THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Data Provenance and Cybersecurity –

Research Challenges and Opportunities

Prof Ryan Ko, UQ Cyber Security

ryan.ko @ uq.edu.au

Ryan’s provenance

- I was born in Singapore, did my Bachelors and PhD at NTU, and have the privilege to work across industry and academia throughout my career.

- Currently: Chair of Cyber Security; Director of UQ Cyber Security; Discipline Leader, Cyber Security & Software Engineering, School of ITEE

- Some highlights include:
  - **Building New Programmes:** Establishing UQ Cyber (2019), Cyber Security Researchers of Waikato (CROW) (2012), NZ Cyber Security Challenge (2014) and the NZ Institute of Security and Crime Science (2017) at the University of Waikato
  - **Curricula Development:** Writing the cyber security ISO standard ISO/IEC 21878, co-authored the (ISC)2 CCSP certification, wrote the NZ level-6 curriculum on cyber security, and started 3 interdisciplinary degree programmes in cyber security: Waikato Uni: (Master of Cyber Security, Master of Security & Crime Science); UQ: Master of Cyber Security
  - **Research across HP Labs and academia:** Invented/co-invented cyber security and provenance tools, e.g. HP security tools and data tracking tools used globally (e.g. DHCPv6 Fuzzer (Kali Linux); Fraud detection for USA Treasury, IRS; cloud provenance tracking for Singapore Government Cloud)
  - **Advising** the INTERPOL, ministers, governments, and several listed companies on cyber security technologies.

UQ CYBER SECURITY

INDUSTRY

AusCERT
and members

THE UNIVERSITY OF QUEENSLAND AUSTRALIA
CREATE CHANGE

UQ ITS

Industry, Governments &
International Organisations

ACADEMIC RESEARCH

EAIT (ITEE)

Science (Maths & Physics)

HASS

BEL

Medicine

Centre for Policy Futures

Inst for Social Science Research

AusCERT Research

EDUCATION & TRAINING

CPD Courses

Bachelor Computer Science
(Cyber Security Major)

Interdisciplinary
Master of Cyber Security
(Launched Aug 2019; any Bachelors degree)

PhD

UQ Cyber Research Areas

- Secure quantum communications
- IoT and cyber physical security
- Secure software engineering
- Critical Infrastructure – Industrial Control Systems (ICS) Security
- Secure communications for space
- Data privacy and user data control
- National security and cyber policies
- Cyber autonomy and automation
- Cyber law and ethics
- Cyber criminology

FORGAN SMITH
BUILDING

Acknowledgements

This talk encompasses data provenance and security research from 2010-2020 at HP Labs, University of Waikato, and the University of Queensland. Further details in references.

Qn: How do you find out who/what is the source of a cyber attack?

Possible sources

Computer Logs?

Telecommunication records?

Network packet captures?

Account login information?

Financial payments? (e.g. for ransomware)

Verified information from trusted intelligence agencies?

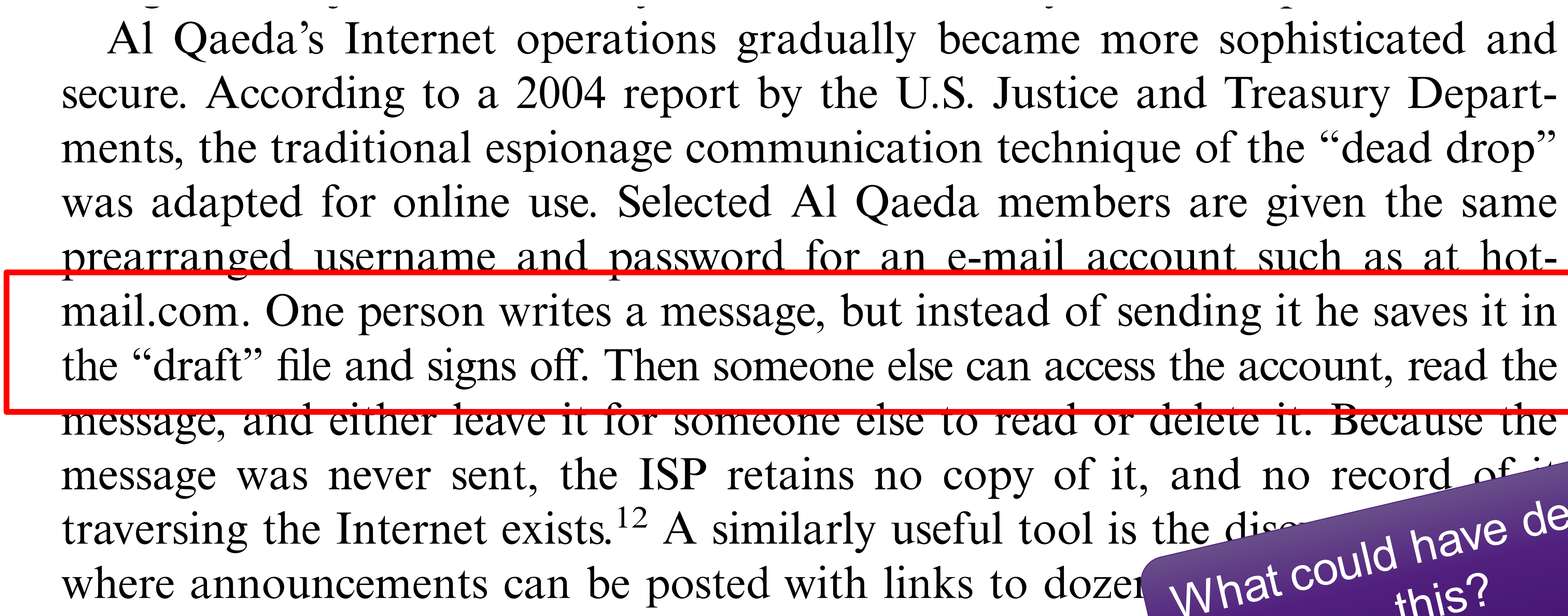
Question: How do you trust these sources?

Qn: Are current systems fit-for-purpose in terms of supporting the attribution of cyber attacks?

Many challenges including a lack of data-centric systems design, encrypted channels, etc.

In this talk, our focus: Data-centric systems design to enable high quality, high-integrity provenance.

September 11, 2001



Al Qaeda's Internet operations gradually became more sophisticated and secure. According to a 2004 report by the U.S. Justice and Treasury Departments, the traditional espionage communication technique of the "dead drop" was adapted for online use. Selected Al Qaeda members are given the same prearranged username and password for an e-mail account such as at hotmail.com. One person writes a message, but instead of sending it he saves it in the "draft" file and signs off. Then someone else can access the account, read the message, and either leave it for someone else to read or delete it. Because the message was never sent, the ISP retains no copy of it, and no record of it traversing the Internet exists.[12] A similarly useful tool is the dis where announcements can be posted with links to dozen

What could have detected this?

Yes, data-centric tools, e.g. provenance tools

Src: Seib and Janbek, Global Terrorism and New Media – The post Al-Qaeda Generation, 2011

The University Of Queensland
Australia
CREATE CHANGE

Defining and Understanding Provenance

Defining Provenance

Across many papers in the literature, provenance is commonly defined as the “derivation history of data”.

Like many IT terminologies, provenance has several overlapping terms linked to it:

- Data lineage, genealogy, information flow, etc.

This mainly stems from the lack of (deep) consideration of the nature of provenance.

My hope is that by the end of this talk, we take provenance (more) seriously. ☺

Observable phenomena

Time

The concept of time is self-evident.

An hour consists of a certain number of minutes, a day of hours and a year of days.

But we rarely think about the fundamental nature of time.

Time is passing non-stop, and we follow it with clocks and calendars. Yet we cannot study it with a microscope or experiment with it. And it still keeps passing. We just cannot say what exactly happens when time passes.

Time is represented through change, such as the circular motion of the moon around Earth. The passing of time is indeed closely connected to the concept of space.

One of the most peculiar qualities of time is the fact that it is measured by motion and it also becomes evident through motion.

Ref: University of Helsinki. "What is time?." ScienceDaily. ScienceDaily, 15 April 2005. <www.sciencedaily.com/releases/2005/04/050415115227.htm>.

Provenance

Similarly, like time, the concept of provenance is self-evident.

Provenance consists of a sequence of actions and sub-actions on an object of interest. The sequence tells us the story of the object.

But we rarely think about the fundamental nature of provenance.

Provenance is passing and generated non-stop, and we follow it with logs, artefacts, and different types of metadata.

Yet we cannot study it with a microscope or experiment with it. And it still keeps passing. We just cannot say what exactly happens when provenance passes.

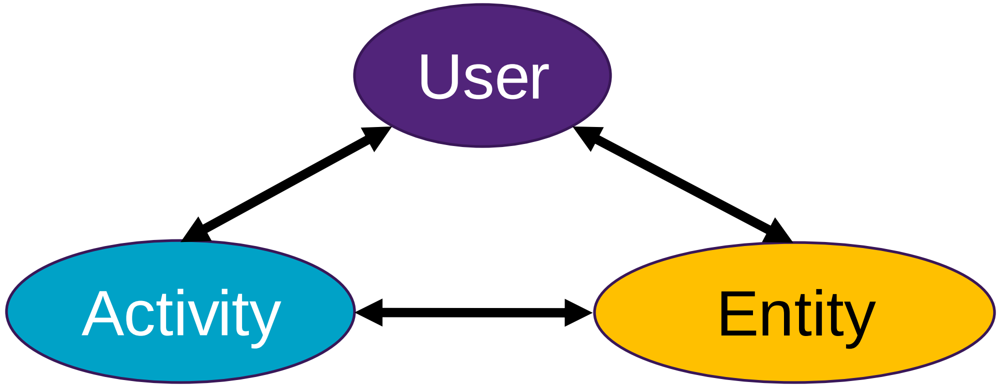
Provenance is represented through change, such as the change of ownership of the Mona Lisa painting. The passing of provenance is indeed closely connected to the concept of data.

One of the most peculiar qualities of provenance is the fact that it is measured by data actions and it also becomes evident through data actions.

If I can describe the role of provenance in cyber security through 1 character, it has to be…

Image src: https://www.insider.com/game-of-thrones-jon-snow-learns-truth-parents-sam-2019-4

Everyday Provenance



Cloud Service Provider	Payroll System	Physical Security Card Access System
Bob opened MSWord to read a document, and copied a paragraph to another document.	Bob initiated a payroll sequence to pay salary from the company account to a personal account.	Bob swiped his card to enter a meeting room.

What has that got to do with cybersecurity?

Attribution

Assurance of data integrity

Trust

Accountability and auditability

Digital Rights Management

Forensics (though strictly speaking, forensics != security)

Proactive Security

Data Provenance in Computer Science & Cyber Security

Evolution of Provenance Research

Since the 1970s, in Computer Science, we have been looking at provenance in the context of:

1970s-1990s: Scientific workflows

1990s-early 2000s: Web (e.g. W3C PROV)

2000s-2010s: Operating Systems (e.g. PASS, SPADE)

2010s-now: Distributed systems such as cloud computing (e.g. Flogger, Progger, S2Logger); Ledgers (Blockchains; DLT)

Now-future: Cyber-physical systems: IoT and Operational Technology (OT)

Provenance Research Opportunities and Challenges

Capturing

- What to capture, where to capture
- Describing and formalizing provenance as an algebra
- Standardising the ‘layers’ of abstraction in the full provenance stack

Storing

- How do we store and where do we store (kernel?)
- How do we handle distributed environments such as the cloud?
- Integrity of provenance metadata (e.g. evidentiary requirements for justice)

Access and Analytics

- Visualising provenance
- Fusion of provenance data
- Inferring user behaviour from provenance (e.g. user actions; Behavioral biometrics)
- Using provenance for prevention and disruption/revocation (e.g. Proactive security)
- Automating compliance for data governance

Case Studies:

Capturing, storing and analyzing Cloud Data Provenance

Cloud Data Provenance – *Related Work*

- *User-Space, Centralized File System Call Monitor*
  - *Local machine - iNotify, swatch, file alteration monitor (FAM)*
  - *Network packet monitoring – snort [Roesch 1999]*
- *File Integrity Checkers as Intrusion Detection*
  - *E.g. TripWire*
- *Virtual Environment Monitors*
  - *E.g. HyTrust CloudControl*
- *Cloud Systems Health and Performance Monitoring*
  - *VMWare vFabric Hyperic, CloudKick*
- *Database provenance*
  - Unable to capture entire end-to-end data flow in clouds
  - However, very useful as part of cloud data provenance linking application to system levels.
- *Traditional Data Provenance Research*
  - *PASS – Provenance Aware Storage System [Muniswamy-Reddy et al. 2009,2010]*
  - SPADE, sysdig, etc.
  - Not tamper-evident
  - Mostly user-space

- Ryan K. L. Ko, Peter Jagadpramana, Bu Sung Lee, “Flogger: A File-centric Logger for Monitoring File Access and Transfers within Cloud Computing Environments”, 10th IEEE TrustCom, 16-18 November 2011, Changsha, China.
- Olive Qing Zhang, Markus Kirchberg, Ryan K. L. Ko, Bu Sung Lee, “How To Track Your Data: The Case for Cloud Computing Provenance”, 3rd IEEE CloudCom 2011, 29 Nov-1 Dec 2011, Athens, Greece.
- Ryan K. L. Ko, Markus Kirchberg, Bu Sung Lee, “From System-Centric Logging to Data-Centric Logging – Accountability, Trust and Security in Cloud Computing”, Defence, Science and Research Conference 2011, IEEE Computer Society, 3-4 August 2011, Singapore.

