SMART Manufacturing & Cyber Security: Foundational ISA/IEC 62443 Standards Evolving with Learning Machines

April 10, 2018 (2^2/2^25/2^13^2)

Bradley D. Taylor, D.Sc.
Department of Electrical Engineering & Computer Science
School of Engineering
The Catholic University of America
contact@VirtualCommonGround.org
Bellingham and Consequences

- Broadcast storm shutdown SCADA and Delayed Leak Detection
  - Loss of View, Loss of Control
- All sensors set to average values and safety systems didn’t actuate
  - Loss of Safety
- Requires revisiting cyber security and safety standards
Last Week

- Attack on Natural Gas Network Shows Rising Cyberthreat (4/6/18)
- Pipeline Firms Hit; Gas Still Flowing (4/4/18)
- “3 of 4 companies operating pipelines admitted they were hit by a cyberattack this week”
SMART Manufacturing Cyber Security Standards Challenges

- Heterogeneous Organizations
  - Manufacturers, Suppliers, Integrators, Governments: *Domain & Size* varies widely

- Heterogeneous Instruments, Automated Control Systems
  - Engineered systems’ *Protocols still compete*

- Heterogeneous Languages: *even within same Natural Language! Interpretation & Communication barriers (Ambiguity & Redundancy)*

- Adversaries: Actors (state & non-state); Mismatched lifespans & evolutionary time constants

- How to avoid an automated *Tower of Babel*?
Outline

- Introduction
- Problem Description
- Foundational & Ongoing IA&CS Cyber Security Standards Development
- Platform for Evolving Threats
- Conclusions & Future Work
ISA 99 Global Coverage: Role, Membership & Industries

- Scope: IACS compromise could result in:
  - “endangerment of public or employee safety
  - environmental protection
  - loss of public confidence
  - violation of regulatory requirements
  - loss of proprietary or confidential information
  - economic loss
  - impact on entity, local, state, or national security”

- 900 members world-wide

- Sector expertise:
  - Chemical Processing
  - Oil & Gas
  - Food & Beverage
  - Energy
  - Pharmaceuticals
  - Water
  - Manufacturing
  - ICS suppliers
  - → Medical ++

- Product:
  - ISA/IEC 62443 series of standards
The Basics

- General Concepts
- Fundamental Concepts
- Foundational Requirements
General Concepts

- Security Context
- Security Objectives
- Least Privilege
- Defense in Depth
- Threat-Risk Assessment
- Supply Chain Security

Source: ISA-62443-1-1, 2nd Edition (Under development)
Fundamental Concepts

- Principal Roles
- Life Cycles
- Zones and Conduits
- Security Levels
- Maturity Assessment
- Security and Safety

Source: ISA-62443-1-1, 2nd Edition (Under development)
Principal Roles

- Product Supplier (PS)
- Integration Provider (IP)
- Asset Owner (AO)
- Maintenance Provider (MP)
- Service Provider (SP)
- System Operator (SO)
- Regulatory Authority (RA)
- Compliance Authority (CA)
Life Cycles

Security Documentation
Security Guidelines
Security Support

Product Development
Product Supplier

Integration / Commissioning
System Integrator

Operation & Maintenance
Asset Owner

Requirements
Zones & Conduits

- A means for defining…
  - How different systems interact
  - Where information flows between systems
  - What form that information takes
  - What devices communicate
  - How fast/often those devices communicate
  - The security differences between system components
- Technology helps, but architecture is more important
Security Levels

Protection against…

4
Intentional Violation Using Sophisticated Means with Extended Resources, IACS Specific Skills & High Motivation

3
Intentional Violation Using Sophisticated Means with Moderate Resources, IACS Specific Skills & Moderate Motivation

2
Intentional Violation Using Simple Means with Low Resources, Generic Skills & Low Motivation

1
Casual or Coincidental Violation
Maturity Assessment

- A means of assessing capability
- Similar to Capability Maturity Models
  - e.g., SEI-CMM
- An evolving concept in the standards
  - Applicability to IACS-SMS
Security & Safety

- Safety: much of the reason for security
  - Presenting consequences
- Much learned from safety community
- Collaboration
  - ISA99-ISA84 joint effort
  - IEC TC65 work group 20
  - ISA Safety and Security Division
Foundational Requirements

- FR 1 – Identification & authentication control
- FR 2 – Use control
- FR 3 – System integrity
- FR 4 – Data confidentiality
- FR 5 – Restricted data flow
- FR 6 – Timely response to events
- FR 7 – Resource availability
Work Products include:
ISA-62443 Series
ISA 99

