual set points, and manual overrides for control valves, heating elements, all the pumps, and the mains-or-microgrid changeover commands. In total, 1252 variables are logged every second.

Data Display and Logging:

The PLC communicates the data to the SCADA PC, which runs the human-machine-interface tool shown in Figure 9. This tool displays the important monitoring variables on the home screen, while the rest of the variables are grouped according to their source and displayed in one of the 23 sub-displays. The HMI also serves to display and acknowledge system alarms and warnings.

Data extraction into a data base is done via the HMI's historical data tool. The sampling time of the data can be chosen at the time of retrieval.

UTRC Middleware:

The UTRC middleware is hosted on a PC on the same network as the SCADA PC. The middleware PC uses OPC to periodically request the current variable values from the SCADA interface, which it parses and stores in another database on its hard drive. The middleware also acts as the interface for any client (e.g., Matlab) to access the data using Simple Object Access Protocol (SOAP).

4.1.2 Hardware and software specification

The prototype of the mini-project has been developed using RStudio Server [2] Version 0.98.1091 running on an Ubuntu Linux 14.04.2 LTS virtual machine.

The virtual machine runs inside a VMware ESXi server 5.5 [3] on an Intel physical rack server model S2600GZ with 24 physical Intel Xeon E5-2697 v2 processors running at a clock speed of 2.7 GHz CPUs, equipped with 250 GB of RAM and 10 TBs of storage. The virtual machine has allocated 32 Virtual CPUs (8 Xeon processors with 4 cores each), 142 GBs of RAM and 2.7 TBs of disk.

This development system has far more computing resources than strictly necessary for an operational system. This was done to de-risk the algorithm test and debug cycle. In a real operational environment it would be possible to tailor the hardware requirements giving due consideration to the detection type required and the amount of data generated by the target SCADA network.

The core components of the prototype are developed using the R programming language [10], which is an interpreted language widely used by data analytics groups in academia and industry. The R interpreter can run in both Windows and Linux.

The R standard library implements a wide variety of statistical and graphical techniques, including linear and nonlinear models, classical statistical tests, time-series analysis, classification, clustering, etc… The language is easily extensible through functions and libraries. Many of R's standard functions are written in R itself, which makes it easy for users to follow the algorithmic choices made. The extensive use of R's library allows for fast prototyping and easy data exploration.

RStudio [2] is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. The server version of RStudio provides a web browser interface to an RStudio instance without the need for any software to be installed on the developer's machine. This provides a consistent development environment for the team to work on while simplifying integration.

The Shinyapp package is an R framework designed to facilitate the construction of R based web applications. The graphical user interface for the SIA for Smart Grids package was developed using this framework.

4.1.3 Connection with SCADA system

To receive the data from the SCADA network the Security Information Analytics machine uses the Secure Shell (SSH) File Transfer Protocol (SFTP) [6] server that is configured to accept files into a configurable system directory. Currently the system processes data in the form of Comma Separated Value (CSV) files [7], but it also supports Excel, SPSS, SAS, Stata and SYSTAT files.

The Cron task scheduler software—the Linux equivalent of Microsoft's Windows Task Scheduler—is used to launch the script that processes the received data. The scheduled time for these jobs depends on the chosen polling rate and the time of data extraction from the SCADA server. The polling rate depends heavily upon the type and quality of detection that is required.

4.1.4 Intersystem API communication

During the first year of the SPARKS project, the potential for integrating multiple mini-projects was identified. This approach could address different threats and cover more complex scenarios like simultaneous multi-facetted attacks. In essence, the system would have a modular structure that can communicate and create a more complete picture of what is happening in the monitored system. Each mini-project should expose some interface to allow for communication with the other systems.

Currently that integration is in the definition state; however the use of Representational State Transfer (REST) API [8] has emerged as the most popular approach for this kind of communication. REST is a software architecture for creating scalable web services and is simpler than SOAP or WSDL-based web services.

RESTful systems typically communicate over the Hypertext Transfer Protocol using HTTP verbs such as GET, POST, PUT, DELETE, etc., which are commonly used by web browsers.