Industry – still lacking something at that time…

However, we need a Cloud-ready, file-centric, detective mechanism which can do the following:

- *Transcend Virtual Machines/ Physical Machines (Full transparency)*
- *Must provide Provenance (History)*
- *Single Auditable View (Empowering Decision Makers)*
- *Efficient storage (Empowering forensics and compliance with retention policies)*
- *Strong Analytics (Instant insights and adoption into future detections)*
- *End-to-End Tracking (Not only tracking of single machines, but* ***entire*** *Cloud)*

HP Labs Flogger/S2Logger

(Tech transferred to HP ArcSight, 2010-2012)

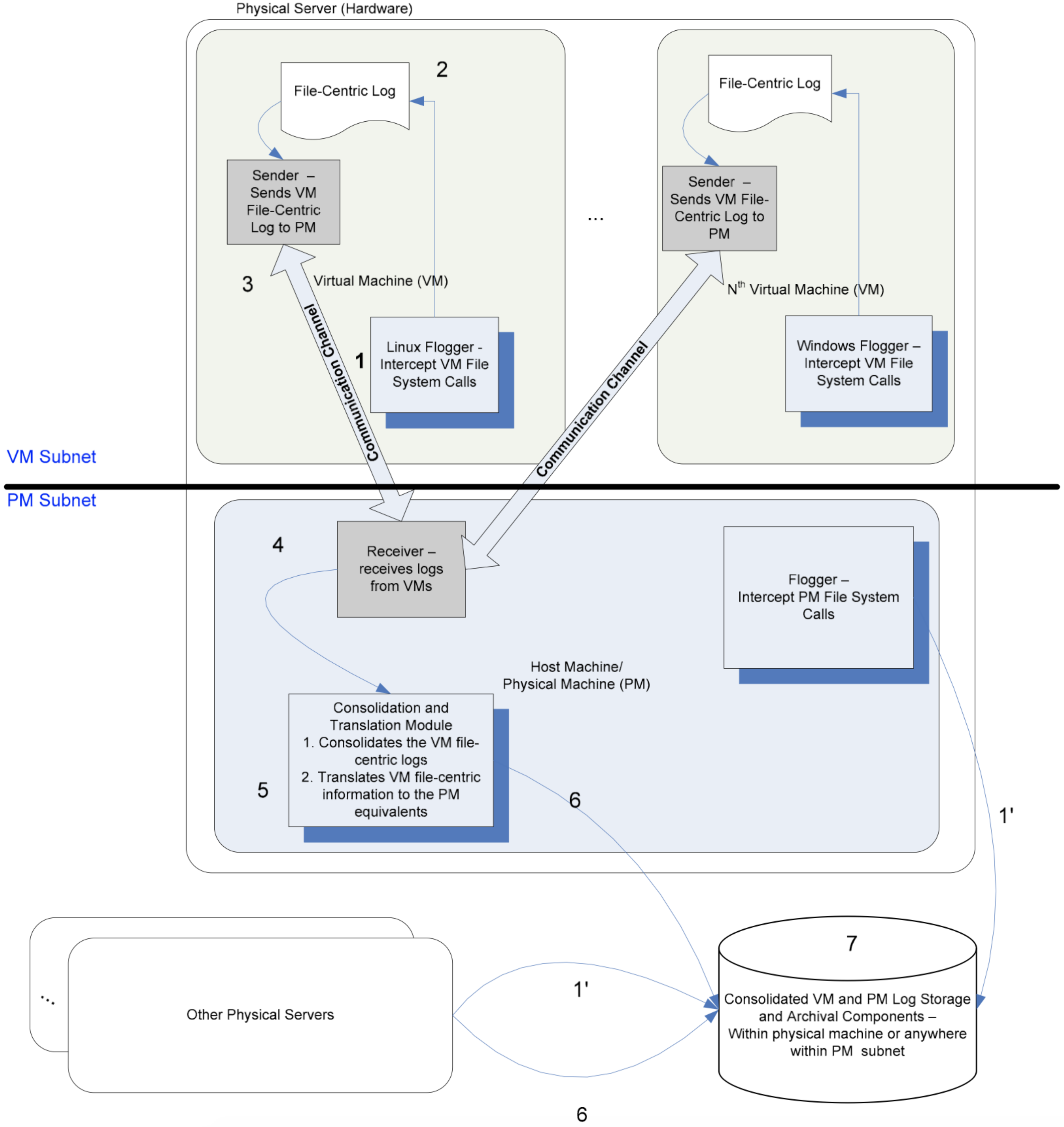
- *Optimised log size and format (space-efficient)*
- *Efficient storage, processing and transfer of logs*
- *Tamper-evidence*
- *Resilient to threats which render logs unusable (e.g. ransomware)*

Progger

(Tech transferred to industry partners, e.g. LayerX; 2013-2017, Industrial control systems HMIs; 2018-2019)

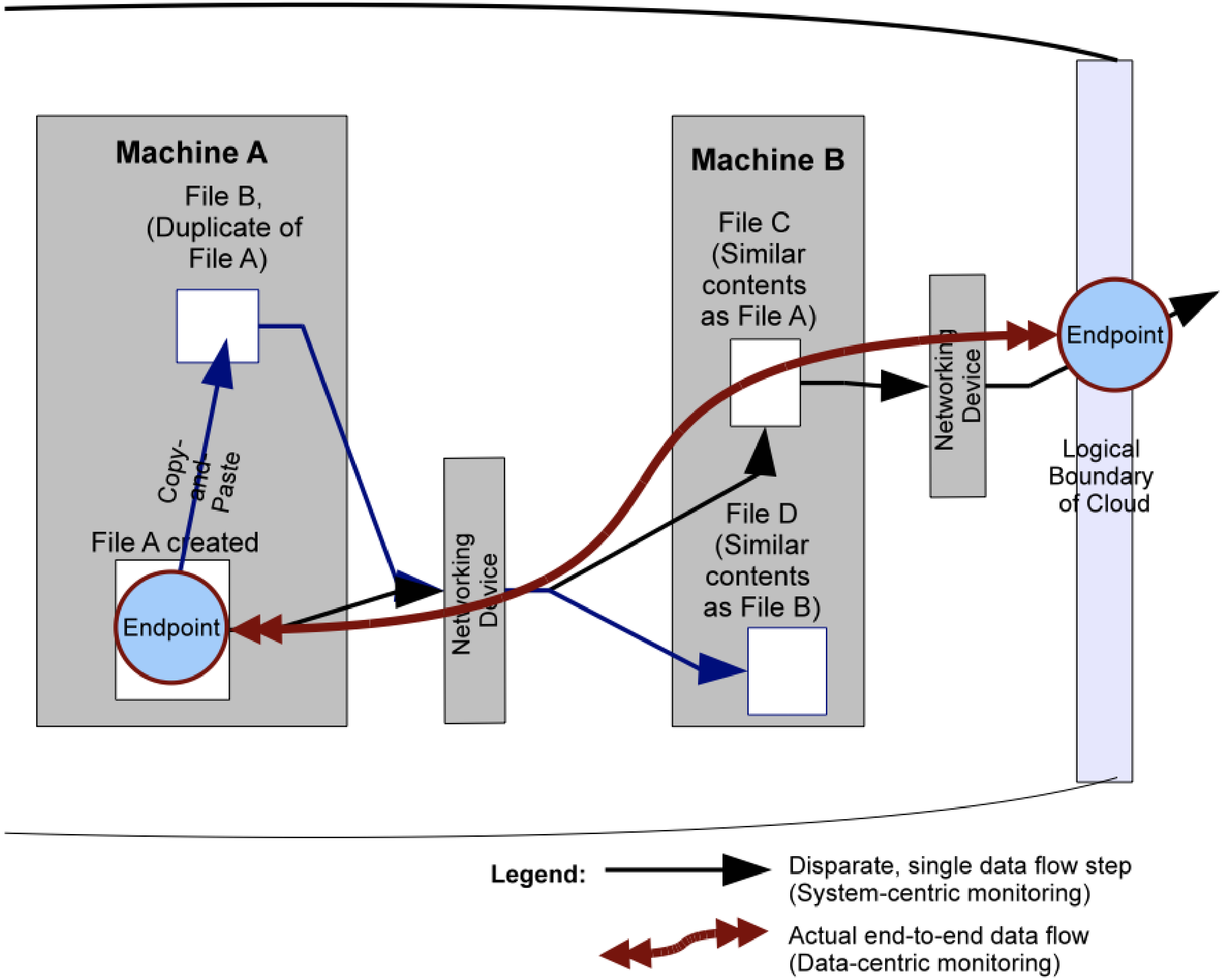
Flogger (2011)

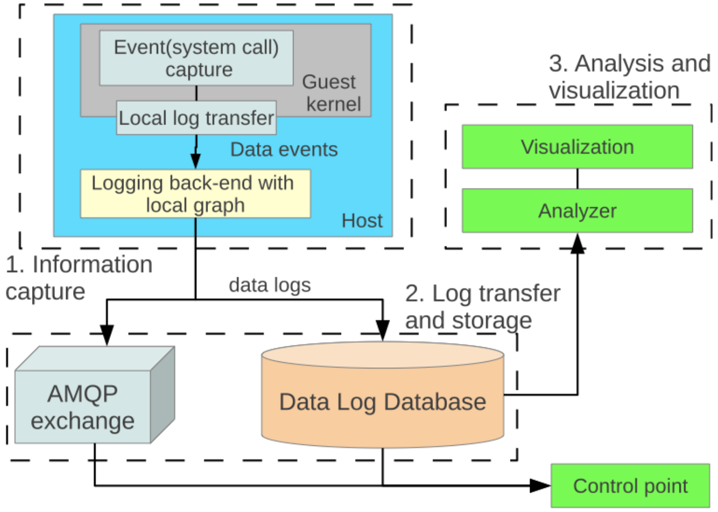
Capturing provenance at kernel space in both physical and virtual machines


Physical Server (Hardware)
File-Centric Log
2
Sender – Sends VM File-Centric Log to PM
3
Virtual Machine (VM)
Communication Channel
1
Linux Flogger - Intercept VM File System Calls
…
File-Centric Log
Sender – Sends VM File-Centric Log to PM
Nth Virtual Machine (VM)
Communication Channel
Windows Flogger – Intercept VM File System Calls
VM Subnet
PM Subnet
4
Receiver – receives logs from VMs
Flogger – Intercept PM File System Calls
Host Machine/ Physical Machine (PM)
Consolidation and Translation Module 1. Consolidates the VM file-centric logs 2. Translates VM file-centric information to the PM equivalents
5
6
1'
…
Other Physical Servers
1'
7
Consolidated VM and PM Log Storage and Archival Components – Within physical machine or anywhere within PM subnet
6

Ref: Ryan K. L. Ko, Peter Jagadpramana, Bu Sung Lee, “Flogger: A File-centric Logger for Monitoring File Access and Transfers within Cloud Computing Environments”, 10th IEEE TrustCom, 16-18 November 2011, Changsha, China.

