Scalable Data Analytics Pipeline for Validation of Real-Time Attack Detection

Eric Badger, Phuong Cao, Alex Withers, Adam Slagell, Zbigniew Kalbarczyk, Ravi Iyer
University of Illinois Urbana-Champaign

Overview

- Introduction/Motivation
- Challenges
- Attack Detection: AttackTagger
- Validation of AttackTagger
- Future Work/Conclusion

Research Problems

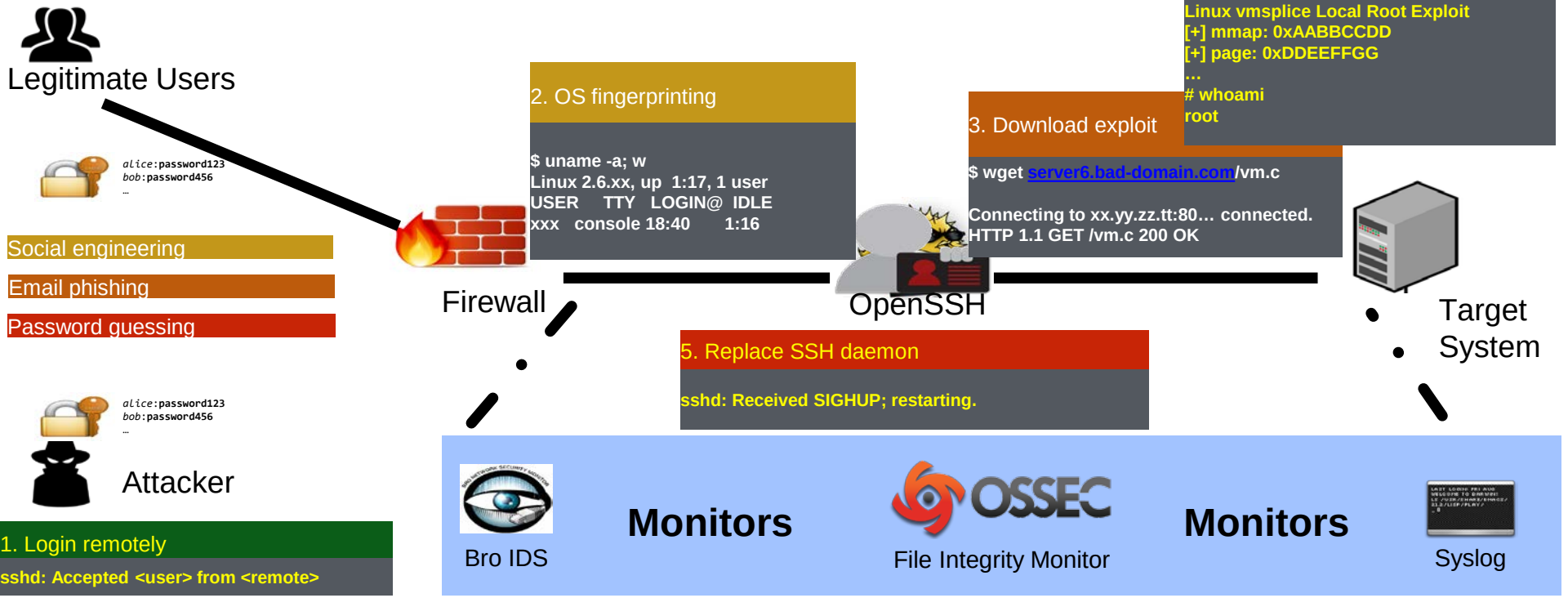
- How can we detect attacks before system misuse?
  High-accuracy, real-time attack detection tools
- How do we validate that our attack detection tools works on real-world data?
- How do we transition attack detection tools from theory to practice?