- 16 years’ effort: ISA/IEC 62443 standards series
- Evolution of standard continues:
  - Application to IoT
  - Devices on Level 0,1
- How do we handle Evolving/Imminent Threats?
- How do we *trust* incoming request?
- How does a human sentry react to a cyber packet sent at speed of light?
Computers

Human Computers at NASA
Learning Machines
Automated Learning Machines
ShOULDERS OF GIANTS
Automated Learning Machines: Helping Securely Share Remote Work

**WORKFLOW/YAWL**

**Ontology/Protégé**

**IOWF/JCoupling**

**Topic Modeling/LDA/LDAWN**

**Trust; But Verify!**
Virtual Common Ground

Organizations are complex. Missions are complex. Communications are complex. Through automating semantic resolution between organizations’ existing systems, we endeavor to simplify the discovery of feasible partnerships. This results in providing seamless communication between organizations, much like robust software module communication.

Learn More  Get Involved
Conclusions & Future Work

- (Automated) Learning Machines need:
  - Structure
  - Good teachers (data)!
- SMART Manufacturing needs:
  - Trustworthy partners
  - Translation assistance
  - OT rather than IT-based Cyber Standards
- We need:
  - Lunch!
  - Questions?
- Can these principles help solve your engineering, management, medical, civil, biological issues?
Acknowledgments

- ISA 99: Eric Cosman & Jim Gilsinn (Co-Chairs); Joe Weiss (Managing Director); Charley Robinson & Eliana Brazda (ISA Staff); committee & slides.
- Research advisor: Professor Shmuel Rotenstreich (b: Germany)
- Numerous global colleagues active development: Pictured
- The Catholic University of America: faculty, facilities &
- Continuing Research: Mr. Khalid Khawaji, Ms. Anh Thai, Mr. Khoi Nguyen, Ms. Cynthia Fioriti, Mr. Ibrahim Al Mubark, Mr. Abdullah Almalki, Mr. Mofaq Alotaibi, Mr. Luke Lepak, Mr. Dominic Abela, Mr. Abdulaziz Alhuthali and Mr. Andrew DeNooyer (CUA EECS Graduate & Undergraduate Students)
- Images: Numerous students, museums & libraries
Backups
Foundational Work: Prior Work, Limitations & Applications

- Task Organization & Automation: Workflow
- Local Knowledge: Organizations & Ontologies
- Process Sharing: Inter-Orgnizational Workflow
- Remote Computational Comprehension
Foundational Work: Workflow

- Major work: Office Automation 1960s & 1970s
- Key principle: decouple business process function & flow logic
- Dimensions (Aalst, 2004); Mining (Wang et al. & Aalst, 2013)
- Development niches (*many*, 200X+)
- Distilled Workflow Elements Model (Unertl et al., 2010)
Foundational Work: Workflow

Primary interfaces (Workflow Management Coalition, 1995):

- **Process Definition Tools**
- **Workflow Enactment Service (Server)**
- **Workflow Engine(s)**
- **Interface 1**: Business Process Definition
- **Interface 2**: Worklist Handler
- **Interface 3**: Tool Agent
- **Interface 4**: Interoperability
- **Interface 5**: Administration & Monitoring Tools

**TYPICAL WEB SERVICES**
Workflow Example (Aircraft Overhaul)

- YAWL (Aalst & Hofstede, 2002; Hofstede, et al., 2010; Adams, et al., 2012): Language & WFMS
  Demonstrates workflow pattern, mathematical & Petri Net bases compliance feasible (previously under contention)
Workflow Example (Component Overhaul)

Specification ID: AircraftOverhaul, Net ID: Overhaul_Component
Workflow Example (Wing Repair)

Specification ID: AircraftOverhaul, Net ID: Wing_Overhaul

Diagram:
- Receive Wing
- Clean and Inspect
- Repair
- Test
- Satisfactory on Receipt
- Replace
- Unsatisfactory Operation
- Satisfactory Operation
Foundational Work: Organizational Behavior