4.2 Security Information Analytics internal structure

The mini-project is composed of two separate modules that take care of different aspects of the SIA detection process:

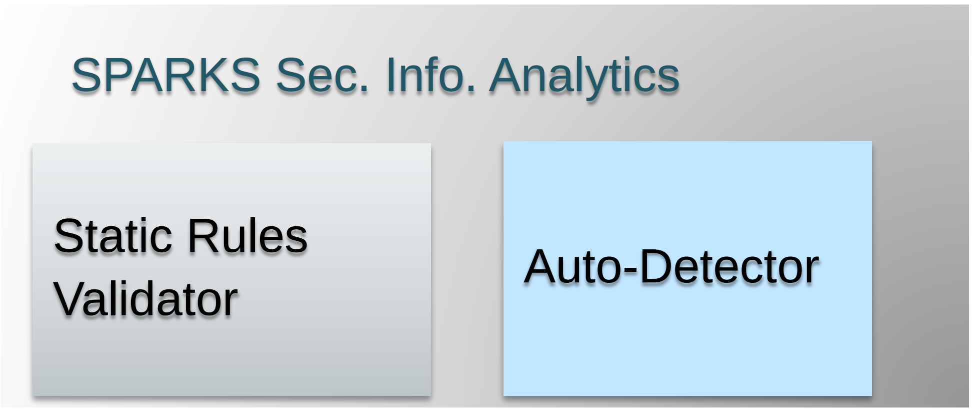


Figure 10: Security Information Analytics internal architecture

- **Static Rules Validator**: Takes into account the static behaviour of the SCADA systems and the validation of the physical rules present in an industrial process.
- **Auto-Detector**: Uses dynamic algorithms to find complex relationships between variables, discovering underlying patterns in the data to determine the state of the system, and normal and anomalous behaviours.

4.2.1 Static Rules Validator

The *Static Rules Validator* was designed to be easily customizable, taking into account that SCADA systems present in critical infrastructures are significantly different depending on when they were designed and the aim of the system. The Static Rules Validator module is composed of three components:

- **Rules List:** A list of rules that the variables in the system must obey. Domain experts are able to customize this list to include any number of rules. The rule definition is done via a simple list of strings, each containing a single physical rule that the SIA system will check. In addition, if there is a priori expectation for the nominal value of variables it is possible to determine if these measured values exhibit outliers. An example of a physical rule is the *Phase A Apparent Power 1* (***J11***) which is calculated as *Phase-Neutral Voltage 1* (***E04***) multiplied by *Phase 1 Current* (***I01***) dived by 1000 is represented in formula form as:

```
(E04 x I01 / 1000) = J11
```

- **Adapter:** This component receives the rules list and maps the formula variables to their respective entries in the dataset. For example, the *Phase A Apparent Power 1* rule will be transformed into:

```
(Value.11LVEM10E04 x Value.11LVEM10I01 / 1000) = Value.11LVEM10J11
```

- **Parser:** The parser applies the rules to the dataset, searching and reporting on any outliers or violations of the rules.