Attack Type: Credential-Stealing Attacks

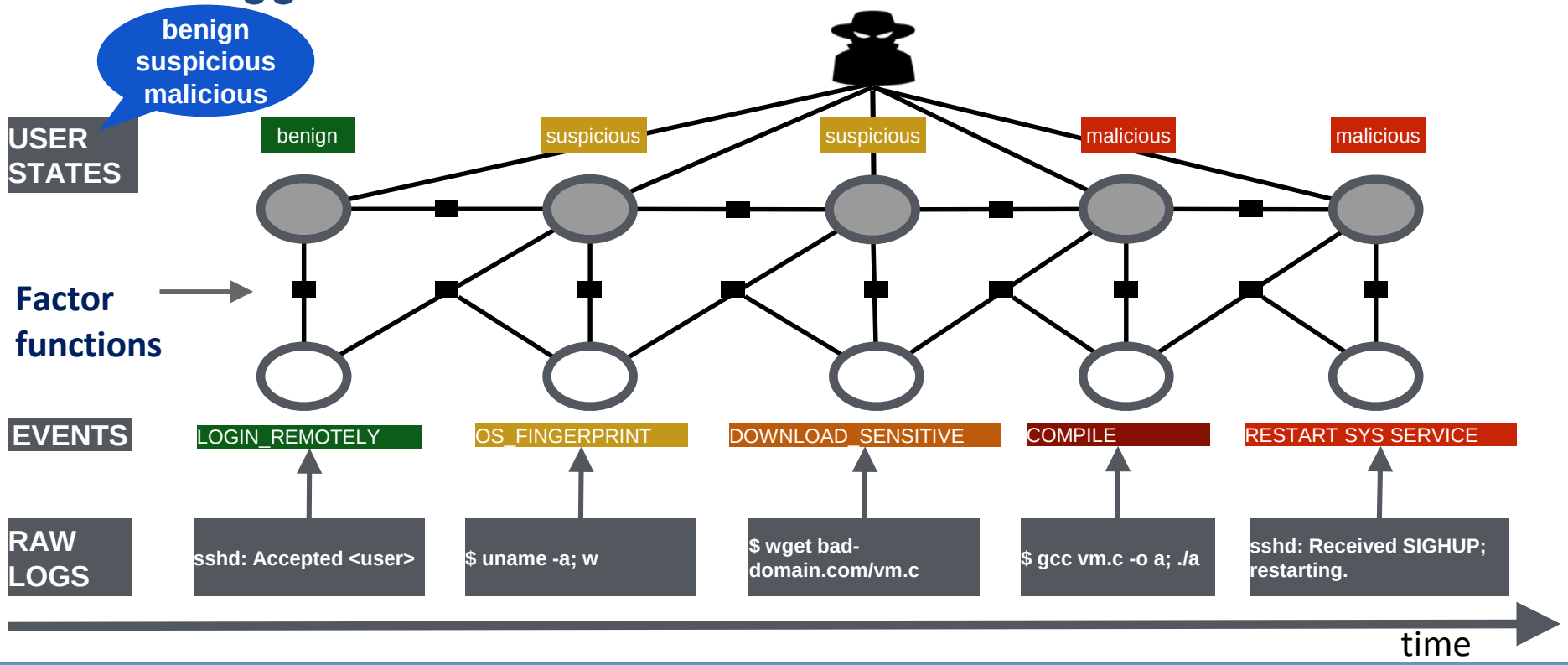
- Definition: An attack where the attacker enters the system with legitimate credentials (e.g. username/password)
  Attacker becomes an insider
- 26% (32/124) of incidents at NCSA over a 5-year period were credential-stealing attacks
- 28% (9/32) of these attacks weren't detected by NCSA monitors

[1] Sharma, A.; Kalbarczyk, Z.; Barlow, J.; Iyer, R., "Analysis of security data from a large computing organization," in *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*