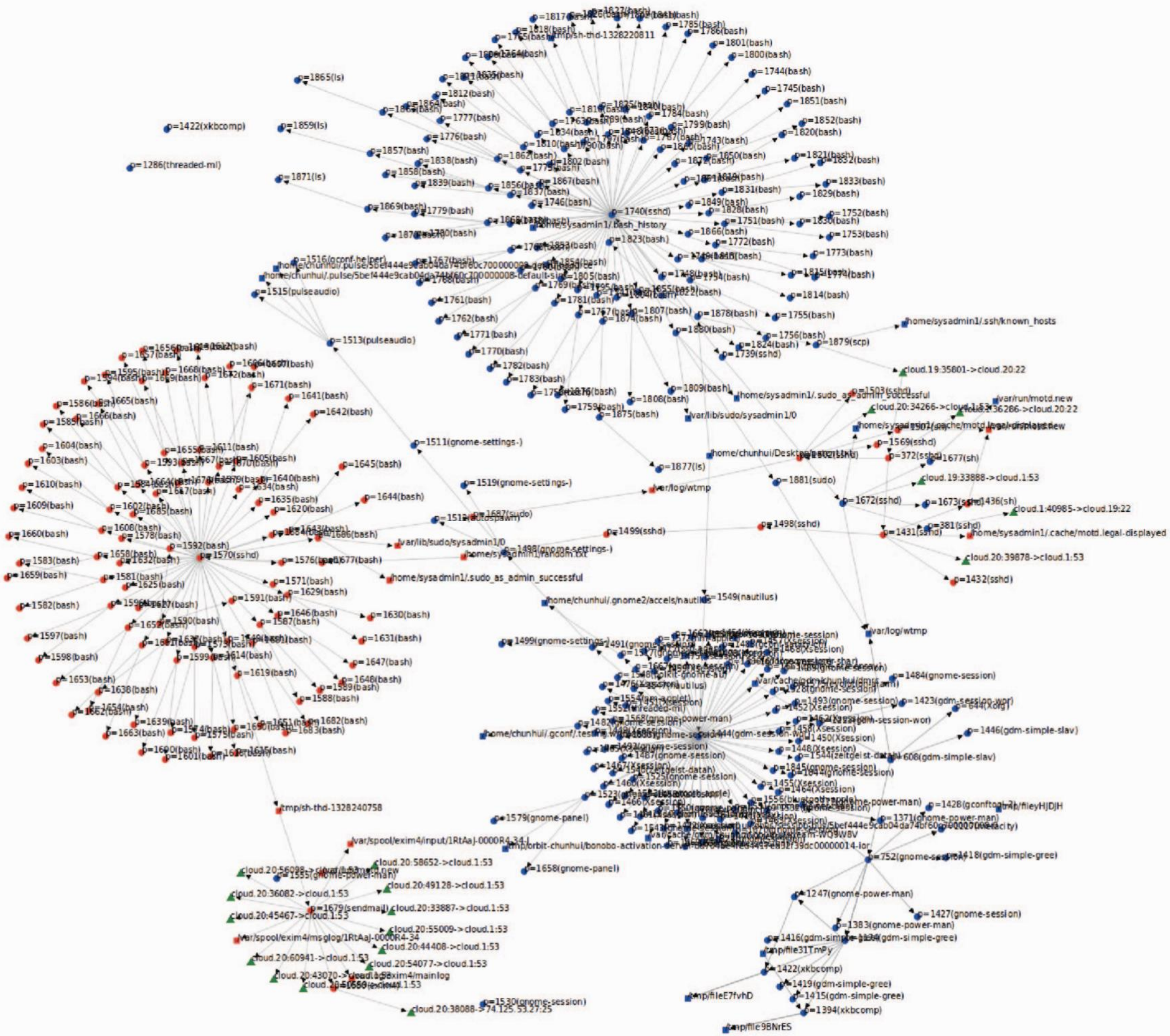
S2Logger (2013) – Capturing Kernel-space Block-level Provenance


Machine A
File B, (Duplicate of File A)
Copy-and-Paste
File A created
Endpoint
Networking Device
Machine B
File C (Similar contents as File A)
File D (Similar contents as File B)
Networking Device
Endpoint
Logical Boundary of Cloud
Legend:
Disparate, single data flow step (System-centric monitoring)
Actual end-to-end data flow (Data-centric monitoring)



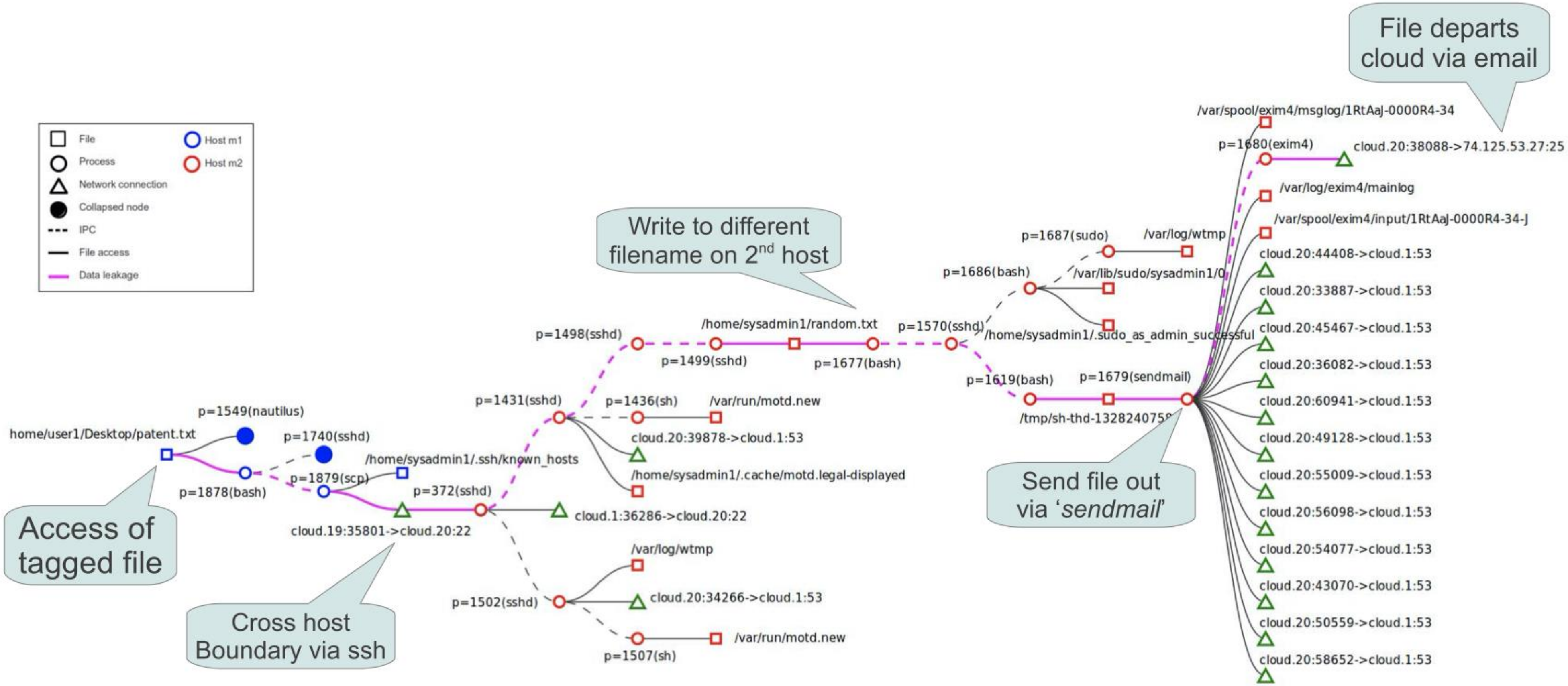
Ref: Chun Hui Suen, **Ryan K. L. Ko**, Y. S. Tan, Peter Jagadpramana and B. S. Lee (2013) "**S2Logger: End-to-End Data Tracking Mechanism for Cloud Data Provenance**." 12th IEEE TrustCom, Melbourne, Australia, 16-18 July **2013**, pp. 594-602. IEEE.

Provenance captured by S2Logger for 1 File transferred across 2 machines in a cloud



Ref: Chun Hui Suen, **Ryan K. L. Ko**, Y. S. Tan, Peter Jagadpramana and B. S. Lee (2013) "**S2Logger: End-to-End Data Tracking Mechanism for Cloud Data Provenance**." 12th IEEE TrustCom, Melbourne, Australia, 16-18 July **2013**, pp. 594-602. IEEE.

S2Logger Visualization of Scenario’s Provenance


□ File
○ Process
△ Network connection
● Collapsed node
--- IPC
— File access
— Data leakage
○ Host m1
○ Host m2
home/user1/Desktop/patent.txt
p=1549(nautilus)
p=1740(sshd)
p=1878(bash)
p=1879(scp)
/home/sysadmin1/.ssh/known_hosts
cloud.19:35801->cloud.20:22
p=372(sshd)
cloud.1:36286->cloud.20:22
p=1431(sshd)
p=1436(sh)
/var/run/motd.new
cloud.20:39878->cloud.1:53
/home/sysadmin1/.cache/motd.legal-displayed
p=1502(sshd)
/var/log/wtmp
cloud.20:34266->cloud.1:53
p=1507(sh)
/var/run/motd.new
p=1498(sshd)
p=1499(sshd)
/home/sysadmin1/random.txt
p=1677(bash)
p=1570(sshd)
p=1686(bash)
p=1687(sudo)
/var/log/wtmp
/var/lib/sudo/sysadmin1/0
/home/sysadmin1/.sudo_as_admin_successful
p=1619(bash)
p=1679(sendmail)
/tmp/sh-thd-132824075
/var/spool/exim4/msglog/1RtAaJ-0000R4-34
p=1680(exim4)
cloud.20:38088->74.125.53.27:25
/var/log/exim4/mainlog
/var/spool/exim4/input/1RtAaJ-0000R4-34-J
cloud.20:44408->cloud.1:53
cloud.20:33887->cloud.1:53
cloud.20:45467->cloud.1:53
cloud.20:36082->cloud.1:53
cloud.20:60941->cloud.1:53
cloud.20:49128->cloud.1:53
cloud.20:55009->cloud.1:53
cloud.20:56098->cloud.1:53
cloud.20:54077->cloud.1:53
cloud.20:43070->cloud.1:53
cloud.20:50559->cloud.1:53
cloud.20:58652->cloud.1:53

File departs cloud via email

Write to different filename on 2nd host

Send file out via ‘*sendmail*’

Access of tagged file

Cross host Boundary via ssh

Ref: Chun Hui Suen, **Ryan K. L. Ko**, Y. S. Tan, Peter Jagadpramana and B. S. Lee (2013) “**S2Logger: End-to-End Data Tracking Mechanism for Cloud Data Provenance**.” 12th IEEE TrustCom, Melbourne, Australia, 16-18 July **2013**, pp. 594-602. IEEE.

Progger (2014-present) – Effective, kernel-space, tamper-evident provenance logger

Built for data traceability and accountability in clouds[1] (and eventually for OT HMI machines[2] in 2019)

Deployed as agent(s) in kernel-spaces of both Virtual and Physical Hosts

Both Windows (File System Mini Filter Driver) and Linux (Kernel Module)

Records CRUD file actions: Create; Read; Update; Delete

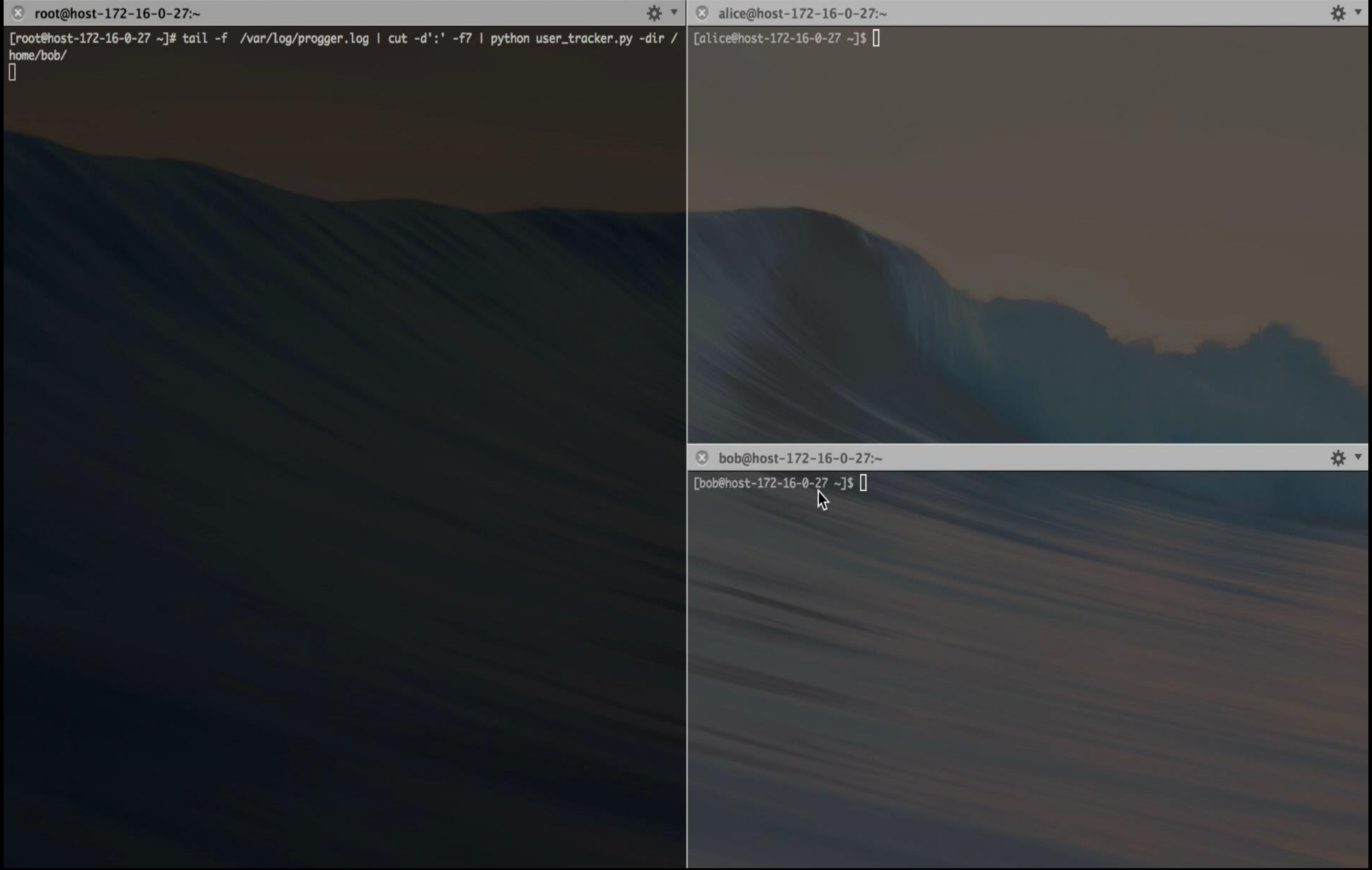
A 'CCTV' for file changes in clouds

Builds on experience since 2010 (i.e. Flogger, S2Logger)