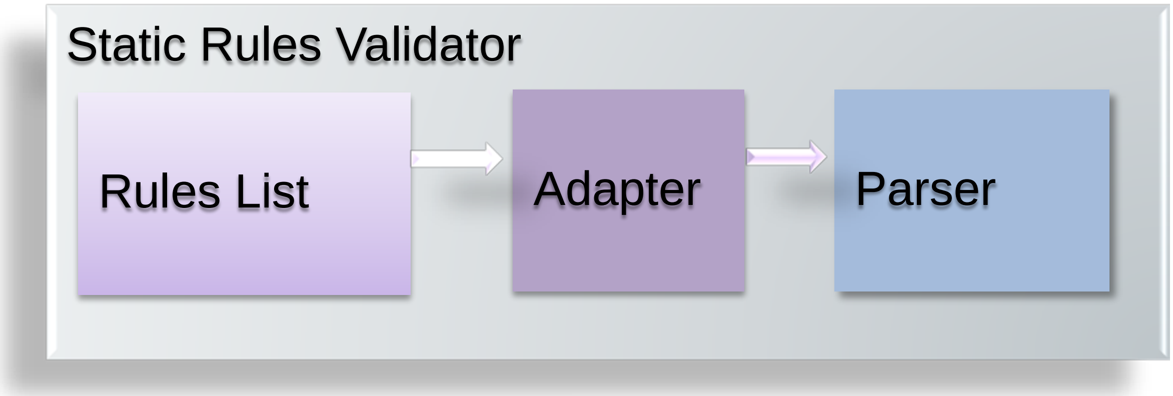


Figure 11: Static Rules Validator internal architecture

This approach of separating the language of the rules from the dataset implementation makes the system more portable. If the system is migrated to a new environment, the system can be adapted by simply editing the rules list and the adapter mapping.

4.2.2 Auto-Detector

This module will use machine-learning techniques to evaluate the entire system state looking for correlations between variables and the system's internal state. This component implements real-time detection capability so it needs to periodically pull data from the monitored system.

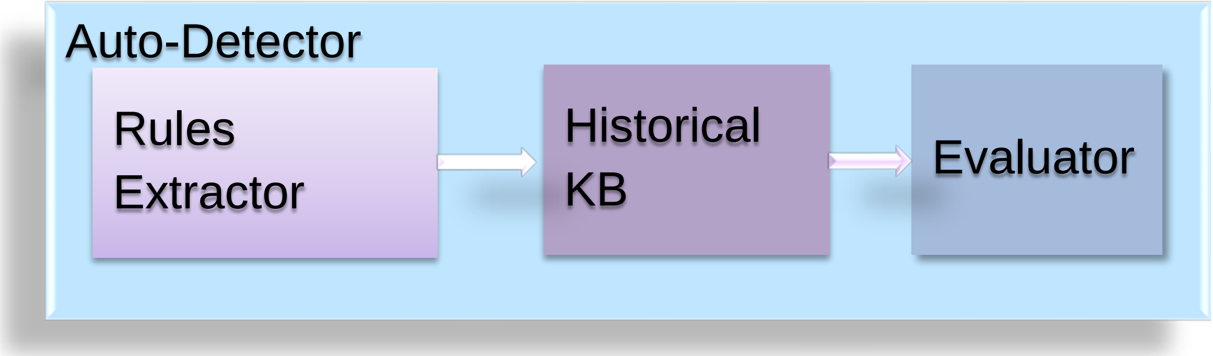


Figure 12: Auto-detector internal architecture

This module is composed of the following components:

- **Rules Extractor:** read data from last SCADA readings received and apply data reduction techniques, it also changes the data format to improve the process performance
- **Historical KB:** extract features form the current SCADA readings and check it against the system historical data;
- **Evaluator:** these modules implement tolerance and correlation checks to reduce False Positive and background detection noise

4.3 Security Data Analytics G.U.I.

The SIA for Smart Grids can operate in online mode, processing the data as it is received in the system, updating the GUI with the new results, showing the “live” state of the system.