Example Credential-Stealing Attack

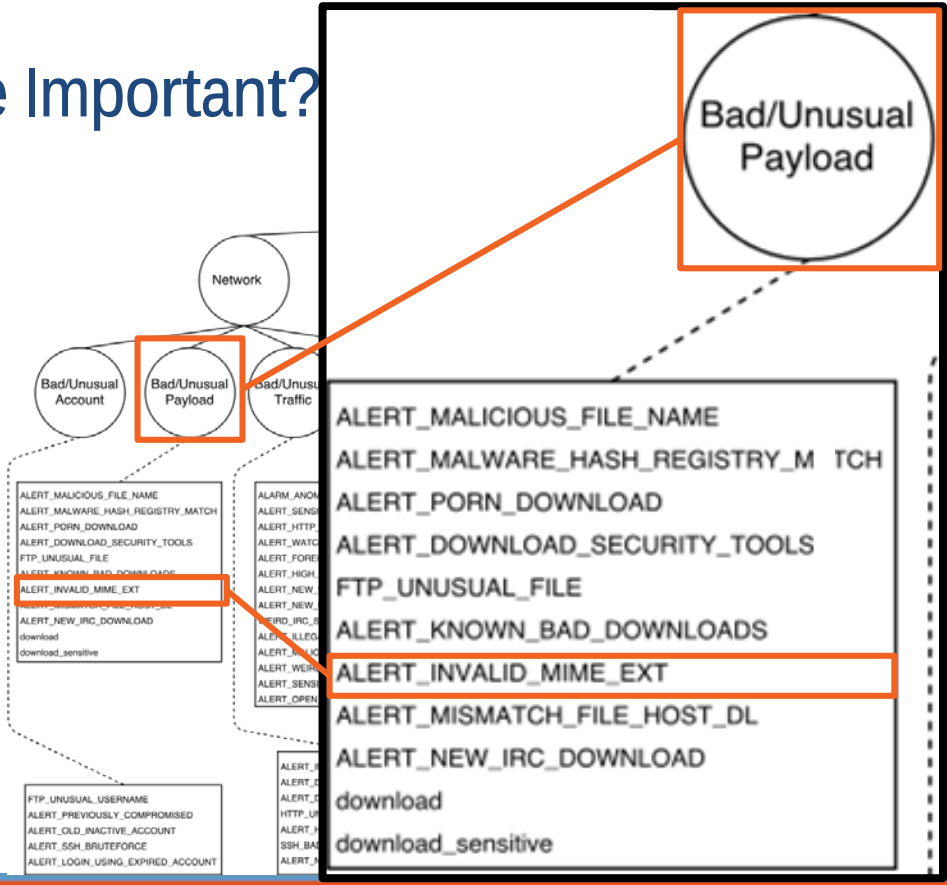


Detecting Attacks Using Factor Graphs: AttackTagger



How Do I Know What Events Are Important?

- We identified over 100 important events related to credential-stealing attacks



[2] P. Cao, K. Chung, Z.Kalbarczyk, R. Iyer, and A. Slagell. Preemptive intrusion detection. HotSoS '14.

[3] P. Cao, E. Badger, Z. Kalbarczyk, R. Iyer, and Adam Slagell. Preemptive intrusion detection: theoretical framework and real-world measurements. HotSoS '15.

AttackTagger Dataset

- Manually extracted data
  - Raw logs
  - Human-written incident reports
- Ideal data
  - No noise
  - Perfect monitors
  - No randomness

[3] P. Cao, E. Badger, Z. Kalbarczyk, R. Iyer, and Adam Slagell. Preemptive intrusion detection: theoretical framework and real-world measurements. HotSoS '15.

Raw logs

11:00:57 sshd: Failed password for root
23:08:26 sshd: Failed password for root
23:08:30 sshd: Failed password for nobody
23:08:38 sshd: Failed password for <user>
23:08:42 sshd: Failed password for root

ALERT_FAILED_PASSWORD
ALERT_FAILED_PASSWORD
ALERT_FAILED_PASSWORD
ALERT_FAILED_PASSWORD
ALERT_FAILED_PASSWORD

Manual Extraction

Human-written incident reports

The security team received ssh suspicious alerts from <machine> for the user <user>. There were also some Bro alerts from the machine <machine>. From the Bro sshd logs the user ran the following commands:

uname -a
unset HISTFILE
wget <xx.yy.zz.tt>/abs.c -O a.c
gcc a.c -o a;

READ_HOST_CONFIGURATION
ALERT_DISABLE_LOGGING
ALERT_DOWNLOAD_SENSITIVE
ALERT_COMPILE_CODE

[3] P. Cao, E. Badger, Z. Kalbarczyk, R. Iyer, and Adam Slagell. Preemptive intrusion detection: theoretical framework and real-world measurements. HotSoS '15.