Ref:

[1] **Ryan K. L. Ko,** Mark A. Will, "**Progger: An Efficient, Tamper-Evident Kernel-Space Logger for Cloud Data Provenance Tracking,**", IEEE CLOUD 2014, Alaska, USA, **2014.**

[2] Ryan K L Ko, Taejun Choi and Christian Anderson-Scott, "Security systems, methods and devices.", NZ Patent 756876 (Filed Aug 2019; PCT filed 2020)



```
[root@host-172-16-0-27 ~]# tail -f  /var/log/progger.log | cut -d':' -f7 | python user_tracker.py -dir /
home/bob/
```

```
[alice@host-172-16-0-27 ~]$
```

```
[bob@host-172-16-0-27 ~]$
```

Progger Architecture (1/2)

- **Linux Progger**
  - *Kernel module*
  - *Kernel versions (v4.xx; also works for 3.xx)*
  - *GPL License*
- Uses tracepoints to record file provenance

  Markers within the Linux kernel

  Hook into running kernel at the point where the marker is located
- Log entries sent to Redis server

  Used as message queue between Progger client and database

Windows Progger

- File System Mini Filter Driver

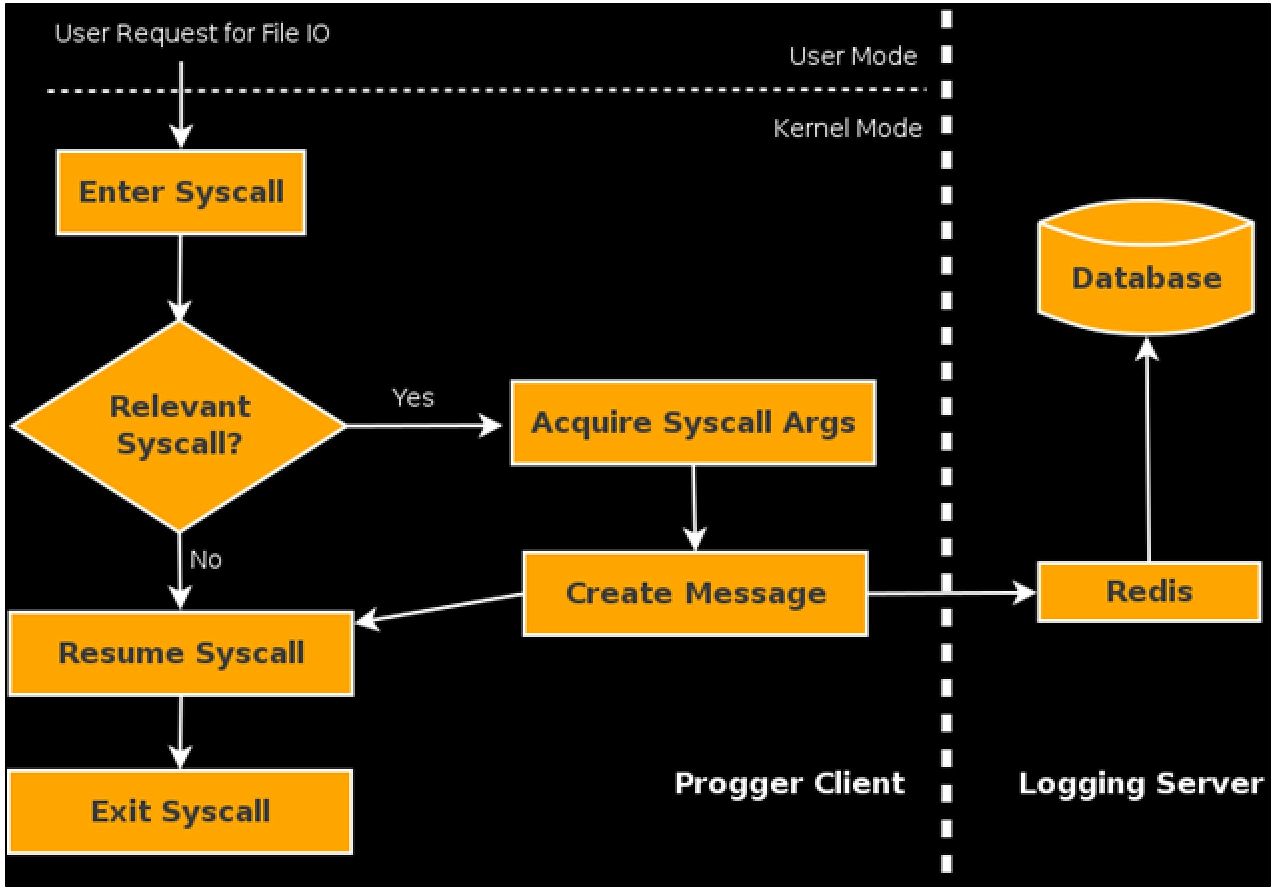
  Runs inside Windows NT Kernel (Version 6.3+)

  Closed sourced
- Loaded into kernel as a Dynamically Linked Device Driver by Filter Manager

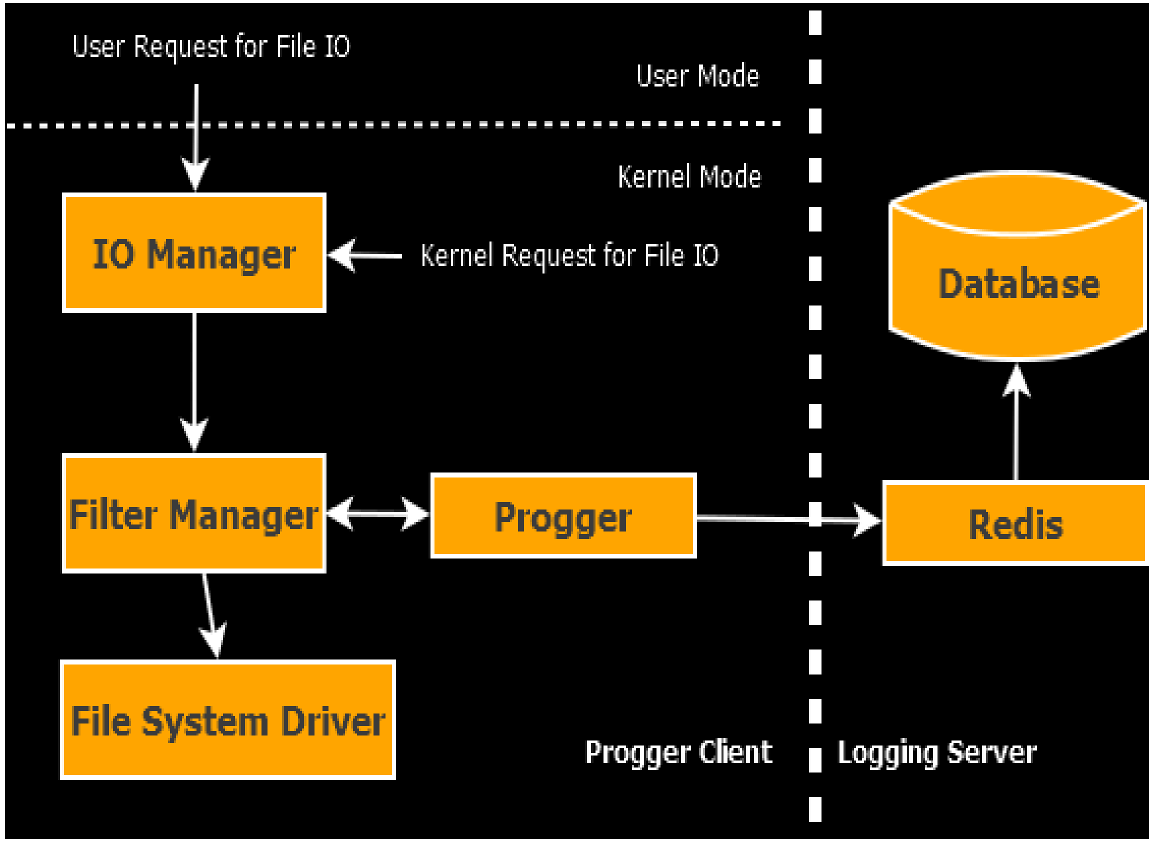
  Communication with Redis directly using TCP socket created using Winsock Kernel

Progger Architecture (2/2)

Linux Progger Architecture

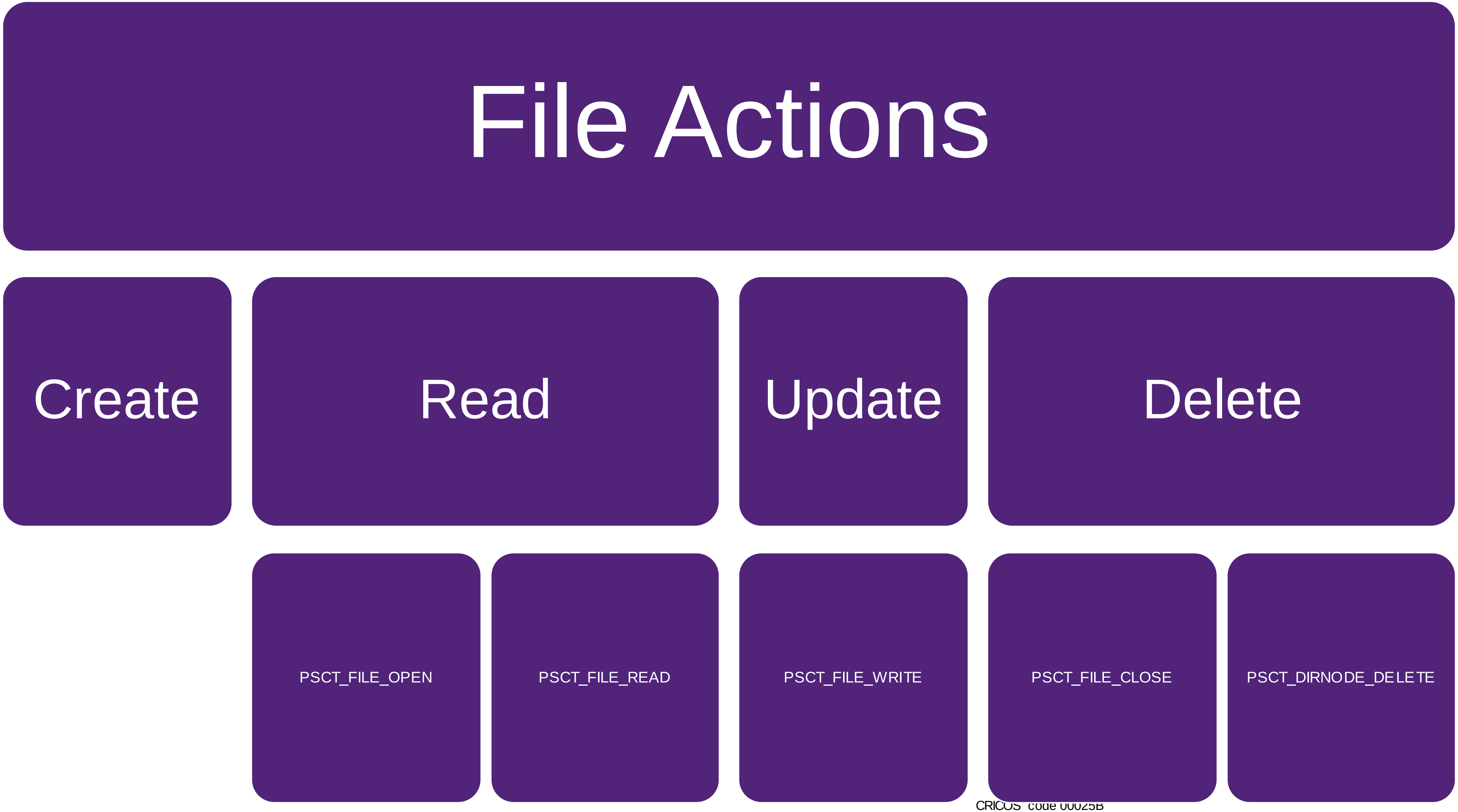


Windows Progger Architecture



Supported System Calls (Sample)

