The Insider Threat to Critical Infrastructure

a social dimension to risk management

Michael Goldsmith
University of Oxford

Sadie Creese, Nick Moffat, Jason Nurse, Jassim Happa, Ioannis Agrafiotis, Phil Legg, Oliver Buckley
Min Chen, Eamonn Maguire, Simon Walton (OeRC)
David Upton, Alix Ellis (Saïd Business School)
Monica Whitty, Gordon Wright, Xin Miao (University of Leicester)
Michael Levi (Cardiff University)

PSCE Forum, Paris
25 November 2014

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Insider Threat

- What do we mean when we talk of ***insider threat***?
  - used either of individual or the danger they pose
- Typically thought of as a rogue employee…
- … but quite possibly not:
  - IT contractors
  - non-IT staff and contractors
    - installing KVM hardware, exploiting unlocked terminals
  - supply-chain partners
  - outsourced data-centre personnel
    - insider as do not need to breach perimeter protections
  - cloud service providers
  - malware?

INSIDERTHREAT

Opportunity and Motive

- An abuse of privileged access:
  - system login credentials, physical access, web-service access…
- A variety of outcomes:
  - destruction / sabotage (e.g. information, physical)
    - potentially disastrous within critical infrastructure
  - theft (e.g. information, financial, physical, fraud)
  - theft for distribution (e.g. IP)
  - dissemination of sensitive information (whistleblowing, *mistake*)
- A variety of motives:
  - financial gain
  - revenge / dissatisfaction with company or management
  - desire for respect (from co-workers / external peer group / self)
  - persuasion / coercion (by family/friends or blackmail/threats)
  - often more than one factor
  - *or indeed none of the above!*

Means

- Abuse of legitimate privileges
  - particularly to breach confidentiality or integrity
- Internal use of exploits to gain unauthorised privileges or to bring down systems
  - facilitated by legitimate access?
- “Stolen” credentials
  - e.g., by shoulder-surfing, unlocked terminals
- Indirect attacks on the system
  - social engineering, blackmail, hardware key-logger, …
- Inadvertent carelessness / recklessness
  - malware-link clicking, succumbing to phishing, BYOI, …

Examples

- Autumn 2012 – US power plant taken offline for 3 weeks by infection inadvertently introduced to turbine-control system via tainted USB stick used by external contractor
  - allegedly son picked up drive-by malware from a dodgy gaming site
- Security guard – with Asperger's syndrome – created high-fidelity model of the building he was responsible for within *Second Life*
  - security of building consequently compromised
- Cloud disaster-recovery company – customer backups corrupted by disgruntled employee
  - only discovered when first customer emergency occurred
  - disastrous for both companies

Challenge

Unlike a normal attack, an insider is entitled to act within the organisation —

- … indeed typically must do so in order to fulfil their job role

How can we assess when "entitled" behaviour becomes – or is likely to become – malicious behaviour?

Corporate Insider Threat Detection Project

- Sponsored by the UK National Cybersecurity Programme
  - with the Centre for the Protection of National Infrastructure (CPNI)
  - ~€2.1M over 30 months
- Collaboration between University of Oxford, University of Leicester and Cardiff University
  - psychology and behavioural analysis led by Leicester
  - criminological analysis led by Cardiff
  - cybersecurity team in Computer Science focus on detection system
  - Oxford e-Research Centre focus on visual-analytics development
  - Saïd Business School focus on education and awareness, and on business-change issues

Literature Review and Interim Survey Results

- Climate and perception of risk:
  - insider attacks are rising; consequences are potentially more significant; phenomenon is widely underreported
  - initial web-based survey finding is that insider-threat detection is not seen as commensurately important nor as part of corporate culture
    - much larger survey (in collaboration with IBM) just concluding…
  - average time to detect internal computer malfeasance in financial services is **33 months**
- Insider-detection practice:
  - most detections of insider attacks rely on people
  - lack of perceived risk inhibits the implementation of good practice
- Management levels of concern:
  - poor level of awareness on the topic – too many "don't knows" in survey
  - increased monitoring of staff may be an issue for managers
  - **36%** of respondents to one published survey do not evaluate their partners' security policies at all

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Conceptual Model

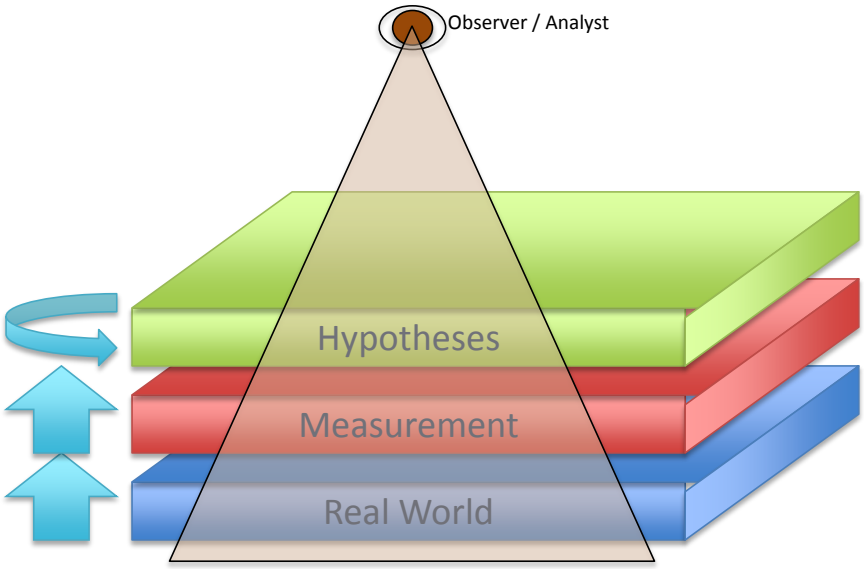
- Identifies the problem space, and the related elements that exist within this space
- Insider Threat is not only a cyber issue
  - therefore, we need to understand the full scope of the problem
- A conceptual model can help to inform which aspects should be considered when implementing a detection system
- Bottom-up reasoning:
  - the data is used to identify suspicious behaviour that alerts the analyst to draw a particular hypothesis
    - machine-learning and data-mining concepts
    - anomaly detection
- Top-down reasoning:
  - the analyst forms their own hypothesis which they want to verify
    - visual analytics and visualisation concepts
    - data exploration

INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Conceptual Model



Hypotheses made regarding the observed potential insider threat

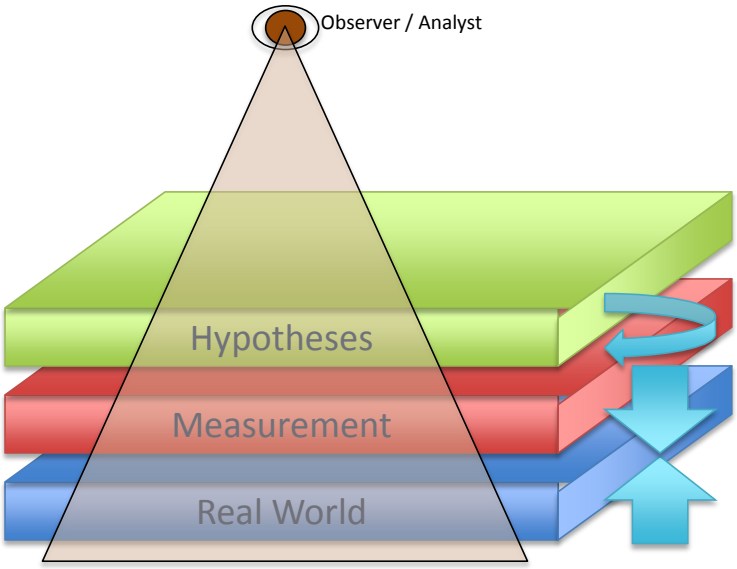
What can one infer about their intent based upon the measured data?

INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Conceptual Model



What if we have an initial hypothesis about an insider's behaviour?