AttackTagger Results

- 74.2% (46/62) malicious users correctly detected as malicious
- 1.52% (19/1,253) benign users incorrectly detected as malicious

Name	*TP*	*TN*	*FP*	*FN*
AttackTagger	*74.2*	*98.5*	*1.5*	*25.8*
Rule Classifier	9.8	96.0	4.0	90.2
Decision Tree	21.0	100.00	0.00	79.0
Support Vector Machine	27.4	100.00	0.00	72.6

[3] P. Cao, E. Badger, Z. Kalbarczyk, R. Iyer, and Adam Slagell. Preemptive intrusion detection: theoretical framework and real-world measurements. HotSoS '15.

How to Extract Important Events

- **Network Monitors**
  - Anything that logs activity between hosts
  - Example: Bro

- **Host Monitors**
  - Anything that logs activity on the host
  - Example: OSSEC

Log Normalization

Auth Logs

```
118 2015-09-29T16:54:39.425503-05:00 whitacre sudo: pam_unix(sudo:session): sess
    ion opened for user root by eric(uid=1000)
119 2015-09-29T16:54:39.499108-05:00 whitacre su[10044]: Successful su for root
    by root
120 2015-09-29T16:54:39.499118-05:00 whitacre su[10044]: + /dev/pts/6 root:root
121 2015-09-29T16:54:39.499296-05:00 whitacre su[10044]: pam_unix(su:session): s
    ession opened for user root by eric(uid=0)
122 2015-09-29T17:00                                    pam_unix(cron:session
    ): session opene
123 2015-09-29T17:0                                     pam_unix(cron:session
    ): session closed for user root
```

1443126106.661021

Bro Notice Logs

```
1 #separator \x09
2 #set_separator  ,
3 #empty_field  (empty)
4 #unset_field  -
5 #path notice
6 #open 2015-09-24-15-20-50
7 #fields  ts  uid id.orig_h id.orig_p id.resp_h id.resp_p fuid  file_mime_type
    file_desc proto note  msg sub src dst p n peer_descr  actions suppress_for
    dropped remote_location.country_code  remote_location.region  remote_locat
  ion.city  remote_location.latitude  remote_location.longitude
8 #types  time  string  addr  port  addr  port  string  string  string  enum
  enum  string  string  addr  addr  port  count string  set[enum] interval  bo
  ol  string  string  string  double  double
9 1443126106.661021 Ceu4Bv3Kd6Pr6MiUa6  130.126.137.23  58190 54.231.133.204
  443 - - - tcp SSL::Invalid_Server_Cert  SSL certificate validation failed wi
  th (unable to get local issuer certificate) CN=*.s3-eu-west-1.amazonaws.com,
  OU=S3-B,O=Amazon.com\\, Inc.,L=Seattle,ST=Washington,C=US 130.126.137.23  54
  .231.133.204  443 - bro Notice::ACTION_LOG  3600.000000 F - - - - -
```

OSSEC Logs

```
3763 ** Alert 1443563681.247829: - syslog,sudo
3764 2015 Sep 29 16:54:41 whitacre->/var/log/auth.log
3765 Rule: 5402 (level 3) -> 'Successful sudo to ROOT executed'
3766 User: eric
3767 2015-09-29T16:54:39.425299-05:00 whitacre sudo:     eric : TTY=pts/6 ; PWD=
     /var/log ; USER=root ; COMMAND=/bin/su
3768
3769 ** Alert 1443563681.248105: mail  - local,syslog,
3770 2015 Sep 29 16:54:41 whitacre->/var/log/auth.log
3771 Rule: 105501 (level 11) -> 'User successfully changed UID to root.'
3772 User: eric
```