- Seminal organization process behavior research (McGrath, 1963)
- Organizational culture shapes & symbols signal deeper meaning
  - Gordon, 1999; Rafaeli & Worline, 2000; Horling & Lesser, 2005; Alvesson, 2011
- Common ground clarifies meaning; internal diversity helps ally new collaborators
  - Weber, 2000; Engeström, 2001; Carroll et al., 2008
- Team communication & cognition: social processes where contextual clarity matters
  - Perin, 1995; Bednar et al., 2007; Fiore et al., 2008; Narayanan et al., 2011
- Known by neighbors kept (Competitors, Customers, Suppliers)
  - Porac et al., 1989; Hodgkinson & Healey, 2011
Foundational Work: Knowledge & Ontologies

- Ontological Computer Science applications: Roles; ECA; DOLCE
  - Gruber, 1993; Smith, 1998; Wagner, 2003; Smith & Grenon, 2004; Bottazzi & Ferrario, 2005 & 2008
- Organizational Ontology: natural knowledge representation bridges internal workflow meaning gap
  - Hodgkinson & Johnson, 1994; Hepp & Roman, 2007
- Knowledge Capture: tacit/explicit; chunking & feature matching strategies
  - McManus et al., 2003; Haynes & Smith, 2008
- Semantic application: encoding methods
  - Miller, 1995; Desouza & Hensgen, 2002; Hirst, 2009
Simple Organizational Ontology Example

- Protégé Ontology design tool (Noy et al., 2000+)
“Simple” Ontology Alignment Example

- Challenge:
  - Variation between organizations
  - Merging *actual* ontologies

- Ontology 1

- Ontology 2

- Aligned Ontology
  - Negotiated

Hameed et.al, Ontology Reconciliation in *Handbook on Ontologies*, 2004
“Simple” Ontology Alignment Example

- Ontology 1
- Ontology 2
- Aligned Ontology and mappings (SUMO, Paliwal et.al, 2012)

Hameed et.al, Ontology Reconciliation in *Handbook on Ontologies*, 2004
Foundational Work: Inter-O rganizational Cooperation

- Teaming Behavior: Cooperation, Social economic interaction, Defined tools & signs, Distributed communication leaky by nature
  - Axelrod, 1984; Schelling, 1978; Carroll et al., 2008; Engeström, 2001; Rentsch et al., 2008
- Protocols: Sequential messaging fits Inter-organizational negotiations
  - Kraus, 2001; Bertino et al., 2004; Hirst, 2002; Aalst et al., 2000 & 2002; Bruno, 2005; Aldred et al., 2005-9; Kuhr et al., 2008
- IOWF:
  - Networked virtual enterprise interoperability unfilled promise
  - Emergent complex systems require robust communication between heterogeneous partners using natural strategies observed
  - Ebers, 1997; Bradley & Nolan, 1998; Johnson, 2001; Papazoglou et al., 2000; Sheth et al., 1997; Stegwee & Rukanova, 2003; Visser et al., 2003; Hofstede et al., 2010
Inter-O rganizational Workflow Primitives

Intra-Organizational Workflow Middleware

- Decoupling middleware 3 communication dims (Aldred, 2005-9)
- JCoupling tool for workflow messaging (Kuhr, 2008 & 2012)

YAWL Message Handling

JCoupling Bridge Architecture
Foundational Work: Computational Comprehension

- **Natural Language**: Hard problem, long automated translation history repairing *language shortcuts* reduces parties' conception gap
  - Weaver, 1949; Bar-Hillel, 1960; Wilks et.al, 1975, 89, 90; Dailey, 1986; Hirst, 2002; Dodig-Crnkovic, 2005

- **Word Sense Disambiguation (WSD)**
  - Approaches vary; facets: sense distinction *granularity*, external knowledge source; context representation; classification method
  - WordNet
    - Tool: manually cataloged words (synonyms, other relations: ontologies)
    - Verbs’ polysemy significantly greater than nouns (verbs, actions, tasks)
    - Semantic similarity distance measurements
  - Metrics
    - Coverage, Precision & Recall
    - Senseval/Semeval international competitions (tri-annual starting 1998)
  - Topic Model-based (next slide) avoids *knowledge acquisition bottleneck*
    - Navigli, 2009; Miller et.al, 1993; Resnik, 1999
Foundational Work: LDA Application to WSD