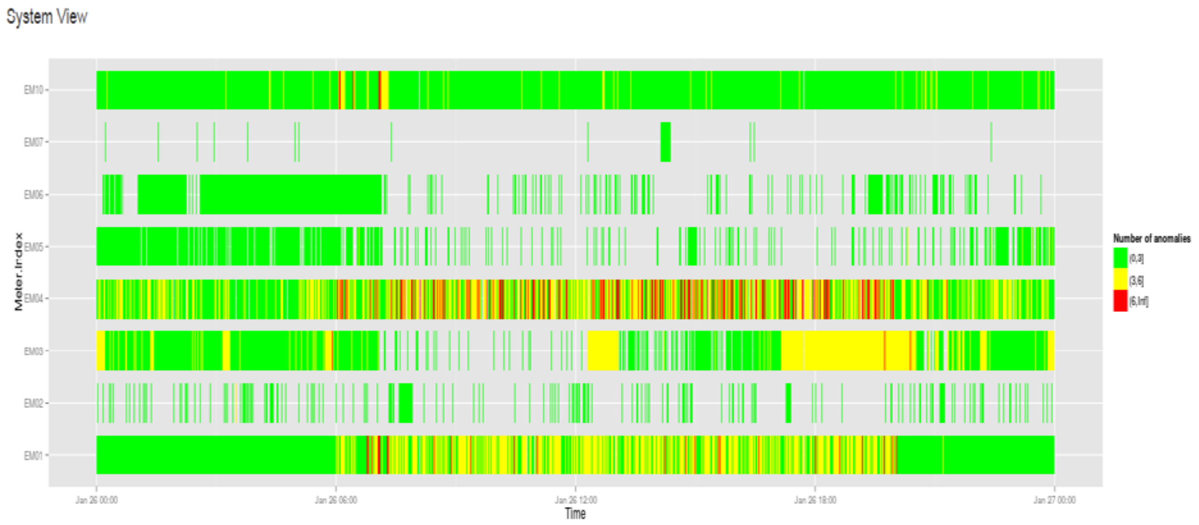


Figure 13 System View section of GUI

The system can also operate in offline mode, where the operator can feed in a previously captured dataset for analysis.

The SIA for Smart Grids GUI is divided into different sections; the first section provides a *System* view of the status of the system (Figure 13). Each row represents the status of each meter throughout the day (00:00–23:59), showing the total number of rules/equations that generate outliers.

Next the status of each meter is shown, providing more detail on the status of each meter. Figure 14 shows the status of the meter “EM01”. The blue crosses represent the number of variables that generate outliers compared to their stated nominal value. In this case, one outlier was generated at approximately 00:00 and 21:00.

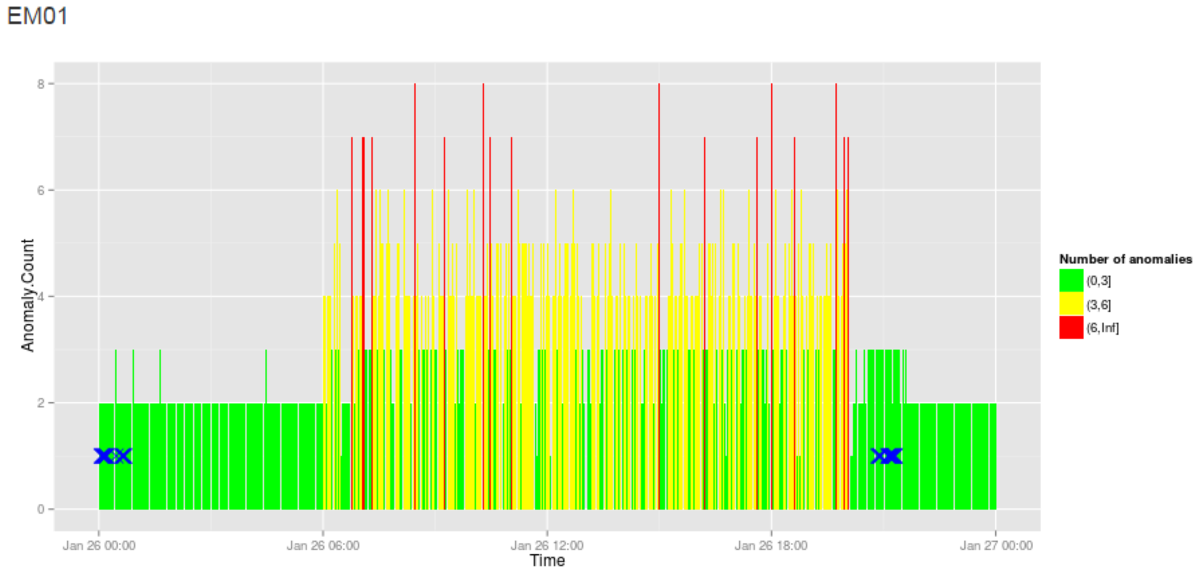


Figure 14: Visualization of the anomaly detection algorithm for meter EM011

4.4 Example of mass disconnection attack detection

The purpose of this analysis is to detect anomalous behaviour in the grid in the form of multiple sensors dropping. For example, an attacker (authorized or otherwise) issues a command which when executed leads to a mass meter remote disconnect as reported in NESCOR FAILURE SCENARIO - AMI.1 [11].