2015-09-29T08:00:06.257580-05:00

RKHunter Logs

```
5582 [16:02:07] System checks summary
5583 [16:02:07] ====================
5584 [16:02:07]
5585 [16:02:07] File properties checks...
5586 [16:02:07] Files checked: 142
5587 [16:02:07] Suspect files: 2
5588 [16:02:07]
5589 [16:02:07] Rootkit checks...
5590 [16:02:07] Rootkits checked : 380
5591 [16:02:07] Possible rootkits: 0
```

Snoopy Logs

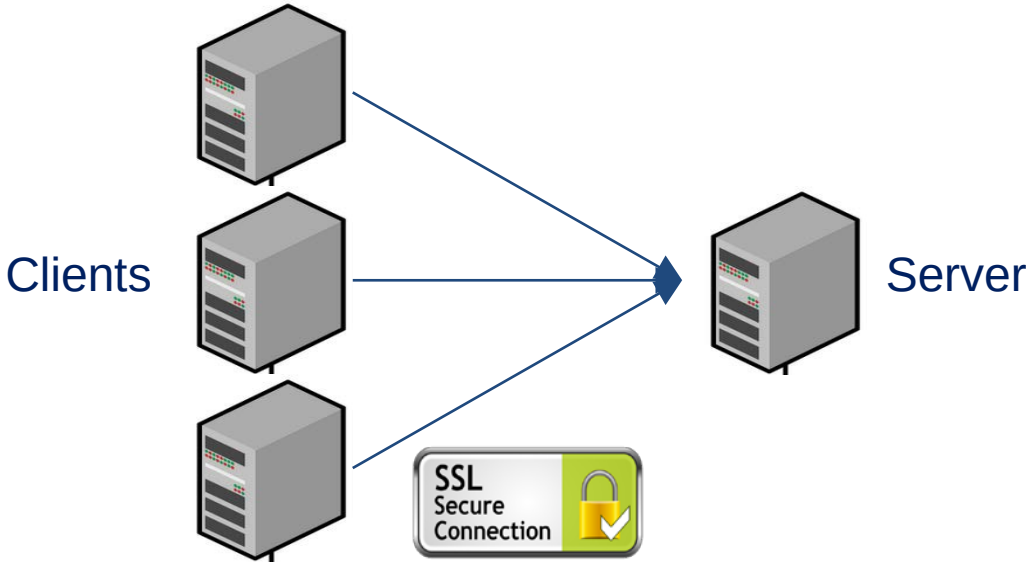
```
4141 2015-09-29T08:00:06.252345-05:00 whitacre snoopy[32190]: [username:root t
     ty_username:(none) uid:0 sid:26590 tty:(none) cwd:/root filename:/bin/una
     me]: uname -r
4142 2015-09-29T08:00:06.254930-05:00 whitacre snoopy[32194]: [username:root t
     ty_username:(none) uid:0 sid:26590 tty:(none) cwd:/root filename:/bin/gre
     p]: grep ^-
4143 2015-09-29T08:00:06.257580-05:00 whitacre snoopy[32197]: [username:root t
     ty_username:(none) uid:0 sid:26590 tty:(none) cwd:/root filename:/bin/egr
     ep]: egrep (^|[^\])[][?*{}]
```