INSIDERTHREAT

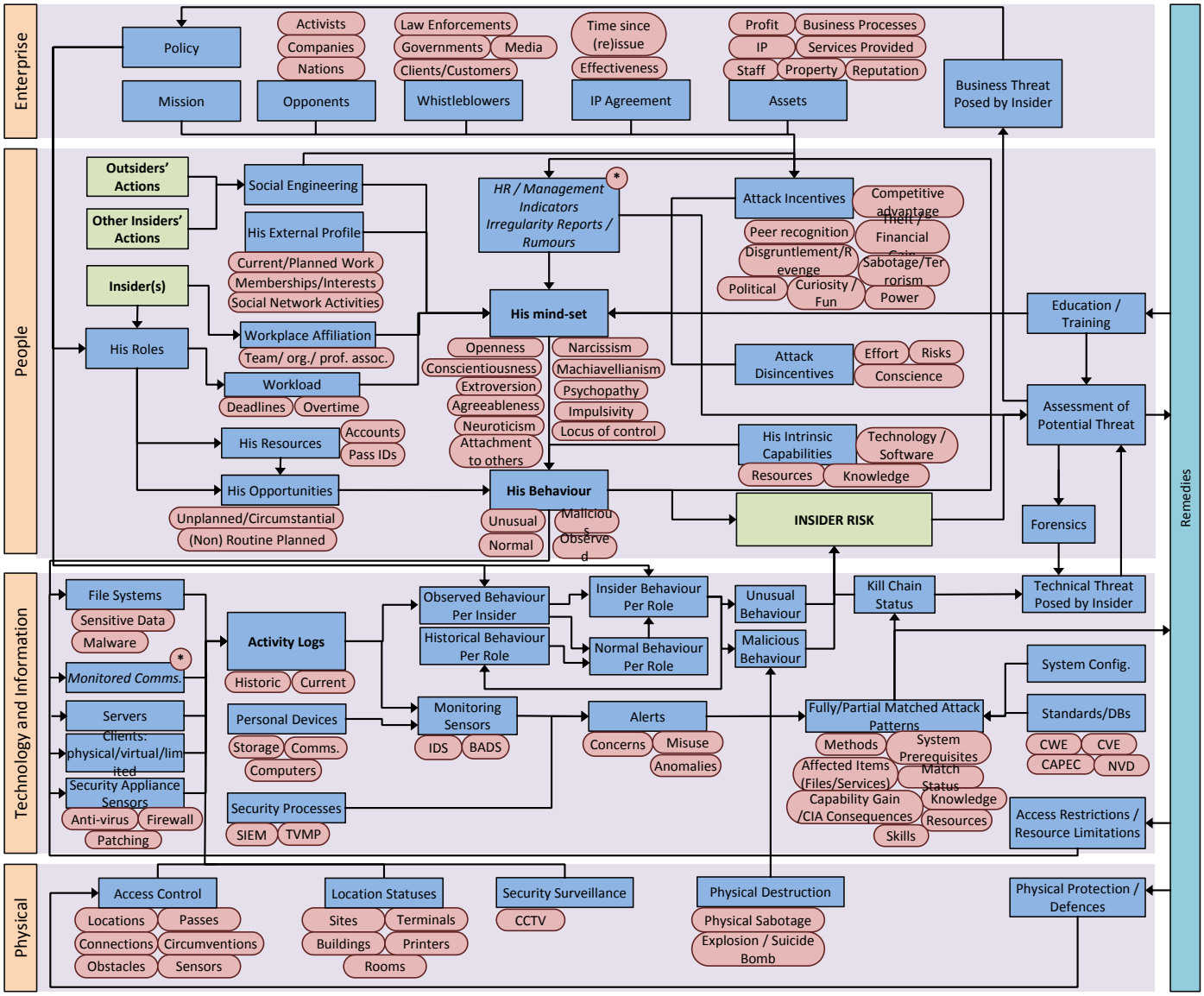
UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Elements of the Model

- At the core of the conceptual model are the elements that exist within the problem space of insider-threat.
- All elements would be present within the real world level of the conceptual model.
- The elements would all be measureable (to some extent) to propagate upwards through the model.

INSIDERTHREAT


Enterprise
People
Technology and Information
Physical
Remedies
Policy
Mission
Activists
Companies
Nations
Opponents
Law Enforcements
Governments
Media
Clients/Customers
Whistleblowers
Time since (re)issue
Effectiveness
IP Agreement
Profit
Business Processes
IP
Services Provided
Staff
Property
Reputation
Assets
Business Threat Posed by Insider
Outsiders' Actions
Other Insiders' Actions
Insider(s)
His Roles
Social Engineering
His External Profile
Current/Planned Work
Memberships/Interests
Social Network Activities
Workplace Affiliation
Team/ org./ prof. assoc.
Workload
Deadlines
Overtime
His Resources
Accounts
Pass IDs
His Opportunities
Unplanned/Circumstantial
(Non) Routine Planned
HR / Management Indicators Irregularity Reports / Rumours
*
His mind-set
Openness
Conscientiousness
Extroversion
Agreeableness
Neuroticism
Attachment to others
Narcissism
Machiavellianism
Psychopathy
Impulsivity
Locus of control
His Behaviour
Unusual
Malicious
Normal
Observed
Attack Incentives
Competitive advantage
Peer recognition
Theft / Financial Gain
Disgruntlement/Revenge
Sabotage/Terrorism
Political
Curiosity / Fun
Power
Attack Disincentives
Effort
Risks
Conscience
His Intrinsic Capabilities
Technology / Software
Resources
Knowledge
INSIDER RISK
Education / Training
Assessment of Potential Threat
Forensics
File Systems
Sensitive Data
Malware
Monitored Comms.
Servers
Clients: physical/virtual/limited
Security Appliance Sensors
Anti-virus
Firewall
Patching
Activity Logs
Historic
Current
Personal Devices
Storage
Comms.
Computers
Security Processes
SIEM
TVMP
Observed Behaviour Per Insider
Historical Behaviour Per Role
Insider Behaviour Per Role
Normal Behaviour Per Role
Monitoring Sensors
IDS
BADS
Alerts
Concerns
Misuse
Anomalies
Unusual Behaviour
Malicious Behaviour
Kill Chain Status
Technical Threat Posed by Insider
System Config.
Standards/DBs
CWE
CVE
CAPEC
NVD
Fully/Partial Matched Attack Patterns
Methods
System Prerequisites
Affected Items (Files/Services)
Match Status
Capability Gain /CIA Consequences
Knowledge
Resources
Skills
Access Restrictions / Resource Limitations
Access Control
Locations
Passes
Connections
Circumventions
Obstacles
Sensors
Location Statuses
Sites
Terminals
Buildings
Printers
Rooms
Security Surveillance
CCTV
Physical Destruction
Physical Sabotage
Explosion / Suicide Bomb
Physical Protection / Defences