- Latent Dirichlet Allocation (LDA: Blei, Ng, Jordan, 2003)
- Finding Scientific Topics (Griffiths & Steyvers, 2004)
- LDA Generative Model:
  - For each topic $k$ of $K$: draw multinomial distribution $\beta_k$ from Dirichlet distribution with parameter $\lambda$
  - For each document $d$ of $D$: draw multinomial distribution $\theta_d$ from Dirichlet distribution with parameter $\alpha$
  - For each word position $n$ of $N$ (in $d$ of $D$): select hidden topic $Z_{d,n}$ from multinomial distribution with parameter $\theta_d$
  - Choose observed word $W_{d,n}$ from distribution $\beta_{Z_{d,n}}$ for that topic
  - Used to mine business contracts for topics (Gao & Singh, 2014)
  - LDAWN compares associated found topics words’ senses, mapping traversed WordNet synsets hypernym paths $\lambda$ speeding convergence
Incorporating Structured Domain Knowledge: Simplified WordNet Wing Hypernymy Example

- **entity**
  - **abstract entity**
    - **physical entity**
      - **device**
        - **{airfoil, control surface}**
        - 37 other kinds
      - **{whole, unit}**
    - **{part, piece}**
      - **body part**
        - **organ**
          - **{part, piece}**
            - **{whole, unit}**
      - **{part, piece}**
        - **{whole, unit}**
          - **{airfoil, control surface}**

Example:
- foot
- wing$^1$
- gland
- elevator
- wing$^2$
- rudder

Hypernym Example:
- entity
- physical entity
- device
- {airfoil, control surface}
- 37 other kinds
IOWF Semantic Mediator

**Concept:**
- Automated workflow systems (like humans) to make *sense* of received information must understand terminology in *context*.
- *Not* of listener, but of speaker.
- Polysemous language overloads individual words’ meaning.
- Implied unique meaning(s) within organization *confound* outsiders.
  - *Port:* Computer Science Department vs. Nautical usage.

**Mediator design pattern solution employs:**
- High level bridge encapsulating myriad workflow systems’ requisites for choreography.
- Low level unstructured computational semantic resolution mechanism providing context.
IOWF Semantic Mediator to Bridge Organizations’ Workflows

Architectural Perspective

Interaction Perspective
Simplified IOWF Example: Outsource & Insourse Wing Repair

Consuming Organization (OS)

Servicing Organization (IS)
Creation of IOWF Semantic Resolution ‘Documents’

Interlocking coordination of Work across varying internal vocabularies of meanings

IOWFSR correlations validated by experience
Sky High Aircraft Maintenance Svc

Sky High Aircraft Ontology

Wing overhaul
OS Workflow

Airframe Best Components

ABC Ontology

Wing repair
IS Workflow

Sequential combination of component organizations’ ontologies & workflows

Interlocking coordination of Work across varying internal vocabularies of meanings

IOWFSR Document

IOWFSR correlations validated by experience
IOWF Semantic Resolution
Document Creation Example

Sky High Aircraft Maintenance Svc

aircraft repair
modernization
{ +
wing overhaul
transport
}

Airframe Best Components
repair supply
aircraft wings
{ +
repair wing
transport
}

IOWFSR Document Vocabulary
aircraft (2) repair
(3) modernization
wing (3) overhaul
(1) transport (2)
supply (1)

Interlocking coordination of Work across varying internal vocabularies of meanings

IOWFSR correlations validated by experience
IOWF Semantic Resolution Corpus Collection and Topic Discovery

IOWFSR Documents

LDAWN analysis: discover document topic mixture correlations

Topic 0
bird
wing
species
veterinarian

Topic 2
aircraft
wing
model
overhaul
Resource Advertisement & Matching: Servicing Organization & Mediator Interaction

Servicing Organization

- Requirement handling similar (Consuming Organization)
- Collection & processing of executed IOWFs similar
Prospective IOWF Topic Distribution

- **bird + plane wing repair**
- **aircraft wing^2 repair**
- **aircraft wing replacement + repair**
- **aircraft wing OS-only**
Magnitude of IOWF Topic Distribution Difference: Prospective – Requesting

- **bird + plane wing repair**
- **aircraft wing^2 repair**
- **aircraft wing replacement + repair**
- **aircraft wing OS-only**
IOWF Semantic Resolution
Generative Model & Inference

Probabilistic Generative Process

Training

Topic 0
- bird
- wing
- species
- veterinarian

Topic 2
- aircraft
- wing
- model
- overhaul

IOWFSR 60003:
- bird\(^0\)
- species\(^0\)
- repair\(^1\)
- wing\(^0\)
- veterinarian\(^0\)

IOWFSR 52200:
- aircraft\(^2\)
- wing\(^2\)
- model\(^2\)
- repair\(^1\)
- overhaul\(^2\)