Windows	Progger Syscall Type	Linux
IRP_MJ_CREATE	PSCT_FILE_OPEN	Sys_open, sys_create
	PSCT_DIRNODE_DELETE	sys_unlink, sys_unlinkat
	PSCT_DIRNODE_CREATE	sys_mkdir
IRP_MJ_READ	PSCT_FILE_READ	sys_read, sys_pread64
IRP_MJ_WRITE	PSCT_FILE_WRITE	sys_write, sys_pwrite64
IRP_MJ_CLEANUP	PSCT_FILE_CLOSE	sys_close
IRP_MJ_SET_INFORMATION (FileDispositionInformation)	PSCT_DIRNODE_DELETE (File)	sys_unlink, sys_unlinkat
	PSCT_DIRNODE_DELETE (Directory)	sys_rmdir
IRP_MJ_SET_INFORMATION(FileRenameInformation)	PSCT_DIRNODE_RENAME (Rename)	sys_rename
IRP_MJ_SET_INFORMATION (FileLinkInformation)	PSCT_DIRNODE_RENAME (Hard Link)	sys_link
IRP_MJ_FILE_SYSTEM_CONTROL, FSCTL_SET_REPARSE_POINT	PSCT_DIRNODE_LINK	sys_link, sys_linkat
IRP_MJ_SET_SECURITY, OWNER_SECURITY_INFORMATION	PSCT_DIRNODE_CHANGE_OWNER	sys_chown, sys_chownat
IRP_MJ_SET_SECURITY, DACL_SECURITY_INFORMATION	PSCT_DIRNODE_CHANGE_PERMISSIONS	Sys_chmod, sys_chmodat
Not yet implemented	PSCT_HANDLE_DUPLICATE	sys_dup, sys_dup2, sys_pipe, sys_pip2



File Actions

Create

Read

Update

Delete

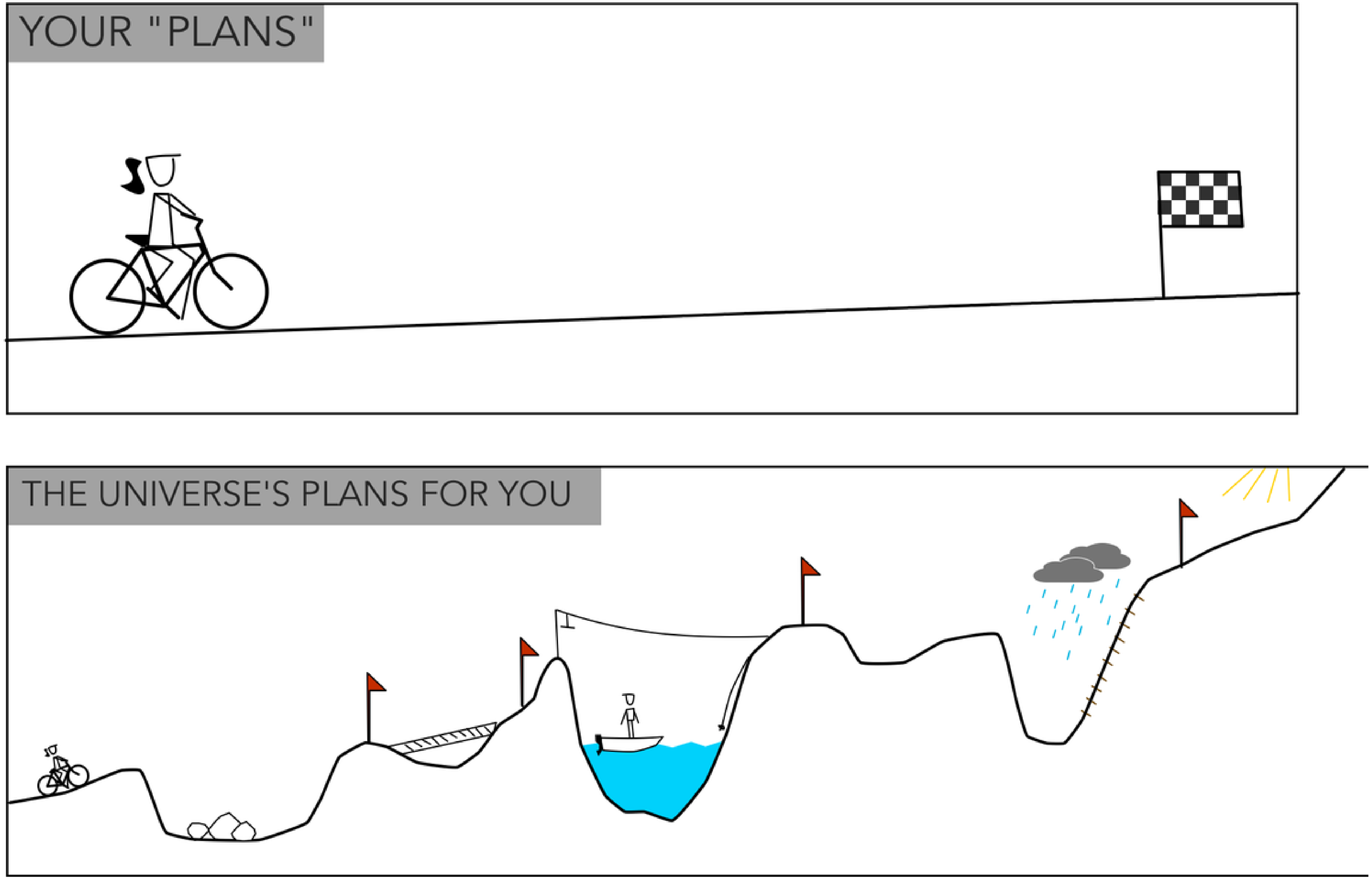
PSCT_FILE_OPEN

PSCT_FILE_READ

PSCT_FILE_WRITE

PSCT_FILE_CLOSE

PSCT_DIRNODE_DELETE



DOGHOUSEDIARIES

Logging Implementation

Initially entries output to log file (text, then binary)

Inefficient

Additional overheads when writing to log file

(plus lessons learned from previous work on HP Labs’ Flogger/S2Logger)

Now logs are sent over a network socket

Avoiding extra IO operations since no logging to disk

Ring buffer to be implemented in the near future in case Redis server cannot be reached

Benchmarking: Efficiency & Load on Current Systems

Progger tracking the installation of a program *(i.e. data CRUD-intensive)*:

- Performance measuring tools say the system ran at 25% CPU usage (worst case) with no RAM issues introduced by Progger.
- Reading dumps from Redis, Progger produced 1.6 Mbps of network traffic without taking into account TCP and Redis overheads.

In one experiment (self-made “Progger killer”):

- 9 million event dataset weighing in at 1.25GB collected over roughly 1 hour and 40 minutes.
- It compresses to 65MB.
- Assuming an entire day of compressed data with high system load that's less than a GB of data produced from 1 machine.

Collection efficiency:

- Running with focus given to the virtual machine over 90% of the time causes 80% of Windows Progger events to be collected in 10-20 microseconds with about 5% that take over 30 microseconds.

Stability:

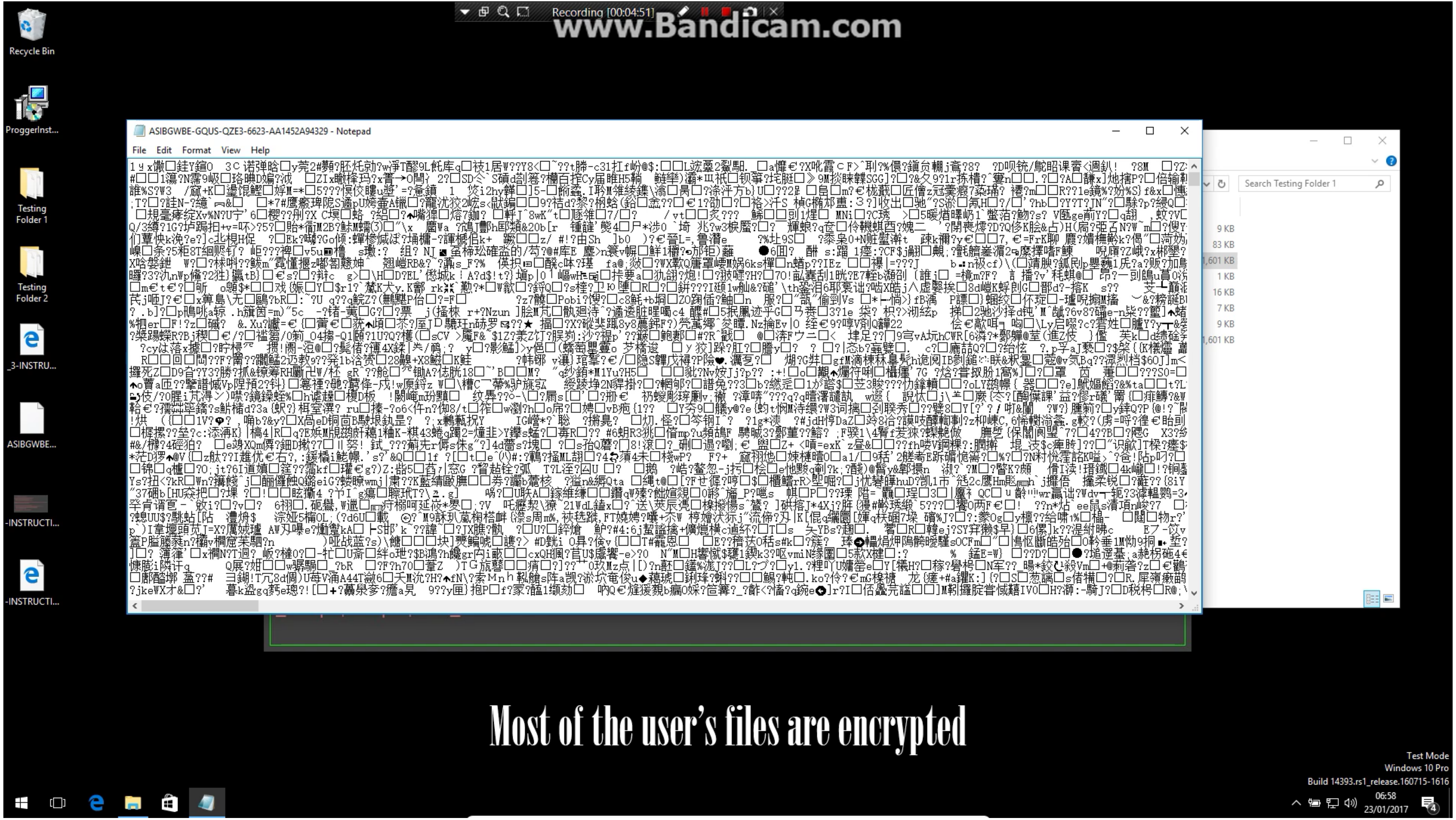
- Progger was executed for data provenance collection in about 180 attack and defence challenge machines in NZ Cyber Security Challenges 2016, 2017. No outages experienced.