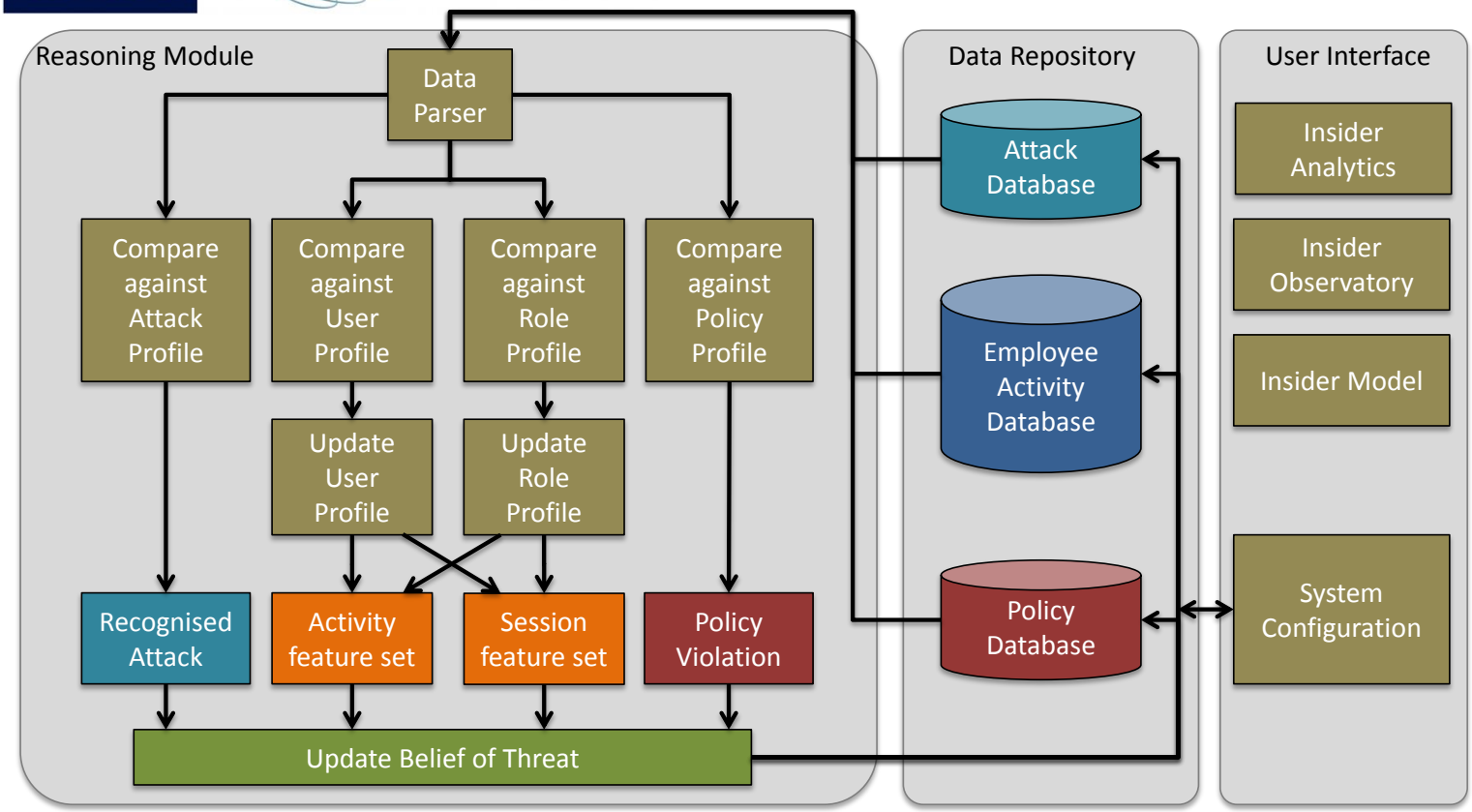
Construction of the Detection Prototype

- IDS-inspired architecture:
  - sensors/monitors, databases, data-mining and attack correlation, visual analytics
- Alerts for both anomaly detection and misuse:
  - machine-learning algorithms to understand normal behaviour
  - data-mining to recognise events (simple or compound) in big data
- Connection between detection algorithms and visual-analytics interface to support semi-supervised learning
- Exploration of performance for subsets of data, attack sensor sources and system configurations
- Validation via experimentation, initially on synthetic data, now real pilot deployment

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Current Architecture



INSIDERTHREAT

Our Approach

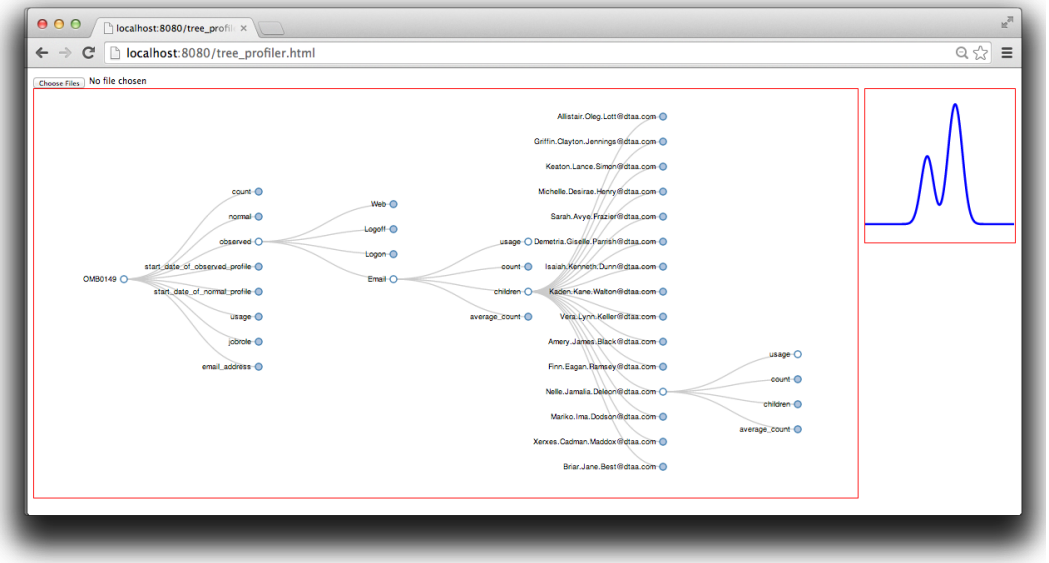
- A probabilistic, generative model of user behaviour:
  - record activities that the user performs
  - attributes associated with these activities
  - time of day/week activities are performed
  - how frequently these activities are performed
- Unsupervised / semi-supervised
  - we do not assume in advance what defines anomalous or threatening behaviour …
  - … but analyst may confirm or reject alerts, updating weights given to future observations
- Online
  - the system learns the user profile in real-time as new data is observed

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Statistical Profiling

- Statistical profiling of employee behaviour.
  - normal vs current
  - individual, role, organisation
- Measure deviation from typical normal usage.
  - unusual logins
  - increased e-mails or web browsing
  - new contacts
  - access of new files on server
  - …



Employee monitoring that does not show deviating behaviour

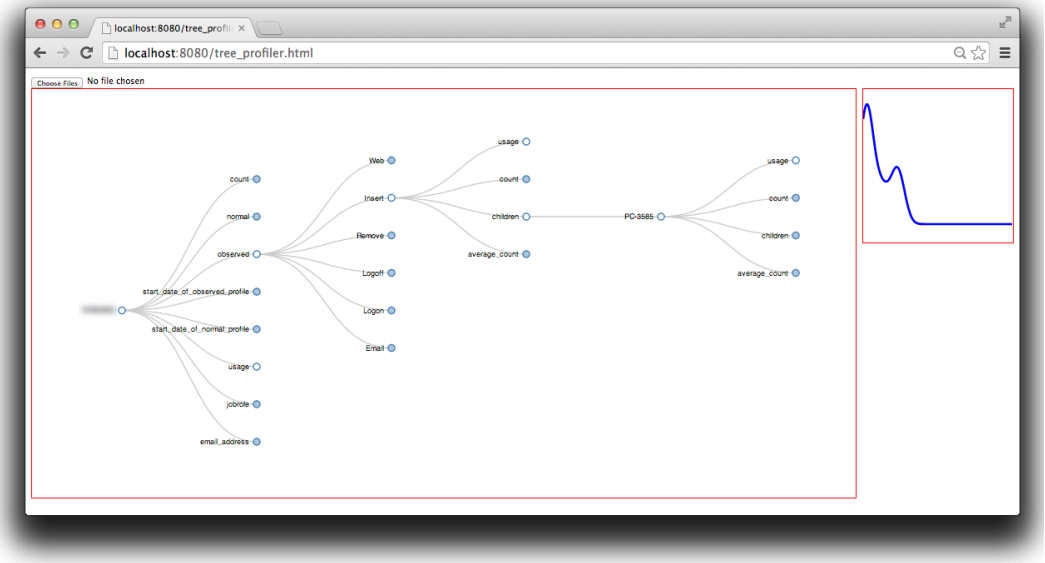
INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Statistical Profiling

- Statistical profiling of employee behaviour
  - normal vs current
  - individual, role, organisation
- Measure deviation from typical normal usage
  - unusual logins
  - increased e-mails or web browsing
  - new contacts
  - access of new files on server
  - …



Employee monitoring that shows suspicious device usage

INSIDERTHREAT

Digging Deeper into Data

- Some activities will also carry *content* that should be incorporated into an employee profile
  - *e-mail message, web site content, file content*
- Whilst not essential for the system, this information could provide greater context to an employee's mind-set
  - what do web browsing habits suggest about an employee?
  - if a file has been modified, what *exactly* has been modified?
  - what does the sentiment of their e-mails suggest about an employee?
- Opens up issues surrounding employee privacy – organisation must decide on level of monitoring desirable
  - privacy-friendlier e-mail monitoring using LIWC profiling?