Log Normalization

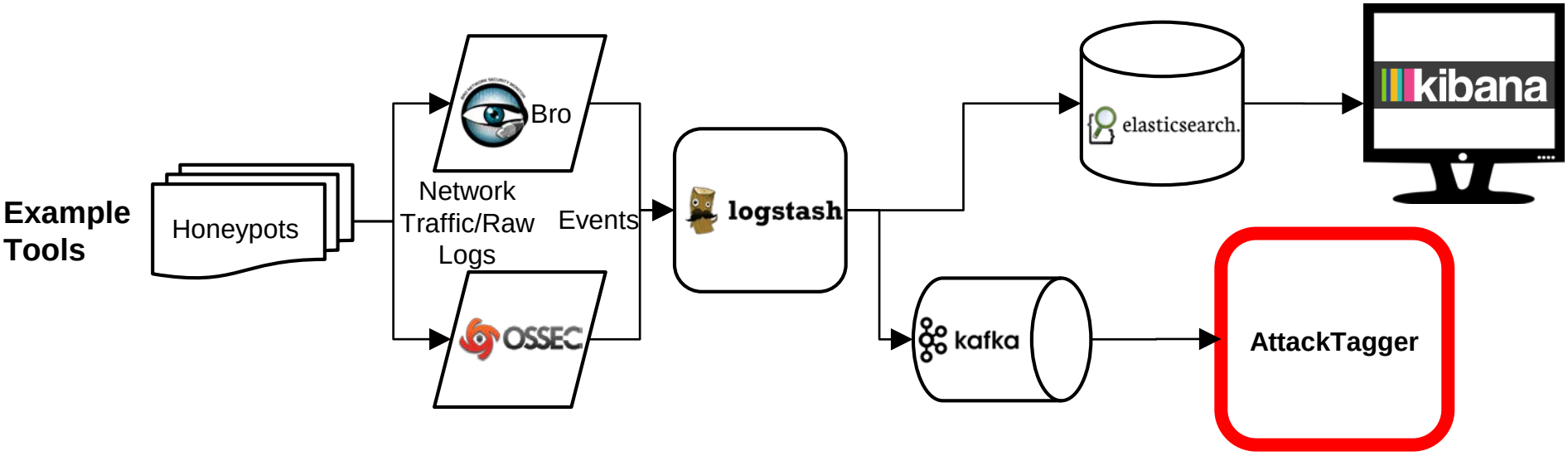
	Timestamp,	IP Address:User,	Event	Extra Info,		Received Timestamp
209	1443457145.717	143.219.0.11:root	ALERT_FAILED_PASSWORD	NaN	NaN	1443461619.483
210	1443457147.510	143.219.0.11:root	ALERT_FAILED_PASSWORD	NaN	NaN	1443461619.490
211	1443457661.505	143.219.0.11:LOGIN	login	NaN	NaN	1443461619.516
212	1443457661.469	143.219.0.11:root	read_host_configuration	NaN	NaN	1443461619.520
213	1443457662.963	143.219.0.11:ubuntu	ALERT_GET_LOGGEDIN_USERS	NaN	NaN	1443461619.536
214	1443461754.305	:,	ALERT_INTERNAL_ADDRESS_SCAN	NaN	NaN	1443461764.436

Log Aggregation

- Multiple clients, single server
- Encryption is necessary
  - Thwart MITM attacks