Initially, dead sensors are grouped together using a moving time-window. If the number of dead sensors in a group exceeds a predefined threshold the cluster is flagged as anomalous. When a sensor ceases to operate, a time counter is started. If another sensor drops within a pre-set time window (e.g., 3 minutes), it is considered to be in the same cluster as the first sensor. The timer is then reset to the time the second sensor died and the count begins again.

The process continues until the clustering time-window lapses with no more dropping sensors. The cluster is then decayed and no more sensors are added to it. Sensors in a decayed cluster are not considered for subsequent clustering. Sensors which come back online are removed from decayed or current clusters and are considered in the clustering process. If the size of the cluster exceeds the pre-set maximum threshold, it is considered anomalous and flagged to the operator immediately. In this case the operator may issue an override disconnection command to stop the attack.

Reporting is done via the web interface, by denoting anomalous clusters in red and benign clusters as green. An example of the interface is shown in Figure 15.

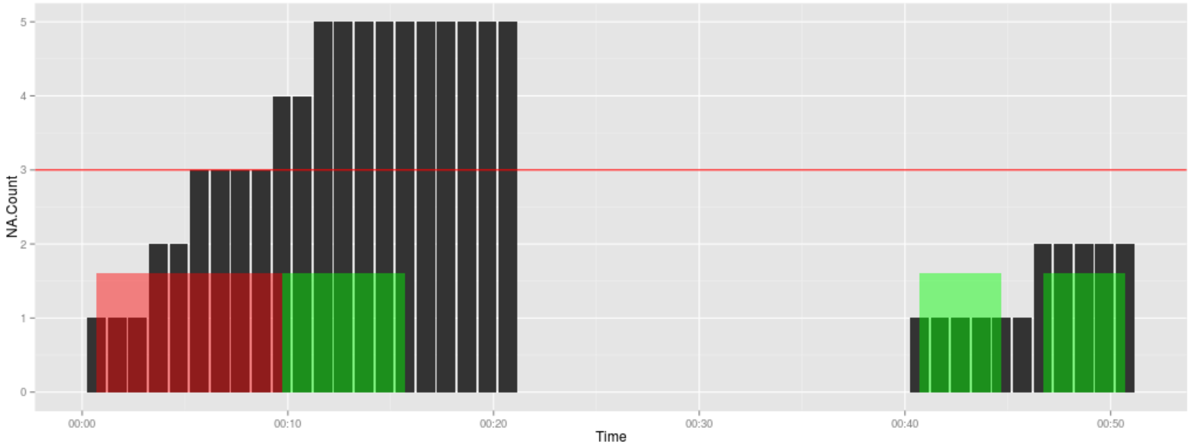


Figure 15: Dashboard output of the cluster detection system.

Shown is the total number of dead sensors as a function (solid bars – *NA.Count* in Figure 15). Anomalous clusters are highlighted in red and benign clusters are shown in green. In the example above a clustering window of 3 minutes was used and an anomalous cluster is defined as any cluster with 3 or more dead sensors.

The clustering algorithm has detected four clusters; an anomalous one, and three benign ones. The anomalous cluster decays with all three of the sensors still down. The second cluster consists of only two sensors, and as such is benign. Note that the sensors associated with the first cluster have not come back online, but they are not included in the second cluster.

5 Conclusion

Once deployed the Security Information Analytics can result in dramatic changes in how decisions are made as well as the rate at which they are executed.

We presented in this deliverable the common security data analytics approach and the system architecture of the SPARKS Security Information Analytics for Smart Grids mini-project.

One of the major issues encountered during the implementation phase of the mini-project was related to system diversity and the absence of dominant standards, architecture or installation practice. To mitigate this issue the system was designed to adapt itself to the target monitoring environment with minimal effort.

As shown previously, the early results obtained from the system are interesting and clearly validate the approach taken. Going forward we will focus our attention on the integration with and the interaction between the other SPARKS mini-projects.