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Profile Metrics

Technical metrics:

#logins
login duration
#unique_logins
earliest_login
latest_login

#usb_insertions
#unique_usb_insertions
#usb_upload_MB
#usb_download_MB

#emails_received
#unique_senders
#new_senders

#emails_sent
#unique_recipients
#new_recipients
earliest_email_sent
latest_email_sent

#files_created
#files_accessed
#unique_files_accessed
#new_files_accessed
#files_modified
#unique_files_modified
#new_files_modified
#files_deleted

email_bag_of_words
files_bag_of_words
website_bag_of_words
email_sentiment

keyboard_biometrics
mouse_biometrics

#websites_visited
#unique_websites
#new_websites
browsing_duration

cpu_usage
memory_usage
network_upload
network_download
processes_running

Physical metrics:

#swipe_card_entries
earliest_swipe_entry
latest_swipe_entry
#keypad_entries
#keyfob_entries

CCTV monitoring

Workstation location
IP address

Activity metrics:

user_new_activity
user_new_attribute
user_time_activity
user_time_attribute
user_count_activity
user_count_attribute
role_new_activity
role_new_attribute
role_time_activity
role_time_attribute
role_count_activity
role_count_attribute

Behavioural metrics:

Openness
Conscientiousness
Extroversion
Agreeableness
Neuroticism

Narcissism
Machiavellianism
Psychopathy

Workplace Affliation
Locus of Control
Attachment to others
Impulsivity

Disgruntlement
Not accepting feedback
Anger management issues
Disengagement
Disregard for Authority
Performance
Stress
Confrontational
Personal Issues
Self-Centeredness
Lack of Dependability
Absenteeism
(Greitzer et al. 2012)

...more to be established with Leicester

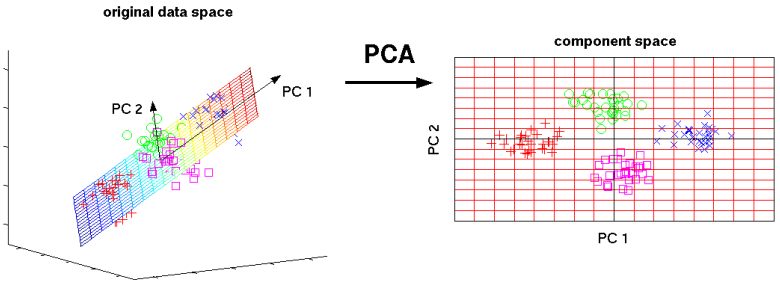
- Two ways to slice data:
  - daily metrics
  - activity-based metrics

INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Anomaly Detection



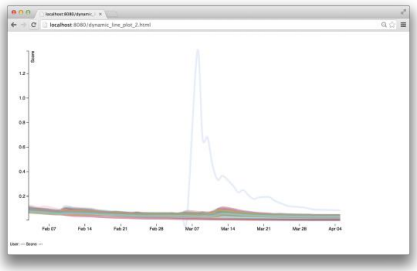
- Principal Component Analysis
  - reduces *n*-D features to ***< n*** components based on variance
  - a user with a suddenly large variance could indicate an anomaly
- Requires a consistent *n*-D feature set for comparison
  - e.g., login count, USB count, e-mail count, file count
  - can include time-based features (e.g., mean, earliest, latest…)
  - can also include ‘new’ accesses from user profile
  - equally suitable for daily or session-based profiling

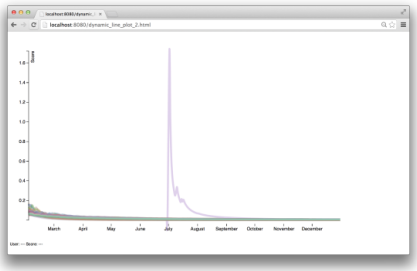
INSIDERTHREAT

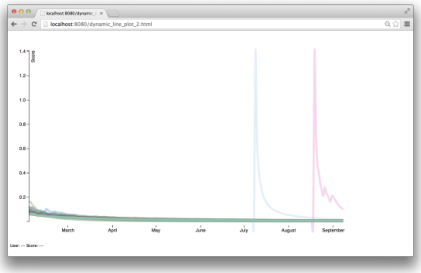
UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Anomaly Detection







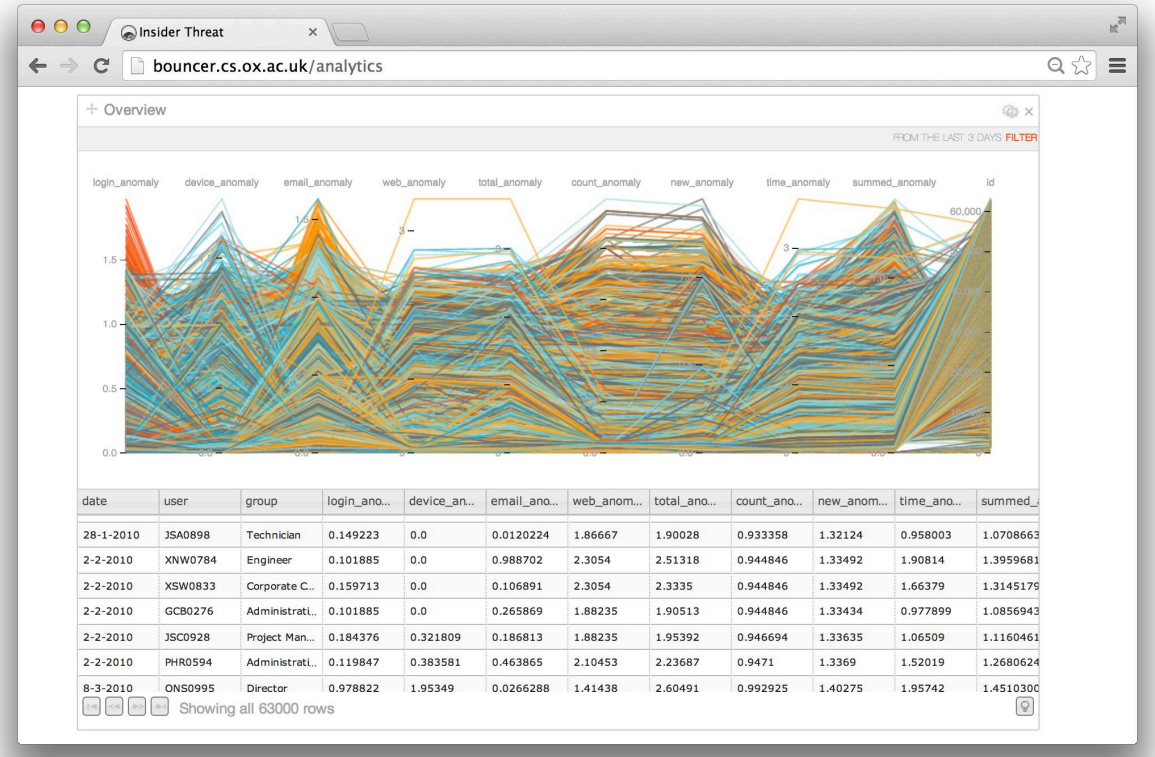
- Measurements are gathered from the employee profile data
- Suspicious behaviour is likely to provoke an anomaly on one or more measurement
- These provide a means to raise alerts about the potential threat posed by a particular individual

INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Analysis of Detection Results

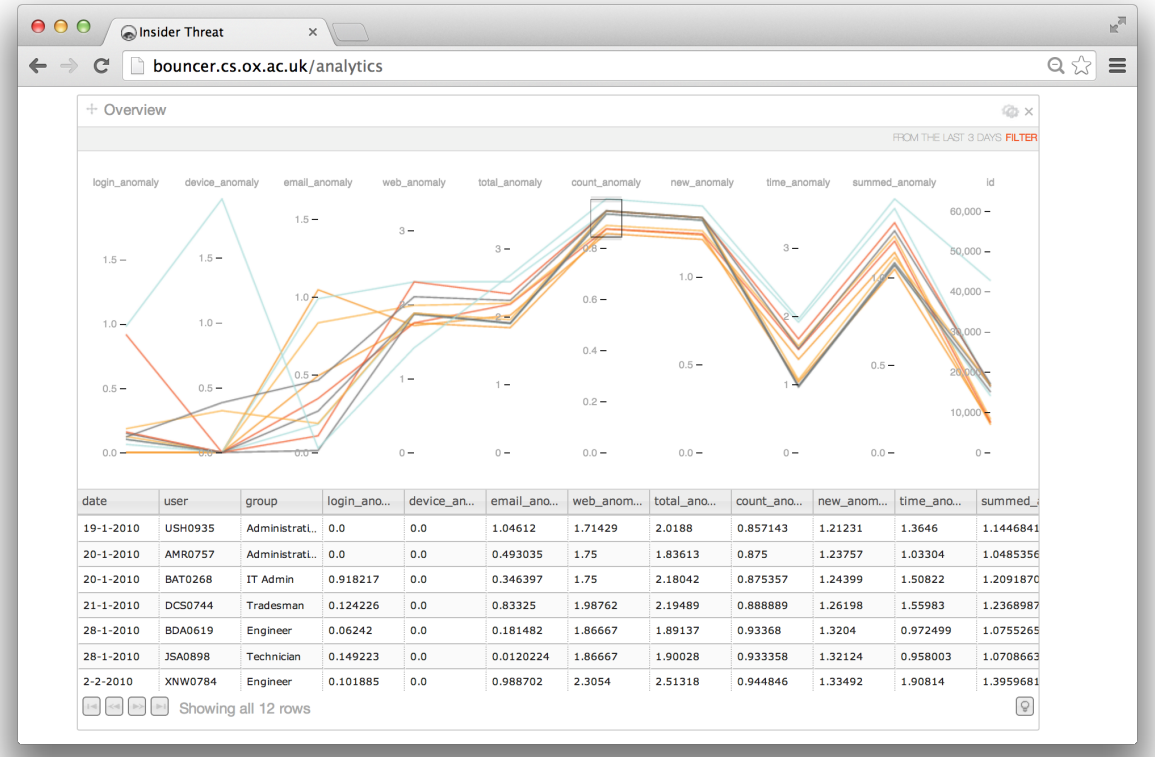


INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Analysis of Detection Results



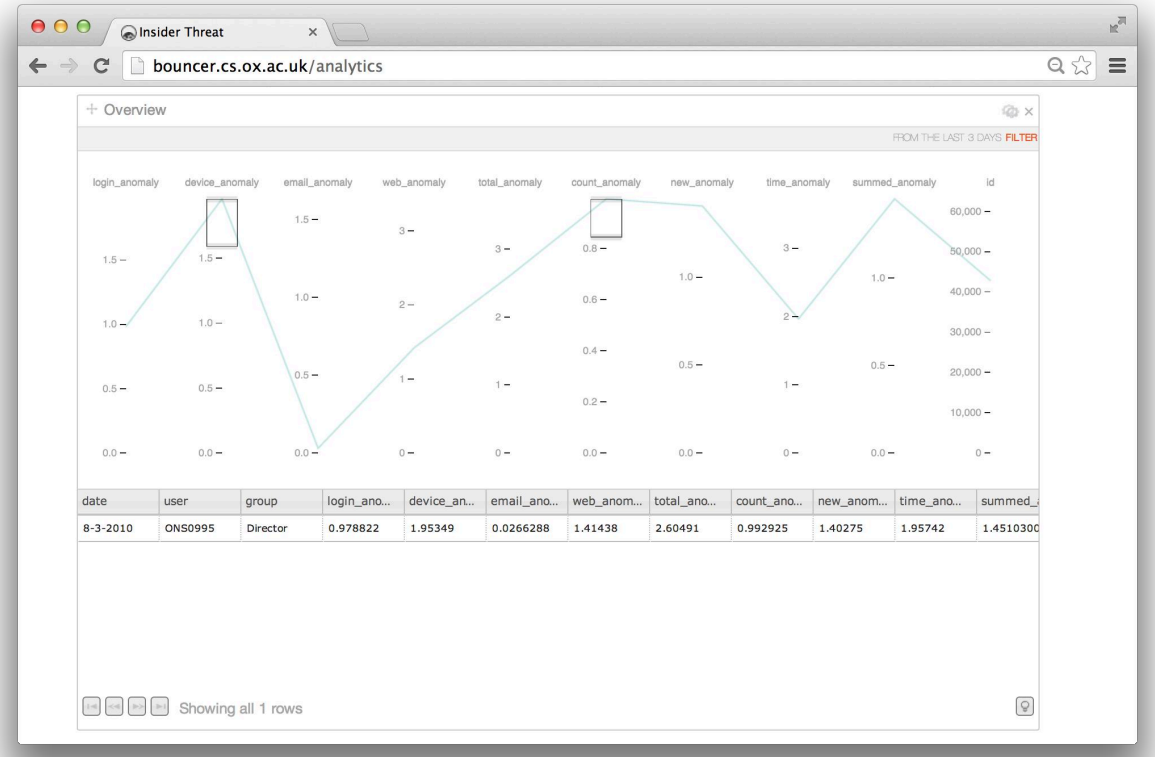
date	user	group	login_ano...	device_an...	email_ano...	web_anom...	total_ano...	count_ano...	new_anom...	time_ano...	summed_
19-1-2010	USH0935	Administrati..	0.0	0.0	1.04612	1.71429	2.0188	0.857143	1.21231	1.3646	1.1446841
20-1-2010	AMR0757	Administrati..	0.0	0.0	0.493035	1.75	1.83613	0.875	1.23757	1.03304	1.0485356
20-1-2010	BAT0268	IT Admin	0.918217	0.0	0.346397	1.75	2.18042	0.875357	1.24399	1.50822	1.2091870
21-1-2010	DCS0744	Tradesman	0.124226	0.0	0.83325	1.98762	2.19489	0.888889	1.26198	1.55983	1.2368987
28-1-2010	BDA0619	Engineer	0.06242	0.0	0.181482	1.86667	1.89137	0.93368	1.3204	0.972499	1.0755265
28-1-2010	JSA0898	Technician	0.149223	0.0	0.0120224	1.86667	1.90028	0.933358	1.32124	0.958003	1.0708663
2-2-2010	XNW0784	Engineer	0.101885	0.0	0.988702	2.3054	2.51318	0.944846	1.33492	1.90814	1.3959681

INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Analysis of Detection Results