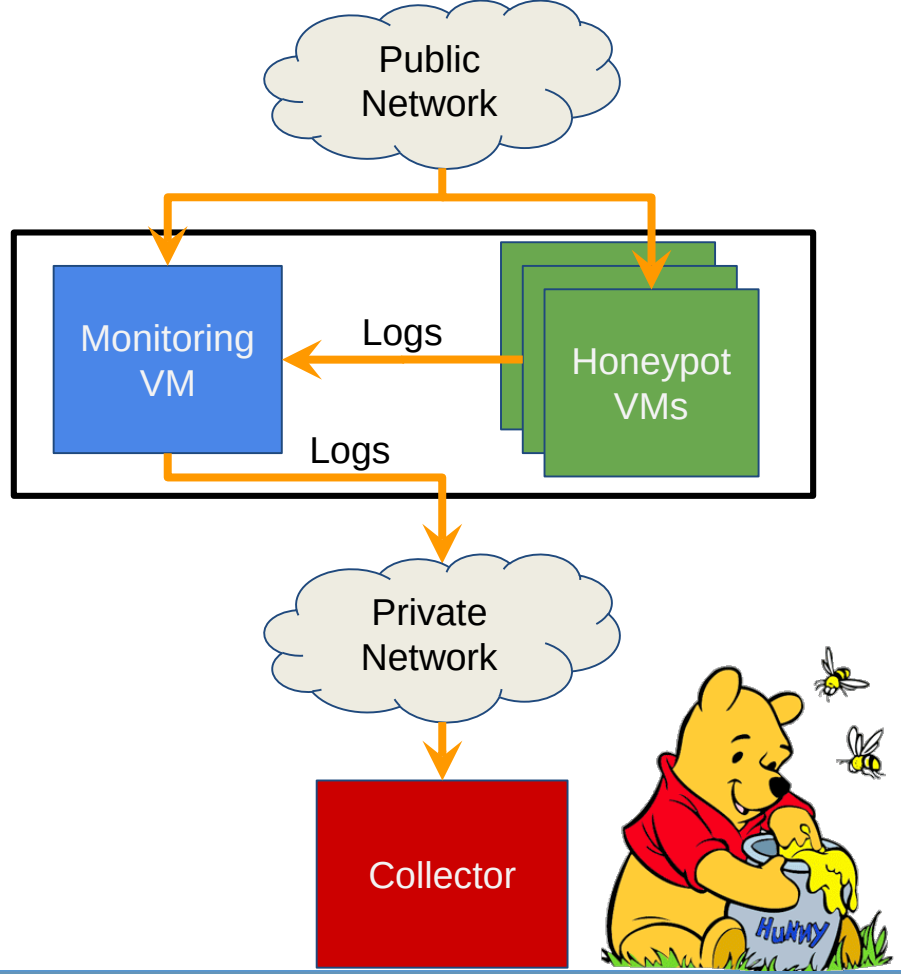


Data Pipeline Design



We Need Data!
Honeypots at NCSA

- NCSA server running several VMs
  - Honeypot VMs
  - Monitoring VM
- Collector (NCSA server)
  - Normalize, aggregate, queue, detect
- Honeypots are low-risk



Preliminary Honeypot Results

- 3 SSH Bruteforce attacks in first 3 days
- Downloaded and ran “/tmp/squid64”
- Attackers beat my monitors! (Well, sort of...)
  - Pushed the malware
  - Immediate file deletion

Where Are We Now?

- Honeypots are online
  - Mining attack data
- Creating targeted attacks
- Upgrading AttackTagger factor functions
- Pipeline performance evaluation underway

Validating AttackTagger in a Real-world Environment

- Compare with theoretical AttackTagger results
- Compare and contrast AttackTagger with different attack detection models

  e.g. Rule-classifier, Bayesian Networks
- Benchmark throughput of events

  Can AttackTagger work in real-time?

Future Work

- Validate AttackTagger using honeypots/pipeline
- Transition entire pipeline into practice at NCSA
- Add additional monitors to data pipeline

  Administrator-generated events/profiles

  Keystroke data (e.g. iSSHD)
- Improve stream-processing of AttackTagger

Conclusion

- Demonstrated attack detection using factor graphs (AttackTagger)
  74.2% true positive
- Designed and implemented data pipeline for real-world validation of attack detection tools

Questions?

Citations

[1] Sharma, A.; Kalbarczyk, Z.; Barlow, J.; Iyer, R., "Analysis of security data from a large computing organization," in *Dependable Systems & Networks (DSN), 2011 IEEE/IFIP 41st International Conference on*

[2] Phuong Cao, Key-whan Chung, Zbigniew Kalbarczyk, Ravishankar Iyer, and Adam J. Slagell. 2014. Preemptive intrusion detection. In *Proceedings of the 2014 Symposium and Bootcamp on the Science of Security* (HotSoS '14). ACM, New York, NY, USA, , Article 21 , 2 pages. DOI=10.1145/2600176.2600197 http://doi.acm.org/10.1145/2600176.2600197

[3] Phuong Cao, Eric Badger, Zbigniew Kalbarczyk, Ravishankar Iyer, and Adam Slagell. 2015. Preemptive intrusion detection: theoretical framework and real-world measurements. In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security* (HotSoS '15). ACM, New York, NY, USA, , Article 5 , 12 pages. DOI=10.1145/2746194.2746199 http://doi.acm.org/10.1145/2746194.2746199

The Honey Pot and The Honey Badger