Progger Log File Design & Compression

Test File 1 (496973 events)	File Size	Compression Ratio
No Compression	42,548 KB	N/A
GZIP Compression (Level9)	4,084 KB	9.6%
Zlib Stream Compression	8,839 KB	20.8%

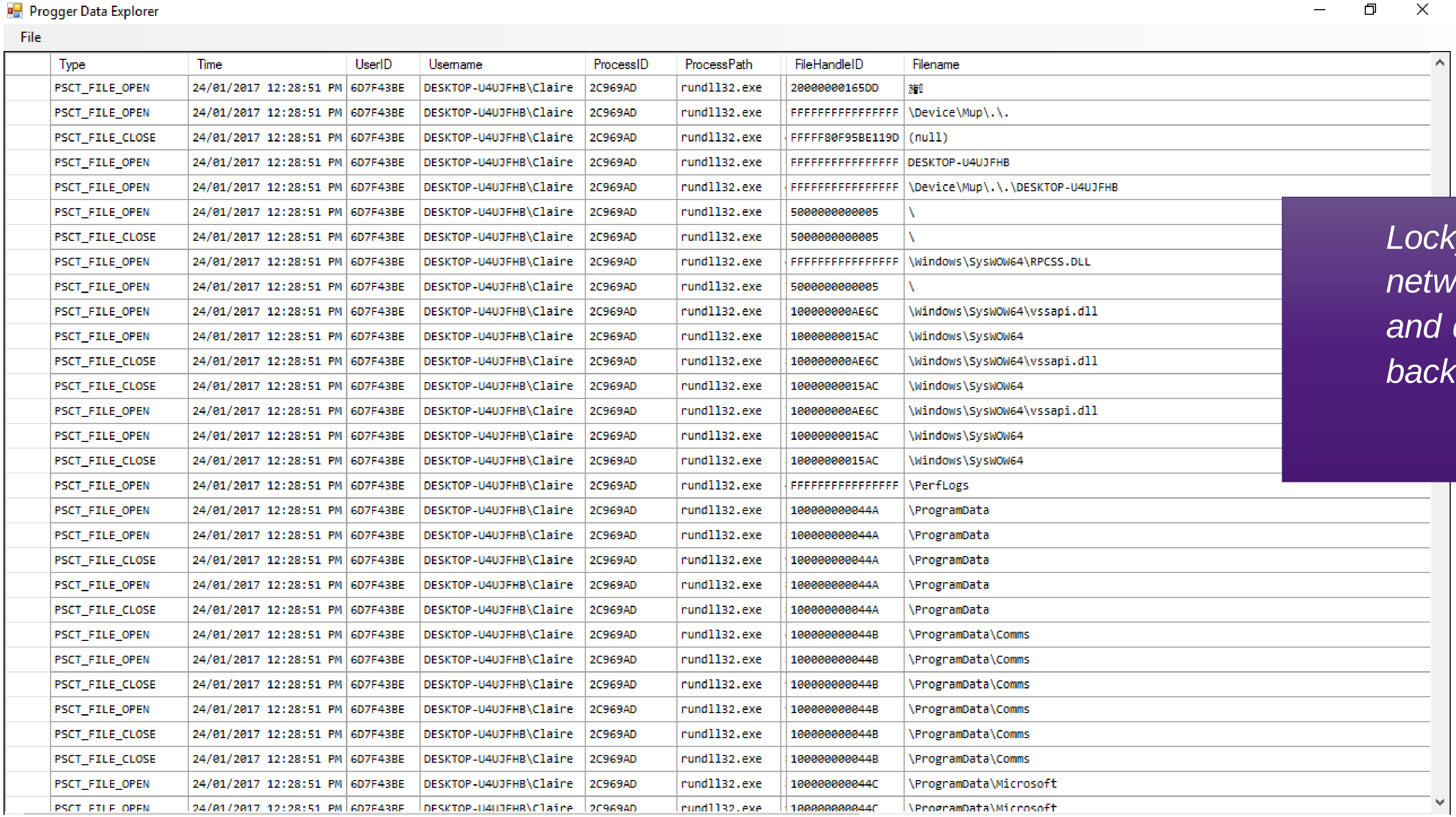
Test File 2 (890208 events)	File Size	Compression Ratio
No Compression	60,701 KB	N/A
GZIP Compression (Level9)	4,679 KB	7.7%
Zlib Stream Compression	13,028 KB	21.5%

Uses of Progger: Recording Locky Ransomware's Provenance (Collaboration with Trend Micro)



Most of the user's files are encrypted

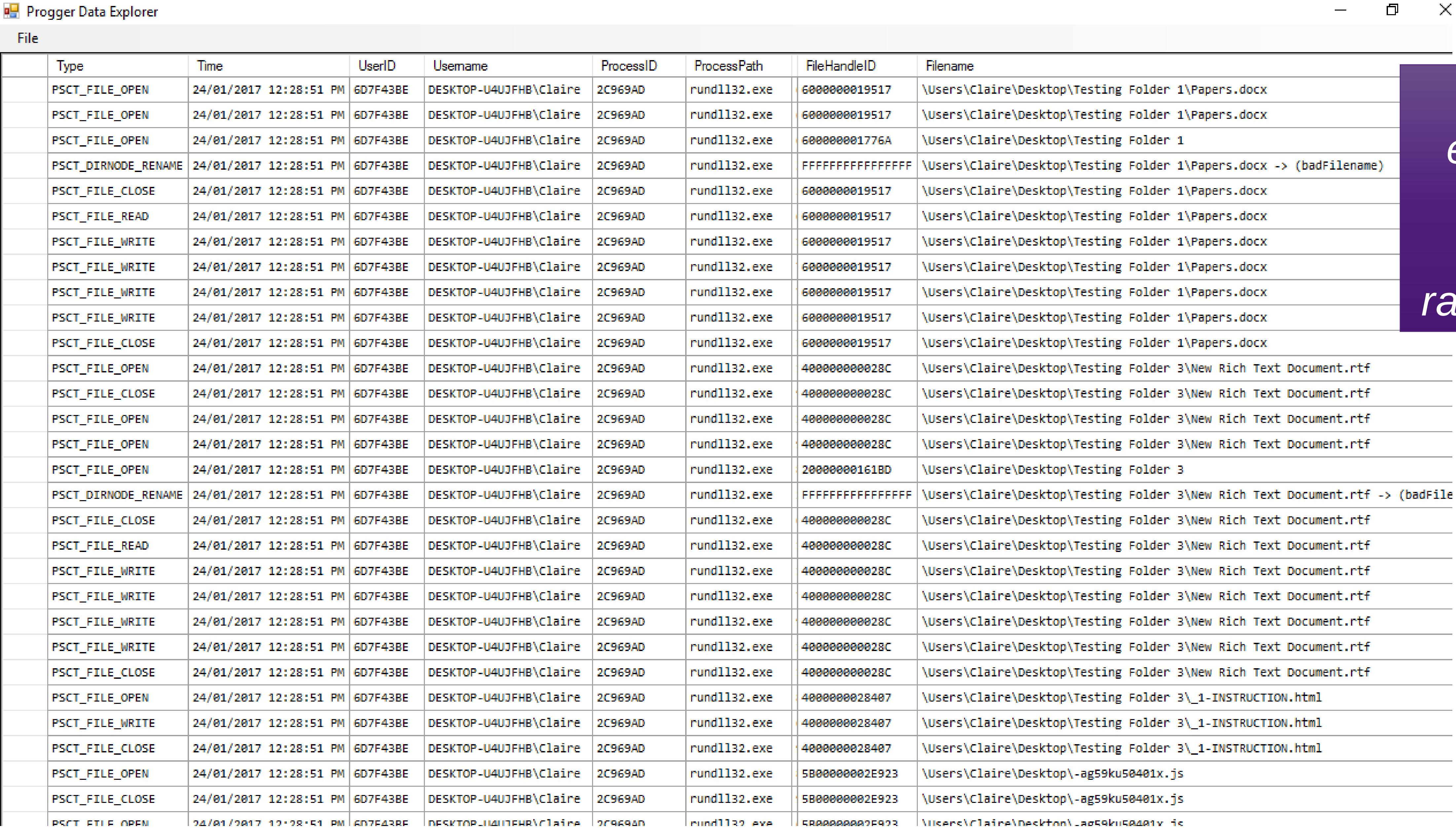
Uses of Progger: Locky Ransomware Provenance



	Type	Time	UserID	Username	ProcessID	ProcessPath	FileHandleID	Filename
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	20000000165DD	瀨
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\Device\Mup\.\.
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFF80F95BE119D	(null)
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	DESKTOP-U4UJFHB
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\Device\Mup\.\.\DESKTOP-U4UJFHB
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5000000000005	\
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5000000000005	\
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\Windows\SysWOW64\RPCSS.DLL
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5000000000005	\
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000AE6C	\Windows\SysWOW64\vssapi.dll
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	10000000015AC	\Windows\SysWOW64
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000AE6C	\Windows\SysWOW64\vssapi.dll
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	10000000015AC	\Windows\SysWOW64
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000AE6C	\Windows\SysWOW64\vssapi.dll
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	10000000015AC	\Windows\SysWOW64
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	10000000015AC	\Windows\SysWOW64
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\PerfLogs
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044A	\ProgramData
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044A	\ProgramData
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044A	\ProgramData
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044A	\ProgramData
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044A	\ProgramData
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044B	\ProgramData\Comms
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044C	\ProgramData\Microsoft
	PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	100000000044C	\ProgramData\Microsoft

Locky enumerating network shares and disabling backups.

Uses of Progger: Locky Ransomware Provenance



Type	Time	UserID	Username	ProcessID	ProcessPath	FileHandleID	Filename
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	600000001776A	\Users\Claire\Desktop\Testing Folder 1
PSCT_DIRNODE_RENAME	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\Users\Claire\Desktop\Testing Folder 1\Papers.docx -> (badFilename)
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_READ	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	6000000019517	\Users\Claire\Desktop\Testing Folder 1\Papers.docx
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	20000000161BD	\Users\Claire\Desktop\Testing Folder 3
PSCT_DIRNODE_RENAME	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	FFFFFFFFFFFFFFFF	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf -> (badFile
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_READ	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	400000000028C	\Users\Claire\Desktop\Testing Folder 3\New Rich Text Document.rtf
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	4000000028407	\Users\Claire\Desktop\Testing Folder 3_1-INSTRUCTION.html
PSCT_FILE_WRITE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	4000000028407	\Users\Claire\Desktop\Testing Folder 3_1-INSTRUCTION.html
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	4000000028407	\Users\Claire\Desktop\Testing Folder 3_1-INSTRUCTION.html
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5B00000002E923	\Users\Claire\Desktop\-ag59ku50401x.js
PSCT_FILE_CLOSE	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5B00000002E923	\Users\Claire\Desktop\-ag59ku50401x.js
PSCT_FILE_OPEN	24/01/2017 12:28:51 PM	6D7F43BE	DESKTOP-U4UJFHB\Claire	2C969AD	rundll32.exe	5B00000002E923	\Users\Claire\Desktop\-ag59ku50401x.js

Locky encrypting files and placing a ransom note.

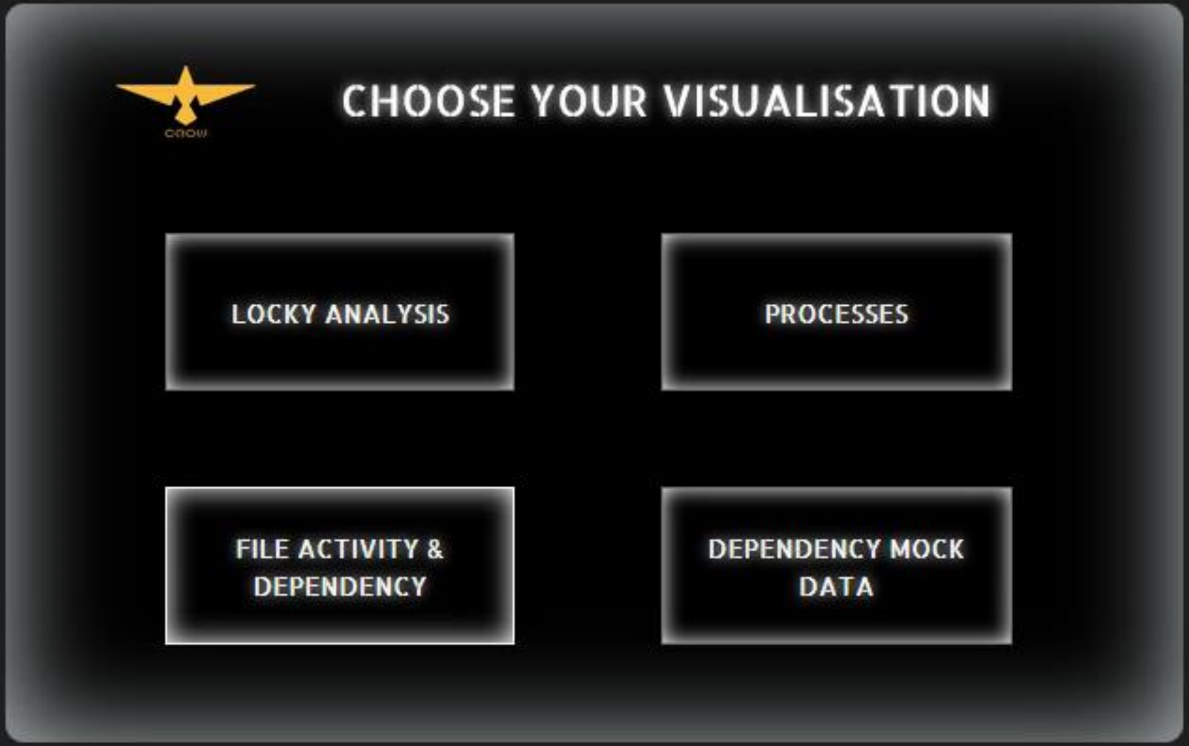
Badfilename = ?