Statistical Inference

Topic 0
- bird
- wing
- species
- veterinarian

Topic 2
- aircraft
- wing
- model
- overhaul

IOWFSR 40640p10:
- aircraft\(^2\)
- bird\(^0\)
- species\(^0\)
- repair\(^1\)
- wing\(^0\)
- wing\(^2\)
- veterinarian\(^0\)

IOWFSR 40640p1:
- aircraft\(^2\)
- wing\(^2\)
- repair\(^1\)
- overhaul\(^2\)

Internally Ambiguous

Internally Consistent

IOWFSR Corpus
Documents

IOWFSR Prospective
Pairings
Conclusions

- Novel framework; exemplary model demonstrates process to automatically resolve semantic ambiguity between organizations’ WFMS without *a priori* knowledge
  - Key to decoupling organizations & resources
  - Easing interoperability means sharing more work to accomplish tasks
  - Flexibility & compatibility of approach supports adaption/adoption
- Enhancing semantic clarity exposes more potential inter-organizational workflow resource & consuming alternatives
  - Publishing alternatives publicly lowers service costs & increases markets
  - R&D partnering assistance
  - Prototype, One-off matching minimizes costs & maximizes markets
- Transfers semantic resolution
  - *From* time of execution
  - *To* planning stage (evaluation of alternatives)
Future Work

- Generalize, automate, test & apply framework to increasingly complex cases
- Formalize IOWFSR Model
- Apply to unlabeled message source identification
- Apply to outstanding Complex System communications issues (How can we apply to solve your engineering / management / medical / biological issue?)
- Dynamic IOWFSR Mechanism to allow modeling interaction of heterogeneous complex system components independently
- Extension of IOWFSR methods beyond natural language to new heterogeneous domains of encoded structural & dynamic knowledge
- How do nano-bot swarms communicate to accomplish a mission?
- Questions? Thanks!
Workflow Management System Example (YAWL)
Organizational Language Barriers to Communication

- Ontology (Workflow components)
  - Language of (given) Organization
  - Subset of Domain, Natural Language

- Languages evolve from particular community seeds
  - Terminals start from instances (proper names: Tailor)

Obs: Task names may vary, even though predicates match
- Generalization rules vary by environment, create unique grammars
- True for clans, societies, organizations
- Grammar commonalities, differences emerge as distinct groups interact

- Industrial orders: “regional ontologies which attempt to define what life, labor, and language are in their own beings” (Foucault, The Order of Things, 1973)

- “Languages are the measure of mankind’s ideas” (Turgot, 1750) in Language as the Key to the Epistemological Labyrinth (Lifschitz, 2004)

- Generated bottom-up; interpreted top-down
  - Direct mappings between ontologies of organizations: intractable computing jungle!
Why Natural Language Workflows?

- **Reasonability**
  - WFMS primarily model organizations of humans automating their work processes
  - Workflow tasks primarily defined using NL

- **Importance**
  - Machine readable dictionaries (MRD) help resolve WF corpus elements (semantic primitives) meaning
  - Linguistic Theory and NL development *slightly* predate advent of Computer Science

- **Extensibility**
  - Handle multiple domains, built from given NL
  - Concept extensible to variety of languages, providing dictionary available (or may be generated)
WSD (Topic Model & MRD) vs. Meta-Ontology Mapping

- Simpler – Façade!
  - “Bag of Terms” analysis internally models latent structure vs.
  - Tagging POS & generating structured syntax meta-trees
- “Automatic” maintenance
  - As terms added, LDA topics migrate slowly over time, maturation
  - Meta-Ontology restructuring might result in incremental versions not resembling each other: revolutionary vs. evolutionary
- Research areas relative success in resolving problem
- Inclusiveness & predictability of semantic resolution
- WordNet empirically defined syn-sets enhance Topic Model’s demonstrated natural clustering affinity
Distinguished from other IO work

- On-going organizational cooperation
  - Shared language evolution may already exist
- Web Services
  - Stateless; complete interface description available
- While many WS applications, limited IOWF investigation; except
  - Meta-model ontology IOWF interoperability approach outlined (Haller, et al. 2005)
  - Semantic ambiguity resolution expanded; actual solution alternatives left to future work (Höfferer, 2007)
- Community-shared standard ontologies IOWF exist
  - Not many; but nice (ex: Medical, ICD 9&10)
- Natural Language vs. Scientific or Grid Workflow
  - NL introduces more ambiguity; broadens application
- Agent-based WFMS
  - Principles don’t require agents
- Custom-developed mapping
  - Hard-coded translations (look-up tables) presume prior knowledge of parties’ languages involved; extensibility intractable
Organizational Diversity