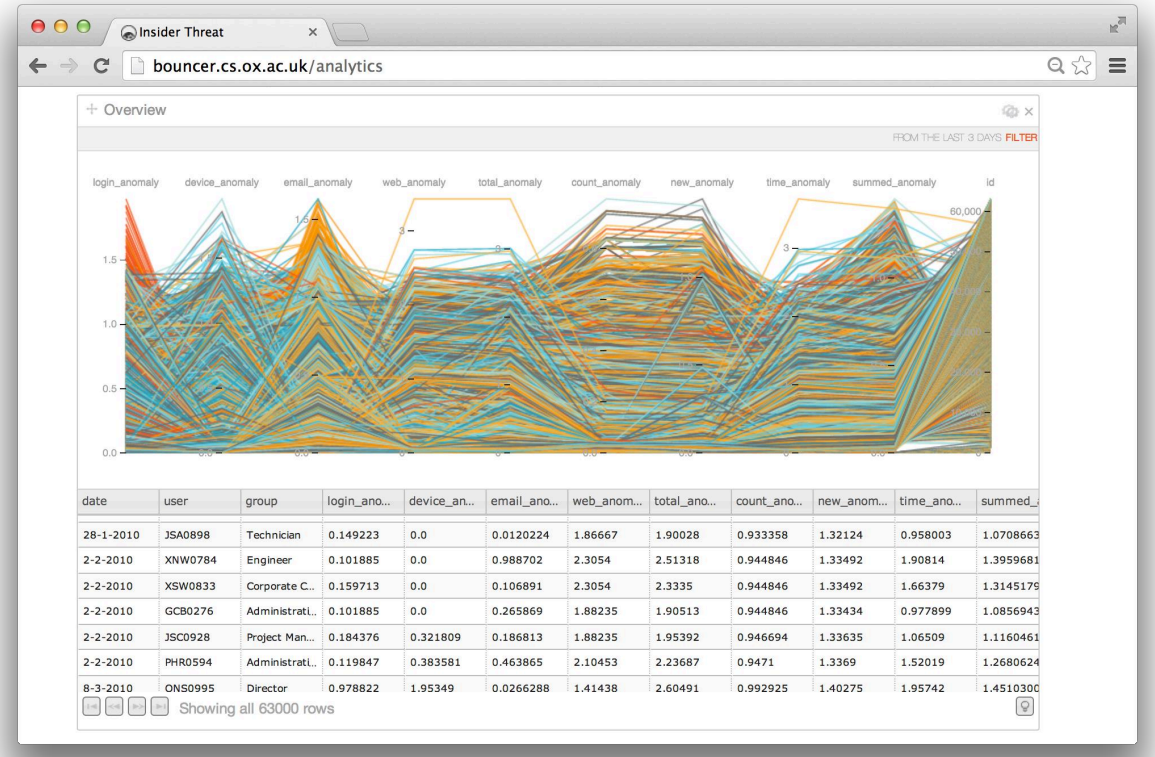


INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Analysis of Detection Results

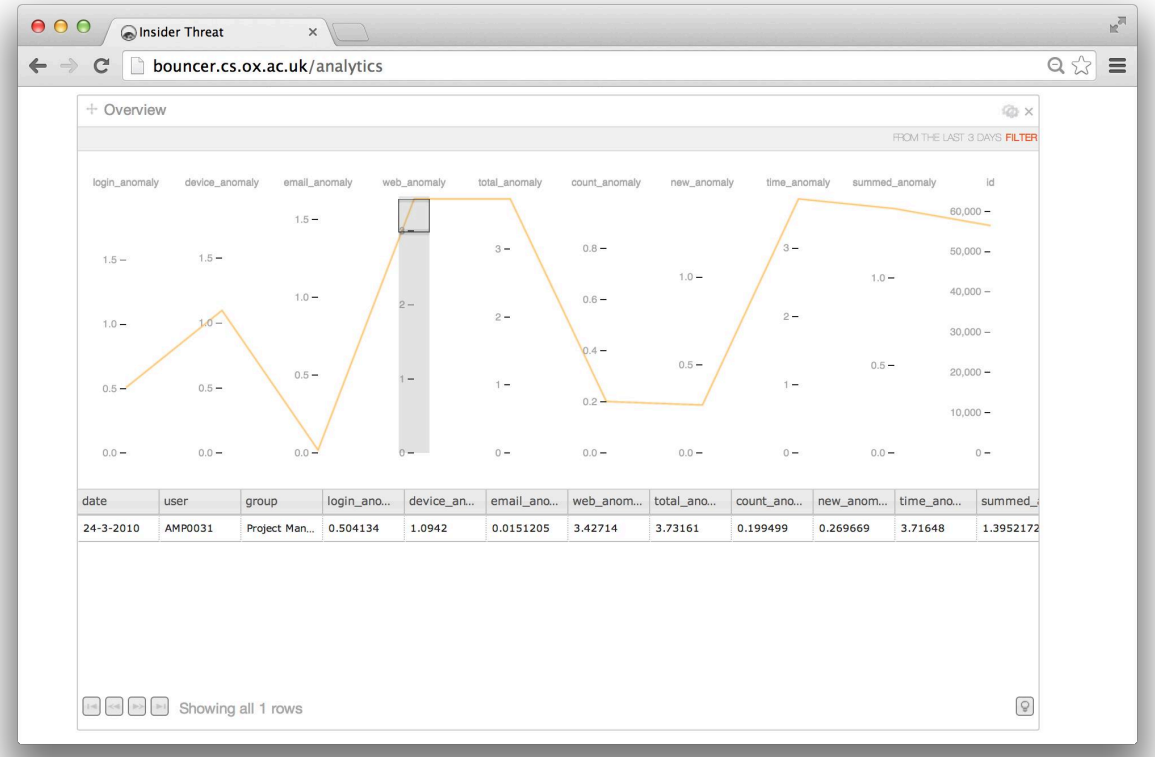


INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Analysis of Detection Results