VISUAL PROGGER

CROW

VISUAL PROGGER

CROW



Jeffery Garae, Ryan K. L. Ko, and Sivadon Chaisiri, "UVisP: User-centric Visualization of Data Provenance with Gestalt Principles", in Proceedings of the 15th IEEE TrustCom -16,Tianjin, China, 23-26 August, IEEE Computer Society, IEEE, 2016.

Why Is This Important?

New early-warning detection technologies → Proactive

- Based on data provenance

E.g. "Rate-Limiting" File Access

Limiting number of files a process/application can access at any given time

Useful in detecting and limiting damage done by "Crypt Locker" type malware (e.g. Locky)

Progger detects multiple file accesses by new/unknown process

Rate-limiting applied based on heuristics using application installation/process creation date

Overcoming the Lack of Datasets

"… we have only collected data for malicious insiders. We have not been able to collect similar data for 'good guys.' … but unfortunately no one has come up with a good way to collect the comprehensive types of data we have without violating employee privacy."

- *M. Cappelli et. al. (2012) The CERT Guide to Insider Threats.*

NZ Cyber Security Challenge

- *NZ Cyber Security Challenge 2015 (50 GB – 2+2 hours – 150 participants)*
- *NZ Cyber Security Challenge 2016 (24 GB – 2+2 hours; more focused)*
- *NZ Cyber Security Challenge 2017 (~20+GB)*
- *NZ Cyber Security Challenge 2018 (~20+GB)*

Real-life cloud provider: LayerX (hosts prominent NZ sites, e.g. realestate.co.nz)

Current work:

UQ's CRP Funding by Data61 → Massive Datasets for Cyber Security Research

Provenance Research Opportunities and Challenges

Capturing

What to capture, where to capture

Standardising the 'layers' of abstraction in the full provenance stack

Layers expose the gaps we still need to address..

Describing and formalizing provenance as an algebra

Storing

How do we store and where do we store (kernel?)

How do we handle distributed environments such as the cloud?

Integrity of provenance metadata (e.g. evidentiary requirements for justice)

Access and Analytics

Visualizing provenance

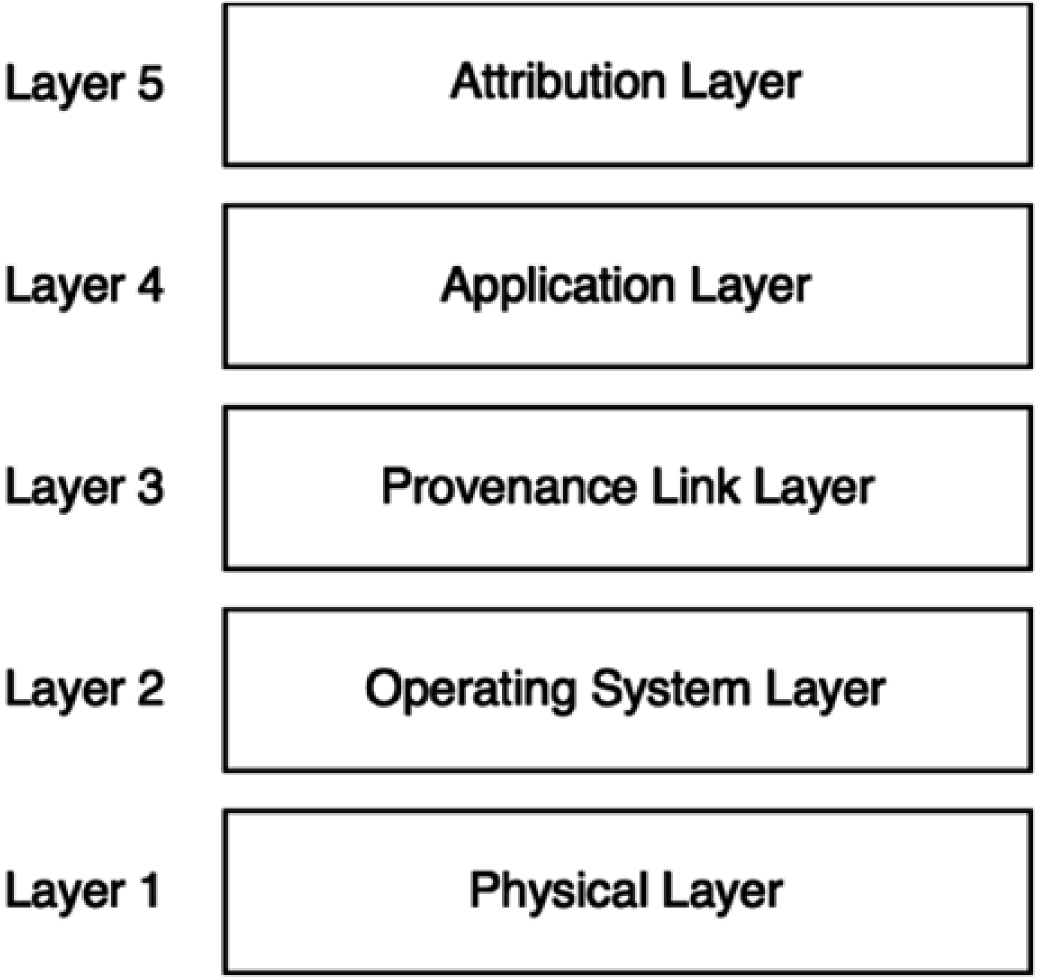
Fusion of provenance data

Inferring user behaviour from provenance → Behavioral biometrics

Using provenance for prevention and disruption/revocation → Proactive security

Automating for data governance

Full Provenance Stack: Five Layers for Complete and Meaningful Provenance (2017)



Layer 5 — Attribution Layer

Layer 4 — Application Layer

Layer 3 — Provenance Link Layer

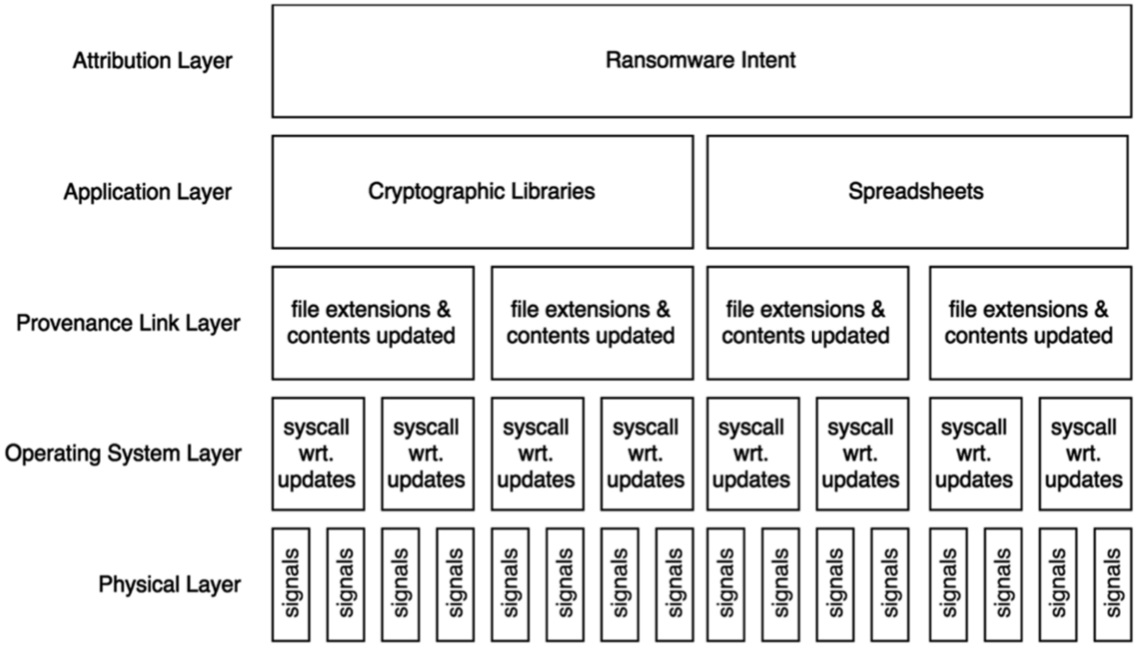
Layer 2 — Operating System Layer

Layer 1 — Physical Layer

Ref: Ko R.K.L., Phua T.W. (2017) The Full Provenance Stack: Five Layers for Complete and Meaningful Provenance. In: Wang G., Atiquzzaman M., Yan Z., Choo KK. (eds) Security, Privacy, and Anonymity in Computation, Communication, and Storage. SpaCCS 2017. Lecture Notes in Computer Science, vol 10658. Springer, Cham.

Using the provenance stack – extracting ransomware requirements

Layer	Application
Layer 5: Attribution	Detecting criminal sequences (e.g. fraud)
Layer 4: Application	Detecting software keyloggers Detecting malware
Layer 3: Provenance Link	Detecting unauthorized access Detecting data manipulation Detecting ransomware attacks
Layer 2: Operating System	Detecting row hammer attacks Detecting privilege escalation
Layer 1: Physical	Detecting BIOS rootkits Detecting hardware intrusion Detecting hardware keyloggers



Attribution Layer: Ransomware Intent

Application Layer: Cryptographic Libraries | Spreadsheets

Provenance Link Layer: file extensions & contents updated | file extensions & contents updated | file extensions & contents updated | file extensions & contents updated

Operating System Layer: syscall wrt. updates | syscall wrt. updates | syscall wrt. updates | syscall wrt. updates | syscall wrt. updates | syscall wrt. updates | syscall wrt. updates | syscall wrt. updates

Physical Layer: signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals | signals

Ransomware 5-Layered Data Provenance Encapsulation Example

Ref: Ko R.K.L., Phua T.W. (2017) The Full Provenance Stack: Five Layers for Complete and Meaningful Provenance. In: Wang G., Atiquzzaman M., Yan Z., Choo KK. (eds) Security, Privacy, and Anonymity in Computation, Communication, and Storage. SpaCCS 2017. Lecture Notes in Computer Science, vol 10658. Springer, Cham.

Using the provenance stack:

Understanding gaps and where current provenance solutions are

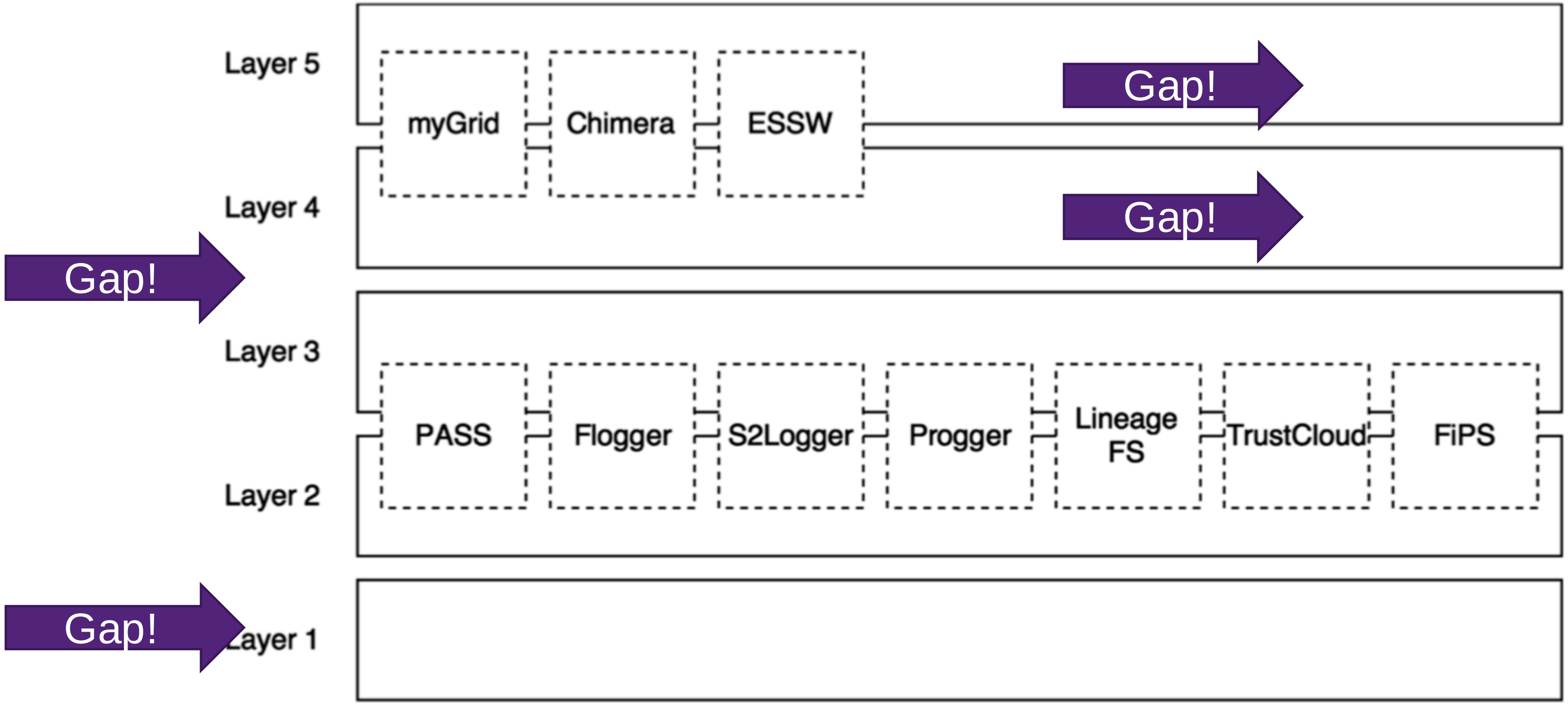


Fig. 6. How does existing systems fit?