It may be possible to create a scenario where we will test all the mini-projects together, using different attacks and in particular simultaneous, multi-facetted attack scenarios.

In these advanced scenarios the integration between SPARKS mini-projects will be crucial for the successful detection and remediation of advanced attacks. Thus the communication interfaces and the data exchanged between systems needs to be careful designed and implemented.

6 References

This section lists the resources referenced in this document.

1. SCADA OPC Server http://en.wikipedia.org/wiki/Open_Platform_Communications
2. Rstudio web site https://www.rstudio.com/
3. VMware ESXi Overview http://www.vmware.com/products/esxi-and-esx/overview
4. Ubuntu 14.04.2 LTS (Trusty Tahr) http://releases.ubuntu.com/14.04/
5. Intel® Server Boards S2600GZ and S2600GL http://www.intel.com/content/www/us/en/motherboards/server-motherboards/server-board-s2600gl-gz.html
6. SSH File Transfer Protocol http://en.wikipedia.org/wiki/SSH_File_Transfer_Protocol
7. Comma-separated values http://en.wikipedia.org/wiki/Comma-separated_values
8. Representational state transfer http://en.wikipedia.org/wiki/Representational_state_transfer
9. Shiny by RStudio http://shiny.rstudio.com/
10. The Comprehensive R Archive Network http://cran.r-project.org/
11. National Electric Sector Cybersecurity Organization Resource (NESCOR). “Analysis of Selected Electric Sector High Risk Failure Scenarios, version 1.” 2013. http://SmartGrid.epri.com/doc/nescor%20detailed%20failure%20scenarios%2009-13%20final.pdf
12. https://www.gedigitalenergy.com/smartgrid/Aug07/PMU_disturbance_recording.pdf
13. ESRI.(2011). Enterprise GIS and the Smart Electric Grid. Retrieved from http://www.esri.com/library/whitepapers/pdfs/enterprise-gis-smart-electric-grid.pdf
14. Ragan, Sean Michael. (2010, December 10) Total system failure completely annihilates power station. *Makezine*. Retrieved from http://makezine.com/2010/12/10/total-system-failure-completely-ann/
15. Tweed, Katharine. (2014, February 6). Attack on California Substation Fuels Grid Security Debate.*IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/energywise/energy/the-smarter-grid/attack-on-california-substation-fuels-grid-security-debate
16. Symantec. (2014, June 30). Dragonfly: Western Energy Companies under Sabotage Threat. Retrievedfrom http://www.symantec.com/connect/blogs/dragonfly-western-energy-companies-under-sabotage-threat
17. Symantec (2014, October 14) Sandworm Windows zero-day vulnerability being actively exploited in targeted attacks. Retrieved from http://www.symantec.com/connect/blogs/sandworm-windows-zero-day-vulnerability-being-actively-exploited-targeted-attacks
18. Govindasaru, Manimaran et al. (2012). Cyber-Physical Systems Security for Smart Grid. Retrieved from http://www.pserc.wisc.edu/documents/publications/papers/fgwhitepapers/Govindarasu_Future_Grid_White_Paper_CPS_Feb2012.pdf
19. Cardenas, Alvaro. (2013). Big Data Analytics (and Security Intelligence) in Smart Grid Applications. *IEEE ISGT Conference*. Retrieved from http://sites.ieee.org/isgt/files/2013/03/Cardenas3C.pdf