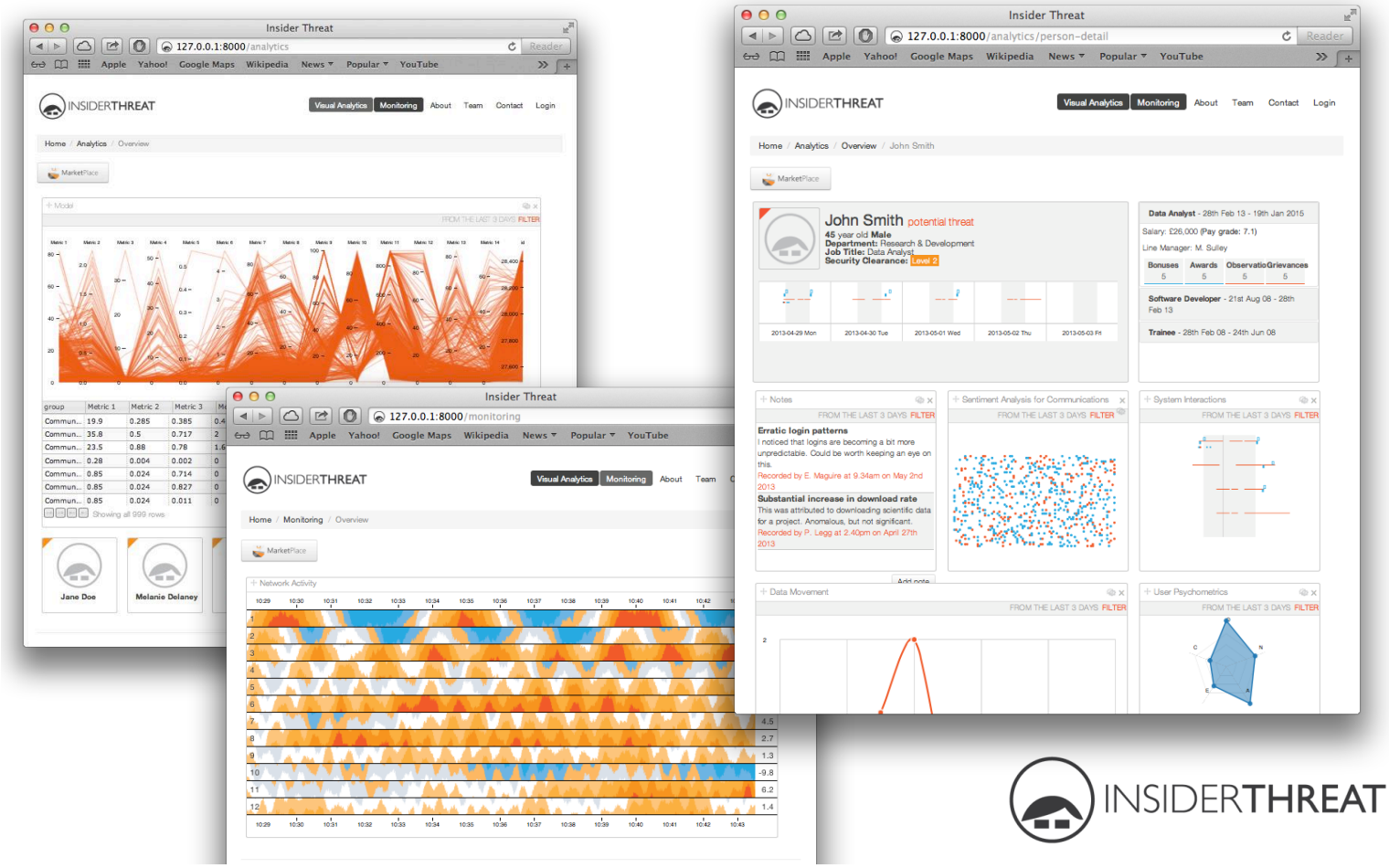


INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Visual Analytics



INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Moving Forward

- We have developed a detection prototype that proves effective for our initial testing on available data sets
- We need to ensure that our system is widely applicable, and can cope with varied scenarios and different organisational data structures in order to be effective
- Currently deploying against real data to experiment on – we also welcome more with real world scenarios who can share anonymized data or experiences to test against
- Folding in cyber indicators of psychological traits and state

INSIDERTHREAT

Thank you for listening!

Professor Michael Goldsmith
michael.goldsmith@cybersecurity.ox.ac.uk
Cyber Security Centre, University of Oxford, UK

E-mail Analysis

- E-mail is perhaps the most expressive presentation of an insider's behaviour and intentions, that can easily be captured in digital form and can be processed for further analysis
- E-mail can show *who* an individual makes contact with and *how* an individual communicates within the workplace (sociolinguistics)
- If an individual begins to vary their patterns of communication, either in terms of *who* or *how* they communicate, could this be indicative of a threat that could be prevented?
- In particular, what if their communication is indicative of some potentially threatening psychological state – such as an increase in tendency towards Narcissism or Machiavellianism?

Preserving Privacy

- Monitoring e-mail content is highly intrusive and a breach of privacy
  - of course, in some professions this breach of privacy may well need to be accepted!
- For further processing of data, it is typical to obtain a series of features that characterise the original data
- Can we obtain a series of features that provide sufficient detail to characterise the e-mail, without the need to breach privacy of the user?
  - at least, until there is strong evidence that the user is a threat!

Linguistic Inquiry Word Count (LIWC)

Dictionary-based text-analysis tool (80 dictionaries):

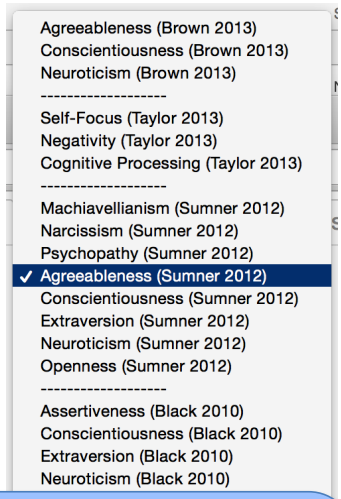
- Linguistic Processes
  - words > 6 letters, pronouns, verbs, tense, negation, swear words
- Psychological Processes
  - family, friends, positive/negative emotions, cognitive, perceptual, relativity
- Current Concerns
  - work, achievement, leisure, home, money, religion, death
- Spoken categories
  - assent, nonfluences, fillers
- Punctuation
  - periods, commas, exclamation marks, emoticons

For a given text, LIWC calculates the percentage of words which occur in each of the 80 dictionaries

Linking psychology to LIWC

There exists considerable research that links LIWC to psychological characteristics

- Neuroticism (Brown 2013):
  - *i_negate_negemo_anx_anger_cogmech_cause_discrep_tentat_certain*
- Self-Focus (Taylor 2013):
  - *ppron_i_we_you*
- Psychopathy (Sumner 2012):
  - *we-_preps-_swear+_family-_posemo-_negemo+_anger+_incl-_percept-_see-_body+_sexual+_relativ-_motion-_time-_work-_death+_filler+_exclam-*
- Assertiveness (Black 2010):
  - *negemo+_achieve+_anger*
- Narcissism (Williams 2003):
  - *sad+_anger+*



But… How do we know the impact that each LIWC category should have on psychological characteristics?

INSIDERTHREAT

Visual Analytics

- We are developing a visual-analytics tool for sociolinguistic e-mail analysis that relates LIWC features to psychological characteristics

- The analyst has control over the impact that each feature has towards a given characteristic, which can then be applied to all users

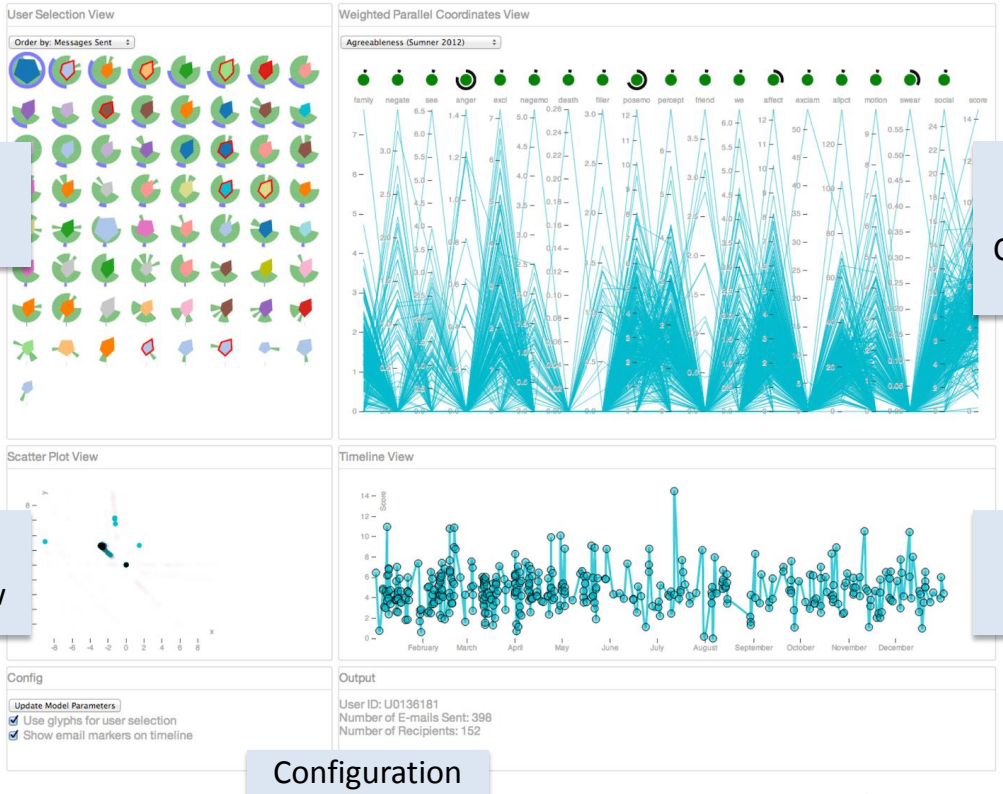
Visual Analytics Requirements

- The system should be able to provide an overview of all users
  - this could potentially be hundreds or thousands of users in a large organisation
- The system should be able to provide detail for comparative assessment of one or more users
  - observation of how psychological characteristics may change over time
- The analyst should be able to interact directly with the analytical model that defines how each LIWC feature contributes towards the assessment
- The analyst should be able to identify which users are currently deemed as a concern requiring further investigation, based on the current state of the model