Ref: Ko R.K.L., Phua T.W. (2017) The Full Provenance Stack: Five Layers for Complete and Meaningful Provenance. In: Wang G., Atiquzzaman M., Yan Z., Choo KK. (eds) Security, Privacy, and Anonymity in Computation, Communication, and Storage. SpaCCS 2017. Lecture Notes in Computer Science, vol 10658. Springer, Cham.

Standardising the Provenance Stack: ISO/IEC PWI 5181 - Data Provenance – Reference Model

ISO/IEC JTC 1/SC 27 WG 4

Editors: Ryan Ko, Brian Cole, Jan de Meer

Goal:

A clearly-defined universal reference model to

- Improve attribution and
- Promote the development and proliferation of protocols and standards

With an increasingly borderless, global environment (e.g. cloud, IoT, devices, supply chains, blockchains, etc.) for data transfer, storage and communications, there is a need to standardize the way provenance is described and classified.

Recap: Provenance Research Opportunities and Challenges

Capturing

- What to capture, where to capture
- Describing and formalizing provenance as an algebra
- Standardising the ‘layers’ of abstraction in the full provenance stack

Storing

- How do we store and where do we store (kernel?)
- How do we handle distributed environments such as the cloud?
- Integrity of provenance metadata (e.g. evidentiary requirements for justice)

Access and Analytics

- Visualizing provenance
- Fusion of provenance data
- Inferring user behaviour from provenance (e.g. user actions; Behavioral biometrics)
- Using provenance for prevention and disruption/revocation (e.g. Proactive security)
- Automating compliance for data governance

Related Provenance Publications For Seminar (1/3)

- Jeffery Garae, **Ryan K. L. Ko**, and Sivadon Chaisiri, "**UVisP: User-centric Visualization of Data Provenance with Gestalt Principles**", in 1st International Workshop on Security and Privacy in Advanced Persistent Threat, in Proceedings of the 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-16), Tianjin, China, 23-26 August, IEEE Computer Society, IEEE, 2016.
- **Ryan K. L. Ko**, Giovanni Russello, Richard Nelson, Shaoning Pang, Aloysius Cheang, Gill Dobbie, Abdolhossein Sarrafzadeh, Sivadon Chaisiri, Muhammad Rizwan Asghar, and Geoffrey Holmes, "**STRATUS: Towards Returning Data Control to Cloud Users**", 6th International Workshop on Trust, Security and Privacy for Big Data (TrustData 2015), held in conjunction with 15th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2015), Zhangjiajie, China, Springer, November 18-20, 2015
- Mohammad Bany Taha, Sivadon Chaisiri, **Ryan K. L. Ko**, **"Trusted Tamper-Evident Data Provenance"** 2015 IEEE International Symposium on Recent Advances of Trust, Security and Privacy in Computing and Communications (RATSP 2015), held in conjunction with 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15), Helsinki, Finland, 20-22 August, 2015.
- Xin Li, Chaitanya Joshi, Yu Shyang Tan, and **Ryan K. L. Ko**, **"Inferring User Actions from Provenance Logs"** 2015 IEEE International Symposium on Recent Advances of Trust, Security and Privacy in Computing and Communications (RATSP 2015), held in conjunction with 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom-15), Helsinki, Finland, 20-22 August, 2015.
- **Ryan K. L. Ko,** Mark A. Will, "**Progger: An Efficient, Tamper-Evident Kernel-Space Logger for Cloud Data Provenance Tracking,**", IEEE Int'l Conference on Cloud Computing (IEEE CLOUD 2014), Alaska, USA, **2014.**

Related Provenance Publications For Seminar (2/3)

- Y. S. Tan, **Ryan K. L. Ko**, Geoff Holmes, “**Security and Data Accountability in Distributed Systems: A Provenance Survey**”, 15th IEEE International Conference on High Performance Computing and Communications (HPCC 2013), Zhangjiajie, China, November 13-15, **2013**, pp. 1-8, IEEE.
- Chun Hui Suen, **Ryan K. L. Ko**, Y. S. Tan, Peter Jagadpramana and B. S. Lee (2013) “**S2Logger: End-to-End Data Tracking Mechanism for Cloud Data Provenance**.” Proc 12th IEEE International Conference of Trust, Security and Privacy in Computing and Communication (TrustCom ’13), Melbourne, Australia, 16-18 July **2013**, pp. 594-602. IEEE.
- Olive Qing Zhang, **Ryan K L Ko**, Markus Kirchberg, Chun Hui Suen, Peter Jagadpramana, Bu Sung Lee, “**How to Track Your Data: Rule-Based Data Provenance Tracing Algorithms**”, 2012 International Symposium on Advances in Trusted and Secure Information Systems (TSIS-2012), in conjunction with 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2012), 25-27 June **2012,** Liverpool, pp.1429-1437.
- Yu Shyang Tan, **Ryan K L Ko**, Peter Jagadpramana, Chun Hui Suen, Markus Kirchberg, Teck Hooi Lim, Bu Sung Lee, Anurag Singla, Ken Mermoud, Doron Keller, Ha Duc, “**Tracking of Data Leaving the Cloud**”, 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2012), 25-27 June **2012**, Liverpool, UK, pp. 137-144.
- Nick Papanikolaou, Siani Pearson, Marco Cassasa Mont, **Ryan K L Ko**, “**Automating Compliance For Cloud Computing Services**”, Security Governance and SLAs in Cloud Computing – CloudSecGov 2012, in conjunction with 2nd International Conference on Cloud Computing and Services Science (CLOSER 2012), 19 April **2012**, Porto, Portugal, pp. 631-637.

Related Provenance Publications For Seminar (3/3)

- **Ryan K. L. Ko**, Peter Jagadpramana, Bu Sung Lee, "**Flogger: A File-centric Logger for Monitoring File Access and Transfers within Cloud Computing Environments**", 3rd IEEE International Workshop on Security in e-Science and e-Research (ISSR 2011), in conjunction with IEEE TrustCom, 16-18 November **2011**, Changsha, China.
- Olive Qing Zhang, Markus Kirchberg, **Ryan K. L. Ko**, Bu Sung Lee, "**How To Track Your Data: The Case for Cloud Computing Provenance**", 3rd IEEE International Conference on Cloud Computing Technology and Science (IEEE CloudCom 2011), 29 Nov – 1 Dec **2011**, Athens, Greece, pp. 446-453.
- **Ryan K. L. Ko**, Markus Kirchberg, Bu Sung Lee, "**From System-Centric Logging to Data-Centric Logging – Accountability, Trust and Security in Cloud Computing**", Defence, Science and Research Conference 2011 – Symposium on Cyber Terrorism, IEEE Computer Society, 3-4 August **2011**, Singapore, pp. 1-4.
- **Ryan K. L. Ko**, Bu Sung Lee, Siani Pearson, "**Towards Achieving Accountability, Auditability and Trust in Cloud Computing**", International Workshop on Cloud Computing- Architecture, Algorithms and Applications (Cloudcomp 2011), Springer Verlag, 22-24 July **2011**, Kochi, India, pp. 5-18.
- **Ryan K. L. Ko**, Peter Jagadpramana, Miranda Mowbray, Siani Pearson, Markus Kirchberg, Qianhui Liang, Bu Sung Lee, "**TrustCloud – A Framework for Accountability and Trust in Cloud Computing**", IEEE 2nd Cloud Forum for Practitioners (IEEE ICFP 2011), IEEE Computer Society, 7-8 July **2011**, Washington DC, USA, pp. 1-5.
- (**Best Paper Award**) Markus Kirchberg, **Ryan K. L. Ko**, Bu Sung Lee, "**From Linked Data to Relevant Data – Time is the Essence**", 1st International Workshop on Usage Analysis and the Web of Data (USEWOD 2011), in conjunction with the 20th Int'l World Wide Web Conference (**WWW 2011**), India, March 28 **2011**.

Other References

[1] R.K.L. Ko, P. Jagadpramana, M. Mowbray, S. Pearson, M. Kirchberg, Q. Liang and B.S. Lee, "TrustCloud - A Framework for Accountability and Trust in Cloud Computing," *Proc. IEEE 2nd Cloud Forum for Practitioners (IEEE ICFP 2011)*, IEEE Computer Society, 2011, pp. 1-5.

[2] R.K.L. Ko, B.S. Lee and S. Pearson, "Towards Achieving Accountability, Auditability and Trust in Cloud Computing," *Proc. International workshop on Cloud Computing: Architecture, Algorithms and Applications (CloudComp2011)*, Springer, 2011, pp. 5-18.

[3] A.J.G. Hey, S. Tansley and K.M. Tolle, *The fourth paradigm: data- intensive scientific discovery*, Microsoft Research Redmond, WA, 2009.

[4] S. Sakr, A. Liu, D. Batista and M. Alomari, "A Survey of Large Scale Data Management Approaches in Cloud Environments," *Communications Surveys & Tutorials, IEEE*, no. 99, pp. 1-26.

[5] R. Love, "Kernel Korner: Intro to iNotify," *Linux Journal*, vol. 2005, no. 139, 2005, pp. 8.

[6] S.E. Hansen and E.T. Atkins, "Automated system monitoring and notification with swatch," USENIX Association's Proceedings of the Seventh Systems Administration (LISA VII) Conference, 1993.

[7] Silicon Graphics International Corp, "File Alteration Monitor (FAM) Overview," 2009; http://oss.sgi.com/projects/fam/.

[8] M. Roesch, "Snort-lightweight intrusion detection for networks," *Proc. 13th Large Installation System Administration Conference (LISA)*, 1999, pp. 229–238.

[9] HyTrust, "HyTrust Appliance," 2010; http://www.hytrust.com/product/overview/.

[10] Fujitsu Research Institute, "Personal data in the cloud: A global survey of consumer attitudes," 2010; http://www.fujitsu.com/downloads/SOL/fai/reports/fujitsu_personal-data- in-the-cloud.pdf.

[11] VMWare Hyperic, "Performance Monitoring for Cloud Services," 2011; http://www.hyperic.com/products/cloud-status-monitoring.

[12] CloudKick, "CloudKick - Cloud Monitoring and Management," 2011; https://www.cloudkick.com/.

[13] Cloud Security Alliance, "Security Guidance for Critical Areas of Focus in Cloud Computing V2.16," 2009; https://cloudsecurityalliance.org/csaguide.pdf.

[14] M Roesch, Snort-lightweight intrusion detection for networks, USENIX LISA 1999.

[15] Olive Qing Zhang, Ryan K L Ko, Markus Kirchberg, Chun Hui Suen, Peter Jagadpramana, Bu Sung Lee, "**How to Track Your Data: Rule-Based Data Provenance Tracing Algorithms**", 2012 International Symposium on Advances in Trusted and Secure Information Systems (TSIS-2012), in conjunction with 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2012), , 25-27 June 2012, Liverpool, UK.

[16] Yu Shyang Tan, Ryan K L Ko, Peter Jagadpramana, Chun Hui Suen, Markus Kirchberg, Teck Hooi Lim, Bu Sung Lee, Anurag Singla, Ken Mermoud, Doron Keller, Ha Duc, "**Tracking of Data Leaving the Cloud**", 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2012), 25-27 June 2012, Liverpool, UK.

[17] K.-K. Muniswamy-Reddy, P. Macko, and M. Seltzer, "Provenance for the cloud," in Proc of the 8th USENIX Conf on File and Storage Technologies.USENIX Association, 2010, pp. 197–210.

[18] K.-K. Muniswamy-Reddy, P. Macko, and M. Seltzer, "Making a cloud provenance-aware," in 1st Workshop on the Theory and Practice of Provenance, 2009.

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Thank you

Prof Ryan Ko | Chair and Director
UQ Cyber Security
The University of Queensland
E: ryan.ko @ uq.edu.au
W: https://researchers.uq.edu.au/researcher/23784

@ryan_kl_ko

https://www.linkedin.com/in/ryan-k-l-ko-38894824/

Thanks to Surya Nepal and Arimdam Pal of Data61 for inviting me to give this talk!

Join us at UQ Cyber Security!

Open positions for PhD, Postdocs and continuing academic positions: https://bit.ly/39uemUP