20. Oltsik, Jon. (2014) Information-Driven Security and RSA Security Analytics and RSA ECAT. *Enterprise Strategy Group.* Retrieved from http://i.crn.com/custom/ESGWhitepaperIntelligenceDrivenSecuritywithRSASecurityAnalyticsandRSAECAT.pdf
21. Feblowitz, Jill et al. (2013). Using Information Intelligence to Improve Projects in the Energy Sector. *IDC*. Retrieved from http://www.emc.com/collateral/analyst-reports/idc-using-info-intelligence-to-improve-projects.pdf
22. Vermesan, Ovidiu et al. (2013). Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems. Retrieved from http://www.internet-of-things-research.eu/pdf/Converging_Technologies_for_Smart_Environments_and_Integrated_Ecosystems_IERC_Book_Open_Access_2013.pdf
23. Searle, Justin. (2012). AMI Penetration Plan. Retrieved from https://media.blackhat.com/bh-eu-12/Searle/bh-eu-12-Searle-Smart_Meters-WP.pdf
24. Verizon. (2014) 2014 Data Breach Report. Retrieved from http://www.verizonenterprise.com/DBIR/2014/reports/rp_Verizon-DBIR-2014_en_xg.pdf
25. Oltsik, Jon. (2013). The Big Data Security Analytics Era is Here. *ESG*. Retrieved from http://www.emc.com/collateral/analyst-reports/security-analytics-esg-ar.pdf
26. Musser, Phil. (2009). Smart Grid Data Management and Analytics; what data to collect, integrate, analyze, act on and report. Retrieved from http://www.burnsmcd.com/Resource_/PressRelease/1386/FileUpload/WhitePaper-SmartGrid-Musser.pdf
27. Microsoft Corporation (n.d.). Windows Management Instrumentation. Retrieved from http://msdn.microsoft.com/en-us/library/aa394582%28v=vs.85%29.aspx
28. Zhang, Yagant et al. (2009) Fault detection based on discriminant analysis theory in electric power system. *IEEE*. Retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5347972&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D5347972
29. Tweed, Katharine. (2012) Fraud should be first priority for smart meter security. Retrieved from http://www.greentechmedia.com/articles/read/fraud-should-be-first-priority-for-smart-meter-security
30. UCTE. (2006). Final Report System Disturbance on 4 November 2006. Retrieved from https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/Final-Report-20070130.pdf
31. Security for Innovation Council. (2011) When Advanced Persistent Threats Go Mainstream. Retrieved from http://www.emc.com/collateral/industry-overview/sbic-rpt.pdf
32. Popovic, Tom et al. (2013). Smart Grid data analytics for digital protective relay event recordings. Retrieved from
33. https://www.academia.edu/7691468/Smart_grid_data_analytics_for_digital_protective_relay_event_recordings
34. IDC, ibid.
35. Feblowitz, Jill. (2013) Making Good on the Promise of Smart Grid: Information Management is Critical. *IDC*. Retrieved from

https://www.emc.com/collateral/software/white-papers/utilities-whitepaper-emc-idc-ei224742.pdf

36. EPRI. (2010). Smart Grid Roadmap and Architecture. Retrieved from http://www.smartgrid.epri.com/doc/cal%20iso%20roadmap_public.pdf
37. RSA, the Security Division of EMC. (2013) Big Data Fuels Intelligence-Driven Security. Retrieved from http://www.emc.com/collateral/industry-overview/big-data-fuels-intelligence-driven-security-io.pdf
38. Popovic, ibid.
39. Popovic, ibid.
40. Swearingen, Michael et al. What you need to know (and don't) about the AURORA vulnerability. Retrieved from http://www.powermag.com/what-you-need-to-know-and-dont-about-the-aurora-vulnerability/
41. Cal ISO. (2010) Smart Grid Roadmap and Architecture. Retrieved from http://www.caiso.com/green/greensmartgrid.html
42. Hwang, Hyeon Keong et al. (2012) . Security Engineering Report on Smart Grids. Retrieved from https://www.academia.edu/2908216/Security_Risk_Analysis_for_Smart_Grid
43. United States Department of Energy (DOE). (2009). Study of security attributes of Smart Grid – Current Cyber Security Issues. Retrieved from http://www.inl.gov/scada/publications/d/securing_the_smart_grid_current_issues.pdf
44. Popovik ibid.
45. IBM. (2010) The state of smarter energy and utilities. Smart Industries Symposium, Barcelona. http://de.slideshare.net/j3juliano/state-of-smarter-utilities
46. EMC. (2010) A 14-minute guide to smarter correspondence with utility customers. Retrieved from: http://www.emc.com/collateral/software/15-min-guide/h5108-15-min-smart-grid-correspondence-gd.pdf