- Maturity
  - *Business Process* definition degree of detail
  - *Continuous Improvement*
- Breadth (specialization) of entity/component
- Domain (community) participation
  - Number & degree of standardization
- Internal topology (hierarchical, P2P, matrix)
- Size matters (flexibility/rigidity of roles & tasks)
- Unique internal knowledge: own Ontology
Standards & implementations vary for WFMS:

- Features, information requirements & outputs (heterogeneity excludes Adaptors)
- Detail (generality): derived from organization
- Environment (hardware/software: supported & required)
- Degree of Automation
- Workflow Interface Support (WFMC, later)
- Representation/extent of organization knowledge, structure
# Lexical Relations

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>synset</td>
<td>word senses expressing (approximately) same meaning</td>
</tr>
<tr>
<td>gloss</td>
<td>textual definition possibly with examples</td>
</tr>
<tr>
<td>antonymy</td>
<td>expresses opposite concept</td>
</tr>
<tr>
<td>pertainymy</td>
<td>adjective of or pertaining to noun (or another adjective)</td>
</tr>
<tr>
<td>nominalization</td>
<td>noun nominalizes verb (flight, fly)</td>
</tr>
<tr>
<td>hypernymy</td>
<td>kind-of or is-a (superclass: wing, airfoil)</td>
</tr>
<tr>
<td>hyponymy</td>
<td>is-a-kind-of (subclass: airplane, jet)</td>
</tr>
<tr>
<td>troponymy</td>
<td>is-a-kind-of, verb (fly, soar)</td>
</tr>
<tr>
<td>meronymy</td>
<td>part-of (airplane, wing)</td>
</tr>
<tr>
<td>holonymy</td>
<td>is-a-part-of (wing, bird)</td>
</tr>
<tr>
<td>entailment</td>
<td>verb is entailed by verb (overhaul, repair)</td>
</tr>
<tr>
<td>similarity</td>
<td>adjective is similar to adjective (satisfactory, acceptable)</td>
</tr>
<tr>
<td>attribute</td>
<td>noun an attribute that adjective expresses value (color, purple)</td>
</tr>
</tbody>
</table>
Co-occurrence Matrix, Topics, Documents & Words

\[
\begin{align*}
\text{documents} & \quad \text{topics} \\
\text{words} & \quad \text{words} \\
\text{normalized co-occurrence matrix} & \quad \Phi \\
\end{align*}
\]

\[
\begin{align*}
\text{topics} & \quad \Theta \\
\text{mixture components} & \quad \text{mixture weights}
\end{align*}
\]
Dirichlet Allocation Example

- Dirichlet: Joint distribution random variables over partitions
- Chinese Restaurant Process (CRP) provides example of clustering
  - \( N \) customers
  - \( K \) tables
  - Initial *Generation* Step: Customer \( w_1 \) sits at Table \( \beta_1 \)
  - Successive steps (\( n = 2..N \)), Customer \( w_n \) sits at Table \( \beta_k \):
    - Occupied Table \( \beta_k \): probability of \( \frac{|\beta_k|}{\alpha + w_n - 1} \)
    - Unoccupied Table \( \beta_{k-max+1} \): probability of \( \frac{\alpha}{\alpha + w_n - 1} \)

Representations:
- Tables are topics
- Customers are words
- Restaurants are documents
- Effect of varying *concentration*, \( \alpha \): \( 10^2 \)…unity…\( 10^{(-2)} \)
Gibbs sampling: Markov chain Monte Carlo posterior inference approximation

- Even fixing $K$ topics, computation of actual posterior intractable
- MCMC biased random walk for 2-dimension r.v. below
- Gibbs sampling (one type of MCMC) explores $K$-dimension distribution space of hidden variables
  - Iterates between
    - Each hidden variables’ conditional distribution (given observations)
    - Current state of other hidden variables
  - $K$ limits topic choice, solution dimension space
- $\alpha$ iteration size; when small (0.001 to 0.01)
  - Restricts # topics per document: sparse
  - Increases convergence time
  - Conversely, large $\alpha$ over-generalizes
- Key insight: use exchangeability
- CRP: Where does customer $w_n$ sit?

Conditions (Rules)

Hierarchy

Disjoint