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Visual Analytics Overview



User Selection View

Weighted Parallel Coordinates View

Feature Space View

Timeline View

Configuration and Status Views

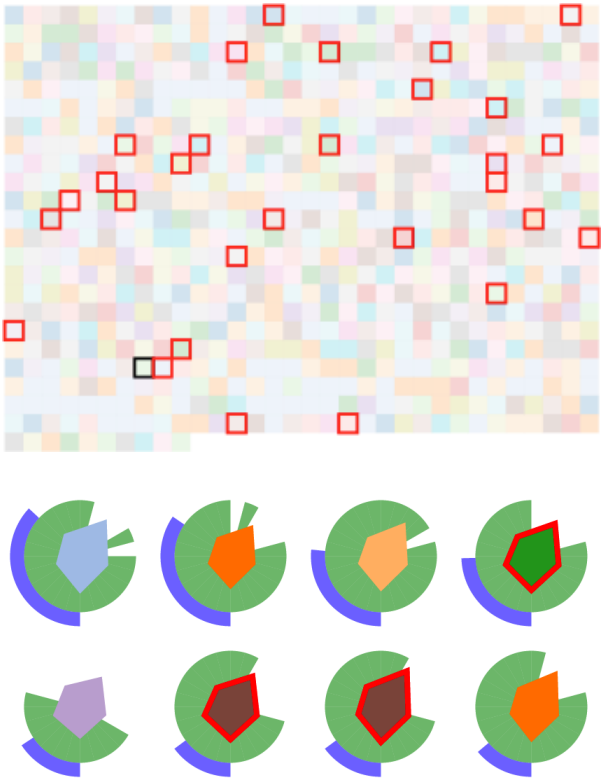
INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

User Selection View

- Two approaches for representing a large number of users:
  - pixel-based visualization
  - glyph-based visualization
- Pixel view shows many users with one attribute (e.g., *#emails*)
- Glyph view shows fewer users, but more attributes (*#emails*, *time of day*, *OCEAN*)
- Black outline shows sender
- Red outline shows recipients

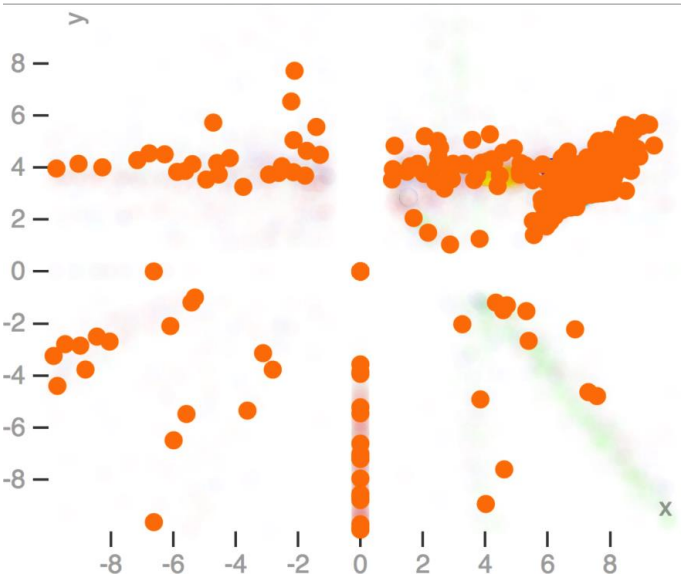


UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Feature Space View

- Each e-mail is captured as a set of LIWC features
- We perform dimensionality reduction (PCA) to observe the similarity between communications
- Spread of data points indicates deviation of communication patterns



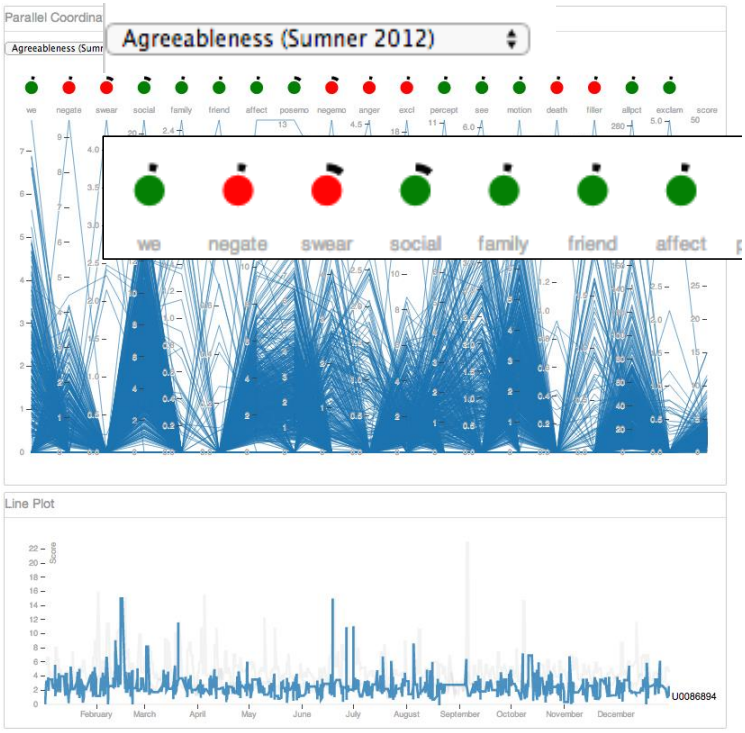
INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Weighted Parallel Coordinates

- Parallel Coordinates shows each e-mail as scored against the LIWC features.
- Each e-mail contributes to a psychological score, based on the weight of each LIWC feature
- User can adjust weights to reconfigure scoring model
- Timeline shows e-mail scores in temporal domain



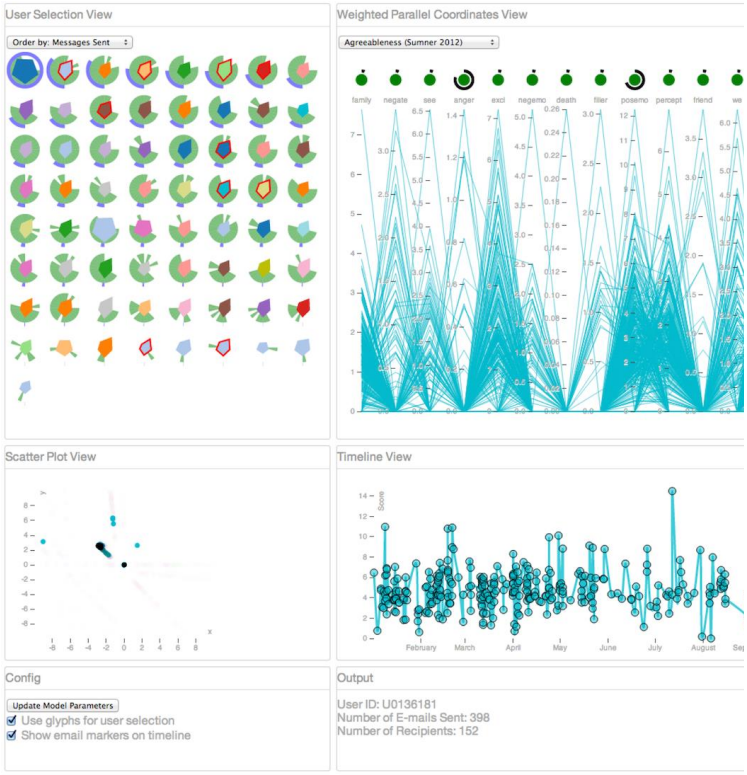
INSIDERTHREAT

UNIVERSITY OF OXFORD

CYBER SECURITY CENTRE

Visual Analytics Workflow

- Analyst can configure psychological models based on the desired impact of LIWC features
- Interaction with the model will update all other views
- User selection can be sorted by *OCEAN*, *#email*, deviation values, etc
- Which users deviate in their behaviour compared against our tuned model?



INSIDERTHREAT

E-mail Analysis Conclusion

- We have presented a proof-of-concept visual-analytics system for analysing behavioural deviations in large volumes of e-mail data from multiple users
- We propose using LIWC as a privacy-preserving scheme for studying behavioural change in e-mail content without explicit need for direct observation
- We are currently deploying our software in a real-world organisation to conduct experimentation on how human behaviours deviate, and how this reflects on the threat they may pose